

Generación de clústeres habitacionales a través de la caracterización y análisis del déficit cualitativo de vivienda en Colombia

Angela Marcela Vargas¹, Jorgue Andrés Buitrago²

¹⁻²Dpto. de Ingeniería y Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

¹angela.vargas@ucentral.edu.co, ²jbuitragoc7@ucentral.edu.com

November 26, 2022

Contents

1	Introducción	3
2	Características del proyecto de investigación	3
2.1	Título del proyecto de investigación	3
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
3	Reflexiones sobre el origen de datos e información	5
3.1	¿Cuál es el origen de los datos e información ?	5
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información?	5
3.3	¿Cuáles son los retos de la información y los datos que utilizará en la base de datos en términos de la calidad y la consolidación?	5
3.4	¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto?	6

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	7
4.2	Diagrama modelo de datos	7
4.3	Imágenes de la Base de Datos	8
4.4	Código SQL - lenguaje de definición de datos (DDL)	10
4.5	Código SQL - Manipulación de datos (DML)	10
4.6	Código SQL + Resultados: Vistas	11
4.7	Código SQL + Resultados: Triggers	11
4.8	Código SQL + Resultados: Funciones	12
4.9	Código SQL + Resultados: procedimientos almacenados	14
5	Bases de Datos No-SQL	16
5.1	Diagrama Bases de Datos No-SQL	16
5.2	SMBD utilizado para la Base de Datos No-SQL	18
6	Conexión Python - MySQL	20
6.1	Modelo segmentación K-Means	21
6.2	Clasificación	24
7	Lecciones aprendidas	25
8	Conclusiones y recomendaciones	26
9	Referencias	27

1 Introducción

El déficit cualitativo de vivienda se entiende como “la medida en que las viviendas existentes cumplen con ciertos estándares mínimos que permitan ofrecer una calidad de vida adecuada a sus habitantes” (CEPAL, 2000). Su correcta medición, apoya de manera significativa la toma de decisiones en materia de políticas públicas que mejoren las condiciones e impacten directamente en el bienestar de los hogares asociado a la calidad de vida. La vivienda “representa la principal inversión y el patrimonio más importante de las familias de ingresos medio y bajo; y, en algunos casos constituye incluso una fuente importante de ingresos.” (CEPAL, 2000), para este fin se cuenta con información recolectada por el DANE a través de encuestas que recogen datos relacionados con la vivienda.

El presente documento hará uso de la Encuesta Nacional de presupuesto de los Hogares – ENPH, este es un instrumento que el país utiliza para estimar ingresos y gastos de los hogares, los cuales sirven para evaluar el nivel, estructura y las tendencias del bienestar económico de los hogares. Además, los flujos de ingresos, los gastos en consumo y el ahorro -consumo futuro-, definen los límites del estilo de vida del hogar y su nivel de bienestar. (DANE, 2020). Cuenta también con un capítulo vivienda y hogares, que tiene como objetivo recolectar información relacionada con características físicas y equipamiento de las viviendas, y también provee el campo de ingresos para poder hacer relacionamiento entre estos dos conceptos. Se pretende entonces, realizar un modelo de base de datos que permita agrupar de forma significativa a los hogares encuestados, para realizar perfilamiento de las características de vivienda según nivel de ingresos.

2 Características del proyecto de investigación

Con la información del capítulo de vivienda y hogares de la Encuesta Nacional de Presupuesto de los Hogares, se pretende dar respuesta a algunas preguntas de interés relacionadas con las características habitacionales de los hogares colombianos, así como también, enlazar estos conceptos a los niveles de ingreso reportados, que permita finalmente hacer recomendaciones prácticas en materia de políticas públicas con base en análisis de clúster y perfiles. Para tal fin, se elaborará un modelo de base de datos que permita segregar la información del cuadernillo, de modo tal que sean aplicables los conceptos aprendidos durante el curso. Es de resaltar el hecho de que la estructura de los datos es apropiada para lograr un modelo entidad relación que lleve al entendimiento de una manera más simple.

2.1 Título del proyecto de investigación

Generación de clústeres habitacionales a través de la caracterización y análisis del déficit cualitativo de vivienda en Colombia.

2.2 Objetivo general

Desarrollar recomendaciones en términos de políticas públicas, a través del análisis del déficit cualitativo de vivienda en Colombia con base en perfiles de grupos elaborados desde la ENPH.

2.2.1 Objetivos específicos

- Definir segmentos de hogares para plantear recomendaciones en términos de políticas públicas de vivienda con enfoque de perfiles.
- Caracterizar las condiciones físicas y de equipamiento de vivienda en los hogares colombianos.
- Evaluar el déficit cualitativo de vivienda en Colombia a través de la ENPH teniendo en cuenta el ingreso de los hogares.

2.3 Alcance

El presente documento tiene un alcance de caracterización de la información para la aplicación de metodología clúster, es decir, se pretende con los datos disponibles generar análisis sobre variables clasificatorias de interés general para perfilar la vivienda de los hogares colombianos en función del nivel de ingresos, la ciudad de residencia, el número de personas que conforman el hogar, etc. Y de esta manera poder evaluar el déficit de vivienda en el país a través de la ENPH, para poder plantear posibles soluciones en materia de política pública con enfoque de perfiles dada la segmentación que arroje el resultado del modelo propuesto.

Esto en el marco de la Encuesta Nacional de presupuestos de los Hogares – ENPH, que fue realizada entre los años 2016 y 2017, cuya vigencia estadística es de 10 años, fue recolectada en 32 ciudades de departamento y 6 cabeceras adicionales. Esta caracterización es aplicable solo para el cuadernillo de vivienda y hogares disponible a nivel de microdatos de forma abierta en la página web del DANE.

2.4 Pregunta de investigación

¿Cómo afecta el nivel de ingresos la evaluación del déficit cualitativo de vivienda de los hogares encuestados en la ENPH?

2.5 Hipótesis

El estrato socioeconómico reportado en la Encuesta Nacional de Presupuestos de los Hogares – ENPH, no determina necesariamente el equipamiento y las condiciones físicas de la vivienda de los colombianos.

3 Reflexiones sobre el origen de datos e información

Los inicios de las mediciones de ingresos y gastos en los hogares colombianos se remontan al período de 1937 – 1940 en cabeza de la Contraloría General, con cobertura para las ciudades de Bogotá y Medellín, desde entonces y en intervalos de 10 años aproximadamente, se han realizado los diferentes esfuerzos para la actualización de esta información en materia de recolección y metodología.

Desde 1967, el Departamento Nacional de Estadística - DANE ha estado a cargo de la implementación de este proceso de investigación, mejorando las estructuras de captación de información y aplicando mejores prácticas sugeridas a nivel internacional por entidades como la OIT, EuroStat, CEPAL y OCDE, así como también, ampliando las coberturas para incluir la población rural, con el fin de entender dinámicas diferenciadas de las grandes ciudades. Entendiendo que dicha información es insumo para la toma de decisiones a nivel de política pública.

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

Encuesta Nacional de Presupuestos de los Hogares – ENPH 2016 – 2017, recolectada por el DANE y correspondiente al cuadernillo 1 cuyo tema principal es vivienda y hogares.

3.2 ¿Cuáles son las consideraciones legales o éticas del uso de la información?

Al ser una base de datos abierta al público para su consulta general, las consideraciones legales y éticas se enmarcan en los principios de los investigadores y sus fines académicos con el uso de la información.

3.3 ¿Cuáles son los retos de la información y los datos que utilizará en la base de datos en términos de la calidad y la consolidación?

Los principales retos de la información y los datos están en el entendimiento, la organización, partición y posterior uso. En primer lugar, el entendimiento de las estructuras del cuadernillo y su presentación en forma de base de datos, segundo la organización de la información relevante para cumplir con los objetivos del proyecto, tercero la partición de las bases de datos adecuadas para establecer un modelo entidad relación eficiente que permita la aplicabilidad de conceptos vistos durante el curso.

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

Eficiencia para el consumo de información con el fin de generar la construcción del conocimiento en torno a los temas de la caracterización física y de equipamiento de las viviendas de los hogares colombianos.

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

Para el desarrollo del presente proyecto se utilizará la herramienta MySQL cuyas características destacables son:

- Disponibilidad en gran cantidad de plataformas y sistemas operativos.
- Soporta gran cantidad de datos y posee diferentes opciones para su almacenamiento.
- Transacciones y claves primarias y foráneas.
- Conectividad segura.
- Utiliza el lenguaje de acceso a datos SQL.
- No necesita requerimientos de software ni hardware de alto rendimiento.
- De código abierto.

4.2 Diagrama modelo de datos

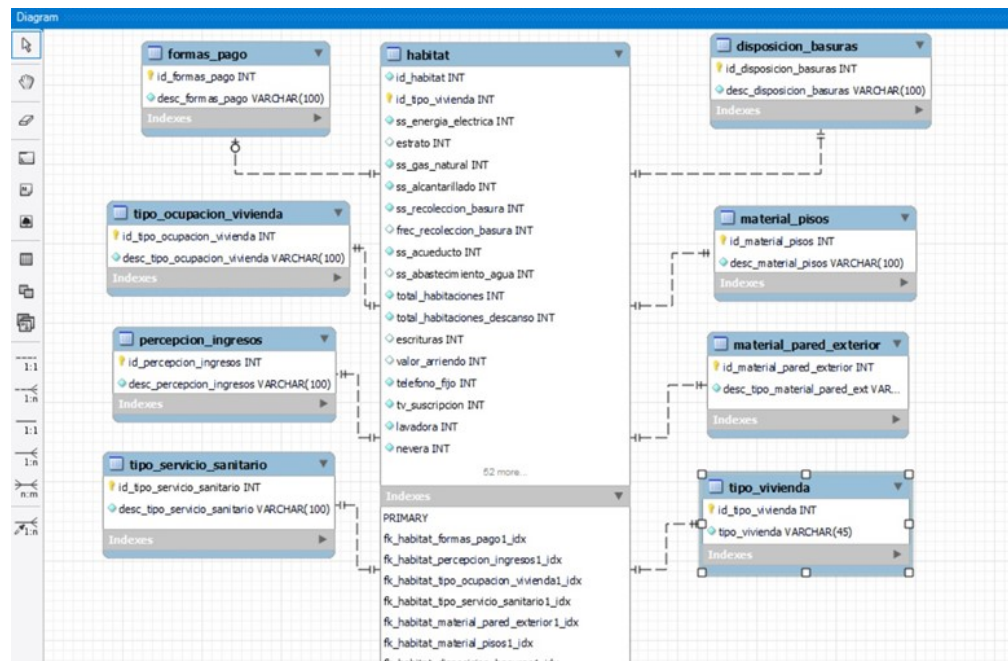


Figura 1. Modelo entidad relación

4.3 Imágenes de la Base de Datos

Model Overview

Tables (9 items)

Add Table

disposicion_basuras

tpo_ocupacion_vivie...

tpo_servicio_sanitario

formas_pago

tpo_vivienda

habitat

material_pared_exterior

material_pisos

percepcion_ingresos

Views (0 items)

Add View

Routines (0 items)

habitat - Table

material_pared_exterior - Table

Table Name: habitat

Schema: mydb

Filter Rows

Edit

Export/Import

Wrap Cell Contents

Apply changes

id_habitat	id_tipo_vivienda	ss_energia_electric	estrato	ss_gas_natural	ss_alcantarillado	ss_recoleccion_bas	frec_recoleccion_bas	ss_acueducto	ss_abastecimiento	total_habitado
118329	ORIENTAL	RURAL DISPERSO	1	1	2	1	2	2	2	2
118331	ORIENTAL	BUCARAMANGA ...	1	5	2	1	1	2	1	1
118332	PACIFICA	TUMACO	1	2	2	1	1	2	2	1
118338	PACIFICA	POPAYAN	1	1	4	1	2	1	1	1
118344	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	2	1	1	1
118346	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	2	2	1	1
118354	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	2	1	1	1
118356	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	2	1	1	1
118362	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	3	2	1	1
118364	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	3	1	1	1
118372	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	3	2	1	1
118373	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	3	1	1	1
118380	CENTRAL	MEDELLIN Y A.M.	1	1	4	1	4	1	1	1
118381	ATLANTICA	BARRANQUILLA	1	1	2	1	1	2	2	1
118383	ATLANTICA	BARRANQUILLA	1	5	1	1	1	2	2	1

Figura 2. Base de datos - Habitat

disposicion_basuras - Table	
Table Name: disposicion_basuras	Schema: mydb
Filter Rows:	Edit: Export/Import: Wrap Cell Content: Apply changes:
id_disposicion_	desc_disposicion_basuras
1	Por recolección pública o privada
2	La tiran a un río, quebrada, caño o la...
3	La tiran a un patio, lote, zanja o baldío
4	La queman o entierran
5	La eliminan de otra forma

material_pared_exterior - Table	
Table Name: material_pared_exterior	Schema: mydb
Filter Rows:	Edit: Export/Import: Wrap Cell Content: Apply changes:
id_material_pai	desc_tipo_material_pared_ext
1	Ladrillo, bloque, material prefabricad...
2	Madera pulida
3	Adobe o tapia pisada
4	Bahareque
5	Madera burda, tabla, tablón

material_pisos - Table	
Table Name: material_pisos	Schema: mydb
Filter Rows:	Edit: Export/Import: Wrap Cell Content: Apply changes:
id_material_pis	desc_material_pisos
1	Tierra, arena
2	Cemento, gravailla
3	Madera burda, tabla, tablón, otro vegetal
4	Baldosín, ladrillo, vinisol, otros materiales sintéticos
5	Mármol
6	Madera pulida
7	Alfombra o tapete de pared a pared

Figura 3. Base de datos - Otras

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

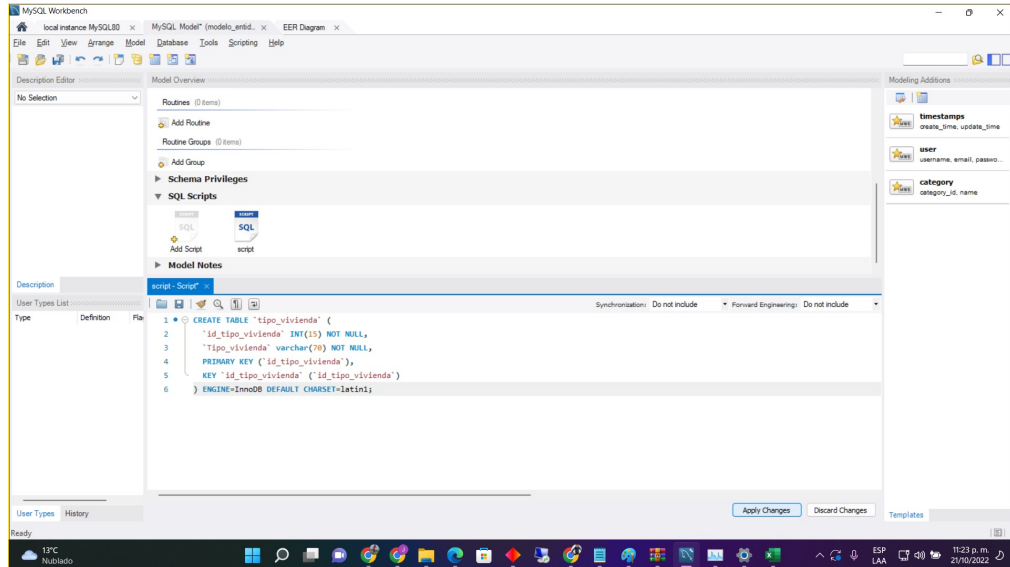


Figura 4. Base de datos - Creación

4.5 Código SQL - Manipulación de datos (DML)

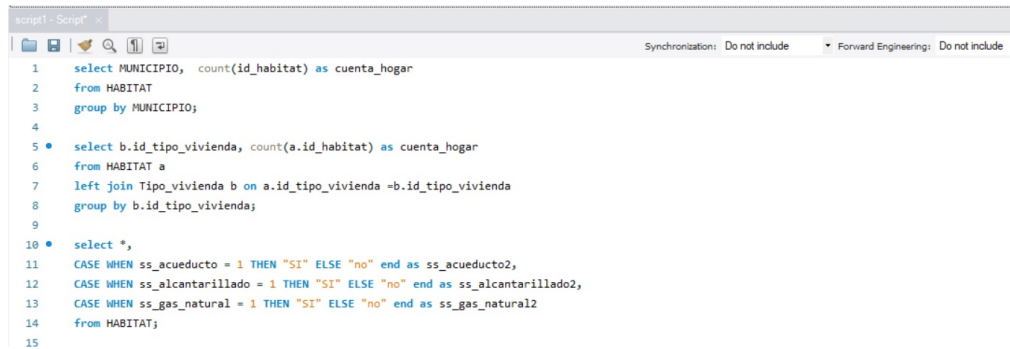
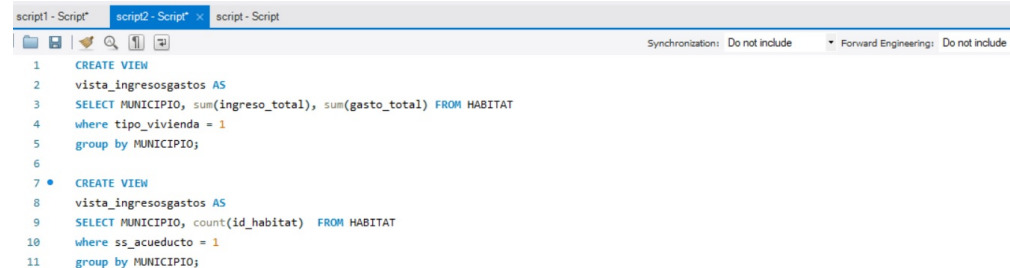


Figura 5. Base de datos - Manipulación

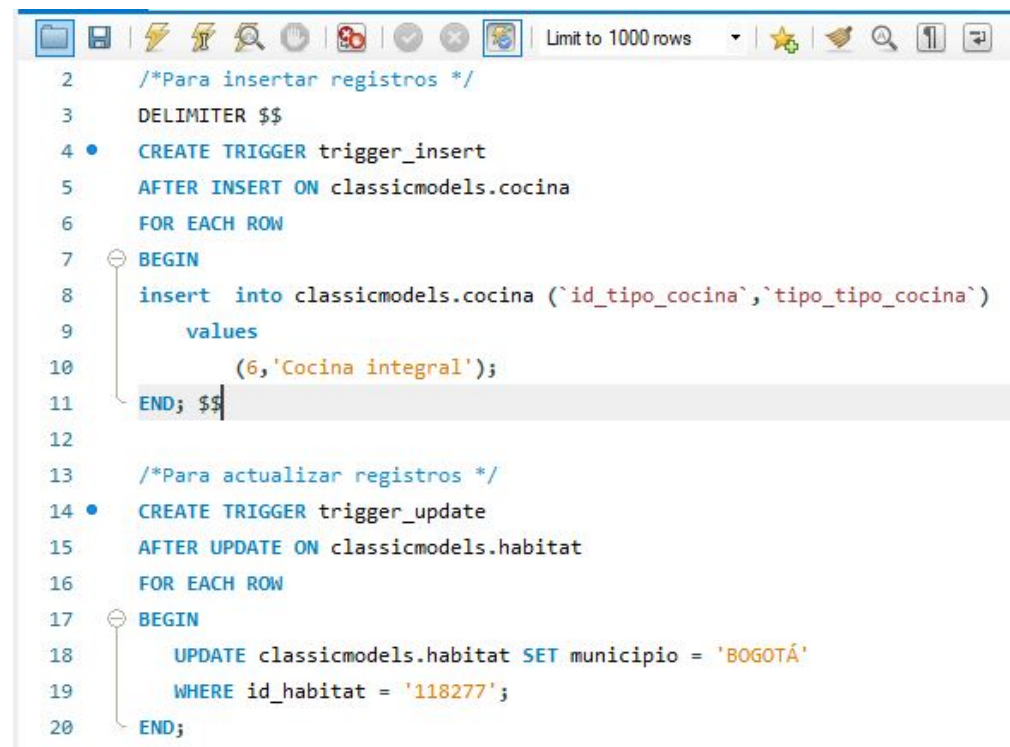
4.6 Código SQL + Resultados: Vistas



```
script1 - Script script2 - Script script - Script
1 CREATE VIEW
2 vista_ingresosgastos AS
3 SELECT MUNICIPIO, sum(ingreso_total), sum(gasto_total) FROM HABITAT
4 where tipo_vivienda = 1
5 group by MUNICIPIO;
6
7 • CREATE VIEW
8 vista_ingresosgastos AS
9 SELECT MUNICIPIO, count(id_habitat) FROM HABITAT
10 where ss_acueducto = 1
11 group by MUNICIPIO;
```

Figura 6. Base de datos - Vistas

4.7 Código SQL + Resultados: Triggers



```
Limit to 1000 rows
2 /*Para insertar registros */
3 DELIMITER $$
4 • CREATE TRIGGER trigger_insert
5 AFTER INSERT ON classicmodels.cocina
6 FOR EACH ROW
7 BEGIN
8 insert into classicmodels.cocina (`id_tipo_cocina`,`tipo_tipo_cocina`)
9 values
10 (6,'Cocina integral');
11 END; $$
12
13 /*Para actualizar registros */
14 • CREATE TRIGGER trigger_update
15 AFTER UPDATE ON classicmodels.habitat
16 FOR EACH ROW
17 BEGIN
18 UPDATE classicmodels.habitat SET municipio = 'BOGOTÁ'
19 WHERE id_habitat = '118277';
20 END;
```

Figura 7. Triggers - insertar,actualizar

```

22      /*Para borrar registros */
23      CREATE TRIGGER trigger_update
24      AFTER UPDATE ON classicmodels.habitat
25      FOR EACH ROW
26      BEGIN
27          DELETE FROM classicmodels.habitat
28          WHERE id_habitat = '118278';
29      END;
30

```

Figura 8. Triggers - Borrar

4.8 Código SQL + Resultados: Funciones

```

CREATE FUNCTION `asigna_si_no` ()
RETURNS TABLE
AS
RETURN(
    create table servicios as select *,
        CASE WHEN ss_acueducto = 1 THEN "SI" ELSE "no" end as ss_acueducto2,
        CASE WHEN ss_alcantarillado = 1 THEN "SI" ELSE "no"
            end as ss_alcantarillado2,
        CASE WHEN ss_recoleccion_basura = 1 THEN "SI" ELSE "no"
            end as ss_gas_natural2
    from classicmodels.habitat
);

```

Figura 9. Funciones - Decodifica servicios públicos

```

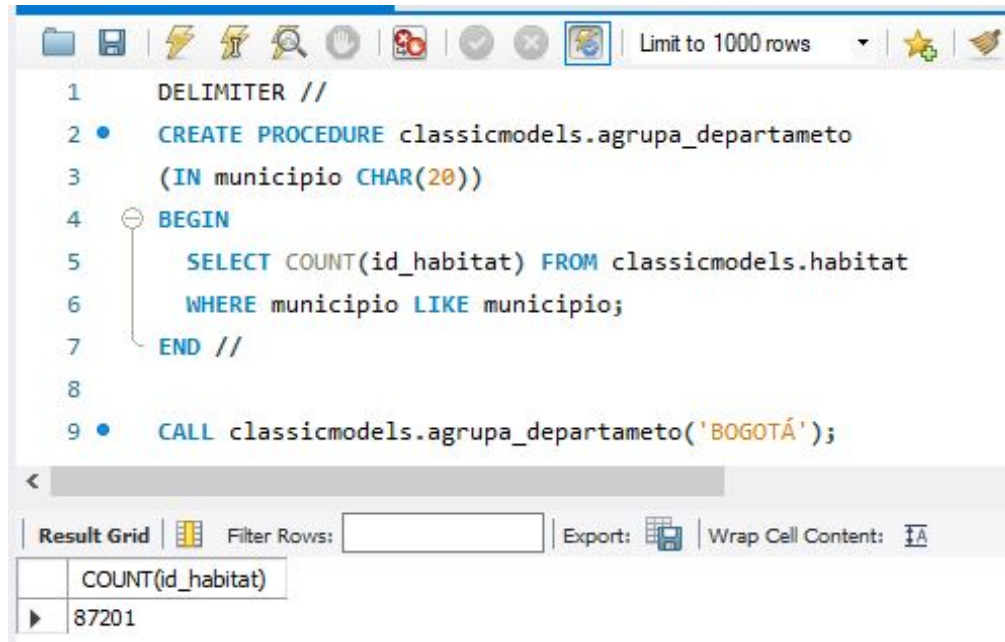
CREATE FUNCTION `asigna_textos` ()
RETURNS TABLE
AS
RETURN(
    create table habitat_textos as select a.*,
        b.tipo_cocina, c.tipo_pared_exterior, d.material_piso,
        e.percepcion_nivel_ingreso, f.tipo_ocupacion_vivienda,
        g.tipo_vivienda
    from classicmodels.habitat as a
    left join cocina as b on a.id_tipo_cocina=b.id_tipo_cocina
    left join material_pared_exterior as c
        on a.id_material_pared_exterior=b.id_material_pared_exterior
    left join material_piso as d on a.id_material_piso=b.id_material_piso
    left join percepcion_nivel_ingreso as e
        on a.id_percepcion_nivel_ingreso=b.id_percepcion_nivel_ingreso
    left join tipo_ocupacion_vivienda as f
        on a.id_tipo_ocupacion_vivienda=b.id_tipo_ocupacion_vivienda
    left join tipo_vivienda as g on a.id_tipo_vivienda=b.id_tipo_vivienda
);

```

tipo_material_piso	percepcion_nivel_ingreso	tipo_ocupacion_vivienda	tipo_vivienda	ss_acueducto
Baldosin, ladrillo, vinisol, otros materiales sintéti...	No alcanzan para cubrir los gastos básicos	En arriendo o subarriendo	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	En arriendo o subarriendo	Casa	si
Baldosin, ladrillo, vinisol, otros materiales sintéti...	Son suficientes para cubrir los gastos básicos	En arriendo o subarriendo	Apartamento	si
Baldosin, ladrillo, vinisol, otros materiales sintéti...	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Apartamento	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	En arriendo o subarriendo	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	En arriendo o subarriendo	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Baldosin, ladrillo, vinisol, otros materiales sintéti...	Son suficientes para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Baldosin, ladrillo, vinisol, otros materiales sintéti...	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Cemento, gravilla	No alcanzan para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si
Baldosin, ladrillo, vinisol, otros materiales sintéti...	Son suficientes para cubrir los gastos básicos	Propia, totalmente pagada	Casa	si

Figura 10. Funciones - Decodifica otras variables

4.9 Código SQL + Resultados: procedimientos almacenados



The screenshot shows a SQL IDE interface. The top toolbar includes icons for file operations, execution, and settings. The main editor displays the following SQL code:

```
1 DELIMITER //
2 CREATE PROCEDURE classicmodels.agrupa_departameto
3 (IN municipio CHAR(20))
4 BEGIN
5     SELECT COUNT(id_habitat) FROM classicmodels.habitat
6     WHERE municipio LIKE municipio;
7 END //
8
9 CALL classicmodels.agrupa_departameto('BOGOTÁ');
```

Below the editor, the 'Result Grid' tab is active, showing the results of the last query:

COUNT(id_habitat)
87201

```
CREATE PROCEDURE classicmodels.servicios_publicos
BEGIN
    select ss_acueducto_texto, ss_alcantarillado_texto,
           ss_recoleccion_basura_texto, count(id_habitat)
    from classicmodels.habitat_textos
    group by ss_acueducto_texto, ss_alcantarillado_texto,
            ss_recoleccion_basura_texto
END;

CALL classicmodels.servicios_publicos;
```



Result Grid				
		Filter Rows:		Export:  Wrap Cell Content: 
	ss_acueducto_texto	ss_alcantarillado_texto	ss_recoleccion_basura_texto	count(id_habitat)
►	si	si	si	69254
	si	no	si	6583
	si	no	no	2142
	no	si	si	2285
	no	no	si	3700
	no	no	no	2652
	no	si	no	62
	si	si	no	523

Figura 11. Procedimiento almacenado - Agrupación y conteo de datos

5 Bases de Datos No-SQL

A través de MongoDB se pretende hacer una colección de imágenes de referencia sobre las características físicas de una vivienda que se considera dentro de la evaluación del déficit cualitativo, para esto se tienen en cuenta los atributos de pisos y paredes, y la estructura física.

5.1 Diagrama Bases de Datos No-SQL

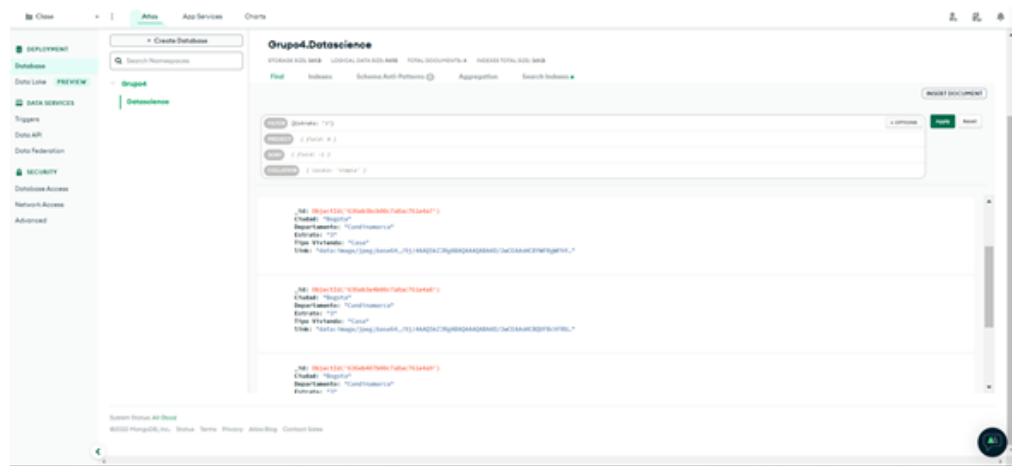


Figura 11. BD NoSQL - Casas en déficit cualitativo

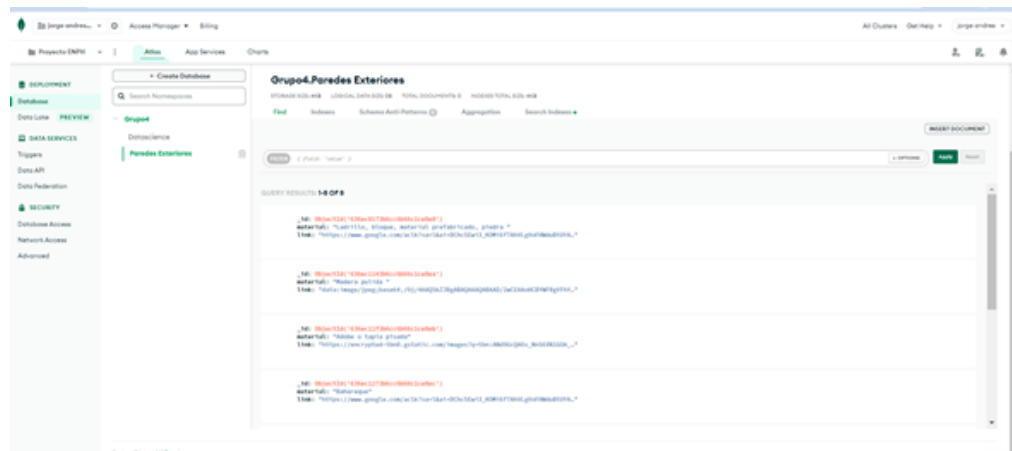


Figura 12. BD NoSQL - Tipos de materiales pisos y paredes

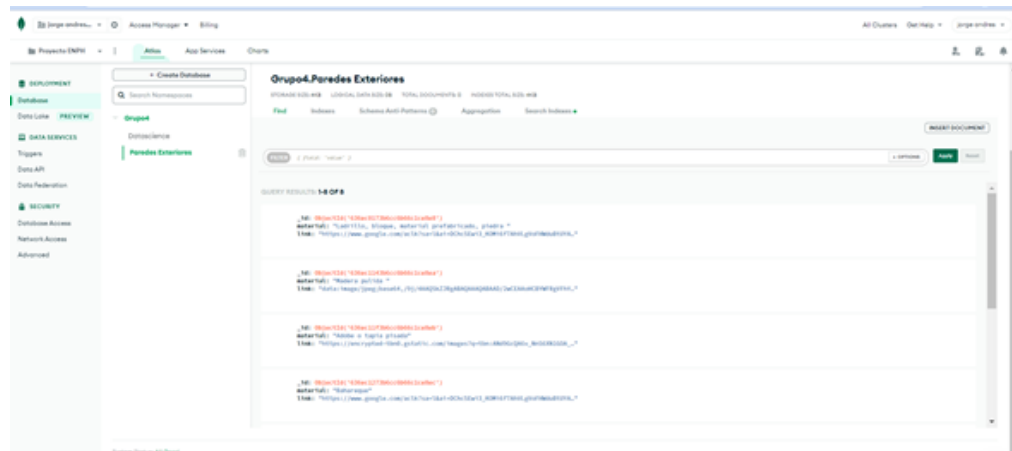


Figura 13. BD NoSQL - Dashboard 1

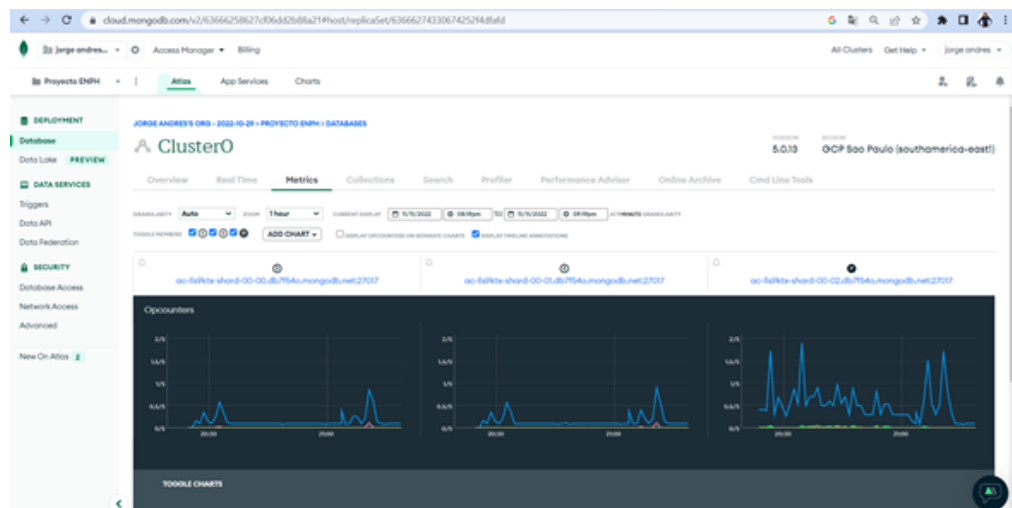


Figura 14. BD NoSQL - Dashboard 2

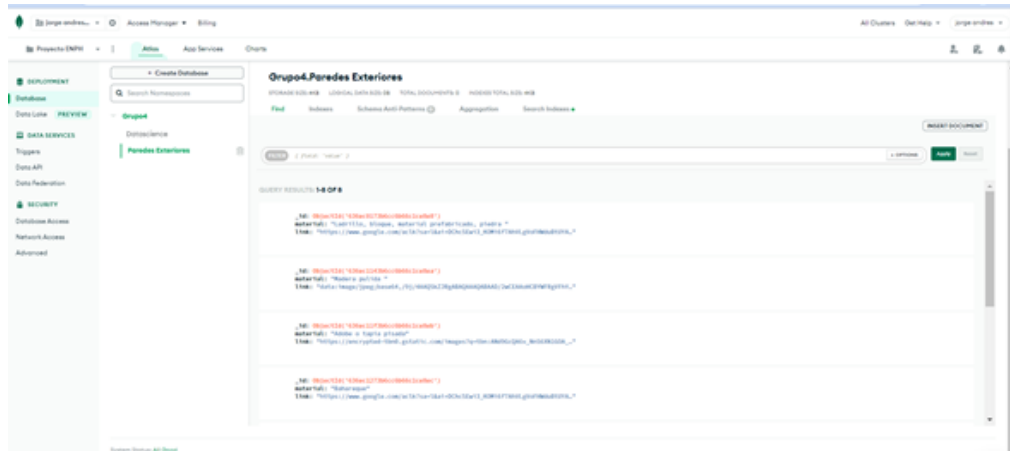


Figura 15. BD NoSQL - Dashboard 3

5.2 SMBD utilizado para la Base de Datos No-SQL

El sistema manejador de bases de datos no-SQL que se implementó fue MongoDB, “este es un sistema de base de datos NoSQL orientado a documentos de código abierto, que en lugar de guardar los datos en tablas lo hace en estructuras de datos BSON (similar a JSON) con un esquema dinámico.” (Robledano, sf)

Características principales:

- Consultas ad hoc.
- Indexación.
- Replicación.
- Balanceo de carga.
- Almacenamiento de archivos.
- Ejecución de JavaScript del lado del servidor.

Ventajas:

- Validación de documentos.
- Motores de almacenamiento integrado.
- Menor tiempo de recuperación ante fallos.

Desventajas:

- No es una solución adecuada para aplicaciones con transacciones complejas.

- No tiene un reemplazo para las soluciones de herencia.
- Aún es una tecnología joven.

6 Conexión Python - MySQL



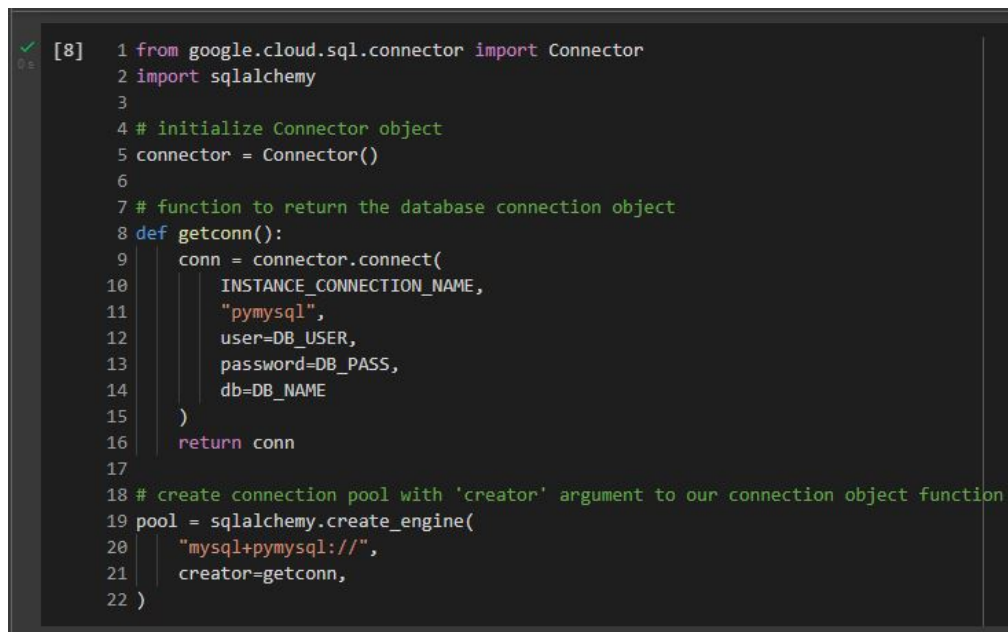
The screenshot shows a Jupyter Notebook interface. On the left, a code cell [6] contains the following Python code:

```
[6] 1 project_id = "guatemala" #@param (type:"string")
    2 region = "US" #@param (type:"string")
    3 instance_name = "guatemala" #@param (type:"string")

    4 # initialize parameters
    5 INSTANCE_CONNECTION_NAME = f"{project_id}:{region}:{instance_name}" # i.e
    6 print(f"Your instance connection name is: {INSTANCE_CONNECTION_NAME}")
    7 DB_USER = "root" #@param (type:"string")
    8 DB_PASS = "*****" #@param (type:"string")
    9 DB_NAME = "guatemala" #@param (type:"string")
```

Below the code cell, the output shows: "Your instance connection name is: guatemala:US:guatemala". On the right, a parameter widget displays the values for the variables: project_id: "guatemala", region: "US", instance_name: "guatemala", DB_USER: "root", DB_PASS: "*****", and DB_NAME: "guatemala".

Figura 16. Python - MySQL - Conexión



The screenshot shows a Jupyter Notebook interface. A code cell [8] contains the following Python code:

```
[8] 1 from google.cloud.sql.connector import Connector
    2 import sqlalchemy
    3
    4 # initialize Connector object
    5 connector = Connector()
    6
    7 # function to return the database connection object
    8 def getconn():
    9     conn = connector.connect(
    10         INSTANCE_CONNECTION_NAME,
    11         "pymysql",
    12         user=DB_USER,
    13         password=DB_PASS,
    14         db=DB_NAME
    15     )
    16     return conn
    17
    18 # create connection pool with 'creator' argument to our connection object function
    19 pool = sqlalchemy.create_engine(
    20     "mysql+pymysql://",
    21     creator=getconn,
    22 )
```

Figura 17. Python - MySQL - Llamado a la BD

```
1 # connect to connection pool
2 with pool.connect() as db_conn:
3
4     # query and fetch ratings table
5     car_personas = db_conn.execute("SELECT * FROM CAR_PERSONAS").fetchall()
6
7 # Create and populate a 5x2 NumPy array.
8 my_data = np.array(car_personas)
9
10 # Create a Python list that holds the names of the two columns.
11 my_column_names = ["ID_EST", "NUM_HOGAR", "ID", "SEXO", "EDAD", "PUEBLOS", "COMUNIDAD", "IDIOMA", "ESTADO_CIVIL"]
12
13 # Create a DataFrame.
14 df_carpersonas = pd.DataFrame(data=my_data, columns=my_column_names)
15
16 # Print the entire DataFrame
17 print(df_carpersonas)
```

ERROR:google.cloud.sql.connector.instance:['guatemala:US:guatemala']: An error occurred while performing refresh. Scheduling a refresh task.
Traceback (most recent call last):
File "/usr/local/lib/python3.7/dist-packages/google/cloud/sql/connector/instance.py", line 402, in _refresh_task

Figura 18. Python - MySQL - Carga de una tabla

6.1 Modelo segmentación K-Means

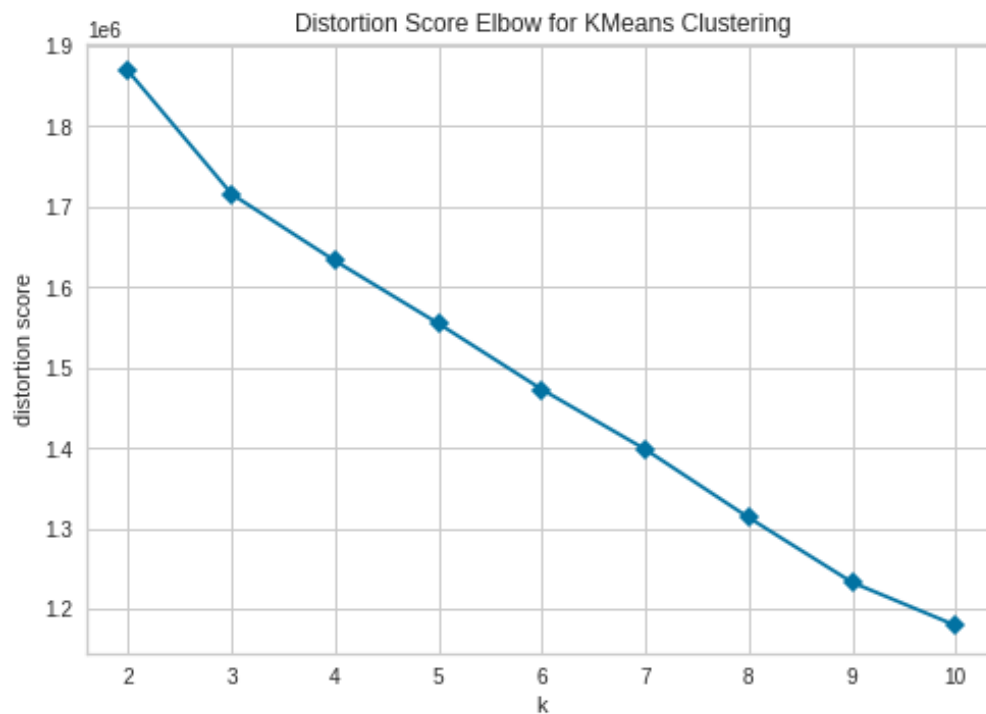


Figura 19. Diagrama de codo

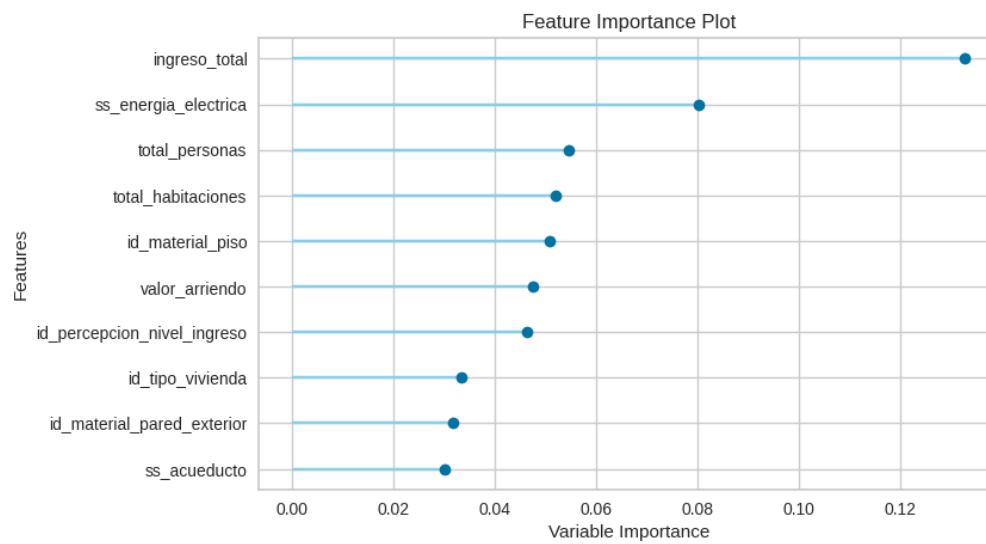


Figura 20. Importancia de variables para el modelo

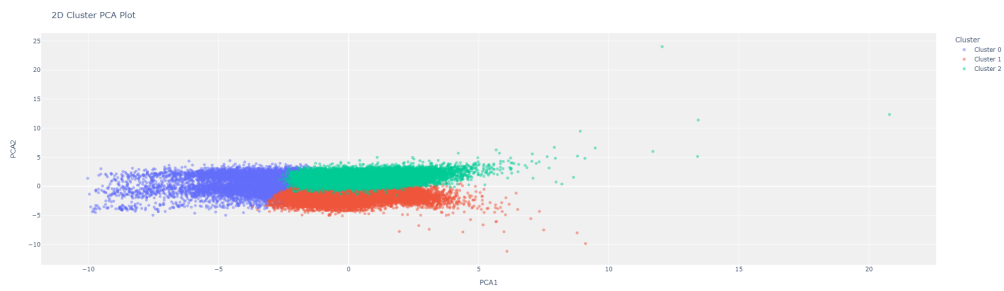


Figura 21. Clústeres

RandomForestClassifier Confusion Matrix

0	198	53	12	3	0	0	0	1
1	1	5408	1823	317	22	2	2	4
2	0	1836	4137	1101	117	23	6	0
3	0	399	1581	2001	230	56	29	0
4	0	30	210	425	231	63	19	0
5	0	6	28	140	75	92	29	0
6	0	3	7	47	31	33	47	0
9	1	33	0	0	0	0	0	17
	0	1	2	3	4	5	6	9

Predicted Class

Figura 22. Matriz de confusión

6.2 Clasificación

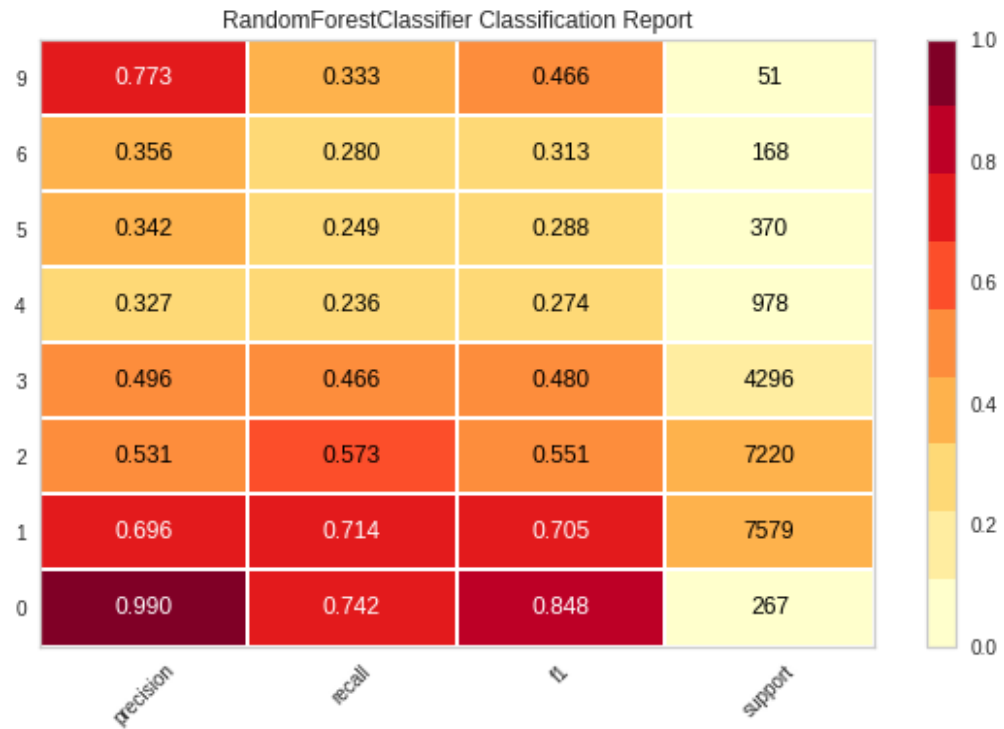


Figura 23. Métricas

7 Lecciones aprendidas

Dentro del desarrollo del presente proyecto se plantearon dos enfoques para las lecciones aprendidas, uno de carácter técnico-funcional y otro de carácter académico-profesional. El primero hace referencia a la aplicabilidad práctica de los conceptos estudiados donde se resalta el entendimiento de la importancia de los modelos entidad relación a la hora de la creación, administración y modificación de las bases de datos SQL.

Dentro de este mismo enfoque, la búsqueda de la eficiencia en la utilización de las herramientas disponibles para el tratamiento de las bases de datos que permitan consultas y transformaciones útiles para el desarrollo de la analítica en torno a la toma de decisiones sobre resultados específicos, es decir, hacer un buen uso de los Triggers, funciones y procesos almacenados. Para esto se considera indispensable tener conocimiento sobre la forma técnica del funcionamiento de estas. Finalmente, se resalta la importancia que toman en la actualidad las bases de datos NoSQL, su comprensión, manejo y amplios campos de aplicabilidad se convierten en una oportunidad para la analítica en general.

En el enfoque académico-profesional, se resaltan los aprendizajes en temas de manejo del tiempo y oportunidad en las entregas de los compromisos adquiridos con el curso, así como también el cambio en el paradigma de las bases de datos estructuradas, ya que se incluyen conceptos nuevos como las bases NoSQL y se amplía el panorama sobre las formas en la que se puede obtener información de una base de datos para futuros proyectos académicos o del campo laboral.

8 Conclusiones y recomendaciones

La variable ingreso representa un alto valor en los análisis asociados al déficit cualitativo de vivienda en Colombia, le siguen el servicio de energía eléctrica y la cantidad de personas que conforman el hogar. Esto en términos de políticas públicas lleva a inferir sobre la importancia del enfoque en términos de empleo, servicios públicos y familia.

En este sentido el enfoque en términos de política pública puede darse desde otra perspectiva diferente a la de vivienda exclusivamente, es decir, se pueden fortalecer las políticas ya existentes como subsidio familia, oportunidades de financiación, garantías para hogares informarles, entre otras (DNP, sf). Esto en el marco de la política nacional de vivienda, y orientar una mayor inversión en aquellas fórmulas que ataquen problemas como el empleo, el acceso a servicios públicos y la familia, por ejemplo, en términos de salud sexual y reproductiva, ya que la cantidad de miembros que conforman el hogar representa un factor importantes en la medición del déficit cualitativo de vivienda.

9 Referencias

1. CEPAL. (Septiembre de 2000). cepal.org. Obtenido de https://www.cepal.org/sites/default/files/publication/files/5080/S00080715_es.pdf
2. DANE. (2009). dane.gov.co. Obtenido de https://www.dane.gov.co/files/investigaciones/fichas/Deficit_vivienda.pdf
3. DANE. (2020). COLOMBIA - Encuesta Nacional de Presupuestos de los Hogares - ENPH - Julio 2016 - Julio 2017. Obtenido de <http://microdatos.dane.gov.co/index.php>
4. DANE. (2020). Metodología General Encuesta Nacional de Presupuestos de Los Hogares – ENPH.
5. DNP. (sf). www.dnp.gov.co. Obtenido de <https://www.dnp.gov.co/programas/vivienda-agua-y-desarrollo-urbano/Vivienda/Paginas/Pol>
6. Gómez, J. (sf). srcodigofuente.es. Obtenido de <https://www.srcodigofuente.es/aprender-sql/triggers-sql>
7. Fedesarrollo. (2020). Tendencia Económica. Obtenido de <https://www.repository.fedesarrollo.org.co/handle/11445/4050>
8. International Household Survey Network - IHSN. (2019). Encuesta Nacional de Presupuestos de los Hogares 2016-2017. Obtenido de <https://catalog.ihsn.org/index.php/catalog/7949>
9. Portafolio. (agosto de 2022). portafolio.co. Obtenido de <https://www.portafolio.co/mis-finanzas/vivienda/en-colombia-se-necesitan-400-mil-viviendas-al-ano-para-acabar-deficit-habitacional-570228>
10. Robledano, A. (sf). openwebinars.net. Obtenido de <https://openwebinars.net/blog/que-es-mongodb/>
11. Saiz, A. (julio de 2022). cre.mit.edu. Obtenido de <https://cre.mit.edu/news-insights/urban-economics-lab-and-la-haus-uncover-extent-of-housing-needs-in-mexico-and-colombia/>