

Actividad en equipos: Análisis de Transacciones

Introducción

En esta actividad, analizaremos el archivo `transacciones.csv`, que contiene información sobre los usuarios y sus transacciones. Las columnas del archivo son las siguientes:

- `id_transaction`
- `date`
- `time`
- `amount`
- `transaction_type`
- `new_user`
- `user_id`
- `status`

Objetivo

Utilizar técnicas de visualización de datos para identificar patrones inusuales.

Realizar al cuatro de estos histogramas propuestos:

- Histograma del monto de las transacciones por hora del día con estado fraudolento de nuevos usuarios (Cuanto dinero históricamente se ha perdido más en las 24 horas del día)
- Histograma de la distribución de nuevos usuarios con una transacción fraudolenta (Cuantos usuarios nuevos tuvieron una transacción fraudolenta vs los usuarios que no son nuevos)
- Histograma del tipo de transacción y el estado de la transacción fraudolenta (Cuantos estados de transacciones fraudolentas tuvieron las transacciones purchase y transfer)
- Histograma de nuevos usuarios de cada mes del año y el monto de una transacción fraudolenta (Cuanto dinero se perdía en cada nuevo usuario por mes cuando el estado de la transacción es fraudolenta)

- Histograma de los meses con transacciones fraudulentas
- Histograma del monto perdido por mes por transacciones de tipo purchase con estado fraudolenta
- Histograma del monto perdido por mes por transacciones de tipo transfer con estado fraudolenta

Ejemplo de Uso de Pandas

A continuación, se muestra un ejemplo básico de cómo cargar el archivo CSV, procesar los datos y graficar algunas métricas. En este ejemplo voy a mostrar la distribución del monto de transacciones fraudulentas por día. Este Histograma lo pueden agregar si desean como parte de su selección de los cuatro histogramas.

1. Cargar el Archivo CSV

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('transacciones.csv')
# Filtrar para obtener solo transacciones fraudulentas y nuevos usuarios
fraudulent_transactions = data[(data['status'] == 'fraudulent')]

# Formato para extraer solo la hora de la columna 'time'
fraudulent_transactions['hour'] = pd.to_datetime(fraudulent_transactions['time'], for

#Se define y se crea la gráfica
plt.figure(figsize=(12, 6)) # Ajusta el tamaño de la figura
sns.histplot(data=fraudulent_transactions, x='hour', weights='amount', bins=24, kde=Tr

#Títulos
plt.title('Distribución del monto de transacciones fraudulentas por hora del día')
plt.xlabel('Hora del día')
plt.ylabel('Monto')

#cuadrícula en la gráfica
plt.grid(True)

#Mostramos la gráfica
```

```
plt.legend()  
plt.show()
```

Entregable

1. Un documento PDF donde adjunten las imagenes de los histograma seleccionados describiendo algun patrón inusuales que hayan notado en la gráfica
2. (EXTRA) Se darán dos participaciones si adjuntan el código que genera dichos histogramas así como un pequeño readme explicando qué gráficas seleccionaron.