

Universidad Nacional Mayor de San Marcos

FACULTAD DE CIENCIAS MATEMÁTICAS

E.P. COMPUTACIÓN CIENTÍFICA



REDES, ARQUITECTURA Y COMUNICACIONES

Documentación de Arquitectura y Flujo de Vida de Datos

Estudiante:

Gallardo Vargas, Ruben Armando

Fecha:

04 - 12 - 2024

Contents

1	Introducción	2
2	Descripción General del Sistema	2
3	Arquitectura del Sistema	2
3.1	Diagrama de Arquitectura	2
3.2	Componentes	2
4	Flujo de Vida de los Datos	2
5	Implementación Técnica	3
5.1	Creación de Tablas en SQL Server	3
5.2	Importación de Datos desde CSV	5
5.3	Conexión a Power BI	11
6	Visualización en Power BI	14
6.1	Construcción de datos	14
6.2	Tipos de Gráficos Implementados	16
6.3	Reporte Realizado	16
7	Conclusiones	16

1 Introducción

Este documento describe la arquitectura del sistema implementado para la carga, transformación y análisis de datos desde archivos CSV a través de SQL Server y su integración con Power BI. También detalla el flujo de vida de los datos desde su origen hasta la visualización final.

2 Descripción General del Sistema

El sistema consta de tres componentes principales:

- **Origen de Datos:** Archivos CSV con datos estructurados.
- **Base de Datos:** SQL Server como herramienta de almacenamiento y transformación.
- **Visualización:** Power BI para la generación de gráficos interactivos.

3 Arquitectura del Sistema

3.1 Diagrama de Arquitectura



3.2 Componentes

1. **Archivos CSV:** Se utilizan como fuente de datos externa. Cada archivo contiene datos delimitados por comas.
2. **SQL Server:** Gestiona las tablas donde se importan los datos, permitiendo realizar consultas y transformaciones.
3. **Power BI:** Carga los datos procesados desde SQL Server para crear visualizaciones dinámicas.

4 Flujo de Vida de los Datos

El flujo de los datos sigue las siguientes etapas:

1. **Ingesta:** Los archivos CSV se cargan en SQL Server mediante scripts de importación.
2. **Transformación:** Dentro de SQL Server, se realizan procesos como normalización, limpieza y preparación de los datos.
3. **Carga:** Los datos procesados se conectan a Power BI mediante un conector ODBC o similar.
4. **Visualización:** Los gráficos y reportes interactivos son generados en Power BI.

5 Implementación Técnica

5.1 Creación de Tablas en SQL Server

Códigos SQL usados para crear las tablas respectivas:

```
-- Tabla de Dimensión para Home Ownership
CREATE TABLE Dim_Person_Home_Ownership (
    home_ownership_id INT PRIMARY KEY IDENTITY(1,1),
    home_ownership VARCHAR(50) NOT NULL
);
```

```
-- Tabla de Dimensión para Loan Intent
CREATE TABLE Dim_Loan_Intent (
    loan_intent_id INT PRIMARY KEY IDENTITY(1,1),
    loan_intent VARCHAR(50) NOT NULL
);
```

```
-- Tabla de Dimensión para Loan Grade
CREATE TABLE Dim_Loan_Grade (
    loan_grade_id INT PRIMARY KEY IDENTITY(1,1),
    loan_grade CHAR(1) NOT NULL
);
```

```
-- Tabla de Dimensión para Loan Status
CREATE TABLE Dim_Loan_Status (
    loan_status_id INT PRIMARY KEY IDENTITY(1,1),
    loan_status INT NOT NULL
);
```

```
-- Tabla de Dimensión para Persona
CREATE TABLE Dim_Person (
    person_id INT PRIMARY KEY IDENTITY(1,1),
    person_age INT NOT NULL,
    person_income INT NOT NULL,
    person_emp_length FLOAT NULL,
    name VARCHAR(100) NOT NULL,
    cb_person_default_on_file CHAR(1) NOT NULL,
```

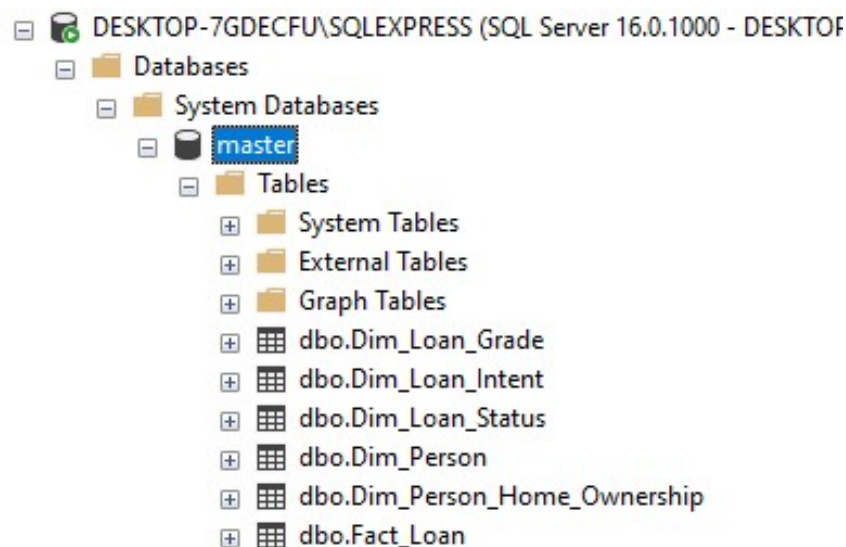
```

        cb_person_cred_hist_length INT NOT NULL
    );

-- Tabla de Hechos para los Préstamos
CREATE TABLE Fact_Loan (
    loan_id INT PRIMARY KEY IDENTITY(1,1),
    person_id INT NOT NULL,
    home_ownership_id INT NOT NULL,
    loan_intent_id INT NOT NULL,
    loan_grade_id INT NOT NULL,
    loan_status_id INT NOT NULL,
    loan_amnt INT NOT NULL,
    loan_int_rate FLOAT NULL,
    loan_percent_income FLOAT NOT NULL,
    FOREIGN KEY (person_id) REFERENCES Dim_Person(person_id),
    FOREIGN KEY (home_ownership_id) REFERENCES Dim_Person_Home_Ownership(home_ownersh
    FOREIGN KEY (loan_intent_id) REFERENCES Dim_Loan_Intent(loan_intent_id),
    FOREIGN KEY (loan_grade_id) REFERENCES Dim_Loan_Grade(loan_grade_id),
    FOREIGN KEY (loan_status_id) REFERENCES Dim_Loan_Status(loan_status_id)
);

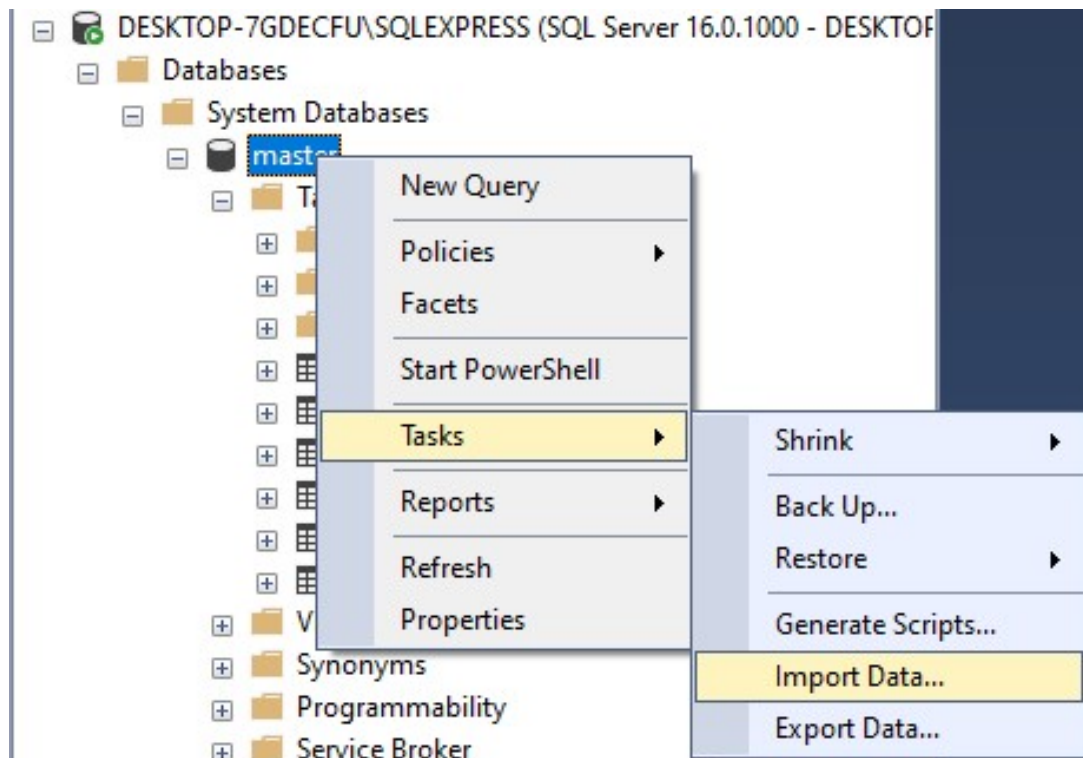
```

Al finalizar podremos observar las tablas creadas:

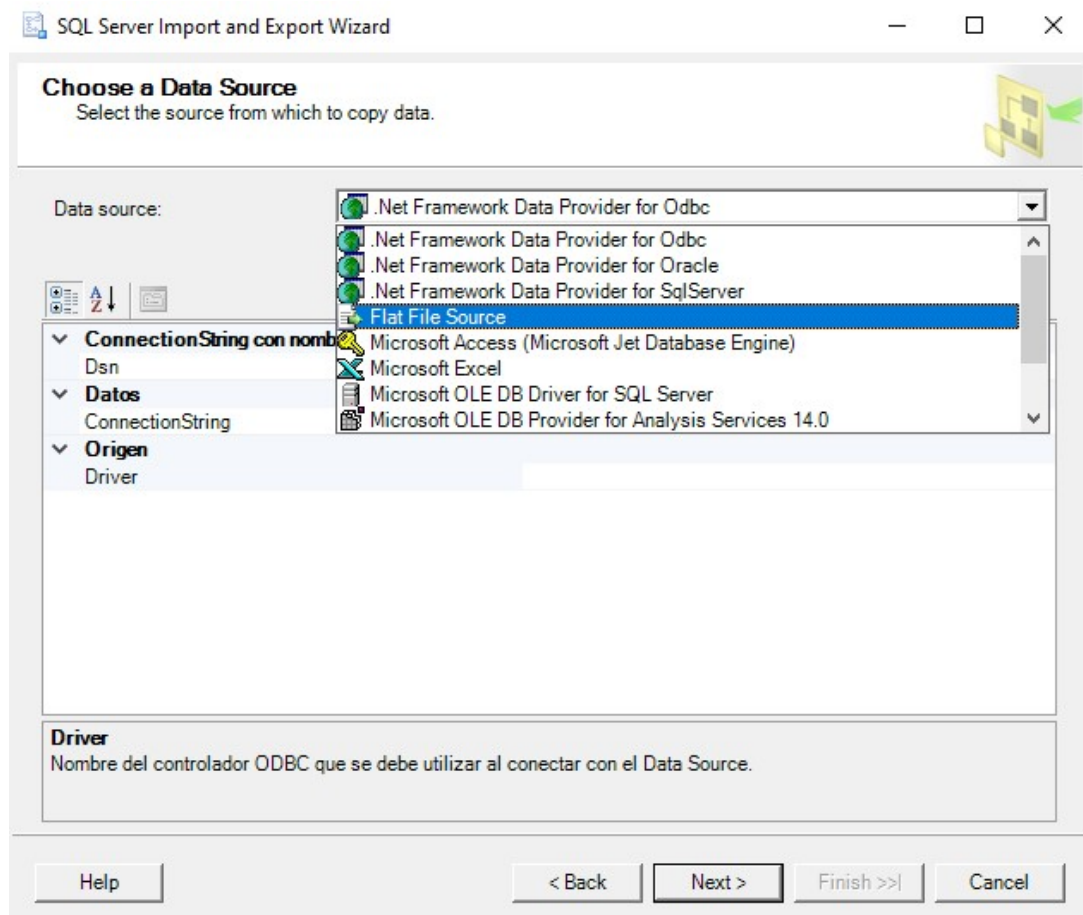


5.2 Importación de Datos desde CSV

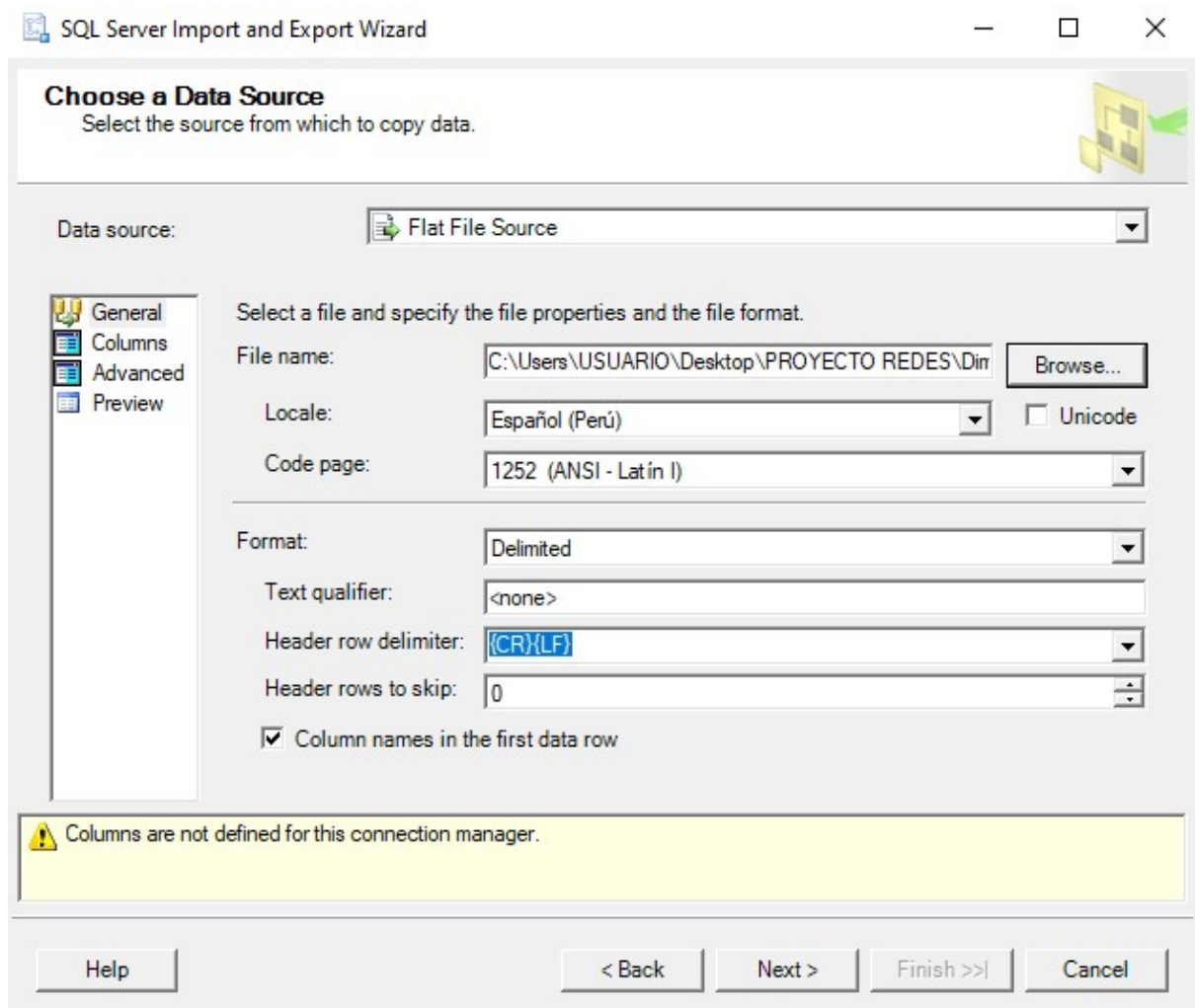
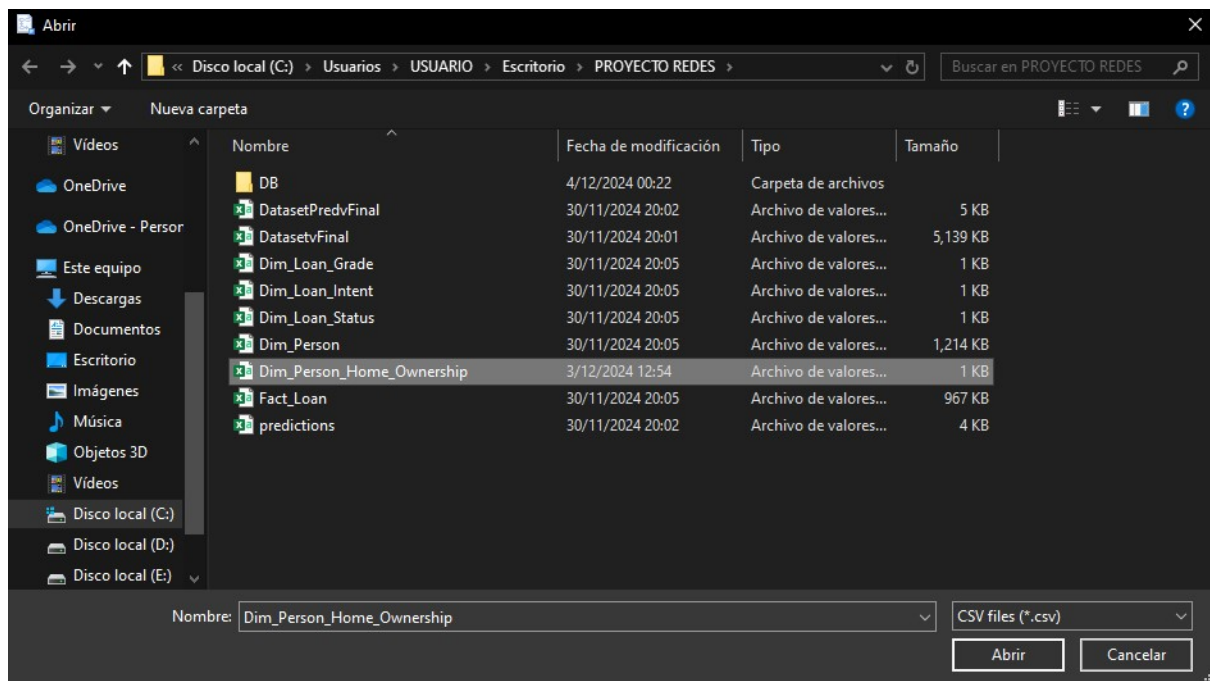
Hacemos clic derecho sobre la BD y seleccionamos Tasks/Import Data...



Seguidamente seleccionamos "Flat File Source" para importar los archivos CSV.



Seleccionamos el archivo CSV correspondiente a la tabla "Dim_Person_Home_Ownership":



Seguidamente, veremos los datos que se importaran ademas de los encabezados de la tabla.

SQL Server Import and Export Wizard

Choose a Data Source

Select the source from which to copy data.

Data source: Flat File Source

Specify the characters that delimit the source file:

Row delimiter: {CR}{LF}

Column delimiter: Comma (,)

Preview rows 2-5:

home_ownership	home_ownership_id
RENT	1
OWN	2
MORTGAGE	3
OTHER	4

Refresh Reset Columns

Help < Back Next > Finish >>| Cancel

Luego seleccionamos "Microsoft OLE DB Provider for SQL Server".

SQL Server Import and Export Wizard

Choose a Destination

Specify where to copy data to.

Destination: .Net Framework Data Provider for Odbc

Flat File Destination

Microsoft Access (Microsoft Jet Database Engine)

Microsoft Excel

Microsoft OLE DB Driver for SQL Server

Microsoft OLE DB Provider for Analysis Services 14.0

Microsoft OLE DB Provider for Oracle

Microsoft OLE DB Provider for Search

Microsoft OLE DB Provider for SQL Server

Connection String con nombre

Dsn

Datos

ConnectionString

Origen

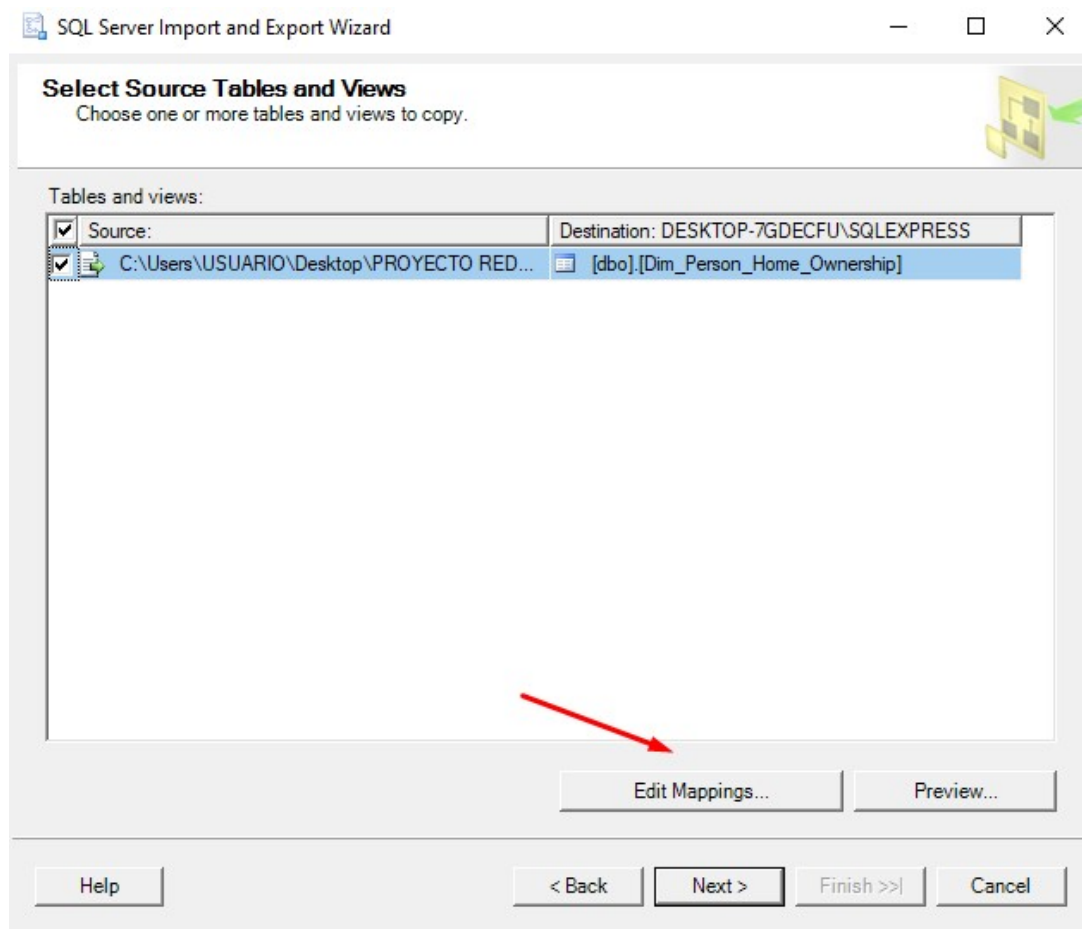
Driver

Driver

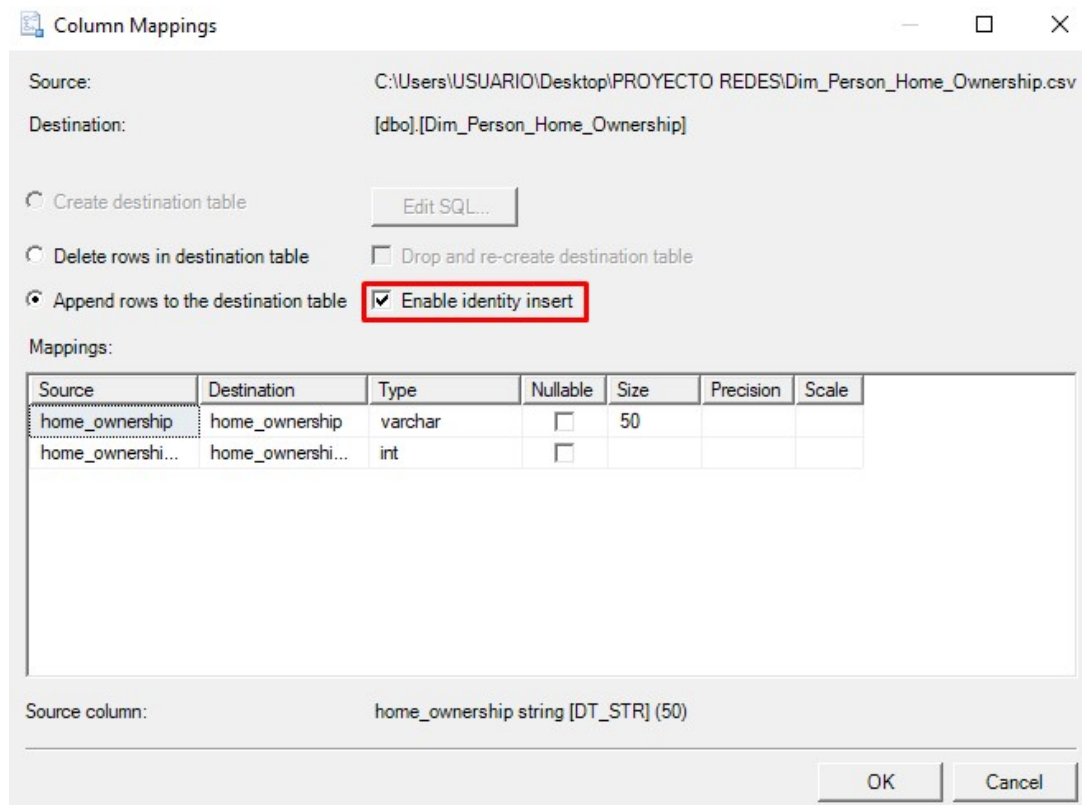
Nombre del controlador ODBC que se debe utilizar al conectar con el Data Source.

Help < Back Next > Finish >>| Cancel

Antes de continuar, seleccionamos "Edit Mappings...":



Y activamos la opción "Enable identity insert":



Finalmente, continuamos dando clic en "Next" hasta llegar al final.

SQL Server Import and Export Wizard

Review Data Type Mapping

Select a table to review how its data types map to those in the destination and how it handles conversion issues.

Table:

Source	Destination
C:\Users\USUARIO\Desktop\PROYECTO REDES\...	[dbo].[Dim_Person_Home_Ownership]

Data type mapping:

Source Column	Source Type	Destination Col...	Destination Type	Convert	On Error	On Truncati...
home_owners...	string [DT_STR]	home_owners...	varchar			
home_owners...	string [DT_STR]	home_owners...	int	<input checked="" type="checkbox"/>	Use Global	Use Global

To view conversion details, double-click the row that contains the column source type to be converted.

On Error (global) Fail

On Truncation (global) Fail

Help < Back Next > Finish >> Cancel

SQL Server Import and Export Wizard

Save and Run Package

Indicate whether to save the SSIS package.

☒ Run immediately

☐ Save SSIS Package

☒ SQL Server

☐ File system

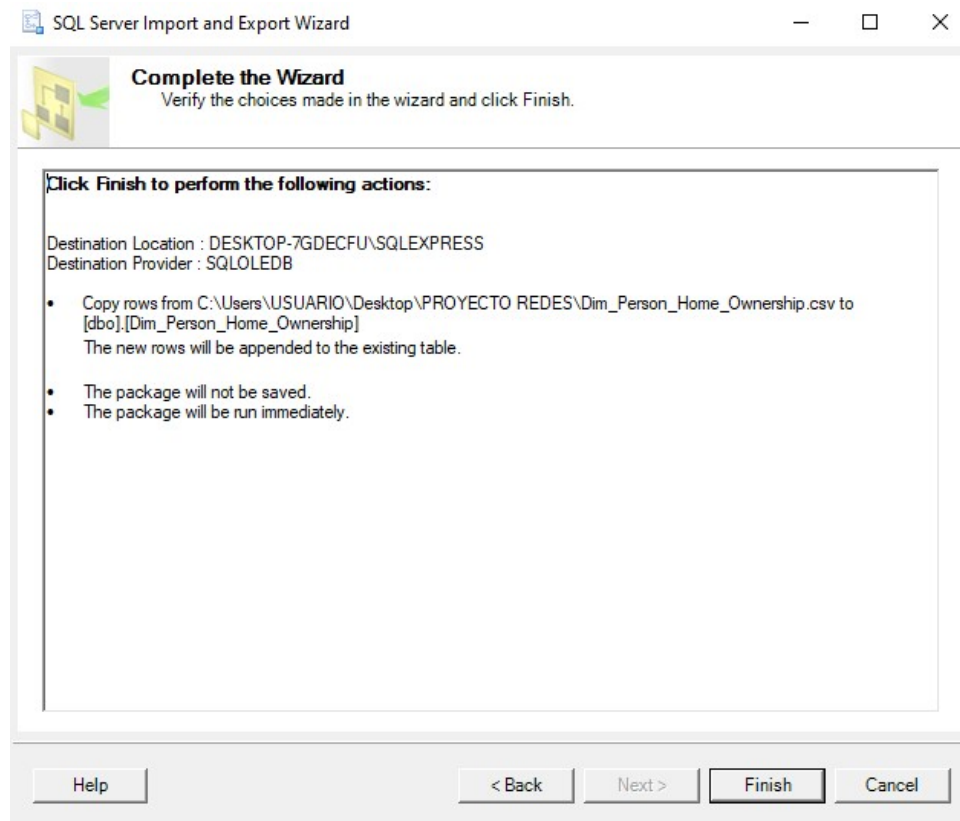
Package protection level:

Encrypt sensitive data with user key

Password:

Retype password:

Help < Back Next > Finish >> Cancel



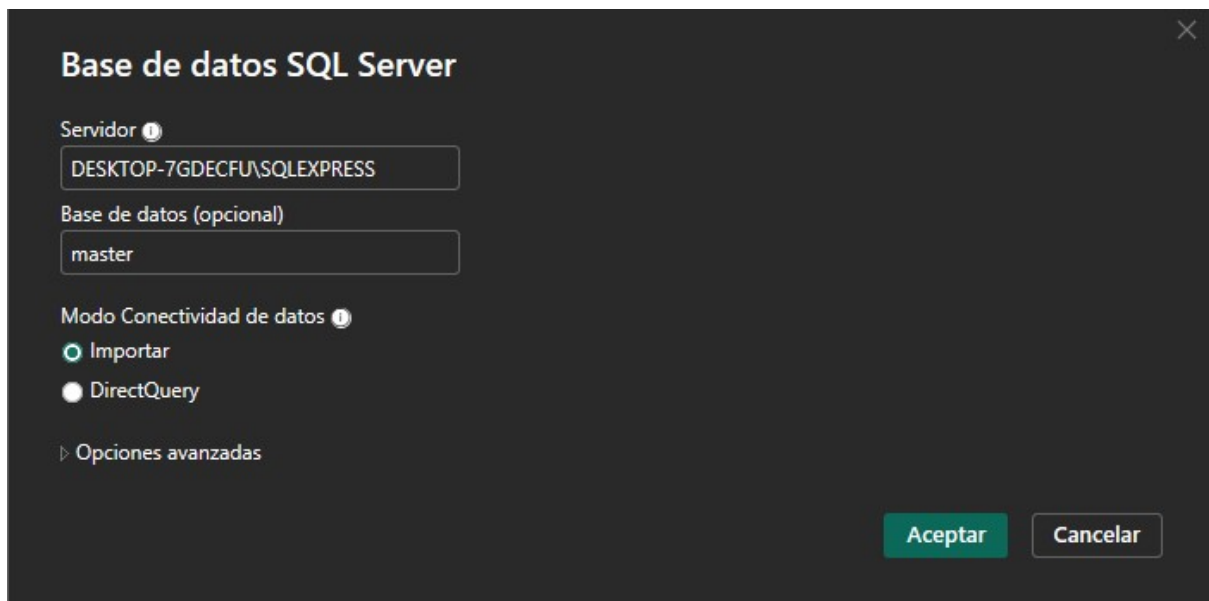
El proceso debe repetirse para cada archivo .csv, con su tabla correspondiente en la BD.

5.3 Conexión a Power BI

Los datos se cargan desde SQL Server:



Conectamos al servidor correspondiente e indicamos la BD a usar:



Base de datos SQL Server

Servidor ⓘ
DESKTOP-7GDECFU\SQLEXPRESS

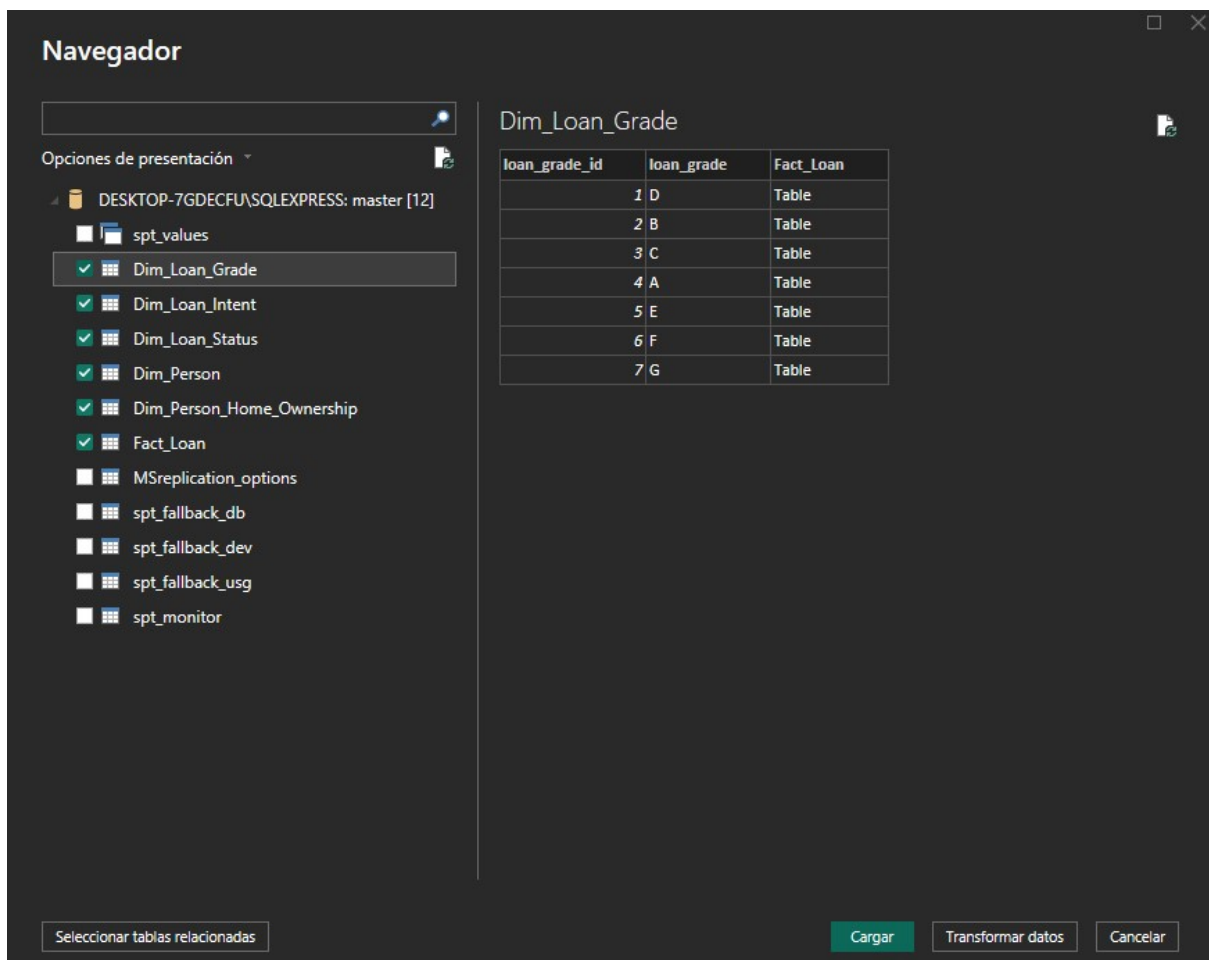
Base de datos (opcional)
master

Modo Conectividad de datos ⓘ
☒ Importar
☐ DirectQuery

▸ Opciones avanzadas

Aceptar Cancelar

Seleccionamos las tablas que vamos a usar para hacer los gráficos:



Navegador

Opciones de presentación ▾

DESKTOP-7GDECFU\SQLEXPRESS: master [12]

- spt_values
- ☒ Dim_Loan_Grade
- ☒ Dim_Loan_Intent
- ☒ Dim_Loan_Status
- ☒ Dim_Person
- ☒ Dim_Person_Home_Ownership
- ☒ Fact_Loan
- MSreplication_options
- spt_fallback_db
- spt_fallback_dev
- spt_fallback_usg
- spt_monitor

Dim_Loan_Grade

loan_grade_id	loan_grade	Fact_Loan
1	D	Table
2	B	Table
3	C	Table
4	A	Table
5	E	Table
6	F	Table
7	G	Table

Seleccionar tablas relacionadas Cargar Transformar datos Cancelar

Para este caso usaremos la tabla "Fact_Loan" para hacer los gráficos correspondientes:

ArchivoInicioAyudaHerramientas de tablasHerramientas de columnas

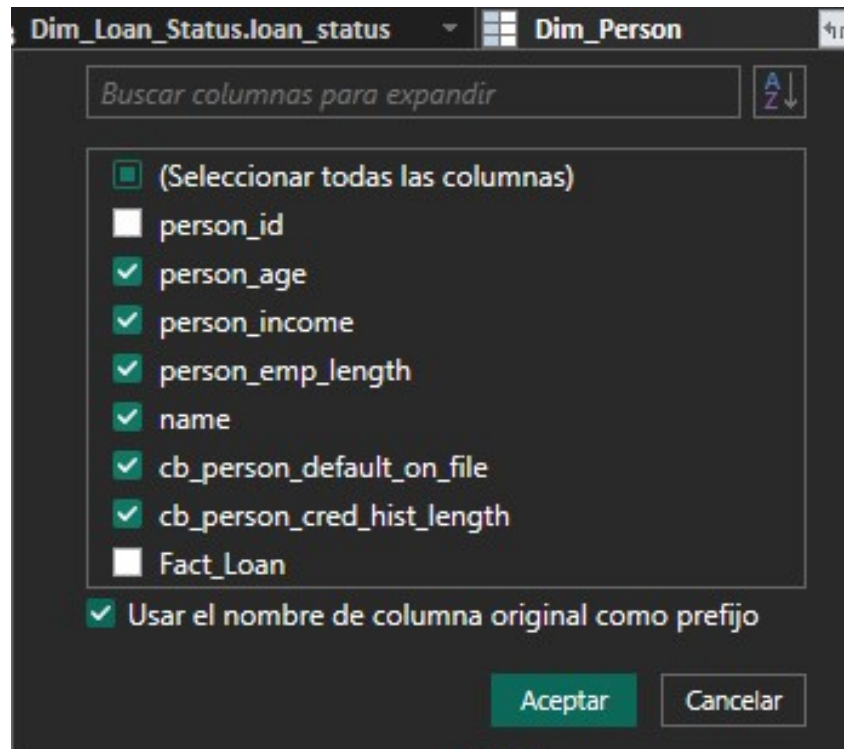
<

6 Visualización en Power BI

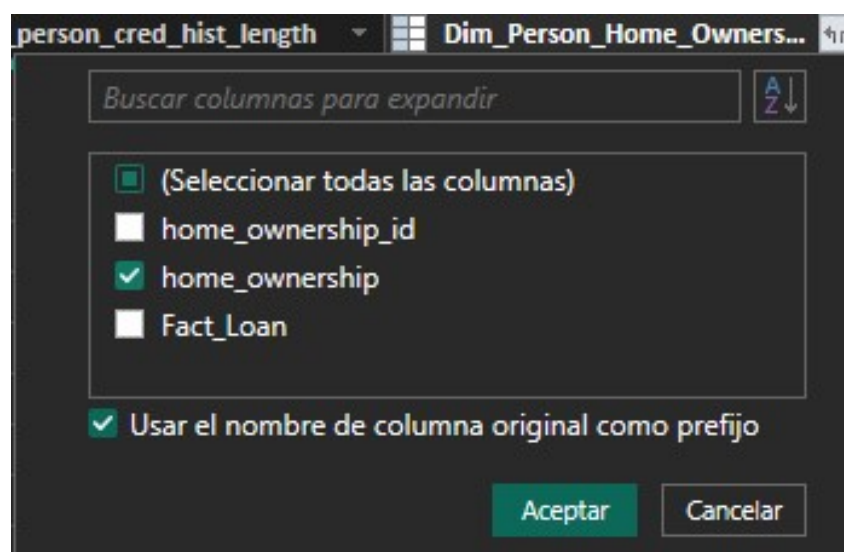
6.1 Construcción de datos

Una vez tengamos las tablas en Power BI, entramos a Power Query para hacer los siguientes pasos:

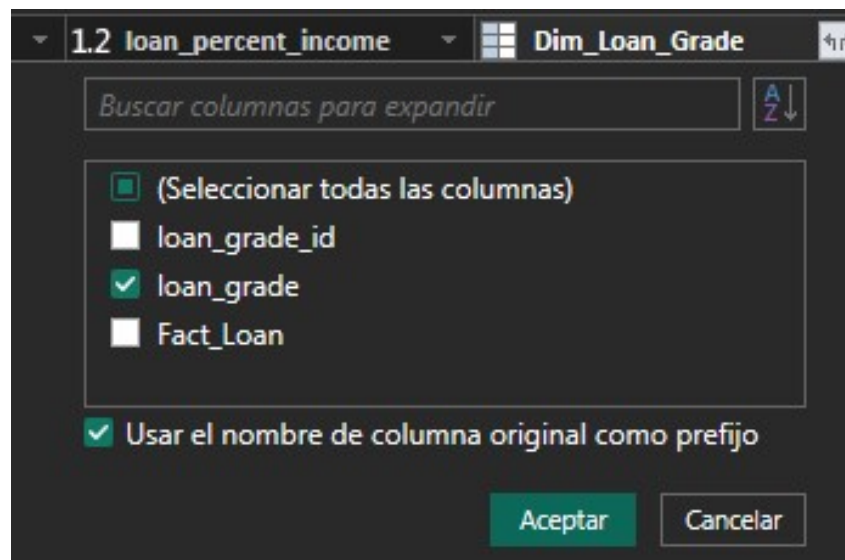
- En Dim.Person elegimos las siguientes opciones.



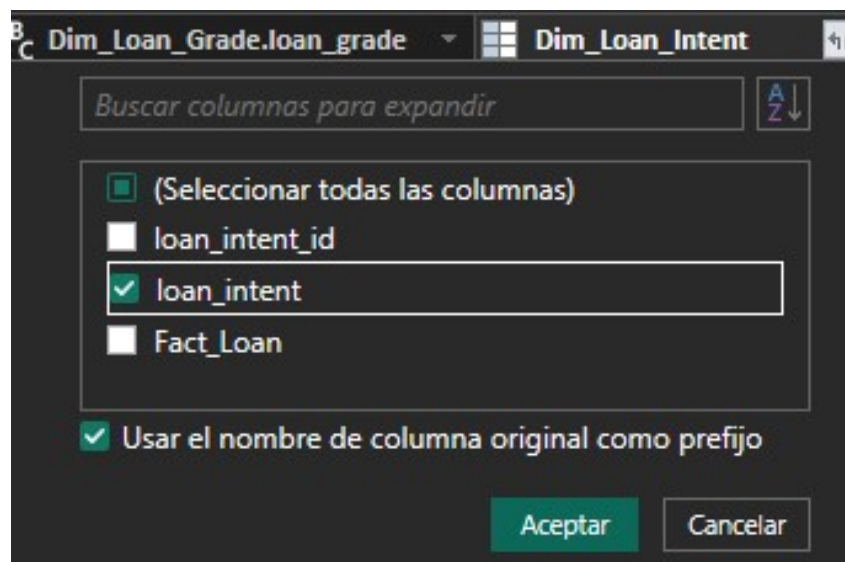
- En Dim.Person.Home.Owershhip elegimos las siguientes opciones.



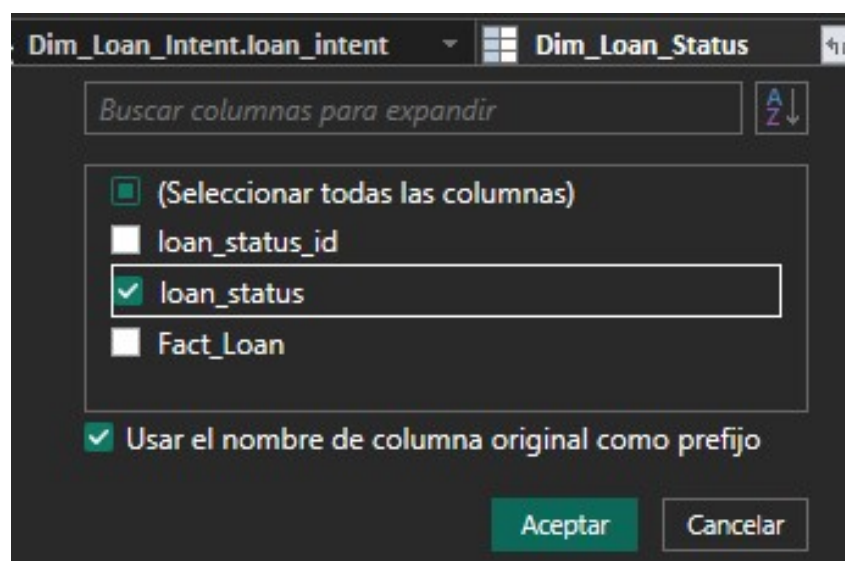
- En Dim.Loan.Grade elegimos las siguientes opciones.



- En Dim.Loan.Intent elegimos las siguientes opciones.



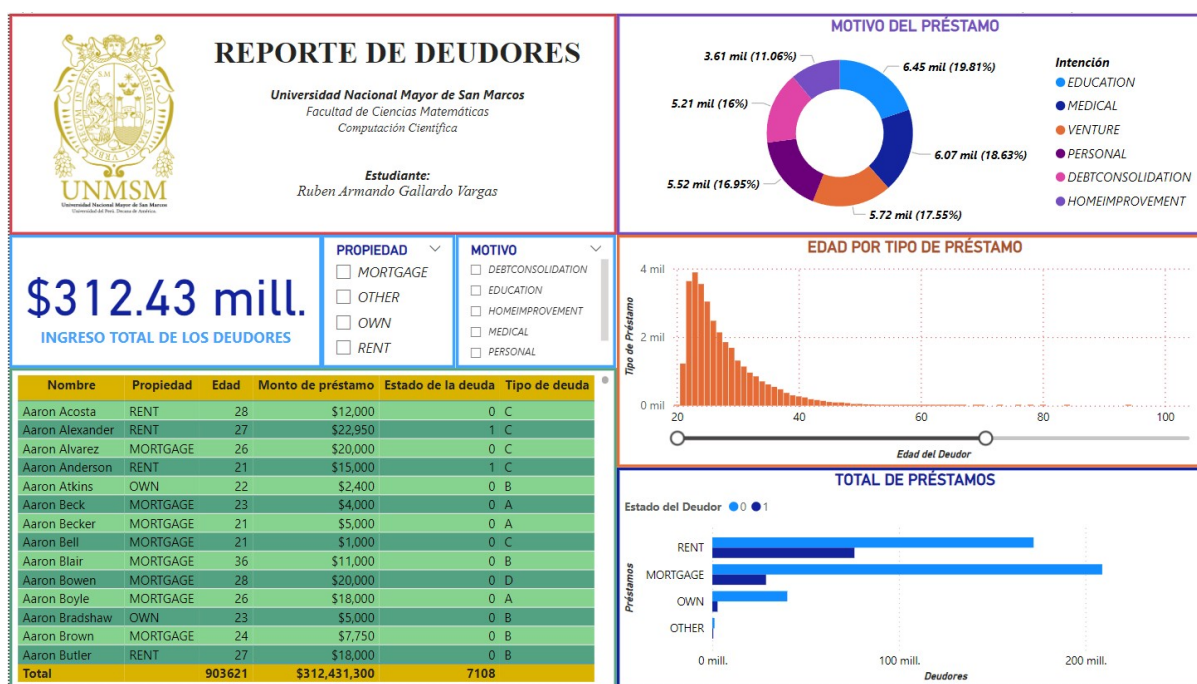
- En Dim.Loan.Status elegimos las siguientes opciones.



6.2 Tipos de Gráficos Implementados

- Gráficos de columnas agrupadas
- Gráfico de anillos
- Barras agrupadas
- Tablas
- Tarjetas

6.3 Reporte Realizado



7 Conclusiones

El sistema implementado cumple con los objetivos de manejar datos desde su origen en archivos CSV hasta su visualización interactiva en Power BI. A continuación, se presentan las conclusiones principales:

- La integración entre SQL Server y Power BI facilita un flujo de datos robusto y eficiente, lo que permite gestionar grandes volúmenes de información sin pérdida de calidad.
- El uso de SQL Server como base de datos intermedia proporciona flexibilidad para realizar transformaciones avanzadas y asegurar la calidad de los datos antes de la visualización.
- La elección de Power BI como herramienta de visualización potencia el análisis al ofrecer gráficos dinámicos y personalizables, que son útiles para la toma de decisiones basada en datos.