

# Análisis Regional de la Influencia de Factores Socioeconómicos en los Resultados de la Prueba Saber 11 en Colombia (Segundo Semestre 2019)

Juan David Bocanegra<sup>1</sup>, María José Castillo<sup>2</sup>, Luisa Guevara<sup>3</sup>, Aura León<sup>4</sup>

Dpto. de Matemáticas  
Universidad Externado de Colombia  
Pregrado Ciencia de Datos  
Curso de Bases de Datos  
Bogotá, Colombia

{<sup>1</sup>juan.bocanegra,<sup>2</sup>maria.castillo,<sup>3</sup>luisa.guevara,<sup>4</sup>aura.leon}@est.ueexternado.edu.co

November 22, 2023

## Contents

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Características del proyecto de investigación que hace uso de Bases de Datos</b>	<b>4</b>
2.1	Título del proyecto de investigación . . . . .	4
2.2	Objetivo general . . . . .	4
2.2.1	Objetivos específicos . . . . .	4
2.3	Alcance . . . . .	4
2.4	Pregunta de investigación . . . . .	4
2.5	Hipótesis . . . . .	4
<b>3</b>	<b>Reflexiones sobre el origen de datos e información</b>	<b>5</b>
3.1	¿Cuál es el origen de los datos e información? . . . . .	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? . . . . .	5
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? . . . . .	5
3.4	¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto? . . . . .	5
<b>4</b>	<b>Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)</b>	<b>6</b>
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto . . . . .	6
4.2	Diagrama modelo de datos . . . . .	6
4.3	Imágenes de la Base de Datos . . . . .	7
4.4	Código SQL - lenguaje de definición de datos (DDL) . . . . .	8
4.5	Código SQL - Manipulación de datos (DML) . . . . .	10
4.6	Código SQL + Resultados: Vistas . . . . .	12
4.7	Código SQL + Resultados: Triggers . . . . .	15
4.8	Código SQL + Resultados: Función 1 . . . . .	16
4.9	Código SQL + Resultados: Procedimientos Almacenados . . . . .	17
<b>5</b>	<b>Bases de Datos No-SQL</b>	<b>19</b>
5.1	Diagrama Bases de Datos No-SQL . . . . .	19
5.2	SMBD utilizado para la Base de Datos No-SQL . . . . .	24

<b>6</b>	<b>Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos</b>	<b>25</b>
6.1	Ejemplo de aplicación de ETL y Bodega de Datos . . . . .	25
6.1.1	Explicación del Código . . . . .	25
6.2	Automatización de Datos . . . . .	26
6.3	Integración de Datos . . . . .	26
<b>7</b>	<b>Próximos pasos (<i>Tercera entrega</i>)</b>	<b>27</b>
<b>8</b>	<b>Lecciones aprendidas (<i>Tercera entrega</i>)</b>	<b>28</b>
<b>9</b>	<b>Referencias Bibliográficas</b>	<b>29</b>

# 1 Introducción

Colombia siempre ha querido mejorar la calidad educativa además de garantizar un acceso equitativo a la educación de todos los niveles, esto por medio de la creación de políticas socioeconómicas, para mejorar la calidad de vida y el desarrollo del país. A pesar de esto, el rendimiento académico de los estudiantes sigue siendo una preocupación, ya que los datos tanto a nivel nacional como a nivel internacional muestran que los colombianos enfrentan dificultades académicas.

Teniendo esto en cuenta, el Instituto Colombiano para el Fomento de la Educación Superior (ICFES) ha tenido un papel fundamental, ya que llevan acabo evaluaciones que aunque no capturan todos los aspectos del conocimiento y habilidades de los estudiantes, sigue siendo la principal herramienta para evaluar la calidad educativa, el rendimiento académico y a su vez es un evaluador del desempeño alcanzado por los colegios.

El presente trabajo se enfoca en el análisis regional de la influencia de factores sociales, personales e institucionales en los resultados obtenidos por los estudiantes colombianos en la prueba Saber 11 en el segundo semestre de 2019. Utilizando metodologías de análisis descriptivo, exploratorio e inferencial, se pretende examinar el impacto de variables como el estrato socioeconómico, el género, la etnia, entre otros, en los resultados académicos. El análisis detallado de los factores que afectan el rendimiento académico, realizado en este trabajo, contribuirá a una comprensión más profunda de los desafíos educativos y proporcionará evidencia empírica relevante para informar las políticas educativas futuras.

## **2 Características del proyecto de investigación que hace uso de Bases de Datos**

### **2.1 Título del proyecto de investigación**

Análisis Regional de la Influencia de Factores Socioeconómicos en los Resultados de la Prueba Saber 11 en Colombia (Segundo Semestre 2019)

### **2.2 Objetivo general**

Examinar el impacto de variables como el estrato socioeconómico, el género, la etnia, entre otros, en los resultados académicos de la prueba Saber 11 en el segundo semestre de 2019 a nivel nacional, abarcando todas las regiones de Colombia.

#### **2.2.1 Objetivos específicos**

- Evaluar la dependencia entre el estrato socioeconómico de los estudiantes y sus resultados en la prueba Saber 11.
- Observar diferencias en los resultados de la prueba Saber 11 según el género de los estudiantes.
- Distinguir el impacto de la etnia en los puntajes de la prueba Saber 11 y explorar diferencias entre grupos étnicos.
- Estudiar cómo variables institucionales, como el tipo de colegio, afectan el rendimiento en la prueba Saber 11.
- Comparar puntajes en la prueba Saber 11 para identificar diferencias regionales significativas.

### **2.3 Alcance**

Se llevará a cabo un estudio a nivel nacional, abarcando todas las regiones de Colombia para obtener una visión completa y representativa de la educación en el país. La muestra de datos se obtuvo de un número considerable de estudiantes que participaron en la prueba Saber 11 2019-2, lo que asegura la fiabilidad de los resultados.

En términos metodológicos, el proyecto emplea diversas técnicas, como análisis descriptivo, exploratorio e inferencial, lo que permitirá un análisis profundo de la relación entre los factores mencionados y los resultados académicos. Estas herramientas estadísticas serán fundamentales para identificar patrones y tendencias en los datos recopilados.

Es relevante destacar que los resultados de este proyecto tienen una gran importancia tanto en el ámbito académico como en la formulación de políticas educativas en Colombia. Los hallazgos proporcionarán información esencial para guiar el diseño de estrategias educativas más efectivas y abordar las disparidades presentes en el sistema educativo del país.

### **2.4 Pregunta de investigación**

¿Cómo influyen los factores sociales, personales e institucionales, como el estrato socioeconómico, el género y la etnia en los resultados académicos de la prueba Saber 11, tomando como referencia los resultados del segundo semestre de 2019 a nivel nacional, abarcando todas las regiones de Colombia?

### **2.5 Hipótesis**

Para comprender la dependencia entre los puntajes y varios factores en las diferentes regiones colombianas, se tendrá un enfoque específico en comparar la región andina con otras, dado que se sostiene que la región andina, debido a sus mayores niveles de presencia estatal, calidad de vida, acceso educativo, menor discriminación, y menor incidencia de problemas asociados a áreas rurales, podría demostrar puntajes superiores en comparación con otras regiones.

### **3 Reflexiones sobre el origen de datos e información**

#### **3.1 ¿Cuál es el origen de los datos e información?**

Es una base de datos anonimizada suministrada por el Instituto Colombiano para la Evaluación de la Calidad de la Educación (ICFES) cuya extracción fue de Datos Abiertos Colombia. Constituida por 546.000 filas y 82 columnas, esta información se deriva de los cuestionarios de las pruebas Saber, las cuales emplean la metodología del Ítem TRI, que permite construir una escala histórica de medición entre periodos, garantizando comparabilidad. Con esta información se construye el índice de nivel socioeconómico (INSE), el cual permite caracterizar a la población de una manera integral, donde se pretende representar la posición económica o estatus socioeconómico.

#### **3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información?**

El manejo y uso de esta información está sujeto a consideraciones éticas y legales fundamentales. Se garantiza la confidencialidad y la privacidad de los datos de los estudiantes, ya que la base de datos se encuentra anonimizada para proteger la identidad de cada individuo. Además, se sigue la normativa de protección de datos y se respeta la legislación vigente, incluyendo el habeas data. Cualquier análisis o divulgación de resultados se realizará cumpliendo con los principios éticos y legales que rigen la investigación académica y estadística, asegurando así el respeto a los derechos y la integridad de los participantes.

#### **3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?**

Los retos en la base de datos del ICFES 2019 incluye garantizar la precisión y consistencia de la información, abordar la complejidad de los datos y lidiar con variables faltantes, como las regiones, que deben clasificarse desde los datos de departamentos. El enfoque incluye procesos de limpieza y validación, documentación detallada y una planificación para la mejora continua en la gestión de datos y el modelo de datos relacionales, asegurando la calidad de los datos para análisis e investigación.

#### **3.4 ¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto?**

En el Sistema de Gestión de Bases de Datos (SMBD), se busca mejorar la agilidad y eficiencia en la gestión de datos, lo que permite ahorrar tiempo gracias a sus mecanismos de consultas rápidas. Esto evita problemas como la duplicación e inconsistencia de datos, lo que podría afectar la integridad de la información. Además, brinda un mayor control sobre las actualizaciones, garantizando el control de acceso a los datos y facilitando la generación automática de informes profesionales.

## 4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

### 4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

Dado que se utilizó SQL como lenguaje de consulta, se optó por Oracle Database como Sistema de Gestión de Bases de Datos (SMBD) debido a varias razones. En primer lugar, Oracle es reconocido por su capacidad de escalar y gestionar grandes conjuntos de datos. Además, Oracle ofrece un sólido soporte para SQL, proporcionando extensiones y características avanzadas que serán valiosas para el análisis. También ofrece un conjunto completo de herramientas, incluyendo Oracle SQL Developer, que facilitará el desarrollo, la administración y el análisis de datos en la base de datos Oracle, razón por la cual será implementado. Además, Oracle tiene una reputación de confiabilidad y estabilidad como uno de los principales proveedores de bases de datos en el mundo, lo que brinda una gran confianza en su capacidad para llevar a cabo el análisis de manera efectiva.

### 4.2 Diagrama modelo de datos

En esta sección, se presenta el Modelo Lógico y el Modelo Relacional de la base de datos que se ha diseñado para nuestro proyecto. Estos modelos son esenciales para comprender la estructura de la base de datos y cómo se relacionan las entidades y atributos.

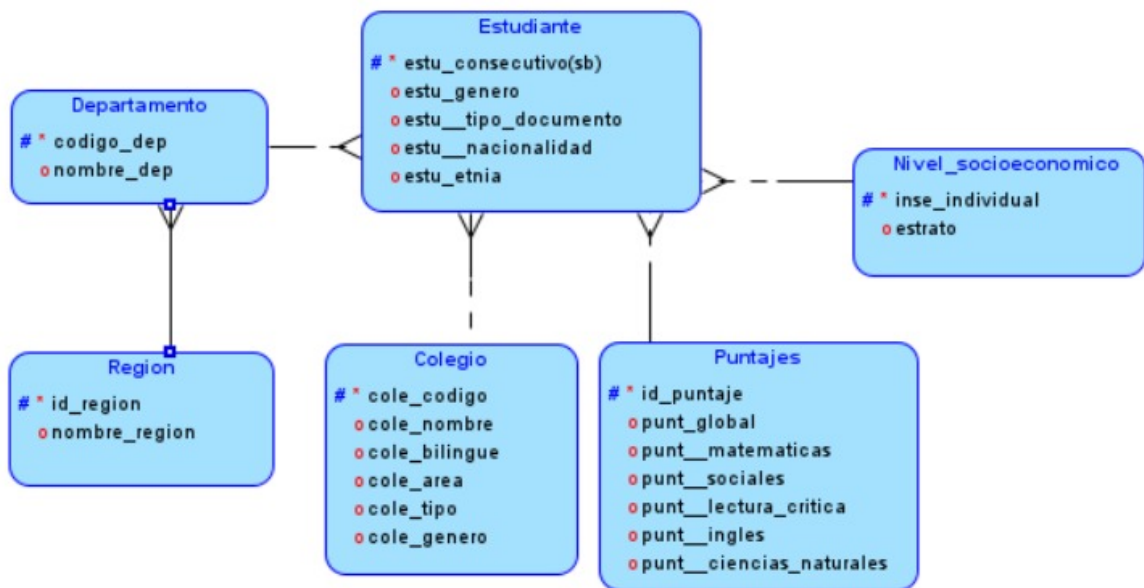


Figure 1: Modelo Lógico de la Base de Datos

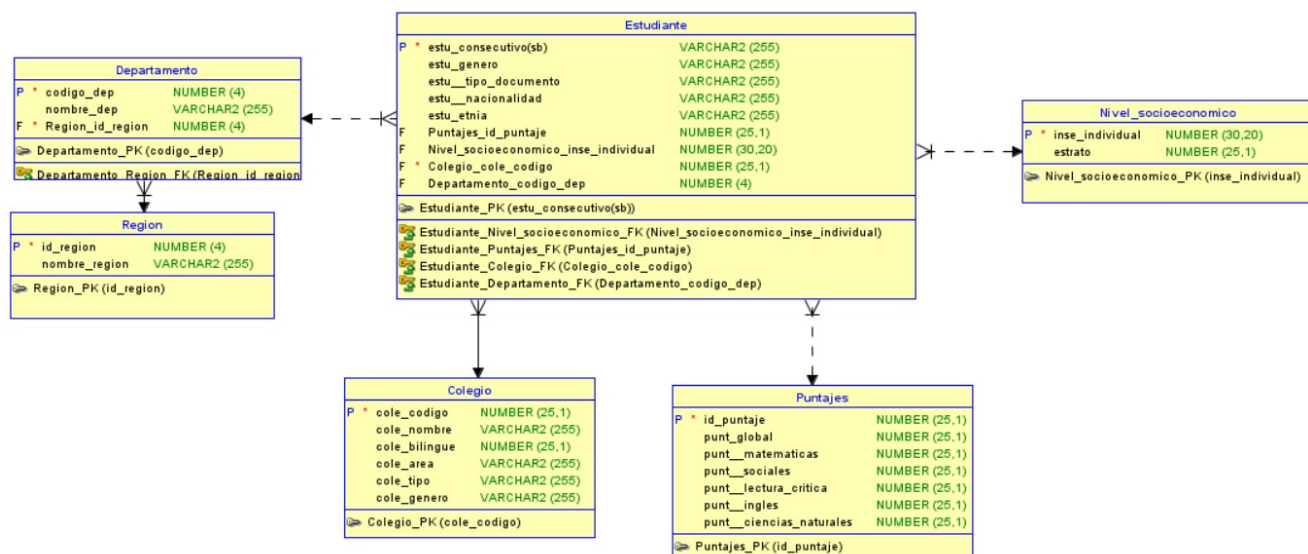


Figure 2: Modelo Relacional de la Base de Datos

### 4.3 Imágenes de la Base de Datos

Inicialmente se realizó una limpieza, se quitaron los valores nulos y para un mejor manejo de los datos se dejó todo en minúscula, como el análisis será por regiones se crea una nueva columna de carácter numérica que tomara valores entre 1-7 donde cada número representa una región o si es extranjero . A continuación, se muestran estas imágenes para brindar una vista más detallada de nuestra base de datos.

COLE_COD_DANE_SEDE	COLE_NOMBRE_SEDE	COLE_BILINGUE	COLE_NATURALEZA	COLE_GENERO
147707000156	col dptal antonio bruges carmona	n	oficial	mixto
311001105383	gimn san angelo	-	no oficial	mixto
113001001719	institucion educativa promocion social de c/gena.	n	oficial	mixto
111001107069	col virginia gutierrez de pineda (ins educ dist)	-	oficial	mixto
111001094897	cent educ dist palermo cedip	n	oficial	mixto
...	...	...	...	...
105440000131	i. e. san jose	n	oficial	mixto
311848000596	gimnasio ingles campestre	n	no oficial	mixto
181001000041	santa teresita	-	oficial	mixto
368872000051	colegio eliseo pinilla rueda	n	oficial	mixto
111001015172	cent educ dist alejandro obregon	n	oficial	mixto

Figure 3: Tabla Colegios

Para la primera tabla se tuvo en cuenta las columnas relacionadas con la entidad Colegio como código dane, nombre del colegio y algunas características como si este es bilingüe, mixto o privado.

U_TIPODOCUMENTO	ESTU_NACIONALIDAD	ESTU_ETNIA	COLE_COD_DANE_SEDE	ESTU_INSE_INDIVIDUAL	COLE_COD_DANE_SEDE	ID_PUNTAJE	CODIGO_DEP
ti	colombia	-	147707000156	50.9421557404314	147707000156	0.0	47
ti	colombia	-	311001105383	59.767864431576	311001105383	1.0	11
ti	colombia	-	113001001719	50.7642383644749	113001001719	2.0	13
ti	colombia	ninguno	111001107069	45.7791353123516	111001107069	3.0	11
ti	colombia	-	111001094897	52.9109464277177	111001094897	4.0	11
...	...	...	...	...	...	...	...
cc	colombia	-	105440000131	47.5807720132296	105440000131	NaN	05
ti	colombia	-	311848000596	57.406418472062	311848000596	NaN	11
ti	colombia	-	181001000041	71.2254308010465	181001000041	NaN	81
ti	colombia	-	368872000051	38.1334451267846	368872000051	NaN	68
ti	colombia	-	111001015172	49.7833058735661	111001015172	NaN	11

Figure 4: Tabla Estudiante

Para la siguiente tabla se tuvo en cuenta las columnas relacionadas con la entidad Estudiante la cual es la principal y contiene toda las llaves foráneas, tipo de documento, nacionalidad, si pertenece a alguna etnia, etc.

ESTU_INSE_INDIVIDUAL	FAMI_ESTRATOVIVIENDA
50.9421557404314	estrato 3
59.767864431576	estrato 3
50.7642383644749	estrato 1
45.7791353123516	estrato 3
52.9109464277177	estrato 3
...	...
47.5807720132296	estrato 2
57.406418472062	estrato 3
71.2254308010465	estrato 2
38.1334451267846	estrato 1
49.7833058735661	estrato 3

Figure 5: Tabla Nivel socioeconómico

Para la siguiente tabla se tuvo en cuenta las columnas relacionadas con la entidad Nivel socioeconómico, contiene el inse individual de cada estudiante y el estrato.

PUNT_GLOBAL	PUNT_MATEMATICAS	PUNT_SOCIALES_CIUADANAS	PUNT_LECTURA_CRITICA	PUNT_INGLES	PUNT_C_NATURALES
208	48	30	47	54.0	37
299	65	59	60	63.0	54
299	57	74	66	64.0	41
309	54	73	62	53.0	61
288	57	57	63	52.0	55
...	...	...	...	...	...
360	78	74	76	58.0	65
360	73	67	75	74.0	72
377	83	77	72	72.0	71
278	61	52	59	46.0	54
365	73	71	76	74.0	72

Figure 6: Tabla Puntajes icfes

Para la siguiente tabla se tuvo en cuenta las columnas relacionadas con la entidad Puntajes, contiene el puntaje global y los puntajes individuales de cada materia evaluada.

ID_REGION	NOMBRE REGION	ESTU_CONSECUTIVO
3	caribe	sb11201940031558
1	andina	sb11201940303339
3	caribe	sb11201940255017
1	andina	sb11201940154799
1	andina	sb11201940280831
...	...	...
1	andina	sb11201940195862
1	andina	sb11201940253002
4	orinoquia	sb11201940364767
1	andina	sb11201940210026
1	andina	sb11201940074679

Figure 7: Tabla de Regiones

Para la siguiente tabla se tuvo en cuenta las columnas relacionadas con la entidad Región, contiene 3 columnas, el id de cada región, nombre de la región y el código único de cada estudiante.

#### 4.4 Código SQL - lenguaje de definición de datos (DDL)

El código SQL proporcionado se generó a partir de Oracle Modeler después de diseñar el modelo entidad-relación. A continuación, se muestra este código para comprender cómo se crea la estructura de la base de datos, específicamente sus tablas y sus relaciones.

Listing 1: Lenguaje de Definición de Datos

```

1 CREATE TABLE colegio (
2     cole_codigo    NUMBER(25, 1) NOT NULL,
3     cole_nombre    VARCHAR2(255),
4     cole_bilingue  NUMBER(25, 1),
5     cole_area      VARCHAR2(255),

```



```

6      cole_tipo      VARCHAR2(255),
7      cole_genero    VARCHAR2(255)
8  );
9
10 -- Definir una clave primaria para la tabla colegio
11 ALTER TABLE colegio ADD CONSTRAINT colegio_pk PRIMARY KEY (cole_codigo);
12
13 -- Crear la tabla departamento
14 CREATE TABLE departamento (
15     codigo_dep      NUMBER(4) NOT NULL,
16     nombre_dep      VARCHAR2(255),
17     region_id_region NUMBER(4) NOT NULL
18 );
19
20 -- Definir una clave primaria para la tabla departamento
21 ALTER TABLE departamento ADD CONSTRAINT departamento_pk PRIMARY KEY (codigo_dep);
22
23 -- Crear la tabla estudiante
24 CREATE TABLE estudiante (
25     estu_consecutivo_sb      VARCHAR2(255) NOT NULL,
26     estu_genero              VARCHAR2(255),
27     estu_tipo_documento      VARCHAR2(255),
28     estu_nacionalidad        VARCHAR2(255),
29     estu_etnia               VARCHAR2(255),
30     puntajes_id_puntaje      NUMBER(25, 1),
31     nivel_socioeconomico_inse_individual NUMBER(30, 20),
32     colegio_cole_codigo      NUMBER(25, 1) NOT NULL,
33     departamento_codigo_dep  NUMBER(4)
34 );
35
36 -- Definir una clave primaria para la tabla estudiante
37 ALTER TABLE estudiante ADD CONSTRAINT estudiante_pk PRIMARY KEY (estu_consecutivo_sb);
38
39 -- Crear la tabla nivel_socioeconomico
40 CREATE TABLE nivel_socioeconomico (
41     inse_individual NUMBER(30, 20) NOT NULL,
42     estrato         NUMBER(25, 1)
43 );
44
45 -- Definir una clave primaria para la tabla nivel_socioeconomico
46 ALTER TABLE nivel_socioeconomico ADD CONSTRAINT nivel_socioeconomico_pk
47 PRIMARY KEY (inse_individual);
48
49 -- Crear la tabla puntajes
50 CREATE TABLE puntajes (
51     id_puntaje      NUMBER(25, 1) NOT NULL,
52     punt_global     NUMBER(25, 1),
53     punt_matematicas NUMBER(25, 1),
54     punt_sociales   NUMBER(25, 1),
55     punt_lectura_critica NUMBER(25, 1),
56     punt_ingles     NUMBER(25, 1),
57     punt_ciencias_naturales NUMBER(25, 1)
58 );
59
60 -- Definir una clave primaria para la tabla puntajes
61 ALTER TABLE puntajes ADD CONSTRAINT puntajes_pk PRIMARY KEY (id_puntaje);
62
63 -- Crear la tabla region
64 CREATE TABLE region (
65     id_region      NUMBER(4) NOT NULL,
66     nombre_region  VARCHAR2(255)
67 );

```

```

68
69 -- Definir una clave primaria para la tabla region
70 ALTER TABLE region ADD CONSTRAINT region_pk PRIMARY KEY (id_region);
71
72 -- Agregar restricciones de clave externa (FK)
73
74 ALTER TABLE departamento
75     ADD CONSTRAINT departamento_region_fk FOREIGN KEY (region_id_region)
76     REFERENCES region (id_region);
77
78 ALTER TABLE estudiante
79     ADD CONSTRAINT estudiante_colegio_fk FOREIGN KEY (colegio_cole_codigo)
80     REFERENCES colegio (cole_codigo);
81
82 ALTER TABLE estudiante
83     ADD CONSTRAINT estudiante_departamento_fk FOREIGN KEY (departamento_codigo_dep)
84     REFERENCES departamento (codigo_dep);
85
86 ALTER TABLE estudiante
87     ADD CONSTRAINT estudiante_nivel_socioeconomico_fk
88     FOREIGN KEY (nivel_socioeconomico_inse_individual)
89     REFERENCES nivel_socioeconomico (inse_individual);
90
91 ALTER TABLE estudiante
92     ADD CONSTRAINT estudiante_puntajes_fk FOREIGN KEY (puntajes_id_puntaje)
93     REFERENCES puntajes (id_puntaje);

```

## 4.5 Código SQL - Manipulación de datos (DML)

El lenguaje de manipulación de datos (DML) es esencial para agregar registros a las tablas y poblar la base de datos con información. A continuación, una muestra del código SQL utilizado para añadir datos a las tablas correspondientes.

Listing 2: Insertar Datos a COLEGIO

```

1 -- Row 36
2 INSERT INTO COLEGIO (COLE_CODIGO, COLE_NOMBRE, COLE_BILINGUE, COLE_AREA, COLE_TIPO,
3 COLE_GENERO) VALUES (105001012106, 'inst educ juan maria cespedes', 0.0, 'urbano', 'oficial',
4 'mixto');
5 --Row 37
6 INSERT INTO COLEGIO (COLE_CODIGO, COLE_NOMBRE, COLE_BILINGUE, COLE_AREA, COLE_TIPO,
7 COLE_GENERO) VALUES (154498000018, 'col jose eusebio caro', 0.0, 'urbano', 'oficial', 'mixto');
8 --Row 38
9 INSERT INTO COLEGIO (COLE_CODIGO, COLE_NOMBRE, COLE_BILINGUE, COLE_AREA, COLE_TIPO,
10 COLE_GENERO) VALUES (108758000015, 'institucion educativa francisco jose de caldas', 0.0,
11 'urbano', 'oficial', 'mixto');
12 --Row 39
13 INSERT INTO COLEGIO (COLE_CODIGO, COLE_NOMBRE, COLE_BILINGUE, COLE_AREA, COLE_TIPO,
14 COLE_GENERO) VALUES (111001014346, 'cent educ dist rodrigo arenas betancourt', 0.0, 'urbano',
15 'oficial', 'mixto');
16 --Row 40
17 INSERT INTO COLEGIO (COLE_CODIGO, COLE_NOMBRE, COLE_BILINGUE, COLE_AREA, COLE_TIPO,
18 COLE_GENERO) VALUES (168001001921, 'ie tec damaso zapata', 0.0, 'urbano', 'oficial', 'mixto');

```

Listing 3: Insertar Datos a DEPARTAMENTO

```

1 --Row 1
2 INSERT INTO DEPARTAMENTO (NOMBRE_DEP, CODIGO_DEP, REGION_ID_REGION) VALUES ('magdalena', 47, 3);
3 --Row 2
4 INSERT INTO DEPARTAMENTO (NOMBRE_DEP, CODIGO_DEP, REGION_ID_REGION) VALUES ('bogota', 11, 1);
5 --Row 3
6 INSERT INTO DEPARTAMENTO (NOMBRE_DEP, CODIGO_DEP, REGION_ID_REGION) VALUES ('bolivar', 13, 3);

```

```

7 --Row 4
8 INSERT INTO DEPARTAMENTO (NOMBRE_DEP, CODIGO_DEP, REGION_ID_REGION) VALUES ('bogota',11,1);
9 --Row 5
10 INSERT INTO DEPARTAMENTO (NOMBRE_DEP, CODIGO_DEP, REGION_ID_REGION) VALUES ('bogota',11,1);

```

Listing 4: Insertar Datos a ESTUDIANTE

```

1
2 --Row 1
3 INSERT INTO ESTUDIANTE (ESTU_CONSECUTIVO_SB, ESTU_GENERO, ESTU_TIPO_DOCUMENTO,
4 ESTU_NACIONALIDAD, ESTU_ETNIA, PUNTAJES_ID_PUNTAJE, NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL,
5 COLEGIO_COLE_CODIGO, DEPARTAMENTO_CODIGO_DEP) VALUES ('sb11201940031558','m','ti',
6 'colombia','- ',0.0,50.9421557404314,147707000156,3);
7 --Row 2
8 INSERT INTO ESTUDIANTE (ESTU_CONSECUTIVO_SB, ESTU_GENERO, ESTU_TIPO_DOCUMENTO,
9 ESTU_NACIONALIDAD, ESTU_ETNIA, PUNTAJES_ID_PUNTAJE, NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL,
10 COLEGIO_COLE_CODIGO, DEPARTAMENTO_CODIGO_DEP) VALUES ('sb11201940303339','m','ti',
11 'colombia','- ',1.0,59.767864431576,311001105383,1);
12 --Row 3
13 INSERT INTO ESTUDIANTE (ESTU_CONSECUTIVO_SB, ESTU_GENERO, ESTU_TIPO_DOCUMENTO,
14 ESTU_NACIONALIDAD, ESTU_ETNIA, PUNTAJES_ID_PUNTAJE, NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL,
15 COLEGIO_COLE_CODIGO, DEPARTAMENTO_CODIGO_DEP) VALUES ('sb11201940255017','m','ti',
16 'colombia','- ',2.0,50.7642383644749,113001001719,3);
17 --Row 4
18 INSERT INTO ESTUDIANTE (ESTU_CONSECUTIVO_SB, ESTU_GENERO, ESTU_TIPO_DOCUMENTO,
19 ESTU_NACIONALIDAD, ESTU_ETNIA, PUNTAJES_ID_PUNTAJE, NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL,
20 COLEGIO_COLE_CODIGO, DEPARTAMENTO_CODIGO_DEP) VALUES ('sb11201940154799','m','ti',
21 'colombia','ninguno',3.0,45.7791353123516,111001107069,1);
22 --Row 5
23 INSERT INTO ESTUDIANTE (ESTU_CONSECUTIVO_SB, ESTU_GENERO, ESTU_TIPO_DOCUMENTO,
24 ESTU_NACIONALIDAD, ESTU_ETNIA, PUNTAJES_ID_PUNTAJE, NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL,
25 COLEGIO_COLE_CODIGO, DEPARTAMENTO_CODIGO_DEP) VALUES ('sb11201940280831','m','ti',
26 'colombia','- ',4.0,52.9109464277177,111001094897,1);

```

Listing 5: Insertar Datos a NIVEL SOCIOECONOMICO

```

1 --Row 1
2 INSERT INTO NIVEL_SOCIOECONOMICO (INSE_INDIVIDUAL, ESTRATO, ESTUDIANTE_ESTU_CONSECUTIVO_SB)
3 VALUES (50.9421557404314,3.0,'sb11201940031558');
4 --Row 2
5 INSERT INTO NIVEL_SOCIOECONOMICO (INSE_INDIVIDUAL, ESTRATO, ESTUDIANTE_ESTU_CONSECUTIVO_SB)
6 VALUES (59.767864431576,3.0,'sb11201940303339');
7 --Row 3
8 INSERT INTO NIVEL_SOCIOECONOMICO (INSE_INDIVIDUAL, ESTRATO, ESTUDIANTE_ESTU_CONSECUTIVO_SB)
9 VALUES (50.7642383644749,1.0,'sb11201940255017');
10 --Row 4
11 INSERT INTO NIVEL_SOCIOECONOMICO (INSE_INDIVIDUAL, ESTRATO, ESTUDIANTE_ESTU_CONSECUTIVO_SB)
12 VALUES (45.7791353123516,3.0,'sb11201940154799');
13 --Row 5
14 INSERT INTO NIVEL_SOCIOECONOMICO (INSE_INDIVIDUAL, ESTRATO, ESTUDIANTE_ESTU_CONSECUTIVO_SB)
15 VALUES (52.9109464277177,3.0,'sb11201940280831');

```

Listing 6: Insertar Datos a PUNTAJES

```

1 --Row 1
2 INSERT INTO PUNTAJES (ID_PUNTAJE, PUNT_GLOBAL, PUNT_MATEMATICAS, PUNT_SOCIALES,
3 PUNT_LECTURA_CRITICA, PUNT_INGLES, PUNT_CIENCIAS_NATURALES)
4 VALUES (0,208,48,30,47,54.0,37);
5 --Row 2
6 INSERT INTO PUNTAJES (ID_PUNTAJE, PUNT_GLOBAL, PUNT_MATEMATICAS, PUNT_SOCIALES,
7 PUNT_LECTURA_CRITICA, PUNT_INGLES, PUNT_CIENCIAS_NATURALES)
8 VALUES (1,299,65,59,60,63.0,54);
9 --Row 3

```

```

10 INSERT INTO PUNTAJES (ID_PUNTAJE, PUNT_GLOBAL, PUNT_MATEMATICAS, PUNT_SOCIALES,
11 PUNT_LECTURA_CRITICA, PUNT_INGLES, PUNT_CIENCIAS_NATURALES)
12 VALUES (2,299,57,74,66,64.0,41);
13 --Row 4
14 INSERT INTO PUNTAJES (ID_PUNTAJE, PUNT_GLOBAL, PUNT_MATEMATICAS, PUNT_SOCIALES,
15 PUNT_LECTURA_CRITICA, PUNT_INGLES, PUNT_CIENCIAS_NATURALES)
16 VALUES (3,309,54,73,62,53.0,61);
17 --Row 5
18 INSERT INTO PUNTAJES (ID_PUNTAJE, PUNT_GLOBAL, PUNT_MATEMATICAS, PUNT_SOCIALES,
19 PUNT_LECTURA_CRITICA, PUNT_INGLES, PUNT_CIENCIAS_NATURALES)
20 VALUES (4,288,57,57,63,52.0,55);

```

Listing 7: Insertar Datos a REGION

```

1 --Row 1
2 INSERT INTO REGION (ID_REGION, NOMBRE_REGION) VALUES (3,'caribe');
3 --Row 2
4 INSERT INTO REGION (ID_REGION, NOMBRE_REGION) VALUES (1,'andina');
5 --Row 3
6 INSERT INTO REGION (ID_REGION, NOMBRE_REGION) VALUES (3,'caribe');
7 --Row 4
8 INSERT INTO REGION (ID_REGION, NOMBRE_REGION) VALUES (1,'andina');
9 --Row 5
10 INSERT INTO REGION (ID_REGION, NOMBRE_REGION) VALUES (1,'andina');

```

## 4.6 Código SQL + Resultados: Vistas

Las vistas ofrecen una perspectiva unificada y eficiente de la información. Cada vista combina datos de múltiples tablas para proporcionar una visión integral de aspectos clave.

Listing 8: Vista 1

```

1 -- VISTA 1:
2
3 -- Se crea una vista la cual combina las tablas COLEGIO y ESTUDIANTE
4 -- por medio de un 'INNER JOIN'
5
6 CREATE VIEW VistaColegiosEstudiantes AS
7
8 -- Se seleccionan columnas especificas
9
10 SELECT
11     C.COLE_CODIGO,
12     C.COLE_NOMBRE,
13     C.COLE_BILINGUE,
14     C.COLE_AREA,
15     C.COLE_TIPO,
16     C.COLE_GENERO,
17     E.ESTU_CONSECUTIVO_SB,
18     E.ESTU_GENERO,
19     E.ESTU_TIPO_DOCUMENTO,
20     E.ESTU_NACIONALIDAD,
21     E.ESTU_ETNIA
22 FROM COLEGIO C
23 INNER JOIN ESTUDIANTE E ON C.COLE_CODIGO = E.COLEGIO_COLE_CODIGO;
24
25 -- Esta vista permite obtener informacion conjunta de colegios y estudiantes.

```

Worksheet		Query Builder						
SELECT * FROM VistaColegiosEstudiantes;								
Script Output x Query Result x								
SQL		Fetches 50 rows in 0,013 seconds						
COLE_CODIGO	COLE_NOMBRE	COLE_BILINGUE	COLE_AREA	COLE_TIPO	COLE_GENERO	ESTU_CONSECUTIVO_SB	ESTU_GENERO	ESTU_TIPO_DOCUMENTO
13	108849000042 institucion educativa nuestra se�sora del transito		urbano	oficial	mixto	sb11201940479565	m	ti
14	108849000042 institucion educativa nuestra se�sora del transito		urbano	oficial	mixto	sb11201940479753	f	ti
15	108849000042 institucion educativa nuestra se�sora del transito		urbano	oficial	mixto	sb11201940398443	f	ti
16	1086380000810 institucion educativa tecnico comercial de sabanalarga		urbano	oficial	mixto	sb11201940379154	m	ti
17	1086380000810 institucion educativa tecnico comercial de sabanalarga		urbano	oficial	mixto	sb11201940559757	m	ti
18	1086380000810 institucion educativa tecnico comercial de sabanalarga		urbano	oficial	mixto	sb11201940379597	f	cc
19	1086380000810 institucion educativa tecnico comercial de sabanalarga		urbano	oficial	mixto	sb11201940379790	f	ti
20	1050450014541.e. san francisco de asis		urbano	oficial	mixto	sb11201940137271	m	ti
21	1050450014541.e. san francisco de asis		urbano	oficial	mixto	sb11201940131514	f	ti
22	2733190011971.e.t. ca�sada rodeo - sede principal		rural	oficial	mixto	sb11201940610060	f	ti
23	125754000331 institucion educativa general santander		urbano	oficial	mixto	sb11201940242821	f	ti
24	320013022891 centro pedagogico estudiantil - sede principal		urbano	no oficial	mixto	sb11201940479649	f	cc
25	320013022891 centro pedagogico estudiantil - sede principal		urbano	no oficial	mixto	sb11201940479643	f	ti
26	320013022891 centro pedagogico estudiantil - sede principal		urbano	no oficial	mixto	sb11201940479440	f	ti
27	247001001472 sede 01 la revuelta		rural	oficial	mixto	sb11201940468204	f	ti
28	154001008606 col bas camilo daza	(null)	urbano	oficial	mixto	sb11201940348741	f	ti
29	154001008606 col bas camilo daza	(null)	urbano	oficial	mixto	sb11201940349951	m	ti
30	154001008606 col bas camilo daza	(null)	urbano	oficial	mixto	sb11201940348743	f	ti
31	1086060001251.e. jose david monteruma recuero - sede principal		urbano	oficial	mixto	sb11201940163983	f	ti
32	113001003061 institucion educativa hermano antonio ramos de la salle		urbano	oficial	mixto	sb11201940104434	m	ti
33	113001003061 institucion educativa hermano antonio ramos de la salle		urbano	oficial	mixto	sb11201940104765	f	ti
34	1735850001001.e.t. perez y aladana - sede principal		urbano	oficial	mixto	sb11201940476552	f	cc
35	1735850001001.e.t. perez y aladana - sede principal		urbano	oficial	mixto	sb11201940475460	f	ti
36	1207100120711.e. andres bello		urbano	oficial	mixto	sb11201940132922	m	ti
37	1207100120711.e. andres bello		urbano	oficial	mixto	sb11201940132125	f	ti

Listing 9: Vista 2

```

1
2 -- VISTA 2
3
4 -- Se crea una vista que une 'ESTUDIANTE' y 'PUNTAJES', mediante 'INNER JOIN'
5 CREATE VIEW VistaPuntajesEstudiantes AS
6
7 -- Se seleccionan columnas especificas
8
9 CREATE VIEW VistaPuntajesEstudiantes AS
10 SELECT
11     E. ESTU_CONSECUTIVO_SB ,
12     P. PUNT_GLOBAL ,
13     P. PUNT_MATEMATICAS ,
14     P. PUNT_SOCIALES ,
15     P. PUNT_LLECTURA_CRITICA ,
16     P. PUNT_INGLES ,
17     P. PUNT_CIENCIAS_NATURALES
18 FROM ESTUDIANTE E
19 INNER JOIN PUNTAJES P ON E. PUNTAJES_ID_PUNTAJE = P. ID_PUNTAJE;
20
21 -- Esta vista proporciona una vision de los puntajes de los estudiantes

```

Worksheet

Query Builder

SELECT \* FROM VistaPuntajesEstudiantes;

Script Output x

Query Result x

SQL | Fetched 50 rows in 0,112 seconds

	ESTU_CONSECUTIVO_SB	PUNT_GLOBAL	PUNT_MATEMATICAS	PUNT_SOCIALES	PUNT_LECTURA_CRITICA	PUNT_INGLES	PUNT_CIENCIAS_NATURALES
1	sb11201940298103	216	44	39	41	45	48
2	sb11201940116453	313	57	68	69	70	54
3	sb11201940083901	208	44	39	41	45	41
4	sb11201940461695	339	73	65	73	56	64
5	sb11201940179789	365	78	68	66	76	79
6	sb11201940278153	278	71	47	56	49	51
7	sb11201940624777	240	52	34	50	57	53
8	sb11201940565280	274	55	50	68	46	49
9	sb11201940156293	230	52	41	45	41	48
10	sb11201940346926	150	28	34	31	25	29
11	sb11201940273657	285	55	60	56	37	64
12	sb11201940469458	327	65	60	70	74	64
13	sb11201940572730	280	59	59	57	42	54
14	sb11201940117695	220	46	38	43	54	46
15	sb11201940129944	308	64	62	53	60	68
16	sb11201940051654	246	61	34	47	66	49
17	sb11201940082910	193	44	27	41	54	37
18	sb11201940113443	305	65	65	54	50	64
19	sb11201940459884	311	55	67	64	68	61
20	sb11201940396839	244	56	41	54	46	45
21	sb11201940027275	246	60	41	49	49	47
22	sb11201940149454	257	51	50	60	41	48
23	sb11201940489273	326	71	66	62	64	62
24	sb11201940164663	289	54	54	68	55	56

Listing 10: Vista 3

```

1
2 -- VISTA 3
3
4 -- Crea una vista que une 'ESTUDIANTE' y 'NIVEL SOCIOECONOMICO'
5
6 CREATE VIEW VistaSocioeconomicoEstudiantes AS
7 SELECT
8     E.ESTU_CONSECUTIVO_SB,
9     N.INSE_INDIVIDUAL,
10    N.ESTRATO
11 FROM ESTUDIANTE E
12 INNER JOIN NIVEL_SOCIOECONOMICO N ON
13 E.NIVEL_SOCIOECONOMICO_INSE_INDIVIDUAL = N.INSE_INDIVIDUAL;
14
15 -- Permite acceder a informacion socioeconomica de los estudiantes.

```

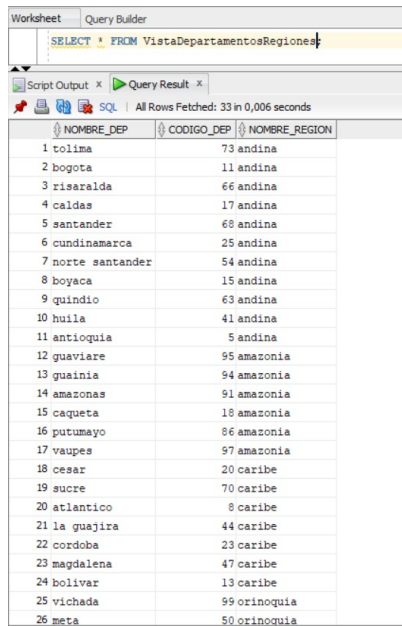
Worksheet		Query Builder	
		SELECT * FROM VistaSocioeconomicoEstudiantes;	
Script Output		Query Result	
		SQL   Fetched 50 rows in 0,066 seconds	
ESTU_CONSECUTIVO_SB	INSE_INDIVIDUAL	ESTRATO	
1 sb11201940016019	32.8509418309449	2	
2 sb11201940518354	61.6412300806444	3	
3 sb11201940317527	35.7389199645696	(null)	
4 sb11201940175803	56.5744952571204	2	
5 sb11201940172132	55.0209606890485	3	
6 sb11201940488663	56.5473408029092	2	
7 sb11201940265266	61.1794339460969	3	
8 sb11201940022853	77.8341504196365	4	
9 sb11201940578111	42.3799248612075	1	
10 sb11201940001069	60.9387530972817	2	
11 sb11201940290113	55.1355814537666	3	
12 sb11201940501474	43.6830233116113	3	
13 sb11201940560058	31.6836151074479	(null)	
14 sb11201940478498	43.3429055396621	3	
15 sb11201940466939	52.6728576101018	1	
16 sb11201940275394	33.3103897592472	1	
17 sb11201940107680	35.5990215303035	1	
18 sb11201940171553	64.2495897322774	2	
19 sb11201940083858	49.5078499071666	2	
20 sb11201940214441	57.8006575361215	4	
21 sb11201940376115	60.6009071695318	4	
22 sb11201940221854	30.6327567865682	1	
23 sb11201940064485	43.8490017357529	1	
24 sb11201940590973	53.0943024976001	3	
25 sb11201940278050	50.3482650290946	(null)	

Listing 11: Vista 4

```

1  -- VISTA 4
2
3  -- Se crea una vista que une 'DEPARTAMENTO' y 'REGION'
4
5  CREATE VIEW VistaDepartamentosRegiones AS
6  SELECT
7      D.NOMBRE_DEP ,
8      D.CODIGO_DEP ,
9      R.NOMBRE_REGION
10 FROM DEPARTAMENTO D
11 INNER JOIN REGION R ON D.REGION_ID_REGION = R.ID_REGION;
12
13
14 -- Esta vista proporciona una relacion entre departamentos y regiones.

```



NOMBRE_DEP	CODIGO_DEP	NOMBRE_REGION
1 tolima	73	andina
2 bogota	11	andina
3 risaralda	66	andina
4 caldas	17	andina
5 santander	68	andina
6 cundinamarca	25	andina
7 norte santander	54	andina
8 boyaca	15	andina
9 quindio	63	andina
10 huila	41	andina
11 antioquia	5	andina
12 guaviare	95	amazonia
13 guainia	94	amazonia
14 Amazonas	91	amazonia
15 caqueta	18	amazonia
16 putumayo	86	amazonia
17 vaupes	97	amazonia
18 cesar	20	caribe
19 sucre	70	caribe
20 atlantico	8	caribe
21 la guajira	44	caribe
22 cordoba	23	caribe
23 magdalena	47	caribe
24 bolivar	13	caribe
25 vichada	99	orinoquia
26 meta	50	orinoquia

## 4.7 Código SQL + Resultados: Triggers

Listing 12: Triggers

```

1  -- TRIGGERS
2  -- Crear el trigger
3  CREATE OR REPLACE TRIGGER mostrar_maximo_puntaje
4  BEFORE INSERT ON puntajes
5  FOR EACH ROW
6  DECLARE
7      max_puntaje NUMBER;
8  BEGIN
9  -- Consultar el valor maximo de los puntajes
10     SELECT MAX(punt_global)
11     INTO max_puntaje
12     FROM puntajes;
13
14 -- Mostrar el valor maximo en pantalla
15     DBMS_OUTPUT.PUT_LINE('El valor maximo de puntaje es: ' || max_puntaje);
16 END;
17 /
18 -- Habilitar la salida del servidor para ver los mensajes del trigger
19 SET SERVEROUTPUT ON;
20

```



```

21 -- Insertar un dato de prueba
22 INSERT INTO puntajes (id_puntaje, punt_global)
23 VALUES (6000, 95);
24
25 -- Eliminar el dato de prueba
26 DELETE FROM puntajes
27 WHERE id_puntaje = 6000 AND punt_global = 95;

```

```

Trigger MOSTRAR_MAXIMO_PUNTAJE compiled

El valor máximo de puntaje es: 416

```

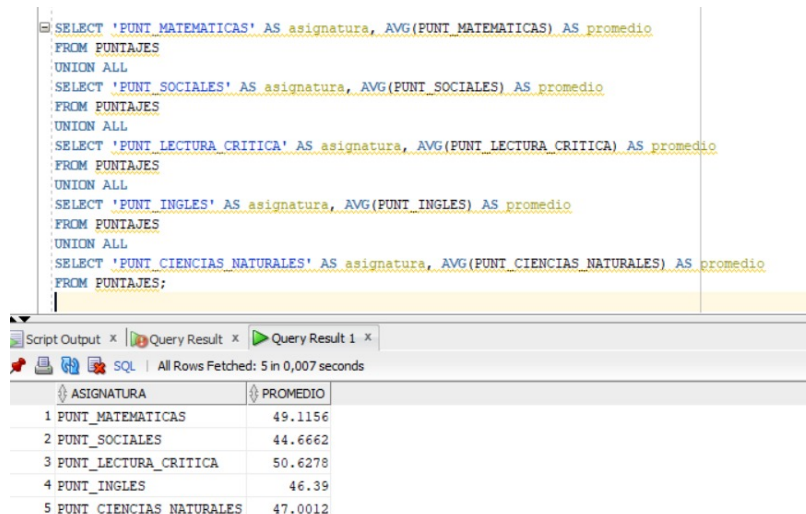
## 4.8 Código SQL + Resultados: Función 1

Listing 13: Funciones

```

1 -- Funciones
2
3 -- Este SQL calcula el promedio de las asignaturas en la tabla PUNTAJES
4
5 -- Cada subconsulta calcula el promedio
6 de una asignatura especifica
7 -- y le asigna un nombre de columna "asignatura" y "promedio".
8
9 -- Luego, se utiliza UNION ALL para combinar los resultados de todas las
10 -- subconsultas en un solo conjunto de resultados.
11
12 SELECT 'PUNT_MATEMATICAS' AS asignatura, AVG(PUNT_MATEMATICAS) AS promedio
13 FROM PUNTAJES
14 UNION ALL
15 SELECT 'PUNT_SOCIALES_CIUDADANAS' AS asignatura, AVG(PUNT_SOCIALES_CIUDADANAS) AS promedio
16 FROM PUNTAJES
17 UNION ALL
18 SELECT 'PUNT_LECTURA_CRITICA' AS asignatura, AVG(PUNT_LECTURA_CRITICA) AS promedio
19 FROM PUNTAJES
20 UNION ALL
21 SELECT 'PUNT_INGLES' AS asignatura, AVG(PUNT_INGLES) AS promedio
22 FROM PUNTAJES
23 UNION ALL
24 SELECT 'PUNT_C_NATURALES' AS asignatura, AVG(PUNT_C_NATURALES) AS promedio
25 FROM PUNTAJES;

```



ASIGNATURA	PROMEDIO
1 PUNT_MATEMATICAS	49.1156
2 PUNT_SOCIALES	44.6662
3 PUNT_LECTURA_CRITICA	50.6278
4 PUNT_INGLES	46.39
5 PUNT_CIENCIAS_NATURALES	47.0012

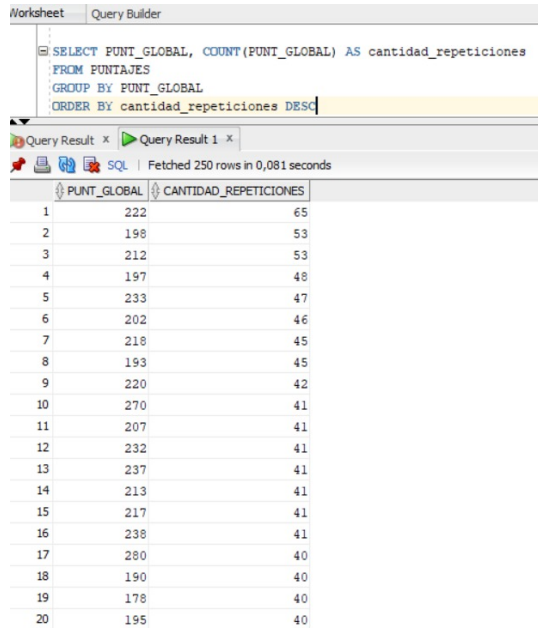


Listing 14: Función 2

```

1 SELECT PUNT_GLOBAL, COUNT(PUNT_GLOBAL) AS cantidad_repeticiones
2 FROM PUNTAJES
3 GROUP BY PUNT_GLOBAL
4 ORDER BY cantidad_repeticiones DESC

```



PUNT_GLOBAL	CANTIDAD_REPETICIONES
1	222
2	198
3	212
4	197
5	233
6	202
7	218
8	193
9	220
10	270
11	207
12	232
13	237
14	213
15	217
16	238
17	280
18	190
19	178
20	195

## 4.9 Código SQL + Resultados: Procedimientos Almacenados

Listing 15: Procedimientos Almacenados

```

1 CREATE OR REPLACE PROCEDURE ver_estudiante(p_estu_consecutivo_sb VARCHAR2) AS
2     v_estu_genero VARCHAR2(255);
3     v_estu_tipo_documento VARCHAR2(255);
4     v_estu_nacionalidad VARCHAR2(255);
5 BEGIN
6     SELECT estu_genero, estu_tipo_documento, estu_nacionalidad
7     INTO v_estu_genero, v_estu_tipo_documento, v_estu_nacionalidad
8     FROM estudiante
9     WHERE estu_consecutivo_sb = p_estu_consecutivo_sb;
10
11     -- Aqui simplemente mostramos la informaci n. Puedes adaptar para mostrar
12     en tu aplicaci n o plataforma.
13     DBMS_OUTPUT.PUT_LINE('Genero: ' || v_estu_genero);
14     DBMS_OUTPUT.PUT_LINE('Tipo Documento: ' || v_estu_tipo_documento);
15     DBMS_OUTPUT.PUT_LINE('Nacionalidad: ' || v_estu_nacionalidad);
16 EXCEPTION                                WHEN NO_DATA_FOUND THEN
17     DBMS_OUTPUT.PUT_LINE('Estudiante no encontrado.');
```

```

18 WHEN OTHERS THEN
19     DBMS_OUTPUT.PUT_LINE('Error al obtener informaci n del estudiante.');
```

```

20 END ver_estudiante;
21 /
22
23 -- Ejemplo para visualizar informaci n del estudiante con el c digo 'E123456':
24 BEGIN
25     ver_estudiante('E123456');
```

```

26 END;
27 /

```

ESTU_GENERO				
-----				
PUNTAJES_ID_PUNTAJE	NIVEL_SOCIOECONOMICO_INSE	INDIVIDUAL	COLEGIO_COLE_CODIGO	DEPARTAMENTO_CODIGO_DEP
-----				
m	0	50.9421557	1.4771E+11	47

Listing 16: Procedimientos Almacenados

```

1 CREATE OR REPLACE PROCEDURE contar_estudiantes(p_total OUT NUMBER) IS
2 BEGIN
3     SELECT COUNT(*)
4     INTO p_total
5     FROM estudiante;
6 EXCEPTION
7     WHEN OTHERS THEN
8         RAISE; -- Re-lanza cualquier error que pueda surgir
9 END contar_estudiantes;
10 /

```

```

Procedure CONTAR_ESTUDIANTES compiled

```

```

Total de estudiantes: 4394

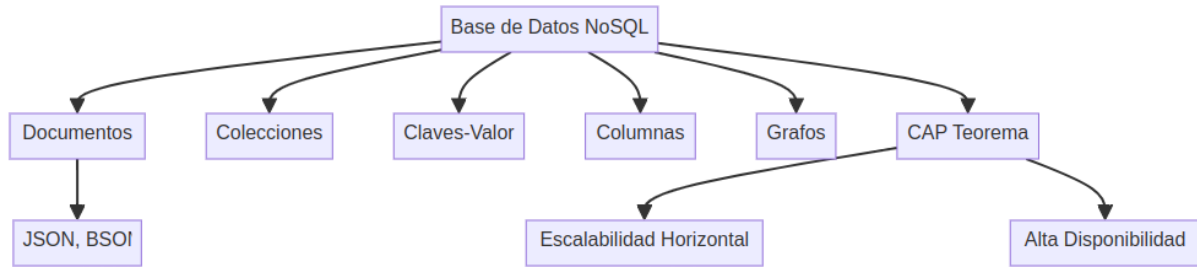
```

## 5 Bases de Datos No-SQL

Conforme la tecnología y las demandas de gestión de información evolucionan, emergen soluciones innovadoras en el ámbito del almacenamiento y procesamiento de datos. Una de estas soluciones es NoSQL, que, según Cáceres et al. (s.f.), ha emergido como una respuesta a las limitaciones de las bases de datos tradicionales, especialmente al enfrentar datos de gran volumen y de naturaleza semántica.

A diferencia de las bases de datos convencionales que se enfocan en estructuras tabulares y relaciones preestablecidas, NoSQL proporciona una diversidad de modelos de datos. Morales y Geovany (2017) resaltan que esta adaptabilidad es esencial para el desarrollo de aplicaciones web contemporáneas. Además, Navas Moya, Soria, Viteri Arias, y Casa Guayta (2018) enfatizan que NoSQL no solo ofrece flexibilidad estructural, sino que también es eficiente en entornos que manejan grandes volúmenes de datos.

Dado esta investigación, es crucial comprender las características y ventajas de NoSQL para maximizar el potencial, incrustando este tipo de tecnología.



### 5.1 Diagrama Bases de Datos No-SQL

Para llevar a cabo la implementación del proyecto, se optó por un enfoque NoSQL, ya que este brinda una mayor flexibilidad en la estructura de la base de datos. Para comenzar con el diseño de la estructura, se ha realizado tanto el modelado conceptual como el lógico de manera conjunta. En este proceso, la intención es definir de manera clara y precisa cómo se relacionan los elementos de datos y cómo se organizará la información en la base de datos, permitiendo comprender y representar eficazmente la lógica y la estructura de los datos de acuerdo con los objetivos del proyecto.

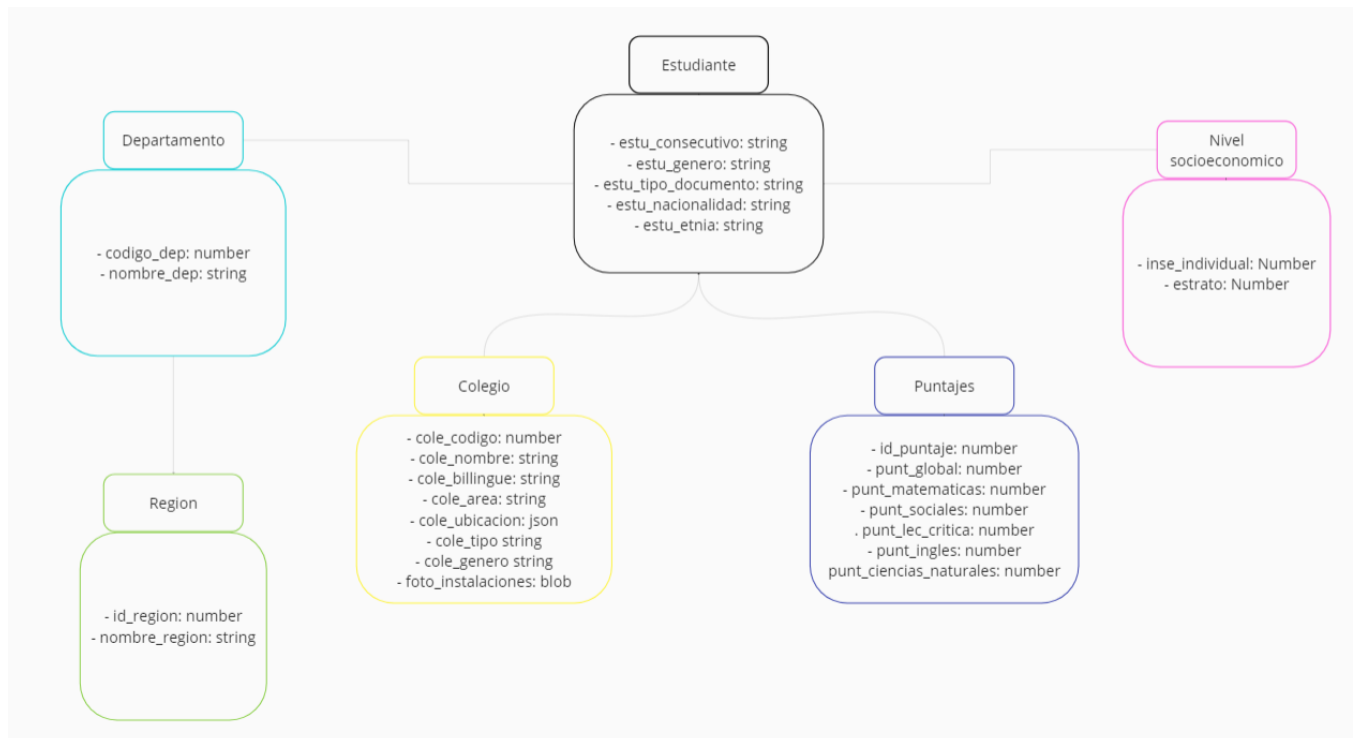


Figure 8: Modelo conceptual y lógico

Luego, para el modelo físico, se llevó a cabo la implementación de cada una de las entidades definidas anteriormente y se ingresaron datos en cada una de estas entidades. Es importante destacar que se agregaron algunos atributos adicionales en ciertas entidades, aprovechando la mayor flexibilidad que ofrece este enfoque.

## Icfes\_2019-2.Colegio

STORAGE SIZE: 36KB   LOGICAL DATA SIZE: 960B   TOTAL DOCUMENTS: 2   INDEXES TOTAL SIZE: 36KB

**Find**   Indexes   Schema Anti-Patterns 0   Aggregation   Search Indexes

**INSERT DOCUMENT**

Filter [🔗](#)   Type a query: { field: 'value' }   **Reset**   **Apply**   [More Options ▶](#)

```

_id: ObjectId('654159fb1db0a5db046d7572')
cole_codi... : "123575001000"
cole_nomb... : "Sagrados corazones"
cole_biling... : BinData(0, 'MA==')
cole_ar... : "Urbana"
cole_ubicaci... : "4°35'45"N 74°04'06"O / 4.5958, -74.0684"
cole_ti... : "No oficial"
cole_gene... : "Mixto"
foto_instalacion... : BinData(0, 'aHR0cHM6Ly93d3cuZ29vZ2x1LmNvbS91cmw/c2E9aSZ1cmw9aHR0cCUzQSUyRiUyRndpa2ltYXBpYS...'

```

```

_id: ObjectId('65415be21db0a5db046d7573')
cole_codi... : 125450001810
cole_nomb... : "Sede la revuelta"
cole_biling... : BinData(0, 'MQ==')
cole_ar... : "Rural"
cole_ti... : "Oficial"
foto_instalaci... : BinData(0, 'aHR0cHM6Ly93d3cuZ29vZ2x1LmNvbS91cmw/c2E9aSZ1cmw9aHR0cHM0E1MkY1MkZ3d3cuYWRTYWdh...'

```

Figure 9: Modelo Físico: Implementación MongoDB de Tabla Colegio

## Icfes\_2019-2.Departamentos

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 219B TOTAL DOCUMENTS: 4 INDEXES TOTAL SIZE: 36KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

More Options

```
_id: ObjectId('654157d41db0a5db046d756a')
codigo_dep: 5
nombre_dep: "Cundinamarca"
```

```
_id: ObjectId('6541583a1db0a5db046d756b')
codigo_dep: "10"
nombre_dep: "Bolívar"
```

```
_id: ObjectId('654158a51db0a5db046d756c')
nombre_dep: "Popayan"
```

```
_id: ObjectId('654158b51db0a5db046d756d')
codigo_dep: "20"
```

Figure 10: Modelo Físico: Implementación MongoDB de Tabla departamentos

## Icfes\_2019-2.Estudiante

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 431B TOTAL DOCUMENTS: 3 INDEXES TOTAL SIZE: 36KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

More Options



```
_id: ObjectId('654147321db0a5db046d7567')
estu_consecuti... : "SB11201940031558"
estu_gene... : "M"
estu_tipo_documen... : "CC"
estu_nacionalid... : "Colombiana"
estu_etnia: "Ninguno"
```



```
_id: ObjectId('654148b11db0a5db046d7568')
estu_consecuti... : "SB11201940303339"
estu_gene... : "F"
estu_tipo_documen... : "TI"
estu_nacionalid... : "Venezolana"
```

```
_id: ObjectId('654149261db0a5db046d7569')
estu_consecuti... : "SB11201940255017"
estu_tipo_documen... : "CC"
estu_nacionalid... : "Colombiana"
```

Figure 11: Modelo Físico: Implementación MongoDB de Tabla estudiante

## Icfes\_2019-2.Nivel Socioeconomico

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 244B TOTAL DOCUMENTS: 4 INDEXES TOTAL SIZE: 36KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

More Options

```
_id: ObjectId('65415c861db0a5db046d7574')
inse_individual: "66827003496001"
estra... : "4"
```

```
_id: ObjectId('65415cd71db0a5db046d7575')
estra... : "6"
```

```
_id: ObjectId('65415ce21db0a5db046d7576')
inse_individual: "285166360762131"
```

```
_id: ObjectId('65415cff1db0a5db046d7577')
inse_individual: "40.8769032514053"
estra... : "1"
```

Figure 12: Modelo Físico: Implementación MongoDB de Tabla Nivel socioeconómico

## lcfes\_2019-2.Puntaje

STORAGE SIZE: 4KB LOGICAL DATA SIZE: 0B TOTAL DOCUMENTS: 0 INDEXES TOTAL SIZE: 4KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

More Options

```
_id: ObjectId('65415d871db0a5db046d7578')
id_puntaje: 266
punt_glob... : 240
punt_matematic... : 52
punt_social... : 65
punt_lec_criti... : 31
punt_ingl... : 74
punt_ciencias_natural... : 21
```

```
_id: ObjectId('65415e8c1db0a5db046d7579')
id_puntaje: "300"
punt_glob... : "234"
punt_social... : "43"
punt_ingl... : "42"
```

Figure 13: Modelo Físico: Implementación MongoDB de Tabla puntaje

## 5.2 SMDB utilizado para la Base de Datos No-SQL

MongoDB fue elegido para el modelado NoSQL debido a varias ventajas. En primer lugar, una de las razones clave es su flexibilidad, ya que permite el almacenamiento de datos en formato BSON, lo que facilita la representación de estructuras de datos complejas. Además, su capacidad de escalar horizontalmente es ideal para aplicaciones que necesitan manejar grandes volúmenes de datos y un alto rendimiento.

Otra ventaja es su capacidad de indexación, que acelera las consultas y la recuperación de datos. MongoDB también soporta la replicación y la fragmentación, lo que garantiza la disponibilidad y la tolerancia a fallos. Además, su modelo de documentos JSON-like facilita la interoperabilidad con lenguajes de programación populares.

El hecho de que MongoDB ofrezca una versión gratuita junto con su licencia paga la convierte en una opción asequible para implementaciones no comerciales. En resumen, MongoDB es adecuado para el modelado NoSQL debido a su flexibilidad, escalabilidad, capacidades de indexación, y su capacidad para gestionar grandes conjuntos de datos, todo a un costo accesible.



## 6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos

### 6.1 Ejemplo de aplicación de ETL y Bodega de Datos

#### 6.1.1 Explicación del Código

1. Extracción de datos: El código utiliza la biblioteca 'csv' para leer el contenido de un archivo CSV. Además utiliza un 'from google.colab import files' ya que este código se realizó en Google Colab.
2. Transformación de datos: Durante la lectura de archivo, se realizan transformaciones, y se verifica la existencia de duplicados.
3. Carga de datos: Los datos transformados se escriben en un nuevo archivo CSV, el cual se guarda con el nombre de 'datos<sub>ETL</sub>.csv'. El archivo se descarga automáticamente para su posterior uso.

Listing 17: ETL

```
import csv
from google.colab import files

def realizar_etl(archivo_entrada, archivo_salida):
    datos_transformados = set() # Conjunto para almacenar los registros únicos
    datos_duplicados = set() # Conjunto para almacenar los registros duplicados

    # Extracción y transformación de datos
    with open(archivo_entrada, 'r') as archivo_csv:
        lector = csv.reader(archivo_csv, delimiter='\t')
        # Cambia el delimitador a '\t' para datos tabulares
        next(lector) # Ignorar la primera fila de encabezados
        for fila in lector:
            # Realizar las transformaciones necesarias aquí, si es necesario
            # Consideramos toda la fila como un conjunto de datos único
            registro = tuple(fila)

            # Verificar si el registro ya existe en el conjunto de datos transformados
            if registro in datos_transformados:
                datos_duplicados.add(registro)
            else:
                datos_transformados.add(registro)

    # Imprimir la cantidad de registros limpiados
    registros_limpiados = len(datos_transformados)
    print("Registros limpiados:", registros_limpiados)

    # Carga de datos en el archivo de salida
    with open(archivo_salida, 'w', newline='') as archivo_csv:
        escritor = csv.writer(archivo_csv, delimiter='\t')
        # Cambia el delimitador a '\t' para datos tabulares
        for registro in datos_transformados:
            escritor.writerow(registro)

    # Descargar el archivo CSV resultante
    files.download(archivo_salida)

    # Imprimir los registros duplicados
    print("Registros duplicados encontrados:")
    for registro in datos_duplicados:
        print(registro)

# Solicitar al usuario que cargue un archivo de entrada
print("Por favor, seleccione el archivo de entrada:")
```

```

archivo_entrada = files.upload()

# Obtener el nombre del archivo cargado
nombre_archivo = list(archivo_entrada.keys())[0]

# Nombre del archivo de salida
archivo_salida = 'datos_ETL.csv'

# Realizar el proceso ETL
realizar_etl(nombre_archivo, archivo_salida)

    Luego de correr el código, se encontró que hubo una limpieza de 13270 registros.

```

## 6.2 Automatización de Datos

En el código anterior se implementó la automatización del código, esto presenta una mejora significativa en el proceso ETL (Extracción, Transformación y Carga) de datos. Únicamente se requiere que el usuario ingrese el nombre del archivo de entrada, ya que el código genera automáticamente un nombre único para el archivo de salida, mejorando la eficiencia y reduciendo la posibilidad de duplicados.

En el caso de no tener automatización, el usuario tendría que seleccionar manualmente el archivo de entrada, ejecutar el código y gestionar la salida; lo cual limita la eficiencia y aumentando el riesgo de errores.

## 6.3 Integración de Datos

En esta etapa, los resultados del proceso ETL se consolidan en un archivo limpio y organizado. Este archivo, generado automáticamente cuyo nombre es 'datos\_ETL', contiene registros únicos y debido a la eliminación de duplicados se garantiza la fiabilidad de la base de datos resultante para impulsar la toma de decisiones y proyectos futuros. Por ende, la integración eficiente de estos datos limpios facilita su utilización en diversos contextos, destacando la calidad y accesibilidad que ofrece este proceso.

## 7 Próximos pasos (*Tercera entrega*)

Seguido a todo el proceso DDL, DML, la implementación a una tecnología No-SQL y por ultimo ETL (Extracción, Transformación y Carga) se necesita un EDA (Análisis Exploratorio de Datos), para este proyecto se hara la conexión con python para hacer dicho análisis. Teniendo esto en cuenta es pertinente obtener unos buenos descriptivos de cada variable, esto es información valiosa respecto a cómo se están comportando los datos. De acuerdo a este análisis univariado y teniendo un objetivo claro desde un inicio se toma la decisión de selección de modelos, para este caso se optó por un modelo de regresión lineal, previamente realizando un análisis univariado y bivariado se considera indispensable para este tipo de modelos una robusta selección e ingeniería de características, por consiguiente se utilizan las distintas métricas para la evalaución de resultados predictivos del modelo.

## 8 Lecciones aprendidas *(Tercera entrega)*

Durante el desarrollo de este proyecto de base de datos, se obtuvieron conocimientos significativos en diversas áreas.

En el ámbito de los Sistemas de Gestión de Bases de Datos (SGBD), especialmente en Oracle, se amplió el entendimiento del diseño de modelos lógicos y relacionales. Se aclararon dudas respecto a las conexiones entre entidades mediante el uso de claves foráneas. Además, se fortaleció la habilidad en la creación de códigos DDL (Data Definition Language) y DML (Data Manipulation Language). Específicamente en DML, se reconoció la importancia de establecer un orden meticuloso al insertar datos en múltiples tablas para evitar conflictos en las conexiones.

En cuanto a vistas, triggers, funciones y procedimientos almacenados, se identificaron y resolvieron confusiones en su creación, lo cual permitió una mayor comprensión sobre su uso y estructuración en proyectos futuros.

En el ámbito de las bases de datos NoSQL, la exploración en la plataforma MongoDB brindó un entendimiento más profundo de su funcionamiento y la manera de plantear diferentes modelos (conceptual, lógico y físico). Se reconoció la flexibilidad que estos modelos ofrecen, permitiendo un escalado más rápido y eficiente.

Por último, en el proceso de ETL (Extracción, Transformación y Carga), se comprendió la importancia de cada uno de estos pasos y la interdependencia entre ellos. Esta comprensión se traduce en la capacidad para realizar trabajos más efectivos, generando resultados superiores en futuros proyectos.

Estos aprendizajes contribuyen significativamente a la mejora de la planificación, ejecución y manejo de futuros proyectos de bases de datos, proporcionando una base sólida para abordar desafíos más complejos con mayor eficacia y conocimiento.

## 9 Referencias Bibliográficas

### References

- [1] Agudelo, A. S., Figueroa, L. A., & Vasquez, L. (2019). Relaciones causales de los factores que afectan el desempeño de los estudiantes en pruebas estandarizadas en Colombia. *Revista espacios*, 40(23).
- [2] Arenas, A., & Submitter, CIEF. (2021). Mejores colegios en Colombia: efecto de las condiciones socioeconómicas sobre el desempeño escolar (Best Schools in Colombia: The Effect of Socioeconomic Conditions on School Performance).
- [3] Bogoya, D. (2006). Evaluación educativa en Colombia. *Seminario internacional de evaluación*, 1, 27.
- [4] Castro Aristizábal, G., Díaz Rosero, M., & Tobar Bedoya, J. (2019). ¿Por qué los colegios privados en Colombia obtienen mejores resultados académicos?
- [5] Caicedo, E. J. Cuéllar, Gómez, S. Y. Guerrero, & López, D. (2016). Propuesta para la construcción de un índice socioeconómico para los estudiantes que presentan las pruebas Saber Pro. *\*Comunicaciones en Estadística*, 9\*(1), 93-106. Universidad Santo Tomás (USTA).
- [6] Chica Gómez, S., Galvis Gutiérrez, D., & Ramirez Hassan, A. (2011). Determinantes del rendimiento académico en Colombia: Pruebas ICFES saber 11, 2009 (academic performance determinants in Colombia: ICFES saber 11, 2009 exam). *Center for Research in Economics and Finance (CIEF), Working Papers*, (11-5).
- [7] Cáceres, N. R., Tolaba, C., Pérez, R. D., Quispe, J. J. M., Rodríguez, C. S., & Sandoval, I. L. (s.f). Benchmarking de bases de datos NoSQL para el almacenamiento de modelos semánticos.
- [8] Gómez, S. M. C., Gutiérrez, D. M. G., & Hassan, A. R. (2010). Academic performance determinants in Colombia: the ICFES tests. *Revista Universidad EAFIT*, 46(160), 48-72.
- [9] Mejía Rojas, D. P. (2017). Correlación entre los resultados de las pruebas Icfes SABER 11° y el rendimiento académico de los estudiantes de Ingeniería y Tecnología que cursan Matemáticas I y Matemáticas I Articulación Primera Asignatura, en el II semestre de 2015 y I semestre de 2016 de la Universidad Tecnológica de Pereira.
- [10] Moncayo Cabrera, M. A. (2016). Determinantes que influyen en el rendimiento académico: un estudio aplicado para Colombia a partir de las pruebas ICFES-SABER 11.
- [11] Morales, I., & Geovany, M. (2017). Estudio de nuevas tecnologías de gestión de bases de datos NOSQL para el desarrollo de aplicaciones web 2.0.
- [12] Moreno, F. J., Ospina Romero, G., & Larios Restrepo, R. (2005). Desempeño de consultas SQL relacionales y objeto-relacionales en Oracle. *\*Ingeniería e Investigación*, 25\*(3), 4-12. Facultad de Ingeniería, Universidad Nacional de Colombia.
- [13] Muñoz, L., Mazon, J.-N., & Trujillo, J. (2010). Systematic review and comparison of modeling ETL processes in data warehouse. En *\*5th Iberian Conference on Information Systems and Technologies\** (pp. 1-6). IEEE.
- [14] Navas Moya, P., Soria, T., Viteri Arias, S., & Casa Guayta, C. (2018). Análisis de impacto y medición de confiabilidad y tiempo en la migración de bases de datos Sql a Nosql.
- [15] Orjuela, J. (2013). Análisis del desempeño estudiantil en las pruebas de Estado para educación media en Colombia mediante modelos jerárquicos lineales. *Ingeniería*, 18(2), 6. Universidad Distrital Francisco José de Caldas.