



Fraud & Risk Data Scientist, LATAM

PayU Colombia

Technical Test

Primera parte

El dataset adjunto contiene datos de transacciones con tarjeta de crédito hechas entre el 1 de enero de 2019 y el 31 de diciembre de 2020. Las columnas incluidas son las siguientes:

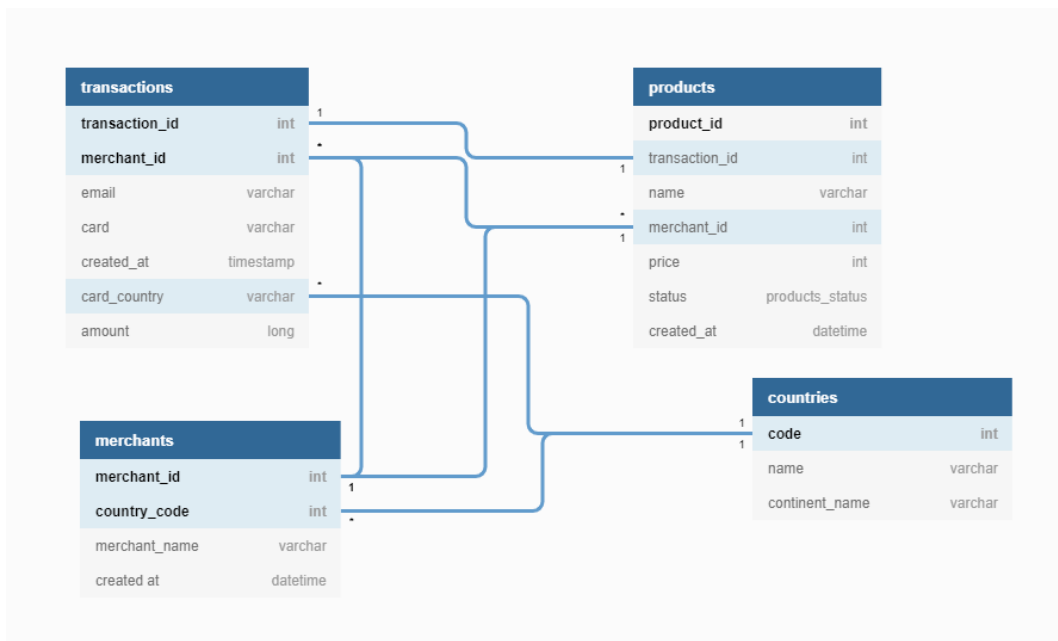
Nombre de la Variable	Descripción
trans_date_trans_time	Fecha de la transacción
cc_num	Tarjeta de crédito
merchant	Nombre del comercio
category	Categoría del comercio
amt	Monto de la transacción
first	Nombre del pagador
last	Apellido del pagador
gender	Género
street	Dirección del pagador
city	Ciudad
state	Estado
zip	Código ZIP
lat	Latitud
long	Longitud
city_pop	Población de la ciudad
job	Ocupación
dob	Fecha de Nacimiento
trans_num	Id de la transacción
unix_time	Tiempo Unix
merch_lat	Latitud del comercio
merch_long	Longitud del comercio
is_fraud	Variable binaria. 1 si la transacción es fraudulenta, 0 si es genuina

Construya un modelo predictivo para la variable `is_fraud` con la información que considere pertinente. Adjunte el código utilizado debidamente comentado, así como un documento con una breve descripción del procedimiento realizado y de los resultados. Incluya en el documento cualquier análisis descriptivo de los datos que considere relevante para el problema de predicción de fraude transaccional.

Se evaluarán más las estrategias utilizadas que los resultados predictivos. Usted puede utilizar el lenguaje que desee, pero preferiblemente trabaje con Python.

Segunda parte

Con base en el siguiente modelo de datos:



1. Genere un query para obtener una tabla que contenga la fecha de creación de la transacción, el monto de la transacción, el nombre del producto, el nombre del comercio y el nombre del país del comercio.
2. Genere un query que calcule la cuenta del número de transacciones por cada mes por cada continente.

Envíe los archivos al correo daniela.sanabria@payu.com. Para resolver cualquier duda puede escribir al mismo correo.

El dataset para resolver el primer punto puede ser encontrado en este enlace:

https://drive.google.com/file/d/1OYco-V6lMBUEi80glb__qNhlZpFZln0q/view?usp=sharing

Si lo considera necesario, puede hacer el ejercicio con una submuestra de los datos.