

**SABER EXPO BI 5: UNA HERRAMIENTA DE INTELIGENCIA DE
NEGOCIOS WEB PARA LA GESTIÓN DE INDICADORES DE LAS
PRUEBAS SABER 5 EN LAS INSTITUCIONES EDUCATIVAS DE LA
SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO**

CEBALLOS ROSERO SEGUNDO DARIO

SANTANDER ARIAS DUBER ALEXANDER

**PROGRAMA DE INGENIERÍA DE SISTEMAS
DEPARTAMENTO DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE NARIÑO
ABRIL, 2022**

**SABER EXPO BI 5: UNA HERRAMIENTA DE INTELIGENCIA DE
NEGOCIOS WEB PARA LA GESTIÓN DE INDICADORES DE LAS
PRUEBAS SABER 5 EN LAS INSTITUCIONES EDUCATIVAS DE LA
SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO**

Autores

CEBALLOS ROSERO SEGUNDO DARIO, daridr8761@gmail.com

SANTANDER ARIAS DUBER ALEXANDER, dasa.deidara@hotmail.com

Informe final de trabajo de grado presentado como requisito parcial para optar al título de
Ingeniero de Sistemas, en modalidad investigación.

Director

PhD. SILVIO RICARDO TIMARÁN PEREIRA SILVIO RICARDO

**PROGRAMA DE INGENIERÍA DE SISTEMAS
DEPARTAMENTO DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE NARIÑO
ABRIL, 2022**

NOTA DE RESPONSABILIDAD

“Las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva del autor.”

Artículo 1ro del Acuerdo No. 324 de octubre 11 de 1966 emanado del Honorable Consejo Directivo de la Universidad de Nariño.

Nota de Aceptación

Firma presidente del jurado

Firma del jurado

Firma del asesor

Firma del co-asesor

RESUMEN

En este trabajo de grado se desarrolló una herramienta web libre de Inteligencia de Negocios, para la gestión de indicadores de las pruebas Saber 5, que presentaron los estudiantes de las instituciones educativas de la subregión de Obando en el departamento de Nariño.

La aplicación está diseñada para los estudiantes que presentan las pruebas Saber, las cuales son evaluaciones externas estandarizadas aplicadas por el Instituto Colombiano para la Evaluación de la Educación -ICFES-, las cuales evalúan el desempeño alcanzado por los estudiantes según las competencias básicas definidas por el Ministerio de Educación Nacional [1].

La arquitectura de Saber Expo BI 5, está compuesta por un módulo Extract, Transform and Load (ETL), que permite extraer la información de la bases de datos del ICFES, aplicarles filtros de limpieza y transformación y cargarlos a un mercado de datos, diseñado bajo el modelo multidimensional en estrella; un módulo de construcción de cubos que permite organizar los datos en dimensiones; un módulo de visualización de resultados para el análisis de la información y un módulo de reportes para la producción de informes por diferentes dimensiones.

Para la construcción de esta herramienta, se siguieron las diferentes etapas de la metodología ágil SCRUM, bajo un modelo de arquitectura Model Template View (MTV), utilizando Django como framework y Python como lenguaje principal, elegido por ser un lenguaje flexible y diseñado para ser fácil de leer [2].

Para el tratamiento de datos dentro de la aplicación fue necesario implementar Pandas, una potente librería la manipulación y análisis de datos logrando analizar a profundidad los detalles del mercado de datos [3].

En el proceso de visualización de datos se utilizó Charts.js, una librería de JavaScript, utilizando el canvas de HTML5, mostrando gráficos indispensables para un sistema de inteligencia de negocios, ágil, y dinámico, mejorando la interacción con el usuario [4].

ABSTRACT

In this degree work, a free Business Intelligence web tool was developed for managing the indicators of the Saber 5 tests, presented by students from educational institutions in the Obando subregion in the department of Nariño.

The application is designed for students who take the Saber tests, which are standardized external evaluations applied by the Colombian Institute for the Evaluation of Education -ICFES-, which evaluate the performance achieved by students according to the basic competencies defined by the Ministry of National Education.

The architecture of Saber Expo BI 5 is composed of an Extract, Transform and Load (ETL) module, which allows extracting the information from the ICFES databases, applying cleaning and transformation filters to them and uploading them to a data market, designed under the multidimensional star model; a module for building cubes that allows you to organize data in dimensions; a results display module for the analysis of the information and a reports module for the production of reports by different dimensions.

For the construction of this tool, the different stages of the SCRUM agile methodology were followed, under a Model Template View (MTV) architecture model, using Django as a framework which uses Python as the main language, which was chosen because it is a language flexible and designed to be easy to read.

For the data treatment within the application, it was necessary to implement Pandas, a powerful data manipulation and analysis library using the Python language with which it was possible to analyze in depth the details of the data market.

In the data visualization process, Charts.js was used, which is a JavaScript library that uses the HTML5 canvas to display essential graphics for a business intelligence system. dynamic improving interaction with the user.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	17
I. GENERALIDADES.....	18
A. MODALIDAD.....	18
B. LÍNEA DE INVESTIGACIÓN	18
C. ALCANCE Y DELIMITACIÓN.....	18
D. DESCRIPCIÓN DEL PROBLEMA.....	18
E. FORMULACIÓN DEL PROBLEMA.....	20
F. OBJETIVOS.....	20
1) Objetivo General.....	20
2) Objetivos Específicos	20
G. JUSTIFICACIÓN	21
II. MARCO TEÓRICO.....	22
A. ANTECEDENTES	22
B. FUNDAMENTOS TEÓRICOS.....	23
1) Inteligencia de negocios	23
2) Mercado de datos	23
3) Dashboard o cuadros de mando.....	24
4) Pruebas ICFES saber 5°.....	24
a) Evaluados	25
b) Evaluación.....	25
c) Metodología para la elaboración de las pruebas	25
5) Análisis Multidimensional OLAP	26
6) Análisis Multidimensional MOLAP.....	27
7) Análisis Multidimensional ROLAP.....	28
8) Análisis Multidimensional HOLAP	28
9) SCRUM	29
a) Product Backlog	30
b) Sprint Planning.....	30
c) Sprint	30

d) Sprint Backlog	31
e) Daily sprint meeting.....	31
f) Demo y retrospectiva.....	31
g) Scrum máster	31
h) Product owner (PO)	31
i) Team	31
10) Metodología de Ralph Kimball	31
11) Planificación	32
12) Análisis de requerimientos	32
13) Modelado dimensional	32
14) Diseño físico	33
15) Entorno de desarrollo.....	33
a) PostgreSQL	33
b) Python	35
c) Django	36
d) PyCharm	37
e) Bootstrap	38
f) Proceso de extracción, transformación y carga (ETL)	40
g) Extracción, carga y transformación (ELT)	41
III. METODOLOGÍA.....	43
A. SPRINT 1: PLANEACIÓN.....	43
B. SPRINT 2: DISEÑO	43
1) Definición de Roles Scrum.....	43
2) Historias de usuario	44
3) Requerimientos funcionales	49
C. SPRINT 3: DESARROLLO DE LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS SABER EXPO BI 5.....	56
1) El modelo.....	56
2) La vista	56
3) La plantilla.....	56
4) Mercado de datos Saber Expo BI 5	57
D. SPRINT 4: CASOS DE PRUEBA	64

E. SPRINT 5: IMPLEMENTACIÓN Y DOCUMENTACIÓN.....	69
IV. RESULTADOS.....	69
A. DISEÑO DEL SISTEMA DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL).....	69
B. DISEÑO MULTIDIMENSIONAL SABER EXPO BI 5	69
1) Fase de extracción	70
2) Integración de Datos.....	71
3) Análisis de la calidad de datos.....	71
4) Limpieza de datos	79
5) Transformación de datos	79
6) Carga de datos	80
C. DISEÑO DE REPOSITORIO DE DATOS	80
1) Diccionario de Datos	81
D. ARQUITECTURA	85
1) Diseño estructural	87
2) Diagrama de paquetes.....	87
a) M = Modelos (Base de datos)	87
b) V = Vista (Funciones).....	88
c) T = Templates (Plantillas).....	88
E. PRUEBAS.....	96
1) Entrar como invitado	96
2) Inicio de sesión	97
3) Cerrar sesión	99
4) Visualización general	99
5) Crear gráfica general	102
6) Crear gráfica por número de estudiante.....	104
7) Crear grafica por puntaje	105
8) Crear gráfica por categoría	108
9) Nueva carga de datos	109
10) Cubos de datos.....	111
a) Número de estudiantes por municipio y por año de presentación	111
b) Número de estudiantes por municipio general.....	112
c) Puntaje área ciencias naturales por municipio	113

d) Puntaje área lenguaje por municipio.....	115
e) Puntaje área de matemáticas por municipio.....	116
f) Puntaje área competencias por municipio	117
g) Número de estudiantes por año de presentación.....	119
11) Gráficas estadísticas adicionales	121
a) Boxplot.....	121
b) Biplot.....	123
V. DISCUSIÓN.....	127
CONCLUSIONES.....	129
RECOMENDACIONES.....	130
REFERENCIAS.....	131
ANEXOS.....	136

FIGURAS

Pág.

Fig. 1. Dashboard y cuadros de mando	24
Fig. 2. Proceso de conversión de datos OLAP	26
Fig. 3. Proceso de conversión de datos con MOLAP.....	27
Fig. 4. Proceso de conversión de datos con ROLAP	28
Fig. 5. Proceso de conversión de datos híbridos con HOLAP	29
Fig. 6. Proceso de aplicación del método SCRUM.....	30
Fig. 7. Proceso de transformación y extracción (ETL)	40
Fig. 8. Proceso de transformación y extracción (ELT)	41
Fig. 9. Modelo vista plantilla	57
Fig. 10. Diseño multidimensional Saber Expo BI 5.....	70
Fig. 11. Saber Expo BI 5	86
Fig. 12. Interacción de la aplicación con la arquitectura MTV	87
Fig. 13. Model template view Expo Saber BI 5	88
Fig. 14. Directorios de plantillas de recepción de datos.....	89
Fig. 15. Directorio de Scripts (JS).....	89
Fig. 16. Directorio de plantillas utilizadas para gráficas y tablas	90
Fig. 17. Directorio de archivos (HTML).....	91
Fig. 18. Directorio de vistas	92
Fig. 19. Directorios con sus respectivas vistas.....	93
Fig. 20. Directorio de rutas.....	94
Fig. 21. Directorio de plantillas.....	94
Fig. 22. Directorio de plantillas rutas de librería y JQuery	95
Fig. 23. Saber Expo BI 5, entrar como invitado.....	96
Fig. 24. Saber Expo BI 5, vista como invitado	97
Fig. 25. Saber Expo BI 5, vista inicio de sesión.....	98
Fig. 26. Saber Expo BI 5, vista administrador	98
Fig. 27. Saber Expo BI 5, cerrar sesión.....	99
Fig. 28. Saber Expo BI 5, vista general dashboard	100
Fig. 29. Saber Expo BI 5, vista general de cubos.....	100
Fig. 30. Saber Expo BI 5, vista general tipos de gráficas	101
Fig. 31. Saber Expo BI 5, vista general por tabla.....	101
Fig. 32. Saber Expo BI 5, crear gráfica general	102
Fig. 33. Saber Expo BI 5, descargar gráfica general	103
Fig. 34. Saber Expo BI 5, documento.pdf gráfica general	103
Fig. 35. Saber Expo BI 5, crear gráfica por número de estudiantes	104
Fig. 36. Saber Expo BI 5, tabla número de estudiantes	104
Fig. 37. Saber Expo BI 5, gráfica de barras por número de estudiantes	105

Fig. 38. Saber Expo BI 5, crear gráfica por puntaje.....	105
Fig. 39. Saber Expo BI 5, crear gráfica de puntaje por institución y municipio	106
Fig. 40. Saber Expo BI 5, crear gráfica de puntaje por año de presentación	106
Fig. 41. Saber Expo BI 5, crear gráfica por puntaje con tipo de gráfica	107
Fig. 42. Saber Expo BI 5, crear gráfica de barras por puntaje	107
Fig. 43. Saber Expo BI 5, crear gráfica por categoría	108
Fig. 44. Saber Expo BI 5, visualizar gráfica por categoría	108
Fig. 45. Saber Expo BI 5, ventana de carga nuevo año.....	109
Fig. 46. Saber Expo BI 5, ventana cargar nuevo año, seleccionar delimitador.....	110
Fig. 47. Saber Expo BI 5, ventana cargar nuevo año, seleccionar archivo	110
Fig. 48. Saber Expo BI 5, ventana cargar nuevo año, archivo cargado	111
Fig. 49. Cubo, gráfica número de estudiantes por año y municipio.....	112
Fig. 50. Cubo, tabla número de estudiantes por año y municipio	112
Fig. 51. Cubo, gráfica número de estudiantes por municipio general.....	113
Fig. 52. Cubo, gráfica número de estudiantes por municipio general.....	113
Fig. 53. Cubo, gráfica puntaje área ciencias por municipio	114
Fig. 54. Cubo, tabla puntaje área ciencias por municipio	114
Fig. 55. Cubo, gráfica puntaje área lenguaje por municipio	115
Fig. 56. Cubo, tabla puntaje área lenguaje por municipio.....	116
Fig. 57. Cubo, gráfica puntaje área matemáticas por municipio.....	117
Fig. 58. Cubo, tabla de puntaje área matemáticas por municipio	117
Fig. 59. Cubo, gráfica puntaje área competencias por municipio	118
Fig. 60. Cubo, gráficas varias puntaje área competencias por municipio	118
Fig. 61. Cubo, tabla puntaje área competencias por municipio	119
Fig. 62. Cubo, gráfica número de estudiantes por año de presentación.....	119
Fig. 63. Cubo, gráficas varias número de estudiantes por año de presentación.....	120
Fig. 64. Cubo, tabla número de estudiantes por año de presentación	120
Fig. 65.Boxplot, menú boxplot y año de prueba a seleccionar	121
Fig. 66.Boxplot, gráfico boxplot del año 2015	122
Fig. 67.Boxplot, gráfico boxplot del año 2016	122
Fig. 68.Boxplot, gráfico boxplot del año 2017	123
Fig. 69.Biplot, menú biplot a seleccionar.....	124
Fig. 70. Biplot, gráfica biplot correlación materia-año	124
Fig. 71. Biplot, gráfica biplot correlación materia-discapacidad	125
Fig. 72. Biplot, gráfica biplot correlación materia-género	125
Fig. 73. Biplot, gráfica biplot correlación materias.....	126
Fig. 74. Biplot, gráfica biplot correlación materias-zona.....	126

TABLAS

Pág.

TABLA I HISTORIA DE USUARIO CREAR CUENTA	44
TABLA II HISTORIA DE USUARIO INICIAR SESIÓN	45
TABLA III HISTORIA DE USUARIO CERRAR SESIÓN	45
TABLA IV HISTORIA DE USUARIO RESTABLECER CONTRASEÑA	45
TABLA V HISTORIA DE USUARIO INGRESAR COMO INVITADO	46
TABLA VI HISTORIA DE USUARIO VISUALIZAR INDICADORES	46
TABLA VII HISTORIA DE USUARIO FILTRAR DATOS	46
TABLA VIII HISTORIA DE USUARIO AGREGAR NUEVO AÑO	47
TABLA IX HISTORIA DE USUARIO SELECCIONAR SEPARADOR	47
TABLA X HISTORIA DE USUARIO DESCARGAR GRÁFICA .PDF	47
TABLA XI HISTORIA DE USUARIO CREAR GRÁFICA GENERAL	48
TABLA XII HISTORIA DE USUARIO GRÁFICA POR PUNTAJE	48
TABLA XIII HISTORIA DE USUARIO GRÁFICA POR CATEGORÍA	48
TABLA XIV HISTORIA DE USUARIO CUBOS DE DATOS	49
TABLA XV REQUERIMIENTO FUNCIONAL. EJECUTAR APLICACIÓN	49
TABLA XVI REQUERIMIENTO FUNCIONAL. INICIAR SESIÓN	50
TABLA XVII REQUERIMIENTO FUNCIONAL. CERRAR SESIÓN	50
TABLA XVIII REQUERIMIENTO FUNCIONAL. CREAR USUARIO	50
TABLA XIX REQUERIMIENTO FUNCIONAL. RESTABLECER CONTRASEÑA	51
TABLA XX REQUERIMIENTO FUNCIONAL. ENTRAR COMO INVITADO	51
TABLA XXI REQUERIMIENTO FUNCIONAL. AGREGAR NUEVO AÑO	52
TABLA XXII REQUERIMIENTO FUNCIONAL. ESCOGER SEPARADOR	52
TABLA XXIII REQUERIMIENTO FUNCIONAL. VISUALIZAR DATOS	53
TABLA XXIV REQUERIMIENTO FUNCIONAL. MOSTRAR DIMENSIONES	53
TABLA XXV REQUERIMIENTO FUNCIONAL. VISTA GLOBAL	54
TABLA XXVI REQUERIMIENTO FUNCIONAL. CREAR GRÁFICA	54
TABLA XXVII REQUERIMIENTO FUNCIONAL. DESCARGAR GRÁFICA	55
TABLA XXVIII REQUERIMIENTO FUNCIONAL. CUBOS	55
TABLA XXIX ATRIBUTOS (DIMENSIÓN ESTUDIANTE)	57
TABLA XXX ATRIBUTOS (DIMENSIÓN INSTITUCIONES)	58
TABLA XXXI ATRIBUTOS (DIMENSIÓN MUNICIPIOS)	58
TABLA XXXII ATRIBUTOS (DIMENSIÓN PRUEBA)	59
TABLA XXXIII ATRIBUTOS (DIMENSIÓN TIEMPO)	59
TABLA XXXIV ATRIBUTOS (DIMENSIÓN FACT SABER 5)	60
TABLA XXXV REPOSITORIO DE DATOS (VALORES PLAUSIBLES)	61
TABLA XXXVI REPOSITORIO DE DATOS (ENTIDADES)	63
TABLA XXXVII REPOSITORIO DE DATOS (ESTABLECIMIENTOS)	63
TABLA XXXVIII REPOSITORIO DE DATOS (MUNICIPIOS)	64

TABLA XXXIX CASO DE PRUEBA INICIO DE SESIÓN	64
TABLA XL CASO DE PRUEBA CREAR USUARIO	65
TABLA XLI CASO DE PRUEBA CERRAR SESIÓN	65
TABLA XLII CASO DE PRUEBA INGRESAR COMO INVITADO	65
TABLA XLIII CASO DE PRUEBA RESTABLECER CONTRASEÑA.....	66
TABLA XLIV CASO DE PRUEBA VISUALIZAR INDICADORES	66
TABLA XLV CASO DE PRUEBA CREAR GRÁFICA GENERAL	66
TABLA XLVI CASO DE PRUEBA DESCARGAR GRÁFICA	67
TABLA XLVII CASO DE PRUEBA GRÁFICA POR PUNTAJE	67
TABLA XLVIII CASO DE PRUEBA GRÁFICA POR CATEGORÍA	68
TABLA XLIX CASO DE PRUEBA VER CUBOS.....	68
TABLA L CASO DE PRUEBA AGREGAR NUEVO AÑO	68
TABLA LI TOTAL DE REGISTROS.....	71
TABLA LII ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN PRUEBA)	72
TABLA LIII ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN MUNICIPIO).....	73
TABLA LIV ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN ESTUDIANTE).....	74
TABLA LV ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN INSTITUCIÓN).....	75
TABLA LVI ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN TIEMPO).....	76
TABLA LVII ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN FACT_SABER5)	76
TABLA LVIII DICCIONARIO DE DATOS (DIM_ESTUDIANTES).....	81
TABLA LIX DICCIONARIO DE DATOS (DIM_MUNICIPIO)	82
TABLA LX DICCIONARIO DE DATOS (DIM_INSTITUCION)	82
TABLA LXI DICCIONARIO DE DATOS (DIM_PRUEBA).....	83
TABLA LXII DICCIONARIO DE DATOS (DIM_TIEMPO)	84
TABLA LXIII DICCIONARIO DE DATOS (FACT_SABER5)	84

INTRODUCCIÓN

En este proyecto se construye una herramienta que brinda información de calidad para soportar la toma de decisiones de las Instituciones Educativas de la Subregión de Obando, en lo relacionado al rendimiento académico de sus estudiantes en las pruebas Saber 5.

Para esto, se hizo un sistema con el cual las instituciones educativas pueden afianzar su conocimiento referente al rendimiento de sus estudiantes, en cuanto a el entendimiento mediante la visualización de datos, pues este brinda un mejor soporte a la explicación de información que en ciertas circunstancias resulta confuso, sobre todo si no se cuenta con conocimientos previos para poder interpretar dicha información. Esta información, que dada su extensión y distribución en diferentes fuentes sin relación tecnológica alguna, hace muy difícil su visualización integrada para una inteligente toma de decisiones, que permita el entendimiento para el mejoramiento de las instituciones, que comprenden los municipios de Aldana, Contadero, Córdoba, Cuaspud, Cumbal, Funes, Guachucal, Gualmatán, Iles, Ipiales, Potosí, Puerres y Pupiales.

Teniendo en cuenta la necesidad de que la información proporcionada sobre las pruebas saber sean más comprensibles, se construyó una herramienta de inteligencia de negocios, para realizar el análisis multidimensional y de visualización de resultados, por medio de gráficas estadísticas y cuadros de mando denominados dashboards, tomando datos que almacenan la información histórica de los estudiantes de las instituciones, que permitió la generación de indicadores, los cuales sirvieron como soporte para la toma de decisiones, en lo relacionado con la formulación de estrategias para afianzar la calidad educativa en las instituciones.

I. GENERALIDADES

A. MODALIDAD

El presente proyecto corresponde a la modalidad de trabajo de investigación, desarrollado por los estudiantes pertenecientes al Grupo de Investigación Aplicado en Sistemas (GRIAS), del departamento de Sistemas de la Universidad de Nariño.

B. LÍNEA DE INVESTIGACIÓN

Este proyecto corresponde a la línea de investigación: Descubrimiento de conocimiento en bases de datos.

C. ALCANCE Y DELIMITACIÓN

Se propuso el desarrollo e implementación de una herramienta de inteligencia de negocios, con la cual se pueda realizar un análisis de la problemática que se presenta en los resultados de las pruebas saber 5 en las instituciones educativas de la subregión de Obando, con el fin de obtener datos relevantes que permita identificar patrones estadísticos, los cuales son fuente de información vital para los municipios que conforman la subregión, proyectando la herramienta a futuro, siendo un apoyo constante al sector educativo.

La Subregión de Obando o Sur es una de las 13 subregiones del departamento de Nariño. Comprende los municipios de Aldana, Contadero, Córdoba, Cuaspud Carlosama, Cumbal, Funes, Guachucal, Gualmatán, Iles, Ipiales, Potosí, Puerres y Pupiales, que abarcan un total de 4.894 kilómetros cuadrados.

D. DESCRIPCIÓN DEL PROBLEMA

El propósito de las pruebas SABER 3.º, 5.º y 9.º es contribuir al mejoramiento de la calidad de la educación colombiana, mediante la realización de evaluaciones periódicas en las que se valoran las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros [5]. Las competencias en Lenguaje, Matemáticas, Ciencias Naturales y Competencias Ciudadanas son

áreas, que a criterio de expertos se constituyen en los componentes principales para desarrollar en los estudiantes, las competencias básicas para la vida y que les permiten alcanzar un desempeño eficaz y eficiente dentro de la sociedad [6].

Los resultados de estas evaluaciones permiten que los establecimientos educativos, las secretarías de educación, el Ministerio de Educación Nacional (MEN) y la sociedad en general conozcan cuáles son las fortalezas y debilidades y, a partir de estas, puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación. Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento.

Uno de los principales problemas que presentan las instituciones educativas de los municipios de la subregión de Obando, es el bajo rendimiento que se viene presentando en los últimos años en las pruebas saber 5 [7]. Además, no se contaba con una información histórica de los resultados obtenidos por los estudiantes de esta región en las pruebas Saber 5, que permita obtener resultados confiables y de calidad para la realización de análisis y soporte de decisiones para el mejoramiento de la calidad educativa en estas instituciones.

Como primer paso para solucionar este problema, el grupo de investigación GRIAS del departamento de Sistemas de la facultad de Ingeniería de la Universidad de Nariño, a través de estudiantes investigadores desarrollaron el proyecto [8] que fue financiado por el sistema de investigaciones de la Universidad de Nariño y como resultado de ello, actualmente se cuenta con un mercado de datos en estrella, que almacena la información histórica de los resultados de los estudiantes de las instituciones educativas de la sub región de Obando que presentaron las pruebas Saber 5 en los años 2014,2015 y 2016.

Con el fin de dar continuidad a este proyecto, fue necesario desarrollar una herramienta que permita a los funcionarios de las instituciones educativas, actualizar fácilmente la información de los siguientes años en los que se presentan estas pruebas y permita además obtener información de calidad a través del análisis de esta información, que permita soportar la toma de decisiones para mejorar el rendimiento académico de los estudiantes de la básica primaria.

En este proyecto de investigación se analizó los requerimientos y necesidades de información estratégica de cada Institución, para generar información de calidad para soportar la toma de decisiones en pro del mejoramiento de la calidad de la educación de estas instituciones educativas. Por otra parte, el sistema se desarrolló bajo software libre y podrá ser utilizado sin costo alguno por cualquier institución educativa que lo requiera.

E. FORMULACIÓN DEL PROBLEMA

¿Es posible obtener información ágil y de calidad a partir de las bases de datos que nos brindan los resultados de las pruebas saber 5, para hacer un análisis confiable, mediante indicadores gráficos de fácil acceso, que puedan contribuir a la calidad educativa de las instituciones de los municipios de la Subregión de Obando?

F. OBJETIVOS

1) Objetivo General

Construir una herramienta web, de Inteligencia de Negocios para la gestión de las pruebas saber 5, ajustado a los requerimientos y necesidades de las diferentes instituciones educativas de educación básica primaria, de los municipios de la subregión de Obando del departamento de Nariño.

2) Objetivos Específicos

- Construir módulo administración que permita el ingreso en sus distintos roles de administrador e invitado.
- Construir módulo de (ETL) carga, transformación y limpieza de datos, que permita añadir un nuevo año al mercado de datos ya establecido.
- Construir módulo de visualización y estadísticas de la información proporcionada por el mercado de datos.
- Construir módulo de visualización de cubos, que permita graficar información estática y relevante.
- Construir módulo de filtrado de datos a visualizar, que permita elegir de forma dinámica consultas categorizadas sobre los datos.

G. JUSTIFICACIÓN

Según la secretaria de educación departamental de Nariño, las sub regiones del departamento de Nariño, cuentan con un bajo promedio en las distintas pruebas realizadas, comprendidas a los estudiantes de 3°,5° y 9° por lo cual se implementaron estrategias para la divulgación y orientación, para mejorar la calidad educativa [9], dicho proceso se ha venido haciendo constantemente, impartiendo a rectores y directivos de las diferentes instituciones educativas que comprenden las sub regiones; sin embargo los resultados han seguido bajando de acuerdo a la clasificación de colegios proporcionada por el ICFES [10]. Para esto, se creó una herramienta web que permitirá a las Instituciones educativas de los municipios de la subregión de Obando, tener información actualizada sobre las pruebas saber 5, que ayudará a la mejora del rendimiento académico de los estudiantes, proporcionando información sobre fortalezas y debilidades, para apoyar la toma de decisiones en caso que se requiera según [11], permitiendo que las directivas de las instituciones, puedan hacer un análisis y determinar en qué áreas del conocimiento existen mayores falencias.

II. MARCO TEÓRICO

A. ANTECEDENTES

En la actualidad toda empresa pública o privada necesita herramientas que le permitan depositar mucha confianza en la toma de decisiones sobre los negocios que efectúa. Para la toma de decisiones se necesita analizar hechos y cifras soportados en información confiable, ya que la competitividad en los negocios crece aceleradamente y por ende las decisiones que se deben tomar en la empresa deben ser más rápidas y deben estar basadas en buenos cimientos y para esto se debe manejar y analizar adecuadamente la información que posee la empresa en el menor tiempo posible ya que la tardanza o el más mínimo error en la información puede ocasionar problemas a la empresa en cuanto a competitividad y agilidad en el mercado [12].

Para poder cumplir con el objeto de estudio se tomó como antecedentes, proyectos, artículos y tesis relevantes, los cuales sirvieron para el desarrollo de la aplicación Saber Expo BI 5.

Para trabajar en la construcción de la aplicación, fue necesario encontrar un previo estudio, el cual sirva como guía para la base de datos que la aplicación va a consumir, se tomó el trabajo de grado [8]. Cuyo objetivo fue construir un mercado de datos para almacenar la información histórica de los estudiantes de las instituciones educativas que presentaron las pruebas saber 5° entre los años 2014 al 2016; para posteriormente aplicar cambios tales como, transformación y limpieza de datos referenciados de [13] para adicionar nuevos años al prototipo de la base de datos de la aplicación, para esto fue necesario aplicar distintos métodos y técnicas de limpieza que ayudaron al entendimiento de conceptos clave para la carga de nuevos datos, teniendo en cuenta el delimitador de las nuevas tablas que se desea añadir al mercado ya existente. Para lo concerniente a consultas en la base de datos se investigó técnicas de minería de datos según [14], cuyos factores se asociaron con el desempeño académico de las diferentes competencias de las pruebas saber, comprendiendo los años del 2014 al 2016; subsiguientemente se hizo el diseño de gráficas estadísticas y cuadros de mando, tomando conceptos de la implementación de un dash board creado y documentado por

[15] proporcionando información clara la creación de cuadros de mando, filtros, categorías, tipos de grafica que son muy necesarios para la creación de este aplicativo.

B. FUNDAMENTOS TEÓRICOS

1) Inteligencia de negocios

La Inteligencia de Negocios es una herramienta bajo la cual diferentes tipos de organizaciones pueden soportar la toma de decisiones basadas en información precisa y oportuna; garantizando la generación del conocimiento necesario que permita escoger la alternativa que sea más conveniente para el éxito de la empresa [16].

Podemos utilizar un sistema de inteligencia de negocios como un importante aliado para almacenar los datos de clientes, empleados, departamentos, compras, ventas, entre otros en aplicaciones, sistemas financieros o fuentes de datos con lo que se pretende tener una mayor ventaja sobre la competencia. Para esto se necesita profundizar el nivel de conocimiento de nuestros clientes, empleados, operaciones para así, tener la capacidad de encontrar patrones de comportamiento, monitorear, rastrear, entender, administrar y contestar aquellas interrogantes que me permitan maximizar el rendimiento de nuestra empresa [17].

Para esta investigación, se extrae un proceso que existe entre el repositorio de datos y la interfaz de acceso al usuario, este es el motor de BI que permite habilitar componentes, administrar consultas, monitorea procesos, cálculos y métricas.

2) Mercado de datos

Un mercado de datos es un elemento sustantivo dentro de la última generación de bases de datos. Un mercado de datos proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una interfaz de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones [18].

3) Dashboard o cuadros de mando

Un cuadro de mando es una vista de información geográfica y de datos que permiten monitorizar eventos, tomar decisiones, informar a otros y ver tendencias. Los cuadros de mando se han diseñado para mostrar varias visualizaciones que trabajan juntas en una sola pantalla. Ofrecen una vista integral de sus datos y proporcionan información clave para tomar decisiones de un vistazo como se muestra en [Fig.1].

Los cuadros de mando se crean de manera que sean intuitivos y fáciles de entender por diferentes equipos y personas de dentro y fuera de una organización [19].

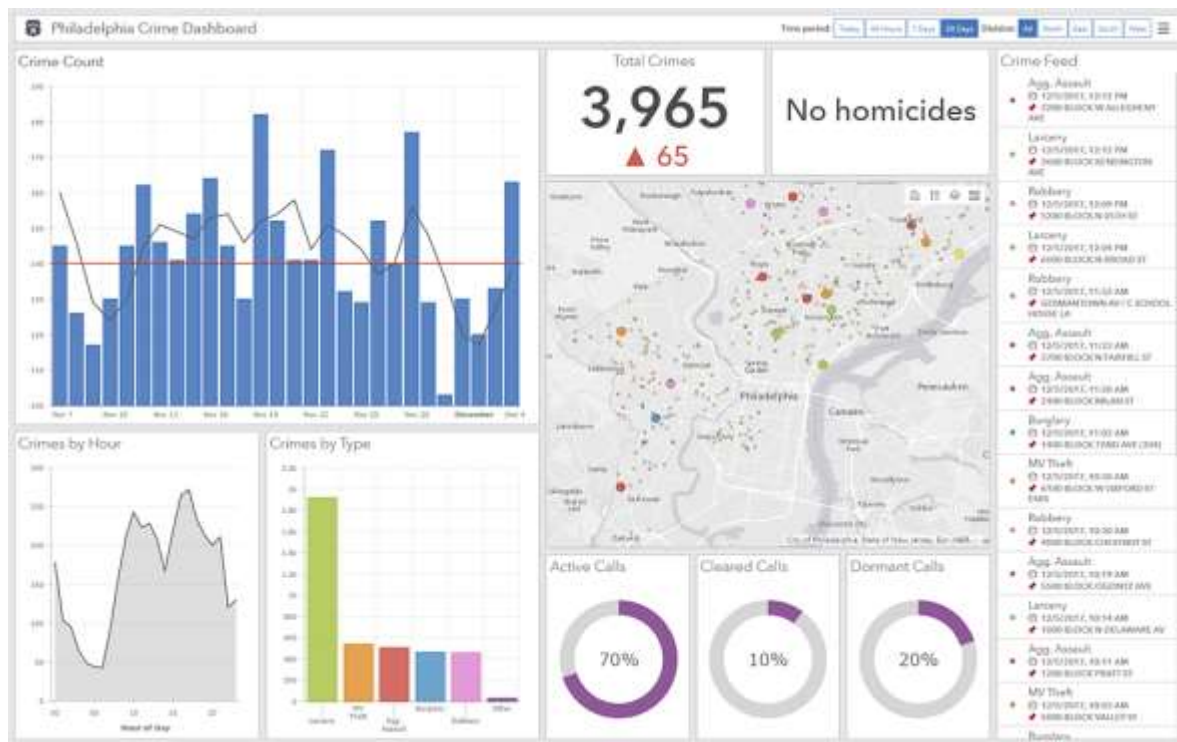


Fig. 1. Dashboard y cuadros de mando

Fuente: Que es un cuadro de mando,2018, enlace <https://sisfor.osinfor.gob.pe/osinfor/help/es/portal/latest/use/what-is-a-dashboard.htm>

4) Pruebas ICFES saber 5°

Las pruebas saber 5° contribuyen al mejoramiento de la calidad de la educación colombiana, mediante la realización de pruebas periódicas en las que se evalúan las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros. Los resultados de estas evaluaciones permiten que los establecimientos educativos, las secretarías de educación, el MEN y la sociedad en general conozcan cuáles son las fortalezas y debilidades y, a partir de estas, puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación. Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento según [1].

a) Evaluados

Saber 5. ° evalúa a los alumnos de 5. ° grado mediante la aplicación periódica de pruebas de competencias básicas que determinan los desempeños de los estudiantes y la calidad educativa de los establecimientos educativos oficiales, privados, urbanos y rurales [20].

b) Evaluación

Las pruebas evalúan las competencias que han desarrollado los estudiantes entre 4. ° y 5. ° grados, acorde con los estándares básicos de competencias establecidos por el MEN, que son los referentes comunes a partir de los cuales es posible establecer qué tanto los estudiantes, y el sistema educativo en su conjunto, están cumpliendo con las expectativas de calidad educativa.

En el contexto escolar las competencias se desarrollan a través del trabajo concreto en una o más áreas. En las pruebas Saber 9° se evalúan las áreas de matemáticas y lenguaje. Si bien estas pruebas no abarcan la totalidad de contenidos ni de estándares definidos para cada área, y algunos logros únicamente pueden valorarse en el ámbito de las actividades escolares, estas pruebas son un indicador fundamental del avance en el proceso de formación en educación básica, y permiten establecer los niveles de desempeño, las fortalezas y debilidades de los estudiantes para, a su vez, plantear estrategias concretas de mejoramiento [21].

c) Metodología para la elaboración de las pruebas

El diseño y construcción de los exámenes Saber se realizan sobre la base de las especificaciones de cada módulo. Estas determinan con exactitud en qué consisten las competencias que se evalúan

y cómo se evalúan. Son diseñadas por el ICFES junto con equipos de expertos de cada área. Las especificaciones se desarrollan siguiendo el Modelo Basado en Evidencias (MBE). De acuerdo con este modelo, en las especificaciones se formalizan, primero, las afirmaciones sobre las competencias que posee un estudiante dado su desempeño en el módulo. Luego, se describen las evidencias que sustentan cada una de las afirmaciones.

Por último, se describen las tareas que se le pide realizar al evaluado para obtener las evidencias que dan sustento a las afirmaciones. De esta manera, la elaboración de las especificaciones garantiza una completa comparabilidad de los exámenes [22].

5) Análisis Multidimensional OLAP

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente [Fig.2].

Un principio clave del OLAP es que los usuarios deberían obtener tiempos de respuesta consistentes para cada vista de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores recalculados son la base de las ganancias de desempeño del OLAP [23].

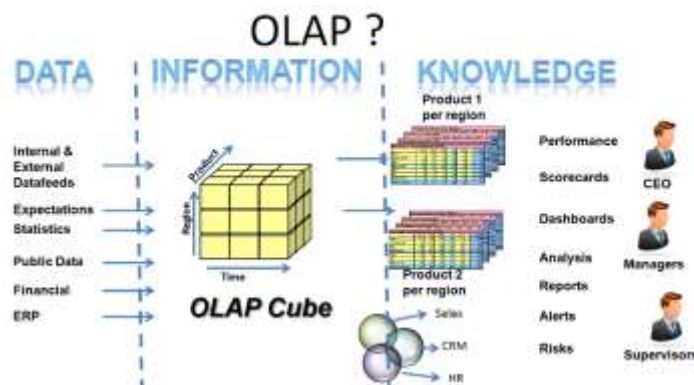


Fig. 2. Proceso de conversión de datos OLAP

Fuente: Proceso de conversión de datos, 2019, enlace <https://sysinfotechs.wordpress.com/2020/03/01/9types-of-most-commonly-used-information-system-in-business-management/>

6) Análisis Multidimensional MOLAP

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales [Fig.3]. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis. [24]

MOLAP Architecture

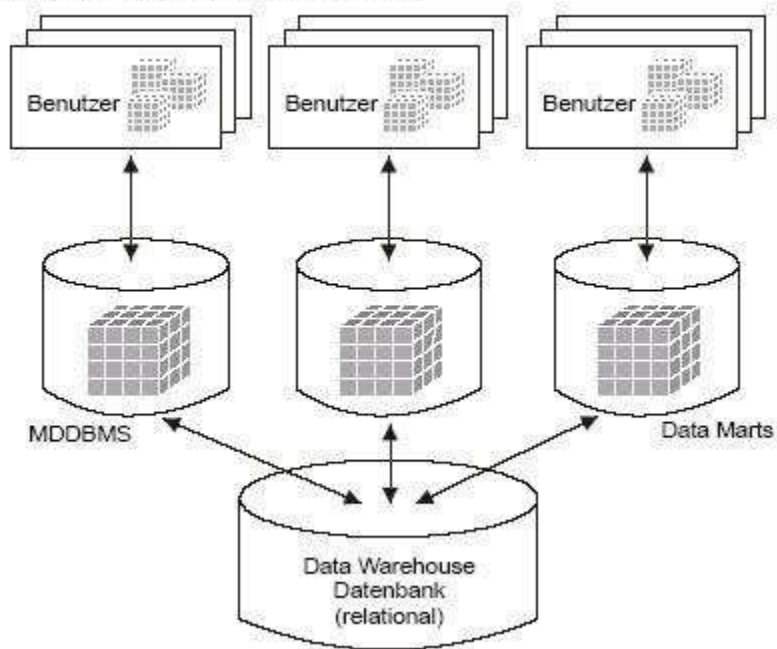


Fig. 3. Proceso de conversión de datos con MOLAP

Fuente: Diferencias entre ROLAP y MOLAP,2020, enlace <https://digitalpicturesimg.blogspot.com/2020/02/molap-and-rolap-difference.html>

7) Análisis Multidimensional ROLAP

La arquitectura ROLAP, accede a los datos almacenados en un data ware house para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales [Fig.4].

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato [25].

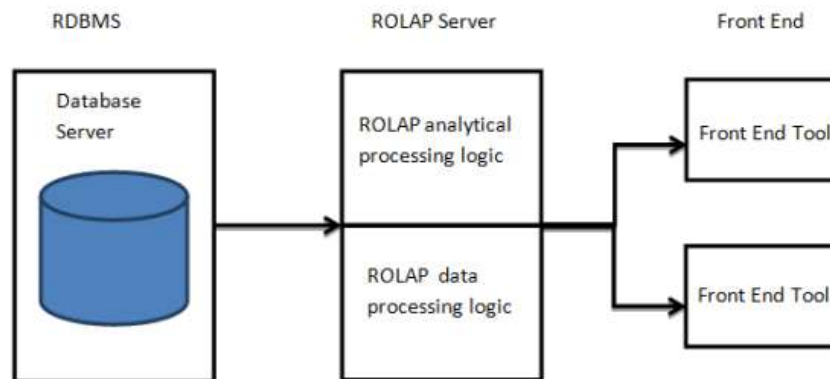


Fig. 4. Proceso de conversión de datos con ROLAP

Fuente: Data warehousing ROLAP, 2020, enlace <https://www.teradatapoint.com/data-warehousing-rolap>

8) Análisis Multidimensional HOLAP

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas, desempeño superior y gran escalabilidad [Fig.5]. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado [26].

HOLAP/MQE/Hybrid architecture

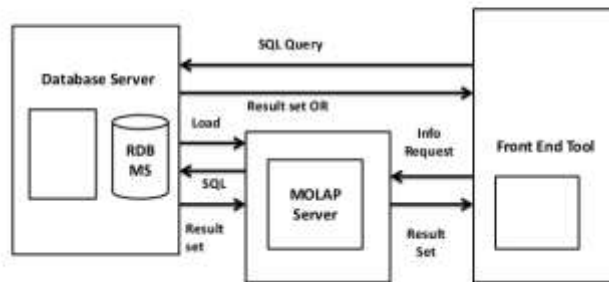


Fig. 5. Proceso de conversión de datos híbridos con HOLAP

Fuente: Arquitectura HOLAP, 2018, enlace https://www.researchgate.net/Figure/HOLAP-architecture-The-main-advantages-of-HOLAP-include-o-High-performance-dimensional_Fig21_319852408

9) SCRUM

Para desarrollar la herramienta de Negocio Inteligente se utiliza la metodología Scrum, la cual es un proceso que permite el manejo de buenas prácticas complementadas unas a otras, para trabajar en equipo el desarrollo optimo y eficaz de un proyecto.

En Scrum se realizan entregas parciales y regulares del producto final, permitiendo tener a gusto al cliente por recibir por bloques parte del producto funcional y así poder detectar fallas o cosas que no gusten y tener la oportunidad de corregirlo [Fig.6].

El sprint es la unidad básica de trabajo para un equipo Scrum, es una simple iteración llevada a cabo por los miembros del equipo. Un equipo puede llevar a cabo varios sprint en el desarrollo del proyecto, un Sprint es como tal un entregable en el desarrollo del proyecto, su tiempo mínimo de entrega es de una semana y el tiempo máximo de 4 semanas.

Scrum es una metodología ágil y flexible para gestionar el desarrollo de software, cuyo principal objetivo es maximizar el retorno de la inversión para su empresa que se basa en construir primero la funcionalidad de mayor valor para el cliente y en los principios de inspección continua, adaptación, autogestión e innovación.

En Scrum se realizan entregas parciales y regulares del producto final, priorizadas por el beneficio que aportan al receptor del proyecto. Por ello, Scrum está especialmente indicado para proyectos en entornos complejos, donde se necesita obtener resultados pronto, donde los requisitos son cambiantes o poco definidos y en donde la productividad es fundamental. [27]

En cada evento realizamos la inspección y adaptación de uno o más artefactos, aunque no solo inspeccionamos y adaptamos estos:

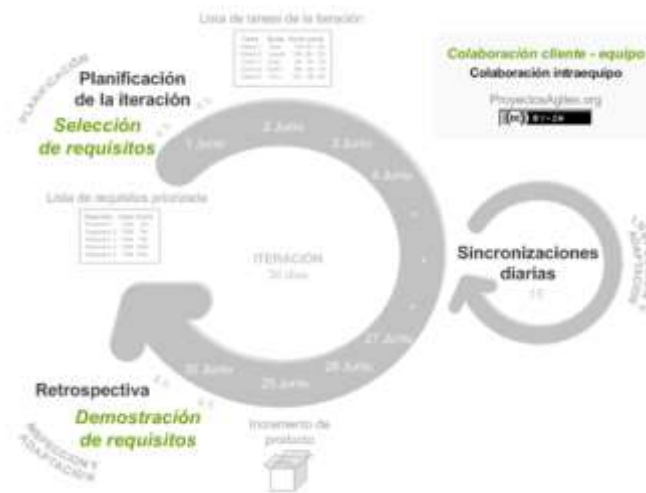


Fig. 6. Proceso de aplicación del método SCRUM

Fuente: Que es SCRUM, 2017, enlace <https://proyectosagiles.org/que-es-scrum/>

a) Product Backlog

Conjunto de requisitos denominados historias descritas en un lenguaje no técnico y priorizados por valor de negocio, o lo que es lo mismo, por retorno de inversión considerando su beneficio y coste. Los requisitos y prioridades se revisan y ajustan durante el curso del proyecto a intervalos regulares.

b) Sprint Planning

Reunión durante la cual el Product Owner presenta las historias del backlog por orden de prioridad. El equipo determina la cantidad de historias que puede comprometerse a completar en ese sprint, para en una segunda parte de la reunión, decidir y organizar cómo lo va a conseguir.

c) Sprint

Iteración de duración prefijada durante la cual el equipo trabaja para convertir las historias del Product Backlog a las que se ha comprometido, en una nueva versión del software totalmente operativo.

d) Sprint Backlog

Lista de las tareas necesarias para llevar a cabo las historias del sprint.

e) Daily sprint meeting

Reunión diaria de cómo máximo 15 min. en la que el equipo se sincroniza para trabajar de forma coordinada. Cada miembro comenta que hizo el día anterior, que hará hoy y si hay impedimentos.

f) Demo y retrospectiva

Reunión que se celebra al final del sprint y en la que el equipo presenta las historias conseguidas mediante una demostración del producto. Posteriormente, en la retrospectiva, el equipo analiza qué se hizo bien, qué procesos serían mejorables y discute acerca de cómo perfeccionarlos.[28].

g) Scrum máster

Persona que lidera al equipo guiándolo para que cumpla las reglas y procesos de la metodología. Gestiona la reducción de impedimentos del proyecto y trabaja con el Product Owner para maximizar el ROI.

h) Product owner (PO)

Representante de los accionistas y clientes que usan el software. Se focaliza en la parte de negocio y él es responsable del ROI del proyecto (entregar un valor superior al dinero invertido). Traslada la visión del proyecto al equipo, formaliza las prestaciones en historias a incorporar en el Product Backlog y las Re prioriza de forma regular.

i) Team

Grupo de profesionales con los conocimientos técnicos necesarios y que desarrollan el proyecto de manera conjunta llevando a cabo las historias a las que se comprometen al inicio de cada sprint. [29]

10) Metodología de Ralph Kimball

Esta investigación involucra la construcción de un mercado de datos y utiliza la metodología propuesta por Kimball denominada Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle), una de las más utilizadas en proyectos de inteligencia de negocios (BI).

Esta metodología contempla seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación, que se describen a continuación [30].

11) Planificación

Determinar el propósito del proyecto, sus objetivos y el alcance. Incluye las acciones típicas de un plan de proyecto:

- Registrar los requerimientos del negocio.
- Identificar las tareas.
- Programar las tareas.
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos.

12) Análisis de requerimientos

Se debe conocer tanto como se pueda del negocio, implica revisar informes, rastrear documentos, dominar los términos y la terminología del negocio.

13) Modelado dimensional

Proceso dinámico e iterativo. Inicia con un modelo dimensional de alto nivel generado a partir de los procesos priorizados de la matriz de requerimientos.

- Establecer el nivel de granularidad.
- Elegir las dimensiones.
- Identificar las tablas de hechos y medidas.
- Modelo gráfico de alto nivel.
- Identificación de atributos de dimensiones y tablas de hechos.
- Implementar el modelo dimensional detallado.
- Prueba del modelo.

- Revisión y validación del modelo.

14) Diseño físico

Se eligen las estructuras que soporten el diseño lógico; busca dar respuesta a preguntas como las siguientes:

- ¿Cómo puede determinar cuán grande será el sistema?
- ¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?
- ¿Cómo se debe configurar el sistema?
- ¿Cuánta memoria y servidores se necesitan?
- ¿Qué tipo de almacenamiento y procesadores son requeridos?
- ¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?

15) Entorno de desarrollo

Para trabajar con el proyecto se necesita tener instalados los siguientes programas y dependencias:

- PostgreSQL
- Python
- Django
- Pycharm
- Visual Studio Code

A continuación, se indica como instalar y configurar correctamente cada uno de ellos.

a) PostgreSQL

PostgreSQL es el motor de base de datos de la actualidad. Lo que lo hace el Sistema Gestor de Bases de Datos de Código libre más potente y robusto del mercado [31].

Ventajas PostgreSQL

- **Instalación ilimitada y gratuita:** Podemos instalarlo en todos los equipos que se desee, independientemente de la plataforma y la arquitectura que usemos, PostgreSQL está disponible para los diferentes SO, Unix, Linux y Windows, en 32 y 64 bits, lo que lo hace un sistema multiplataforma y de gran escalado.
- **Gran escalabilidad:** Nos permite configurar PostgreSQL en cada equipo según el hardware; Por esto, es capaz de ajustarse al número de CPU y a la cantidad de memoria disponible de forma óptima. Con ello logramos una mayor cantidad de peticiones simultáneas a la base de datos de forma correcta.
- **Estabilidad y confiabilidad:** Tiene más de 20 años de desarrollo activo y en constante mejora. No se han presentado nunca caídas de la base de datos que estén documentadas gracias a su capacidad de establecer un entorno de Alta disponibilidad en donde podemos hacer tareas de mantenimiento o recuperación sin bloquear completamente el sistema.
- **Potencia y Robustez:** PostgreSQL cumple en su totalidad con la característica ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad en español). Por ello permite que las transacciones no interfieran unas con otras. Con ello se garantiza la información de las Bases de Datos y que los datos perduren en el sistema.
- **Extensibilidad:** tenemos a nuestra disponibilidad una gran variedad de extensiones distribuidas por el grupo de desarrolladores de PostgreSQL. Estas extensiones pueden ser lenguajes de programación, tales como, Perl, Java, Python, C++ y muchos más.

Desventajas de PostgreSQL

- **Base de datos pequeñas:** Es relativamente lento en inserciones y actualizaciones en bases de datos pequeñas, PostgreSQL está diseñado para ambientes de alto volumen, por ello la velocidad de respuesta pueda parecer lenta en comparación con bases de datos de pequeño tamaño.

- **Soporte oficial:** No cuenta con un soporte en línea o telefónico, por el contrario, cuenta con foros oficiales donde los usuarios pueden exponer sus dudas que responden, es de resaltar que la comunidad PostgreSQL es una de las más activas del mercado.
- **Sintaxis:** La sintaxis de algunos de sus comando o sentencias puede llegar a no ser intuitiva si no tienes un nivel medio de conocimientos en lenguaje SQL.

b) Python

Es un lenguaje de programación versátil multiplataforma y multiparadigma que se destaca por su código legible y limpio. Una de las razones de su éxito es que cuenta con una licencia de código abierto que permite su utilización en cualquier escenario.

Python es ideal para trabajar con grandes volúmenes de datos ya que, el ser multiplataforma, favorece su extracción y procesamiento, por eso lo eligen las empresas de Big Data. A nivel científico, tiene una gran biblioteca de recursos con especial énfasis en las matemáticas para aspirantes a programadores en áreas especializadas. Y si esto fuera poco, es posible crear videojuegos, aunque no es tan eficiente como Java o C# [32].

Ventajas de Python

- **Simplificado y rápido:** Este lenguaje simplifica mucho la programación, es un gran lenguaje para scripting.
- **Elegante y flexible:** El lenguaje ofrece muchas facilidades al programador al ser fácilmente legible e interpretable.
- **Programación sana y productiva:** Es sencillo de aprender, con una curva de aprendizaje moderada. Es muy fácil comenzar a programar y fomenta la productividad.
- **Ordenado y limpio:** es muy legible y sus módulos están bien organizados.
- **Portable:** Es un lenguaje muy portable. Podemos usarlo en prácticamente cualquier sistema de la actualidad.

- **Comunidad:** Cuenta con un gran número de usuarios. Su comunidad participa activamente en el desarrollo del lenguaje.

Desventajas de Python

- **Hosting:** La mayoría de los servidores no tienen soporte a Python, y si lo soportan, la configuración es un poco difícil.
- **Librerías incluidas:** Algunas librerías que trae por defecto no son del gusto de amplio de la comunidad, y optan a usar librerías de terceros.

c) Django

Django es un framework web de alto nivel que permite el desarrollo rápido de sitios web seguros y mantenibles. Desarrollado por programadores experimentados, Django se encarga de gran parte de las complicaciones del desarrollo web, por lo que puedes concentrarte en escribir tu aplicación sin necesidad de reinventar la rueda. Es gratuito y de código abierto, tiene una comunidad próspera y activa, una gran documentación y muchas opciones de soporte gratuito y de pago [33].

Características:

- **Completo:** Django sigue la filosofía "Baterías incluidas" y provee casi todo lo que los desarrolladores quisieran que tenga "de fábrica". Porque todo lo que necesitas es parte de un único "producto", todo funciona a la perfección, sigue principios de diseño consistentes y tiene una amplia y actualizada documentación.
- **Versátil:** Django puede ser para construir casi cualquier tipo de sitio web, desde sistemas manejadores de contenidos y wikis, hasta redes sociales y sitios de noticias. Puede funcionar con cualquier framework en el lado del cliente.
- **Seguro:** Django ayuda a los desarrolladores evitar varios errores comunes de seguridad al proveer un framework que ha sido diseñado para "hacer lo correcto" para proteger el sitio web automáticamente. Por ejemplo, Django, proporciona una manera segura de administrar

cuentas de usuario y contraseñas, evitando así errores comunes como colocar informaciones de sesión en cookies donde es vulnerable.

- **Escalable:** Tiene en cuenta una clara separación entre las diferentes partes significa que puede escalar para aumentar el tráfico al agregar hardware en cualquier nivel, servidores de cache, servidores de bases de datos o servidores de aplicación.
- **Mantenible:** El código de Django está escrito usando principios y patrones de diseño para fomentar la creación de código mantenible y reutilizable. En particular, utiliza el principio No te repitas "Don't Repeat Yourself" (DRY) para que no exista una duplicación innecesaria, reduciendo la cantidad de código.
- **Portable:** Django está escrito en Python, el cual se ejecuta en muchas plataformas. Lo que significa que no está sujeto a ninguna plataforma en particular, y puede ejecutar sus aplicaciones en muchas distribuciones de Linux, Windows y Mac OS X [34].

d) PyCharm

Es un IDE o entorno de desarrollo integrado multiplataforma utilizado para desarrollar en el lenguaje de programación Python. Proporciona análisis de código, depuración gráfica, integración con VCS / DVCS y soporte para el desarrollo web con Django, entre otras bondades. PyCharm es desarrollado por la empresa JetBrains y debido a la naturaleza de sus licencias tiene dos versiones, la Community que es gratuita y orientada a la educación y al desarrollo puro en Python y la Professional, que incluye más características como el soporte a desarrollo web con varios precios [35].

Características:

- Resaltado de sintaxis, sangría automática y formateo de código, finalización de código, formateador de código, resaltado de errores sobre la marcha, análisis del código a medida que se escribe y detector de código duplicado.
- Depurador gráfico en local.

- Refactorizaciones rápidas y seguras válidas para nombres de fichero, funciones, constantes, clases, propiedades, métodos, parámetros y variables locales y globales.
- Busca en todas partes, encontrar usos de cualquier símbolo, ir a clase/archivo/símbolo, ir a declaración, marcadores personalizados.
- Integración con sistemas de control de versiones: Git, Mercurial, CVS, Subversion, GitHub.
- Integración con documentación: reStructuredText y Google.
- Integración con PyQt y PyGTK.
- Gestión de paquetes.
- Integración nativa con Virtualenv/Buildout.
- Consola de Python.
- Soporte para XML, HTML, YAML, JSON, RelaxNG.
- Terminal local.

e) Bootstrap

Bootstrap, es un framework originalmente creado por Twitter, que permite crear interfaces web con CSS y JavaScript, cuya particularidad es la de adaptar la interfaz del sitio web al tamaño del dispositivo en que se visualice. Es decir, el sitio web se adapta automáticamente al tamaño de una PC, una Tablet u otro dispositivo [36].

Esta técnica de diseño y desarrollo se conoce como Responsive Design o Diseño Adaptativo

Este Framework te evita el tener que preocuparte por las medias queries y los porcentajes en tus CSS para hacer una web Responsive, facilitando la programación del sitio.

Además, se basa en la simplicidad de sus interfaces, lo cual es una tendencia del mercado, en las que tiende a diseño plano, botones grandes, etc; para facilitar la usabilidad en los dispositivos más pequeño, atendiendo a la Experiencia de Usuario.

Actualmente Bootstrap es uno de los Framework más usados.

Ventajas

- Se puede diseñar una web jugando con sus elementos compuestos por diferentes combinaciones de HTML, CSS y Javascript.
- Utiliza HTML5, CSS3, jQuery o GitHub, entre otros.
- Incluye Grid system: muy útil para maquetar por columnas.
- Sus plantillas son de sencilla adaptación responsive.
- Se integra con librerías JavaScript.
- Usa Less: un lenguaje de las hojas de estilo CSS preparado para enriquecer los estilos de la web.
- Cuenta con una documentación completa fácil de comprender.
- Facilita enormemente el diseño de interfaces y además incluye por defecto una plantilla bastante optimizada.

Desventajas

- Se recomienda trabajar con Bootstrap desde el inicio de un proyecto, ya que si quieres incluir el framework en un trabajo ya iniciado algunos estilos podrían “romperse” y se tendría que ajustar a como se tenía en un principio, y eso puede ser un poco tedioso y molesto.
- Es complicado, cambiar de versión si has realizado modificaciones profundas.
- No es ligero, además, para algunas funcionalidades, será necesario tener que usar Javascript y jQuery.
- El diseño se debe adaptar a un grid de 12 columnas, que se modifican según el dispositivo. Es un posible problema puesto que Bootstrap por defecto trae anchos, márgenes y altos de línea, que no permiten realizar cambios específicos [37].

f) Proceso de extracción, transformación y carga (ETL)

Extracción, transformación y carga (ETL) es una canalización de datos que se utiliza para recopilar datos de varios orígenes, transformar los datos según las reglas de negocio y cargarlos en un almacén de datos de destino. El trabajo de transformación en ETL tiene lugar en un motor especializado y, a menudo, implica el uso de tablas de almacenamiento temporal para conservar los datos temporalmente a medida que estos se transforman y, finalmente, se cargan en su destino [Fig.7].

La transformación de datos que tiene lugar a menudo conlleva varias operaciones como filtrado, ordenación, agregación, combinación de datos, limpieza de datos, des duplicación y validación de datos [38].

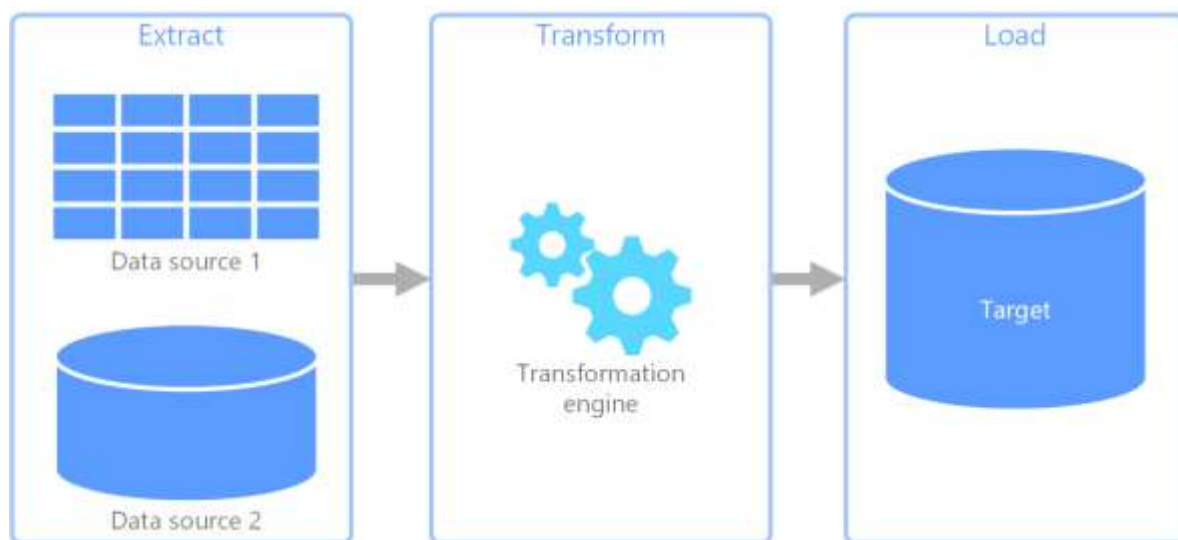


Fig. 7. Proceso de transformación y extracción (ETL)

Fuente: Limpieza y preparación de datos, 2017, enlace <https://www.institutoamericano.es/limpieza-y-preparacion-de-datos-en-microsoft-azure/>

Frecuentemente, las tres fases del proceso ETL se ejecutan en paralelo para ahorrar tiempo. Por ejemplo, mientras se extraen datos, puede que esté funcionando un proceso de transformación sobre los datos ya recibidos y de preparación para la carga, y puede que empiece a funcionar un proceso

de carga sobre los datos preparados, en lugar de tener que esperar a que termine todo el proceso de extracción.

g) Extracción, carga y transformación (ELT)

Extracción, carga y transformación (ELT) difiere de ETL solo por la ubicación en la que se realiza la transformación. En la canalización de ELT, la transformación se produce en el almacén de datos de destino. En lugar de usar un motor de transformación independiente, las funcionalidades de procesamiento del almacén de datos de destino se utilizan para transformar los datos [Fig.8]. Esto simplifica la arquitectura ya que permite quitar el motor de transformación de la canalización. Otra ventaja de este enfoque es que al escalar el almacén de datos de destino también se escala el rendimiento de la canalización de ELT. No obstante, ELT solo funciona bien si el sistema de destino tiene la suficiente potencia para transformar los datos de forma eficaz [39].

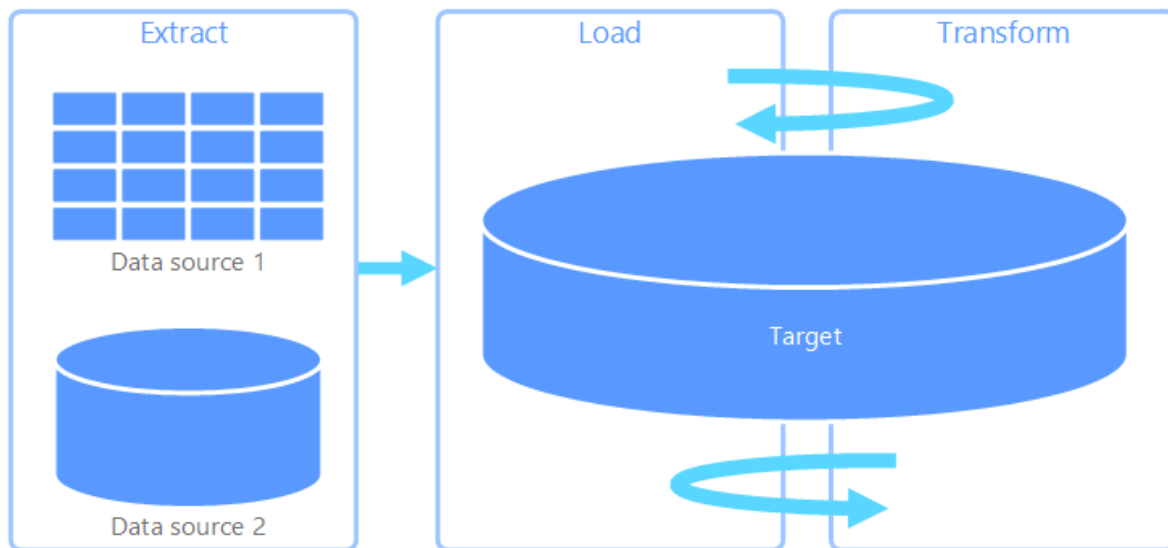


Fig. 8. Proceso de transformación y extracción (ELT)

Fuente: Extracción, carga y transformación (ELT), 2017, enlace <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl>

El punto clave de ELT es que el almacén de datos que se usa para realizar la transformación es el mismo almacén de datos en el que se consumen en última instancia los datos. Este almacén de

datos lee directamente desde el almacenamiento escalable, en lugar de cargar los datos en su propio almacenamiento propietario. Este enfoque omite el paso de copia de datos que se da en ETL. Esta es una operación que lleva mucho tiempo en el caso de conjuntos de datos grandes.

III. METODOLOGÍA

La investigación fue de tipo descriptivo bajo el enfoque cuantitativo. Para la construcción de la herramienta web Saber Expo BI 5de limpieza, transformación y visualización de datos se utilizó la metodología ágil de desarrollo Scrum el cual mediante los Sprints permite realizar una investigación ágil y eficiente. Para la herramienta de inteligencia de negocios se propuso 5 Sprints que se profundizan posteriormente.

A. SPRINT 1: PLANEACIÓN

Para la etapa de planeación fue necesario investigar conceptos indispensables para el desarrollo del proyecto, conceptos como mercado de datos, inteligencia de negocios, Dashboard y cuadros de mandos, análisis multidimensional; conocimiento con el cual se puede interpretar los datos proporcionados por los repositorios de base de datos del ICFES.

B. SPRINT 2: DISEÑO

Para el diseño de la herramienta fue necesario la construcción de una bodega de datos, datos que son proporcionados directamente desde la plataforma ICFES; estos repositorios de datos cuentan con 7 carpetas en las cuales se encuentran archivos .zip que contienen archivos planos en los cuales reposan información de pruebas, estudiantes, municipios e instituciones de toda Colombia. Para el diseño se tiene en cuenta tanto los tipos de gráficos óptimos para la interpretación de la información contenida en los repositorios como la previa construcción de la bodega de datos de los años 2014, 2015, 2016.

1) Definición de Roles Scrum

Para el desarrollo de la aplicación se aplicó la metodología SCRUM previamente mencionada. Para esto se definieron diferentes roles de trabajo, esto con el fin de hacer un trabajo conjunto, en equipo, con la finalidad de obtener un mejor resultado en cuanto a calidad y eficiencia, para este proyecto se definieron los siguientes roles:

- **Scrum Master:** Alexander Santander Arias encargado de que se cumpla con los Sprints propuestos para el proyecto.
- **Scrum Team:** Segundo Dario Ceballos, desarrollador encargado de la parte técnica en Django y Python.
- **Product Owner:** Grupo De Investigación Aplicada En Sistemas (GRIAS)
- **Stakeholder:** La persona, grupo o asociación interesada en usar la aplicación, ya sea usuario general o usuario administrador.
 - **Usuario Administrador:** Usuario que cuenta con privilegios especiales dentro de la aplicación que solo él puede visualizar como la carga de nueva data al banco de datos del aplicativo y crear usuarios.
 - **Usuario General:** Usuario que puede visualizar y consultar información sobre las pruebas saber 5°.

2) Historias de usuario

Una historia de usuario es una explicación general e informal de una función de software, escrita desde la perspectiva del usuario final o cliente, para esto se entrevistó a dos instituciones ubicadas en el municipio de Ipiales, Centro Educativo El Charco y la Institución Educativa Agroindustrial los Pastos, con las que se logró articular elementos para la construcción de las historias de usuario.

TABLA I
HISTORIA DE USUARIO CREAR CUENTA.

HISTORIA DE USUARIO N°1	
ID: HU-001	Usuario: Administrador
Nombre de la historia: Crear nueva cuenta	
Descripción: Se necesita crear una cuenta para ser administrador	
Resultado: Crear una cuenta administrador	Restricción: Ser usuario registrado

TABLA II
HISTORIA DE USUARIO INICIAR SESIÓN.

HISTORIA DE USUARIO N°2	
ID: HU-002	Usuario: Administrador
Nombre de la historia: Iniciar sesión	
Descripción: Se necesita administrador que inicie sesión para entrar a la aplicación	
Resultado: Acceso a la aplicación	Restricción: Estar registrado

TABLA III
HISTORIA DE USUARIO CERRAR SESIÓN

HISTORIA DE USUARIO N°3	
ID: HU-003	Usuario: Administrador
Nombre de la historia: Cerrar sesión	
Descripción: Se necesita administrador que cierre la aplicación	
Resultado: Terminar sesión	Restricción: Estar registrado

TABLA IV
HISTORIA DE USUARIO RESTABLECER CONTRASEÑA.

HISTORIA DE USUARIO N°4	
ID: HU-004	Usuario: Administrador
Nombre de la historia: Restablecer contraseña	
Descripción: Se necesita recuperar la contraseña olvidada	
Resultado: Nueva contraseña	Restricción: Estar registrado

TABLA V

HISTORIA DE USUARIO INGRESAR COMO INVITADO.

HISTORIA DE USUARIO N°5	
ID: HU-005	Usuario: Administrador, General
Nombre de la historia: Ingresar	
Descripción: Se necesita ingresar a la plataforma	
Resultado: Ingreso a la aplicación	Restricción: Ninguna

TABLA VI

HISTORIA DE USUARIO VISUALIZAR INDICADORES.

HISTORIA DE USUARIO N°6	
ID: HU-006	Usuario: Administrador, general
Nombre de la historia: Visualizar indicadores	
Descripción: Se necesita visualizar datos de gráficos de dashboard	
Resultado: Visualizar Dashboard	Restricción: Ninguna

TABLA VII

HISTORIA DE USUARIO FILTRAR DATOS.

HISTORIA DE USUARIO N°7	
ID: HU-007	Usuario: Administrador, general
Nombre de la historia: Filtrar datos	
Descripción: Se necesita generar gráficos según los indicadores requeridos	
Resultado: Gráficos filtrados	Restricción: Indicador

TABLA VIII

HISTORIA DE USUARIO AGREGAR NUEVO AÑO.

HISTORIA DE USUARIO N°8	
ID: HU-008	Usuario: Administrador
Nombre de la historia: Agregar nuevo	
Descripción: Se necesita agregar nuevos datos al mercado de datos	
Resultado: Nuevos datos Saber 5°	Restricción: Año de subida

TABLA IX.

HISTORIA DE USUARIO SELECCIONAR SEPARADOR.

HISTORIA DE USUARIO N°9	
ID: HU-009	Usuario: Administrador
Nombre de la historia: Seleccionar separador	
Descripción: Se necesita seleccionar separador para los campos del nuevo año	
Resultado: Tabla subida por separador	Restricción: selección de separador

TABLA X

HISTORIA DE USUARIO DESCARGAR GRÁFICA .PDF.

HISTORIA DE USUARIO N°10	
ID: HU-0010	Usuario: Administrador, general
Nombre de la historia: Descargar grafica pdf	
Descripción: Se necesita descargar grafica consultada	
Resultado: Archivo pdf de la grafica	Restricción: Ninguna

TABLA XI

HISTORIA DE USUARIO CREAR GRÁFICA GENERAL.

HISTORIA DE USUARIO N°11	
ID: HU-011	Usuario: Administrador, general
Nombre de la historia: Crear grafica general	
Descripción: Se necesita crear grafica general según las métricas requeridas	
Resultado: Grafica filtrada por métricas	Restricción: Depende de las métricas

TABLA XII

HISTORIA DE USUARIO GRÁFICA POR PUNTAJE.

HISTORIA DE USUARIO N°12	
ID: HU-012	Usuario: Administrador
Nombre de la historia: Grafica puntaje	
Descripción: Se necesita que la aplicación grafique por puntaje	
Resultado: Grafica filtrada por puntaje	Restricción: Ninguna

TABLA XIII

HISTORIA DE USUARIO GRÁFICA POR CATEGORÍA.

HISTORIA DE USUARIO N°13	
ID: HU-013	Usuario: Administrador
Nombre de la historia: Grafica categoría	
Descripción: Se necesita que la aplicación grafique por categoría	
Resultado: Grafica filtrada por categoría	Restricción: Ninguna

TABLA XIV

HISTORIA DE USUARIO CUBOS DE DATOS.

HISTORIA DE USUARIO N°14	
ID: HU-014	Usuario: Administrador
Nombre de la historia: Cubos	
Descripción: Se necesita que la aplicación grafique distinto cubos	
Resultado: Grafica de cubos	Restricción: Ninguna

3) Requerimientos funcionales

Los requerimientos funcionales son las descripciones explícitas del comportamiento que debe tener una solución de software referente a la información que debe manejar. Los requerimientos funcionales fueron realizados tomando en cuenta las historias de usuario.

TABLA XV

REQUERIMIENTO FUNCIONAL. EJECUTAR APLICACIÓN.

Código	RF01	Nombre	Ejecutar la aplicación
Característica	El software deberá ejecutarse.		
Descripción	Al momento de ejecutar la aplicación se mostrará la interfaz de usuario		
Prioridad	Alta		
Entradas	Orden de ejecución	Salidas	Login Saber BI 5
Actores	Usuario, administrador		
Precondición	Dirección URL	Postcondición	Inicio de aplicación

TABLA XVI

REQUERIMIENTO FUNCIONAL. INICIAR SESIÓN.

Código	RF02	Nombre	Iniciar sesión
Característica	Requerimiento para iniciar sesión en el aplicativo		
Descripción	Pantalla de logeo para administrador		
Prioridad	Alta		
Entradas	Usuario, contraseña	Salidas	Inicio de sesión pre- registro
Actores	Administrador		
Precondición	Ser usuario registrado	Postcondición	Entrada a página principal del aplicativo

TABLA XVII

REQUERIMIENTO FUNCIONAL. CERRAR SESIÓN.

Código	RF03	Nombre	Cerrar sesión
Característica	Requerimiento para cerrar sesión en el aplicativo		
Descripción	Se cierra la sesión devolviendo al usuario a la pantalla de logeo		
Prioridad	Alta		
Entradas		Salidas	Pantalla de login
Actores	Administrador		
Precondición	Ser usuario registrado	Postcondición	Salir de la aplicación

TABLA XVIII

REQUERIMIENTO FUNCIONAL. CREAR USUARIO.

Código	RF04	Nombre	Crear usuario
Característica	Crea nuevo administrador		
Descripción	Permite crear un usuario administrador para el sistema		

Código	RF04	Nombre	Crear usuario
Prioridad	Alta		
Entradas	Usuario, correo, contraseña	Salidas	Creación de cuenta
Actores	Administrador		
Precondición	No estar registrado	Postcondición	Registrado correctamente

TABLA XIX

REQUERIMIENTO FUNCIONAL. RESTABLECER CONTRASEÑA.

Código	RF05	Nombre	Restablecer contraseña
Característica	Restablece contraseña		
Descripción	Permite al administrador recuperar su contraseña		
Prioridad	Alta		
Entradas	Correo del administrador	Salidas	Contraseña modificada
Actores	Administrador		
Precondición	Ser usuario registrado	Postcondición	Restablece el ingreso

TABLA XX

REQUERIMIENTO FUNCIONAL. ENTRAR COMO INVITADO.

Código	RF06	Nombre	Entrada como invitado
Característica	Permite a cualquier tipo de usuario entrar a la aplicación		
Descripción	Un usuario general puede visualizar los datos		
Prioridad	Alta		
Entradas		Salidas	Visualización de datos
Actores	Usuario general		
Precondición		Postcondición	

TABLA XXI

REQUERIMIENTO FUNCIONAL. AGREGAR NUEVO AÑO.

Código	RF07	Nombre	Agregar nuevo año
Característica	Selecciona las tablas que se desean ingresar		
Descripción	Sube las tablas del nuevo año en que se presentaron las pruebas saber 5 y se le aplica ETL		
Prioridad	Alta		
Entradas	Archivo .txt de la tabla estudiantes y valores plausibles	Salidas	Mensaje de carga exitosa o error en el archivo de carga
Actores	Administrador		
Precondición	Tablas estudiantes y valores plausibles proporcionadas por el ICFES	Postcondición	Subida correcta de los nuevos datos

TABLA XXII

REQUERIMIENTO FUNCIONAL. ESCOGER SEPARADOR.

Código	RF08	Nombre	Escoger separador
Característica	Selecciona el separador para los datos		
Descripción	Escoge el separador de los datos presentes en las tablas		
Prioridad	Alta		
Entradas	Archivo .txt de la tabla estudiantes y valores plausibles	Salidas	Archivo .txt separado por el delimitador elegido
Actores	Administrador		
Precondición	Tablas estudiantes y valores plausibles proporcionadas por el ICFES	Postcondición	Subida correcta de datos delimitados por la opción seleccionada por el admin

TABLA XXIII

REQUERIMIENTO FUNCIONAL. VISUALIZAR DATOS.

Código	RF09	Nombre	Visualizar Datos
Característica	Visualiza datos presentes en la aplicación		
Descripción	Permite observar los indicadores		
Prioridad	Alta		
Entradas		Salidas	Visualización Dashboard
Actores	Administrador, usuario general		
Precondición		Postcondición	Visualización de indicadores

TABLA XXIV

REQUERIMIENTO FUNCIONAL. MOSTRAR DIMENSIONES.

Código	RF10	Nombre	Mostrar dimensiones
Característica	Muestra los atributos de cada dimensión		
Descripción	Se visualiza todos los datos de las dimensiones contenidas en la aplicación		
Prioridad	Media		
Entradas	Seleccionar pestaña dimensiones	Salidas	Los datos que se desean visualizar de las dimensiones
Actores	Administrador, usuario general		
Precondición		Postcondición	Tablas de las dimensiones usadas con sus respectivos filtros

TABLA XXV

REQUERIMIENTO FUNCIONAL. VISTA GLOBAL.

Código	RF11	Nombre	Indicadores pantalla principal
Característica	Datos de interés		
Descripción	Se visualiza datos globales e importantes de interés sobre las pruebas ICFES saber 5°		
Prioridad	Alta		
Entradas		Salidas	Datos importantes y graficas estáticas sobre distintos indicadores
Actores	Administrador, usuario general		
Precondición		Postcondición	Visualización de indicadores

TABLA XXVI

REQUERIMIENTO FUNCIONAL. CREAR GRÁFICA.

Código	RF12	Nombre	Crear gráfica
Característica	Crea grafica según requerimiento de usuario		
Descripción	Permite crear grafica según sea la necesidad del usuario utilizando diferentes tipos de métricas e indicadores		
Prioridad	Alta		
Entradas	Puntaje, categoría, Año, Municipio, Institución, tipo de gráfica	Salidas	Gráfica resultante según los filtros seleccionados
Actores	Administrador, usuario general		
Precondición		Postcondición	Visualización gráfica de indicadores

TABLA XXVII

REQUERIMIENTO FUNCIONAL. DESCARGAR GRÁFICA.

Código	RF13	Nombre	Descargar gráfica
Característica	Genera archivo PDF		
Descripción	Permite generar un archivo PDF de la gráfica		
Prioridad	Alta		
Entradas	Gráfica resultante según los filtros seleccionados	Salidas	Archivo PDF de la gráfica
Actores	Administrador, usuario general		
Precondición	Tener una consulta previa de filtros	Postcondición	Visualización gráfica de indicadores en archivo PDF

TABLA XXVIII

REQUERIMIENTO FUNCIONAL. CUBOS.

Código	RF14	Nombre	Cubos
Característica	Visualiza cubos		
Descripción	Permite visualizar graficas con los indicadores más relevantes del dashboard.		
Prioridad	Alta		
Entradas	Tipo de gráfica	Salidas	Visualización de cubo
Actores	Administrador, usuario general		
Precondición		Postcondición	Visualización gráfica de cubos

C. SPRINT 3: DESARROLLO DE LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS SABER EXPO BI 5

En esta etapa se definieron las distintas funcionalidades con las cuales cuenta la herramienta, para esto se trabajó con distintos paquetes, librerías, entre otras herramientas de creación de software enfocado principalmente a la utilización del framework Django que trabaja con el modelo-plantilla-vista web de alto nivel, muy popular y funcional, tiene como características principales el ser un framework de código abierto, escrito bajo el lenguaje de programación Python; la vista, el modelo y la plantilla se define de la siguiente manera:

1) El modelo

El modelo se considera como la fuente única y definitiva de los datos que posee la aplicación desarrollada, y que permite ejecutar las operaciones. En la arquitectura MTV (Model Template View) se establece que el modelo se encuentra en forma de clases, por lo que en Python se puede indicar y controlar el comportamiento de los datos almacenados.

2) La vista

La vista es considera como el elemento de la aplicación que contiene la lógica de negocio; necesaria para devolver una respuesta hacia el cliente que la solicita, también procesa las peticiones o solicitudes que accederán al modelo para poder entregar u obtener los datos.

3) La plantilla

En la plantilla se debe de decidir la forma en la que se presentarán los datos devueltos por la vista en el navegador web; utilizando estilos CSS o brindando dinamismo a través de JavaScript [40] como se representa en [Fig.9].



Fig. 9. Modelo vista plantilla

Fuente: Configuración de rutas de arquitectura Django, 2018, <https://knepublishing.com/index.php/KnE-Engineering/article/view/5921/11329#Figures>

4) Mercado de datos Saber Expo BI 5

La herramienta de inteligencia de negocios Saber Expo BI 5 cuenta con una base de datos y una estructura de datos definida en 5 dimensiones; en la cual se tienen en cuenta con una base de datos propuesta con los siguientes atributos:

TABLA XXIX

ATRIBUTOS (DIMENSIÓN ESTUDIANTE).

Atributo	Tipo
id_estudiante	integer
estu_consecutivo	character varying
estu_grupo	character varying
estu_n	character varying
estu_estrato	character varying
estu_grado	character varying
estusexo	character varying
estu_disenso	character varying

TABLA XXX.

ATRIBUTOS (DIMENSIÓN INSTITUCIONES).

Atributo	Tipo
id_institucion	integer
ins_codigo_dane	character varying
ins_id_ente	character varying
ins_nombre	character varying
ins_zona	character varying
ins_sector	character varying
ins_tipo_estab	character varying
ins_calendario	character varying
ins_nivel_socio	character varying

TABLA XXXI

ATRIBUTOS (DIMENSIÓN MUNICIPIOS).

Atributo	Tipo
muni_id_ente	integer
muni_nombre	character varying
muni_munexclu	character varying
muni_depar	character varying

TABLA XXXII

ATRIBUTOS (DIMENSIÓN PRUEBA).

Atributo	Tipo
id_prueba	integer
prue_estu_consecutivo	character varying
prue_sn_lenguaje	character varying
prue_sn_matematicas	character varying
prue_sn_ciencias_naturales	character varying
prue_sn_competencias	character varying
des_lenguaje	character varying
des_matematicas	character varying
des_ciencias_naturales	character varying
des_competencias	character varying

TABLA XXXIII

ATRIBUTOS (DIMENSIÓN TIEMPO)

Atributo	Tipo
id_tiempo	integer
prueba	character varying
anio	character varying

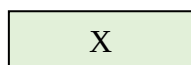
TABLA XXXIV
ATRIBUTOS (DIMENSIÓN FACT SABER 5)

Atributo	Tipo
id_estudiante	integer
muni_id_ente	integer
id_institucion	integer
id_prueba	integer
id_hoja	character varying
numero_estu	integer
puntaje_lenguaje	numeric
puntaje_matematicas	numeric
puntaje_ciencias	numeric
puntaje_competencias	numeric
numero_lenguaje	integer
numero_matematicas	integer
numero_ciencias	integer
numero_competencias	integer
promedio	numeric
Id_tiempo	integer

En las siguientes tablas se presentan los atributos de los Valores plausibles, Entidades territoriales certificadas, Establecimientos y Municipios existentes en los repositorios que comprende los años de 2014 a 2017 que tienen datos indispensables para la agregación de la nueva bodega de datos.

En las siguientes tablas se analizan los atributos que están y no están presentes en cada uno de los repositorios.

Está presente



No está presente



TABLA XXXV

REPOSITORIO DE DATOS (VALORES PLAUSIBLES)

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
estu_consecutivo	X	X	X	X	ID público del estudiante
aplicación	X	X	X	X	No identificado
calendario	X	X	X	X	Calendario del establecimiento
departamento	X	X	X	X	Código DANE del departamento al que pertenece el establecimiento educativo
dissenso	X	X	X	X	Marca de discapacidad cognitiva
enteterr	X	X	X	X	Código DANE de la entidad a la que pertenece el establecimiento educativo
establecimiento	X	X	X	X	Código DANE del establecimiento educativo al que pertenece el estudiante
estrato	X	X	X	X	Estrato al que pertenecen el establecimiento dentro del marco muestral
grado	X	X	X	X	Grado del estudiante
grupo	X	X	X	X	Grupo/salón al que pertenecen los estudiantes
id_hoja	X	X	X	X	Código de identificación de hoja de respuestas
jornada	X	X	X	X	Código de la sede jornada al que pertenece el estudiante
leng_copietas	X	X	X	X	Indicador de copia en lenguaje
leng_score1	X	X	X	X	Score1-valor plausible 1 de lenguaje
leng_score2	X	X	X	X	Score2-valor plausible 2 de lenguaje
leng_score3	X	X	X	X	Score3-valor plausible 3 de lenguaje
leng_score4	X	X	X	X	Score4-valor plausible 4 de lenguaje
leng_score5	X	X	X	X	Score5-valor plausible 5 de lenguaje
leng_weight	X	X	X	X	Peso muestral del lenguaje
mate_copietas	X	X	X	X	Indicador de copia en matemáticas
mate_score1	X	X	X	X	Score1-valor plausible 1 de matemáticas
VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
mate_score2	X	X	X	X	Score2-valor plausible 2 de matemáticas
mate_score3	X	X	X	X	Score3-valor plausible 3 de matemáticas

mate_score4	X	X	X	X	Score4-valor plausible 4 de matemáticas
mate_score5	X	X	X	X	Score5-valor plausible 5 de matemáticas
mate_weight	X	X	X	X	Peso muestral de matemáticas
cien-copietas	X				Indicador de copia de ciencias
cien_score1	X				Score1-valor plausible 1 de ciencias
cien_score2	X				Score2-valor plausible 2 de ciencias
cien_score3	X				Score3-valor plausible 3 de ciencias
cien_score4	X				Score4-valor plausible 4 de ciencias
cien_score5	X				Score5-valor plausible 5 de ciencias
cien_weight	X				Peso muestral de ciencias
comp_copietas		X	X		Indicador de copia de competencias
comp_score1		X	X		Score1-valor plausible 1 de competencias
comp_score2		X	X		Score2-valor plausible 2 de competencias
comp_score3		X	X		Score3-valor plausible 3 de competencias
comp_score4		X	X		Score4-valor plausible 4 de competencias
comp_score5		X	X		Score5-valor plausible 5 de competencias
comp_weight		X	X		Peso muestral de competencias
modeloedu	X	X	X	X	No definido
municipio	X	X	X	X	Código DANE al que pertenece el establecimiento educativo
n	X	X	X	X	Matricula del grupo al que pertenecen
nivel	X	X	X	X	Nivel socioeconómico del establecimiento
sector	X	X	X	X	Sector del establecimiento
sexo	X	X	X	X	Sexo del estudiante
zona	X	X	X	X	Zona de la jornada
zonastab	X	X	X	X	Zona del establecimiento

TABLA XXXVI

REPOSITORIO DE DATOS (ENTIDADES)

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	X	Identificador del ente territorial
nombre	X	X	X	X	Nombre del ente territorial
munexclu	X	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	X	Tipo de ente territorial 1=ETC, 2=DPTO,3= MPIO

TABLA XXXVII

REPOSITORIO DE DATOS (ESTABLECIMIENTOS)

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
cod_dane	X	X	X	X	Código DANE del establecimiento educativo
id_municipio	X	X	X	X	Código DANE del municipio al que pertenece el establecimiento educativo
id_ente	X	X	X	X	id de la entidad territorial
nombre	X	X	X	X	Nombre del establecimiento educativo reportado en el DUE
zona	X	X	X	X	Zona donde la mayoría de la población atendida por el establecimiento educativo se ubica 1=urbano 2=rural
sector	X	X	X	X	Naturaleza administrativa de establecimiento educativo según 1=oficial urbano 2=oficial rural 3=no oficial
tipo_estab	X	X	X	X	Tipo de establecimiento 1=Oficial urbano 2=Oficial rural 3=no oficial
calendario	X	X	X	X	Calendario del establecimiento
nivel_socio	X	X	X	X	NSE asignado de acuerdo a la clasificación realizada con puntajes por medios

TABLA XXXVIII

REPOSITORIO DE DATOS (MUNICIPIOS)

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	X	Identificador del ente territorial
nombre	X	X	X	X	Nombre del ente territorial
munexclu	X	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	X	Tipo del ente territorial posibles valores 3

D. SPRINT 4: CASOS DE PRUEBA

En este sprint se definen los casos de prueba que servirán como pautas esenciales para las distintas funcionalidades de la herramienta con las cuales se podrán realizar por etapas el diseño del sistema.

TABLA XXXIX

CASO DE PRUEBA INICIO DE SESIÓN

ID	CP-01	Nombre	Prueba de Inicio de sesión.
Caso de uso	Iniciar sesión.		
Entrada	Nombre Usuario, contraseña.		
Salida Esperada	- Pantalla inicial Saber BI5. - Error al iniciar sesión.		

TABLA XL

CASO DE PRUEBA CREAR USUARIO

ID	CP-02	Nombre	Prueba crear usuario.
Caso de uso	Crear usuario.		
Entrada	Nombre de usuario, e-mail, contraseña.		
Salida Esperada	• Pantalla inicial Saber BI5. • Su cuenta se creó satisfactoriamente. • Error al crear cuenta.		

TABLA XLI

CASO DE PRUEBA CERRAR SESIÓN

ID	CP-03	Nombre	Prueba cerrar sesión.
Caso de uso	Cerrar sesión.		
Entrada	Ninguna.		
Salida Esperada	• ¿Está seguro que quiere salir de su cuenta? • Sesión finalizada.		

TABLA XLII

CASO DE PRUEBA INGRESAR COMO INVITADO

ID	CP-04	Nombre	Prueba ingreso como invitado.
Caso de uso	Ingresar como invitado		
Entrada	Ninguna.		
Salida Esperada	Pantalla inicial Saber BI5.		

TABLA XLIII

CASO DE PRUEBA RESTABLECER CONTRASEÑA

ID	CP-05	Nombre	Prueba restablecer contraseña.
Caso de uso	Restablecer contraseña		
Entrada	E-mail del usuario, nombre de usuario		
Salida Esperada	• ¡Su contraseña ha sido actualizada con éxito! • Error al restablecer contraseña.		

TABLA XLIV

CASO DE PRUEBA VISUALIZAR INDICADORES

ID	CP-06	Nombre	Prueba de visualización de indicadores
Caso de uso	Visualizar datos.		
Entrada	Indicadores propuestos por el usuario.		
Salida Esperada	Visualización de los indicadores propuestos por el usuario.		

TABLA XLV

CASO DE PRUEBA CREAR GRÁFICA GENERAL

ID	CP-07	Nombre	Prueba de grafica general.
Caso de uso	Crear grafica general.		
Entrada	• Tipo métrica • Tipo de visualización • Puntaje, Institución, Municipio, Año.		
Salida Esperada	Gráfica general.		

TABLA XLVI

CASO DE PRUEBA DESCARGAR GRÁFICA

ID	CP-08	Nombre	Prueba descargar gráfica.
Caso de uso	Descargar gráfica.		
Entrada	Elegir la gráfica que se desea guardar.		
Salida Esperada	Archivo .pdf de la gráfica requerida.		

TABLA XLVII

CASO DE PRUEBA GRÁFICA POR PUNTAJE

ID	CP-09	Nombre	Prueba crear gráfica por puntaje.
Caso de uso	Crear gráfica por puntaje.		
Entrada	• Puntaje • Tipo de grafica • Año • Municipio • Institución		
Salida Esperada	Gráfica por puntaje.		

TABLA XLVIII

CASO DE PRUEBA GRÁFICA POR CATEGORÍA

ID	CP-10	Nombre	Crear gráfica por categoría.
Caso de uso	Crear gráfica por categoría.		
Entrada	• Categoría • Tipo de grafica • Año, Municipio, Institución		
Salida Esperada	Gráfica por categoría.		

TABLA XLIX

CASO DE PRUEBA VER CUBOS

ID	CP-11	Nombre	Prueba de cubos.
Caso de uso	Visualizar cubos.		
Entrada	La entrada del cubo que se desee visualizar.		
Salida Esperada	Vista de los cubos seleccionados.		

TABLA L

CASO DE PRUEBA AGREGAR NUEVO AÑO

ID	CP-12	Nombre	Prueba agregar nuevo año.
Caso de uso	Agregar nuevo año.		
Entrada	• Archivo que se desee agregar formato .txt. • Elegir delimitador de atributos.		
Salida Esperada	• Su archivo se subió correctamente. • Error al subir nuevo archivo.		

E. SPRINT 5: IMPLEMENTACIÓN Y DOCUMENTACIÓN

Como Sprint final se realizó la respectiva documentación con respecto a los manuales de usuario y manuales de desarrollador, en los cuales se explica a fondo cada función que realiza Saber Expo BI 5, esta información fue establecida en los anexos A y B del presente proyecto.

IV. RESULTADOS

A. DISEÑO DEL SISTEMA DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL)

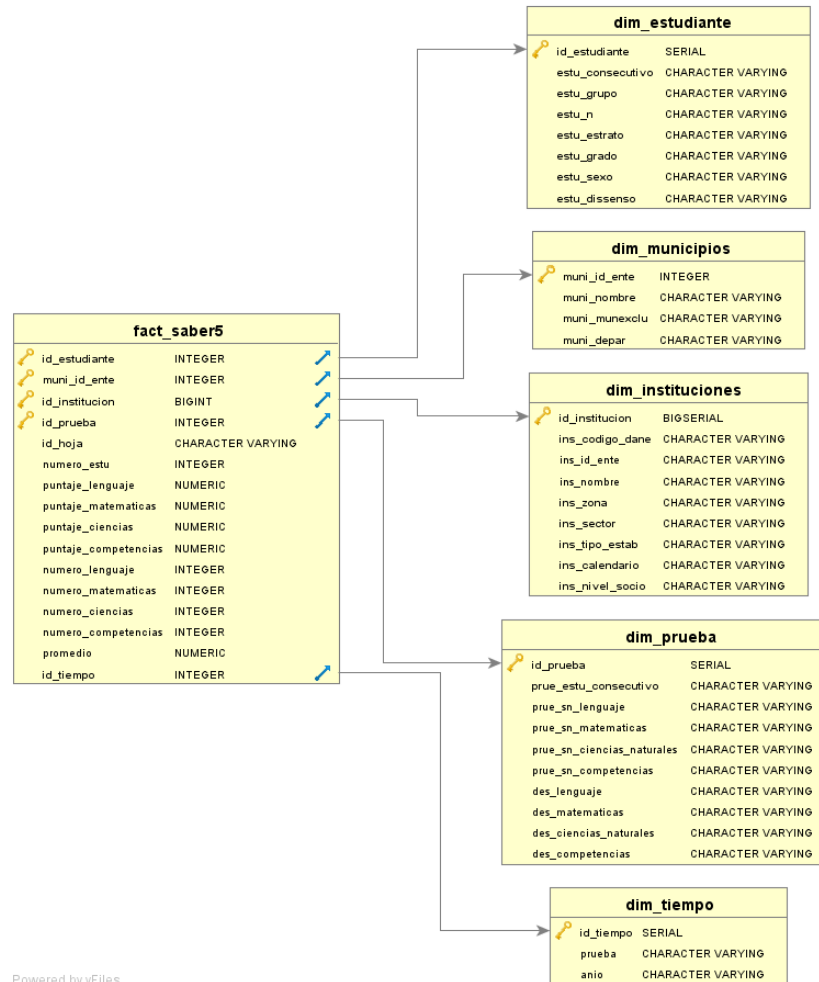
Para trabajar se usó el motor de bases de datos relacionales PostgreSQL Versión 11.5, inicialmente se usaron tres repositorios de datos suministrados por el FTP del ICFES, cada uno con información referente a las pruebas Saber 5° para los años 2014, 2015, 2016 y 2017.

A continuación, se describe el proceso de Extracción, Transformación y Carga para construir y añadir los nuevos datos.

B. DISEÑO MULTIDIMENSIONAL SABER EXPO BI 5

Para el diseño de la aplicación se decidió trabajar con diseño multidimensional con un esquema de estrella, este tipo de esquema de base de datos relacional consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

En [Fig.10], se muestra un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones. Las ramas situadas al final de los enlaces que conectan las tablas indican una relación de muchos a uno entre la tabla de hechos y cada tabla de dimensiones.



Powered by yFiles

Fig. 10. Diseño multidimensional Saber Expo BI 5

Fuente: Propia, 2021

1) Fase de extracción

El propósito general de esta fase fue construir un repositorio de datos limpio y estandarizado a partir de los datos fuentes, entregados para tal fin se desarrollaron tres guías de trabajo.

2) Integración de Datos

Para la creación del repositorio Saber Expo BI 5, fue necesario revisar los diccionarios de datos e identificar los atributos de cada una de ellas ya que no existía un 100% de coincidencia entre el documento y las columnas de las tablas.

TABLA LI
TOTAL DE REGISTROS

Repositorio	No. de registros
Valores_plausibles_2014_2017	2'938.315
Entidades	95
Establecimientos	15.786
Municipios	1.060
Estudiantes_2017	776.436
TOTAL REGISTROS	3'731.692

3) Análisis de la calidad de datos

Análisis calidad de datos de la bodega para identificar el número de registros nulos, no nulos y distintos cuenta con un total de **3'731.692** registros, se tuvo en cuenta las 5 dimensiones con el propósito de realizar un análisis de la calidad de los datos e identificar que los datos contenidos sean fiables.

TABLA LII

ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN PRUEBA)

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_prueba	13875	0	13875	"40" "41" "42" "43" "44"
2	prue_estu_consecutivo	13875	0	13875	"V201450046303" "V201450046304" "V201450046306" "V201450046309"
3	prue_sn_lenguaje	9056	4819	2	"CON_COPIA" "NO_COPIA"
4	prue_sn_matematicas	9170	4705	2	"CON_COPIA" "NO_COPIA"
5	prue_sn_ciencias	6051	7824	2	"CON_COPIA" "NO_COPIA"
6	prue_sn_competencias	3011	10864	2	"CON_COPIA" "NO_COPIA"
7	des_lenguaje	9056	4819	0	'INSUFICIENTE' 'MINIMO' 'SATISFACTORIO' 'AVANZADO'
8	des_matematicas	9170	4705	0	'INSUFICIENTE' 'MINIMO' 'SATISFACTORIO' 'AVANZADO'

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
9	des_ciencias_naturales	6051	7824	0	'INSUFICIENTE' 'MINIMO' 'SATISFACTORIO' 'AVANZADO'
10	des_competencias	3011	10864	0	'INSUFICIENTE' 'MINIMO' 'SATISFACTORIO' 'AVANZADO'

TABLA LIII
ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN MUNICIPIO)

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	muni_id_ente	13	0	13	"52022" "52210" "522152" "52224" "52227"
2	muni_nombre	13	0	13	"ALDANA" "CONTADERO" "CORDOBA" "CUASPUD" "CUMBAL "
3	muni_munexclu	13	0	2	"ETC" "SIN_DATO"
4	muni_depar	13	0	1	"NARINO"

TABLA LIV

ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN ESTUDIANTE)

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_estudiante	13875	0	13875	"50" "51" "52" "53"
2	estu_consecutivo	13875	0	13875	"V201450046302" "V201450046303" "V201450046304" "V201450046305"
3	estu_grupo	13875	0	3	"01" "02" "99"
4	estu_n	13875	0	112	"1" "10" "100" "101"
5	estu_estrato	13875	0	12	99E1209 99E1219 99E9999 99O1109
6	estu_grado	13875	0	1	"5"
7	estu_sexo	13875	0	3	FEMENINO MASCULINO NO_ESPECIFICA
8	estu_disenso	13875	0	2	"N" "S"

TABLA LV

ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN INSTITUCIÓN)

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_institucion	358	0	358	"39" "40" "41" "42"
2	ins_codigo_dane	358	0	358	"152022000084" "152210000261" "152215000138" "152224000019"
3	ins_id_ente	358	0	2	48 89
4	ins_nombre	358	0	336	CENTRO EDUCATIVO CASA FRIA CENTRO EDUCATIVO ARVELA
5	ins_zona	358	0	2	RURAL URBANO
6	ins_sector	358	0	2	NO_OFICIAL OFICIAL
7	ins_tipo_estab	358	0	3	NO_OFICIAL OFICIAL_RURAL OFICIAL_URBANO
8	ins_calendario	358	0	1	A
9	ins-nivel_socio	358	0	5	1 2 3 4

TABLA LVI

ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN TIEMPO)

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	Id_tiempo	13875	0	0	"52022"
					"52210"
					"522152"
					"52224"
					"52227"
2	prueba	13875	0	0	"V201450046302"
					"V201450046303"
					"V201450046304"
					"V201450046305"
					"V201450046306"
3	anio	13875	0	0	"2014"
					"2015"
					"2016"
					"2017"

TABLA LVII

ANÁLISIS DE CALIDAD DE DATOS (DIMENSIÓN FACT_SABER5)

No.	Atributos	No nulos	Nulos	Valores distintos	Promedio	Desviación	Valores máximos	Valores mínimos
1	id_estudiante	13875	0	"35"				
				"36"				
				"37"				
				"38"				
				"39"				

No.	Atributos	No nulos	Nulos	Valores distintos	Promedio	Desviación	Valores máximos	Valores mínimos
2	muni_id_ente	13875	0	"52022" "52210" "52215" "52224" "52227"				
3	id_institucion	13875	0	"150" "151" "163" "167" "168"				
4	id_prueba	13875	0	"1" "55" "60" "70" "100"				
5	id_hoja	13875	0	"V201450046302" "				
6	numero_estu	13875	0	1				
7	puntaje_lenguaje	9056	4819	"1" "1,2" "1,25" "1,33" "1,38"	0.97	0.76	4	0
8	puntaje_matematicas	9170	4705	"146" "152" "152,67" "152,67" "153" "154"	0.98	0.77	4	0

No.	Atributos	No nulos	Nulos	Valores distintos	Promedio	Desviación	Valores máximos	Valores mínimos
9	puntaje_ciencias	6051	7824	"1,45" "146" "147" "148" "148,78"	0.53	0.77	4	0
10	puntaje_competencias	3011	10864	"150,104" "150,943" "152" "163"	0.28	0.64	3.1	0
11	numero_lenguaje	13875	0	"0" "1"				
12	numero_matematicas	13875	0	"0" "1"				
13	numero_ciencias	13875	0	"0" "1"				
14	numero_competencias	13875	0	"0" "1"				
15	promedio	13875	0	"1,45" "146" "147" "148"				
16	Id_tiempo	13875	0	"1" "55" "60" "70" "100"				

4) Limpieza de datos

El proceso de limpieza de datos incluye la validación y corrección de los mismos, generando un mercado de datos confiable; para esto se indaga sobre información incompleta o con errores, puesto que pueden crear conclusiones falsas lo que puede ocasionar problemas en el proceso de toma de decisiones y brinda ventajas tales como:

- Minimiza el riesgo de los proyectos, especialmente en los relacionados con Tecnologías de la Información.
- Ahorro de tiempo y recursos, haciendo un mejor uso de la infraestructura tecnológica y sistemas para explotar su información.
- Toma de decisiones de negocio oportunas, en base a información confiable, validada y limpia.

Partiendo de una bodega de datos previamente tratada, el proceso de limpieza de datos se aplicará a la tabla valores_plausibles2017, para posteriormente ser agregada al datamart en los que se comprende datos de los años 2014, 2015, 2016.

5) Transformación de datos

En la fase de transformación es donde se produce la operación más crítica, el objetivo de esta fase fue construir un repositorio estandarizado y con datos de calidad que fuesen usado como fuente de carga del datamart, este procedimiento se realizó en dos fases:

- **Transformación y normalización:** Se definen los datos que necesarios, con los cuales se estandarizan, se dan formato y se almacenan para posteriormente definir las etapas sucesivas; en el proyecto se aplica a la tabla valores_plausibles2017, en el año 2017 en adelante se cambió la forma en la cual el puntaje se expresaba, simplemente cambiando su forma decimal por ejemplo un puntaje de 0.32 en 2016 se expresaría como 3.2 en 2017 por lo que los puntajes de los años anteriores fueron multiplicados x 10 para mover el punto decimal y que los nuevos puntajes de 2017 y posteriores no presentaran ningún problema al ser agregados a la tabla.

- **Eliminación de datos:** Se excluyen y/o eliminan los datos redundantes sobre la cual se efectuaron rutinas para eliminar atributos que no son requeridos en el Datamart. Finalmente se obtuvo una tabla de 13875 filas.

6) Carga de datos

Para el proceso de carga de los datos en el mercado de datos, solo se consideraron aquellos registros asociados a las instituciones de la Subregión de Obando a partir de las dimensiones cargadas.

El paso siguiente fue cargar los datos desde temp_plausibles. A cada una de las tablas del mercado de datos (5 dimensiones y 1 tabla de hechos).

El proceso anterior se realiza para la función “nuevo año”. Para esto, es necesario realizar la limpieza de estos datos, elaborar un nuevo proceso de transformación y eliminación, antes de importar las bases de datos directamente de los archivos descargados del ICFES, para que puedan coincidir con la estructura de la nueva base de datos.

C. DISEÑO DE REPOSITORIO DE DATOS

El repositorio de datos de la aplicación se compone de un esquema el cual está conformado por 5 dimensiones y 1 tabla de hechos, realizando un diccionario de datos asociados a cada dimensión del diseño multidimensional, que fueron necesarios para el diseño del sistema de extracción, transformación y carga (ETL).

1) Diccionario de Datos

TABLA LVIII

DICCIONARIO DE DATOS (DIM_ESTUDIANTES)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	id_estudiante	Llave primaria identificador del estudiante	1,2,3...	SERIAL
1	estu_consecutivo	ID público del estudiante	VAAAAGXXXXXXXXX donde V = constante, AAAA = Año, G = grado y XXXXXXXX = consecutivo	CHARACTER VARYING
2	estu_grupo	Salón al que pertenece los estudiantes	99 = censal, <>99 = control	CHARACTER VARYING
3	estu_n	Matricula grupo al que pertenece	1,2,3...	CHARACTER VARYING
4	estu_estrato	Estrato al que pertenece el establecimiento		CHARACTER VARYING
5	estu_grado	Grado del estudiante	5	CHARACTER VARYING
6	estu_sexo	Sexo del estudiante	1=Masculino, 2=Femenino, 3=No especifica	CHARACTER VARYING
7	estu_disenso	Marca de discapacidad cognitiva	0 = sin discapacidad, 1 = con discapacidad	CHARACTER VARYING

TABLA LIX
DICCIONARIO DE DATOS (DIM_MUNICIPIO)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	muni_id_ente	Identificador del municipio	Llave tabla dim_municipios	INTEGER
1	muni_nombre	Nombre del municipio	Cadena	CHARACTER VARYING
2	muni_nunexclu	Identificador de certificación del municipio	1=ETC, 2=DPTO, 3=SIN_DATO	CHARACTER VARYING
3	Muni_depar	Nombre del departamento	Cadena	CHARACTER VARYING

TABLA LX
DICCIONARIO DE DATOS (DIM_INSTITUCION)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	Id_institucion	Llave primaria identificador de la institución	1,2,3...	SERIAL
1	Ins_Codigo_Dane	Código Dane de la institución	Número (12)	CHARACTER VARYING
2	Ins_id_ente	Código DANE de la entidad a la que pertenece establecimiento educativo	Número (2)	CHARACTER VARYING
3	Ins_nombre	Nombre de la institución	Cadena	CHARACTER VARYING
4	Ins_zona	Zona donde se ubica la mayoría de la población atendida	1 = Urbana, 2 = Rural	CHARACTER VARYING
5	Ins_sector	Sector del establecimiento	1 = Oficial, 2 = No oficial	CHARACTER VARYING
6	Ins_tipo_esta	Sector y zona de establecimiento	1=OFICIAL_URBANO, 2=OFICIAL_RURAL, 3=NO_OFICIAL	CHARACTER VARYING

	Atributo	Descripción	Tipo Campo	Tipo Atributo
7	Ins_calendario	Calendario del establecimiento	A, B	CHARACTER VARYING
8	Ins_nivel_socio	Nivel socioeconómico del establecimiento	1, 2, 3, 4, 5	CHARACTER VARYING

TABLA LXI

DICCIONARIO DE DATOS (DIM_PRUEBA)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	Id_prueba	Llave primaria identificador de la prueba	1,2,3...	SERIAL
1	Estu_Consecutivo	Id público del estudiante	VAAAAGXXXXXXXXX donde V = constante, AAAA = Año, G = grado y XXXXXXXX = consecutivo	CHARACTER VARYING
2	sn_lenguaje	Número de copias en lenguaje	0=NO_COPIA, 1=COPIA	CHARACTER VARYING
3	sn_matematicas	Número de copias en matemáticas	0=NO_COPIA, 1=COPIA	CHARACTER VARYING
4	sn_ciencias_naturales	Número de copias en ciencias naturales	0=NO_COPIA, 1=COPIA	CHARACTER VARYING
5	sn_competencias	Número de copias en competencias	0=NO_COPIA, 1=COPIA	CHARACTER VARYING
6	des_lenguaje	Desempeño de lenguaje	0 a 1=INSUFICIENTE 1,1 a 2=MINIMO 2,1 a 3=SATISFACTORIO 3,1 a 4=AVANZADO	NUMERIC
7	des_matematicas	Desempeño de matemáticas	0 a 1=INSUFICIENTE 1,1 a 2=MINIMO 2,1 a 3=SATISFACTORIO 3,1 a 4=AVANZADO	NUMERIC

	Atributo	Descripción	Tipo Campo	Tipo Atributo
8	des_ciencias_naturales	Desempeño de ciencias naturales	0 a 1=INSUFICIENTE 1,1 a 2=MINIMO 2,1 a 3=SATISFACTORIO 3,1 a 4=AVANZADO	NUMERIC
9	des_competencias	Desempeño de competencias	0 a 1=INSUFICIENTE 1,1 a 2=MINIMO 2,1 a 3=SATISFACTORIO 3,1 a 4=AVANZADO	NUMERIC

TABLA LXII

DICCIONARIO DE DATOS (DIM_TIEMPO)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	Id_tiempo	Llave primaria identificador del tiempo	Llave tabla dim_tiempo	INTEGER
1	prueba	Id público del estudiante	VAAAAGXXXXXXXXX donde V = constante, AAAA = Año, G = grado y XXXXXXXX = consecutivo	CHARACTER VARYING
2	anio	Año de presentación de la prueba	2014, 2015, 2016, 2017	CHARACTER VARYING

TABLA LXIII

DICCIONARIO DE DATOS (FACT_SABER5)

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	Id_estudiante	Llave foránea	1,2,3...	INTEGER
PK	Muni_id_ente	Llave foránea	Número (5)	INTEGER

	Atributo	Descripción	Tipo Campo	Tipo Atributo
PK	Id_institucion	Llave foránea	1,2,3...	INTEGER
PK	Id_prueba	Llave foránea	1,2,3...	INTEGER
1	Id_hoja	Código de identificación de hoja de respuestas	VAAAAGXXXXXXXXX donde V = constante, AAAA = Año, G = grado y XXXXXXXX = consecutivo	TEXT
2	Numero_estu	Matrícula del grupo al que pertenecen	1,2...	INTEGER
3	puntaje_lenguaje	Puntaje de lenguaje	Número (6)	NUMERIC
4	puntaje_matematicas	Puntaje de matemáticas	Número (6)	NUMERIC
5	puntaje_ciencias	Puntaje de ciencias	Número (6)	NUMERIC
6	puntaje_competencias	Puntaje de competencias	Número (6)	NUMERIC
7	Numero_lenguaje	Presento o no la prueba de lenguaje	0,1	INTEGER
8	Numero_matematicas	Presento o no la prueba de lenguaje	0,1	INTEGER
9	Numero_ciencias	Presento o no la prueba de lenguaje	0,1	INTEGER
10	Numero_competencias	Presento o no la prueba de lenguaje	0,1	INTEGER
11	promedio	Promedio global de todas las competencias	Número (1,2)	NUMERIC
12	Id_tiempo	Llave foránea	1,2,3...	INTEGER

D. ARQUITECTURA

SABER EXPO BI-5, permite al usuario enviar una dirección URL por medio del navegador y así buscar la respectiva ruta en un mapa de URLs, aquí cada dirección está asociada a una vista en la cual si necesita algún dato consulta al modelo y de igual forma este realiza una consulta a la base de datos donde podemos encontrar la información de las pruebas saber 5 de los diferentes años, después se envía los datos a una plantilla con la consulta lista , mostrando al usuario en forma entendible para él, en una gráfica o tabla para su respectivo análisis [Fig.11].

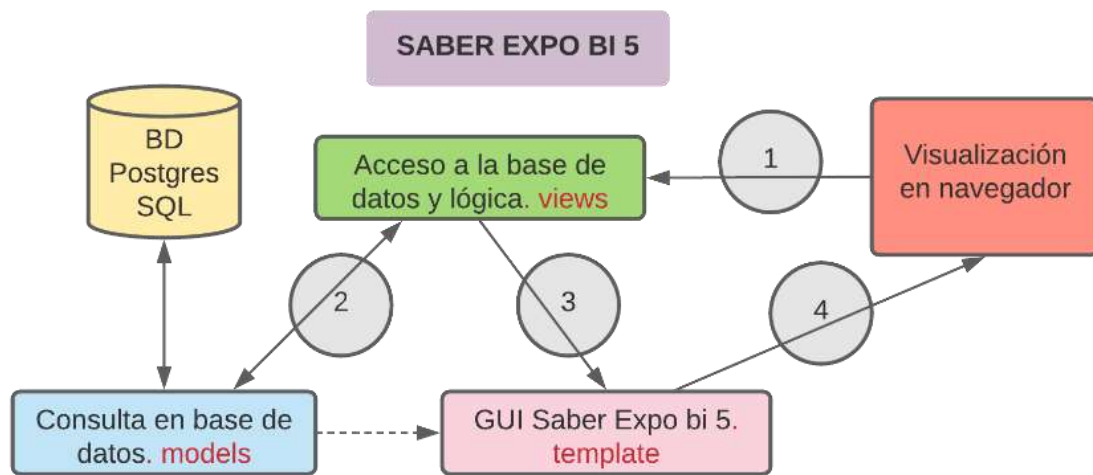


Fig. 11. Saber Expo BI 5

Fuente: Propia, 2021

Se maneja una arquitectura MTV (Model-Template-View), con ello se corre un servidor que permite cargar en el navegador la aplicación y poder realizar las respectivas consultas a la base de datos manejando formato JSON para agilizar la carga de los datos y con ello visualizar de manera óptima los datos en una plantilla con su respectiva grafica o tabla; el Navegador manda una solicitud, luego la vista interactúa con el modelo para obtener datos mientras la vista llama a la plantilla, esta plantilla renderiza la respuesta a la solicitud del navegador y la muestra, para comprender de una forma clara se puede observar [Fig.12].

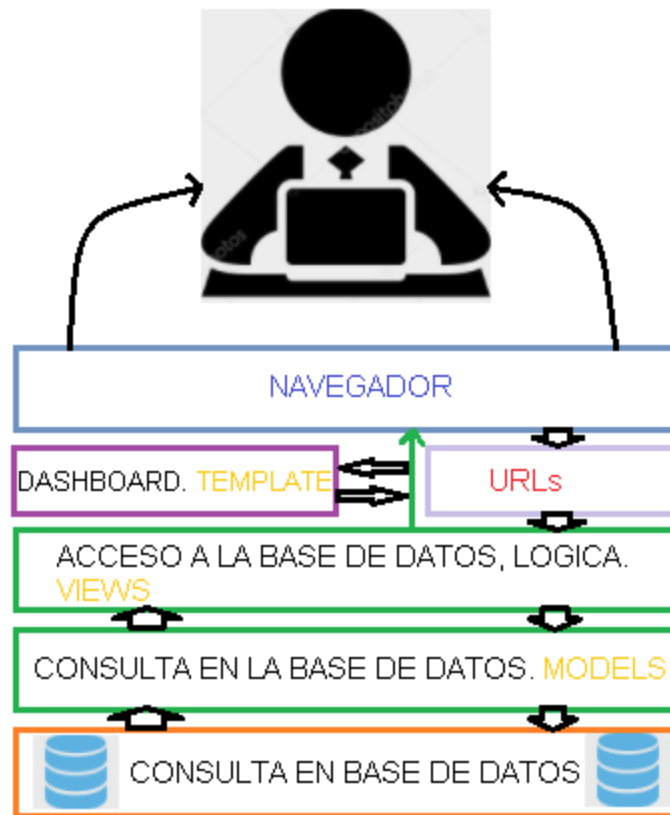


Fig. 12. Interacción de la aplicación con la arquitectura MTV

Fuente: Que es el patrón MTV, 2017, enlace <https://espifreelancer.com/mtv-django.html>

1) Diseño estructural

Aquí se explica las características estructurales de un sistema general, la arquitectura y distribución de artefactos de software requerida.

2) Diagrama de paquetes

En el desarrollo de la aplicación se maneja Python con un Framework el cual es un entorno Web llamado Django y nos permite trabajar con una estructura MVT (Modelo Vista Témplate).

A continuación, se detalla el manejo de dicha estructura:

a) M = Modelos (Base de datos)

Capa de interacción y comunicación que muestra un archivo donde podemos observar las clases y detalles de las diferentes dimensiones que se manipula en el proyecto.

b) V = Vista (Funciones)

Permite interactuar con el modelo y es donde encontramos la lógica con la cual realizamos las diferentes operaciones para generar los diferentes resultados.

c) T = Templates (Plantillas)

Son los diferentes archivos que se encuentran en la capa de presentación y nos permite mostrar los resultados a los diferentes usuarios.

A continuación, se observa las rutas o directorios donde se encuentran cada uno de los diferentes archivos de configuración para el desarrollo de la aplicación.

La imagen nos permite observar de forma general los directorios con sus respectivos nombres en los cuales se encuentran las Plantillas, Vistas, Modelos y Scripts además archivos de configuración para su conexión entre los diferentes archivos utilizados en la aplicación [Fig.13].

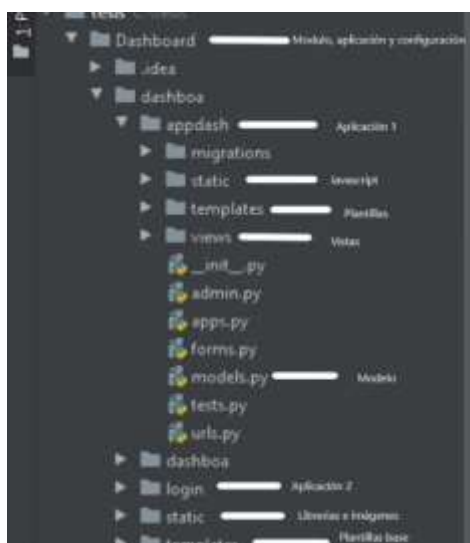


Fig. 13.Model template view Expo Saber BI 5

Fuente: Propia, 2021

En la siguiente imagen podemos visualizar los diferentes directorios donde se encuentran todos los archivos Scripts para cada plantilla los cuales permite recibir los datos que vienen desde el modelo conectado con la vista y la BD [Fig.14].

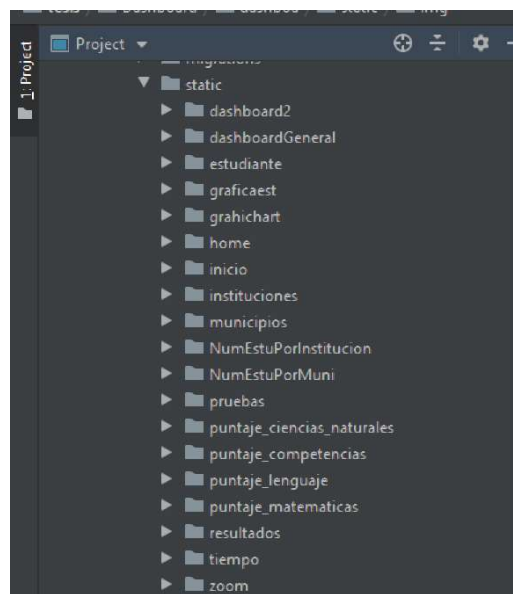


Fig. 14. Directorios de plantillas de recepción de datos

Fuente: Propia, 2021

En las dos siguientes imágenes podemos observar desplegados los diferentes directorios que contienen cada uno de los archivos Scripts (JS) que son utilizados con cada plantilla y vista para la visualización de los datos [Fig.15].

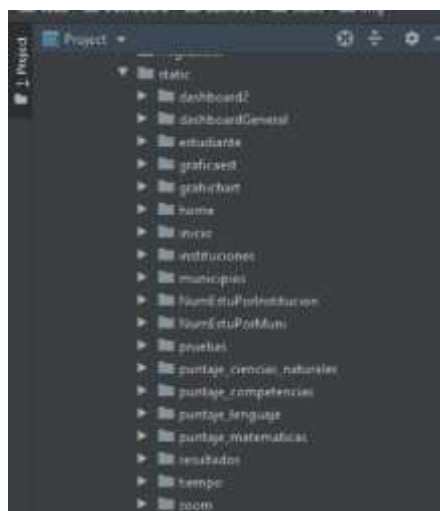


Fig. 15. Directorio de Scripts (JS)

Fuente: Propia, 2021

En la siguiente grafica se encuentran las plantillas utilizadas para la visualización de las gráficas y tablas para el usuario [Fig.16].

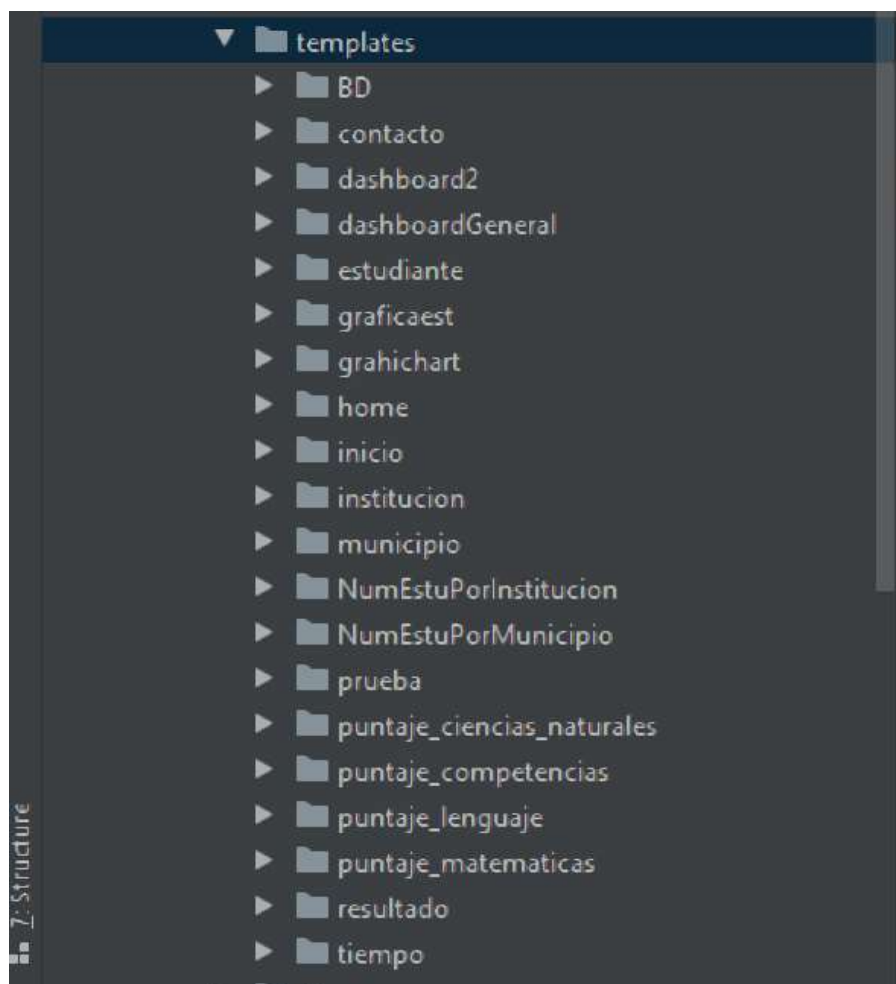


Fig. 16. Directorio de plantillas utilizadas para gráficas y tablas

Fuente: Propia, 2021

Ahora se visual los diferentes directorios desplegados con sus respectivos archivos con extensión (HTML) [Fig.17].

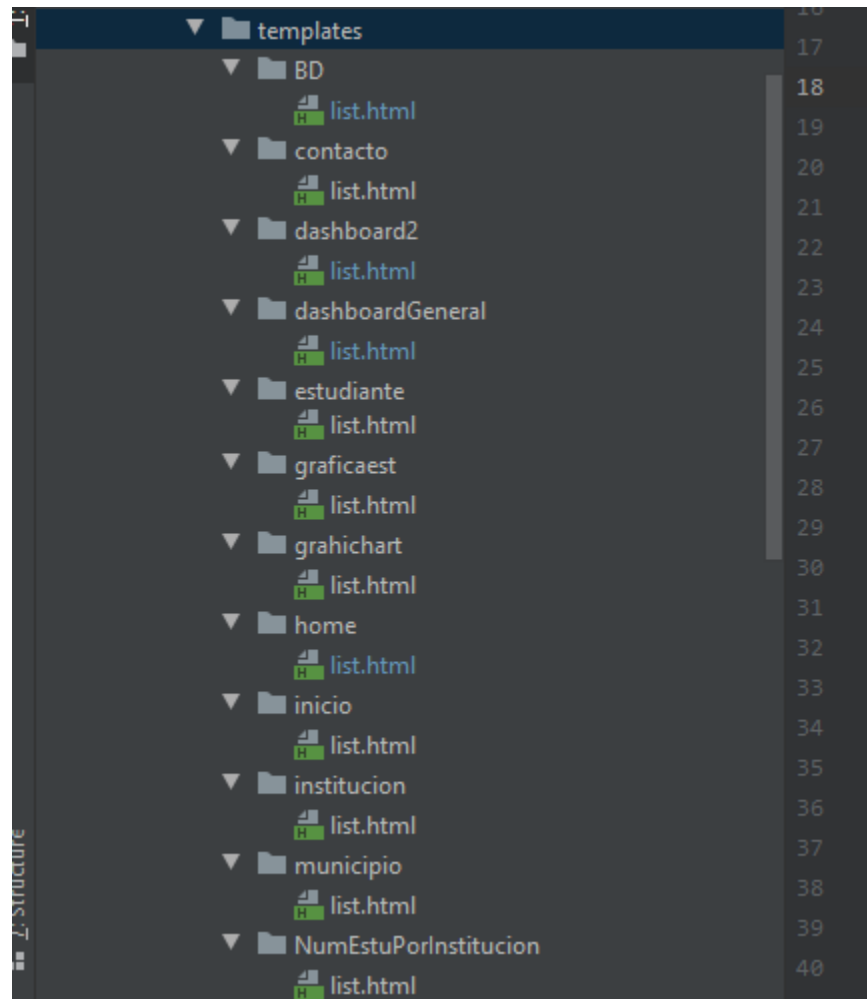


Fig. 17. Directorio de archivos (HTML)

Fuente: Propia, 2021

En la siguiente imagen podemos visualizar los directorios que contienen las diferentes vistas utilizadas para el manejo de la lógica en la aplicación [Fig.18].

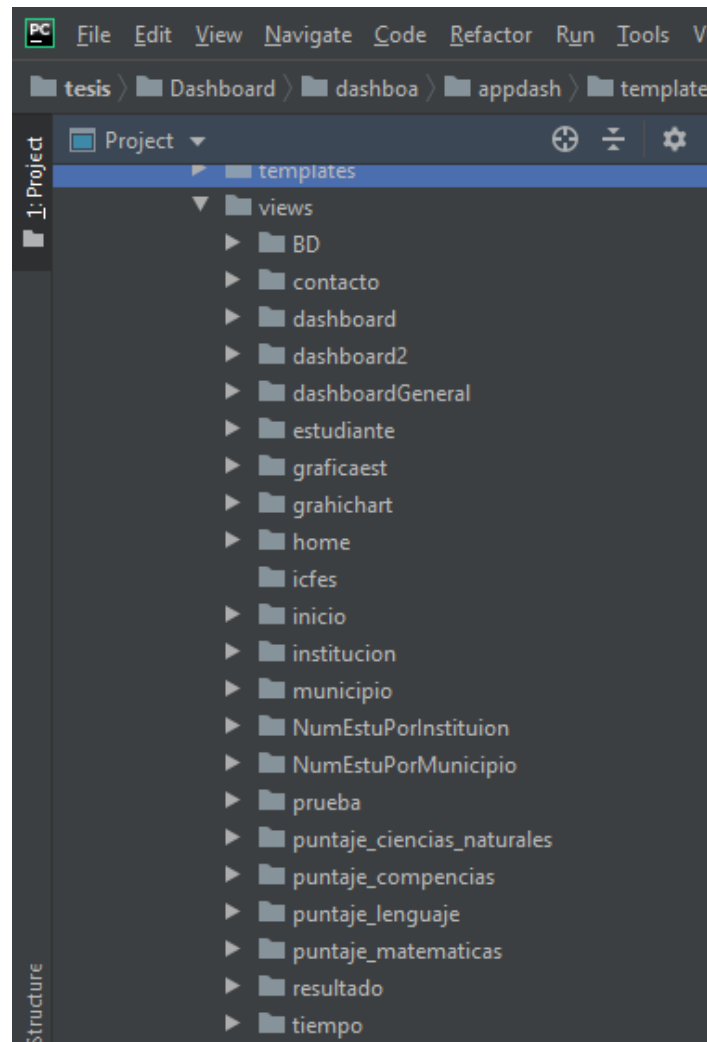


Fig. 18. Directorio de vistas

Fuente: Propia, 2021

En la siguiente grafica podemos observar desplegados los diferentes directorios donde se encuentras los archivos que contienen las respectivas vistas [Fig.19].

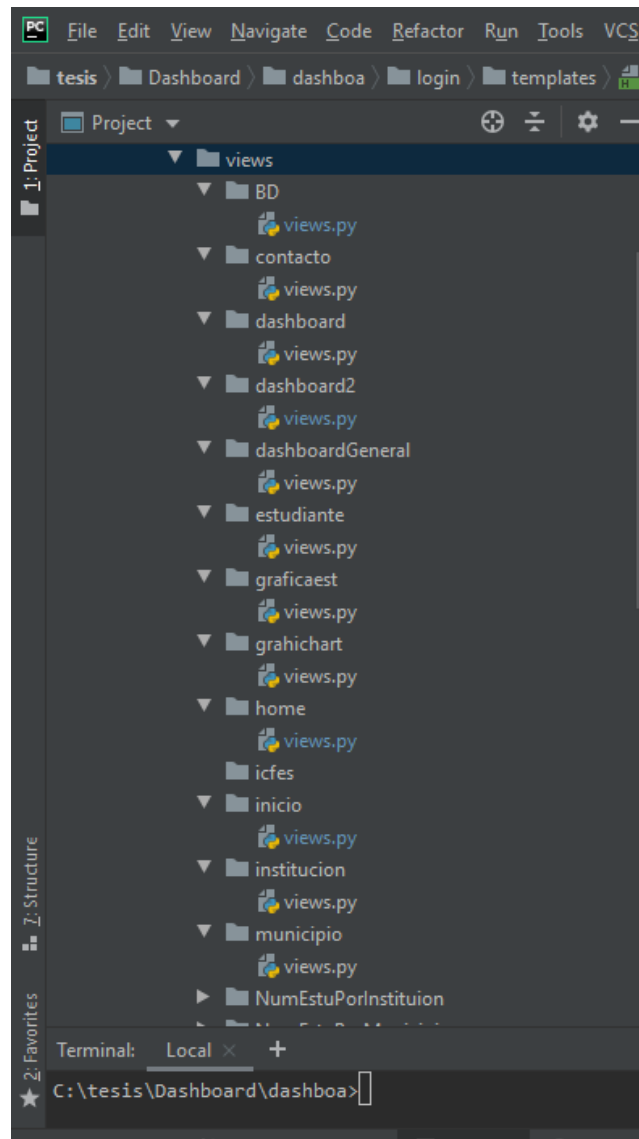


Fig. 19. Directorios con sus respectivas vistas

Fuente: Propia, 2021

En la siguiente imagen podemos observar el directorio dashboard el cual contiene el archivo urls.py en el cual podemos encontrar las diferentes rutas que permiten conectar con las diferentes vistas para tener acceso a la información de la base de datos según la lógica descrita en cada vista [Fig.20].

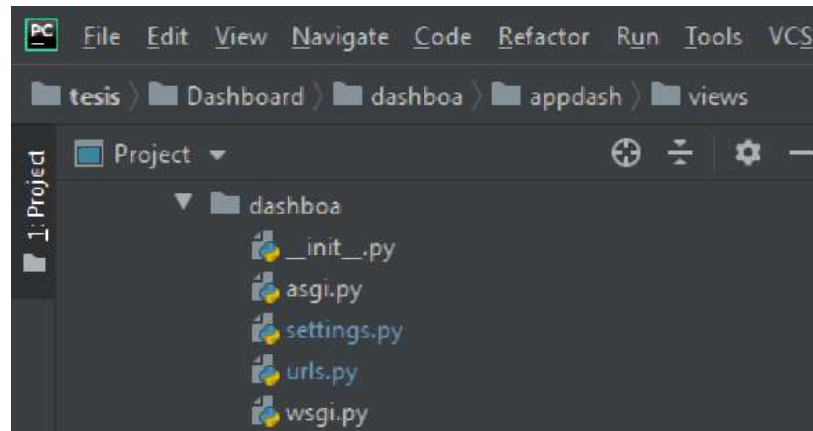


Fig. 20. Directorio de rutas

Fuente: Propia, 2021

En el directorio login podemos encontrar la plantilla en el directorio templates y la vista en el archivo views.py los cuales permiten la configuración del login de usuario para administrador e invitado en la aplicación [Fig.21].

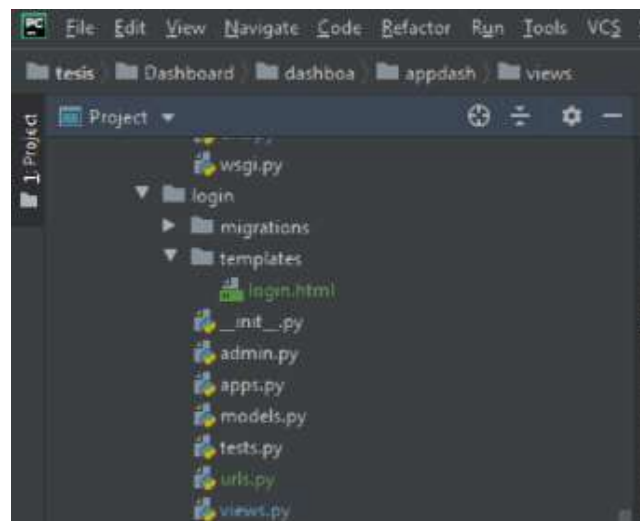


Fig. 21. Directorio de plantillas

Fuente: Propia, 2021

En la siguiente imagen podemos encontrar desplegado el directorio o carpeta templates el cual contiene las diferentes plantillas bases que permite tener las diferentes rutas de librerías y JQuery que se utilizan para toda la aplicación con sus respectivas conFiguraciones, así también un pie de página y un encabezado como herencia para todas las plantillas que se visualicen para el usuario [Fig.22].

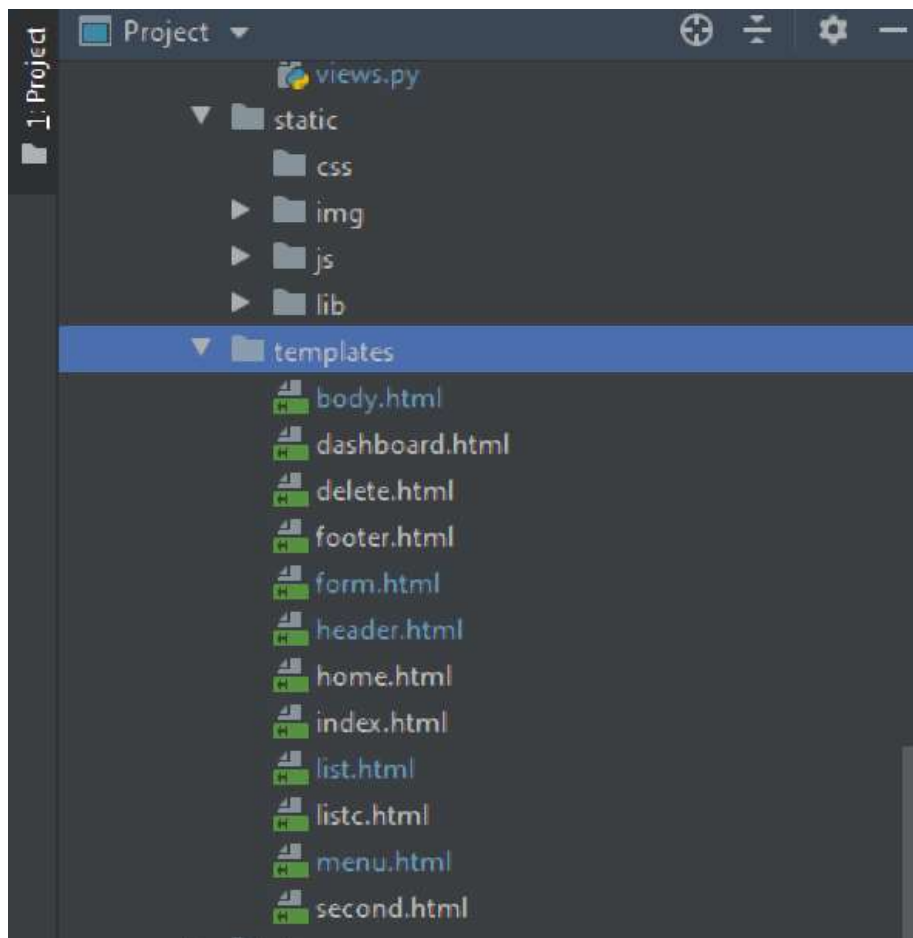


Fig. 22. Directorio de plantillas rutas de libreria y JQuery

Fuente: Propia, 2021

E. PRUEBAS

Para realizar el análisis de la creación de la herramienta Saber Expo BI 5 es necesario contar con un equipo adecuado; esto por la cantidad de información y el procesamiento que esta consume, de igual manera, los intérpretes y herramientas para la realización del software; para esto se contó con un pc portátil Asus x450c con un procesador Intel(R) Core(TM) i5-3337U CPU @ 1.80GHz 1.80 GHz que cuenta con una gráfica NVIDIA 720 de 2GB con una memoria RAM de 8GB y un disco duro mecánico de 1TB; se tomó cada uno de los casos de uso para programar la herramienta.

1) Entrar como invitado

El modo de entrar como invitado sirve para que no todos los usuarios puedan acceder a herramientas importantes de la aplicación como la subida de nuevos datos, para esto se implementó un botón de entrar como invitado en la pantalla de inicio de la herramienta [Fig.23].



Fig. 23. Saber Expo BI 5, entrar como invitado

Fuente: Propia, 2021

Una vez el invitado se autentica y puede visualizar la siguiente ventana, en donde está presente la vista general de inicio y el menú proporcionado para el invitado [Fig.24].



Fig. 24. Saber Expo BI 5, vista como invitado

Fuente: Propia, 2021

2) Inicio de sesión

Este caso se permite a los usuarios validar su identidad ante el sistema, muestra en pantalla un cuadro de dialogo donde se digita el login de usuario y contraseña para poder ingresar a la aplicación, para esto los campos cuentan con una validación con la cual si el usuario tiene un registro previo puede acceder a la siguiente pantalla de la herramienta, si el usuario no tiene este registro o los campos son llenados con información incorrecta, un mensaje de error de logeo se desplegará [Fig.25].



Fig. 25. Saber Expo BI 5, vista inicio de sesión

Fuente: Propia, 2021

Como particularidad, al iniciar sesión se añade al menú principal, una opción en donde el administrador es capaz de cargar nueva información al mercado de datos ya establecido [Fig.26].

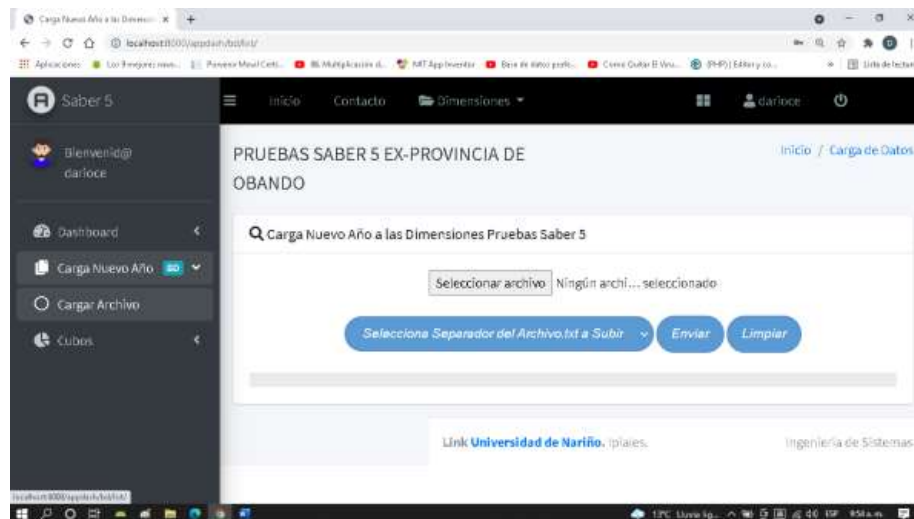


Fig. 26. Saber Expo BI 5, vista administrador

Fuente: Propia, 2021

3) Cerrar sesión

Para realizar esta prueba, antes se debe ser administrador registrado, a lado superior derecho de la aplicación se muestra el rol que ocupa en la en la herramienta, sea administrador o invitado, si se elige esta opción se despliega la opción de cerrar sesión en donde se pregunta al usuario si está seguro de querer salir de la aplicación; en caso de que si, se llevara de regreso a la pantalla de inicio [Fig.27].

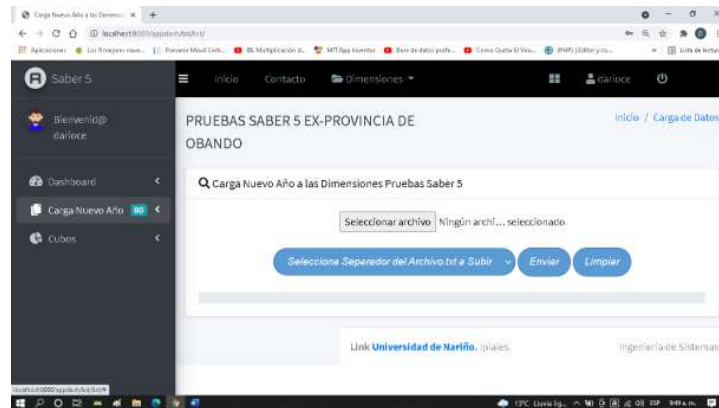


Fig. 27. Saber Expo BI 5, cerrar sesión

Fuente: Propia, 2021

4) Visualización general

Esta visualización se encuentra al entrar a la aplicación, independientemente si el usuario es administrador o invitado, en este Dashboard se muestra información global como número de estudiantes, número de instituciones, y datos analizados hasta la fecha y un puntaje general en función de los estudiantes de cada municipio de la ex provincia de Obando como se puede ver en [Fig.28], [Fig.29], [Fig.30], [Fig.31].

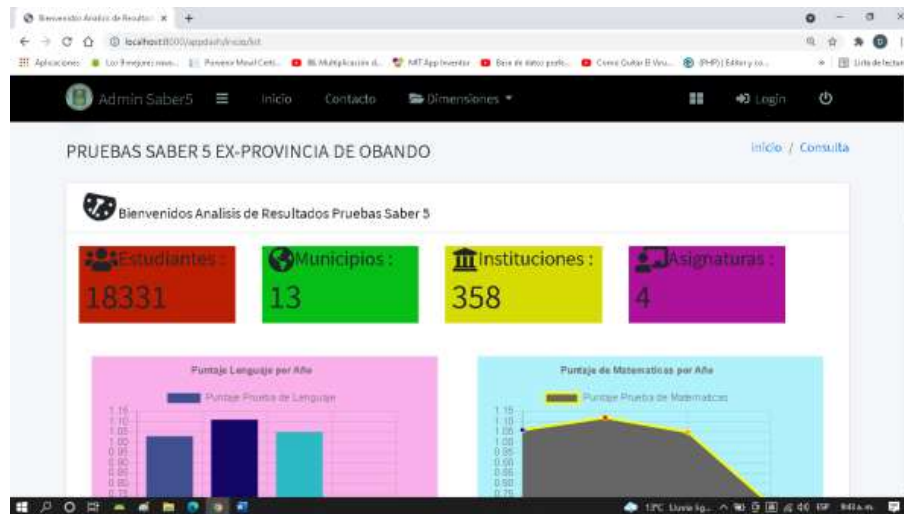


Fig. 28 Saber Expo BI 5, vista general dashboard

Fuente: Propia, 2021

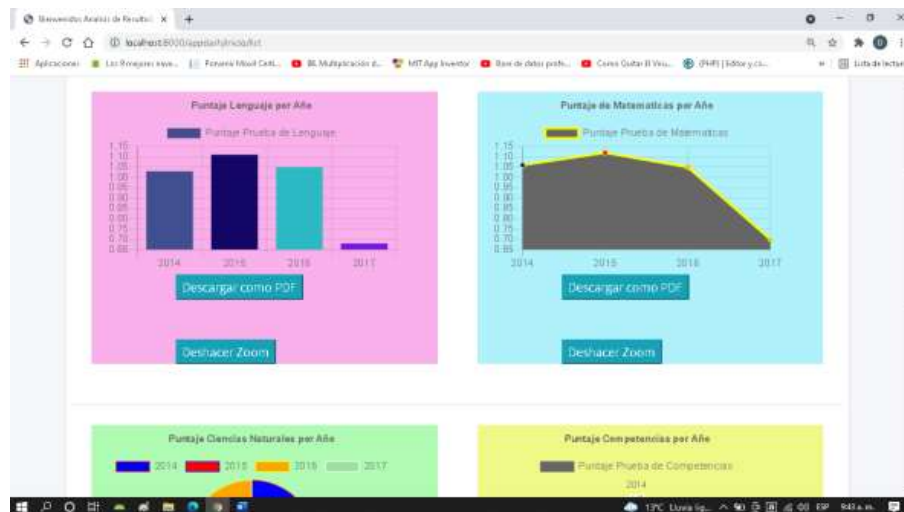


Fig. 29 Saber Expo BI 5, vista general de cubos

Fuente: Propia, 2021



Fig. 30. Saber Expo BI 5, vista general tipos de gráficas

Fuente: Propia, 2021

The dashboard displays a table with the following data:

Años	Puntaje Estudiantes Ex-provincia de Obando
2014	1,672
2015	1,700
2016	1,580
2017	1,660

Mostrando registros del 1 al 4 de un total de 4 registros

Fig. 31. Saber Expo BI 5, vista general por tabla

Fuente: Propia, 2021

5) Crear gráfica general

Para crear esta gráfica se debe elegir dashboard en el menú de inicio y este desplegará opciones en las cuales se podrá escoger entre diferentes tipos de categorías, métricas y filtros que se desee visualizar para ver en el dashboard, a esto se le añade una opción de tipo de grafica con la cual los resultados se presentan a conveniencia del usuario, si este desea existe la opción de guardar gráfica, con la cual se genera un archivo.pdf de la misma, se pueden observar en [Fig.32], [Fig.33], [Fig.34].

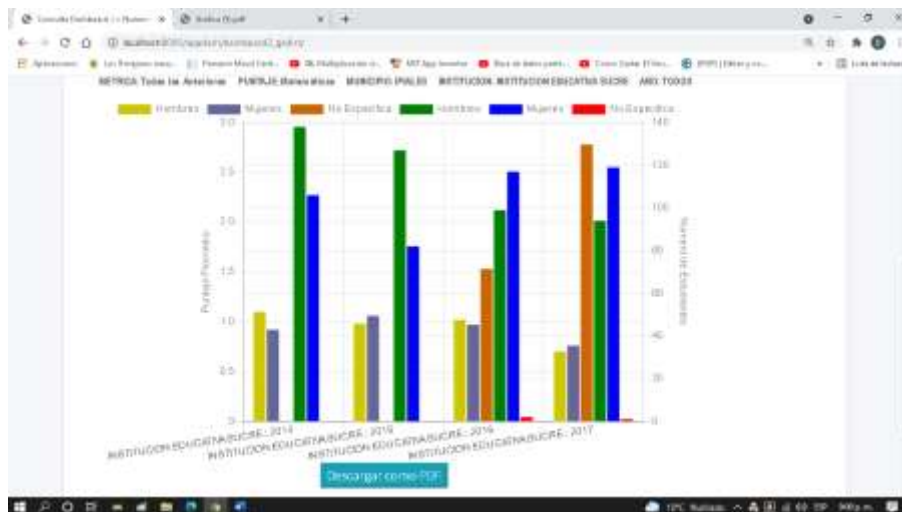


Fig. 32. Saber Expo BI 5, crear gráfica general

Fuente: Propia, 2021



Fig. 33. Saber Expo BI 5, descargar gráfica general

Fuente: Propia, 2021

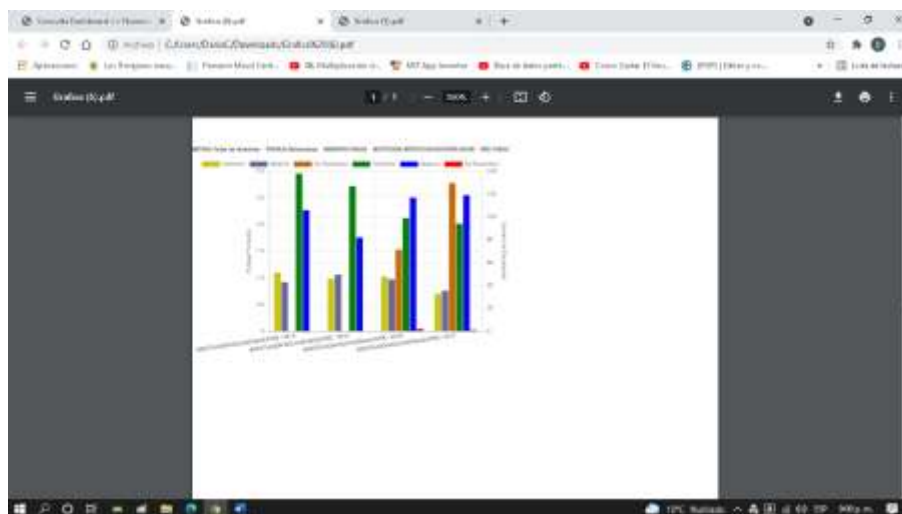


Fig. 34. Saber Expo BI 5, documento.pdf gráfica general

Fuente: Propia, 2021

6) Crear gráfica por número de estudiante

Se debe elegir en el menú dashboard la opción número de estudiante en el cual se presentan opciones de filtros como son; módulo, categoría el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf, vistos en [Fig.35], [Fig.36], [Fig.37].

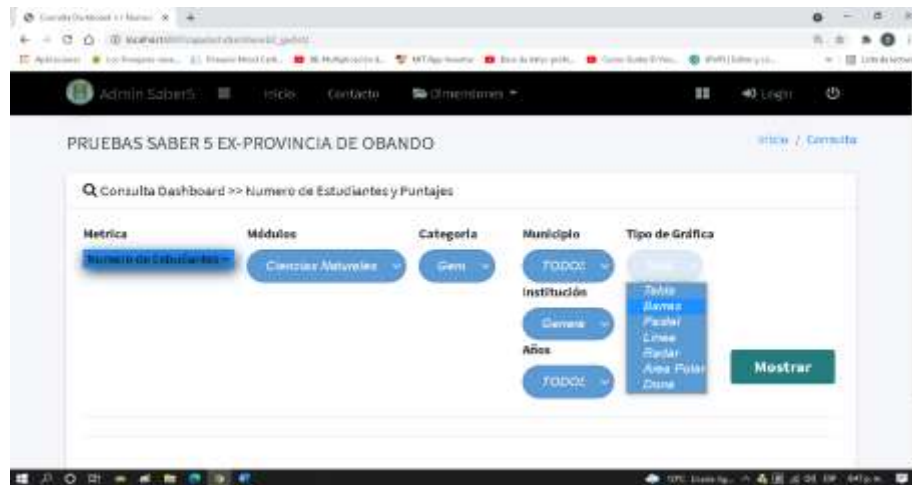


Fig. 35. Saber Expo BI 5, crear gráfica por número de estudiantes

Fuente: Propia, 2021

The screenshot shows the same dashboard with the filters set to 'Métrica: Número de Estudiantes', 'Módulos: Matemáticas', 'Categoría: Tipo de Establecimiento', 'Municipio: TODOS', 'Institución: General', and 'Año: TODOS'. Below the filters, a table displays the data for the number of students across different municipalities and years.

	ALDANA 2014	ALDANA 2015	ALDANA 2016	ALDANA 2017	CONTADERO 2014	CONTADERO 2015	CONTADERO 2016	CONTADERO 2017
Oficial Rural	142	169	121	150	42	47	57	39
No Oficial	0	0	0	0	0	0	0	0
Oficial Urbano	0	0	0	0	38	50	41	25

Fig. 36. Saber Expo BI 5, tabla número de estudiantes

Fuente: Propia, 2021

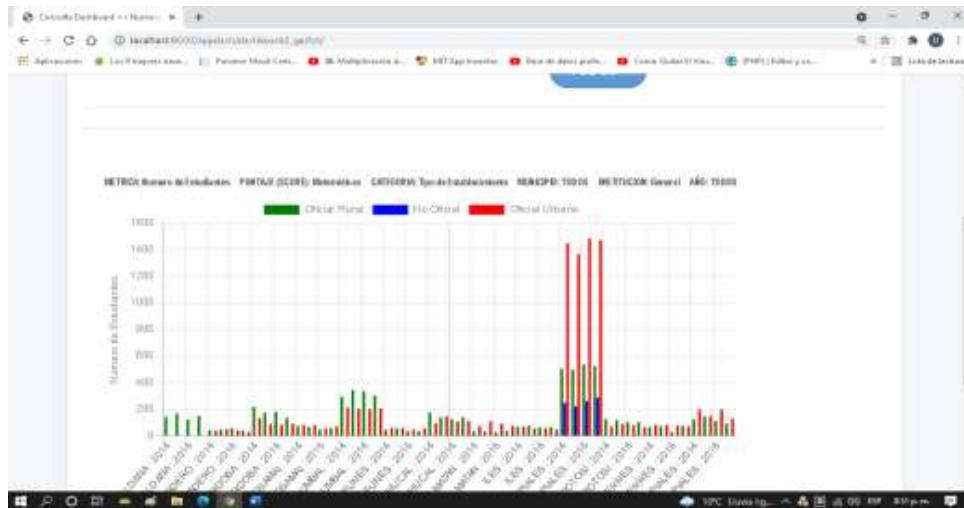


Fig. 37. Saber Expo BI 5, gráfica de barras por número de estudiantes

Fuente: Propia, 2021

7) Crear grafica por puntaje

Se debe elegir en el menú dashboard la opción puntaje en el cual se presentan opciones de filtros como son; el área de la que se desea el puntaje, el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf, se pueden observar en [Fig.38], [Fig.39], [Fig.40], [Fig.41], [Fig.42].



Fig. 38. Saber Expo BI 5, crear gráfica por puntaje

Fuente: Propia, 2021

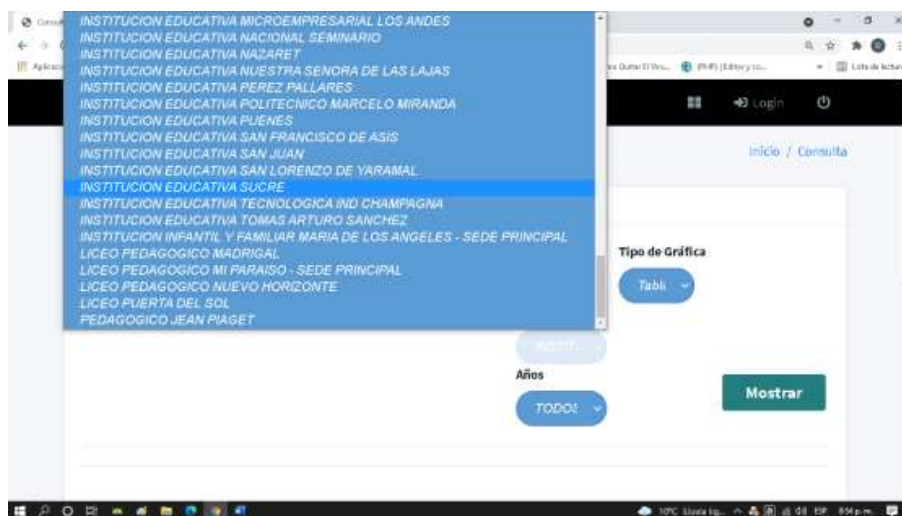


Fig. 39. Saber Expo BI 5, crear gráfica de puntaje por institución y municipio

Fuente: Propia, 2021

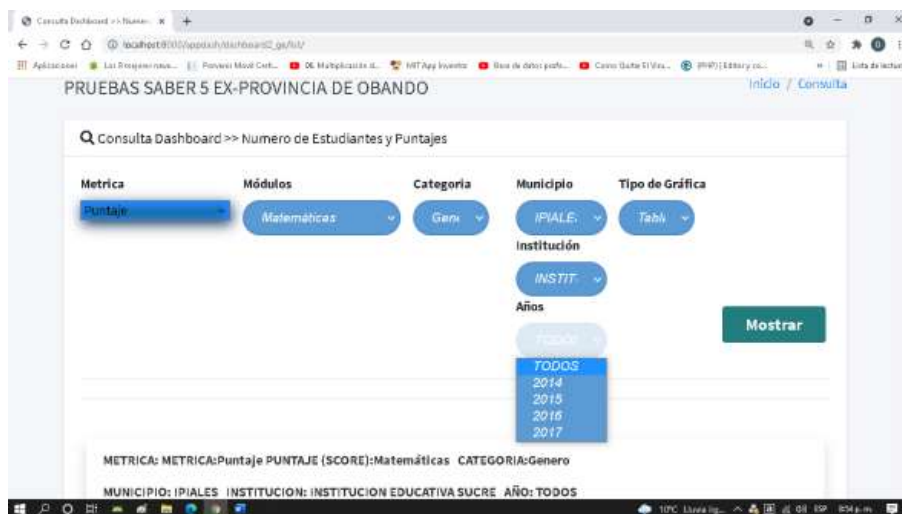


Fig. 40. Saber Expo BI 5, crear gráfica de puntaje por año de presentación

Fuente: Propia, 2021



Fig. 41. Saber Expo BI 5, crear gráfica por puntaje con tipo de gráfica

Fuente: Propia, 2021



Fig. 42. Saber Expo BI 5, crear gráfica de barras por puntaje

Fuente: Propia, 2021

8) Crear gráfica por categoría

Se debe elegir en el menú dashboard la opción puntaje en el cual se presentan opciones de filtros como son; categoría, el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf, se pueden observar en [Fig.43], [Fig.44].

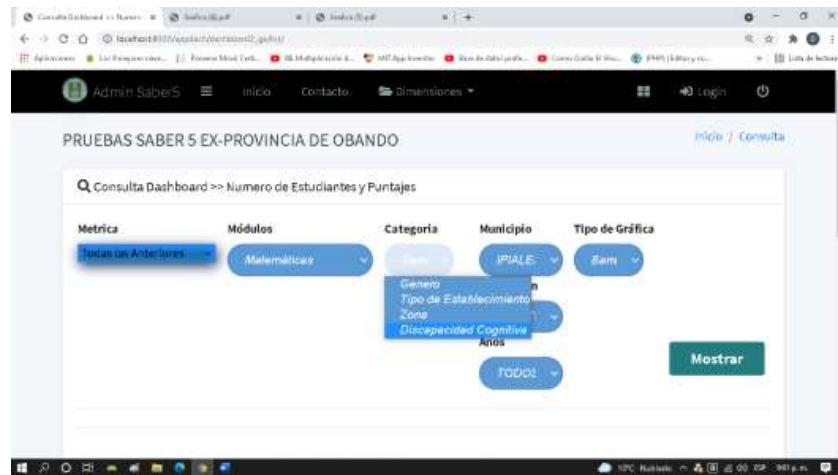


Fig. 43. Saber Expo BI 5, crear gráfica por categoría

Fuente: Propia, 2021

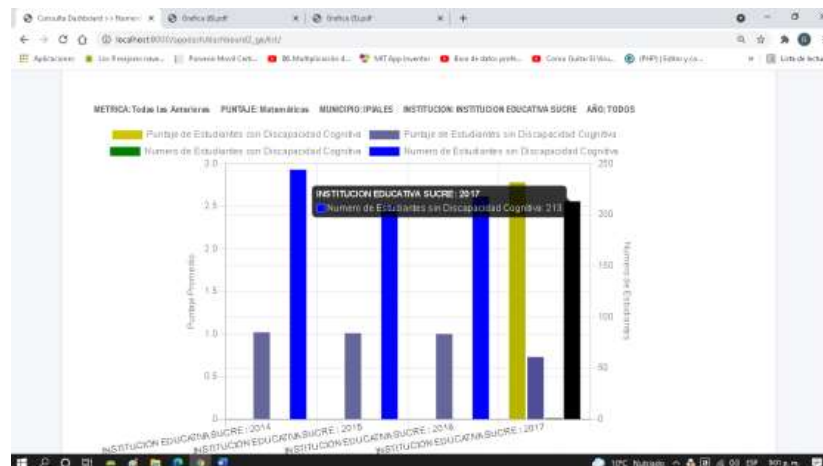


Fig. 44. Saber Expo BI 5, visualizar gráfica por categoría

Fuente: Propia, 2021

9) Nueva carga de datos

Para los usuarios administradores se les presenta la opción de cargar nueva base de datos, esta se encuentra en el menú subir nueva base de datos, en este proceso se toma la tabla valores plausibles del repositorio brindado por la página oficial del ICFES el cual es un archivo .txt; cuenta con la opción de elegir agregar nueva base de datos la cual permite buscar el archivo proporcionado por el ICFES una vez sea descargado, una vez marcado se procede a elegir el delimitador con el que se separa los atributos de la nueva base de datos y se procede a subir el nuevo archivo, cabe resaltar que al momento de subir este archivo, la aplicación estandarizará, limpiará, eliminará y transformará el archivo .txt que se sube para añadirlo como un nuevo año a la bodega de datos de la aplicación, se pueden observar en [Fig.45], [Fig.46], [Fig.47], [Fig.48].

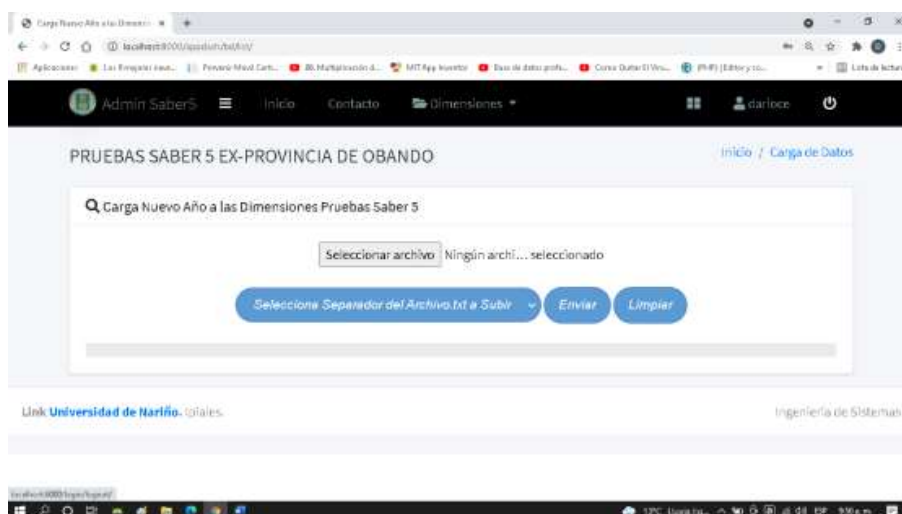


Fig. 45. Saber Expo BI 5, ventana de carga nuevo año

Fuente: Propia, 2021

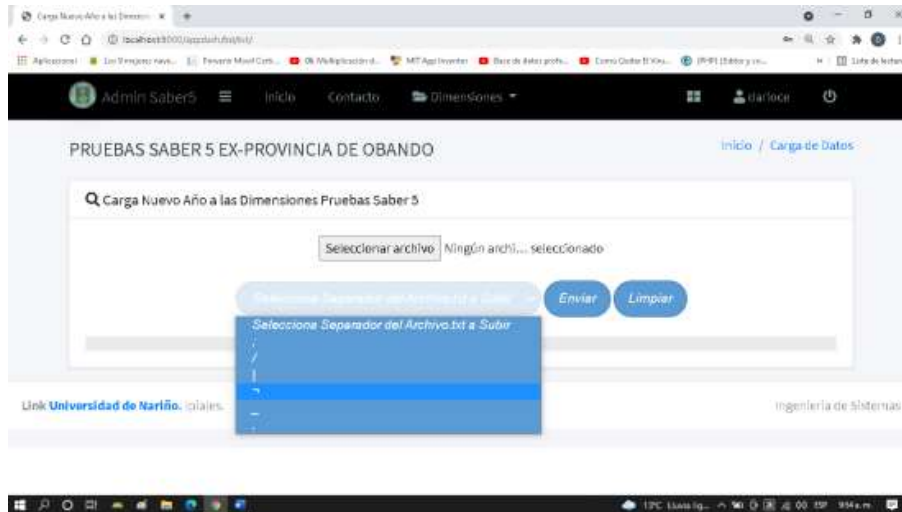


Fig. 46. Saber Expo BI 5, ventana cargar nuevo año, seleccionar delimitador

Fuente: Propia, 2021

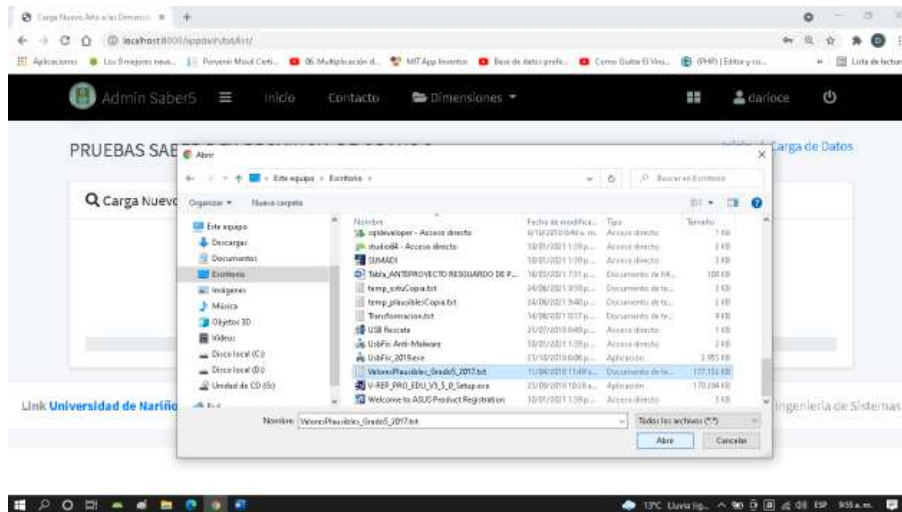


Fig. 47. Saber Expo BI 5, ventana cargar nuevo año, seleccionar archivo

Fuente: Propia, 2021

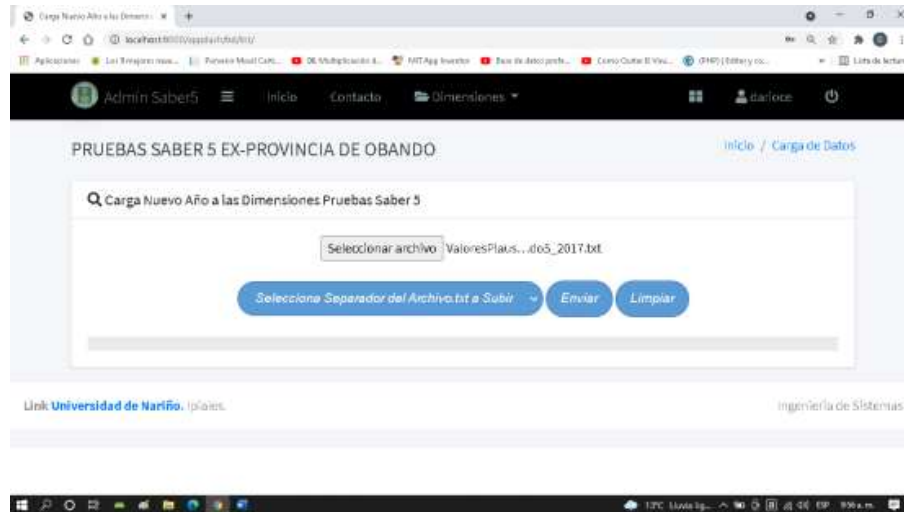


Fig. 48. Saber Expo BI 5, ventana cargar nuevo año, archivo cargado

Fuente: Propia, 2021

10) Cubos de datos

a) Número de estudiantes por municipio y por año de presentación

Cubo que brinda información sobre la población de estudiantes que presentaron las pruebas en determinado año [Fig.49], [Fig.50], para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, COUNT(id_estudiante) as suma FROM dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY 1,2 ORDER BY 1,2
```



Fig. 49. Cubo, gráfica número de estudiantes por año y municipio

Fuente: Propia, 2021

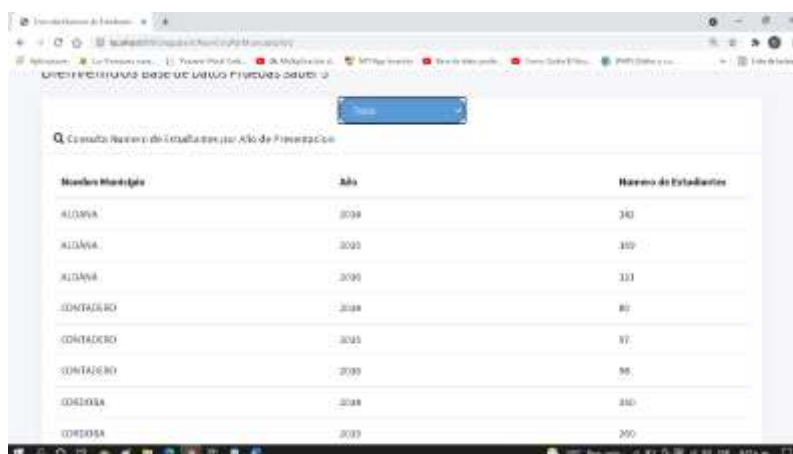


Fig. 50. Cubo, tabla número de estudiantes por año y municipio

Fuente: Propia, 2021

b) Número de estudiantes por municipio general

Cubo que brinda información sobre los estudiantes que presentaron la prueba por municipio, [Fig.51], [Fig.52]., para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre,COUNT(id_estudiante) as suma FROM dim_municipios JOIN
fact_saber5 USING(muni_id_ente) JOIN dim_estudiante USING(id_estudiante) GROUP BY 1
ORDER BY 1
```

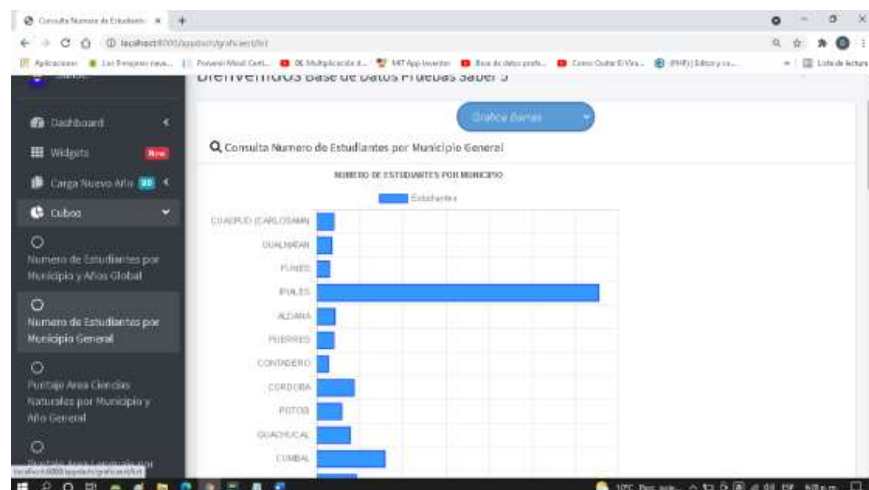


Fig. 51. Cubo, gráfica número de estudiantes por municipio general

Fuente: Propia, 2021

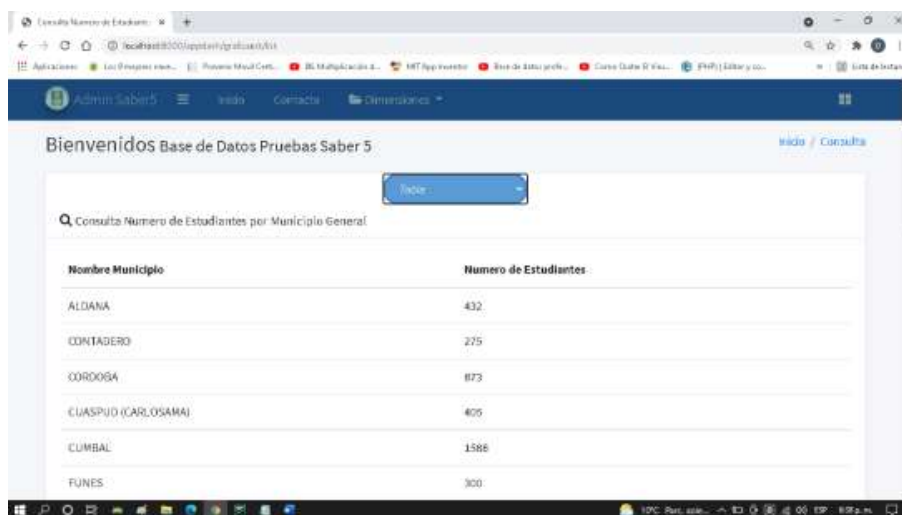


Fig. 52. Cubo, gráfica número de estudiantes por municipio general

Fuente: Propia, 2021

c) Puntaje área ciencias naturales por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de ciencias naturales de cada municipio tomando todos los años en que se realizaron las pruebas, [Fig.53], [Fig.54]., para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_ciencias),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```

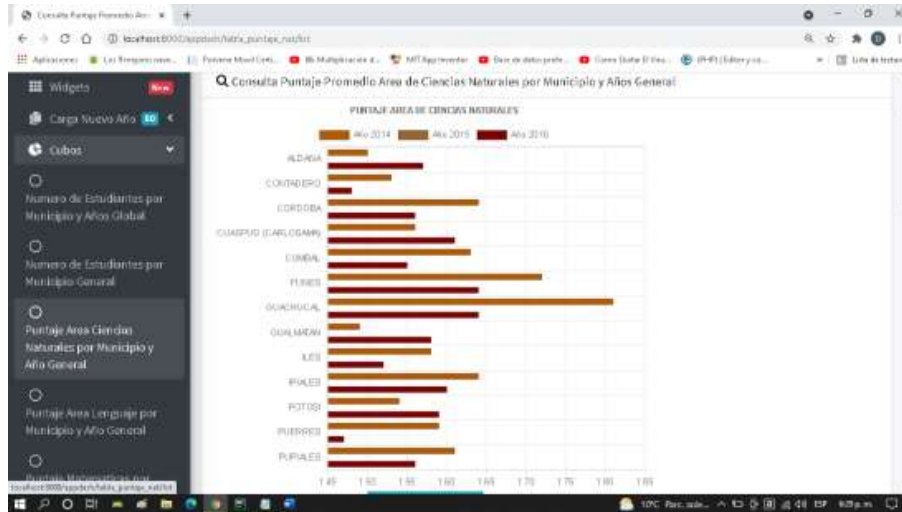


Fig. 53. Cubo, gráfica puntaje área ciencias por municipio

Fuente: Propia, 2021

Nombre Municipio	Año de Presentación	Puntaje Promedio Ciencias Naturales
ALDANA	2014	1,50
CONTADERO	2014	1,53
CORDOBA	2014	1,56
CUASPUG (CARLOSAMA)	2014	1,59
EMBAL	2014	1,62
FUNES	2014	1,65

Fig. 54. Cubo, tabla puntaje área ciencias por municipio

Fuente: Propia, 2021

d) Puntaje área lenguaje por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de lenguaje de cada municipio tomando todos los años en que se realizaron las pruebas, [Fig.55], [Fig.56], para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_lenguaje),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```



Fig. 55. Cubo, gráfica puntaje área lenguaje por municipio

Fuente: Propia, 2021

Consulta Puntaje Promedio Área de Lenguaje por Municipio y Años General

Nombre Municipio	Año de Presentación	Puntaje Promedio Lenguaje
ALDANA	2014	1,33
CONTADERO	2014	1,55
CORDOBA	2014	1,64
CUASPUD (CARLOSAMA)	2014	1,90
CUMBAL	2014	1,63
FUNES	2014	1,77

Fig. 56. Cubo, tabla puntaje área lenguaje por municipio

Fuente: Propia, 2021

e) Puntaje área de matemáticas por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de matemáticas de cada municipio tomando todos los años en que se realizaron las pruebas, [Fig.57], [Fig.58], para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_matematicas),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```



Fig. 57. Cubo, gráfica puntaje área matemáticas por municipio

Fuente: Propia, 2021

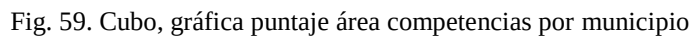
Nombre Municipio	Año de Presentación	Puntaje Promedio Matemáticas
ALDANA	2014	1,50
CONTADERO	2014	1,54
CORDOBA	2014	1,56
CUASPUD (CARLOSAMA)	2014	1,54
CUMBAL	2014	1,56
FUNES	2014	1,78

Fig. 58. Cubo, tabla de puntaje área matemáticas por municipio

Fuente: Propia, 2021

f) Puntaje área competencias por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de competencias cada municipio tomando todos los años en que se realizaron las pruebas, [Fig.59], [Fig.60], [Fig.61], para esto se utilizó la siguiente sentencia SQL:



Hydrocarbon d., MIT App Inventor Best in class profile.



118

Bienvenidos Base de Datos Pruebas Saber 5

Consulta Partaje Promedio Área de Competencias por Municipio y Año de Presentación

Nombre Municipio	Año de Presentación	Partaje Promedio Competencias
AIJANA	2014	None
CONTADERO	2014	None
CORDOBA	2014	None
ELIASPITO (CARLOSAMA)	2014	None
ELUMBAL	2014	None
FUNES	2014	None

Fig. 61. Cubo, tabla puntaje área competencias por municipio

Fuente: Propia, 2021

g) Número de estudiantes por año de presentación

Cubo que brinda información sobre el número de estudiantes que presentaron la prueba saber 5 en función al año de presentación, [Fig.62], [Fig.63], [Fig.64], se utilizó la siguiente sentencia SQL:

SELECT anio, COUNT(id_estudiante) as suma FROM fact_saber5 GROUP BY 1,1 ORDER BY 1,1

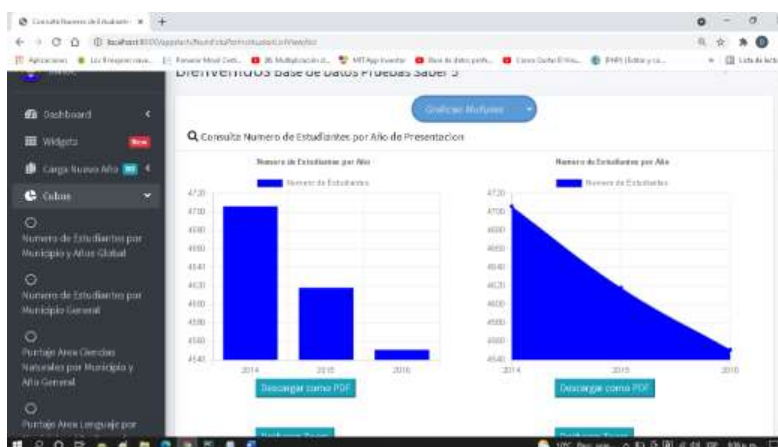


Fig. 62. Cubo, gráfica número de estudiantes por año de presentación

Fuente: Propia, 2021

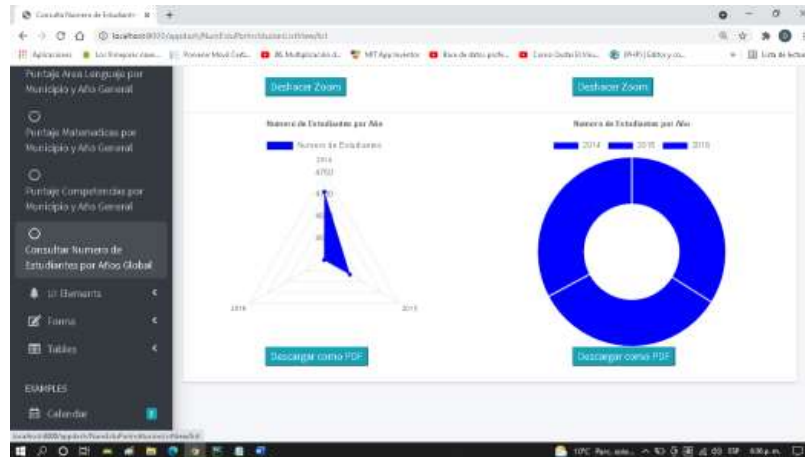


Fig. 63. Cubo, gráficas varias número de estudiantes por año de presentación

Fuente: Propia, 2021

The figure shows a web interface with a table titled 'Bienvenidos Base de Datos Pruebas Saber 5'. The table has two columns: 'Años' and 'Número de Estudiantes'. The data is as follows:

Años	Número de Estudiantes
2014	4706
2015	4818
2016	4553

Fig. 64. Cubo, tabla número de estudiantes por año de presentación

Fuente: Propia, 2021

11) Gráficas estadísticas adicionales

Existen forma de presentación estadística más compleja, destinada fundamentalmente a resaltar aspectos de la distribución, observando patrones en una o más series de datos cuantitativos a la vez, para esta representación se eligieron los tipos de gráficos boxplot y biplot, por la facilidad con la cual se pueden observar datos de distintas categorías al tiempo.

a) Boxplot

Los diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente [41].

Para ingresar se debe elegir la opción estadística en el menú de opciones y marcar la opción boxplot, posteriormente elegir el año que se desea mostrar cómo se observa en [Fig.65].



Fig. 65.Boxplot, menú boxplot y año de prueba a seleccionar

Fuente: Propia, 2021

En [Fig.66], [Fig.67], [Fig.68], podemos observar la implementación de boxplot a los puntajes de las pruebas saber 5 categorizadas por año.



Fig. 66.Boxplot, gráfico boxplot del año 2015

Fuente: Propia, 2021



Fig. 67.Boxplot, gráfico boxplot del año 2016

Fuente: Propia, 2021



Fig. 68.Boxplot, gráfico boxplot del año 2017

Fuente: Propia, 2021

b) Biplot

Los **Biplots** son un tipo de gráfico exploratorio usado en estadística, se trata de una generalización multivariante de un diagrama de dispersión de dos o más variables, el biplot aproxima la distribución de una muestra multivariante en un espacio de dimensión reducida, normalmente de dimensión dos, y superpone sobre la misma representación de las variables sobre las que se mide la muestra. [42].

Para ingresar se debe elegir la opción estadística en el menú de opciones y marcar la opción biplot, cabe aclarar que no se hace ninguna búsqueda de categoría o filtro, puesto que se encuentran todas las variables y categorías que son representados por los PCA dentro de la programación del biplot, [Fig.69].



Fig. 69. Biplot, menú biplot a seleccionar

Fuente: Propia, 2021

En [Fig.70] podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto al año, con una variabilidad del 54%.

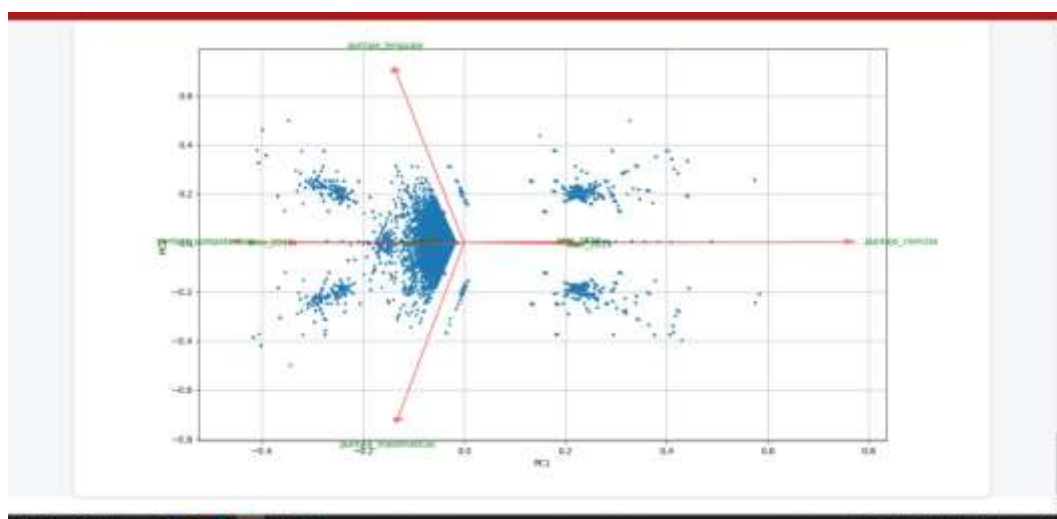


Fig. 70. Biplot, gráfica biplot correlación materia-año

Fuente: Propia, 2021

En [Fig.71] podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a la discapacidad, con una variabilidad del 65%.

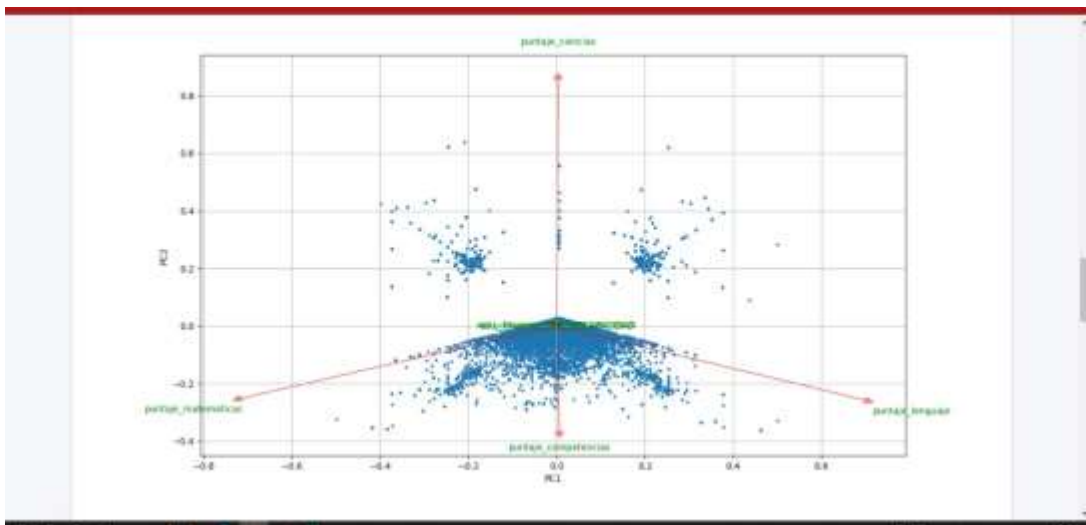


Fig. 71. Biplot, gráfica biplot correlación materia-discapacidad

Fuente: Propia, 2021

En [Fig.72] podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a género, con una variabilidad del 53%.

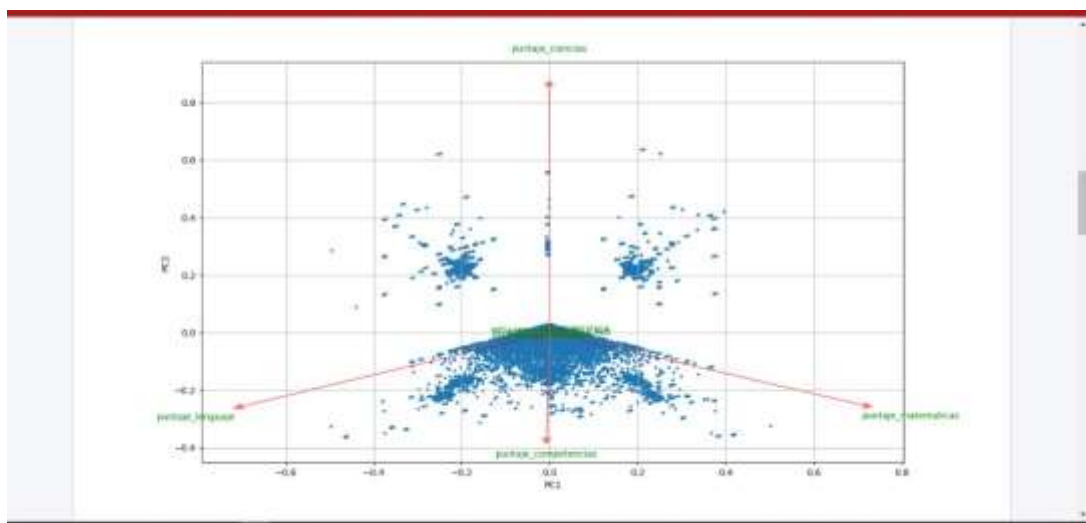


Fig. 72. Biplot, gráfica biplot correlación materia-género

Fuente: Propia, 2021

En [Fig.73] podemos observar la implementación de biplot, a la correlación de puntaje de las diferentes materias, con una variabilidad del 65%.

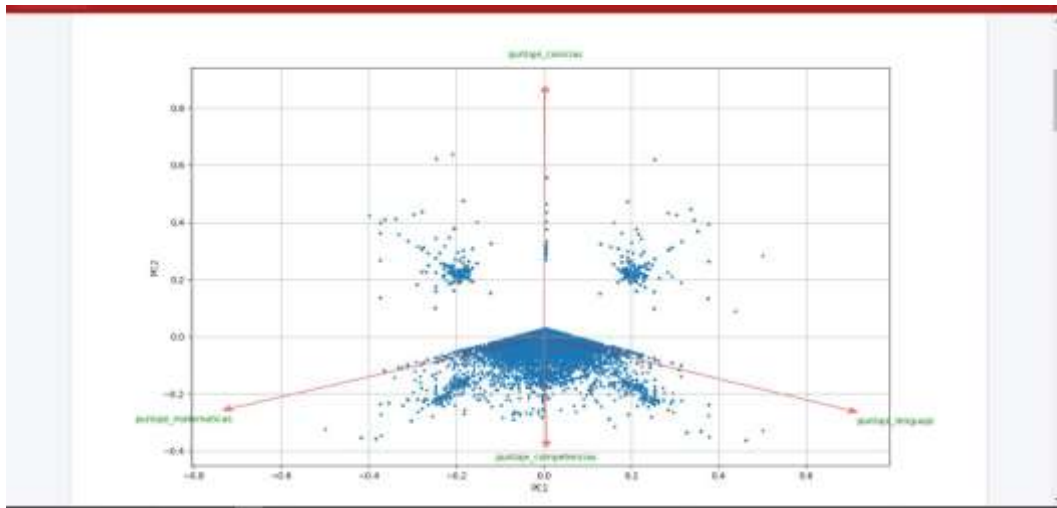


Fig. 73. Biplot, gráfica biplot correlación materias

Fuente: Propia, 2021

En [Fig.74] podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a la zona, con una variabilidad del 45%.

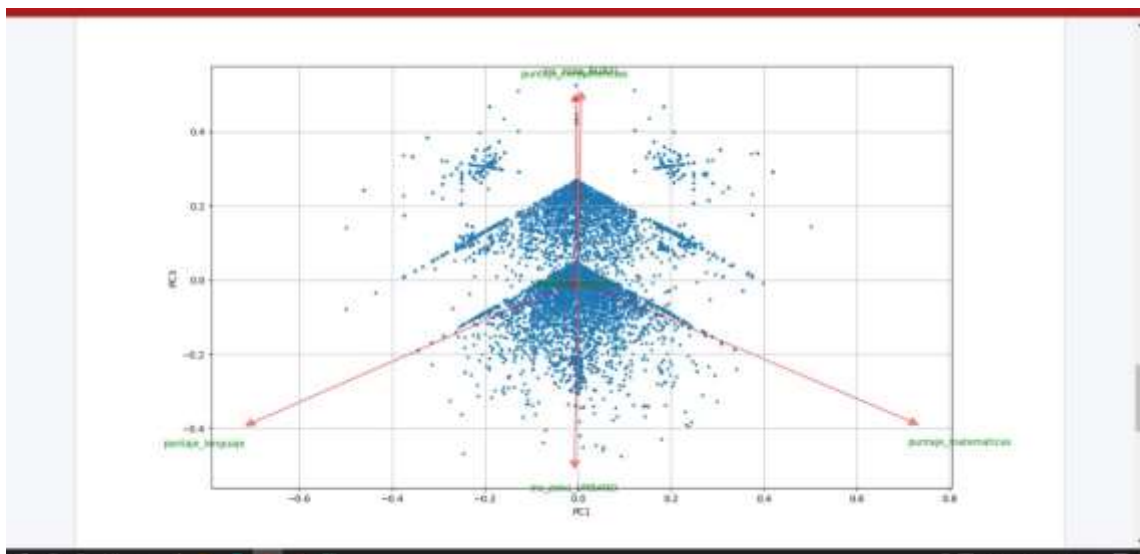


Fig. 74. Biplot, gráfica biplot correlación materias-zona

Fuente: Propia, 2021

V. DISCUSIÓN

En el presente trabajo de investigación se toman distintas ideas y recursos de varias fuentes, las cuales desempeñan un rol importante en la construcción de cada módulo con los que la aplicación cuenta. Para poder cumplir con el objeto de Saber Expo BI 5, se tomó como antecedentes, proyectos, artículos y tesis relevantes, los cuales sirvieron para el desarrollo de la aplicación Saber Expo BI 5.

El primer paso para la construcción de la aplicación, fue basarse en una base de datos construida previamente en el trabajo de grado [8], esta base de datos cuenta con información de las pruebas saber 5 comprendidas entre los años del 2012 a 2016, sin embargo, en el presente trabajo de grado se pensó en construir un módulo en el cual se pueda agregar un nuevo año en el banco de datos ya existente. Para que esto ocurra, los campos de las tablas en las cuales se encuentra dicha información de las pruebas deben ser idénticos, no obstante, campos como los puntajes de los estudiantes, aunque si estaban presentes, el rango con el cual se calificaba las asignaturas de las pruebas cambió desde el año 2017 en adelante, por lo cual fue necesario estandarizar esta información. Teniendo los campos listos, se procedió con la programación del botón (agregar nuevo año), para esto el botón debe ser capaz de aplicar una limpieza y transformación; en este Sprint, fue necesario tomar referencias de [13], cuyo objetivo fue aplicar cambios tales como, transformación y limpieza de datos (ETL) a las pruebas saber 11, sin embargo, surgió nuevamente el problema de la estandarización, puesto que en el mercado de datos tomado de [8] los atributos y tipos de datos no eran consistentes con el nuevo año a agregar, así se identificó el tipo de dato y tipo de atributo de cada campo en cada tabla y mediante consultas SQL fue posible hacer la programación del botón (agregar nuevo año), teniendo como función subir una nueva base de datos con un nuevo año al mercado ya existente, no sin antes aplicar una transformación y una limpieza, para posteriormente ser cargada y adicionada a este banco de datos. Para esto fue necesario aplicar distintos métodos y técnicas de limpieza que ayudaron al entendimiento de conceptos clave para la carga de nuevos datos, teniendo en cuenta el delimitador de las nuevas tablas que se desea añadir al mercado ya existente. Para lo concerniente a consultas en la base de datos, se investigó técnicas

de minería de datos según [14], cuyos factores se asociaron con el desempeño académico de las diferentes competencias de las pruebas saber, comprendiendo los años del 2014 al 2017. Esta información es vital para el desarrollo del proyecto, ya que se puede identificar factores con los cuales se puedan hacer una mejor toma de decisiones. La información tomada de [14] era muy precisa, sin embargo, sólo podía ser entendida por personas, las cuales tengan un mínimo conocimiento en este ámbito, por lo cual se piensa en el siguiente paso que es simplificar y optimizar este tipo de información, para que cualquier persona pueda identificar dichos factores que puedan afianzar su entendimiento, en cuanto al rendimiento de las pruebas saber 5 en cada establecimiento. Para esto, se pensó en un diseño de graficas estadísticas, las cuales proporcionen esta facilidad de entendimiento; se tomó conceptos de la implementación de un dash board creado y documentado por [15], proporcionando información clara en la creación de cuadros de mando, filtros, categorías, tipos de gráfica que son muy necesarios para la creación de este aplicativo, sin embargo, este trabajo de investigación contaba con pocos tipos de gráfica, con lo cual fue necesario investigar más a fondo sobre la programación de cuadros de mando para aumentar considerablemente los tipos de gráfica que cualquier usuario quisiera visualizar.

Finalmente se creó un aplicativo basado en distintos trabajos de investigación, tomando y mejorando aspectos con los cuales dichos trabajos no contaban, ya sea por extensión o por complejidad de los mismos.

CONCLUSIONES

- La implementación de una aplicación de inteligencia de negocios como lo es Saber Expo BI5, para el seguimiento del rendimiento académico en cuanto a las competencias evaluadas en las pruebas saber 5, permite analizar la información de manera óptima y precisa, logrando así, descubrir información relevante a la ya brindada por el ICFES.
- La representación gráfica de los indicadores de las pruebas, permite tener un mayor control sobre el estado y rendimiento de las instituciones de la ex provincia de Obando, con lo cual, es mucho más fácil tomar decisiones, teniendo toda la información de forma centralizada y organizada, características brindadas por la creación e implementación de los cuadros en mando en la aplicación.
- En las instituciones educativas de la exprovincia y gracias a entrevistas realizadas a escuelas en su mayoría del sector rural, se evidencio que no cuentan con una método o herramienta que les permitan conocer información de primera mano, en relación al rendimiento académico de sus estudiantes en las pruebas Saber 5.
- A medida que se trabajó con el lenguaje de programación Python, se notó su gran simplicidad, flexibilidad y facilidad de aprendizaje a diferencia de otros lenguajes, caso fue la creación de los módulos de filtrado y visualización de datos, con los cuales a pesar de la poca documentación acerca del uso de librerías importantes como pandas y matplotlib, indispensables para el manejo de datos, fue aceptable el avance de la creación del código, gracias a la facilidad que presenta su estructura de modelo-vista-plantilla.
- Este proyecto permitió aplicar los conocimientos adquiridos durante toda la formación académica, junto con el proceso investigativo desarrollado en el grupo GRIAS, debido a la experiencia adquirida y la constante asesoría por parte de los docentes e investigadores del grupo.

RECOMENDACIONES

- Se recomienda a las instituciones educativas de la exprovincia de Obando utilizar esta herramienta para soportar la toma de decisiones.
- Agregar a la herramienta los datos de los demás municipios del departamento de Nariño.
- Se recomienda el uso de la aplicación en instituciones, para que estas sean capaces de tomar decisiones, plantear objetivos y estrategias de mejoramiento en pro del avance académico de sus estudiantes.

REFERENCIAS

- [1] Mineducación,” Pruebas Saber”, 2018. [Online]. Disponible: https://www.mineducacion.gov.co/1759/w3-article-397384.html?_noredirect=1.
- [2] Datademia, “¿Qué es Python?”, 2020. [Online]. Disponible: <https://datademia.es/blog/que-es-python>.
- [3] Profile, “Introducción a Pandas”, 2021. [Online]. Disponible: <https://profile.es/blog/pandas-python/>.
- [4] Programación .net, “Creación de gráficos animados con Chart JS”, 2021. [Online]. Disponible: https://programacion.net/articulo/crea_impresionantes_graficos_animados_con_chart_js_1623.
- [5] ICFES, Pruebas Saber 3°, 5° y 9°, 2015. [Online]. Disponible: <https://www.icfes.gov.co/documents/20143/176813/Guia%20pruebas%20saber%203%205%209%20lineamientos%20para%20las%20aplicaciones%20muestral%20y%20censal%202015%202%20v2.pdf>.
- [6] Posada, J. y Mendoza, F, Determinantes del logro académico de los estudiantes de grado 11 en el periodo 2008 2010. Una perspectiva de género y región. Estudios sobre calidad de la educación en Colombia, ICFES, Ministerio de Educación. Bogotá, Colombia. 2014.
- [7] ICFES, “Resultados Pruebas Saber 5°”, 2015. [Online]. Disponible: <https://www.icfes.gov.co/web/guest/resultados-saber-pro>.
- [8] B. A. Bucheli Caicedo, B. Y. Imbacuan Yandun, “Un mercado de datos para el análisis multidimensional de las pruebas saber 5 de las instituciones educativas de los municipios de la subregión de Obando del departamento de Nariño” Trabajo de grado, Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de Nariño, Ipiales, 2018.

- [9] Secretaría de Educación Departamental, “Evaluación de estudiantes”, 2016. [Online]. Disponible: <http://www.sednarino.gov.co/SEDNARINO12/index.php/es/descargassed/2557-evaluacion-estudiantes>.
- [10] Mineducación,” Clasificación de establecimientos y sedes”, 2014. [Online]. Disponible: <https://www.icfes.gov.co/documents/20143/193495/Clasificacion+de+establecimientos+y+sedes+Saber+11.pdf/2f177381-3c38-6b20-f5da-272dba42b412>.
- [11] Educación y ciudad, “Calidad educativa: una mirada a la escuela y al maestro en Colombia”, 2019. [Online]. Disponible: <https://revistas.idep.edu.co/index.php/educacion-y-ciudad/article/view/2120>.
- [12] KIO Networks, “Las ventajas competitivas de business intelligence”, 2020. [Online]. Disponible en: <https://www.kionetworks.com/blog/data-center/las-ventajas-competitivas-de-business-intelligence>.
- [13] A. D. Benavides Rosero, D. A. Villarreal Moreno, D. J. Villarreal Moreno, “Transformación estructural de datos en el proceso de inteligencia de negocios para las pymes de la región” Trabajo de grado, Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de Nariño, Ipiales, 2020.
- [14] L. M. Gómez Melo, Y. A. Jaramillo Córdoba, “Descubrimiento de factores asociados al desempeño en las pruebas saber 5 con técnicas descriptivas de minería de datos” Trabajo de grado, Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de Nariño, Pasto, 2017.
- [15] D. A. Aza Trujillo, D. H. Muñoz Quenguan “Un dashboard web para la gestión de indicadores de las pruebas saber 11 para las instituciones educativas de la subregión de Obando del departamento de Nariño” Trabajo de grado, Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de Nariño, Ipiales, 2020.
- [16] Scientia Et Technica, “Inteligencia de Negocios BI”, 2010. [Online]. Disponible: https://www.redalyc.org/pdf/849/Resumenes/Resumen_84917316060_1.pdf.

- [17] ORACLE, “¿Qué es la inteligencia de negocios?”, 2015. [Online]. Disponible: https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf.
- [18] STUDOCU, “Mercado de datos, conceptos y desarrollo”, 2011. [Online]. Disponible: <https://www.studocu.com/latam/document/universidad-latina-de-costarica/historia-del-derecho/dialnet-mercado-de-datos-4835866/3016999>.
- [19] ArcGIS, “Que es un cuadro de mando”, 2017. [Online]. Disponible: <https://doc.arcgis.com/es/dashboards/get-started/what-is-a-dashboard.htm>.
- [20] ICFES, “¿A quiénes evaluamos?”, 2017. [Online]. Disponible: <https://www.icfes.gov.co/documents/20143/1353827/Guia+de+orientacion+saber+5+2017.pdf/844f0d1e-e176-2ae9-6afa-165a1f28677f>.
- [21] ICFES, “Evaluación”, 2017. [Online]. Disponible: <https://www.icfes.gov.co/documents/20143/1353827/Guia+de+orientacion+saber+9+2017.pdf/fd46960-c1d4-96b2-ef0d-78b4c885bfcc>.
- [22] ICFES, “Metodología de pruebas por competencias”, 2017. [Online]. Disponible: <https://www.icfes.gov.co/documents/20143/495243/Guia%20de%20orientacion%20competencias%20especificas%20modulo%20de%20ensenar%20saber%20pro-2017.pdf>.
- [23] Sinnexus, “Persistencia HOLAP, MOLAP, ROLAP”, 2017. [Online]. Disponible: https://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
- [24] Tec_faq, “MOLAP”, 2018. [Online]. Disponible: <https://www.tech-faq.com/molap.html>
- [25] Dataprix, “Sistema ROLAP”, 2018. [Online]. Disponible: <https://www.dataprix.com/es/olap-molap-y-rolap>.
- [26] 1Library, “OLAP procesamiento analítico en línea”, 2018. [Online]. Disponible: <https://1library.co/article/olap-procesamiento-anal%C3%ADtico-l%C3%ADnea-sistema-soporte-decisiones-tecnolog%C3%ADa.y6eln8nz>.

- [27] ProyectosAgiles, “¿Qué es SCRUM?”, 2018. [Online]. Disponible: <https://proyectosagiles.org/que-es-scrum/>.
- [28] Softeng, “Procesos y roles SCRUM”, 2019. [Online]. Disponible: <https://www.softeng.es/es-es/empresa/metodologias-de-trabajo/metodologia-scrum/proceso-roles-de-scrum.html>.
- [29] Integrait, “Roles de SCRUM y sus responsabilidades”, 2020. [Online]. Disponible: <https://integrait.com.mx/blog/roles-de-scrum/>.
- [30] Rivadera, “la metodología Kimball para el diseño de almacenes de datos”, 2012. [Online]. Disponible: <https://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>
- [31] Todo sobre PostgreSQL, “Ventajas y desventajas”, 2018. [Online]. Disponible: <https://www.todopostgresql.com/ventajas-y-desventajas-de-postgresql/>.
- [32] Open Webinars, “Que es Python, características, evolución y futuro”, 2019. [Online]. Disponible: <https://openwebinars.net/blog/que-es-python/>.
- [33] MDN web Docs, “Introducción a Django”, 2019. [Online]. Disponible: <https://developer.mozilla.org/es/docs/Learn/Server-side/Django/Introduction>.
- [34] GoConq, “Características de Django”, 2019. [Online]. Disponible: https://cdn.goconqr.com/en/p/24640049?dont_count=true&frame=true&fs=true.
- [35] Jet brains, “Que es Pycharm”, 2019. [Online]. Disponible: <https://www.jetbrains.com/es-es/pycharm/>.
- [36] ARWEB, “Bootstrap y sus características”, 2019. [Online]. Disponible: <https://www.arweb.com/blog/%C2%BFque-es-bootstrap-y-como-funciona-en-el-diseno-web/>.
- [37] Horizonte, “Ventajas y desventajas de usar bootstrap”, 2018. [Online]. Disponible: <https://soyhorizonte.com/blog/ventajas-y-desventajas-de-usar-bootstrap/>.
- [38] Microsoft Docs, “Extracción, transformación y carga de dato (ETL)”, 2020. [Online]. Disponible: <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl>.

- [39] TechTarget, “ETL vs ELT dos opciones para extracción, carga y transformación de datos”, 2018. [Online]. Disponible: <https://searchdatacenter.techtarget.com/es/definicion/ETL-vs-ELT-dos-opciones-para-la-extraccion-carga-y-transformacion-de-datos>.
- [40] UNIWEBSIDAD, “El patrón de deiseño MTV”, 2018. [Online]. Disponible: <https://uniwebsidad.com/libros/django-1-0/capitulo-5/el-patron-de-diseno-mtv>.
- [41] Estadística para todos, “Diagrama de Caja y Bigotes”, 2008. [Online]. Disponible: <https://www.estadisticaparatodos.es/taller/graficas/cajas.html>.
- [42] Biplot usual, “Notas para biplot clásico”, 2016. [Online]. Disponible: <http://biplot.usal.es/multibiplot/documentation/notas-sobre-biplot-clasico-.pdf>.

ANEXOS

SABER EXPO BI 5: UNA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS WEB PARA LA GESTIÓN DE INDICADORES DE LAS PRUEBAS SABER 5 EN LAS INSTITUCIONES EDUCATIVAS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO

Anexo A

MANUAL DE DESARROLLADOR



INTRODUCCIÓN

En este anexo se describe la documentación técnica de programación, incluyendo la instalación del entorno de desarrollo, la estructura de la aplicación, su compilación, la configuración de los diferentes servicios de integración utilizados.

METODOLOGIA DE DESARROLLO

La investigación fue de tipo descriptivo bajo el enfoque cuantitativo. Para la construcción de la herramienta web Saber Expo BI 5 de limpieza, transformación y visualización de datos se utilizó la metodología ágil de desarrollo Scrum el cual mediante los Sprints permite realizar una investigación ágil y eficiente. Para la herramienta de inteligencia de negocios se propuso 5 Sprints que se profundizan posteriormente.

SPRINT	DETALLES
SPRINT 1. Planeación.	La apropiación de los conceptos básicos para el desarrollo del proyecto tales como: Inteligencia de Negocios. Mercado de datos. Cuadros de mando. Análisis multidimensional OLAP. Diseño y desarrollo de cuadros de mandos.
SPRINT 2. Diseño herramienta.	Se determinó los diferentes tipos de gráficos más adecuados y el maquetado de la vista en virtud del concepto de cada indicador, además se realizó el completo análisis del comportamiento del sistema.

SPRINT 3. Desarrollo herramienta.

En este Sprint se comenzó a definir los distintos paquetes necesarios para el desarrollo de las funcionalidades del aplicativo iniciando con todo lo referente a la visualización de los indicadores que abarca la aplicación haciendo uso de la librería Charts.js junto con otras utilidades. Lo siguiente fue la construcción del diseño multidimensional asociado al mercado de datos, realizando un diccionario de datos asociados a cada dimensión del diseño multidimensional, que fueron necesarios para el diseño del sistema de extracción, transformación y carga (ETL) para la creación de las bodegas de datos.

SPRINT 4. Realización de pruebas.

En esta fase se definieron y aplicaron diferentes tipos de pruebas para mejorar la calidad de la aplicación.

SPRINT 5. Implementación y documentación.

Se subió el aplicativo a un servidor el cual se encuentra en constante actualización y mejoramiento gracias a la herramienta Git Por último se realizó la respectiva documentación con respecto a los manuales de usuario y manuales del programador.

ENTORNO DE DESARROLLO

Para trabajar con el proyecto se necesita tener instalados los siguientes programas y dependencias:

- PostgreSQL
- Python
- Django
- Pycharm
- Visual Studio Code

A continuación, se indica como instalar y configurar correctamente cada uno de ellos.

PostgreSQL

PostgreSQL es el motor de base de datos de la actualidad. Lo que lo hace el Sistema Gestor de Bases de Datos de Código libre más potente y robusto del mercado.

Ventajas PostgreSQL

- **Instalación ilimitada y gratuita:** Podemos instalarlo en todos los equipos que se desee, independientemente de la plataforma y la arquitectura que usemos, PostgreSQL está disponible para los diferentes SO, Unix, Linux y Windows, en 32 y 64 bits, lo que lo hace un sistema multiplataforma y de gran escalado.
- **Gran escalabilidad:** Nos permite configurar PostgreSQL en cada equipo según el hardware; Por esto, es capaz de ajustarse al número de CPU y a la cantidad de memoria disponible de forma óptima. Con ello logramos una mayor cantidad de peticiones simultáneas a la base de datos de forma correcta.
- **Estabilidad y confiabilidad:** Tiene más de 20 años de desarrollo activo y en constante mejora. No se han presentado nunca caídas de la base de datos que estén documentadas

gracias a su capacidad de establecer un entorno de Alta disponibilidad en donde podemos hacer tareas de mantenimiento o recuperación sin bloquear completamente el sistema.

- **Potencia y Robustez:** PostgreSQL cumple en su totalidad con la característica ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad en español). Por ello permite que las transacciones no interfieran unas con otras. Con ello se garantiza la información de las Bases de Datos y que los datos perduren en el sistema.
- **Extensibilidad:** tenemos a nuestra disponibilidad una gran variedad de extensiones distribuidas por el grupo de desarrolladores de PostgreSQL. Estas extensiones pueden ser lenguajes de programación, tales como, Perl, Java, Python, C++ y muchos más.

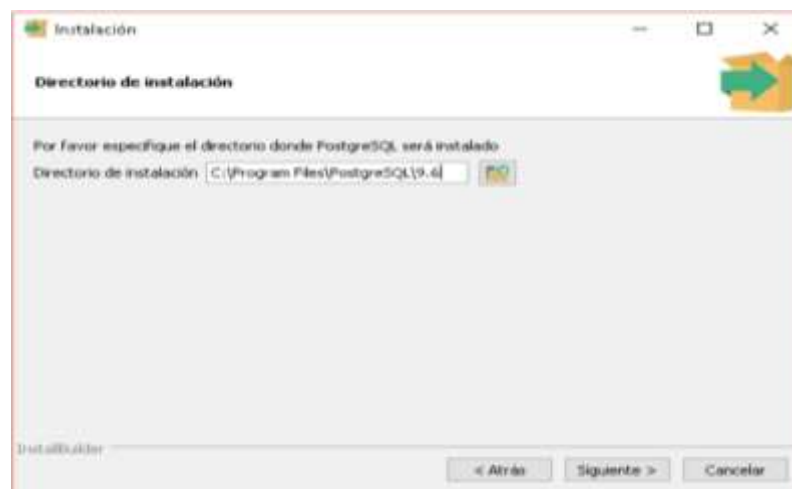
Desventajas de PostgreSQL

- **Base de datos pequeñas:** Es relativamente lento en inserciones y actualizaciones en bases de datos pequeñas, PostgreSQL está diseñado para ambientes de alto volumen, por ello la velocidad de respuesta pueda parecer lenta en comparación con bases de datos de pequeño tamaño.
- **Soporte oficial:** No cuenta con un soporte en línea o telefónico, por el contrario, cuenta con foros oficiales donde los usuarios pueden exponer sus dudas que responden, es de resaltar que la comunidad PostgreSQL es una de las más activas del mercado.
- **Sintaxis:** La sintaxis de algunos de sus comando o sentencias puede llegar a no ser intuitiva si no tienes un nivel medio de conocimientos en lenguaje SQL.

Instalación de PostgreSQL

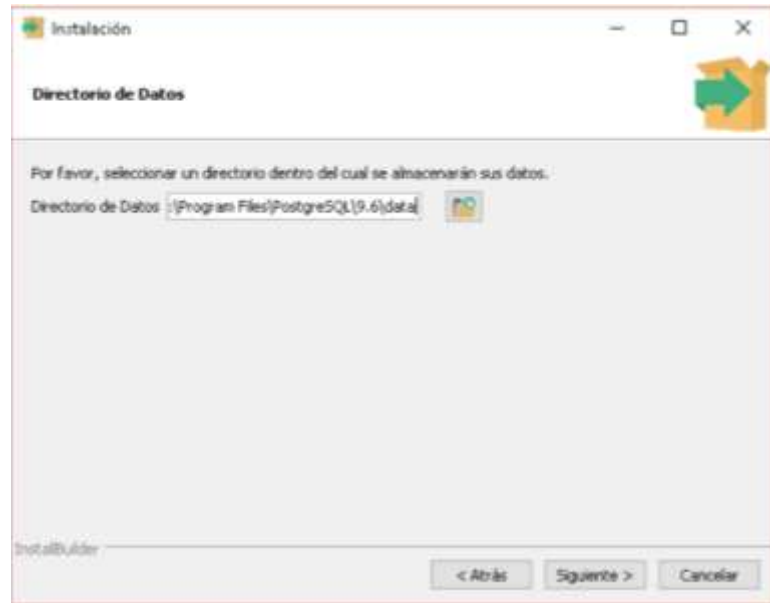
- Para instalar PostgreSQL nos dirigimos a nuestro navegador y accedemos a la siguiente URL: <https://www.enterprisedb.com/downloads/postgres-postgresql-downloads> nos direccionará a otra página donde podremos descargarnos la versión correcta tanto para nuestras necesidades como para el tipo de sistema operativo que tengamos instalado.

- En nuestro caso es la versión 13.1 para Windows de 64 bits. Una vez descargado el archivo de instalación lo ejecutamos y si tenemos activado el control de cuentas de usuario nos mostrará una advertencia con el texto "¿Desea permitir que este programa realice cambios en el equipo?", pulsaremos "Sí" para continuar con la instalación de PostgreSQL; Se iniciará el asistente para instalar PostgreSQL, pulsaremos "Siguiente".
- Indicaremos los complementos que queremos instalar y la carpeta de instalación de PostgreSQL, donde se guardarán los ejecutables, librerías y ficheros de configuración de PostgreSQL.

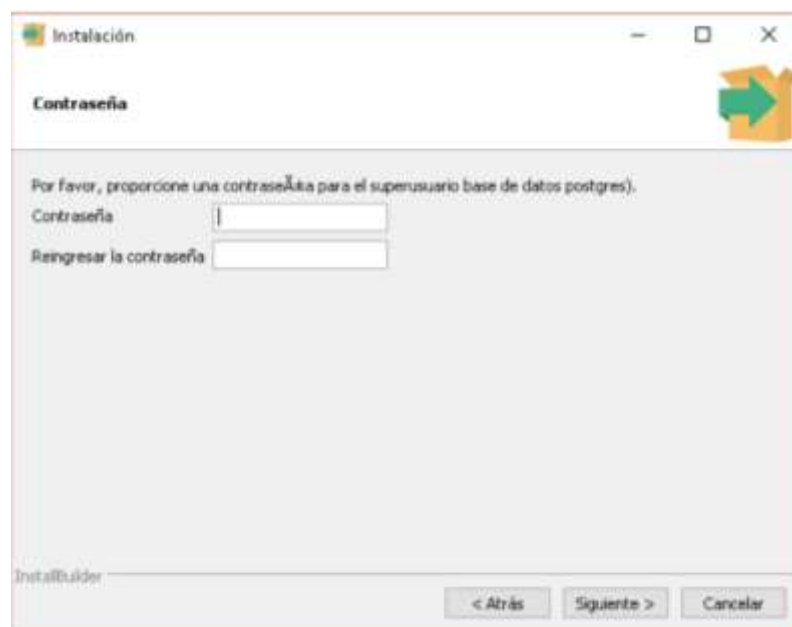


- Indicaremos también la carpeta donde se guardarán los datos por defecto de PostgreSQL.

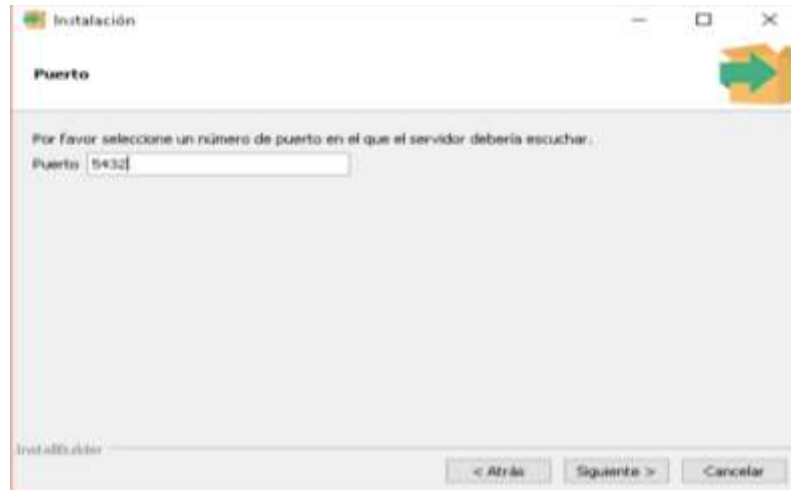




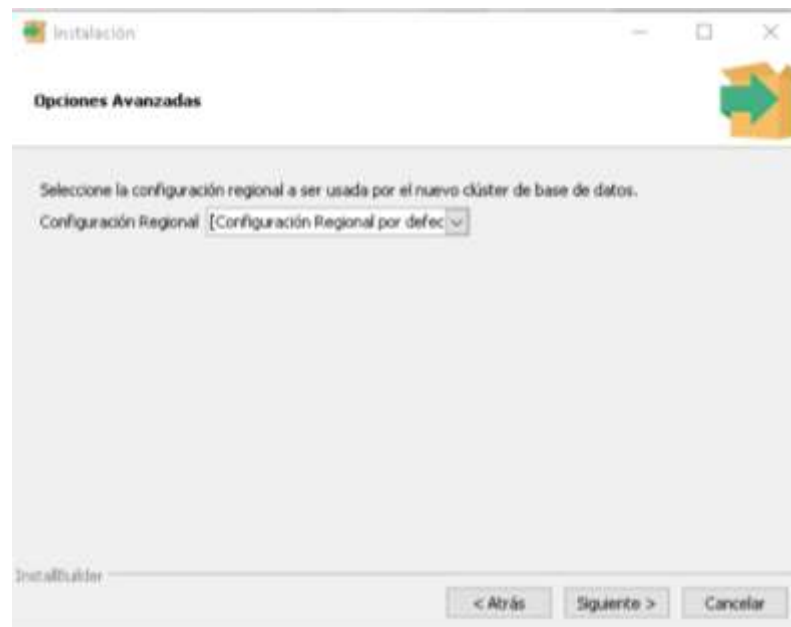
- Introduciremos la contraseña para el súper usuario "postgres" que será con el que iniciemos sesión para administrar la base de datos, si tenemos clave de cuenta en Windows introduciremos nuestra clave.



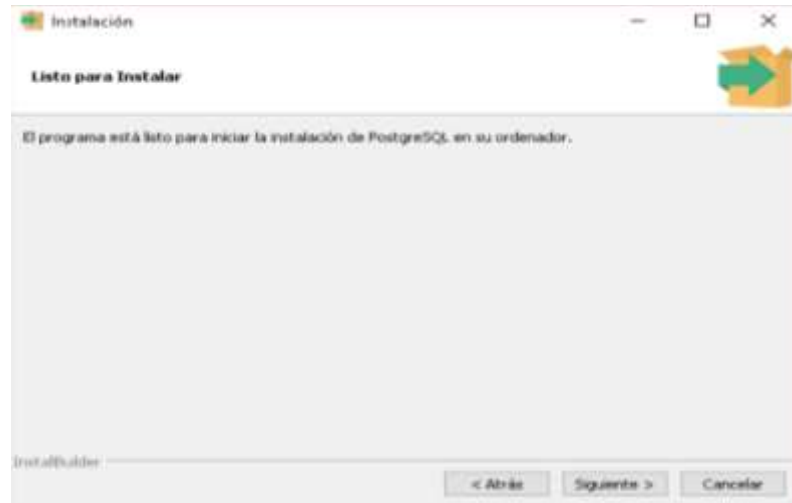
- Introduciremos el puerto de escucha para la conexión con el servidor PostgreSQL, por defecto el 5432.



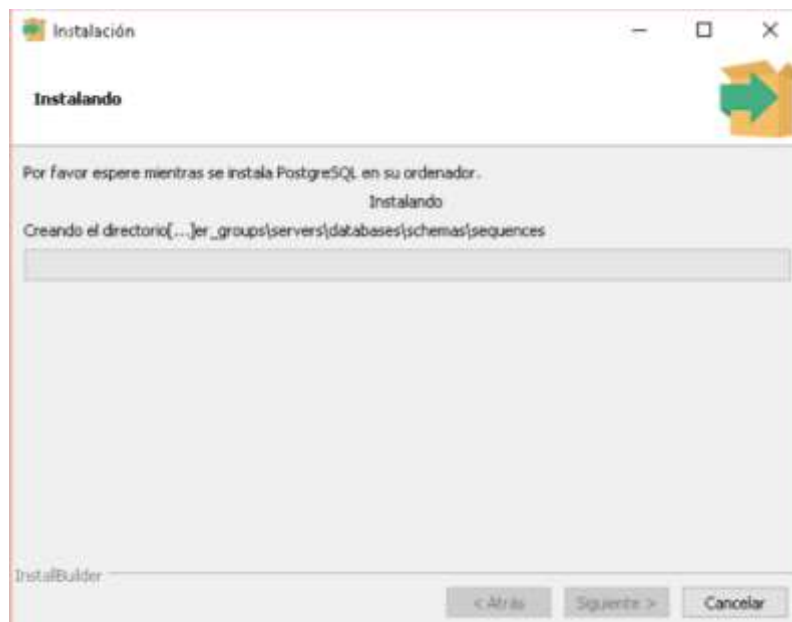
- Seleccionaremos la configuración regional.



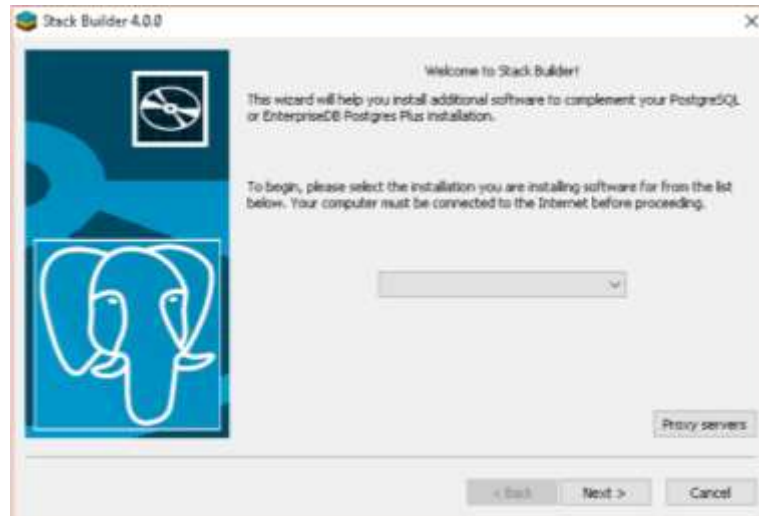
- Pulsaremos "Siguiente" para iniciar la instalación definitiva.



- Se iniciará el asistente para instalar el motor de base de datos PostgreSQL, que creará las carpetas oportunas, copiará los ficheros necesarios y creará el servicio Windows para iniciar de forma automática el motor de base de datos.



- Una vez finalizada la instalación el asistente nos dará la posibilidad de ejecutar Stack Builder, aplicación que nos permitirá instalar otros componentes y herramientas para PostgreSQL.



- Si hemos marcado la opción de Stack Builder, se iniciará, seleccionaremos "PostgreSQL on port 5432". Pulsaremos "Finish" (en nuestro caso cancelaremos Stack Builder pues no instalaremos más componentes).



Python

Es un lenguaje de programación versátil multiplataforma y multiparadigma que se destaca por su código legible y limpio. Una de las razones de su éxito es que cuenta con una licencia de código abierto que permite su utilización en cualquier escenario.

Python es ideal para trabajar con grandes volúmenes de datos ya que, al ser multiplataforma, favorece su extracción y procesamiento, por eso lo eligen las empresas de Big Data. A nivel científico, tiene una gran biblioteca de recursos con especial énfasis en las matemáticas para aspirantes a programadores en áreas especializadas. Y si esto fuera poco, es posible crear videojuegos, aunque no es tan eficiente como Java o C#.

Ventajas de Python

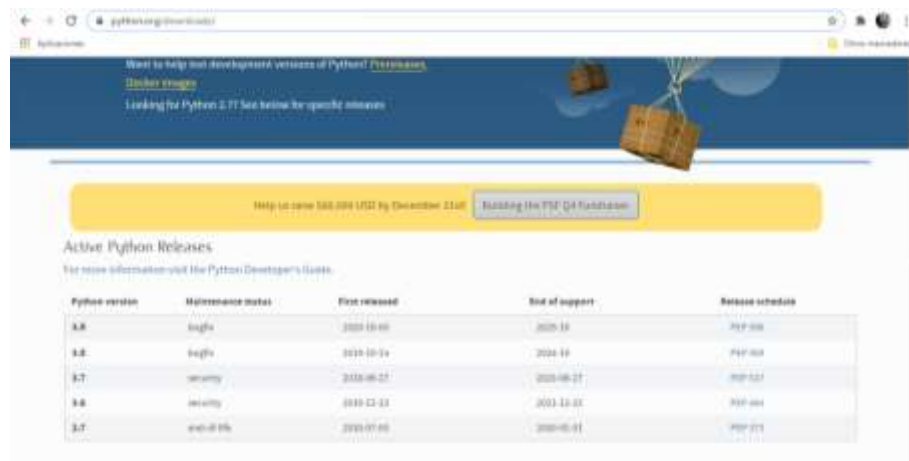
- **Simplificado y rápido:** Este lenguaje simplifica mucho la programación, es un gran lenguaje para scripting.
- **Elegante y flexible:** El lenguaje ofrece muchas facilidades al programador al ser fácilmente legible e interpretable.
- **Programación sana y productiva:** Es sencillo de aprender, con una curva de aprendizaje moderada. Es muy fácil comenzar a programar y fomenta la productividad.
- **Ordenado y limpio:** es muy legible y sus módulos están bien organizados.
- **Portable:** Es un lenguaje muy portable. Podemos usarlo en prácticamente cualquier sistema de la actualidad.
- **Comunidad:** Cuenta con un gran número de usuarios. Su comunidad participa activamente en el desarrollo del lenguaje.

Desventajas de Python

- **Hosting:** La mayoría de los servidores no tienen soporte a Python, y si lo soportan, la configuración es un poco difícil.
- **Librerías incluidas:** Algunas librerías que trae por defecto no son del gusto de amplio de la comunidad, y optan a usar librerías de terceros.

Instalación de Python

- Para la instalación accedemos desde nuestro navegador al sitio oficial de Python desde la siguiente URL <https://www.python.org/downloads/> y descargamos el instalador ejecutable de Python según los requerimientos de nuestro computador.



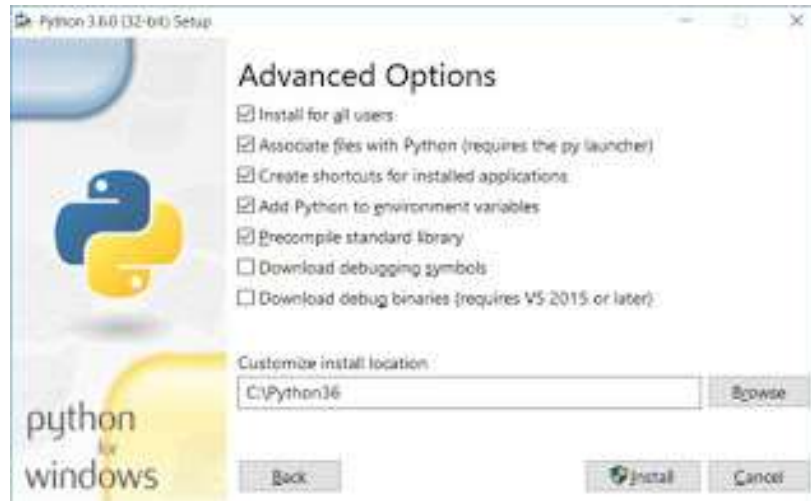
- Iniciar la instalación ejecutando el archivo descargado Pytho.exe con doble clic. Si es necesario, confirmar la ejecución en la ventana de advertencia, una vez iniciado el instalador, en la ventana Install Python activar las casillas de las opciones: Install launcher for all users (recommended) y Add Python 3.6 to PATH. Después, continuar seleccionando la opción Customize installation. Choose Location and features.



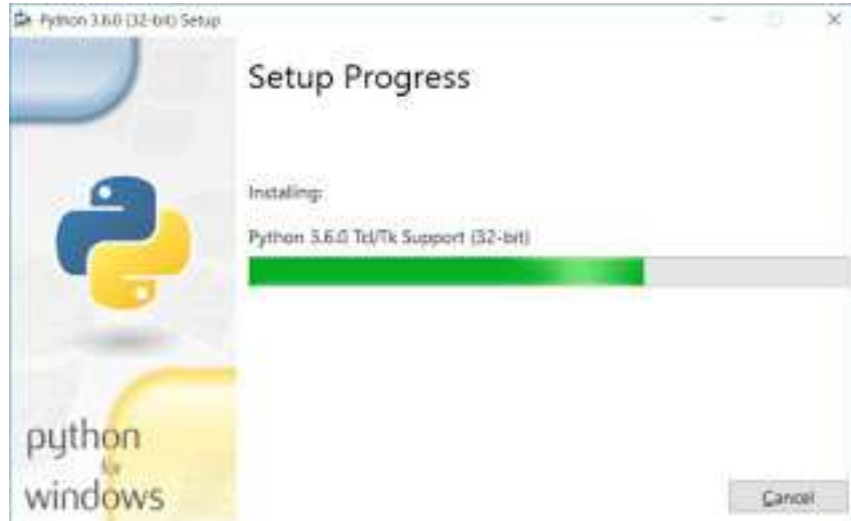
- En la ventana Optional features verificar que están activas todas las opciones y hacer clic en el botón Next.



- En la ventana Advanced Options verificar que están activas las opciones de la imagen, escribir la ruta del directorio de instalación "C:\Python36" (o aceptar la ruta por defecto) y comenzar la instalación haciendo clic en el botón Install.



- A continuación, después de unos segundos de espera, comenzará el proceso de instalación de Python.

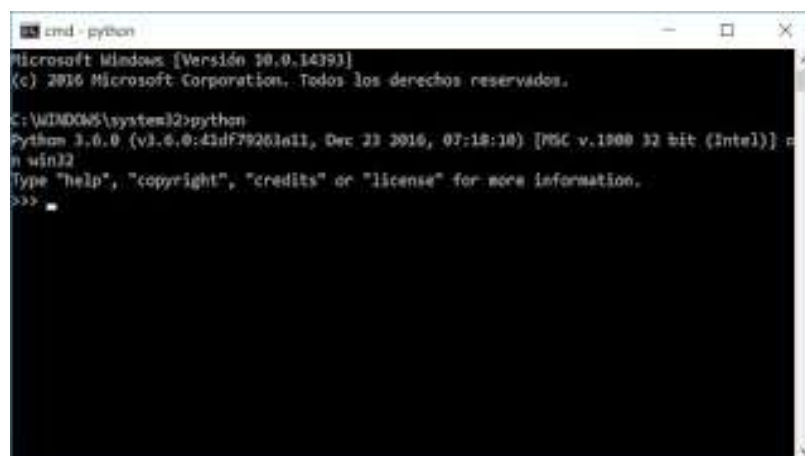


- En la ventana Setup was successful, una vez que ha concluido el proceso de instalación hacer clic en el botón Close. Desde esta ventana es posible acceder a un tutorial online de

Python, a la documentación oficial del lenguaje y a información con las novedades de la presente versión.



- Para probar Python acceder a la línea de comandos ejecutando el programa cmd.exe o Símbolo del sistema. Después, escribir "python" y aceptar para iniciar el entorno interactivo de Python donde se puede verificar la versión de Python instalada y es posible introducir comandos Python. Para finalizar, introducir quit() o presionar Ctrl+D.



Django

Django es un framework web de alto nivel que permite el desarrollo rápido de sitios web seguros y mantenibles. Desarrollado por programadores experimentados, Django se encarga de gran parte de las complicaciones del desarrollo web, por lo que puedes concentrarte en escribir tu aplicación sin necesidad de reinventar la rueda. Es gratuito y de código abierto, tiene una comunidad próspera y activa, una gran documentación y muchas opciones de soporte gratuito y de pago.

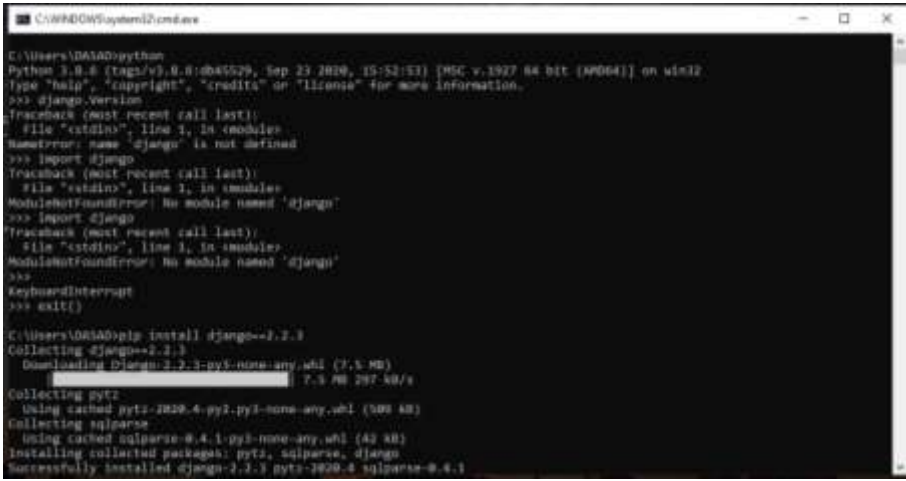
Características:

- **Completo:** Django sigue la filosofía "Baterías incluidas" y provee casi todo lo que los desarrolladores quisieran que tenga "de fábrica". Porque todo lo que necesitas es parte de un único "producto", todo funciona a la perfección, sigue principios de diseño consistentes y tiene una amplia y actualizada documentación.
- **Versátil:** Django puede ser para construir casi cualquier tipo de sitio web, desde sistemas manejadores de contenidos y wikis, hasta redes sociales y sitios de noticias. Puede funcionar con cualquier framework en el lado del cliente.
- **Seguro:** Django ayuda a los desarrolladores evitar varios errores comunes de seguridad al proveer un framework que ha sido diseñado para "hacer lo correcto" para proteger el sitio web automáticamente. Por ejemplo, Django, proporciona una manera segura de administrar cuentas de usuario y contraseñas, evitando así errores comunes como colocar informaciones de sesión en cookies donde es vulnerable
- **Escalable:** Tiene en cuenta una clara separación entre las diferentes partes significa que puede escalar para aumentar el tráfico al agregar hardware en cualquier nivel, servidores de cache, servidores de bases de datos o servidores de aplicación.

- **Mantenible:** El código de Django está escrito usando principios y patrones de diseño para fomentar la creación de código mantenible y reutilizable. En particular, utiliza el principio No te repitas "Don't Repeat Yourself" (DRY) para que no exista una duplicación innecesaria, reduciendo la cantidad de código.
- **Portable:** Django está escrito en Python, el cual se ejecuta en muchas plataformas. Lo que significa que no está sujeto a ninguna plataforma en particular, y puede ejecutar sus aplicaciones en muchas distribuciones de Linux, Windows y Mac OS X.

Instalación de Django

Para instalar Django, se debió instalar Python previamente, pues este es un framework que está escrito en Python y basta con abrir el intérprete de comandos de Windows y escribir el comando “pip install django==2.2.3” y se comenzaran a descargar los repositorios de Django en nuestro ordenador.



```
C:\Users\DA5AD>python
Python 3.8.1 (tags/v3.8.1:ba55579, Sep 23 2019, 15:52:51) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> django.VERSION
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'django' is not defined
>>> import django
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'django'
>>> import django
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'django'
>>>
KeyboardInterrupt
>>> exit()

C:\Users\DA5AD>pip install django==2.2.3
Collecting django==2.2.3
  Downloading Django-2.2.3-py3-none-any.whl (7.5 MB)
    7.5 MB 297 kB/s
Collecting pytz
  Using cached pytz-2020.4-py3-none-any.whl (500 kB)
Collecting sqlparse
  Using cached sqlparse-0.4.1-py3-none-any.whl (42 kB)
Installing collected packages: pytz, sqlparse, django
Successfully installed django-2.2.3 pytz-2020.4 sqlparse-0.4.1
```

Para comprobar que Django esté instalado escribimos en el interprete de comandos la palabra “Python” y dentro de Python escribimos “import django” y luego “django.VERSION”, si esta todo correcto podremos visualizar la versión de Django que hemos instalado.

```
C:\Users\DA5AD>python
>>>
KeyboardInterrupt
>>> exit()

C:\Users\DA5AD>pip install django==2.2.3
Collecting django==2.2.3
  Downloading Django-2.2.3-py3-none-any.whl (7.5 MB)
    100% |#####| 7.5 MB 287 KB/s
Collecting pytz
  Using cached pytz-2020.4-py3-none-any.whl (509 KB)
Collecting sqlparse
  Using cached sqlparse-0.4.1-py3-none-any.whl (42 KB)
Installing collected packages: pytz, sqlparse, django
Successfully installed django-2.2.3 pytz-2020.4 sqlparse-0.4.1
WARNING: You are using pip version 19.3.1; however, version 20.2.1 is available.
You should consider upgrading via the 'C:\Users\DA5AD\AppData\Local\Programs\Python\Python38\python.exe -m pip install --upgrade pip' command.

C:\Users\DA5AD>python
Python 3.8.6 (tags/v3.8.6:2b45529, Sep 22 2020, 15:32:13) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> import django
>>> django.VERSION
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: module 'django' has no attribute 'VERSION'
>>> django.VERSION
(2, 2, 3, 'final', 0)
>>>
```

PyCharm

Es un IDE o entorno de desarrollo integrado multiplataforma utilizado para desarrollar en el lenguaje de programación Python. Proporciona análisis de código, depuración gráfica, integración con VCS / DVCS y soporte para el desarrollo web con Django, entre otras bondades. PyCharm es desarrollado por la empresa JetBrains y debido a la naturaleza de sus licencias tiene dos versiones, la Community que es gratuita y orientada a la educación y al desarrollo puro en Python y la Professional, que incluye más características como el soporte a desarrollo web con varios precios.

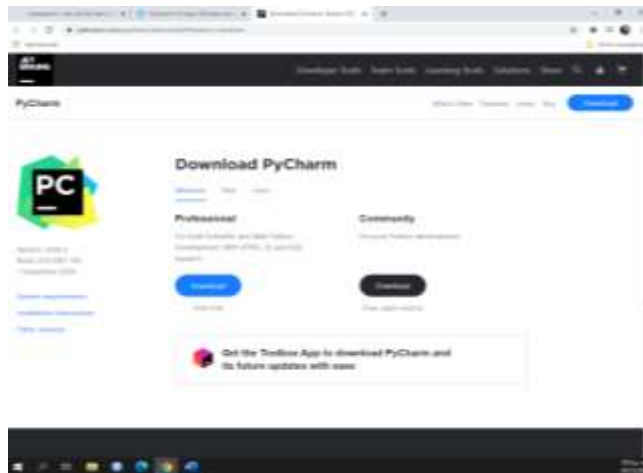
Características:

- Resaltado de sintaxis, sangría automática y formateo de código, finalización de código, formateador de código, resaltado de errores sobre la marcha, análisis del código a medida que se escribe y detector de código duplicado.
- Depurador gráfico en local.
- Refactorizaciones rápidas y seguras válidas para nombres de fichero, funciones, constantes, clases, propiedades, métodos, parámetros y variables locales y globales.

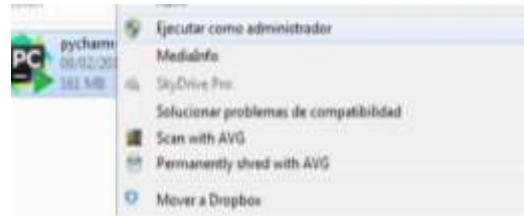
- Busca en todas partes, encontrar usos de cualquier símbolo, ir a clase/archivo/símbolo, ir a declaración, marcadores personalizados.
- Integración con sistemas de control de versiones: Git, Mercurial, CVS, Subversion, GitHub.
- Integración con documentación: reStructuredText y Google.
- Integración con PyQt y PyGTK.
- Gestión de paquetes.
- Integración nativa con Virtualenv/Buildout.
- Consola de Python.
- Soporte para XML, HTML, YAML, JSON, RelaxNG.
- Terminal local.

Instalación de PyCharm

- La versión gratuita Pycharm Community se puede descargar de la siguiente dirección:
<https://www.jetbrains.com/pycharm/download/>



- Para poder usar PyCharm es necesario tener una licencia, podemos elegir la versión community en la cual debemos ingresar una licencia estudiantil que podremos gestionar en la siguiente página, el único requisito es tener una cuenta de correo institucional; la licencia se puede obtener de la siguiente URL:
<https://www.jetbrains.com/es-es/commnity/education/?fromMenu#students>
- Ejecuta como administrador el archivo que se acaba de descargar.



- Pulsar siguiente.



- Elegir la ruta donde se desee instalar y pulsar siguiente.





- Pulsar siguiente.



- Seleccionamos la carpeta jetbrains y pulsamos instalar.



- Seleccionamos la opción de correr el programa y pulsamos finalizar.

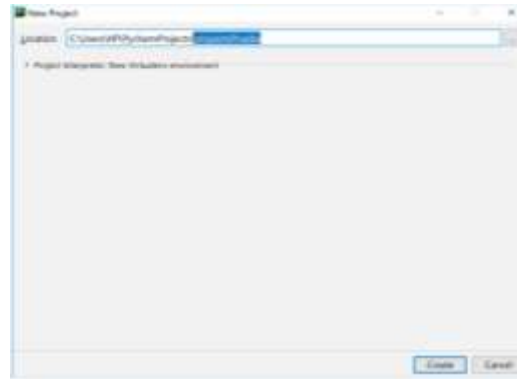


Crear nuevo proyecto PyCharm con Python

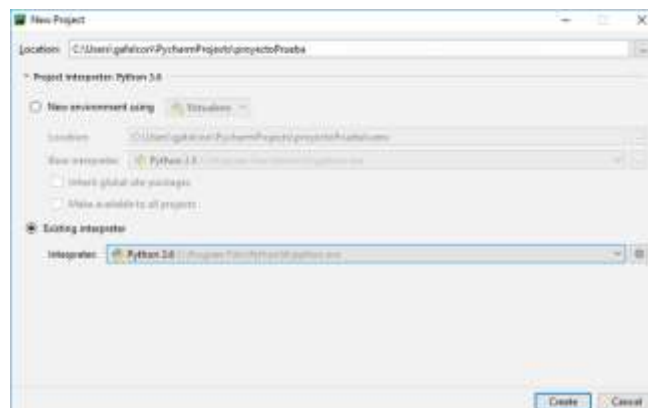
- Después de haber instalado PyCharm procederemos a crear un nuevo proyecto.



- En la pantalla siguiente, escribir un nuevo nombre a nuestro nuevo proyecto. en este caso, hemos escrito “proyectoPrueba”

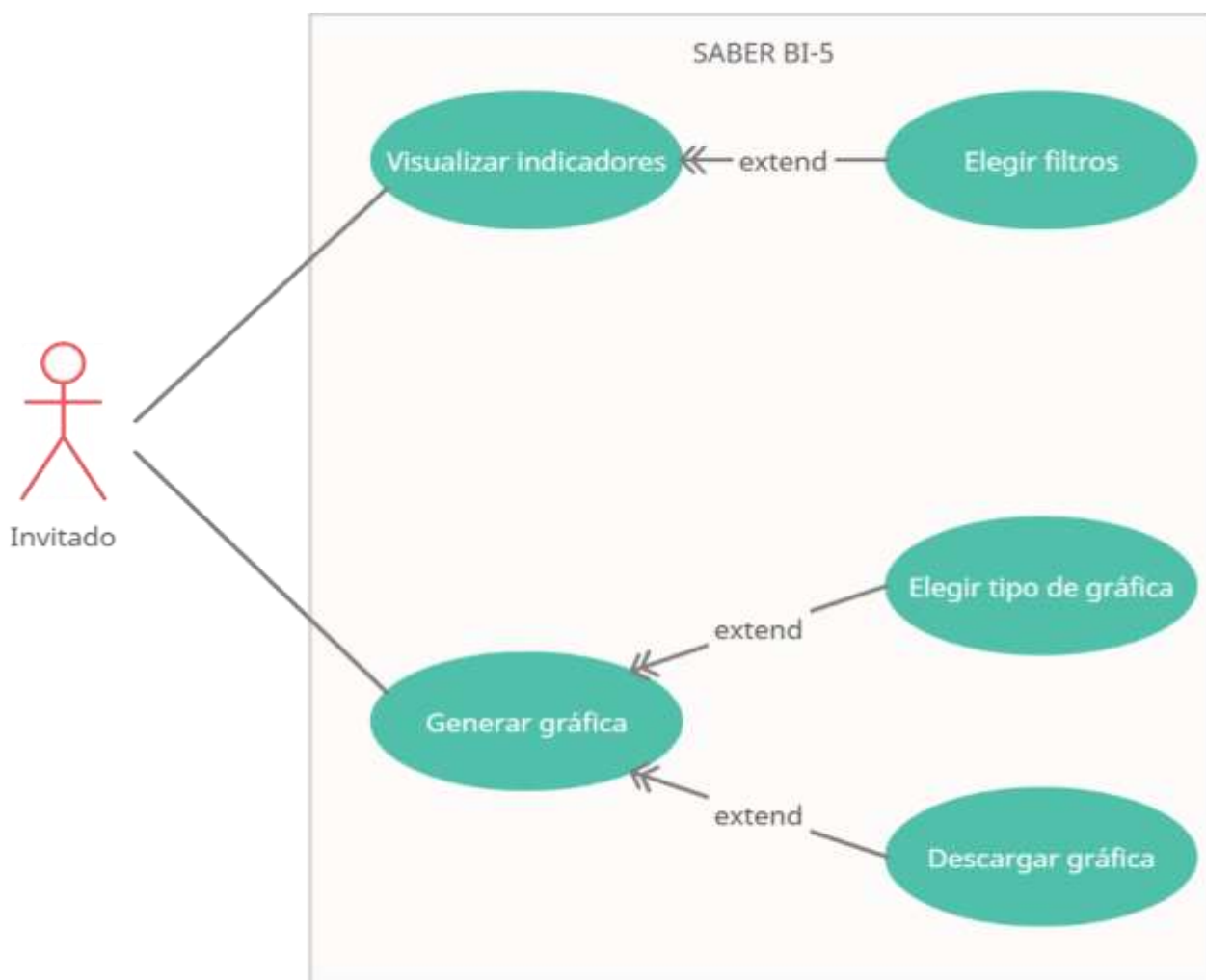


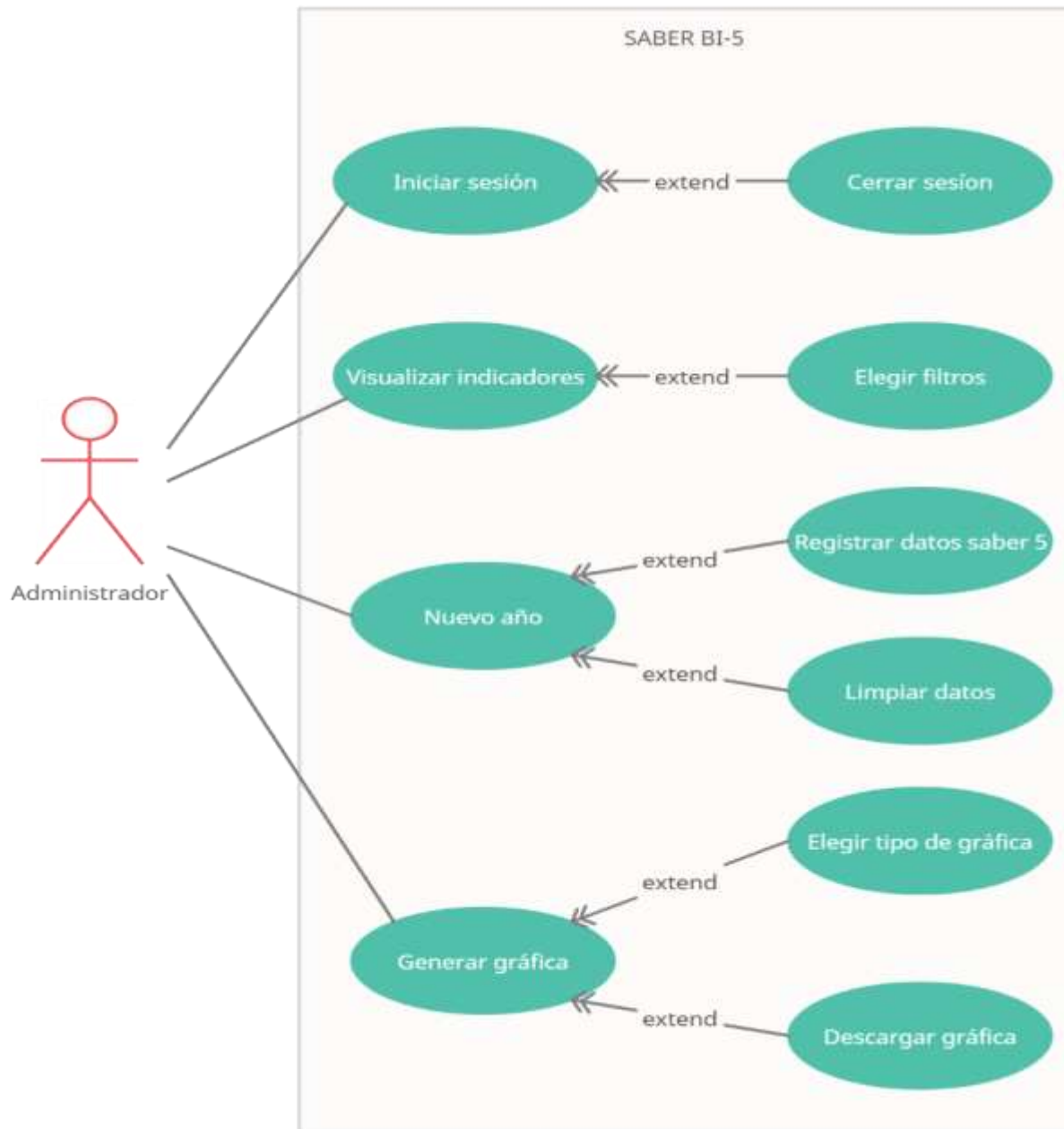
- En la opción Project Interpreter debemos seleccionar el intérprete de Python a utilizar en el proyecto. Pycharm nos presenta dos opciones:
 - ✓ Nuevo ambiente (New Environment): Si seleccionamos esta opción, pycharm instalará un intérprete de python exclusivo para el proyecto a crear. De esta manera podremos tener diferentes intérpretes de python para cada proyecto.
 - ✓ Intérprete existente (Recomendado): Pycharm nos permitirá escoger el intérprete de python que tengamos instalado en nuestra computadora. Se recomienda utilizar esta opción que utilizará el intérprete de Python que instalamos previamente.
- Una vez seleccionado el intérprete, hacemos click en “Crear” y se creará un nuevo proyecto.



DISEÑO FUNCIONAL

Las características funcionales del sistema se ilustran por medio de diagramas de Casos de Uso, los cuales se concentran en dos roles que interactúan con el sistema, el usuario general y el usuario administrador representados en los siguientes diagramas.





DISEÑO ESTRUCTURAL

Aquí se explica las características estructurales de un sistema general, la arquitectura y distribución de artefactos de software requerida.



Diagrama De Paquetes

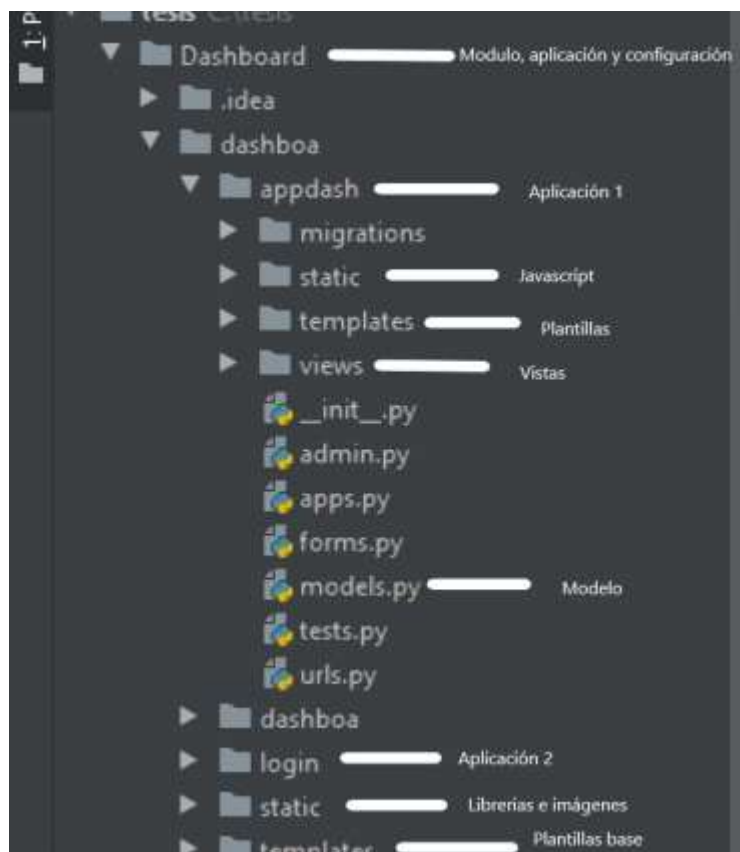
En el desarrollo de la aplicación se maneja Python con un Framework el cual es un entorno Web llamado Django y nos permite trabajar con una estructura MTV (model-template-view),

A continuación, se detalla el manejo de dicha estructura:

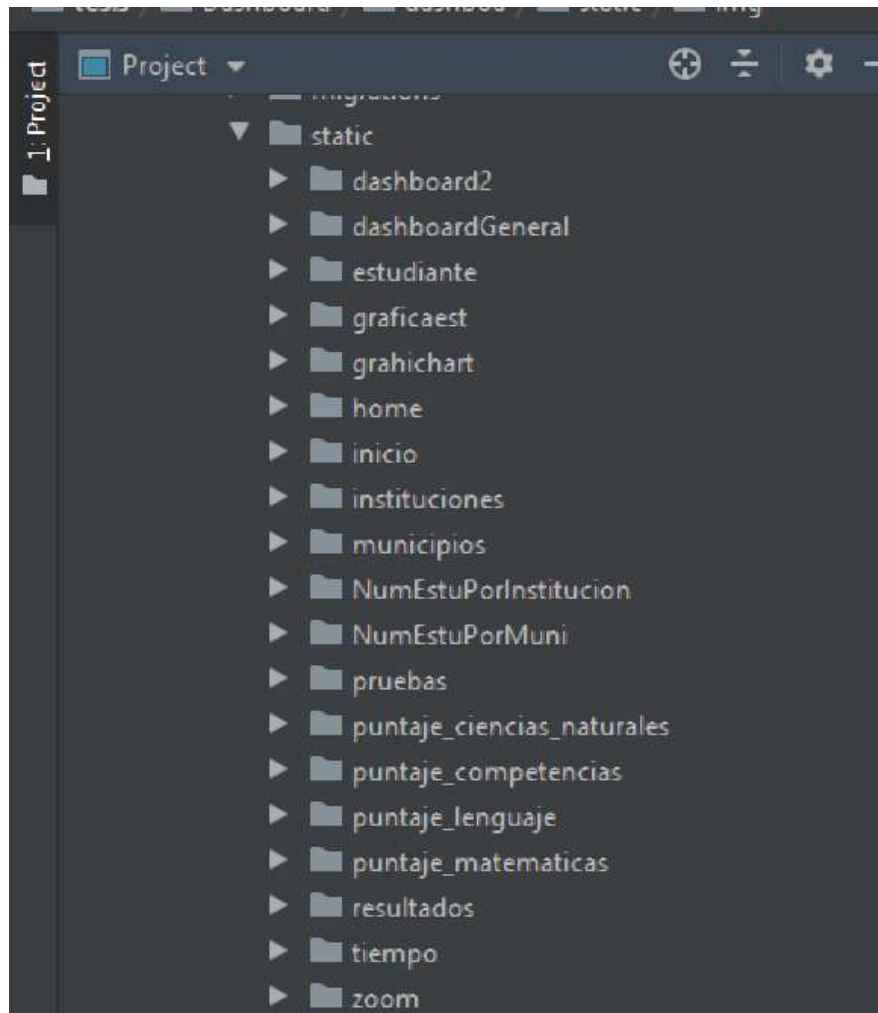
M = Modelos (Base de datos): Capa de interacción y comunicación que muestra un archivo donde podemos observar las clases que detallan las diferentes dimensiones manejadas en el proyecto.

V = Vista (Funciones): permite interactuar con el modelo y es donde encontramos la lógica con la cual realizamos las diferentes operaciones para generar los diferentes resultados.

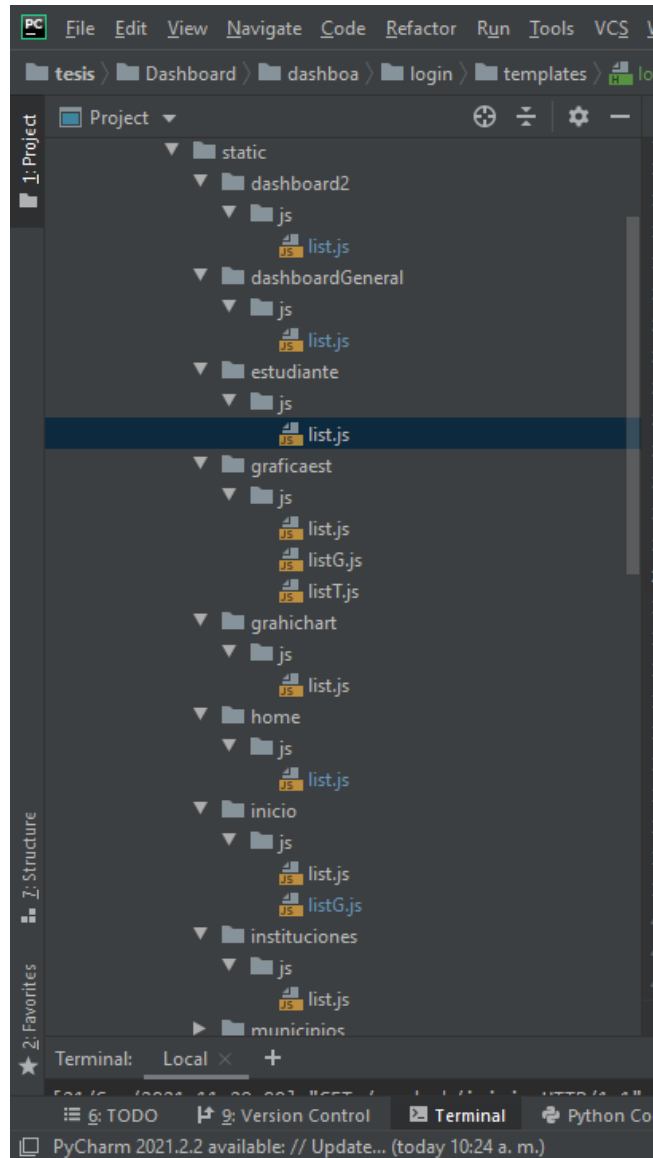
T = Templates (Plantillas): son los diferentes archivos que se encuentran en la capa de presentación y nos permite mostrar los resultados a los diferentes usuarios. A continuación, se observa las rutas o directorios donde se encuentran cada uno de los diferentes archivos de configuración para el desarrollo de la aplicación. La imagen nos permite observar de forma general los directorios con sus respectivos nombres en los cuales se encuentran las Plantillas, Vistas, Modelos y Scripts además archivos de configuración para su conexión entre los diferentes archivos utilizados en la aplicación.



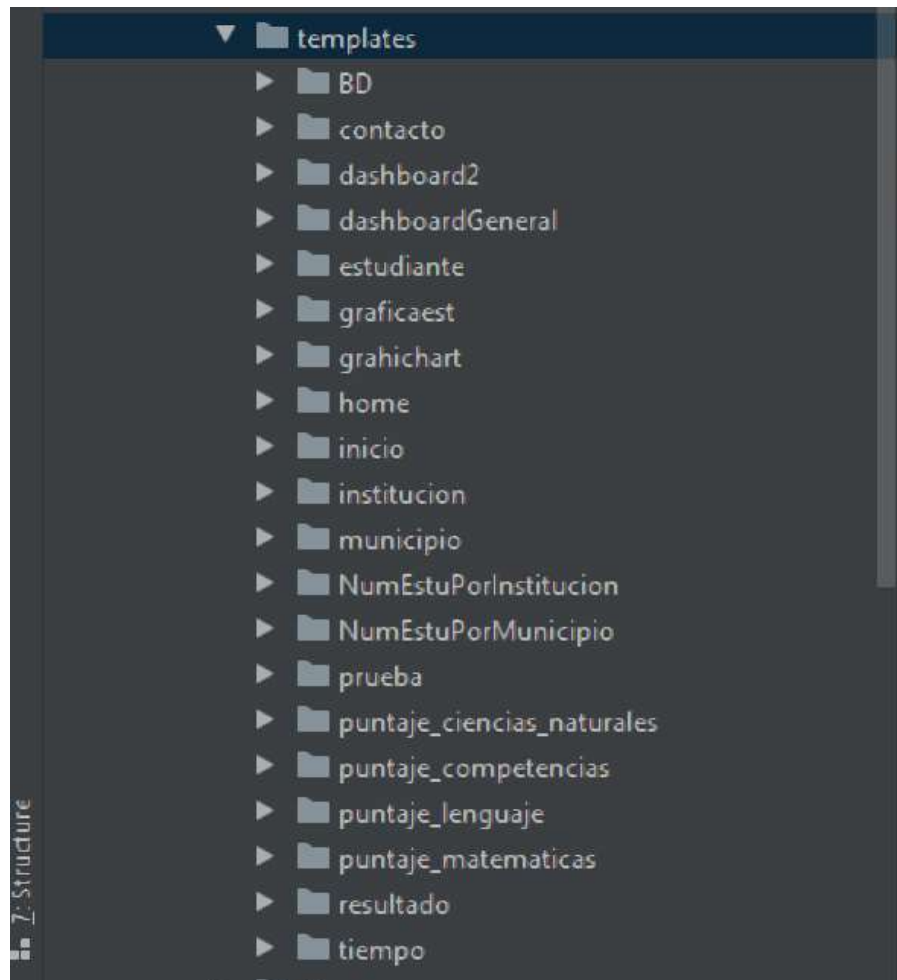
En la siguiente imagen podemos visualizar los diferentes directorios donde se encuentran todos los archivos Scripts para cada plantilla los cuales permite recibir los datos que vienen desde el modelo conectado con la vista y la BD.



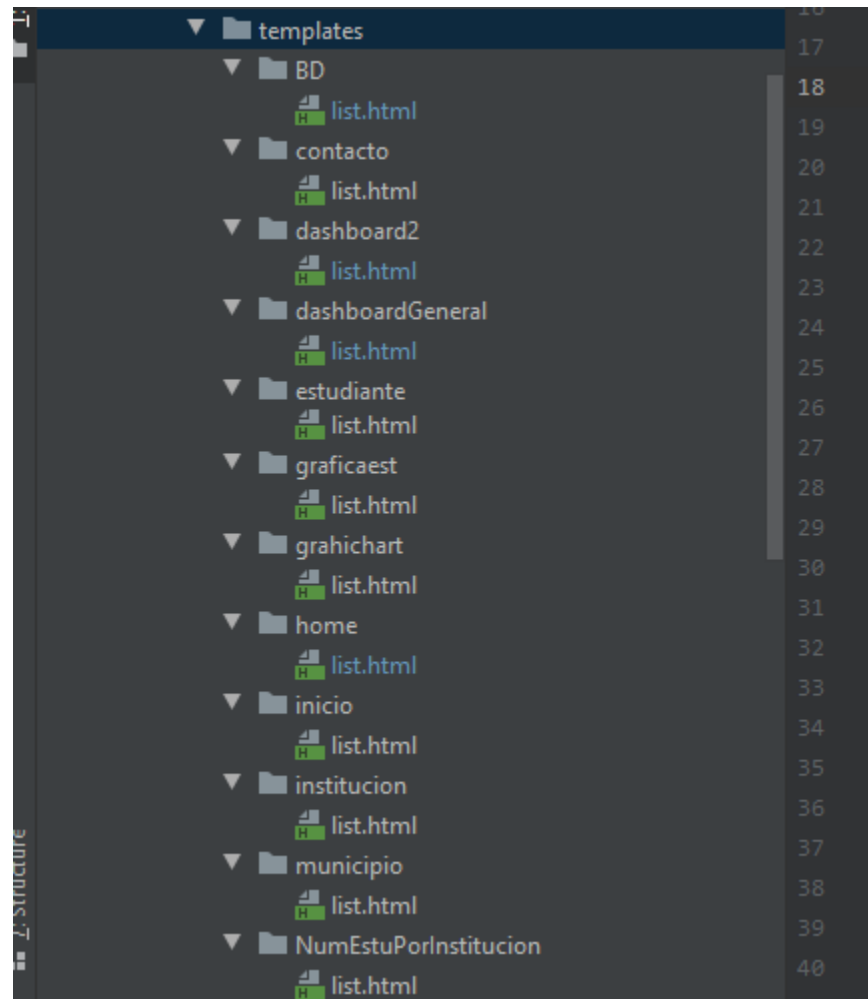
En la siguiente imagen podemos observar desplegados los diferentes directorios que contienen cada uno de los archivos Scripts (JS) que son utilizados con cada plantilla y vista para la visualización de los datos.



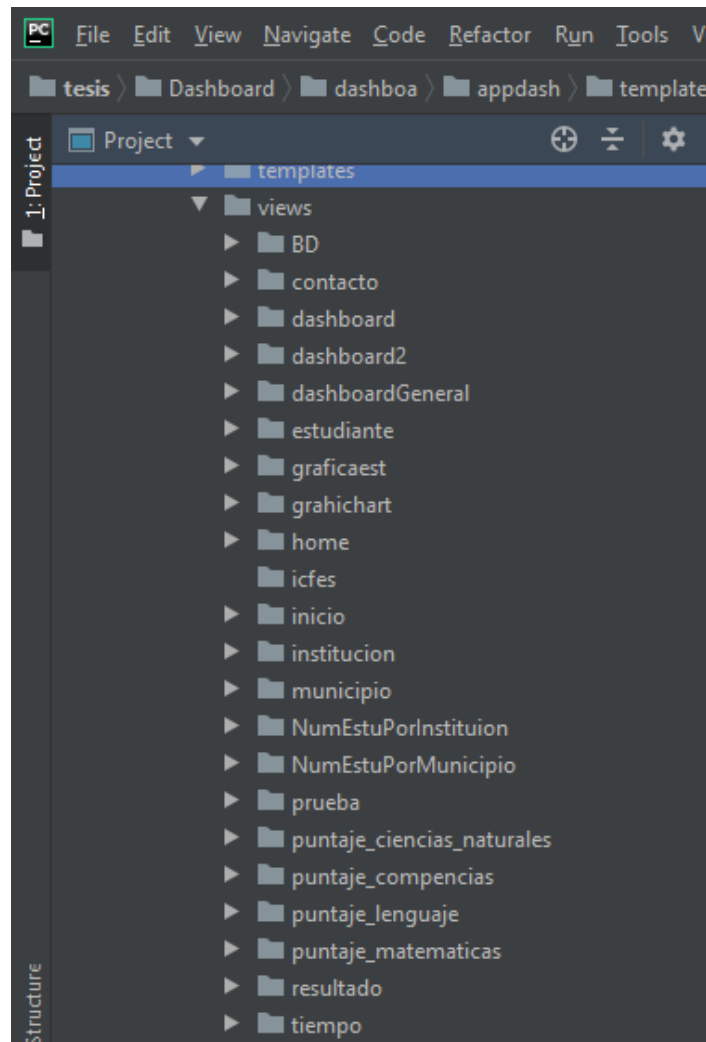
En la siguiente grafica se encuentran las plantillas utilizadas para la visualización de las gráficas y tablas para el usuario.



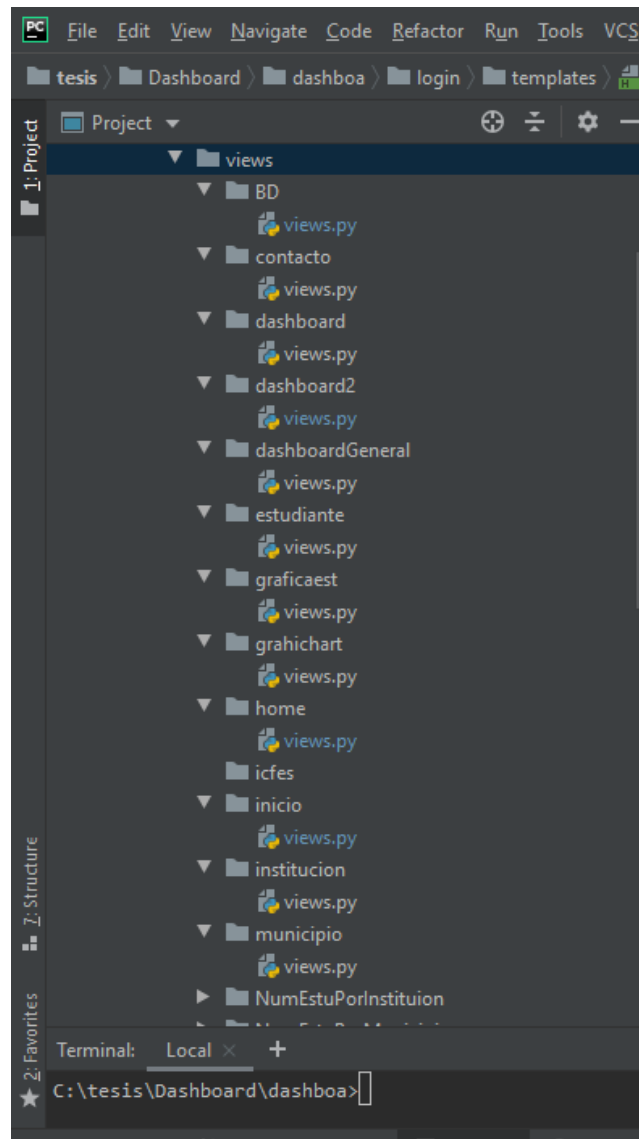
Ahora se visual los diferentes directorios desplegados con sus respectivos archivos con extensión (HTML).



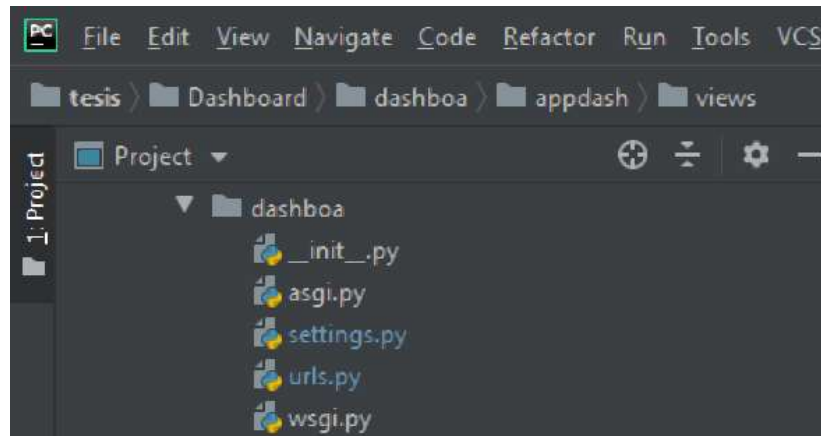
En la siguiente imagen podemos visualizar los directorios que contienen las diferentes vistas utilizadas para el manejo de la lógica en la aplicación.



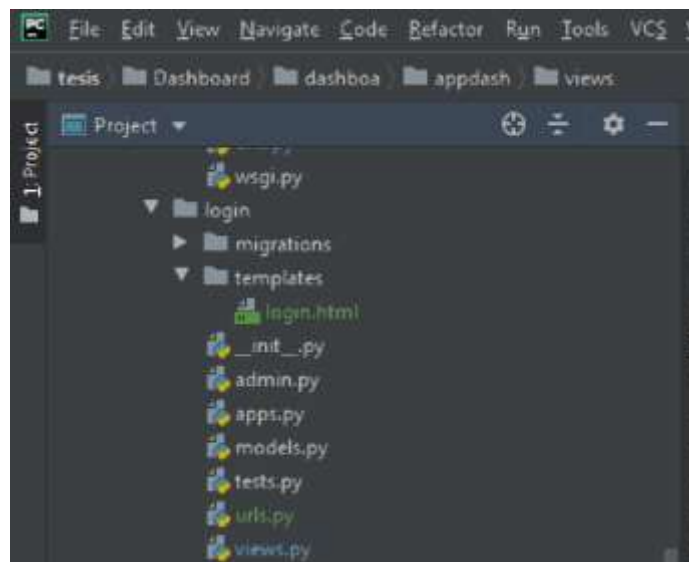
En la siguiente grafica podemos observar desplegados los diferentes directorios donde se encuentran los archivos que contienen las respectivas vistas.



En la siguiente imagen podemos observar el directorio dashboard el cual contiene el archivo urls.py en el cual podemos encontrar las diferentes rutas las cuales permiten conectar con las diferentes vistas para tener acceso a la información de la base de datos según la lógica descrita en cada vista.

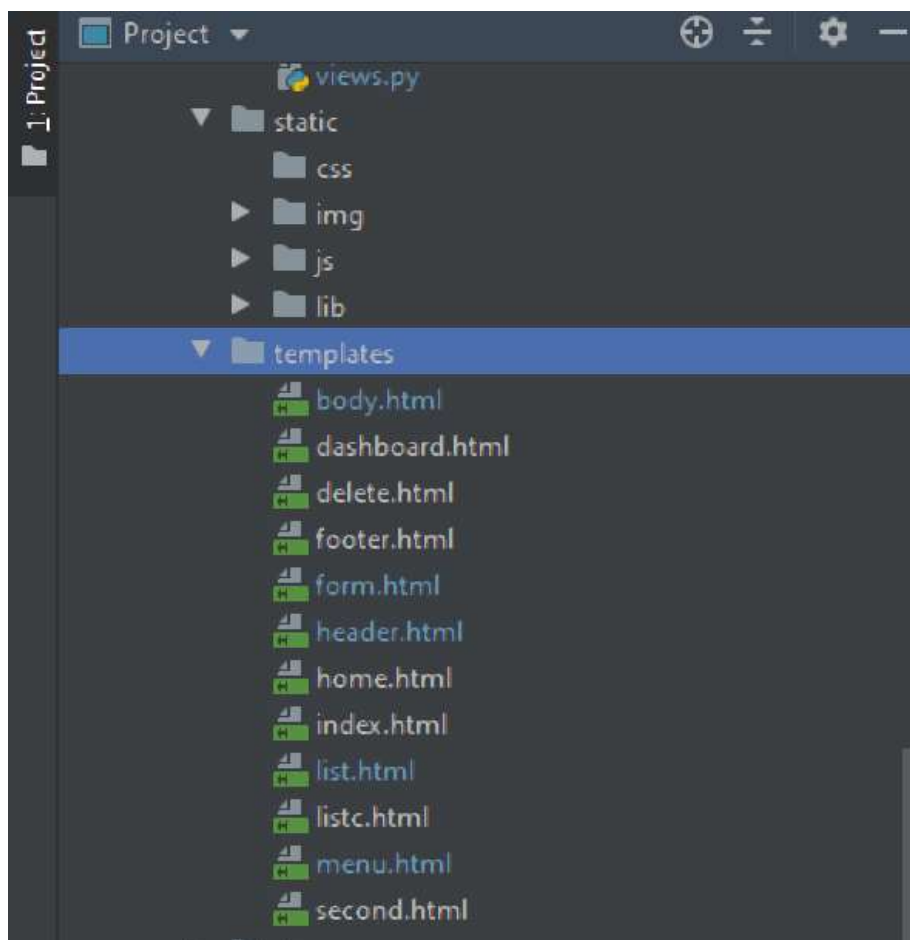


En el directorio login podemos encontrar la plantilla en el directorio templates y la vista en el archivo views.py los cuales permiten la configuración del login de usuario para administrador e invitado en la aplicación.



En la siguiente imagen podemos encontrar desplegado el directorio o carpeta témpate el cual contiene las diferentes plantillas bases que permite tener las diferentes rutas de librerías y JQuery

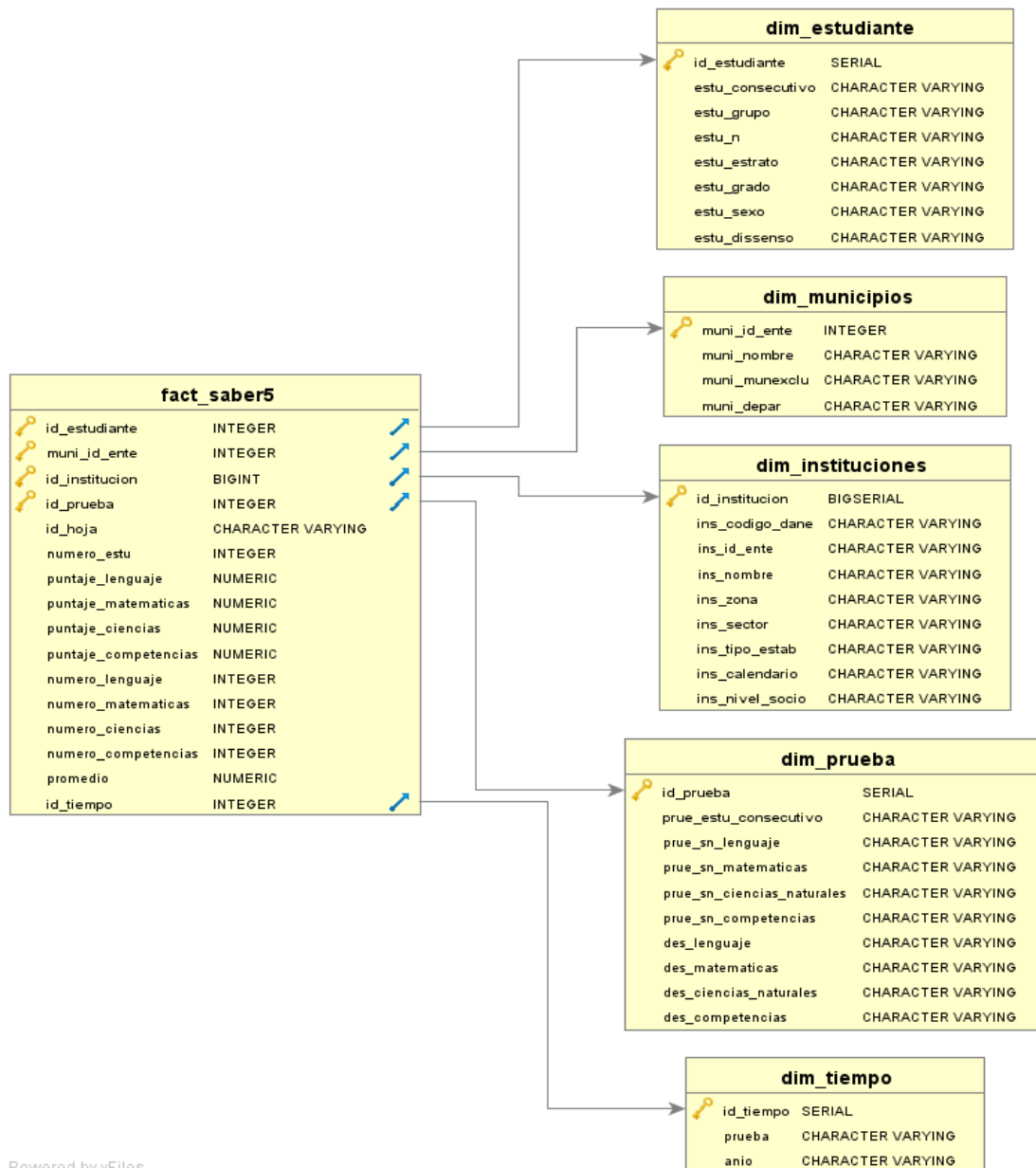
que se utilizan para toda la aplicación con sus respectivas configuraciones, así también un pie de página y un encabezado como herencia para todas las plantillas que se visualicen para el usuario.



MODELADO DE DATOS

Para el diseño de la aplicación se decidió trabajar con diseño multidimensional con un esquema de estrella, este tipo de esquema de base de datos relacional consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

En la siguiente Figura se muestra un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones. Las ramas situadas al final de los enlaces que conectan las tablas indican una relación de muchos a uno entre la tabla de hechos y cada tabla de dimensiones con sus respectivas llaves primarias y llaves foráneas.



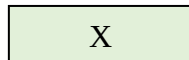
Powered by yFiles

MERCADO DE DATOS

En las siguientes tablas se presentan los atributos de los Valores plausibles, Entidades territoriales certificadas, Establecimientos y Municipios existentes en los repositorios que comprende los años de 2014 a 2017 que tienen datos indispensables para la agregación de la nueva bodega de datos.

En las siguientes tablas se analizan los atributos que están y no están presentes en cada uno de los repositorios.

Está presente



No está presente



Repositorio de datos (Valores plausibles).

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
estu_consecutivo	X	X	X	X	ID público del estudiante
aplicación	X	X	X	X	No identificado
calendario	X	X	X	X	Calendario del establecimiento
departamento	X	X	X	X	Código DANE del departamento al que pertenece el establecimiento educativo
dissenso	X	X	X	X	Marca de discapacidad cognitiva
enteterr	X	X	X	X	Código DANE de la entidad a la que pertenece el establecimiento educativo
establecimiento	X	X	X	X	Código DANE del establecimiento educativo al que pertenece el estudiante
estrato	X	X	X	X	Estrato al que pertenecen el establecimiento dentro del marco muestral
grado	X	X	X	X	Grado del estudiante
grupo	X	X	X	X	Grupo/salón al que pertenecen los estudiantes
id_hoja	X	X	X	X	Código de identificación de hoja de respuestas

jornada	X	X	X	X	Código de la sede jornada al que pertenece el estudiante
leng_copietas	X	X	X	X	Indicador de copia en lenguaje
leng_score1	X	X	X	X	Score1-valor plausible 1 de lenguaje
leng_score2	X	X	X	X	Score2-valor plausible 2 de lenguaje
leng_score3	X	X	X	X	Score3-valor plausible 3 de lenguaje
leng_score4	X	X	X	X	Score4-valor plausible 4 de lenguaje
leng_score5	X	X	X	X	Score5-valor plausible 5 de lenguaje
leng_weight	X	X	X	X	Peso muestral del lenguaje
mate_copietas	X	X	X	X	Indicador de copia en matemáticas
mate_score1	X	X	X	X	Score1-valor plausible 1 de matemáticas
mate_score2	X	X	X	X	Score2-valor plausible 2 de matemáticas
mate_score3	X	X	X	X	Score3-valor plausible 3 de matemáticas
mate_score4	X	X	X	X	Score4-valor plausible 4 de matemáticas
mate_score5	X	X	X	X	Score5-valor plausible 5 de matemáticas
mate_weight	X	X	X	X	Peso muestral de matemáticas
cien-copietas	X				Indicador de copia de ciencias
cien_score1	X				Score1-valor plausible 1 de ciencias
cien_score2	X				Score2-valor plausible 2 de ciencias
cien_score3	X				Score3-valor plausible 3 de ciencias
cien_score4	X				Score4-valor plausible 4 de ciencias
cien_score5	X				Score5-valor plausible 5 de ciencias
cien_weight	X				Peso muestral de ciencias
comp_copietas		X	X		Indicador de copia de competencias
comp_score1		X	X		Score1-valor plausible 1 de competencias
comp_score2		X	X		Score2-valor plausible 2 de competencias
comp_score3		X	X		Score3-valor plausible 3 de competencias
comp_score4		X	X		Score4-valor plausible 4 de competencias
comp_score5		X	X		Score5-valor plausible 5 de competencias
comp_weight		X	X		Peso muestral de competencias
modeloedu	X	X	X	X	No definido

municipio	X	X	X	X	Código DANE al que pertenece el establecimiento educativo
n	X	X	X	X	Matricula del grupo al que pertenecen
nivel	X	X	X	X	Nivel socioeconómico del establecimiento
sector	X	X	X	X	Sector del establecimiento
sexo	X	X	X	X	Sexo del estudiante
zona	X	X	X	X	Zona de la jornada
zonastab	X	X	X	X	Zona del establecimiento

Repositorio de datos (Entidades).

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	X	Identificador del ente territorial
nombre	X	X	X	X	Nombre del ente territorial
munexclu	X	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	X	Tipo de ente territorial 1=ETC, 2=DPTO,3= MPIO

Repositorio de datos (Establecimientos).

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
cod_dane	X	X	X	X	Código DANE del establecimiento educativo
id_municipio	X	X	X	X	Código DANE del municipio al que pertenece el establecimiento educativo
id_ente	X	X	X	X	id de la entidad territorial
nombre	X	X	X	X	Nombre del establecimiento educativo reportado en el DUE
zona	X	X	X	X	Zona donde la mayoría de la población atendida por el establecimiento educativo se ubica 1=urbano 2=rural
sector	X	X	X	X	Naturaleza administrativa de establecimiento educativo según 1=oficial urbano 2=oficial rural 3=no oficial

tipo_estab	X	X	X	X	Tipo de establecimiento 1=Oficial urbano 2=Oficial rural 3=no oficial
calendario	X	X	X	X	Calendario del establecimiento
nivel_socio	X	X	X	X	NSE asignado de acuerdo a la clasificación realizada con puntajes por medios

Repositorio de datos (Municipios).

VARIABLES	2014	2015	2016	2017	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	X	Identificador del ente territorial
nombre	X	X	X	X	Nombre del ente territorial
munexclu	X	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
tipo	X	X	X	X	Tipo del ente territorial posibles valores 3

ATRIBUTOS DE REPOSITORIO DE DATOS

Atributos (dimensión estudiante).

id_estudiante	integer
estu_consecutivo	character varying
estu_grupo	character varying
estu_n	character varying
estu_estrato	character varying
estu_grado	character varying
estu_sexo	character varying
estu_disenso	character varying

Atributos (dimensión instituciones).

id_institucion	integer
----------------	---------

ins_codigo_dane	character varying
ins_id_ente	character varying
ins_nombre	character varying
ins_zona	character varying
ins_sector	character varying
ins_tipo_estab	character varying
ins_calendario	character varying
ins_nivel_socio	character varying

Atributos (dimensión municipios).

muni_id_ente	Integer
muni_nombre	character varying
muni_munexclu	character varying
muni_depar	character varying

Atributos (dimensión prueba).

id_prueba	Integer
prue_estu_consecutivo	character varying
prue_sn_lenguaje	character varying
prue_sn_matematicas	character varying
prue_sn_ciencias_naturales	character varying
prue_sn_competencias	character varying
des_lenguaje	character varying
des_matematicas	character varying
des_ciencias_naturales	character varying
des_competencias	character varying

Atributos (dimensión tiempo).

id_tiempo	Integer
prueba	character varying
anio	character varying

Atributos (dimensión fact saber 5).

id_estudiante	Integer
muni_id_ente	Integer
id_institucion	Integer
id_prueba	Integer
id_hoja	character varying
numero_estu	Integer
puntaje_lenguaje	Numeric
puntaje_matematicas	Numeric
puntaje_ciencias	Numeric
puntaje_competencias	Numeric
numero_lenguaje	Integer
numero_matematicas	Integer
numero_ciencias	Integer
numero_competencias	Integer
promedio	Numeric
Id_tiempo	Integer

ANÁLISIS Y CALIDAD DE DATOS

Análisis calidad de datos de la bodega para identificar el número de registros nulos, no nulos y distintos cuenta con un total de **3'731.692** registros, se tuvo en cuenta las 5 dimensiones con el propósito de realizar un análisis de la calidad de los datos e identificar que los datos contenidos sean fiables

Análisis de calidad de datos (dimensión prueba).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_prueba	13875	0	13875	"40" "41" "42" "43" "44"
2	prue_estu_consecutivo	13875	0	13875	"V201450046302" "V201450046303" "V201450046304" "V201450046306" "V201450046309"
3	prue_sn_lenguaje	9056	4819	2	"CON_COPIA" "NO_COPIA"
4	prue_sn_matematicas	9170	4705	2	"CON_COPIA" "NO_COPIA"
5	prue_sn_ciencias	6051	7824	2	"CON_COPIA" "NO_COPIA"
6	prue_sn_competencias	3011	10864	2	"CON_COPIA" "NO_COPIA"
7	des_lenguaje	9056	4819	0	'INSUFICIENTE'

8	des_matematicas	9170	4705	0	'MINIMO'
					'SATISFACTORIO'
					'AVANZADO'
9	des_ciencias_naturales	6051	7824	0	'INSUFICIENTE'
					'MINIMO'
					'SATISFACTORIO'
10	des_competencias	3011	10864	0	'AVANZADO'
					'INSUFICIENTE'
					'MINIMO'
					'SATISFACTORIO'
					'AVANZADO'

Análisis de calidad de datos (dimensión municipio).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	muni_id_ente	13	0	13	"52022"
					"52210"
					"522152"
					"52224"
					"52227"
2	muni_nombre	13	0	13	"ALDANA"
					"CONTADERO"
					"CORDOBA"
					"CUASPUD"
					"CUMBAL"

3	muni_munexclu	13	0	2	"ETC" "SIN_DATO"
4	muni_depar	13	0	1	"NARINO"

Análisis de calidad de datos (dimensión estudiante).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_estudiante	13875	0	13875	"50" "51" "52" "53" "54"
2	estu_consecutivo	13875	0	13875	"V201450046302" "V201450046303" "V201450046304" "V201450046305" "V201450046306"
3	estu_grupo	13875	0	3	"01" "02" "99"
4	estu_n	13875	0	112	"1" "10" "100" "101" "103"
5	estu_estrato	13875	0	12	99E1209 99E1219 99E9999 99O1109

					99O1209
6	estu_grado	13875	0	1	"5"
7	estu_sexo	13875	0	3	FEMENINO MASCULINO NO_ESPECIFICA
8	estu_disenso	13875	0	2	"N" "S"

Análisis de calidad de datos (dimensión institución).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_institucion	358	0	358	"39" "40" "41" "42" "43"
2	ins_codigo_dane	358	0	358	"152022000084" "152210000261" "152215000138" "152224000019" "152227000028"
3	ins_id_ente	358	0	2	48 89
4	ins_nombre	358	0	336	-CENTRO EDUCATIVO SAN MARTIN CENTRO EDUCATIVO CASAFRIA

					CENTRO EDUCATIVO ARVELA
					CENTRO EDUCATIVO CUAS
					CENTRO EDUCATIVO GUAN
5	ins_zona	358	0	2	RURAL
					URBANO
6	ins_sector	358	0	2	NO_OFICIAL
					OFICIAL
7	ins_tipo_estab	358	0	3	NO_OFICIAL
					OFICIAL_RURAL
					OFICIAL_URBANO
8	ins_calendario	358	0	1	A
9	ins-nivel_socio	358	0	5	1
					2
					3
					4
					5

Análisis de calidad de datos (dimensión tiempo).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	Id_tiempo	13875	0	0	"52022"
					"52210"
					"522152"
					"52224"
					"52227"

2	prueba	13875	0	0	"V201450046302"
					"V201450046303"
					"V201450046304"
					"V201450046305"
					"V201450046306"
3	anio	13875	0	0	"2014"
					"2015"
					"2016"
					"2017"

Análisis de calidad de datos (dimensión fact_saber5).

No.	Atributos	No nulos	Nulos	Valores distintos	Valores distintos
1	id_estudiante	13875	0	13875	"35"
					"36"
					"37"
					"38"
					"39"
2	muni_id_ente	13875	0	13	"52022"
					"52210"
					"52215"
					"52224"
					"52227"
3	id_institucion	13875	0	324	"150"
					"151"
					"163"
					"167"
					"168"
4	id_prueba	13875	0	13875	"1"

					"55"
					"60"
					"70"
					"100"
5	id_hoja	13875	0	13875	"V201450046302"
6	numero_estu	13875	0	1	1
7	puntaje_lenguaje	9056	4819	151	"1"
					"1,2"
					"1,25"
					"1,33"
					"1,38"
8	puntaje_matematicas	9170	4705	154	"146"
					"152"
					"152,67"
					"153"
					"154"
9	puntaje_ciencias	6051	7824	102	"1,45"
					"146"
					"147"
					"148"
					"148,78"
10	puntaje_competencias	3011	10864	72	"150,104"
					"150,943"
					"152"
					"163"
11	numero_lenguaje	13875	0	2	"0"
					"1"
12	numero_matematicas	13875	0	2	"0"
					"1"
13	numero_ciencias	13875	0	2	"0"

					"1"
14	numero_competencias	13875	0	2	"0"
					"1"
15	promedio	13875	0	0	""1,45"
					"146"
					"147"
					"148"
16	Id_tiempo	13875	0	0	"1"
					"55"
					"60"
					"70"
					"100"

SABER EXPO BI 5: UNA HERRAMIENTA DE INTELIGENCIA DE
NEGOCIOS WEB PARA LA GESTIÓN DE INDICADORES DE LAS
PRUEBAS SABER 5 EN LAS INSTITUCIONES EDUCATIVAS DE LA
SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO

Anexo B

MANUAL DE USUARIO



INTRODUCCIÓN

En este anexo se muestran conceptos y guías de utilización necesarias referente al funcionamiento del aplicativo, en el que se encuentran funcionamiento de interfaz, descripción de funcionalidades de menús igualmente la interacción de botones y significancia de gráficas e interpretación de datos.

FUNCIONAMIENTO

La siguiente es la documentación que permite guiar a los usuarios finales en el uso e interacción con el aplicativo desarrollado; para esto la aplicación tiene una vista por roles los cuales permiten acceder a funcionalidades particulares, los roles con los que cuenta el sistema son rol de invitado y rol de administrador. EN el siguiente documento se explica detenidamente el funcionamiento de cada módulo.

Entrar como invitado

El modo entrar como invitado sirve para que no todos los usuarios puedan acceder a funciones específicas de la aplicación, para esto se implementó un botón de entrar como invitado en la pantalla de inicio de la herramienta como se puede observar en la imagen el usuario tendrá que elegir la opción sesión invitado para poder acceder al aplicativo.

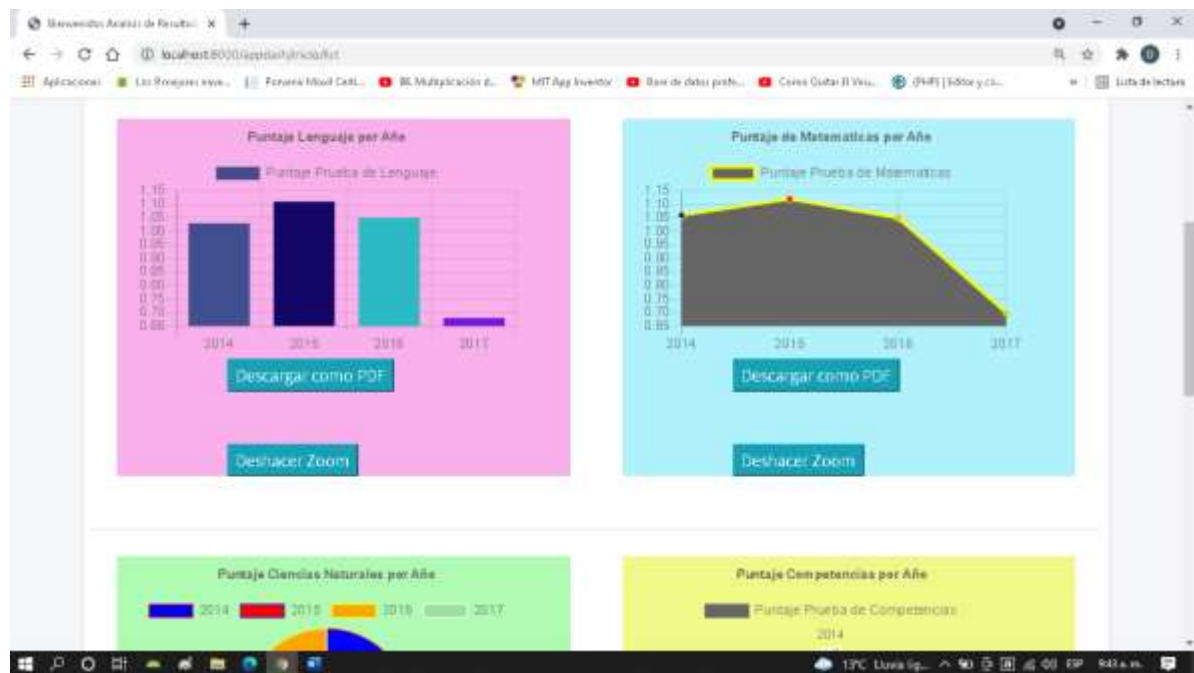
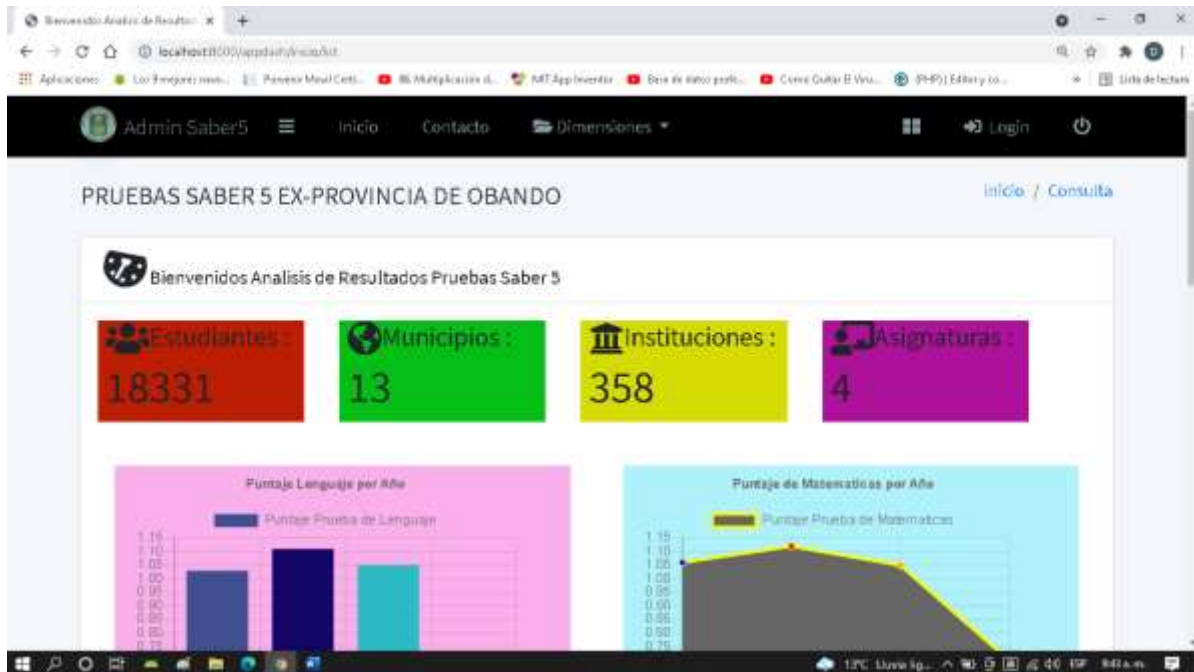


Una vez el invitado se autentica y puede visualizar la siguiente ventana, en donde está presente la vista general de inicio y el menú proporcionado para el invitado.



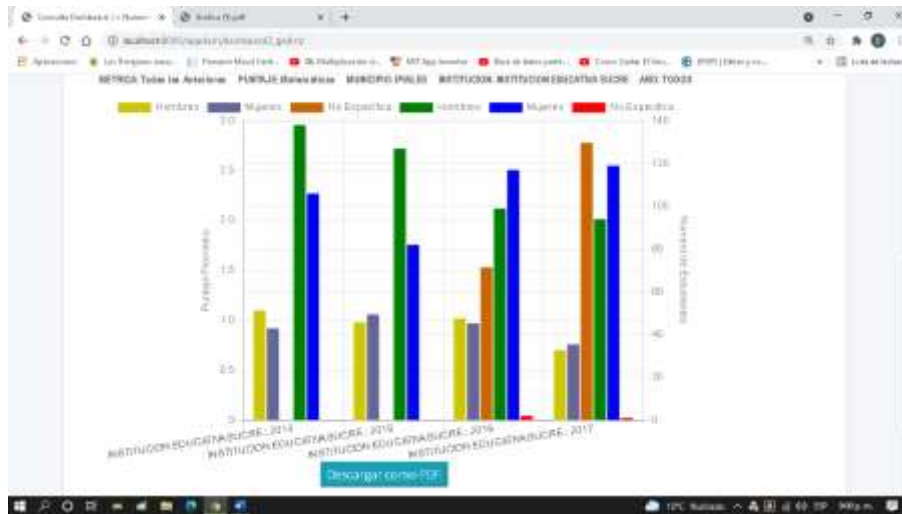
Visualización general

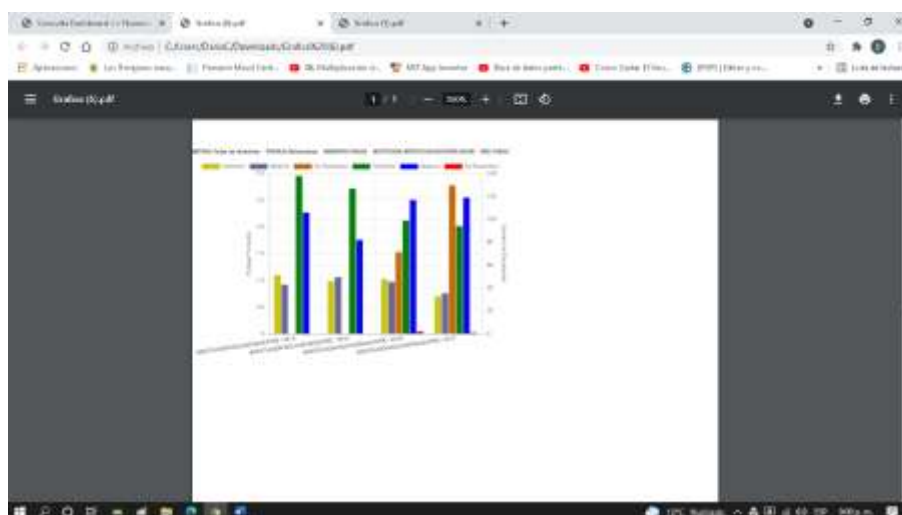
Esta visualización se encuentra al entrar a la aplicación, independientemente si el usuario es administrador o invitado, en este Dashboard se muestra información global como número de estudiantes, numero de instituciones, y datos analizados hasta la fecha y un puntaje general en función de los estudiantes de cada municipio de la ex provincia de Obando.



Crear grafica general

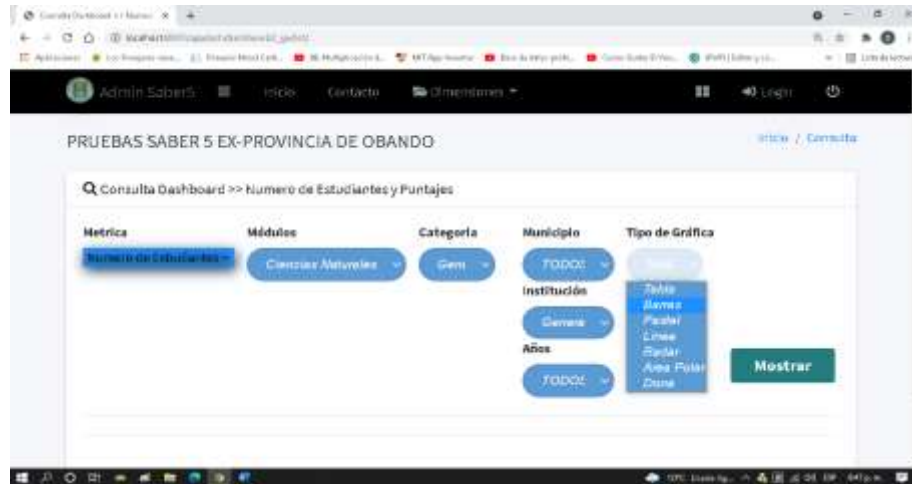
Para crear esta grafica se debe elegir dashboard en el menú de inicio y este desplegará opciones en las cuales se podrá escoger entre diferentes tipos de categorías, métricas y filtros que se desee visualizar para ver en el dashboard, a esto se le añade una opción de tipo de grafica con la cual los resultados se presentan a conveniencia del usuario, si este desea existe la opción de guardar gráfica, con la cual se genera un archivo.pdf de la misma.





Crear grafica por número de estudiante

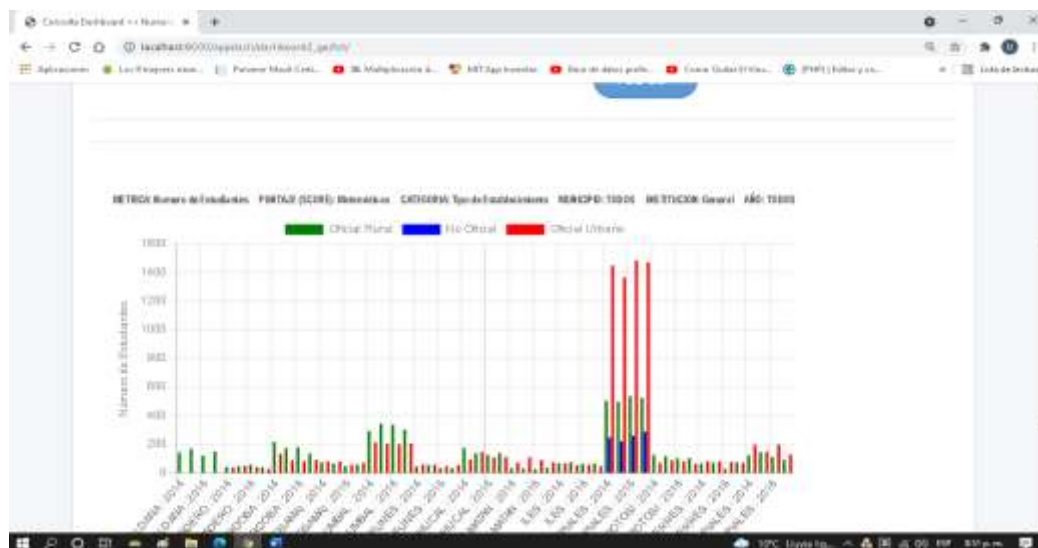
Se debe elegir en el menú dashboard la opción número de estudiante en el cual se presentan opciones de filtros como son; modulo, categoría el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf



METRICA: METRICA:Numero de Estudiantes PUNTAJE (SCORE):Matemáticas CATEGORIA:Tipo de Establecimiento

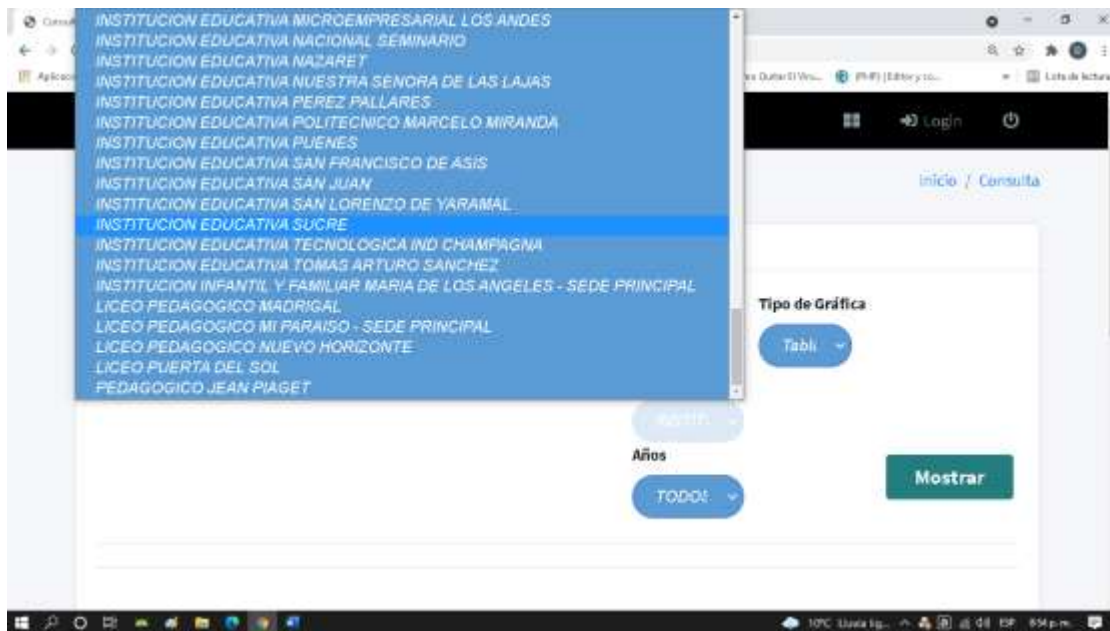
MUNICIPIO: TODOS INSTITUCIÓN: General AÑO: TODOS

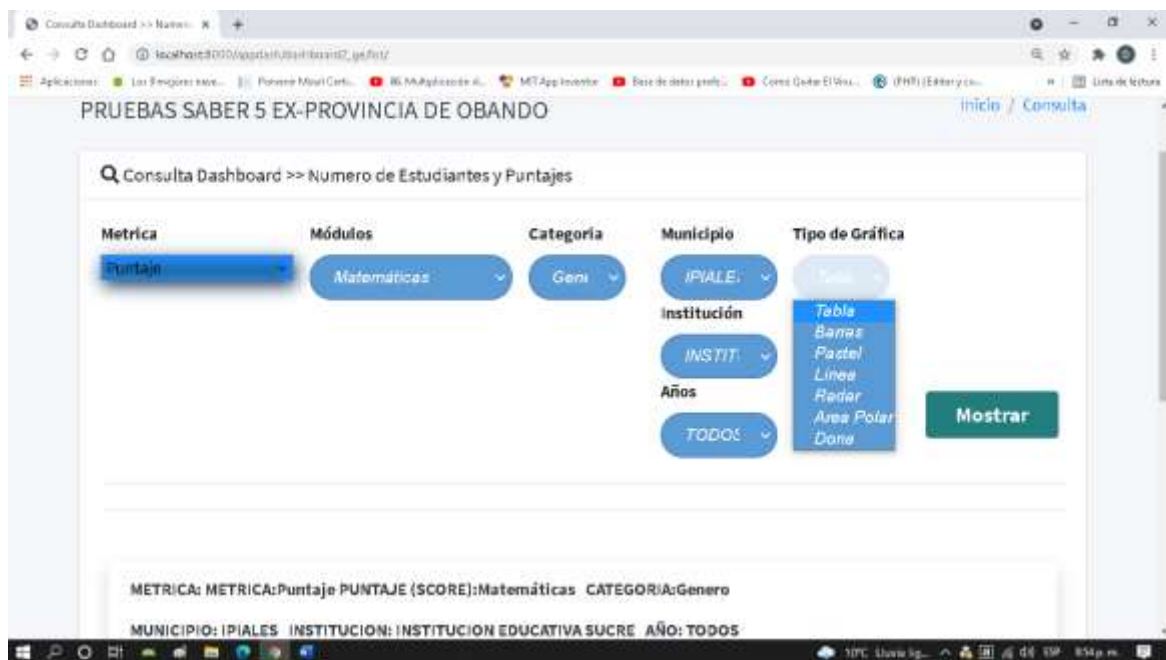
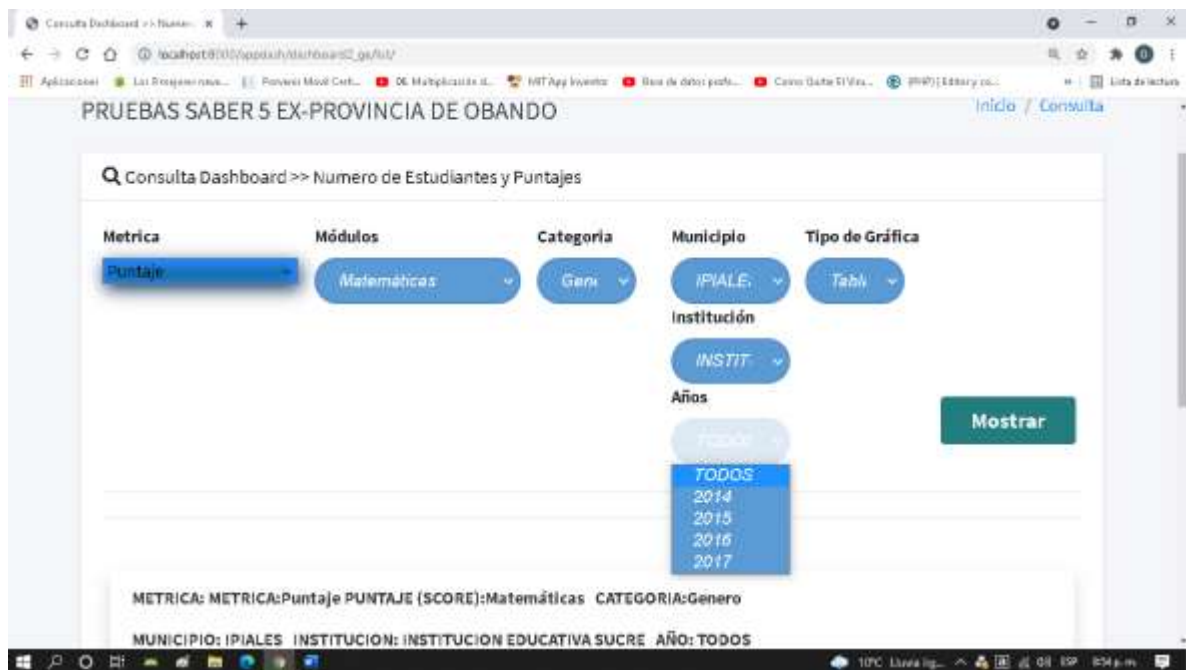
	ALDANA : 2014	ALDANA : 2015	ALDANA : 2016	ALDANA : 2017	CONTADERO : 2014	CONTADERO : 2015	CONTADERO : 2016	CONTADERO : 2017
Oficial Rural	142	169	121	150	42	47	57	39
No Oficial	0	0	0	0	0	0	0	0
Oficial Urbano	0	0	0	0	38	50	41	26

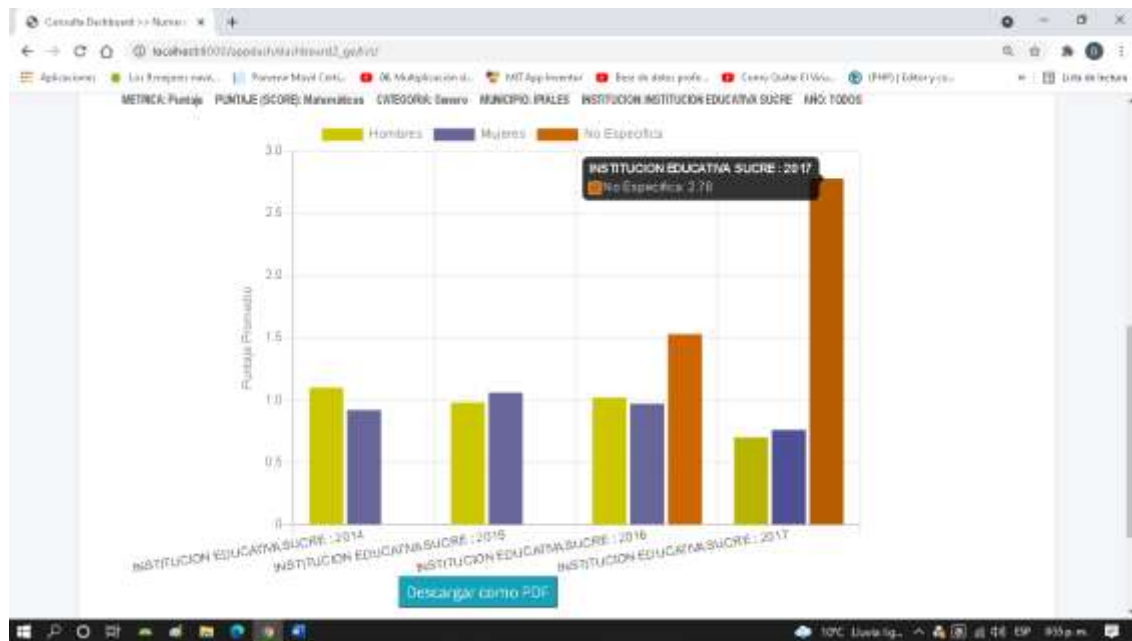


Crear grafica por puntaje

Se debe elegir en el menú dashboard la opción puntaje en el cual se presentan opciones de filtros como son; el área de la que se desea el puntaje, el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf

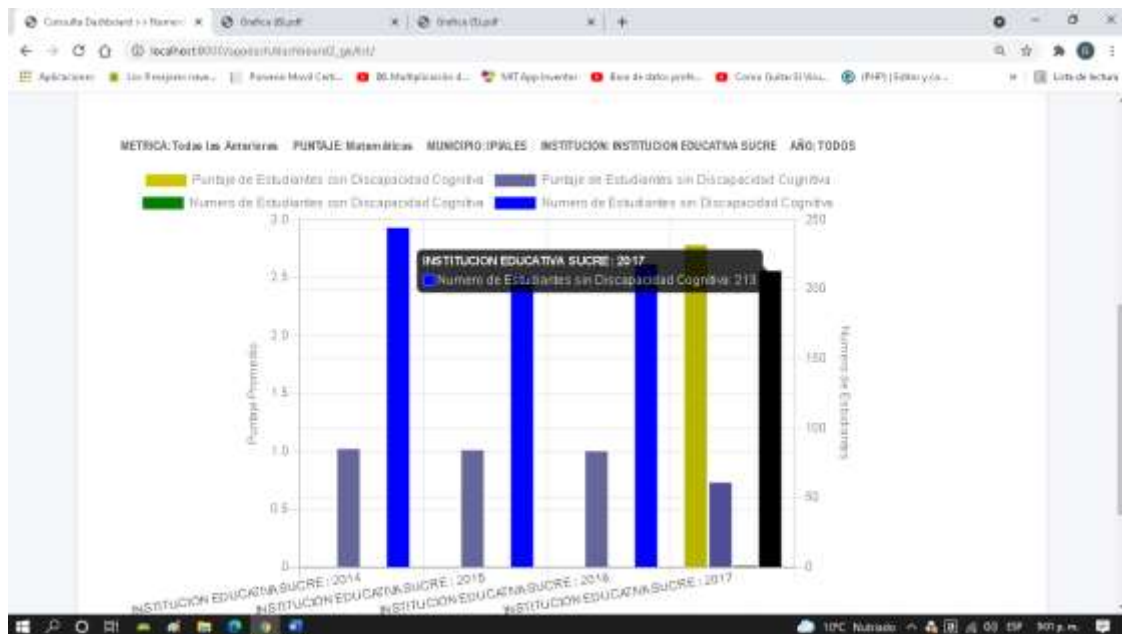
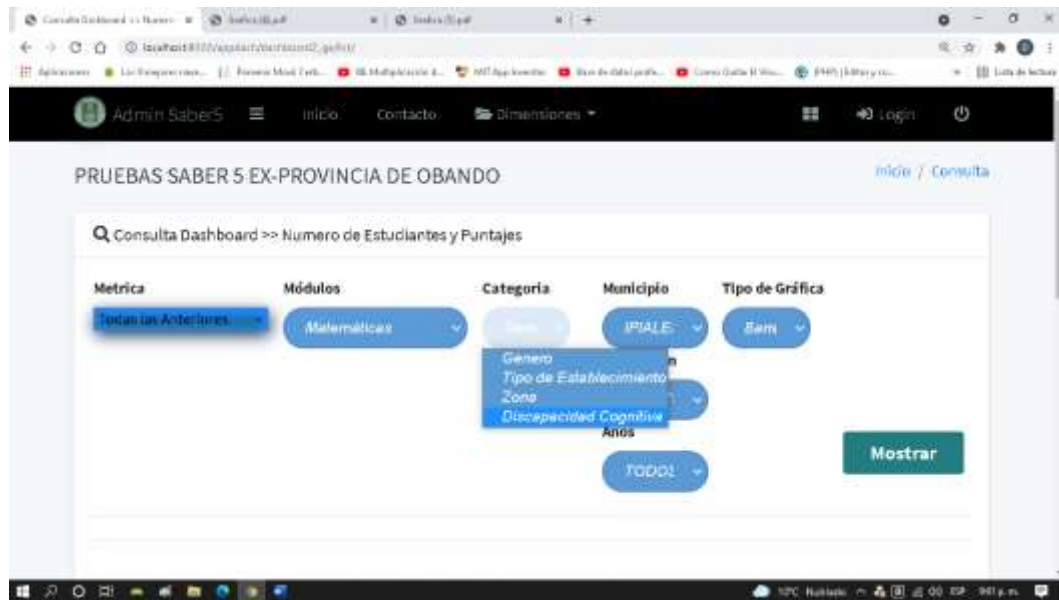






Crear grafica por categoría

Se debe elegir en el menú dashboard la opción puntaje en el cual se presentan opciones de filtros como son; categoría, el año y el municipio, una vez se eligen los filtros deseados, se continúa con el tipo de gráfica y esta se podrá descargar en un archivo con formato .pdf



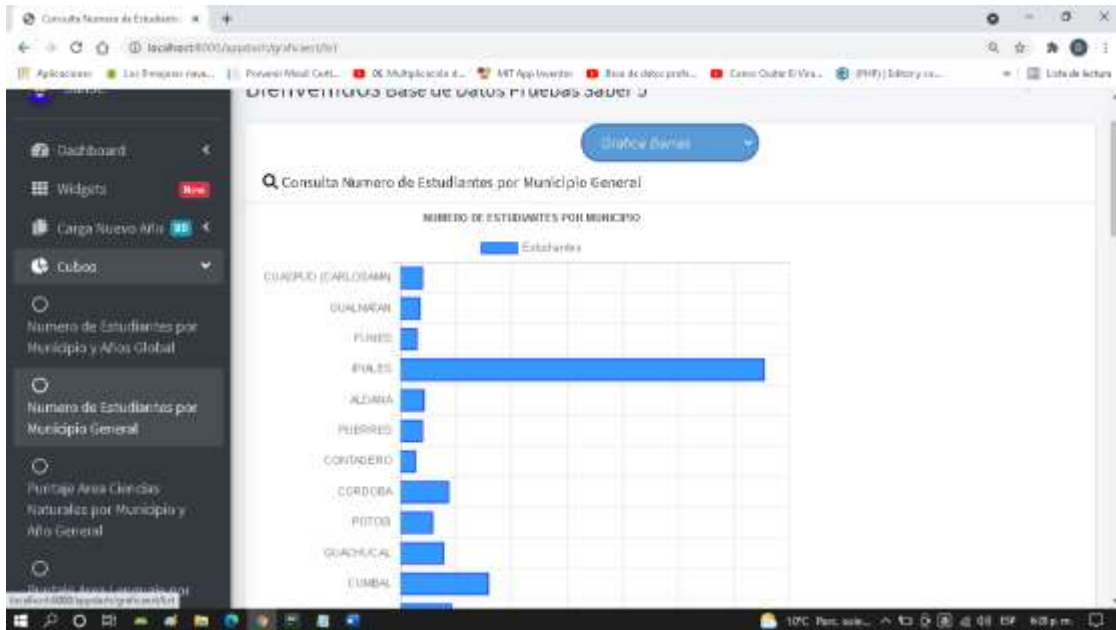
Ubicados en el menú en la sección cubos, este permite visualizar los cubos prediseñados con consultas comunes y de alta relevancia.

Cubo que brinda información sobre la población de estudiantes que presentaron las pruebas en determinado año.



Número de estudiantes por municipio general

Cubo que brinda información sobre los estudiantes que presentaron la prueba por municipio.

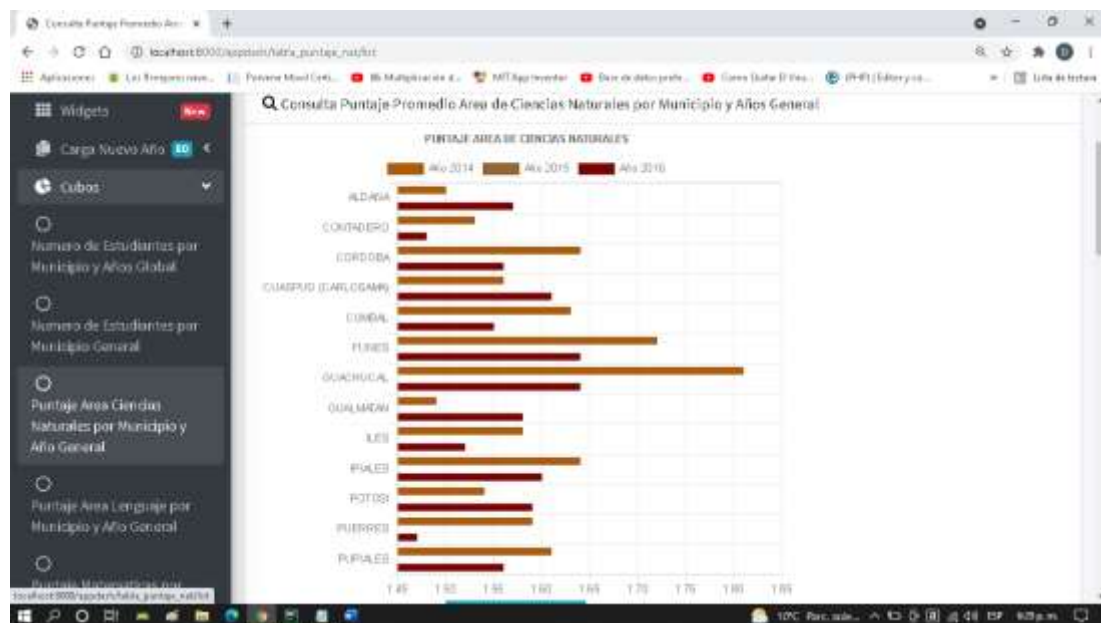


A table view of the same data, titled 'Bienvenidos Base de Datos Pruebas Saber 5'. It includes a search bar and a table with two columns: 'Nombre Municipio' and 'Numero de Estudiantes'.

Nombre Municipio	Numero de Estudiantes
ALDANA	432
CONTADERO	275
CORDOSA	873
CUASPUD (CARLOSAMA)	405
CUMBAL	1586
FUNES	300

Puntaje área ciencias naturales por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de ciencias naturales de cada municipio tomando todos los años en que se realizaron las pruebas.



Bienvenidos Base de Datos Pruebas Saber 5

Admin Sabaró

Inicio / Consulta

Tabla

Consulta Puntaje Promedio Área de Ciencias Naturales por Municipio y Años General

Nombre Municipio	Año de Presentación	Puntaje Promedio Ciencias Naturales
ALDANA	2014	1,50
CONTADERO	2014	1,53
CORDOBA	2014	1,58
CUASPUD (CARLOSAMA)	2014	1,56
CUMBAI	2014	1,53
FUNES	2014	1,72

Puntaje área lenguaje por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de lenguaje de cada municipio tomando todos los años en que se realizaron las pruebas.



Bienvenidos Base de Datos Pruebas Saber 5

Admin: Sabero

Inicio / Consulta

Consulta Puntaje Promedio Area de Lenguaje por Municipio y Años General

Nombre Municipio	Año de Presentación	Puntaje Promedio Lenguaje
ALDANA	2014	1,53
CONTADERO	2014	1,55
CORDOBA	2014	1,64
CUASPUD (CARLOSAMA)	2014	1,50
CUMBAL	2014	1,63
FUNES	2014	1,77

Puntaje área de matemáticas por municipio

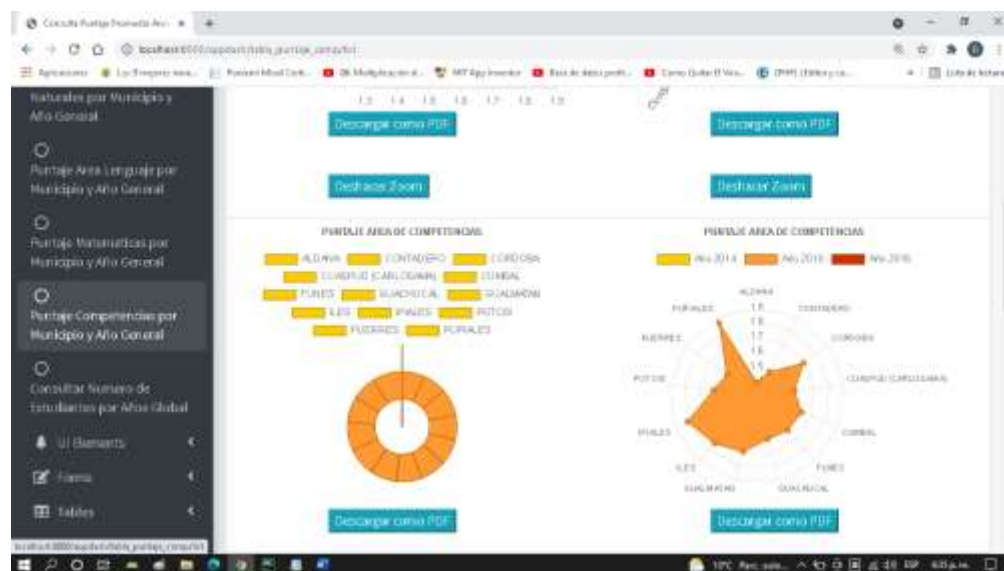
Cubo que brinda información sobre el puntaje obtenido en el área de matemáticas de cada municipio tomando todos los años en que se realizaron las pruebas.



Nombre Municipio	Año de Presentación	Puntaje Promedio Matemáticas
ALDANA	2014	1,56
CONTADERO	2014	1,54
CÓRDOBA	2014	1,54
CUASPUJ (CARLOSAMA)	2014	1,54
CUMBAL	2014	1,56
FUNES	2014	1,78

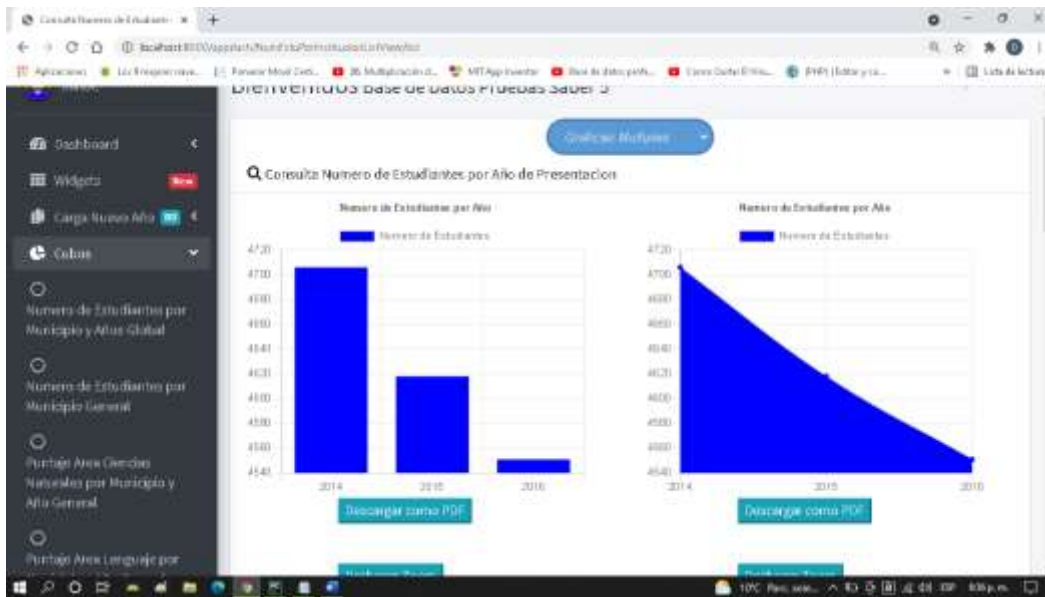
Puntaje área competencias por municipio

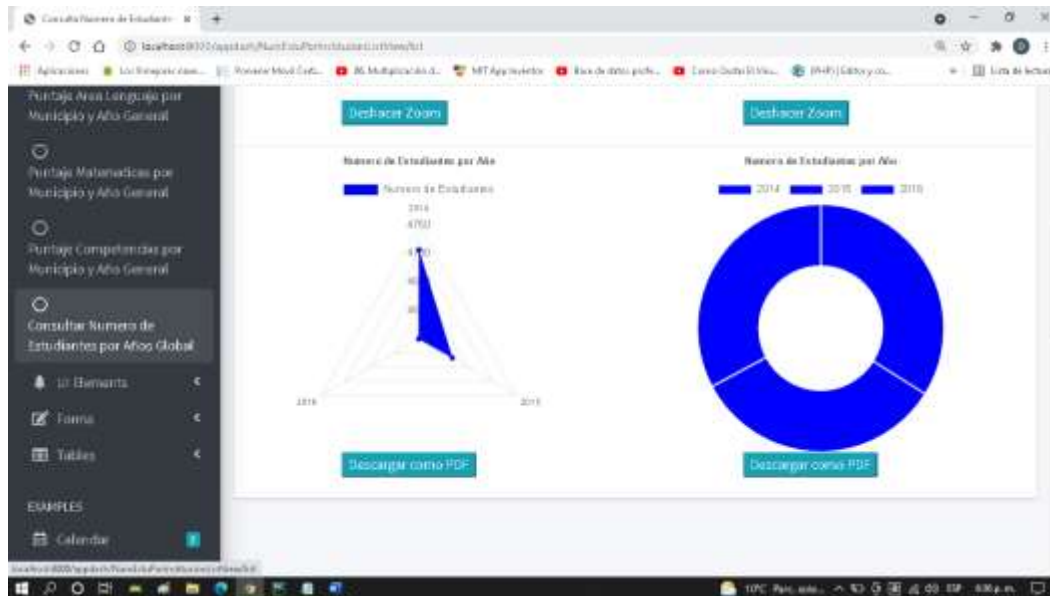
Cubo que brinda información sobre el puntaje obtenido en el área de competencias cada municipio tomando todos los años en que se realizaron las pruebas.



Número de estudiantes por año de presentación

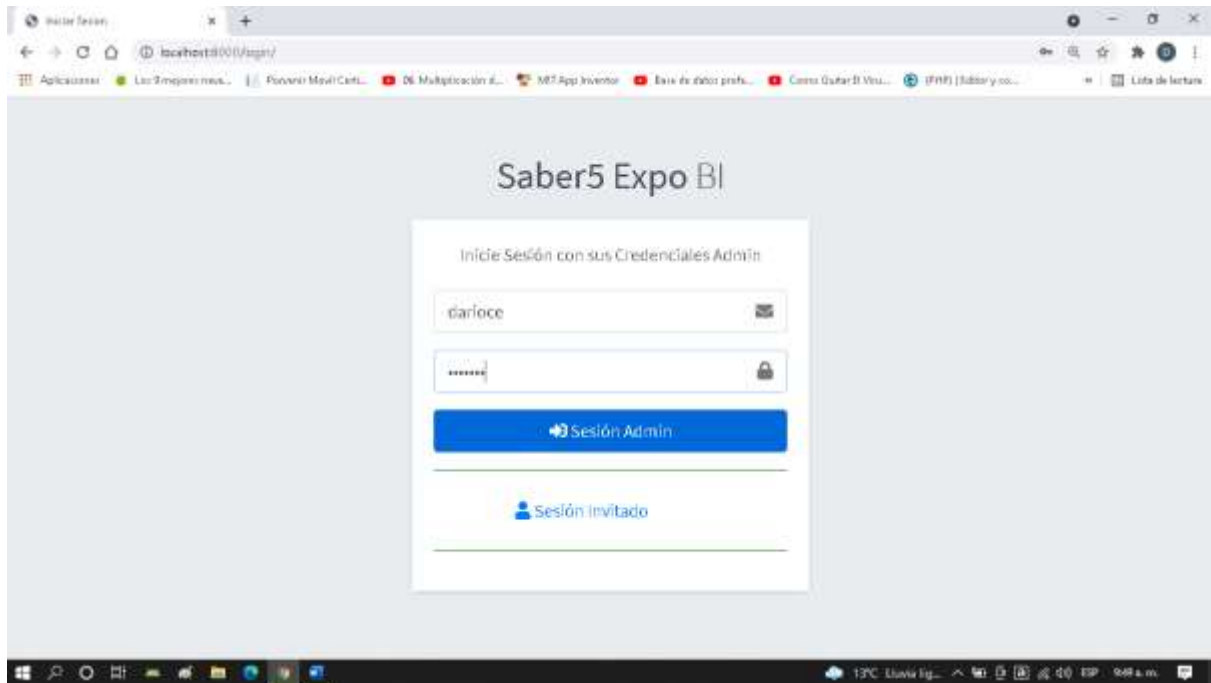
Cubo que brinda información sobre el número de estudiantes que presentaron la prueba saber 5 en función al año de presentación.



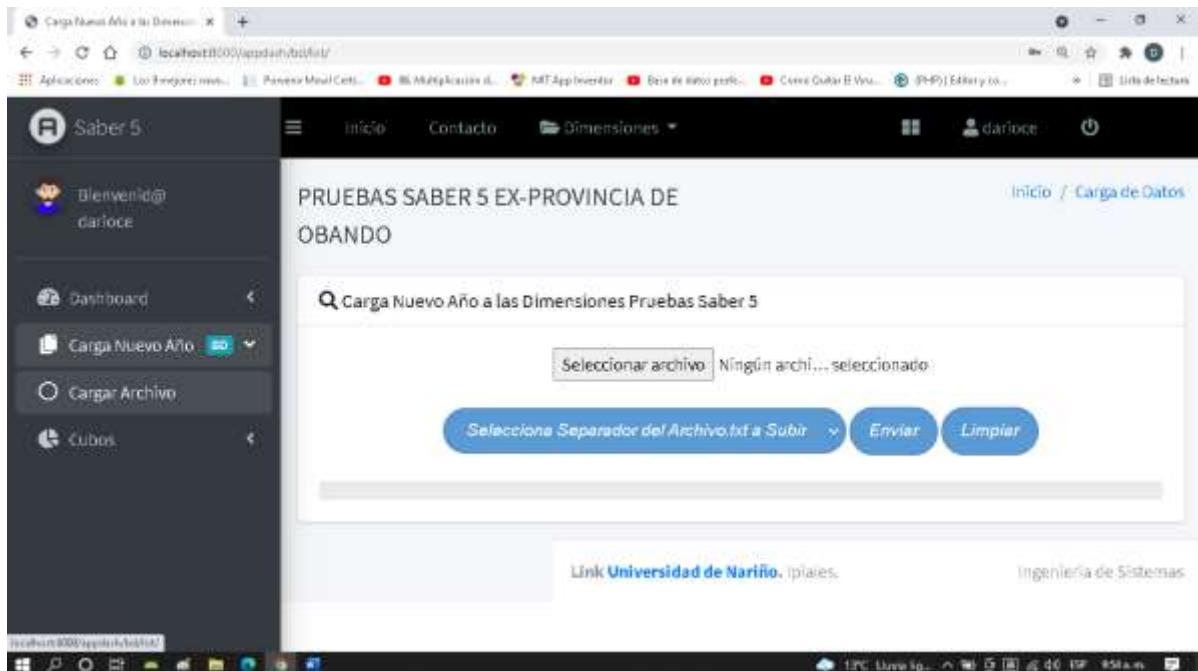


Entrar como administrador

Este caso se permite a los usuarios validar su identidad ante el sistema, muestra en pantalla un cuadro de dialogo donde se digita el login de usuario y contraseña para poder ingresar a la aplicación, para esto los campos cuentan con una validación con la cual si el usuario tiene un registro previo puede acceder a la siguiente pantalla de la herramienta, si el usuario no tiene este registro o los campos son llenados con información incorrecta, un mensaje de error de autenticación se desplegará.

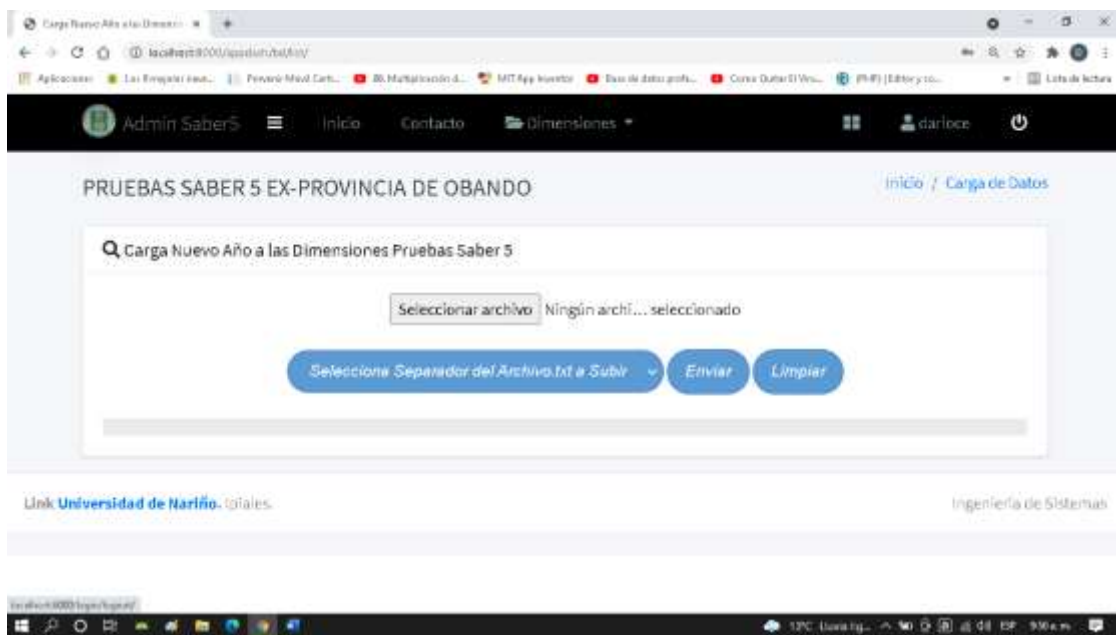


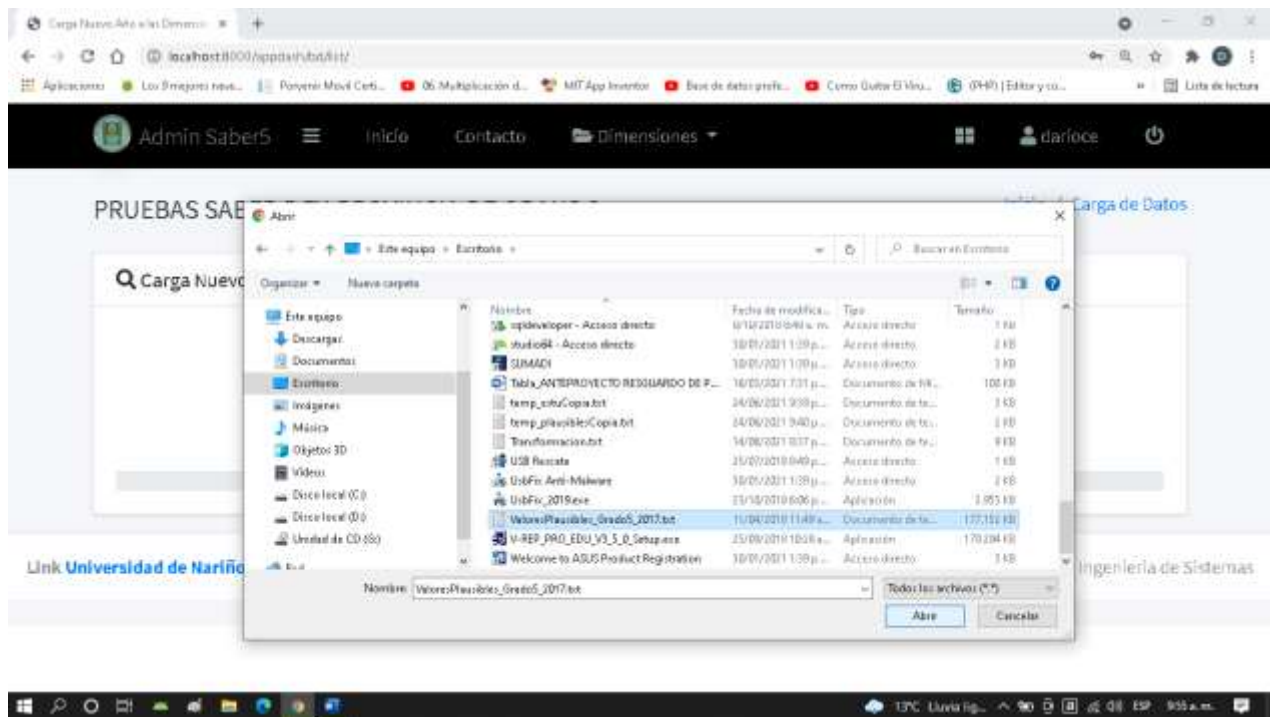
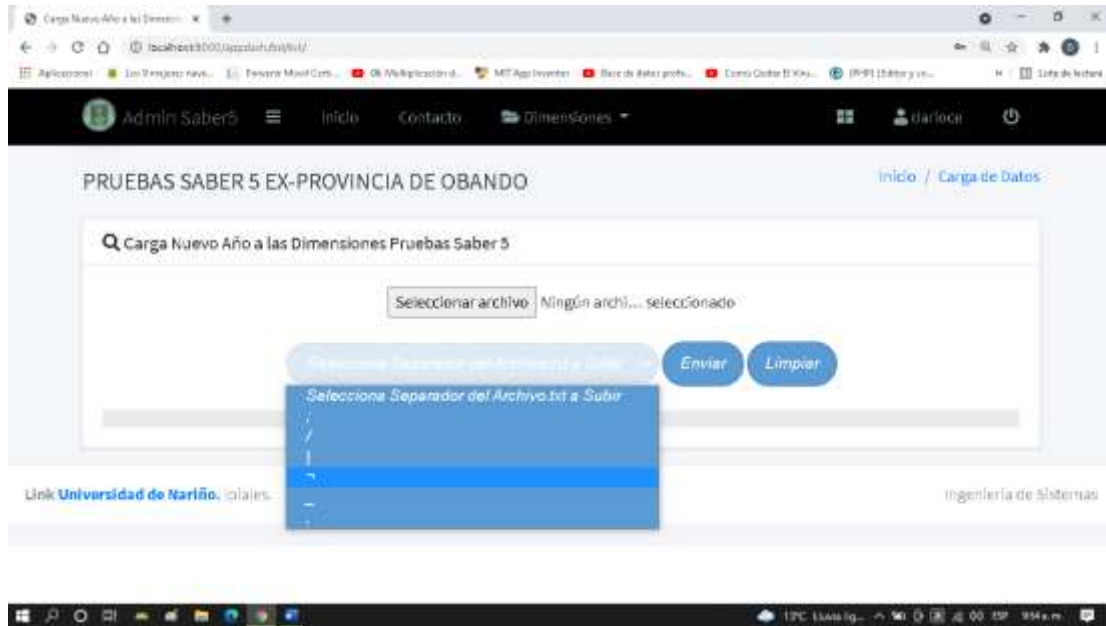
Al iniciar sesión como administrador se añade al menú principal una opción en donde el administrador es capaz de cargar nueva información al mercado de datos ya establecido.

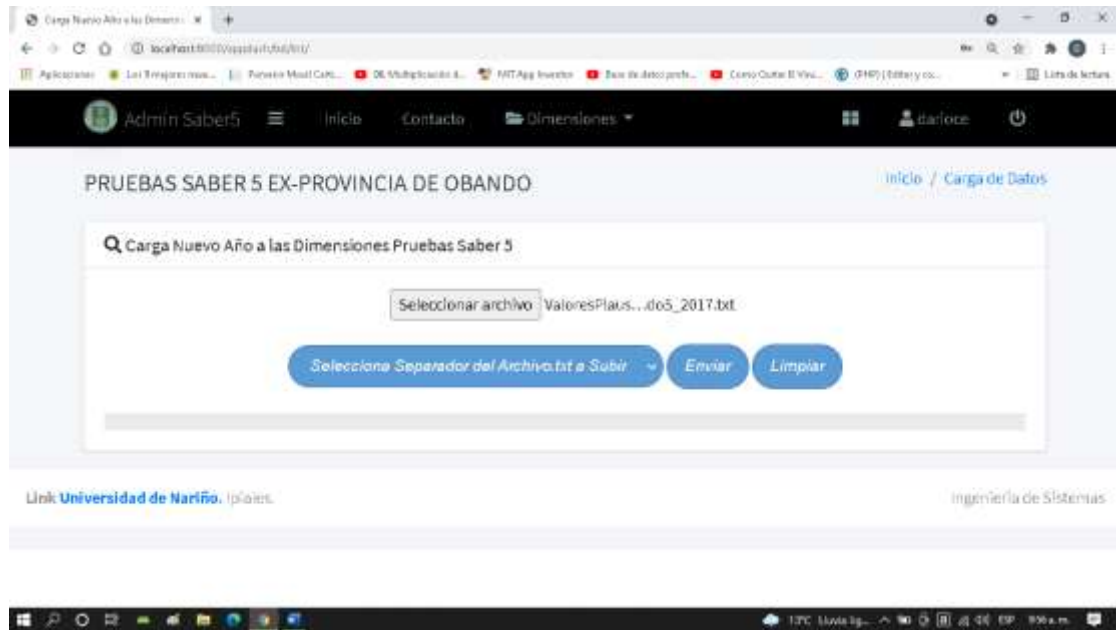


Nueva carga de datos

Para los usuarios administradores se les presenta la opción de cargar nueva base de datos, esta se encuentra en el menú subir nueva base de datos, en este proceso se toma la tabla valores plausibles del repositorio brindado por la página oficial del ICFES el cual es un archivo .txt; cuenta con la opción de elegir agregar nueva base de datos la cual permite buscar el archivo proporcionado por el ICFES una vez sea descargado, una vez marcado se procede a elegir el delimitador con el que se separa los atributos de la nueva base de datos y se procede a subir el nuevo archivo, cabe resaltar que al momento de subir este archivo la aplicación estandarizará, limpiará, eliminará y transformará el archivo .txt que se sube para añadirlo como un nuevo año a la bodega de datos de la aplicación.

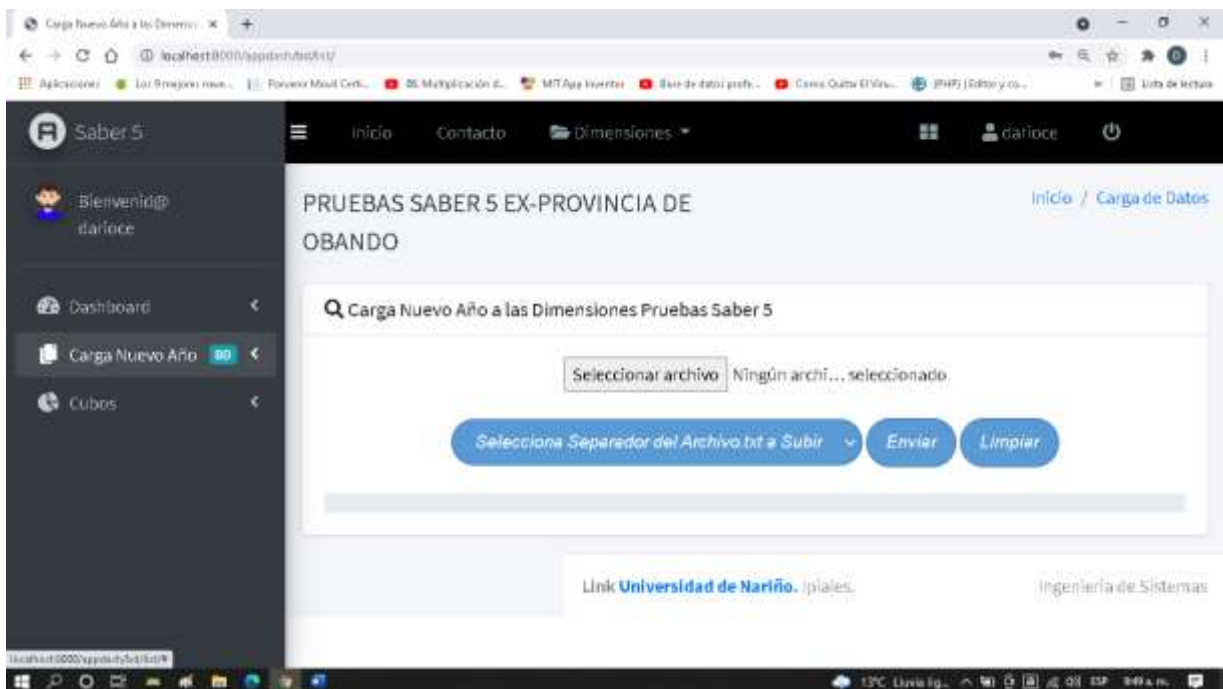






Cerrar sesión

Para realizar esta prueba, antes se debe ser administrador registrado, a lado superior derecho de la aplicación se muestra el rol que ocupa en la en la herramienta, sea administrador o invitado, si se elige esta opción se despliega la opción de cerrar sesión en donde se pregunta al usuario si está seguro de querer salir de la aplicación; en caso de que si, se llevara de regreso a la pantalla de inicio.

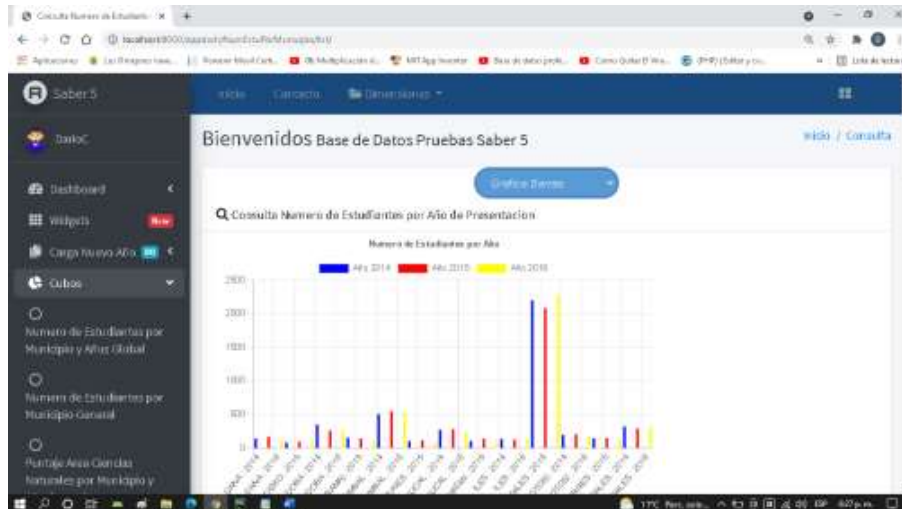


Cubos de datos

a) Número de estudiantes por municipio y por año de presentación

Cubo que brinda información sobre la población de estudiantes que presentaron las pruebas en determinado año, para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, COUNT(id_estudiante) as suma FROM dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY 1,2 ORDER BY 1,2
```



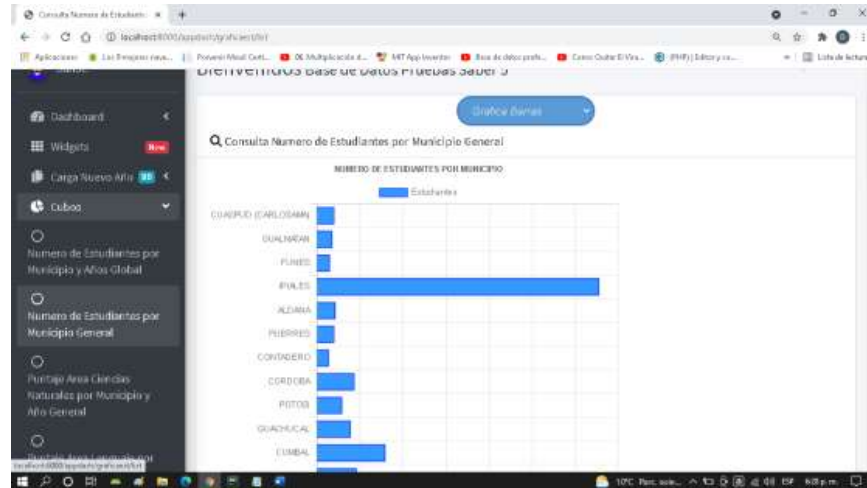
Consulta Número de Estudiantes por Año de Presentación

Municipio	Año	Número de Estudiantes
ALVARA	2014	140
ALVARA	2015	169
ALVARA	2016	321
CONTADOR	2014	80
CONTADOR	2015	97
CONTADOR	2016	98
CONTADOR	2014	80
CONTADOR	2015	97
CONTADOR	2016	98
CONTADOR	2014	80
CONTADOR	2015	97
CONTADOR	2016	98

b) Número de estudiantes por municipio general

Cubo que brinda información sobre los estudiantes que presentaron la prueba por municipio, para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre,COUNT(id_estudiante) as suma FROM dim_municipios JOIN
fact_saber5 USING(muni_id_ente) JOIN dim_estudiante USING(id_estudiante) GROUP BY 1
ORDER BY 1
```



c) Puntaje área ciencias naturales por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de ciencias naturales de cada municipio tomando todos los años en que se realizaron las pruebas, para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_ciencias),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```


Cubo que brinda información sobre el puntaje obtenido en el área de matemáticas de cada municipio tomando todos los años en que se realizaron las pruebas, para esto se utilizó la siguiente sentencia SQL:

```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_matematicas),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```



f) Puntaje área competencias por municipio

Cubo que brinda información sobre el puntaje obtenido en el área de competencias cada municipio tomando todos los años en que se realizaron las pruebas, para esto se utilizó la siguiente sentencia SQL:

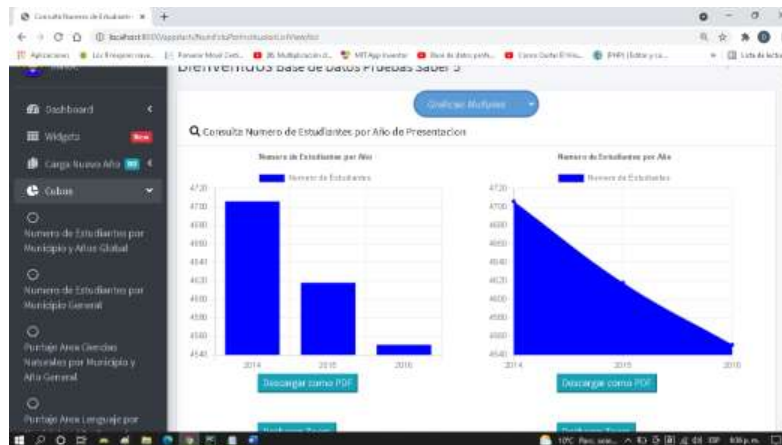
```
SELECT muni_nombre, anio, ROUND(AVG(puntaje_competencias),2) AS PROMEDIO FROM
dim_municipios JOIN fact_saber5 USING(muni_id_ente) GROUP BY anio,muni_nombre
ORDER BY anio,muni_nombre
```




g) Número de estudiantes por año de presentación

Cubo que brinda información sobre el número de estudiantes que presentaron la prueba saber 5 en función al año de presentación, se utilizó la siguiente sentencia SQL:

```
SELECT anio, COUNT(id_estudiante) as suma FROM fact_saber5 GROUP BY 1,1 ORDER BY 1,1
```



Gráficas estadísticas adicionales

Existen forma de presentación estadística más compleja, destinada fundamentalmente a resaltar aspectos de la distribución, observando patrones en una o más series de datos cuantitativos a la vez, para esta representación se eligieron los tipos de gráficos boxplot y biplot, por la facilidad con la cual se pueden observar datos de distintas categorías al tiempo.

a) Boxplot

Los diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría. Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente [41].

Para ingresar se debe elegir la opción estadística en el menú de opciones y marcar la opción boxplot, posteriormente elegir el año que se desea mostrar cómo se observa en [Fig.65].



Podemos observar la implementación de boxplot a los puntajes de las pruebas saber 5 categorizadas por año.



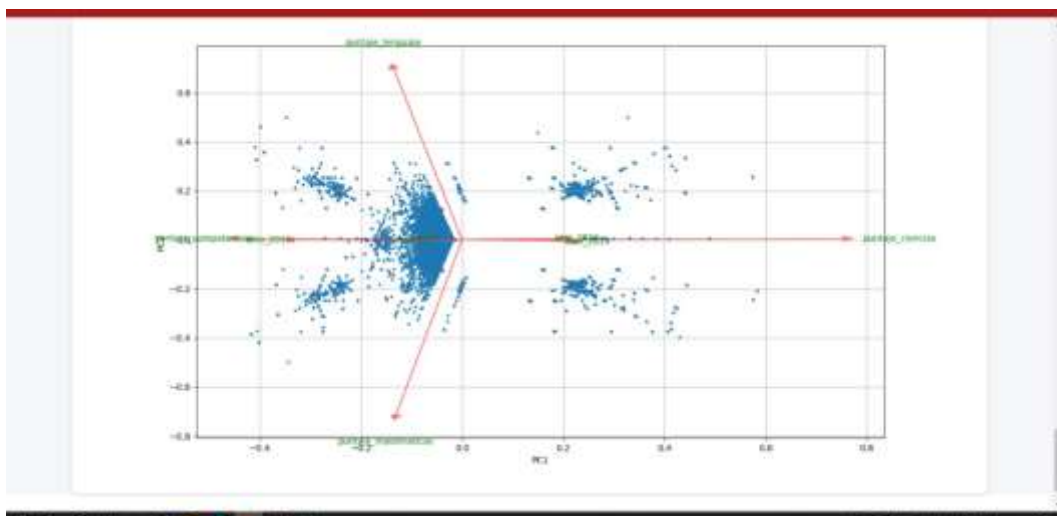
b) Biplot

Los **Biplots** son un tipo de gráfico exploratorio usado en estadística, se trata de una generalización multivariante de un diagrama de dispersión de dos o más variables, el biplot aproxima la distribución de una muestra multivariante en un espacio de dimensión reducida, normalmente de dimensión dos, y superpone sobre la misma representación de las variables sobre las que se mide la muestra. [42].

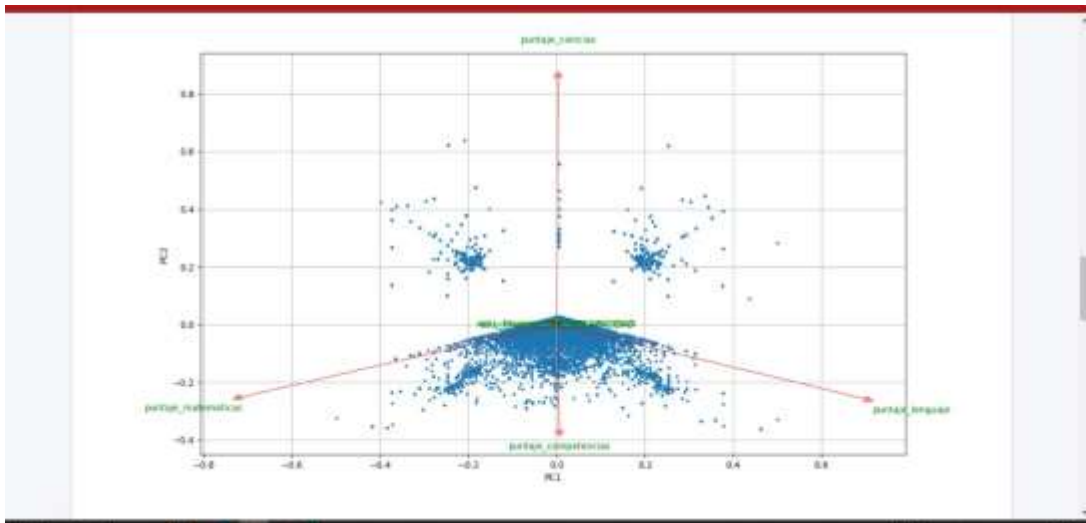
Para ingresar se debe elegir la opción estadística en el menú de opciones y marcar la opción biplot, cabe aclarar que no se hace ninguna búsqueda de categoría o filtro, puesto que se encuentran todas las variables y categorías que son representados por los PCA dentro de la programación del biplot.



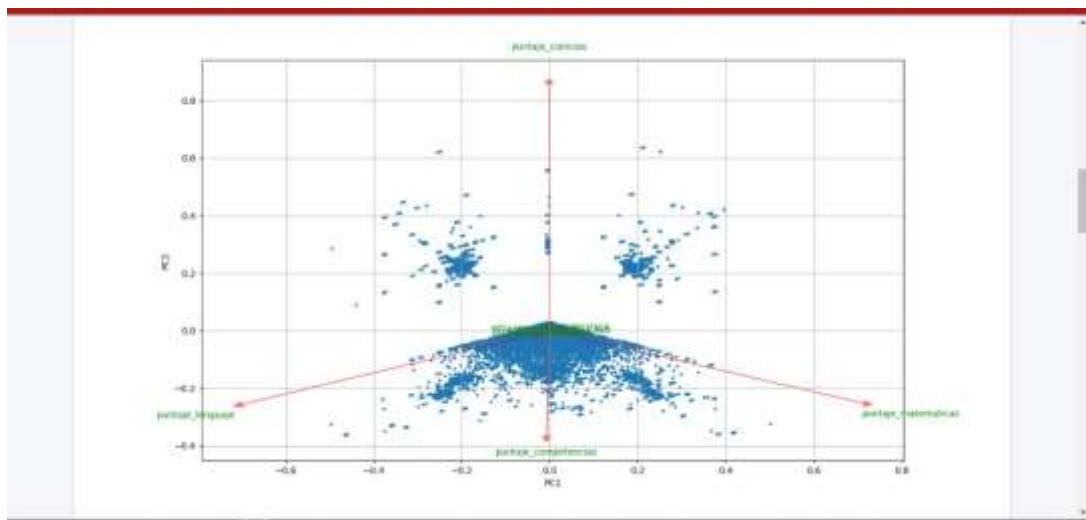
Podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto al año, con una variabilidad del 54%.



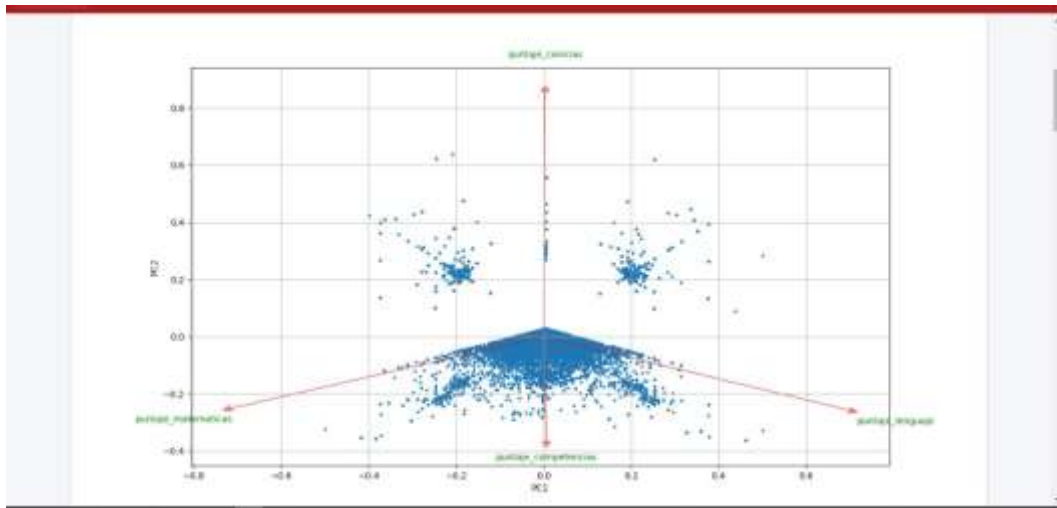
Podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a la discapacidad, con una variabilidad del 65%.



Podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a género, con una variabilidad del 53%.



Podemos observar la implementación de biplot, a la correlación de puntaje de las diferentes materias, con una variabilidad del 65%.



Podemos observar la implementación de biplot, a la correlación de puntaje de materias respecto a la zona, con una variabilidad del 45%.

