

Análisis estadístico para apoyar la toma de decisiones en las políticas públicas educativas con base en los resultados de las pruebas saber 11 (2019-2, 2020-1).

Oscar Alberto Castrillón¹, Diego Armando Gamba², Hector Estiven Rodriguez³

¹⁻³Facultad de matemáticas e ingeniería

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

{¹ocastrillonb@ucentral.edu.co,²dgambah@ucentral.edu.co}, ³hrodriguez4@ucentral.edu.co

November 19, 2022

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	5
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>) . .	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	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? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(<i>Primera entrega</i>)	6
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (<i>Primera entrega</i>)	6
4.2	Diagrama modelo de datos (<i>Primera entrega</i>)	6
4.3	Imágenes de la Base de Datos (<i>Primera entrega</i>)	7
4.4	Código SQL - lenguaje de definición de datos (DDL) (<i>Primera entrega</i>)	8

4.5	Código SQL - Manipulación de datos (DML) (<i>Primera entrega</i>)	9
4.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	9
4.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	11
4.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	12
4.9	Código SQL + Resultados: procedimientos almacenados (<i>Primera entrega</i>)	12
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	14
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	14
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	18
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	19
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	19
7	Lecciones aprendidas (<i>Tercera entrega</i>)	20
8	Bibliografía	21

1 Introducción (Max 250 Palabras) - (Primera entrega)

En la actualidad un tema primordial que está estrictamente relacionado con las futuras posibilidades de acceder a la educación superior que tienen los estudiantes próximos a salir de la educación media son los resultados del examen de estado (Icfes-saber 11), realizados en 2 cortes anuales. Estos se encuentran estipulados como requisito para poder acceder al grado de bachiller académico, no obstante, el trasfondo en la utilización de estas pruebas es descrito a través de los objetivos buscados por el ministerio de educación el cual va en relación con 3 aspectos fundamentales(Decreto N° 869, 17 de Marzo 2010): el primero va en caminado hacia la comprobación de las competencias adquiridas por los estudiantes y a su vez, que los mismos tengan elementos para la realización de su autoevaluación para el correcto desarrollo de su proyecto de vida. El segundo aspecto hace referencia a brindarle a las instituciones de educación media información relevante con base en los resultados obtenidos de los estudiantes para lograr una consolidación y/o reorientación de sus prácticas pedagógicas, además de ello también ofrece información para las instituciones de educación superior (IES) sobre el nivel de desarrollo de las competencias que tiene las personas que deseen ingresar a uno de sus programas ofertados. Por último, el tercer aspecto que tiene en cuenta el ministerio de educación para la realización de las pruebas es el poder contar con insumos relevantes que sirvan como marco estratégico para la generación de políticas educativas, entre las cuales se encuentran beneficios como becas y/o créditos condenables para los mejores puntajes.

En busca de mejorar la toma de decisiones con base en los anteriores objetivos planteados por el ministerio de educación surge el actual proyecto de investigación, donde se realiza un análisis estadístico el cual pretende orientar los esfuerzos que se lleven a cabo en las políticas públicas educativas, las cuales den respuesta a las necesidades encontradas en concordancia con en los resultados obtenidos en las pruebas saber 11.

2 Características del proyecto de investigación (Max 500 Palabras) - (Primera entrega)

2.1 Título del proyecto de investigación (Max 100 Palabras) - (Primera entrega)

Análisis estadístico para apoyar la toma de decisiones en las políticas públicas educativas con base en los resultados de las pruebas saber 11 (2019-2 y 2020-1).

2.2 Objetivo general (Max 100 Palabras) - (Primera entrega)

Analizar las causas y efectos de los resultados en las pruebas saber 11 del segundo semestre del 2019 y primer semestre del 2020 en concordancia con las políticas públicas educativas establecidas, a través de un análisis estadístico el cual vaya orientado a servir como insumo para apoyar la toma de decisiones informada.

2.2.1 Objetivos específicos (Max 100 Palabras) - (Primera entrega)

- Identificar que variables se relacionan directa e indirectamente con el puntaje obtenido en las pruebas saber 11.
- Determinar cuál es la relación existente entre los estudiantes que sacaron mejores puntajes.
- Aplicar la regresión logística para determinar que estudiantes aplican al programa de generación E.
- Analizar como las políticas públicas educativas pueden mejorar el rendimiento de los estudiantes con base en los factores más relevantes encontrados en los análisis estadísticos.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

De acuerdo con Hernández et al. (2014), el proyecto de investigación puede darse de manera continua optando por los 4 alcances que delimitan las estrategias de investigación y los procedimientos a utilizar para su correcta ejecución. Estos 4 alcances son:

- Exploratorio: indaga sobre temas en específicos con base en teorías encontradas, su utilización suele darse en base a temas o problemas de investigación poco estudiados.
- Descriptivo: su objetivo va en caminado a recoger información de cada una de las variables de manera independiente, sin llegar a analizar como se correlacionan entre ellas.
- Correlacional: va en búsqueda de la relación existente entre 2 o más variables de la muestra seleccionada.
- Explicativo: como su nombre lo indica, va encaminado a explicar que conexión puedan tener las variables y el porqué de dicha relación, analizando las causalidades de los temas estudiados.

El presente proyecto de investigación es realizado hasta el alcance explicativo, donde no solo se describe cada una de las variables, sino también se buscará una relación idónea de los resultados de las pruebas saber 11 con las demás variables a través del análisis estadístico, analizando que relación presenta y las causas y efectos que puedan llegar a tener.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cuáles factores son los más relevantes en el rendimiento de los estudiantes en las pruebas saber 11 realizadas en el segundo semestre del 2019 y primer semestre del 2020 para la toma de decisiones en las políticas públicas educativas?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

Los principales aspectos que pueden inferir en el rendimiento de los estudiantes estarían directamente relacionados con el estrato socioeconómico, región del país, tipo de colegio, jornada, internet disponible y cantidad de integrantes por familia

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (Primera entrega)

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (Primera entrega)

Los datos utilizados para este proyecto provienen de las bases de datos de la página de datos abiertos de Colombia, información aportada por el instituto colombiano para la evaluación de la educación (ICFES), ubicado en Bogotá. Consta de 82 columnas y 546.212 filas, que contiene información sobre el estudiante en cuanto a sus datos personales, información socioeconómica del contexto del estudiante, información sobre la institución educativa y por supuesto los resultados obtenidos en su prueba de estado, de los diferentes departamentos de Colombia. Además, una columna sobre pertenencia o no al programa del gobierno llamado generación e.

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (Primera entrega)

El manejo de los datos esta regulado por la ley 1712 del año 2014, en donde se reglamenta el acceso a la información pública y se le obliga a las organizaciones a permitir el acceso de manera libre a los datos, asimismo la ley 1581 2012 y en la constitución política en el artículo 15 en donde se hace referencia a la naturaleza de los datos y para este caso, se al manejo de los datos públicos, lo que permite poder trabajar sobre la base datos y hacer el respectivo proceso de analítica.

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? (Max 100 Palabras) - (Primera entrega)

Las bases son demasiado robustas, cuenta con una gran cantidad de datos que aportan información valiosa que contribuye al objetivo del proyecto, por lo que los principales retos esta en empezar a seleccionar los datos de la base principal y empezar a generar las bases secundarias, asimismo, generar las conexiones y las relaciones teniendo en cuenta las llaves y las variables que contiene cada una las bases que se van a construir y que no quede información cruzada, para tener una buena relación de las variables con las bases de datos.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (Primera entrega)

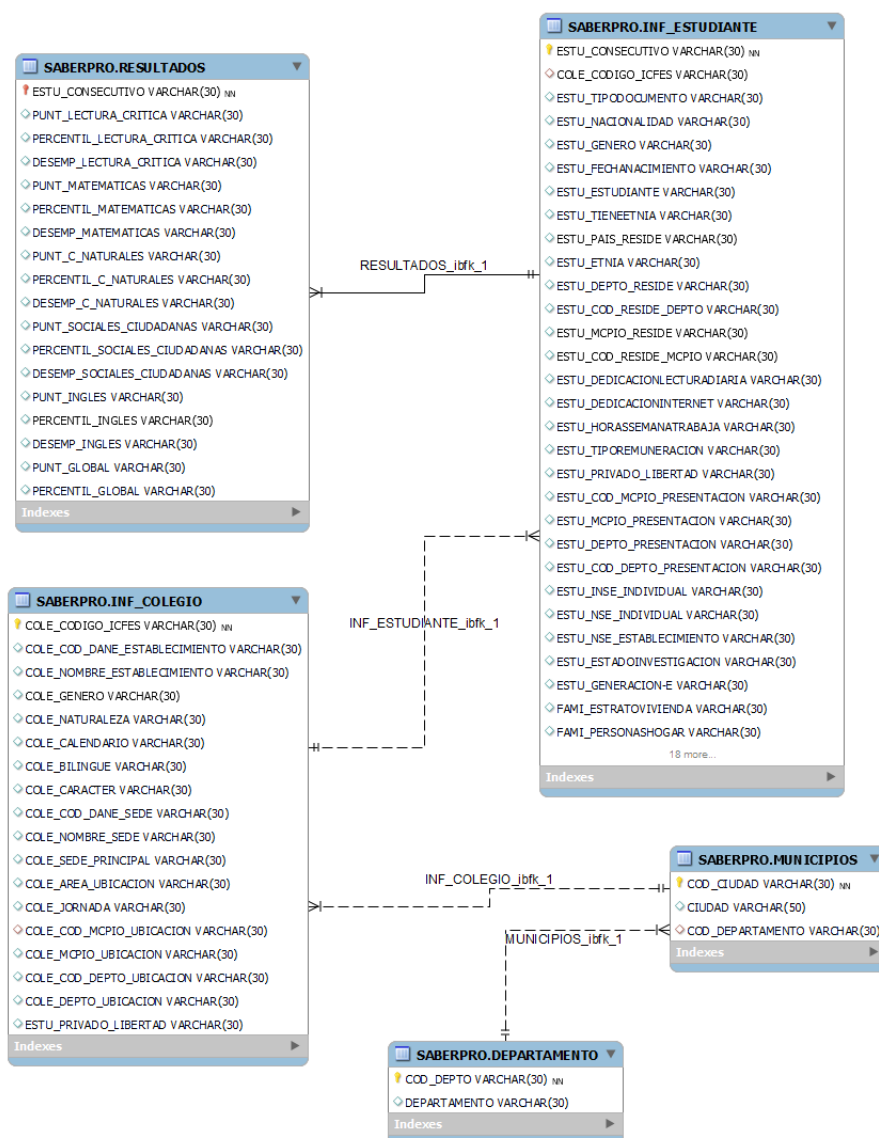
Tener una organización de la información por medio de DBMS, permitirá, definir un esquema conceptual teniendo en cuenta las diferentes variables (pues se tiene una base de datos bastante robusta), con los que cuenta la base y de esta forma poder recuperar, insertar, modificar, eliminar y acceder a la información siguiendo una ruta de trabajo dependiendo de la base que se requiera, asimismo permitirá administrar la información según los requerimientos y necesidades del proceso de análisis que se esté llevando, teniendo en cuenta las relaciones establecidas entre las bases.

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

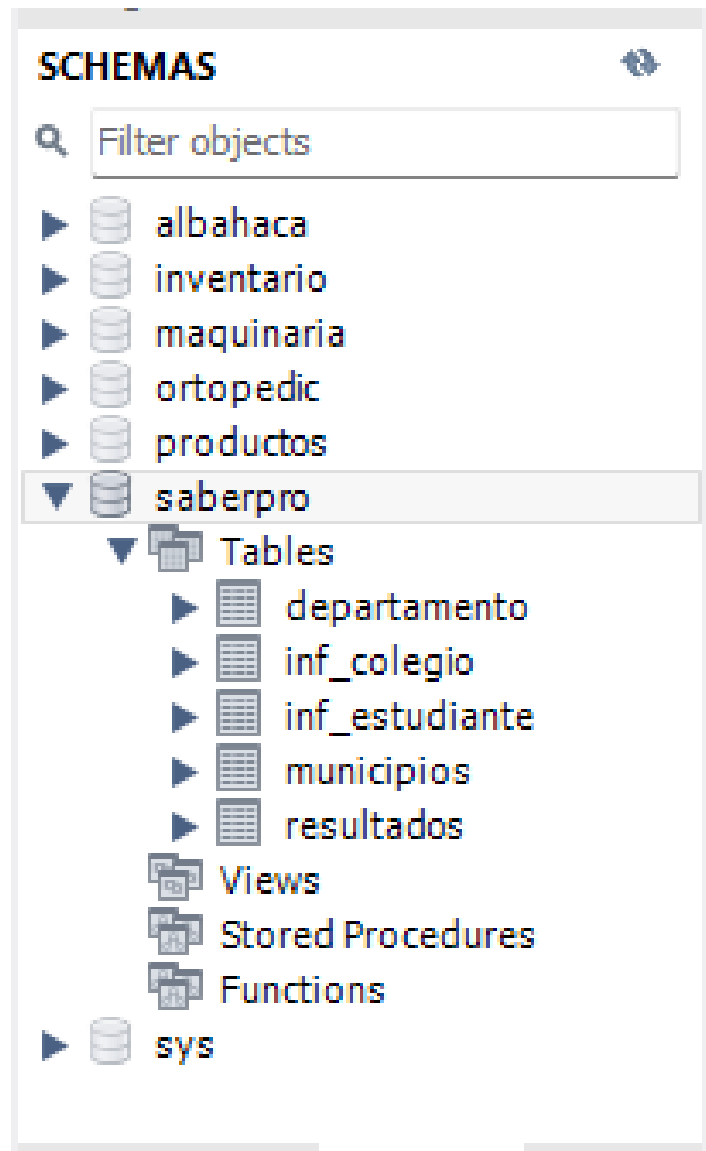
4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (Primera entrega)

El Sistema Manejador de Base de Datos que se utiliza en este proyecto es MySQL Workbench 8.0, es una herramienta para trabajar con bases de datos relacionales, que permite trabajar con los datos de una manera organizada, proporcionando características de modelamiento de datos, seguridad y administración de usuarios y manipulación de datos a través de lenguaje SQL. Esta disponible en sistemas operativos Windows, Linux y Mac OS X.

4.2 Diagrama modelo de datos (Primera entrega)



4.3 Imágenes de la Base de Datos (*Primera entrega*)



4.4 Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)

```
CREATE SCHEMA SABERPRO;

USE SABERPRO;

CREATE TABLE `DEPARTAMENTO` (
  `COD_DEPTO` NVARCHAR(30),
  `DEPARTAMENTO` NVARCHAR(30),
  PRIMARY KEY (`COD_DEPTO`)
);

CREATE TABLE `MUNICIPIOS` (
  `COD_CIUADAD` NVARCHAR(30),
  `CIUADAD` NVARCHAR(50),
  `COD_DEPARTAMENTO` NVARCHAR(30),
  PRIMARY KEY (`COD_CIUADAD`),
  FOREIGN KEY (`COD_DEPARTAMENTO`) REFERENCES `DEPARTAMENTO` (`COD_DEPTO`)
);

CREATE TABLE `INF_COLEGIO` (
  `COLE_CODIGO_ICFES` NVARCHAR(30),
  `COLE_COD_DANE_ESTABLECIMIENTO` NVARCHAR(30),
  `COLE_NOMBRE_ESTABLECIMIENTO` NVARCHAR(30),
  `COLE_GENERO` NVARCHAR(30),
  `COLE_NATURALEZA` NVARCHAR(30),
  `COLE_CALENDARIO` NVARCHAR(30),
  `COLE_BILINGUE` NVARCHAR(30),
  `COLE_CARACTER` NVARCHAR(30),
  `COLE_COD_DANE_SEDE` NVARCHAR(30),
  `COLE_NOMBRE_SEDE` NVARCHAR(30),
  `COLE_SEDE_PRINCIPAL` NVARCHAR(30),
  `COLE_AREA_UBICACION` NVARCHAR(30),
  `COLE_JORNADA` NVARCHAR(30),
  `COLE_COD_MCPIO_UBICACION` NVARCHAR(30),
  `COLE_MCPIO_UBICACION` NVARCHAR(30),
  `COLE_COD_DEPTO_UBICACION` NVARCHAR(30),
  `COLE_DEPTO_UBICACION` NVARCHAR(30),
  `ESTU_PRIVADO_LIBERTAD` NVARCHAR(30),

  PRIMARY KEY (`COLE_CODIGO_ICFES`),
  FOREIGN KEY (`COLE_COD_MCPIO_UBICACION`) REFERENCES `MUNICIPIOS` (`COD_CIUADAD`)
);
```



```

CREATE TABLE `RESULTADOS` (
  `ESTU_CONSECUTIVO` NVARCHAR(30),
  `PUNT_Lectura_CRITICA` NVARCHAR(30),
  `PERCENTIL_Lectura_CRITICA` NVARCHAR(30),
  `DESEMP_Lectura_CRITICA` NVARCHAR(30),
  `PUNT_MATEMATICAS` NVARCHAR(30),
  `PERCENTIL_MATEMATICAS` NVARCHAR(30),
  `DESEMP_MATEMATICAS` NVARCHAR(30),
  `PUNT_C_NATURALES` NVARCHAR(30),
  `PERCENTIL_C_NATURALES` NVARCHAR(30),
  `DESEMP_C_NATURALES` NVARCHAR(30),
  `PUNT_SOCIALES_CIUDADANAS` NVARCHAR(30),
  `PERCENTIL_SOCIALES_CIUDADANAS` NVARCHAR(30),
  `DESEMP_SOCIALES_CIUDADANAS` NVARCHAR(30),
  `PUNT_INGLES` NVARCHAR(30),
  `PERCENTIL_INGLES` NVARCHAR(30),
  `DESEMP_INGLES` NVARCHAR(30),
  `PUNT_GLOBAL` NVARCHAR(30),
  `PERCENTIL_GLOBAL` NVARCHAR(30),

  PRIMARY KEY (`ESTU_CONSECUTIVO`),
  FOREIGN KEY (`ESTU_CONSECUTIVO`) REFERENCES `INF_ESTUDIANTE`(`ESTU_CONSECUTIVO`)
);

```

4.5 Código SQL - Manipulación de datos (DML) *(Primera entrega)*

4.6 Código SQL + Resultados: Vistas *(Primera entrega)*

```





CREATE VIEW `PROMEDIO_RESULTADOS_POR_MUNICIPIO` AS
SELECT
  AVG(PUNT_GLOBAL) AS PROMEDIO, M.CIUDAD AS Municipio
FROM
  resultados A
  INNER JOIN
  inf_estudiantes B ON B.ESTU_CONSECUTIVO = A.ESTU_CONSECUTIVO
  INNER JOIN
  municipios M ON B.ESTU_COD_MCPIO_PRESENTACION = M.COD_CIUDAD
GROUP BY M.CIUDAD

```

```

1 • SELECT * FROM promedio_resultados_por_municipio
2 ORDER BY 1 DESC

```

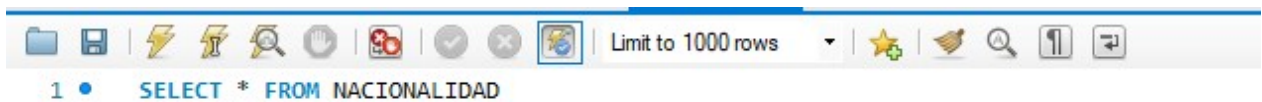
Result Grid   Filter Rows: <input type="text"/>			Export: 	Wrap Cell Content: 
	PROMEDIO	Municipio		
▶	348	PUERTO COLOMBIA		
	347	CALI		
	344	GARZÓN		
	334.5	OCAÑA		
	333	MIRANDA		
	329	RIONEGRO		
	324	SAN ANDRÉS		
	321	CUBARRAL		
	320.33333333333333	BUCARAMANGA		
	319	DUITAMA		
	318	ARMENIA		
	309	PAMPLONA		
	308	YOPAL		
	305	FUNZA		
	302	SANTA MARTA		
	300	SAHAGÚN		
	297	LA MESA		
	297	TUNJA		

promedio_resultados_por_muni... x

```

• CREATE VIEW `NACIONALIDAD` AS
  SELECT
    COUNT(ESTU_CONSECUTIVO) AS NoEstudiantes,
    ESTU_NACIONALIDAD AS Nacionalidad
  FROM
    saberpro.inf_estudiantes
  WHERE
    ESTU_NACIONALIDAD <> 'COLOMBIA';

```



Result Grid		Filter Rows:	Export:	Wrap Cell Content:
NoEstudiantes	Nacionalidad			
2	VENEZUELA			

4.7 Código SQL + Resultados: Triggers (Primera entrega)

```

DELIMITER //
CREATE TRIGGER HISTORICO_BORRADOS BEFORE DELETE ON inf_colegio
FOR EACH ROW
BEGIN
    INSERT INTO log_inf_colegio(COLE_CODIGO_ICFES, COLE_COD_DANE_ESTABLECIMIENTO, COLE_NOMBRE_ESTABLECIMIENTO )
    VALUES
    (old.COLE_CODIGO_ICFES, old.COLE_COD_DANE_ESTABLECIMIENTO, old.COLE_NOMBRE_ESTABLECIMIENTO);
END //
DELIMITER ;

```

```

1 • INSERT INTO inf_colegio (COLE_CODIGO_ICFES, COLE_COD_DANE_ESTABLECIMIENTO)
2   VALUES ('123456789', '123456789');
3
4 • SELECT * FROM inf_colegio WHERE COLE_CODIGO_ICFES='123456789';
5
6 • DELETE FROM inf_colegio WHERE COLE_CODIGO_ICFES='123456789';
7
8 • SELECT * FROM log_inf_colegio;

```

Result Grid						Filter Rows:	Edit:	Export/Import:	Wrap Cell Content:
	COLE_CODIGO_ICFES	COLE_COD_DANE_ESTABLECIMIENTO	COLE_NOMBRE_ESTABLECIMIENTO	COLE_GENERO	COLE_NATURALEZA				
▶	123456789	123456789	NULL	NULL	NULL				
*	NULL	NULL	NULL	NULL	NULL				

4.8 Código SQL + Resultados: Funciones (Primera entrega)

```
DELIMITER //
```

- **CREATE FUNCTION** devolverPuntaje(consecutivo **varchar**(250)) **RETURNS DECIMAL**(9,2)

```
BEGIN
    declare puntaje decimal(9,2);
    SELECT PUNT_GLOBAL
    INTO puntaje
    FROM saberpro.resultados
    WHERE ESTU_CONSECUTIVO = consecutivo;
    RETURN puntaje;
END //
```

```
DELIMITER ;
```



```
1 • select devolverPuntaje1(ESTU_CONSECUTIVO) from inf_estudiante where COLE_CODIGO_ICFES='108361';
2
3
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	devolverPuntaje1(ESTU_CONSECUTIVO)			
▶	321.00			

4.9 Código SQL + Resultados: procedimientos almacenados (Primera entrega)

```
DELIMITER $$
```

- **CREATE PROCEDURE** obtenerPromedioColegio(**IN** codigo_colegio **VARCHAR**(255))

```
BEGIN
    SELECT AVG(PUNT_GLOBAL) as Promedio,
    COLE_NOMBRE_ESTABLECIMIENTO as Nombre
    FROM resultados A
    inner join inf_estudiante B on B.ESTU_CONSECUTIVO=A.ESTU_CONSECUTIVO
    inner join inf_colegio C on C.COLE_CODIGO_ICFES=B.COLE_CODIGO_ICFES
    WHERE C.COLE_CODIGO_ICFES = codigo_colegio;
END$$
```

```
DELIMITER
```

```
1 • call obtenerPromedioColegio ('108381')
2
```

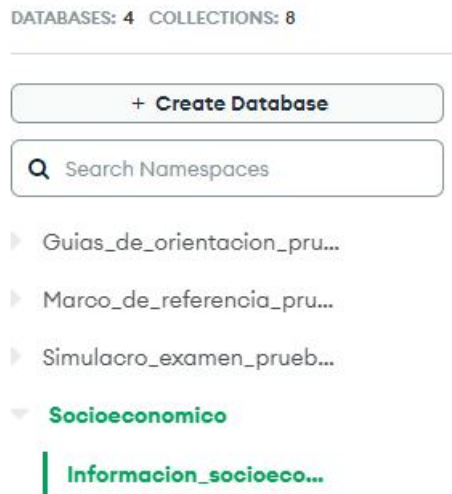
Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	Promedio	Nombre			
▶	321	COLEGIO RETOS			

5 Bases de Datos No-SQL (*Segunda entrega*)

5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

La organización en la base de datos consta de diferentes documentos en pdf, divididos según el tipo de información que aporta, por tanto se tiene 4 bases (Simulacro pruebas, marco de referencia, Guía de orientación y información socioeconómica) que cuentan con los archivos que se pudo descargar de las bases de datos del ICFES y los que estan disponibles en la web.

Distribución bases de datos.



la base simulacro exámen consta de 5 subbases dividadas por materia y cada una con sus respectivos archivos (simulacros)

Simulacro_examen_pruebas_saber_pro_11

LOGICAL DATA SIZE: 3.09KB STORAGE SIZE: 180KB INDEX SIZE: 180KB TOTAL COLLECTIONS: 5 CREATE COLLECTION

Collection Name	Documents	Logical Data Size	Avg Document Size	Storage Size	Indexes	Index Size	Avg Index Size
Simulacro_examen_ciencias	2	652B	326B	36KB	1	36KB	36KB
Simulacro_examen_competencias	2	630B	315B	36KB	1	36KB	36KB
Simulacro_examen_ingles	2	594B	297B	36KB	1	36KB	36KB
Simulacro_pruebas_saber11	2	670B	335B	36KB	1	36KB	36KB
Simulacro_examen_comprensión	2	618B	309B	36KB	1	36KB	36KB

Cada subbase esta organizada con información del colegio donde se puede aplicar el simulacro, el código del estudiantes y el enlace que lleva al documento que se encuentra guardado en drive.

Simulacro_examen_pruebas_saber_pro_11.Simulcro_examen_comprensión

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 618B TOTAL DOCUMENTS: 2 INDEXES TOTAL SIZE: 36KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

INSERT DOCUMENT

FILTER { field: 'value' } OPTIONS Apply Res

QUERY RESULTS: 1-2 OF 2

```
{
  "_id": "ObjectId('63785ddbd289925d4aa29c78')",
  "Año": "2021",
  "Archivo": "https://drive.google.com/file/d/1AyMkewR1s3tMFRuiyQc_J8whyr2g8IWZ/view...",
  "Asignatura": "Lectura crítica",
  "Colegio": "Gimnasio los pinos",
  "Estrato_estudiante": "3",
  "Id_estudiante": "20200120145"
}
```

Simulacro_examen_pruebas_saber_pro_11.Simulacro_pruebas_saber11

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 670B TOTAL DOCUMENTS: 2 INDEXES TOTAL SIZE: 36KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

INSERT DOCUMENT

FILTER { field: 'value' } OPTIONS Apply

```
{
  "Colegio": "Gimnasio los pinos",
  "Asignatura": "Matemáticas",
  "Cantidad_de_preguntas": "30",
  "Id_estudiante": "20201020145",
  "Nat_colegio": "No oficial",
  "Estrato_estudiante": "3",
  "Jornada": "Completa",
  "Archivo": "https://drive.google.com/file/d/1AH4uTsyUa0fcqCnikNgkAcPl2uKKi_2T/view..."
}
```

La base guía de orientación tiene información sobre la información que otorga el icfes a los colegios para que se preparen para la presentación de la prueba, dando a conocer las diferentes competencias que evalúa y la forma en que se va a realizar la prueba.

Guías_de_orientacion_pruebas_saber_11

LOGICAL DATA SIZE: 342B STORAGE SIZE: 36KB INDEX SIZE: 20KB TOTAL COLLECTIONS: 1

CREATE COLLECTION

Collection Name	Documents	Logical Data Size	Avg Document Size	Storage Size	Indexes	Index Size	Avg Index Size
Guías_de_orientacion	2	342B	171B	36KB	1	20KB	20KB

Estos documentos estan organizados por año y teniendo en cuenta el calendario que maneja el colegio, pues el icfes genera estas guías de orientación para colegios de calendario A y de calendario B.

Guías_de_orientacion_pruebas_saber_11.Guías_de_orientacion

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 342B TOTAL DOCUMENTS: 2 INDEXES TOTAL SIZE: 20KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes

INSERT DOCUMENT

FILTER { field: 'value' }

OPTIONS

Apply

Reset

QUERY RESULTS: 1-2 OF 2

```
_id: ObjectId('63786057d289925d4aa29c7d')
Año: "2021"
Archivo: "https://drive.google.com/file/d/1bQLmiIXkTmFKlUABkAXpQy3z-ttgm9XK/view..."
Cant_pag: "78"
Fuente: "ICFES"
```

Asimismo, los colegios cuentan con otro documento que el icfes proporciona y son los marcos de referencia, en donde se habla específicamente de cada una de las asignaturas, contando éstas con su propio marco, la información relevante que se presenta está centrada en el tipo de pregunta que se va evaluar, las diferentes habilidades y competencias de cada una de las asignaturas.

Marco_de_referencia_pruebas_saber_11

LOGICAL DATA SIZE: 446B STORAGE SIZE: 36KB INDEX SIZE: 36KB TOTAL COLLECTIONS: 1

CREATE COLLECTION

Collection Name	Documents	Logical Data Size	Avg Document Size	Storage Size	Indexes	Index Size	Avg Index Size
Marco_de_referencia	2	446B	223B	36KB	1	36KB	36KB

De esta manera la información que se proporciona del marco de referencia; es la fecha de publicación y la asignatura como principales datos.

Marco_de_referencia_pruebas_saber_11.Marco_de_referencia

STORAGE SIZE: 36KB LOGICAL DATA SIZE: 446B TOTAL DOCUMENTS: 2 INDEXES TOTAL SIZE: 36KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

INSERT DOCUMENT

FILTER { field: 'value' }

OPTIONS

Apply

Reset

QUERY RESULTS: 1-2 OF 2

```
_id: ObjectId('63785f3ed289925d4aa29c7b')
Año: "2021"
Archivo: "https://drive.google.com/file/d/1xZZFBlbCamZtCikUGSkSHVuT95IWpXs/view_"
Calendario: "A"
Cant_pág: "86"
Fuente : "ICFES"
Asignatura: "Ciencias naturales"
```

Por último se quiso agregar algo de información sobre el cuestionario socioeconómico que solucionan los estudiantes en el momento de encontrar la prueba, aunque no se encontró algún tipo de documento del icfes en este ambito, porque todo remitian a la guía de orientación, se dejó documentos que hablen sobre la construcción de las preguntas, propuestos por un grupo de estudiantes de la universidad Santo Tomas, en donde especifican que se debe tener como criterio para generar el cuestionario socioeconómico, por tanto, la información que se incluye del éste esta relacionada con información bibliográfica.

Socioeconomico

LOGICAL DATA SIZE: 0B STORAGE SIZE: 4KB INDEX SIZE: 4KB TOTAL COLLECTIONS: 1

CREATE COLLECTION

Collection Name	Documents	Logical Data Size	Avg Document Size	Storage Size	Indexes	Index Size	Avg Index Size
Informacion_socioeconomico	0	0B	0B	4KB	1	4KB	4KB

Socioeconomico.Informacion_socioeconomico

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

Find Indexes Schema Anti-Patterns ⓘ Aggregation Search Indexes ●

INSERT DOCUMENT

FILTER { field: 'value' } ▶ OPTIONS Apply Reset

QUERY RESULTS: 1-1 OF 1

```
{
  "_id": "ObjectId('63786119d289925d4aa29c7f')",
  "Año": "2016",
  "Título": "Propuesta para la construcción ...",
  "Autores": "Cuellar, Guerrero, López",
  "Editorial": "Universidad Santo Tomas",
  "ciudad": "Bogotá",
  "Archivo": "https://drive.google.com/file/d/1l357SctMce-hpcndFq_0_mUkYYfPVzVj/view..."
}
```

5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

Para el presente proyecto de investigación se utilizará una base de datos NoSQL, la cual consiste en simulacros y guías de orientación en formatos pdf que pueden leer y realizar los estudiantes próximos a presentar las pruebas saber 11, esto con el objetivo de contribuir en su preparación y mejorar así, los resultados en la presentación de dichas pruebas. Para el tratamiento de este tipo de datos se utilizará MongoDB, el cual es un sistema manejador de bases de datos NoSQL de gran reconocimiento debido a que cuenta con 37.000 clientes en más de 100 países y con aproximadamente 300 millones de descargas (MongoDB, 2022), esto se logra debido a que entre sus características podemos encontrar que es de código abierto (gracias a esto está disponible en todos los sistemas operativos), se pueden realizar todo tipo de consultas, soporta varios lenguajes de programación (entre ellos los más destacados como Python y Java) y no requiere de grandes recursos (puede ser instalado en ordenadores sin mayores prestaciones). MongoDB es un sistema de base de datos basado en documentos, cada registro en la base de datos es un documento en un formato muy similar al JSON llamado BSON, permitiendo una estructura de datos mucho más flexible y variable que en una base de datos relacional. MongoDB ofrece una gran escalabilidad y flexibilidad, y un modelo de consultas e indexación avanzado, disponible en opciones On-premise y Cloud, en el caso de este proyecto se utiliza la opción de almacenamiento en la nube con MongoDB Atlas.

6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (*Tercera entrega*)

6.1 Ejemplo de aplicación de ETL y Bodega de Datos (*Tercera entrega*)

7 Lecciones aprendidas (*Tercera entrega*)

8 Bibliografía

- Constitución Política de Colombia. Artículo 15. 1991. (Colombia).
- Datos Abiertos. (28 de febrero de 2020). Resultados Saber Pro Competencias Específicas 2019-2. Instituto Colombiano para la Evaluación de la educación – ICFES. Recuperado de <https://www.datos.gov.co/Educacion/Resultados-Saber-Pro-Competencias-Especificas-2019/6kwm-9788>
- Decreto N° 869 del 17 de marzo del 2010. Por el cual se reglamenta el Examen de Estado de la Educación Media, ICFES – SABER 11°. 17 de marzo de 2010.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). Metodología de la investigación (6ª. Ed.). Mc Graw Hill Education.
- Ley Estatutaria 1581 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales. 17 de octubre de 2012.
- Ley 1712 de 2014. Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones. 6 de marzo de 2014. D.O. N° 49.084.
- MongoDB Inc. (2022). ¿Que es MongoDB?. Recuperado de <https://www.mongodb.com/es/what-is-mongodb>
- MySQL. (2022). Products-MySQL Enterprise Edition. Recuperado de <https://www.mysql.com/products/workbench/>