

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

MAESTRÍA EN CIENCIAS DE DATOS

BASES DE DATOS RELACIONALES

MAESTRO: JOSE ALBERTO BENAVIDES VAZQUEZ

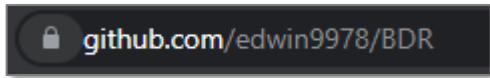
TAREA #1

ALUMNO: EDWIN MARTIN ROMERO SILVA

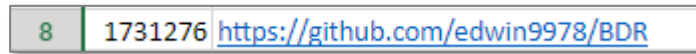
MATRICULA: 1731276

Tarea 1

- Crear un repositorio público de Github



- Compartir el repositorio en el Tems correspondiente.



- Describir una base de datos y sus relaciones de manera no estructurada. Agregar el tipo de dato que tendrá cada uno de los atributos.

Elegí un Dataset de Kaggle llamado **Credit Card Transactions**, contiene 3 archivos:

1) credit_card_transactions-ibm_v2:

Contiene más de 4 millones de transacciones desde el 2015 al 2020, cada registro es único y tiene una etiqueta que indica si la transacción es Fraudulenta o no.

Este base contiene 2 variables importantes, User y Card, estas indican cual cliente realizó la transacción y con cual tarjeta (Cada cliente puede tener mas de 1 tarjeta). Estas 2 variables nos servirán para relacionar esta tabla con la tabla 2).

Estas son las variables que contiene la tabla y el tipo de dato que tendrá cada atributo:

Variable	Tipo
Is Fraud?	String
User	Int
Card	Int
Year	Int
Month	Int
Day	Int
Time	String
Amount	Float
Use Chip	String
Merchant Name	String
Merchant City	String
Merchant State	String
Zip	Float
MCC	Int
Errors?	String

2) sd254_cards:

Contiene más de 6 mil tarjetas de débito y crédito, podríamos decir que el ID único de esta tabla es Card Number, pero luego de explorar las tablas noté que para relacionar esta tabla con la tabla 1) se deben utilizar las variables User y CARD_INDEX, ambas.

Estas son las variables que contiene la base y el tipo de dato que tendrá cada atributo:

Variable	Tipo
User	Int
CARD INDEX	Int
Card Brand	String
Card Type	String
Card Number	Int
Expires	Date
CVV	Int
Has Chip	String
Cards Issued	Int
Credit Limit	Float
Acct Open Date	Date
Year PIN last Changed	Int
Card on Dark Web	String

3) sd254_users:

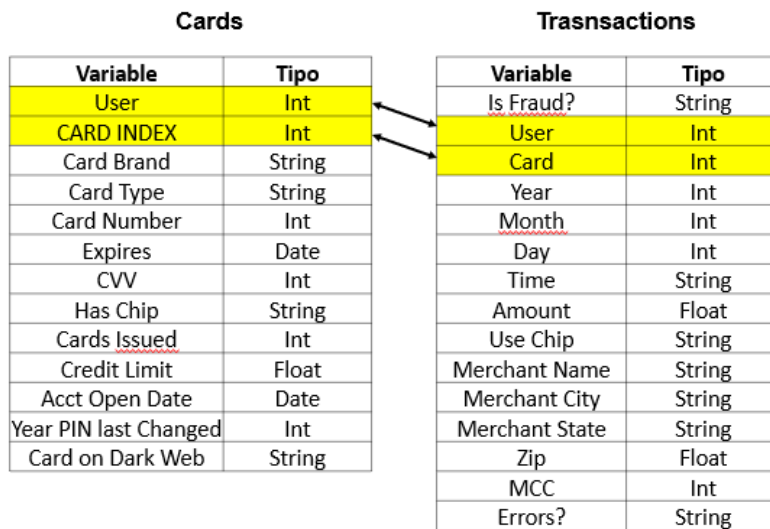
Contiene 2 mil registros de usuarios. Contiene variables como fecha de nacimiento, edad, dirección, ingresos anuales, etc.

Luego de explorar la base noté que el ID único es la variable Person (Es el nombre del cliente o usuario), pero desafortunadamente esta base no puede relacionarse con las 2 anteriores ya que no contiene la variable User y la variable Person no está presente en las bases 1 y 2.

Estas son las variables que contiene la base y el tipo de dato que tendrá cada atributo:

Variable	Tipo
Person	String
Current Age	Int
Retirement Age	Int
Birth Year	Int
Birth Month	Int
Gender	String
Address	String
Apartment	String
City	String
State	String
Zipcode	Int
Latitude	float
Longitude	float
Per Capita Income - Zipcode	float
Yearly Income - Person	float
Total Debt	float
FICO Score	Int
Num Credit Cards	Int

La relación entre las bases 1 y 2 se puede representar de la siguiente forma:



- Investigar diferentes SGBD, elegir uno y describirlo.

Elegí MySQL ya que según lo que leí, es uno de los mas utilizados y me gusta que haya muchos foros que hablen al respecto, ya que eso facilita buscar en internet cuando me surge alguna duda.

MySQL es el sistema de gestión de bases de datos relacional de código abierto más popular, está basado en lenguaje SQL, al igual que muchos otros SGBD.

Funciona como interfaz entre el usuario y la información almacenada en una base de datos.

