

Business case: Rappi

Brenda Michelle Reyes Camacho



Overview

Se proporcionó una base de datos
“ds_challenge_apr2021.csv”
la cual contiene el registro de
26,975 operaciones con sus
respectivas descripciones (21
variables)

Tasks

- Hacer un análisis exploratorio
- Categorizar a los clientes
- Implementar un modelo que detecte las operaciones fraudulentas
- Explicar que tan bueno es el modelo.

Información general

**2 - 30 enero
del 2020**

**Periodo de
evaluación**

4,000

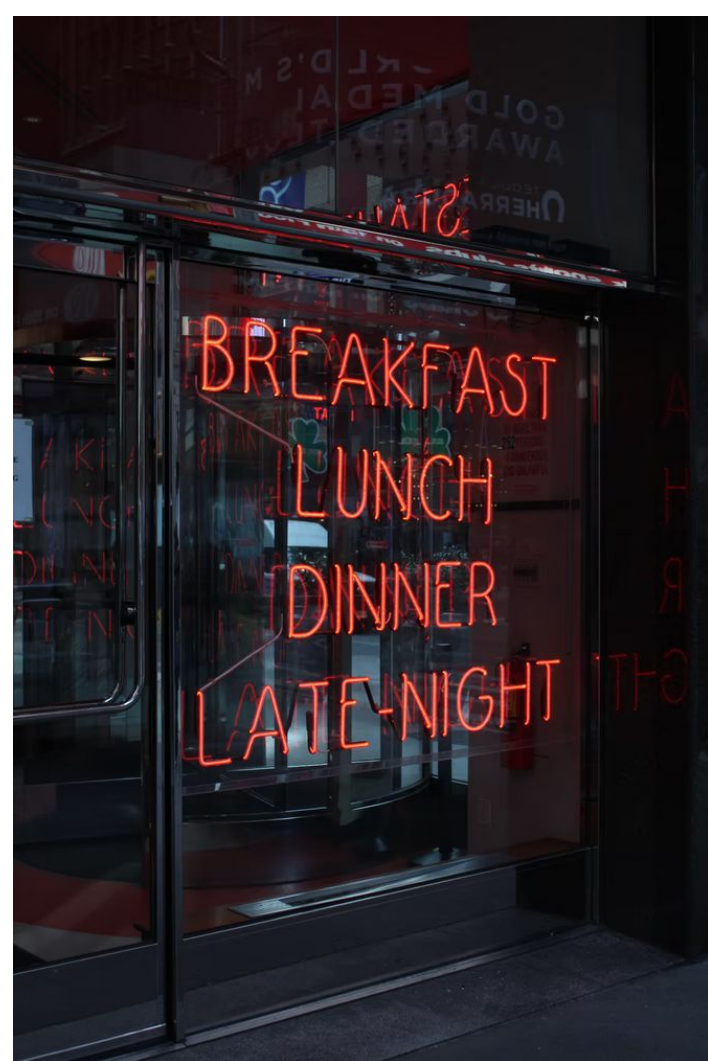
Usuarios

26,975

Transacciones

\$499.70

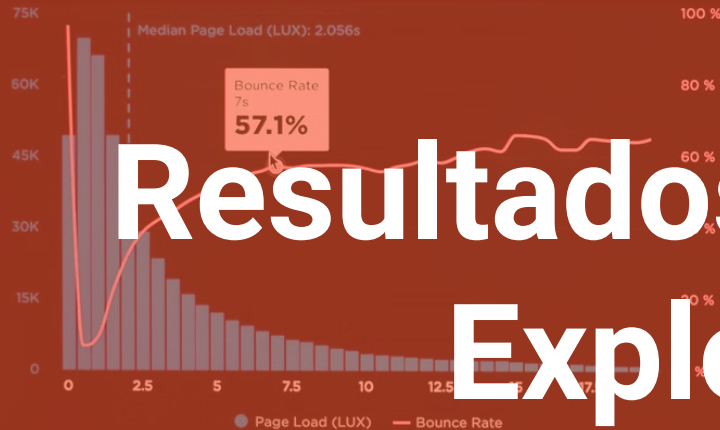
**Monto promedio
por transacción**



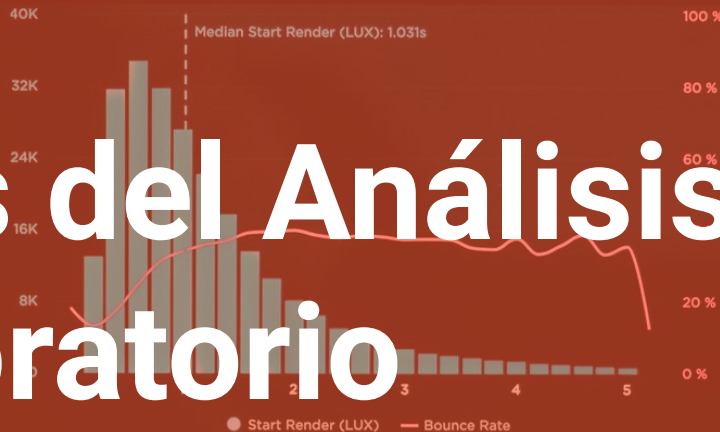
USERS: LAST 7 DAYS USING MEDIAN ▾



LOAD TIME VS BOUNCE RATE



START RENDER VS BOUNCE RATE



PAGE VIEWS VS ONLOAD

Page Load (LUX)

0.7s

1s

0.8s

0.6s

0.4s

Page Views (LUX)

2.7Mpvs

Bounce Rate (LUX)

40.6%

500K 100%

400K 80%

300K 60%

200K 40%

OPTIONS

SESSIONS

Sessions (LUX)

479K

4 pvs

3.2 pvs

2.4 pvs

1.6 pvs

Session Length (LUX)

17min

OPTIONS

PVs Per Session (LUX)

2pvs

100K 40 min

80K 32 min

60K 24 min

40K



Resultados del análisis Exploratorio

El análisis exploratorio de los datos mostró dos cosas importantes:

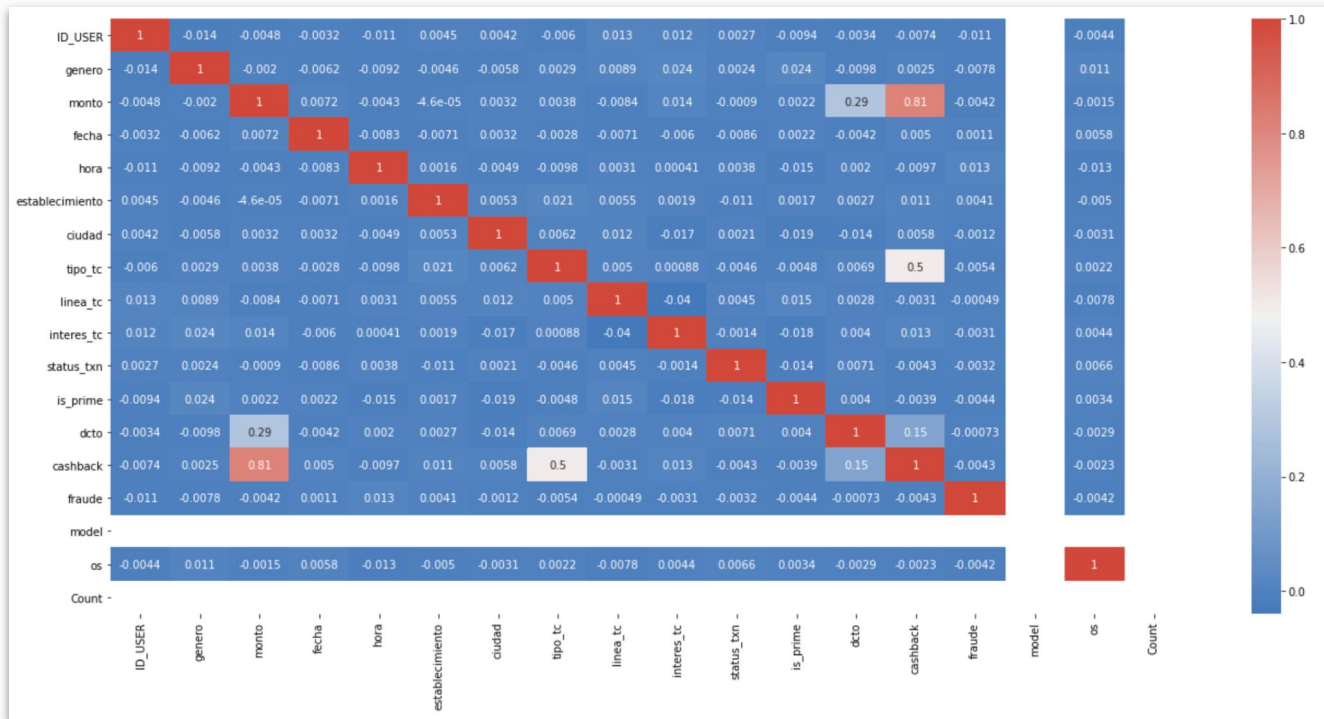
- Se tiene un problema imbalanceado, por lo que se tendrá que generar data para implementar el modelo.
- Las variables por sí solas no determinan la probabilidad de que una operación sea fraudulenta o no.

No obstante, se pueden destacar los siguientes valores de las operaciones **fraudulentas**:

- **Género:** Masculino (47.9%)
- **Día de la semana:** Jueves (17.9%)
- **Hora:** 7,19,22 y 23 hrs (22.47%)
- **Establecimiento:** Desconocido (39.14%) y MPago (12.84%)
- **Ciudad:** Desconocido (43.7%) y Mérida (14.81%)
- **Tipo tc:** Física (71.48%)
- **Status txn:** Aceptada (70.37%)
- **Prime:** False (87.65%)
- **os:** Android (26.67%)

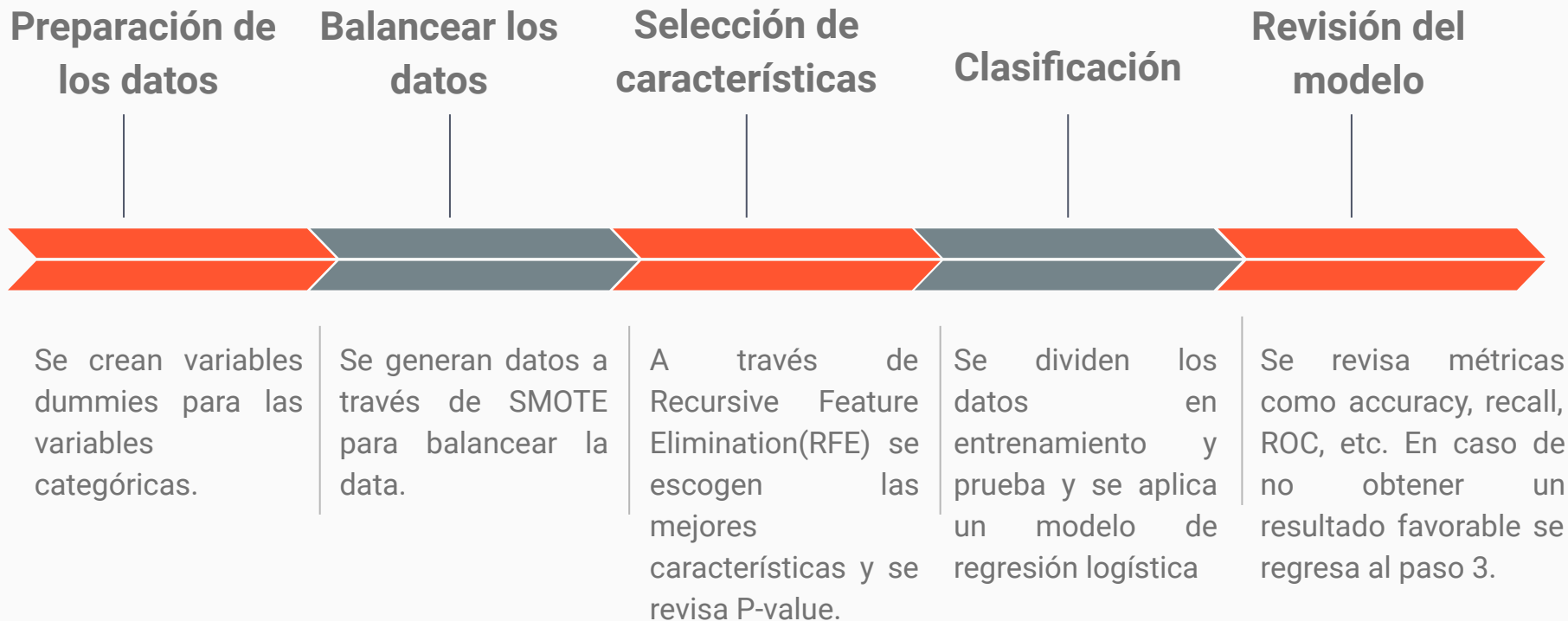
Correlación entre variables y fraude

El análisis exploratorio mostró a grandes rasgos que no hay variables que por sí solas estén altamente correlacionadas con el fraude. Y esto se confirma en la gráfica de la izquierda.



Implementación del modelo

Pasos de la implementación

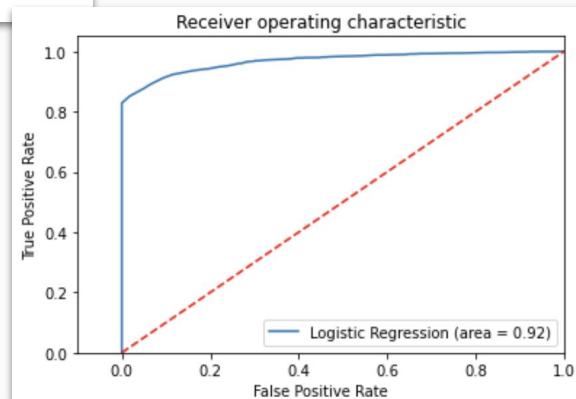
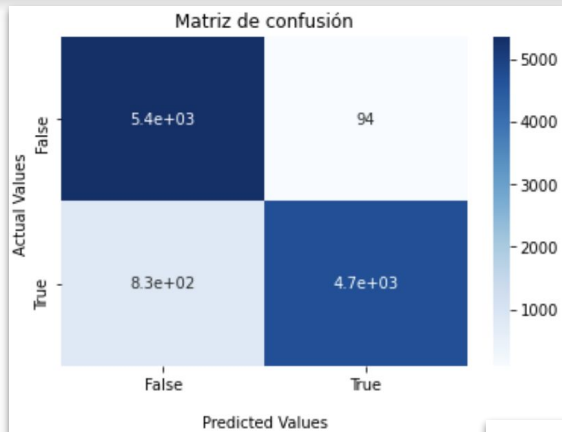


Resultados

92%

Accuracy

El modelo fue bueno clasificando las operaciones fraudulentas y esto se debe a que se balancearon los datos y se eliminaron las variables que perjudicaban o que no mejoraban al modelo.





Gracias
