

Prueba Técnica – Información demográfica Clientes y campaña de tarjetas de crédito.

Analítico Nivel 1 - LdC Analítica Pymes y Empresas Bancolombia.

Simón Correa Marín.

Para la realización de esta prueba técnica se utilizaron diferentes herramientas que sirven de apoyo a la hora de realizar un proyecto de ingeniería de datos. A continuación explico las herramientas y dejo los enlaces correspondientes. Por favor revisar cada uno ya que allí se encuentra más a detalle el análisis y procesamiento de datos realizado.

- **GitHub:** Se creó un proyecto privado en GitHub donde está toda la estructura de carpetas y desarrollo del proyecto. [GitHub Prueba Técnica - Analítico BC](#)
- **Canva:** Allí se realizó la presentación de los resultados de las consultas y la visualización. [Presentación Prueba Analítica](#)
- **VSCode:** IDE de desarrollo del proyecto.
- **SSMS (SQL Server Management Studio):** Gestor de bases de datos utilizado.
- **Power BI:** Herramienta de visualización utilizada.
- **Mermaid:** Diagramador online donde se realizó el diagrama ER.

1. Revisar estas tablas y determinar si ya se tienen identificados todos los catálogos o si se podrían crear otros adicionales.

Antes de realizar la identificación de los catálogos y proponer adicionales se realizó un EDA (Exploratory Data Analysis) para analizar y conocer los datos en busca de nulos, duplicados, atípicos y otras anomalías o patrones identificables en este primer overview.

Revisar el archivo del repositorio – [exploracion.ipynb](#)

Se realizaron las siguientes modificaciones y tratamientos a los datos:

- Se realizó una exploración y análisis de todas las tablas proporcionadas, con foco en la estructura, calidad de los datos y relaciones entre entidades.
- Se aplicó un análisis exploratorio automatizado con ydata-profiling, lo cual permitió detectar anomalías, valores atípicos, errores tipográficos y oportunidades de mejora en la calidad de los datos.
- Todas las anomalías identificadas fueron corregidas o tratadas, incluyendo:
 - Corrección del valor mal escrito "Fenemino" en la columna desc_genero.
 - Conversión de la columna fecha_nacimiento a formato datetime y eliminación de outliers en años improbables (como 1800 y 2050).
 - Eliminación de registros inconsistentes como personas menores de edad con cédula de ciudadanía (CC).

- Filtrado de clientes con ingresos inusualmente bajos (menores a \$100.000), que podrían corresponder a datos faltantes o erróneos.
 - Limpieza de registros en catálogos como cat_ciudad con valores "Sin Informacion" o códigos 0.
 - Se identificó que campos como desc_departamento contienen nombres de países, lo que sugiere una mezcla de niveles territoriales en los catálogos.
- Se propone la creación de nuevos catálogos derivados de campos como desc_genero, desc_departamento y desc_pais, actualmente embebidos en otras tablas, para facilitar futuros análisis y reportes.
 - Se plantea establecer una estructura más robusta para los catálogos, mejorando la trazabilidad y consistencia de los datos mediante llaves foráneas bien definidas.

Para observar todos los archivos con los cambios realizados por favor dirigirse al link de GitHub:

[GitHub Prueba Técnica - Analítico BC](#)

Propuesta de nuevos catálogos

Revisar el archivo del repositorio – *catalogos.ipynb*

Se identificó que el campo cod_ciudad en cat_ciudad tiene una estructura jerárquica compuesta que permite derivar niveles geográficos como país, departamento, municipio y ciudad. Por tanto, se propone crear catálogos adicionales para facilitar la segmentación y el análisis regionalizado de los clientes.

Segmento	Significado	Ejemplo (169068296006)
PPP	País (o tipo de sistema)	169
DDD	Departamento (en Colombia)	068
MMM	Municipio o ciudad núcleo	296
CCC	Localidad, barrio o vereda	006

Propuesta de nuevos catálogos, su llave primaria y campos sugeridos.

Tabla	Clave	Campos sugeridos
cat_genero	cod_genero	desc_genero
cat_departamento	cod_departamento	desc_departamento, cod_pais
cat_pais	cod_pais	desc_pais
cat_region	cod_region	desc_region, cod_pais

Ahora, el modelo de datos ha sido normalizado mediante la creación de catálogos jerárquicos, permitiendo segmentaciones por país, región, departamento, ciudad, género, y zona comercial. Esto facilita el análisis por filtros en herramientas de BI y garantiza una arquitectura escalable.

2. Crear una metadata para todas las tablas con la siguiente estructura PK (Clave principal): cualquier campo de identificador que se encuentra en la base de datos principal o catálogos.

# fuente de dato	Nombre tabla	Nombre campo	Tipo de dato	Longitud	PK (si/no)	FK (si/no)
1	clientes	id	int		si	no
1	clientes	cod_tip_doc	int		no	si
1	clientes	cod_genero	varchar	1	no	si
1	clientes	nombre	varchar	65	no	no
1	clientes	fecha_nacimiento	date		no	no
1	clientes	ingresos	decimal	12.2	no	no
1	clientes	cod_sucursal	int		no	si
2	cat_tip_doc	cod_tip_doc	int		si	no
2	cat_tip_doc	desc_tip_doc	varchar	20	no	no
3	cat_sucursal	cod_sucursal	int		si	no
3	cat_sucursal	desc_sucursal	varchar	50	no	no
3	cat_sucursal	cod_pais_region_zona	varchar	7	no	si
3	cat_sucursal	cod_ciudad	varchar	12	no	si
4	cat_ciudad	cod_ciudad	varchar	12	si	no
4	cat_ciudad	desc_ciudad	varchar	35	no	no
4	cat_ciudad	cod_departamento	varchar	3	no	si
4	cat_ciudad	cod_pais	varchar	3	no	si
5	cat_zona	cod_pais_region_zona	varchar	7	si	no
5	cat_zona	desc_zona	varchar	50	no	no
5	cat_zona	cod_region	varchar	2	no	si
5	cat_zona	cod_pais	varchar	3	no	si
6	cat_genero	cod_genero	varchar	1	si	no
6	cat_genero	desc_genero	varchar	15	no	no
7	cat_pais	cod_pais	varchar	3	si	no
7	cat_pais	desc_pais	varchar	20	no	no
8	cat_departamento	cod_departamento	varchar	3	si	no
8	cat_departamento	desc_departamento	varchar	25	no	no
8	cat_departamento	cod_pais	varchar	3	no	si
9	cat_region	cod_region	varchar	2	si	no
9	cat_region	desc_region	varchar	30	no	no

Metadata Tablas

3. Construir un diagrama de base de datos relacional en el que se muestren las tablas y sus relaciones entre ellas.

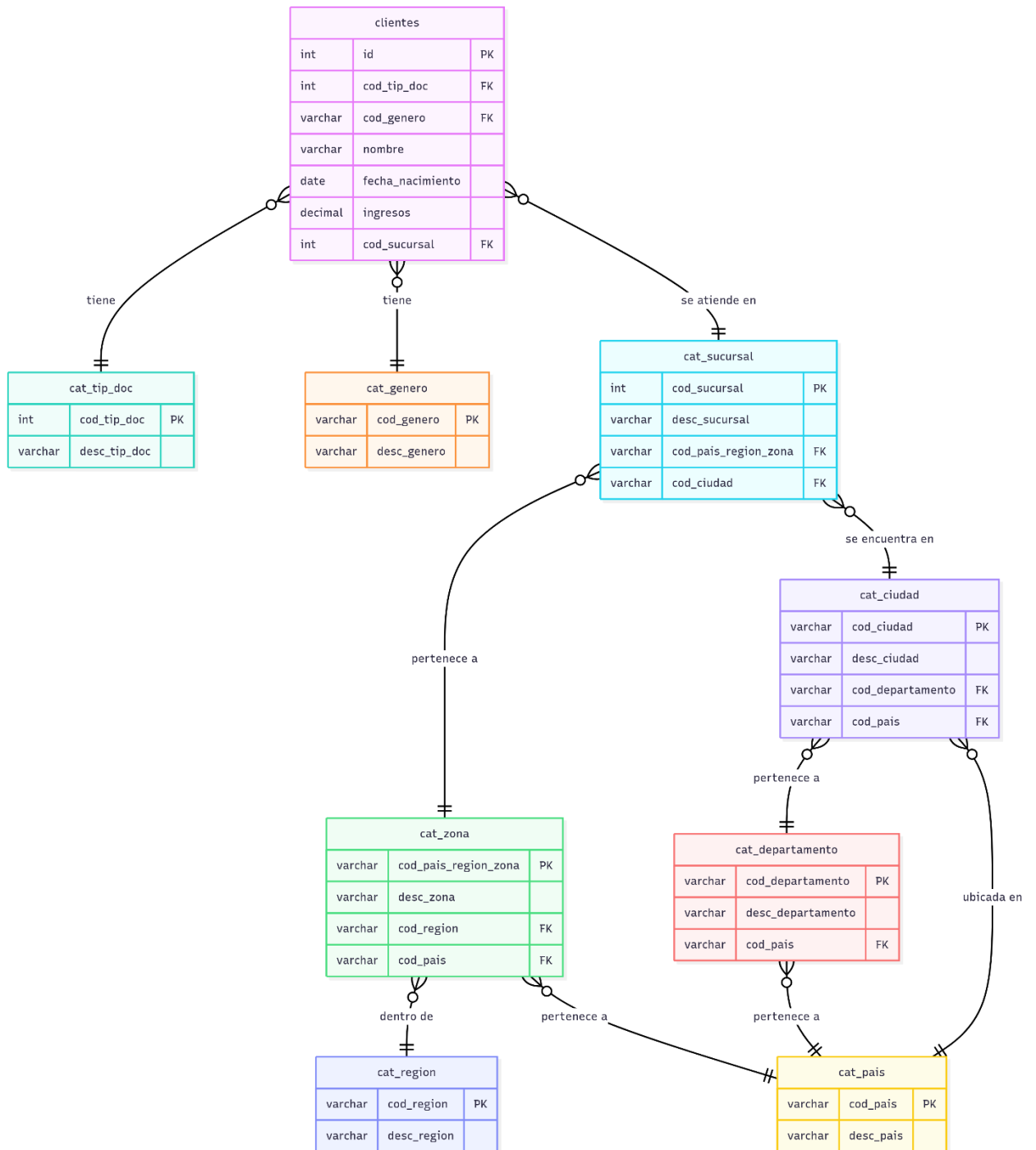
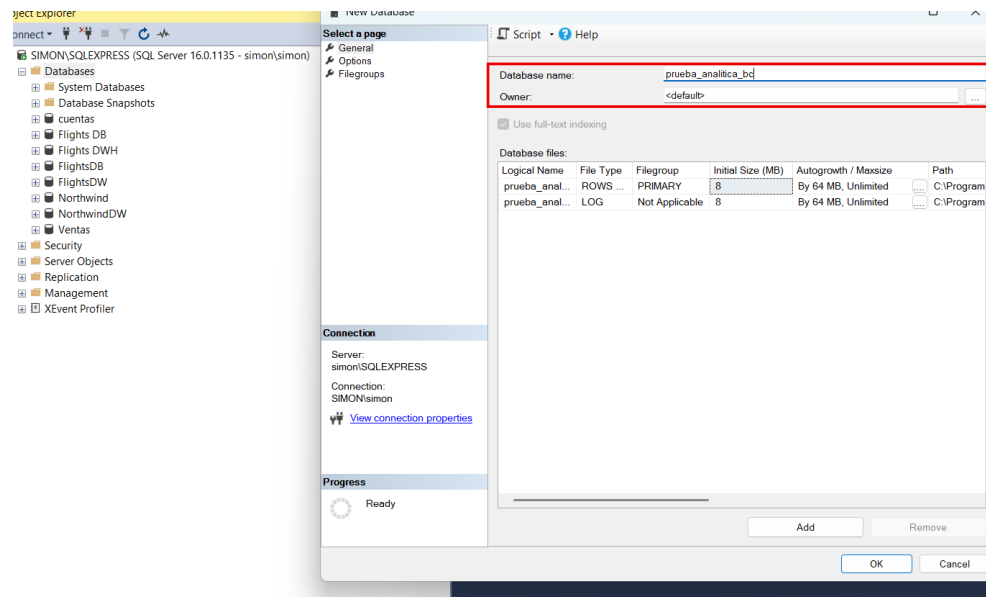


Diagrama ER realizado en mermaid.com por Simón Correa Marín.

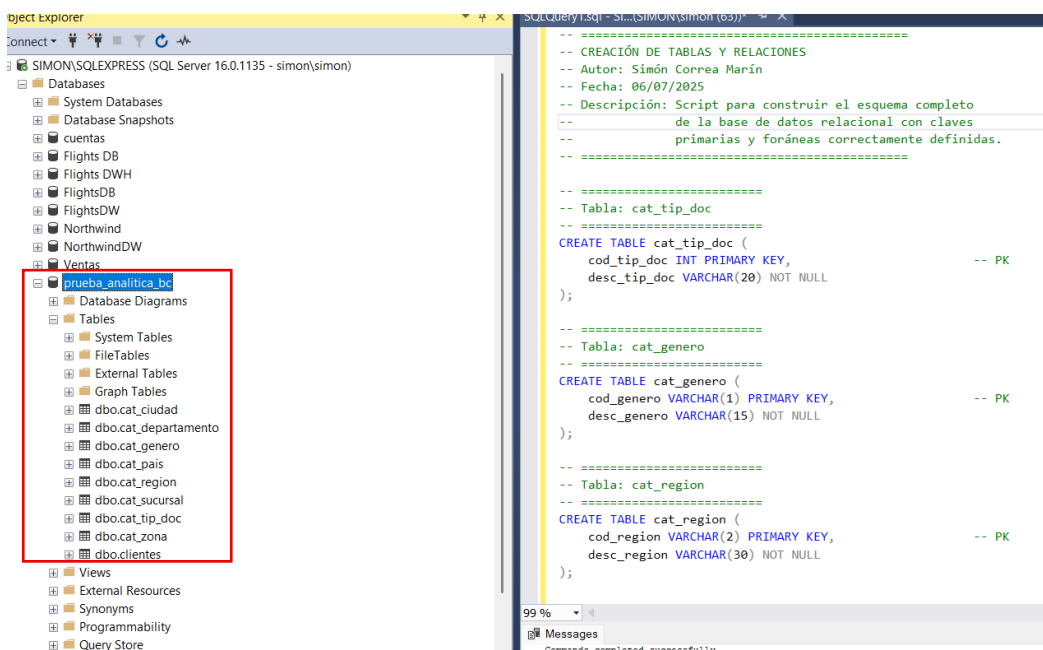
4. Crear esta estructura en algún manejador de bases de datos (estructurado o no estructurado) como SQL Server, MySQL, DB2, Oracle, MongoDB, Hadoop, etc. Inclusive se podría en Access o en Excel (y crear los queries mediante macros, que también es válido). Es importante mostrar y explicar bien cómo se crearon las tablas y como se insertaron los registros en ellas.

Para efectos de este ejercicio opté por utilizar el database manager de SQL SERVER de Microsoft. SQL Server Management Studio (SSMS).

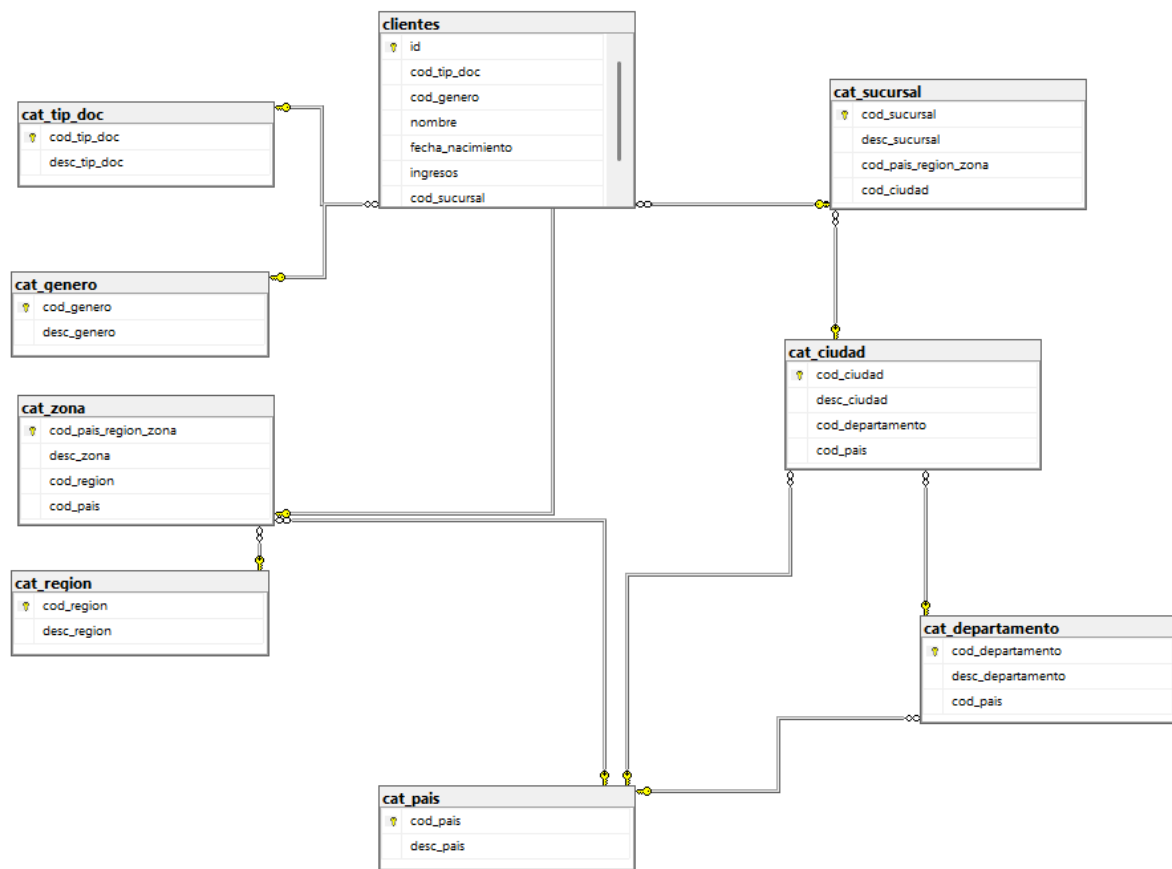
Creamos la conexión al servidor y creamos la base de datos.



Luego de tener el script **schema.sql** de creación del esquema de la base de datos los ejecutamos en un query en SQL Server y de esta manera se crea nuestro esquema sin datos aún.



Ahora podemos ver el diagrama ER creado a partir del esquema definido anteriormente.



Alimentemos el esquema ahora, los datos fueron divididos y exportados anteriormente a partir de la propuesta que se realizó de nuevos catálogos.

La imagen muestra un entorno de desarrollo con una interfaz de usuario y un panel de código.

Interfaz de Usuario:

- Se muestra una lista de archivos en el directorio `data`, con una subcarpetal `csv` que contiene los siguientes archivos: `cat_ciudad.csv`, `cat_departamento.csv`, `cat_genero.csv`, `cat_pais.csv`, `cat_region.csv`, `cat_sucursal.csv`, `cat_tip_doc.csv`, `cat_zona.csv`, `clientes.csv`, `tablas_finales.xlsx`, `tablas_limpias_prueba.xlsx` y `tablas_prueba_ing_datos.xlsx`.

Panel de Código:

```
import os

# Ruta de salida "Ruta": Unknown word.
ruta_csv = "../data/csv/" "ruta": Unknown word.
os.makedirs(ruta_csv, exist_ok=True) # Crea la carpeta

# Se exporta cada tabla como archivo CSV "exporta
for nombre, df in tablas.items(): "nombre": Unkn
    ruta_archivo = os.path.join(ruta_csv, f"{nombre}
    df.to_csv(ruta_archivo, index=False, encoding='u

print(f"Archivos CSV exportados exitosamente en: {ru

[9] ✓ 0.7s
```

El mensaje de salida indica: "Archivos CSV exportados exitosamente en: ../data/csv/"

Ahora podemos alimentar cada tabla con sus respectivos registros mediante el wizard que nos destina SSMS para ello.

Preview Data

This operation analyzed the input rows.

cod_tip_doc	desc_tip_doc
6	Id Extranjer...
7	Id Extranjer...
0	Carné Diplo...
8	Fideicomiso
4	T.I
1	C.C
5	Pasaporte
3	NIT
2	C.E
9	Registro Civil

Preview Data

This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.

id	cod_tip_doc	cod_genero	nombre	fecha_nacimix	ingresos	cod
1092347697	1	M	Edgar Javier...	1995-02-04	30589665	202
1007686547	1	M	Jordan Stev...	1986-07-17	26600681	184
71494890	1	M	Fabio De Je...	1983-03-22	6184024	368
1032490466	1	F	Ivonne Alej...	1991-04-23	12402821	378
71196443	1	M	Alex Albert...	1987-04-26	9715700	185
77176360	1	M	Olver Jose G...	1960-12-11	22029506	292
66814186	1	F	Yaneth Jara...	1981-06-09	28552690	699
42983979	1	F	Gloria Marg...	1968-11-01	9485800	699
70047535	1	M	Francisco Ja...	1966-11-08	23103321	23
1075239249	1	F	Jennifer Pl...	1975-04-15	3635983	477
900751472	3	N	Comercializ...	1989-05-25	13263269	299
1017268732	1	F	Carolina Via...	1991-02-27	26519518	876
1013666942	1	M	Anderson Q...	2002-02-28	12233170	212
717933	2	F	Zaideth Del...	1949-01-10	15245709	706
1010070289	1	F	Julieith Paol...	1956-05-19	2993839	37
1023905225	1	F	Lizeth Giova...	1950-12-09	29451311	364
55238680	1	F	Daniela ALE...	1988-08-03	31606816	753
3412661	1	M	Ivan Dario V...	1969-05-21	16214050	647
52993543	1	F	Dominga D...	1959-03-23	19541187	771
5104230317	5	M	Asnoib Ric...	1995-09-09	17534569	392

☒ Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous values may be dropped.

Basados en la tabla de metadatos que realizamos anteriormente definimos las reglas y restricciones para cada tabla incluidas sus claves primarias y foráneas.

Modify Columns

This operation generated the following table schema. Please verify if schema is accurate, and if not, please make any changes.

Column Name	Data Type	Primary Key	Allow Nulls
id	int	<input checked="" type="checkbox"/>	<input type="checkbox"/>
cod_tip_doc	int	<input type="checkbox"/>	<input type="checkbox"/>
cod_genero	nvarchar(1)	<input type="checkbox"/>	<input type="checkbox"/>
nombre	nvarchar(65)	<input type="checkbox"/>	<input type="checkbox"/>
fecha_nacimiento	date	<input type="checkbox"/>	<input type="checkbox"/>
ingresos	decimal(12, 2)	<input type="checkbox"/>	<input type="checkbox"/>
cod_sucursal	int	<input type="checkbox"/>	<input type="checkbox"/>

Database Diagrams

prueba_analitica_bc

- Database Diagrams
 - Tables
 - System Tables
 - FileTables
 - External Tables
 - Graph Tables
 - dbo.cat_ciudad
 - dbo.cat_departamento
 - dbo.cat_genero
 - dbo.cat_pais
 - dbo.cat_region
 - dbo.cat_sucursal
 - dbo.cat_tip_doc
 - dbo.cat_zona
 - dbo.clientes
 - Columns
 - id (PK, bigint, not null)
 - cod_tip_doc (FK, int, not null)
 - cod_genero (nvarchar(1), not null)
 - nombre (nvarchar(65), not null)
 - fecha_nacimiento (date, not null)
 - ingresos (decimal(12,2), not null)
 - cod_sucursal (FK, int, not null)

5. Genere las siguientes columnas adicionales:

- Nombres:** esta columna debe contener los nombres completos de los clientes
- Apellidos:** Esta columna debe contener los dos apellidos de los clientes
- Edad del cliente**

Esta es nuestra tabla maestra antes de generar las columnas.

```
SELECT *  
FROM clientes
```

id	cod_tip_doc	cod_genero	nombre	fecha_nacimiento	ingresos	cod_sucursal
3422	1	M	Enrique Florez Faillace	1994-06-05	19983828.00	81
3486	1	M	Cesar Alberto Cortes De La Espriella	1996-06-23	3305365.00	320
3495	1	F	Cliente A Depurar	1997-07-11	20394115.00	351
3552	1	M	Armando Calderon Lesmes	1945-07-25	27624743.00	384
3637	1	M	Jesus Maria Rodriguez Alvarez	1946-11-17	14860576.00	769
3735	1	M	Gonzalo Gaitan Quijano	1957-07-02	19320026.00	503
3796	1	M	Guillermo Ruiz Lara	1979-07-09	3758767.00	448
3821	1	M	Leonte Rojas Castro	1945-03-13	1339311.00	521
3826	1	M	Gustavo Moreno Camacho	1991-08-06	7663363.00	106
3882	1	M	Roberto Jose Escallon Ricaurte	1954-05-03	20730393.00	83
3898	1	M	Alberto Enrique Schotborgh Schot...	1988-03-01	14261834.00	399
4080	1	M	Jorge Abelardo Ayala Betancourt	1991-02-27	14182755.00	307
4091	1	M	Alejandro Dishington Gonzalez	1977-05-01	11719872.00	41
4202	1	M	Hector Jose Trujillo Cardenas	1997-07-15	7855208.00	790
4299	1	M	Alvaro Palacio Rudas	1976-09-12	26594201.00	896
4310	1	M	Fernando Corral Maldonado	2001-11-01	21645924.00	392
4479	1	M	Enrique Urdaneta Holguin	2002-06-13	22023005.00	621
4568	1	M	Romulo Gonzalez Trujillo	1981-03-21	7848604.00	848

Lo primero que se realizó fue capitalizar el nombre a mayúsculas, de esta manera es más sencillo crear las reglas de validación para dividirlo en nombres y apellidos.

```
-- Normalizar la columna nombre a mayúsculas y quitar espacios extra  
UPDATE clientes  
SET nombre = UPPER(LTRIM(RTRIM(nombre)));
```


Así nos queda

id	cod_tip_doc	cod_genero	nombre	fecha_nacimiento	ingresos	cod_sucursal
3422	1	M	ENRIQUE FLOREZ FAILLACE	1994-06-05	19983828.00	81
3486	1	M	CESAR ALBERTO CORTES DE LA ESPRIELLA	1996-06-23	3305365.00	320
3495	1	F	CLIENTE A DEPURAR	1997-07-11	20394115.00	351
3552	1	M	ARMANDO CALDERON LESMES	1945-07-25	27624743.00	384
3637	1	M	JESUS MARIA RODRIGUEZ ALVAREZ	1946-11-17	14860576.00	769
3735	1	M	GONZALO GAITAN QUIJANO	1957-07-02	19320026.00	503
3796	1	M	GUILLERMO RUIZ LARA	1979-07-09	3758767.00	448
3821	1	M	LEONTE ROJAS CASTRO	1945-03-13	1339311.00	521
3826	1	M	GUSTAVO MORENO CAMACHO	1991-08-06	7663363.00	106
3882	1	M	ROBERTO JOSE ESCALLON RICAURTE	1954-05-03	20730393.00	83
3898	1	M	ALBERTO ENRIQUE SCHOTBORGH SCHOTBORGH	1988-03-01	14261834.00	399
4080	1	M	JORGE ABELARDO AYALA BETANCOURT	1991-02-27	14182755.00	307
4091	1	M	ALEJANDRO DISHINGTON GONZALEZ	1977-05-01	11719872.00	41
4202	1	M	HECTOR JOSE TRUJILLO CARDENAS	1997-07-15	7855208.00	790
4299	1	M	ALVARO PALACIO RUDAS	1976-09-12	26594201.00	896
4310	1	M	FERNANDO CORRAL MALDONADO	2001-11-01	21645924.00	392
4479	1	M	ENRIQUE URDANETA HOLGUIN	2002-06-13	22023005.00	621
4568	1	M	ROMULO GONZALEZ TRUJILLO	1981-03-21	7848604.00	848

Ahora podemos crear las columnas que se nos piden.

Creamos las siguientes columnas vacías: nombres, apellidos, edad y es_empresa.

```
-- Se crean las columnas adicionales
ALTER TABLE clientes ADD
    nombres NVARCHAR(100),
    apellidos NVARCHAR(100),
    edad INT;

-- Crear columna temporal para detectar si es empresa
ALTER TABLE clientes ADD es_empresa BIT;

-- Se actualiza la columna de nombres y apellidos según número de palabras. La lógica es la siguiente:
-- 1. Si el nombre tiene cuatro palabras o más, se considera que las primeras dos son nombres y las últimas dos apellidos.
-- 2. Si el nombre tiene tres palabras, se toma la primera como nombre y las otras dos como apellidos.
-- 3. Si el nombre tiene dos palabras, la primera es nombre y la segunda apellido.
```

Messages

Commands completed successfully.

Completion time: 2025-07-07T01:27:24.8882798-05:00

Para la columna **es_empresa** se definieron algunas palabras o caracteres clave que tienen algunos registros que pertenecen a empresas o pymes.

```
-- Marcar como empresa si contiene palabras clave o caracteres especiales
UPDATE clientes
SET es_empresa = CASE
    -- Si contiene indicios de empresa Y NO contiene ciertos apellidos/personas
    WHEN (
        nombre LIKE '% SAS%'
        OR nombre LIKE '% S.A.S%'
        OR nombre LIKE '% S.A%'
        OR nombre LIKE '% LTDA%'
        OR nombre LIKE '% LTD%'
        OR nombre LIKE '% CIA%'
        OR nombre LIKE '% S. EN C%'
        OR nombre LIKE '% &%'
        OR nombre LIKE '%.%'
        OR nombre LIKE '%[0-9]%'
    )
    AND nombre NOT LIKE '% SASTOQUE%'
    AND nombre NOT LIKE '% SASSO%'
    AND nombre NOT LIKE '% SASCHA%'
    AND nombre NOT LIKE '% SASSON%'
    AND nombre NOT LIKE '% CIANCI%'
    THEN 1
    ELSE 0
END;
```

Según mi query hay aproximadamente 1167 empresas en el dataset.

Results

Messages

	id	cod_tin_doc	cod_genero	nombre	fecha_nacimiento	ingresos	cod_sucursal	nombres	apellidos	edad	es_empresa
1	800085853	3	N	JUNTA DE ACCION COMUNAL DEL BARRIO 31 DE MARZO	2018-02-17	29865618.00	765	NULL	NULL	NULL	1
2	800091336	3	N	PROEZA CONSULTORES SAS	1992-12-16	31612034.00	124	NULL	NULL	NULL	1
3	800099478	3	N	JAVIER L VILLEGAS P Y CIA S EN C S	1952-01-01	29473775.00	762	NULL	NULL	NULL	1
4	800211079	3	N	EDIFICIO NEPTUNO 1	2017-07-16	11981241.00	672	NULL	NULL	NULL	1
5	800225529	3	N	DISEÑOS EXCLUSIVOS Y CIA LTDA	1966-04-12	29613740.00	501	NULL	NULL	NULL	1
6	800236451	3	N	INVERSIONES PUINAL LTDA	1996-03-12	27815622.00	878	NULL	NULL	NULL	1
7	801007275	3	N	TELECOMUNICACIONES Y SERVICIOS LTDA	1960-08-12	23854117.00	616	NULL	NULL	NULL	1
8	802025277	3	N	U T PANAMER DREDGING AND SURVEYS SERV DE DRAG Y C...	1945-07-24	19924481.00	845	NULL	NULL	NULL	1
9	804004915	3	N	SODIMAN SAS	1955-07-19	1650402.00	129	NULL	NULL	NULL	1
10	804016761	3	N	COMPANY SEGUROS Y CIA LTDA	2018-10-29	5590612.00	696	NULL	NULL	NULL	1
11	805009143	3	N	CONJ. RESIDENCIAL CARAMANTA PROPIEDAD HORIZONTAL	1986-04-02	29252361.00	459	NULL	NULL	NULL	1
12	805023365	3	N	B E B COMERCIAL LTDA	1960-11-17	15137478.00	919	NULL	NULL	NULL	1
13	805024564	3	N	NEGOCIOS C DE BOTERO & COMPAÑIA S EN C	1961-06-19	15407312.00	88	NULL	NULL	NULL	1
14	805028169	3	N	JHON KELLY JEANS SAS	1990-01-19	28236298.00	209	NULL	NULL	NULL	1
15	806005558	3	N	PESICAR LTDA	1979-10-21	14943692.00	689	NULL	NULL	NULL	1
16	806013901	3	N	AIRE INDUSTRIAL DEL CARIBE LTDA	2005-01-02	29994501.00	631	NULL	NULL	NULL	1
17	807010061	3	N	COMERCIALIZADORA INTERNACIONAL ENIGMA DE MODA LT...	1979-06-23	30595889.00	572	NULL	NULL	NULL	1
18	810007297	3	N	CONSTRUCTORA MANIZALES SAS	1995-06-02	6359190.00	28	NULL	NULL	NULL	1
19	811004826	3	N	EDIFICIO LAS FUENTES NRO 5	2003-06-19	1088766.00	923	NULL	NULL	NULL	1
20	811013633	3	N	C & C LLANTAS S A	1998-10-09	22685993.00	78	NULL	NULL	NULL	1
21	811018041	3	N	M S CONSTRUCCIONES SAS	1997-01-26	21760982.00	247	NULL	NULL	NULL	1
22	811027385	3	N	C.E.I. CONSTRUCCIONES E INTERVENTORIA E.A.T.	1991-07-24	15170389.00	600	NULL	NULL	NULL	1
23	811027481	3	N	GRUPO CARVACARO SAS	1979-02-09	10385489.00	750	NULL	NULL	NULL	1
24	811029672	3	N	ELECTROENLACES S.A.S	2013-08-09	8456171.00	336	NULL	NULL	NULL	1

Query executed successfully.

SIMON\SQLXPRESS (16.0 RTM) SIMON\simon (51) prueba_analitica_bc 00:00:00 1,167 rows

Para rellenar las columnas nombres y apellidos se tuvieron en cuenta varias consideraciones:

- Si es empresa solo se llena el nombre y el apellido se pone como NULL.
- Si el nombre tiene cuatro palabras o más, se considera que las primeras dos son nombres y las últimas dos apellidos.
- Si el nombre tiene tres palabras, se toma la primera como nombre y las otras dos como apellidos.
- Si el nombre tiene dos palabras, la primera es nombre y la segunda apellido.

```
-- Lógica para nombres y apellidos (Si es empresa solo queda el nombre)
UPDATE clientes
SET
  nombres = CASE
    WHEN es_empresa = 1 THEN nombre -- Si es empresa, se guarda completo en nombres
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) >= 3 THEN
      CONCAT(PARSENAME(REPLACE(nombre, ' ', '.'), 4), ' ', PARSENAME(REPLACE(nombre, ' ', '.'), 3))
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) = 2 THEN
      PARSENAME(REPLACE(nombre, ' ', '.'), 3)
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) = 1 THEN
      PARSENAME(REPLACE(nombre, ' ', '.'), 2)
    ELSE NULL
  END,
  apellidos = CASE
    WHEN es_empresa = 1 THEN NULL -- Si es empresa, no se llena el campo apellidos
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) >= 3 THEN
      CONCAT(PARSENAME(REPLACE(nombre, ' ', '.'), 2), ' ', PARSENAME(REPLACE(nombre, ' ', '.'), 1))
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) = 2 THEN
      CONCAT(PARSENAME(REPLACE(nombre, ' ', '.'), 2), ' ', PARSENAME(REPLACE(nombre, ' ', '.'), 1))
    WHEN LEN(nombre) - LEN(REPLACE(nombre, ' ', '')) = 1 THEN
      PARSENAME(REPLACE(nombre, ' ', '.'), 1)
    ELSE NULL
  END;

-- Calcular la edad con base en la fecha de nacimiento
```

(42914 rows affected)

Completion time: 2025-07-07T01:40:42.3121511-05:00

Luego, para calcular la edad del cliente se resto le fecha de hoy y la fecha de nacimiento del cliente.

```
-- Calcular la edad con base en la fecha de nacimiento
UPDATE clientes
SET edad = DATEDIFF(YEAR, fecha_nacimiento, GETDATE());
```

Messages

(42914 rows affected)

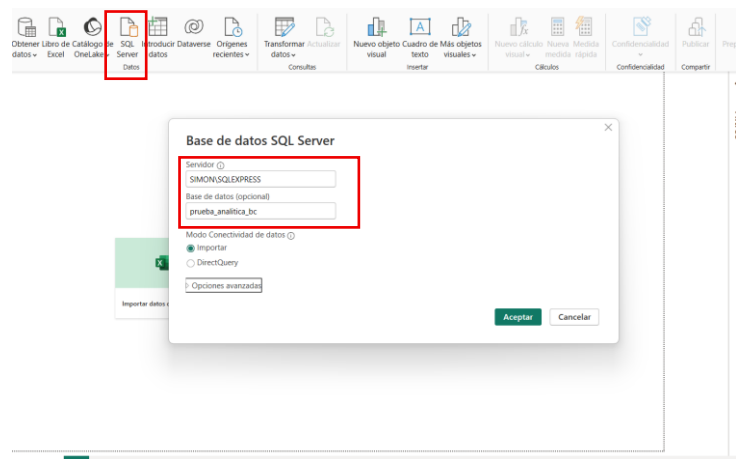
Completion time: 2025-07-07T01:53:36.5039968-05:00

Este sería el resultado final con los nuevos campos agregados.

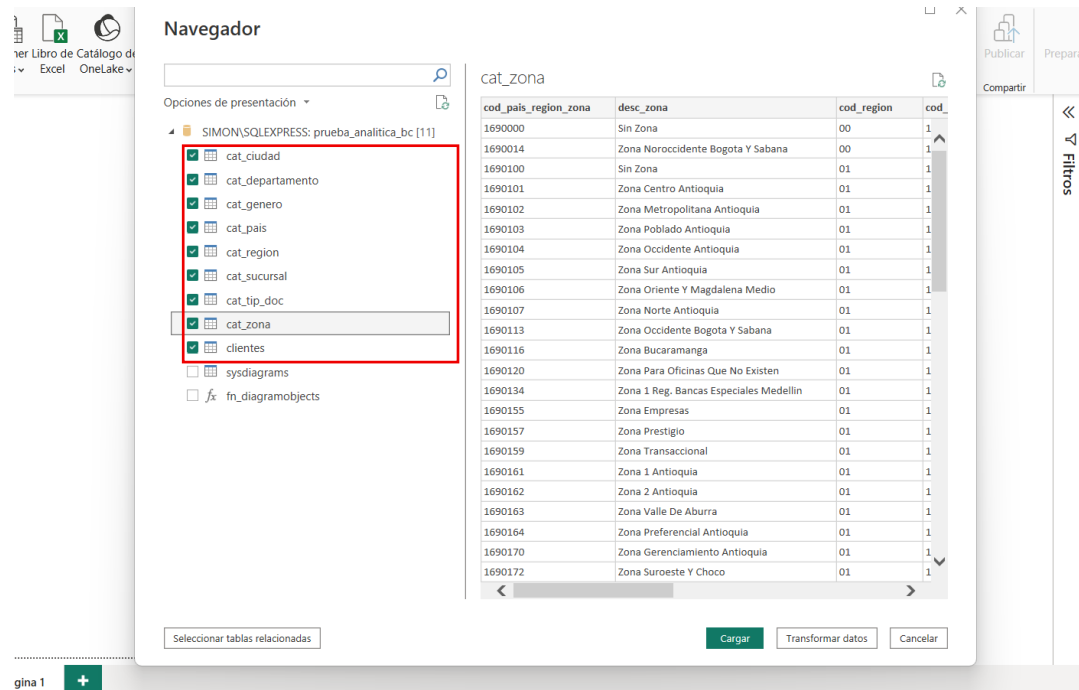
Results Messages											
	id	cod_tip_doc	cod_genero	nombre	fecha_nacimiento	ingresos	cod_sucursal	nombres	apellidos	edad	es_empresa
7	3796	1	M	GUILLERMO RUIZ LARA	1979-07-09	3758767.00	448	GUILLERMO	RUIZ LARA	46	0
8	3821	1	M	LEONTE ROJAS CASTRO	1945-03-13	1339311.00	521	LEONTE	ROJAS CASTRO	80	0
9	3826	1	M	GUSTAVO MORENO CAMACHO	1991-08-06	7663363.00	106	GUSTAVO	MORENO CAMACHO	34	0
10	3882	1	M	ROBERTO JOSE ESCALLON RICAURTE	1954-05-03	20730393.00	83	ROBERTO JOSE	ESCALLON RICAURTE	71	0
11	3898	1	M	ALBERTO ENRIQUE SCHOTBORGH SCHOTBORGH	1988-03-01	14261834.00	399	ALBERTO ENRIQUE	SCHOTBORGH SCHOTBORGH	37	0
12	4080	1	M	JORGE ABELARDO AYALA BETANCOURT	1991-02-27	14182755.00	307	JORGE ABELARDO	AYALA BETANCOURT	34	0
13	4091	1	M	ALEJANDRO DISHINGTON GONZALEZ	1977-05-01	11719872.00	41	ALEJANDRO	DISHINGTON GONZALEZ	48	0
14	4202	1	M	HECTOR JOSE TRUJILLO CARDENAS	1997-07-15	7855208.00	790	HECTOR JOSE	TRUJILLO CARDENAS	28	0
15	4299	1	M	ALVARO PALACIO RUDAS	1976-09-12	26594201.00	896	ALVARO	PALACIO RUDAS	49	0
16	4310	1	M	FERNANDO CORRAL MALDONADO	2001-11-01	21645924.00	392	FERNANDO	CORRAL MALDONADO	24	0
17	4479	1	M	ENRIQUE URDANETA HOLGUIN	2002-06-13	22023005.00	621	ENRIQUE	URDANETA HOLGUIN	23	0
18	4568	1	M	ROMULO GONZALEZ TRUJILLO	1981-03-21	7848604.00	848	ROMULO	GONZALEZ TRUJILLO	44	0
19	4607	1	M	FERNANDO HENAO SAENZ	1980-01-21	29124501.00	488	FERNANDO	HENAO SAENZ	45	0
20	4858	1	M	ROSENDO TORRES CRUZ	1944-07-01	4374787.00	902	ROSENDO	TORRES CRUZ	81	0
21	4952	2	M	LUIS AUGUSTO MAZARIEGOS GODOY	1950-05-29	29798136.00	298	LUIS AUGUSTO	MAZARIEGOS GODOY	75	0
22	5228	1	M	ALFONSO LEMUS	1949-06-11	28443446.00	781	ALFONSO	LEMUS	76	0

6. Se quiere hacer una campaña de tarjeta de crédito para estos clientes, los directivos de la organización quieren saber algunos datos para dirigir mejor la campaña, realice las consultas y muéstrelas de forma visual para presentárselas (En caso de encontrar datos atípicos, incoherentes o faltantes realice el tratamiento que considere pertinentes y explique).

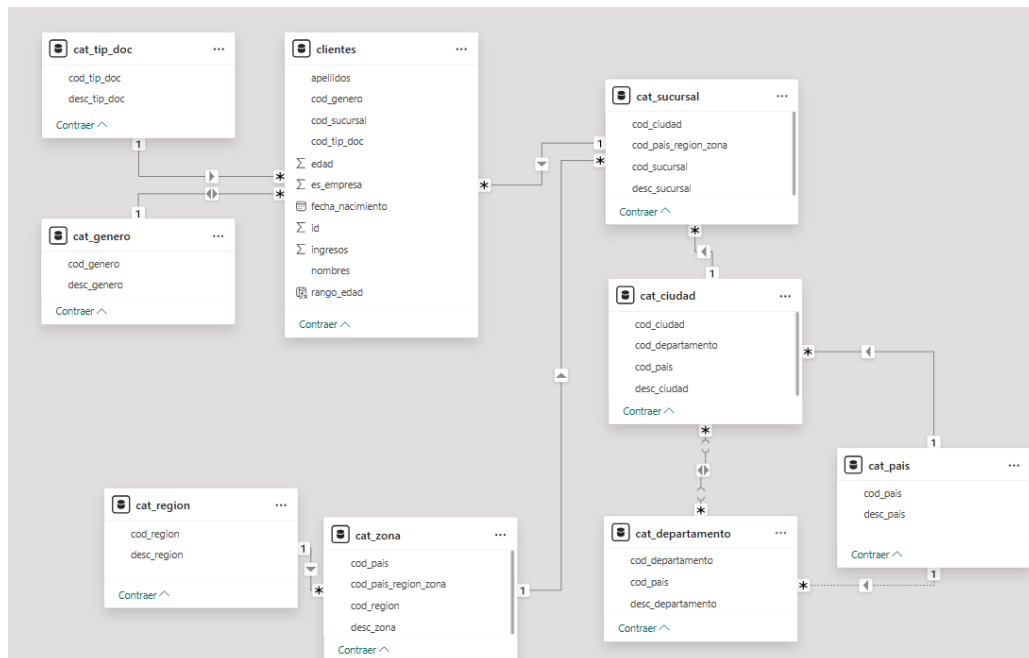
El proceso de consultas y visualización se realizó en Power Bi. A continuación, una breve explicación de la conexión a la base de datos y lectura de datos. Nos conectamos desde Power Bi a SSMS e ingresamos el server y el nombre de la BD.



Elegimos las tablas y las importamos al modelo de Power BI.



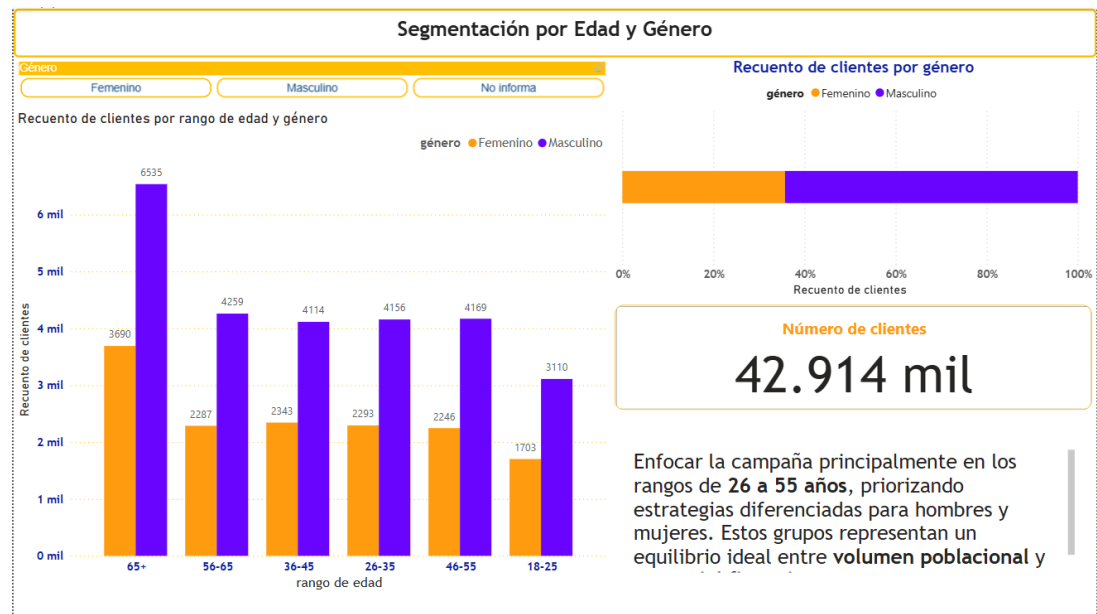
Modelo de datos en Power BI.



Nota: El tablero se encuentra en GitHub también en la carpeta dashboard.

- a. **Identifique los rangos de edad a los que se les podría dirigir la campaña para tener un mayor público objetivo, agrupados por género “Masculino” y “Femenino”.**

Realicé un diagrama de barras para contar cuantos clientes hay por rango de edad y por género. Un filtro para poder ver los datos por género, también un recuento de clientes por genero para ver el porcentaje de hombre y mujeres en el dataset y un card que muestra el total de clientes.



Rangos de edad recomendados para enfocar la campaña (segmentación por edad y género)

- El grupo 65+ tiene la mayor cantidad de clientes (más de 10.000 en total, combinando géneros). Sin embargo, los grupos 36-45, 46-55 y 26-35 también presentan una alta concentración de clientes, en especial del género masculino.

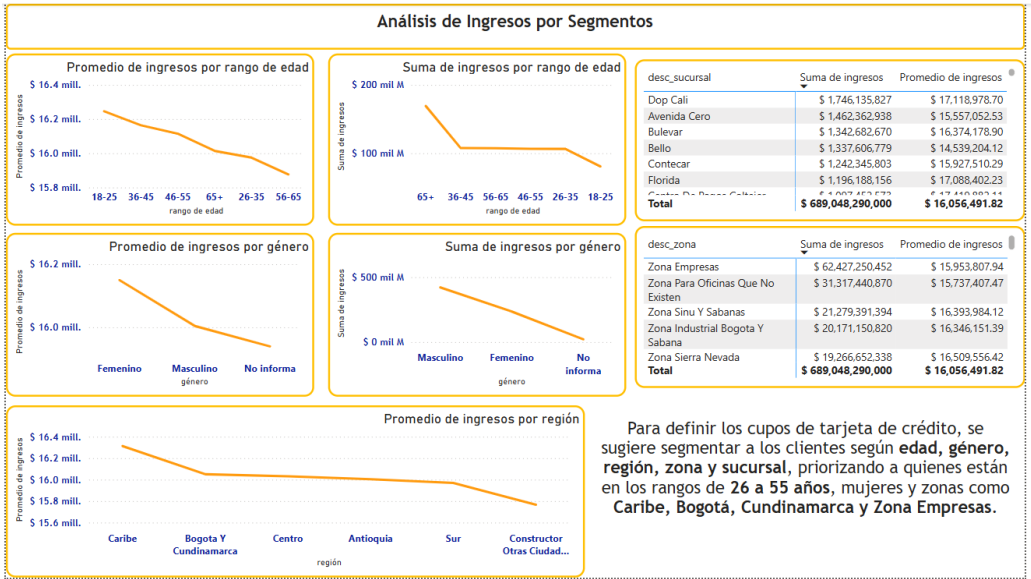
¿Cuales son entonces los segmentos ideales para la campaña?

- Aunque el grupo 65+ tiene más registros, no necesariamente es el público más activo para tarjetas de crédito. Por lo general, los productos de este tipo están más alineados con personas entre 26 y 55 años, ya que tienen ingresos más estables, están en etapa de consumo activo, suelen tener responsabilidades financieras (hogar, hijos, negocios, etc.).

Género predominante: Se observa una mayoría significativa del género masculino en todos los rangos, por lo que la campaña debe ser diseñada con enfoque masculino, sin descuidar al público femenino (especialmente en 65+ y 56-65).

- b. Se quiere saber cuáles son los Ingresos totales y promedios de los clientes, agrúpelas por las variables que considere sea relevante para definir un aproximado de los cupos que se podrían otorgar en esta campaña, comente que otras variables que no están en la base de datos pueden servir para discriminar mejor los ingresos.

Realicé gráficos de línea ya que es la manera mas práctica de mostrar un análisis de ingresos en este caso, lo hice tanto para promedios como para suma de ingresos discriminado por varias variables como rango de edad, género, región. Además creé dos matrices para desplegar los mismos datos pero para sucursales y zonas.



- El promedio de ingresos más alto está en el grupo de 18-25 años (aproximadamente \$16,4 millones), lo cual puede deberse a algunos valores atípicos (por ejemplo, pocos jóvenes con altos ingresos).
- Los grupos de 26 a 65 años tienen promedios ligeramente menores, pero con sumas totales más altas, reflejando una base de clientes más amplia y estable.
- El grupo 65+ representa la mayor suma total de ingresos, probablemente por tener más clientes, aunque con ingresos moderados.

Por Género: Las mujeres tienen mayores ingresos promedio que los hombres (aprox \$16,2M vs aprox \$16,0M). Sin embargo, los hombres suman más ingresos totales, lo que indica un mayor número de clientes masculinos.

Por Región: La región Caribe lidera en promedio de ingresos (aprox \$16,4M). Le siguen Bogotá y Cundinamarca, y Centro, lo cual sugiere un enfoque fuerte en estas zonas por potencial económico.

Por Zona y Sucursal: La Zona Empresas y la sucursal Dop Cali lideran tanto en suma como en promedio de ingresos. Estas áreas podrían recibir límites de cupo más altos y estrategias de fidelización premium.

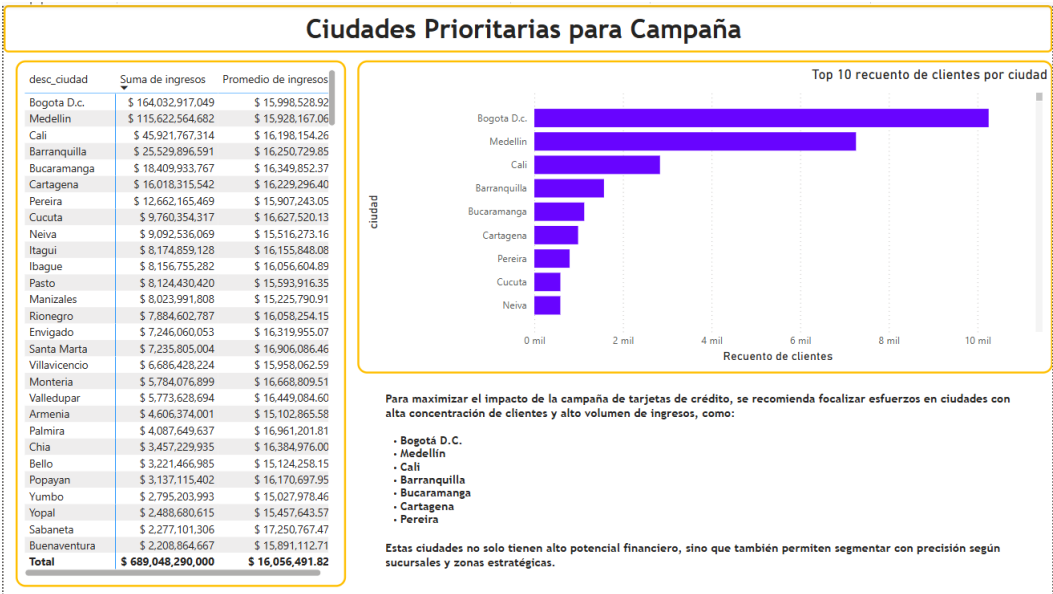
Variables adicionales sugeridas (no disponibles en la base):

- Ocupación o tipo de empleo (ej. independiente, empleado público, estudiante).
- Nivel educativo.
- Historial crediticio / Score de riesgo.
- Número de productos financieros activos.
- Comportamiento de pago / Mora histórica.
- Estado civil y dependientes económicos.
- Ciudad o estrato socioeconómico más detallado.

Estas variables nos permitirían una discriminación más precisa del perfil económico, reduciendo riesgos en la asignación de cupos y mejorando la eficacia de la campaña.

c. Muestre las ciudades que considera se debería enfocar la campaña.

Para el punto C, realicé una tabla con la ciudad, su suma de ingresos y promedio de ingresos agrupado. También un chart que indica el top 10 ciudades por volumen de clientes. Esto ayuda a decidir que ciudades priorizar teniendo en cuenta volumen de clientes e ingresos.



Top Ciudades por Suma de Ingresos

- Bogotá D.C. lidera ampliamente con más de \$164 mil millones en ingresos.
- Le siguen Medellín y Cali, con ingresos superiores a \$45 mil millones.

- Otras ciudades destacadas por su volumen económico incluyen Barranquilla, Bucaramanga, y Cartagena.

Top Ciudades por Número de Clientes:

- Bogotá y Medellín también encabezan el recuento de clientes, lo que sugiere alto potencial de penetración y retorno en estas ciudades.
- Las siguientes ciudades más densas en clientes son: Cali, Barranquilla, Bucaramanga, Cartagena, y Pereira.

- d. **Se quiere saber si hay clientes de la zona “Zona Atlántico” que vivan en la ciudad “Barranquilla” y qué clientes son, ya que a estos clientes se les va a dar unas condiciones especiales.**

Se realizó una tabla filtrada con información general de clientes a los que se les van a dar condiciones especiales. Filtros: ciudad Barranquilla, Zona Atlantico.

Clientes especiales en Zona Atlántico y Barranquilla									
id	tipo doc	ciudad	zona	nombres	apellidos	edad	género	ingresos	sucursal
5398	C.C	Barranquilla	Zona Atlantico	BADIH MERHEG	MERHEG RAUCHADI	49	Masculino	\$ 16.871.719	Avenida Kennedy
7635	C.C	Barranquilla	Zona Atlantico	JAIME ANTONIO	MARQUEZ CUERVO	28	Masculino	\$ 23.992.001	Avenida Kennedy
10736	C.C	Barranquilla	Zona Atlantico	ALVARO SALAZAR	SALAZAR VELASCO	65	Masculino	\$ 15.142.183	Avenida Kennedy
14868	C.C	Barranquilla	Zona Atlantico	JAIME CORREDOR	CORREDOR ARJONA	46	Masculino	\$ 22.349.175	Avenida Kennedy
34510	C.C	Barranquilla	Zona Atlantico	LUIS HUMBERTO	VALDERRAMA NUÑEZ	21	Masculino	\$ 1.295.181	Avenida Kennedy
51379	C.C	Barranquilla	Zona Atlantico	JOSE JOAQUIN	GOMEZ AFANADOR	79	Masculino	\$ 14.520.353	Avenida Kennedy
67470	C.C	Barranquilla	Zona Atlantico	JOSE MARIO	GOMEZ RUEDA	78	Masculino	\$ 14.045.341	Avenida Kennedy
69345	C.C	Barranquilla	Zona Atlantico	PANTALEO MASSARO	MASSARO CICARDI	27	Masculino	\$ 25.733.598	Avenida Kennedy
73358	C.C	Barranquilla	Zona Atlantico	ENRIQUE OLIVERIO	LONDOÑO SANCHEZ	56	Masculino	\$ 27.737.245	Avenida Kennedy
84461	C.C	Barranquilla	Zona Atlantico	JUAN SUSAIETA	SUSAIETA OGUETA	31	Masculino	\$ 11.934.792	Avenida Kennedy
108747	C.C	Barranquilla	Zona Atlantico	ALICIA MARTIN DE CURE	DE CURE	69	Femenino	\$ 6.847.006	Avenida Kennedy
115838	C.C	Barranquilla	Zona Atlantico	JORGE FORCADAS	FORCADAS FELIU	65	Masculino	\$ 25.491.542	Avenida Kennedy
118875	C.C	Barranquilla	Zona Atlantico	JOSE ANTONIO	PEDRAZA PICON	75	Masculino	\$ 19.403.125	Avenida Kennedy
121848	C.C	Barranquilla	Zona Atlantico	JAIME ENRIQUE	VILLAMIL LEON	27	Masculino	\$ 6.386.328	Avenida Kennedy
157545	C.C	Barranquilla	Zona Atlantico	PANAGULA FAGRESOS DE MINERVINI	DE MINERVINI	45	Femenino	\$ 6.849.398	Avenida Kennedy
162806	C.C	Barranquilla	Zona Atlantico	JUAN ALBERTO	POLO FIGUEROA	68	Masculino	\$ 7.567.541	Avenida Kennedy
163478	C.C	Barranquilla	Zona Atlantico	CELIMO GONZALEZ	GONZALEZ RECIO	53	Masculino	\$ 3.284.281	Avenida Kennedy
165492	C.C	Barranquilla	Zona Atlantico	ALFREDO GARCIA	GARCIA ORDOÑEZ	51	Masculino	\$ 10.368.729	Avenida Kennedy
180605	C.C	Barranquilla	Zona Atlantico	JAIME ROMERO	ROMERO MELO	38	Masculino	\$ 22.996.121	Avenida Kennedy
197259	C.C	Barranquilla	Zona Atlantico	JOSE HIPOLITO	NEUTA PEÑA	31	Masculino	\$ 19.064.557	Avenida Kennedy
331597	C.C	Barranquilla	Zona Atlantico	CARLOS HECTOR	DIAZ CARRERO	38	Masculino	\$ 2.974.501	Avenida Kennedy
463710	C.C	Barranquilla	Zona Atlantico	JUANA ESPINOSA	ESPINOSA AGUILAR	59	Femenino	\$ 7.476.943	Avenida Kennedy
504037	C.C	Barranquilla	Zona Atlantico	EGBERTO OSPINA	OSPINA BARRIOS	42	Masculino	\$ 16.506.679	Avenida Kennedy
504849	C.C	Barranquilla	Zona Atlantico	ALFONSO URIBE	URIBE MELGUIZO	53	Masculino	\$ 26.272.573	Avenida Kennedy
Total	C.C	Barranquilla	Zona Atlantico	TRES SESENTA ESPECIALISTAS EN GESTION DE RIESGOS	ABDALA SERNA	42062	Femenino	\$ 14.245.397.420	Avenida Kennedy

Clientes que habitan en la Zona Atlántico, específicamente en Barranquilla y su información general que incluye: Tipo de documento, nombres, apellidos, edad, género, ingresos totales mensuales y su sucursal.

7. (opcional) Un comercial reportó una alerta de fraude para un cliente de nombre Juan Gomez, sin embargo no presentó mucha información al respecto, genere un listado con toda la información (ciudad, región, sucursal, zona, país, nit, nombre completo, fecha_nacimiento, ingresos) de los clientes que tienen este nombre para presentársela al comercial y que él pueda identificar quién es el cliente que presenta la alerta. En caso de realizar este punto adicione al documento el procedimiento para solucionarlo.

Se realizó una tabla filtrando por nombres y apellidos de tal manera que el nombre contenga la palabra “Juan” y el apellido “Gomez”.

Alerta de Fraude - Cliente con nombre Juan Gomez									
id	tipo doc	ciudad	zona	nombres	apellidos	edad	género	ingresos	sucursal
94291131	C.C	A.lopez P.	Sin Zona	JUAN ESTEBAN	GOMEZ MONTOYA	69	Masculino	\$ 30,982,807	20 De Julio
1082809731	C.C	A.lopez P.	Sin Zona	JUAN GABRIEL	TAFUR GOMEZ	33	Masculino	\$ 30,866,669	20 De Julio
1118239966	C.C	A.lopez P.	Sin Zona	JUAN DAVID	GOMEZ VALENCIA	38	Masculino	\$ 29,888,902	20 De Julio
2929502	C.C	A.lopez P.	Sin Zona	JUAN FRANCISCO	GUTIERREZ GOMEZ	75	Masculino	\$ 29,875,129	20 De Julio
3319728	C.C	A.lopez P.	Sin Zona	JUAN GUILLERMO	MESA GOMEZ	55	Masculino	\$ 29,612,492	20 De Julio
1120583562	C.C	A.lopez P.	Sin Zona	JUAN JACOBO	OSORIO GOMEZ	41	Masculino	\$ 29,117,903	20 De Julio
1045026592	C.C	A.lopez P.	Sin Zona	JUAN DAVID	GIRALDO GOMEZ	47	Masculino	\$ 27,998,374	20 De Julio
3611490	C.C	A.lopez P.	Sin Zona	JUAN NEPOMUCENO	ZULUAGA GOMEZ	56	Masculino	\$ 27,263,314	20 De Julio
1764596	C.C	A.lopez P.	Sin Zona	JUAN JOSE	GARCIA GOMEZ	56	Masculino	\$ 24,980,169	20 De Julio
99121010720	C.C	A.lopez P.	Sin Zona	JUAN PABLO	BURBANO GOMEZ	17	Masculino	\$ 23,451,425	20 De Julio
91075156	C.C	A.lopez P.	Sin Zona	JUAN JOSE	VARGAS GOMEZ	76	Masculino	\$ 22,258,932	20 De Julio
3438221	C.C	A.lopez P.	Sin Zona	JUAN CARLOS	GOMEZ MESA	33	Masculino	\$ 21,800,746	20 De Julio
1090423618	C.C	A.lopez P.	Sin Zona	CRISTIAN JUAN	GOMEZ VECINO	66	Masculino	\$ 16,428,536	20 De Julio
1088326042	C.C	A.lopez P.	Sin Zona	JUAN NICOLAS	GOMEZ URIBE	75	Masculino	\$ 14,603,774	20 De Julio
5018011544	C.C	A.lopez P.	Sin Zona	JUAN MANUEL	GONZALEZ GOMEZ	56	Masculino	\$ 14,047,487	20 De Julio
94535748	C.C	A.lopez P.	Sin Zona	JUAN CARLOS	MORALES GOMEZ	81	Masculino	\$ 13,076,537	20 De Julio
1083929113	C.C	A.lopez P.	Sin Zona	JUAN DAVID	GOMEZ ALDANA	37	Masculino	\$ 12,832,596	20 De Julio
4277442	C.C	A.lopez P.	Sin Zona	JUAN CARLOS	GOMEZ MORA	78	Masculino	\$ 12,150,260	20 De Julio
71228379	C.C	A.lopez P.	Sin Zona	JUAN CARLOS	URIBE GOMEZ	73	Masculino	\$ 11,944,044	20 De Julio
1092541330	C.C	A.lopez P.	Sin Zona	JUAN ESTEBAN	CUBAQUE GOMEZ	54	Masculino	\$ 10,002,363	20 De Julio
1059816422	C.C	A.lopez P.	Sin Zona	JUAN MANUEL	RODRIGUEZ GOMEZ	20	Masculino	\$ 9,911,774	20 De Julio
1053618249	C.C	A.lopez P.	Sin Zona	JUAN PABLO	GOMEZ RODRIGUEZ	69	Masculino	\$ 8,563,248	20 De Julio
1017191951	C.C	A.lopez P.	Sin Zona	JUAN CAMILO	GOMEZ OSORIO	53	Masculino	\$ 7,178,933	20 De Julio
3353955	C.C	A.lopez P.	Sin Zona	JUAN DE DIOS GOMEZ ARISTIZABAL	GOMEZ ARISTIZABAL	56	Masculino	\$ 7,054,762	20 De Julio
Total	C.C	A.lopez P.	Sin Zona	CRISTIAN JUAN	BURBANO GOMEZ	1637	Masculino	\$ 482,338,450	20 De Julio

Registros para presentar al director relacionados con una alerta de fraude para un cliente de nombre Juan Gomez. Se filtra por nombres que contengan la palabra "Juan" y apellidos que contengan la palabra "Gomez" y se obtienen los siguientes registros.

Filtros

Buscar

Filtros de esta página

apellidos

contiene 'GOMEZ'

nombres

contiene 'JUAN'

Agregar campos de datos ...

Filtros de todas las páginas

Agregar campos de datos ...