CREDIT CARD FRAUD DETECTION

BY:
AISHWARYA GYANJYOTI

Agenda

Part 1: Objective

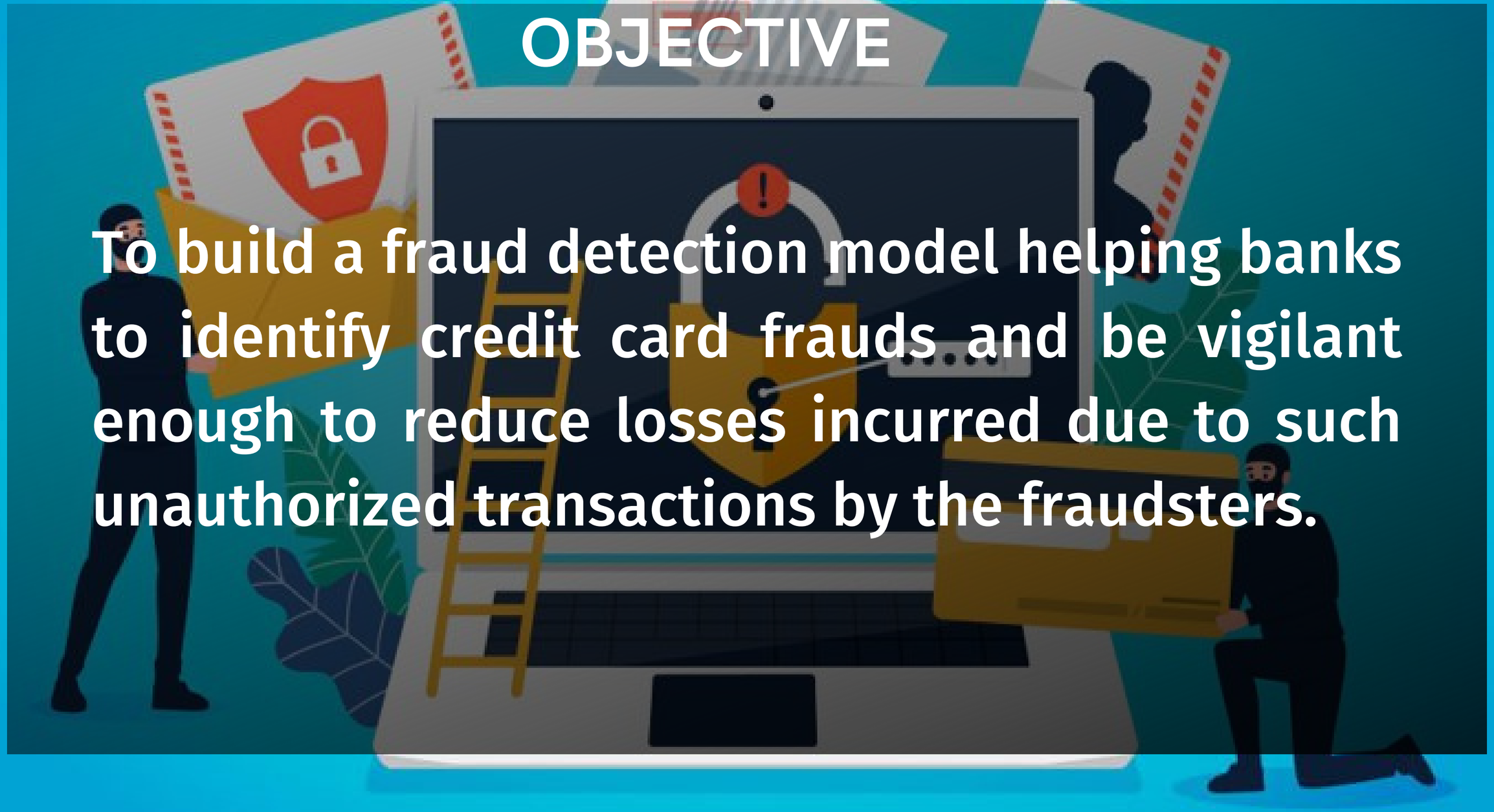
Part 2: Data Understanding

Part 3: Exploratory Data Analysis

Part 4: Cost Benefit Analysis

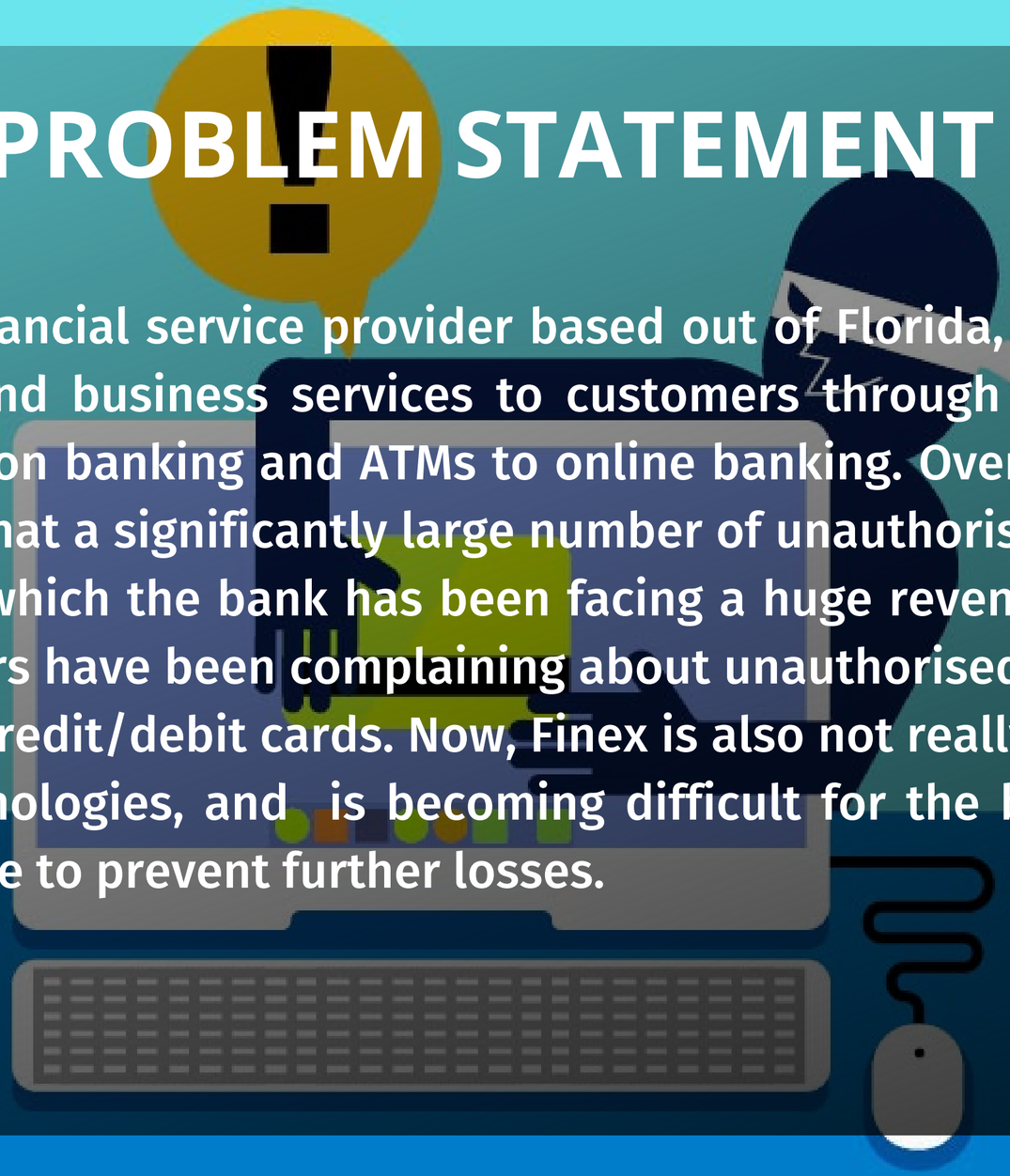
Part 5: Methodology

OBJECTIVE



To build a fraud detection model helping banks to identify credit card frauds and be vigilant enough to reduce losses incurred due to such unauthorized transactions by the fraudsters.

PROBLEM STATEMENT



Finex is a leading financial service provider based out of Florida, US offering a wide range of products and business services to customers through different channels, ranging from in-person banking and ATMs to online banking. Over the last few years, Finex has observed that a significantly large number of unauthorised transactions are being made, due to which the bank has been facing a huge revenue and profitability crisis. Many customers have been complaining about unauthorised transactions being made through their credit/debit cards. Now, Finex is also not really equipped with the latest financial technologies, and is becoming difficult for the bank to track these data breaches on time to prevent further losses.

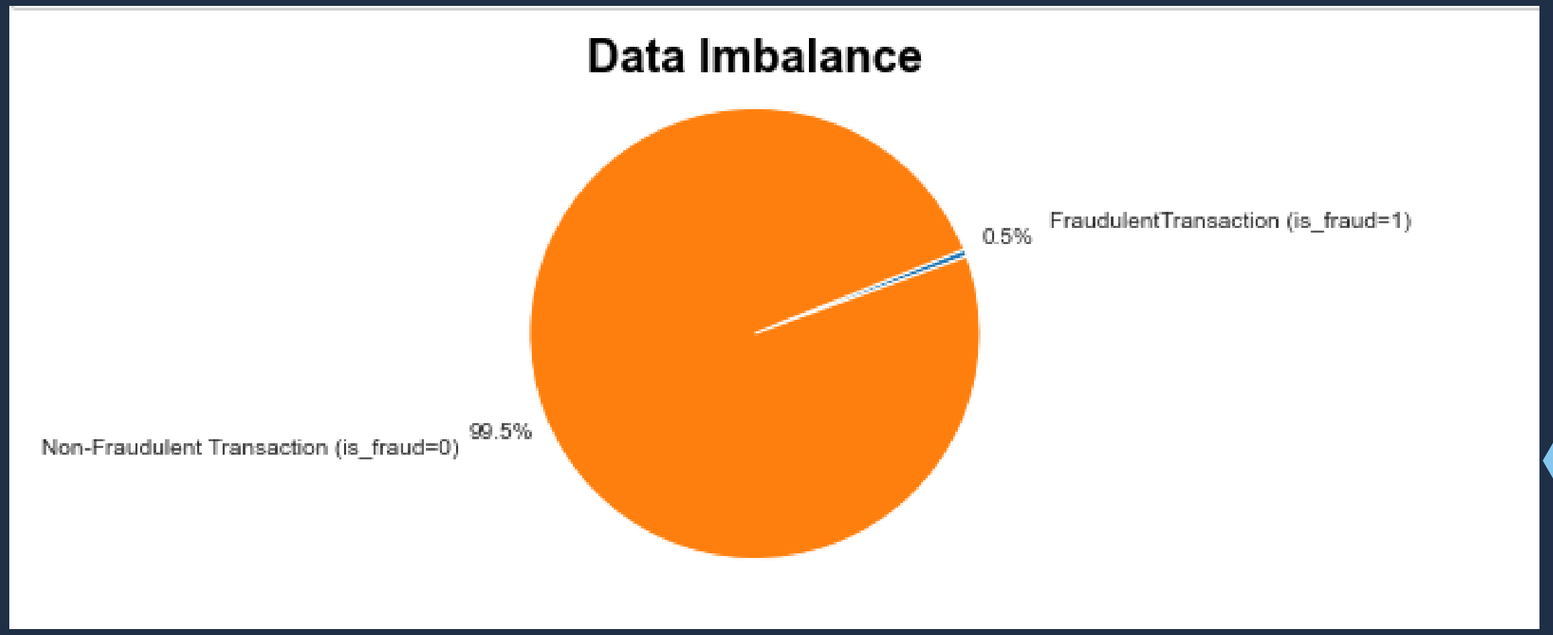
DATA UNDERSTANDING

Consisting of two datasets i.e.,fraud train and fraud test. Further, it's been combined.

The combined dataset i.e., cred fraud df has 1852394 rows and 22 columns (after dropping unnamed column though it is irrelevant).

1842743 (99.5%) transactions are non-fraudulent and 9651 (0.5%) transactions seem to be fraudulent - Reason of high data imbalance.

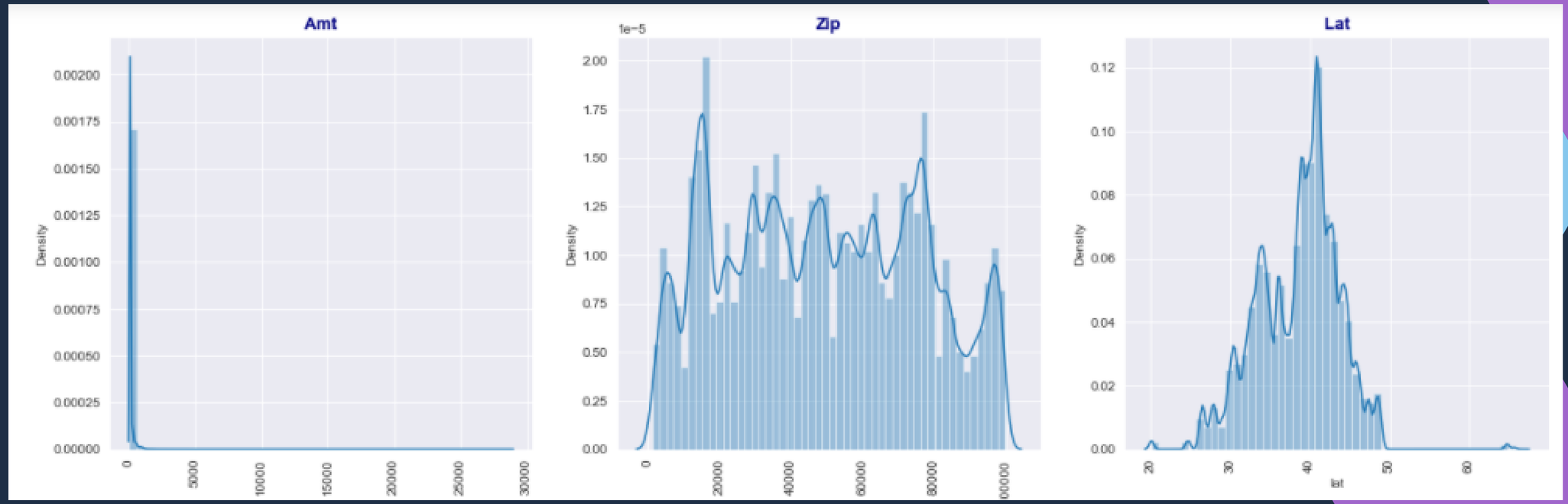

EXPLORATORY
DATA ANALYSIS



99.5% of transactions are non-fraud while rest 0.5% refers to fraud.

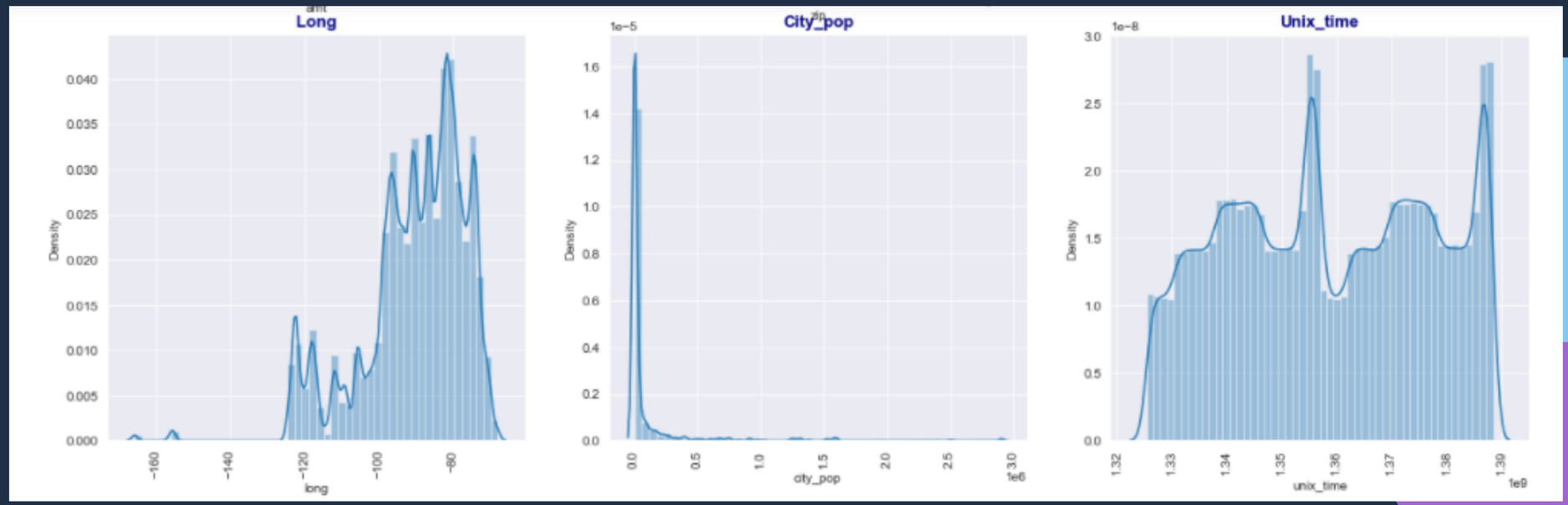
UNIVARIATE ANALYSIS

NUMERICAL COLUMNS



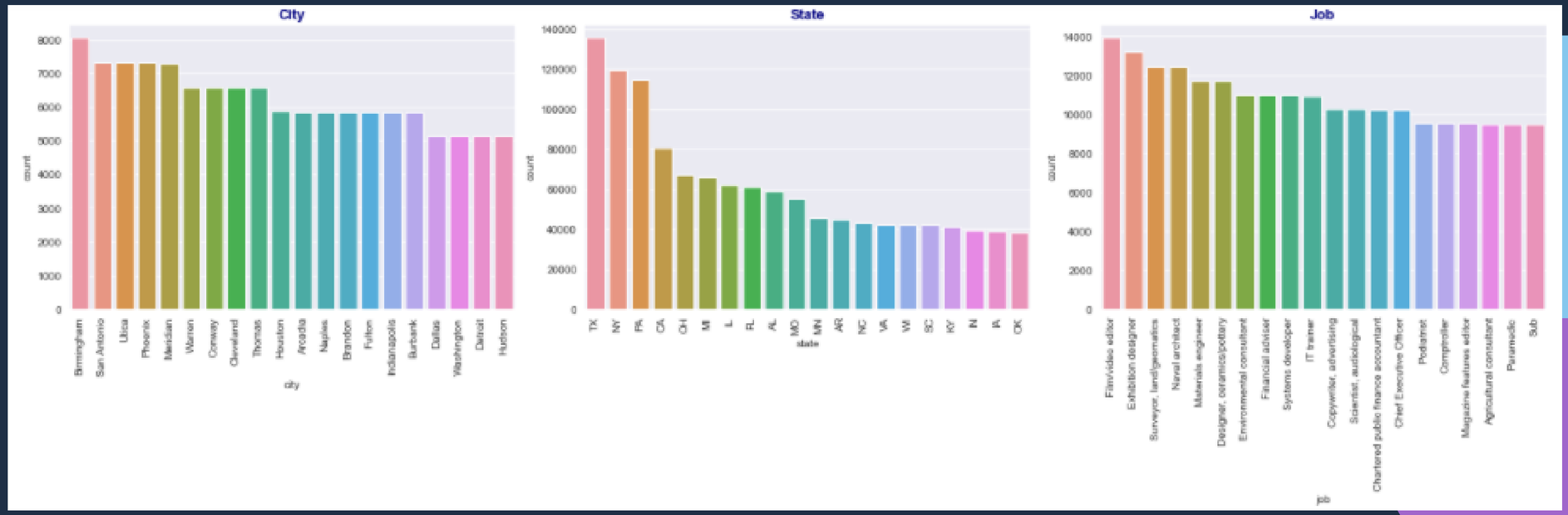
UNIVARIATE ANALYSIS

NUMERICAL COLUMNS



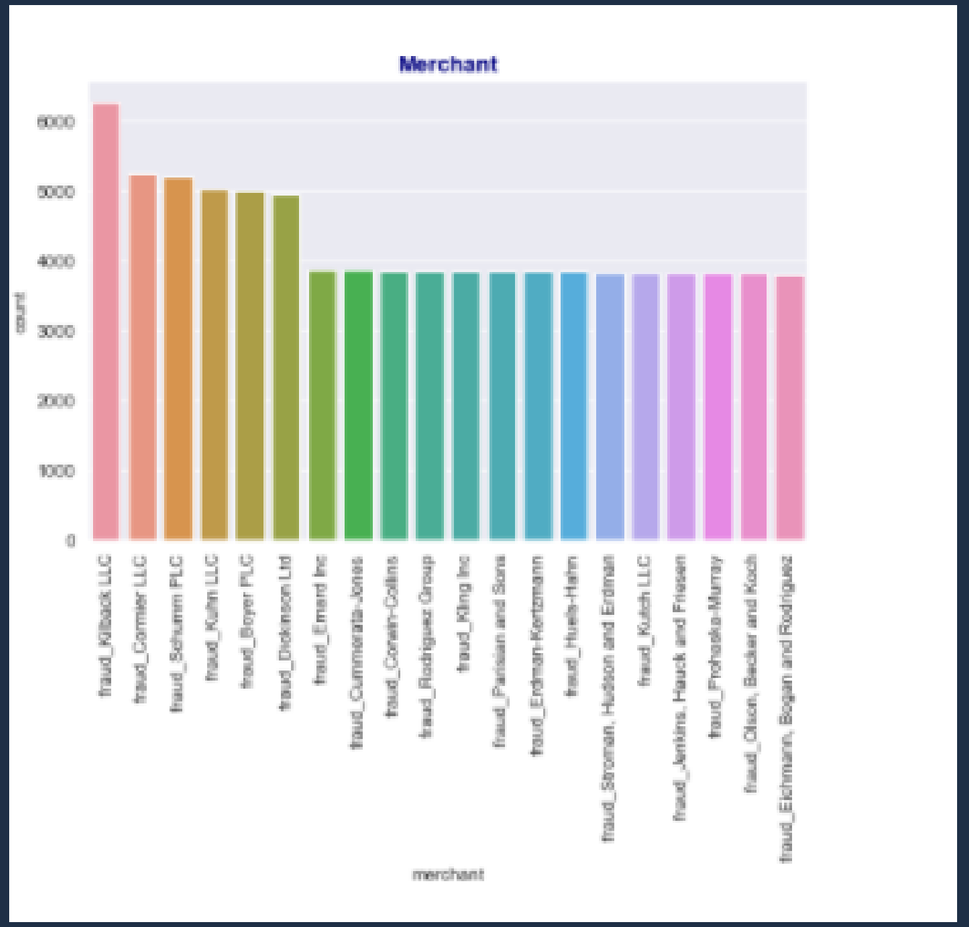
UNIVARIATE ANALYSIS

CATEGORICAL COLUMNS



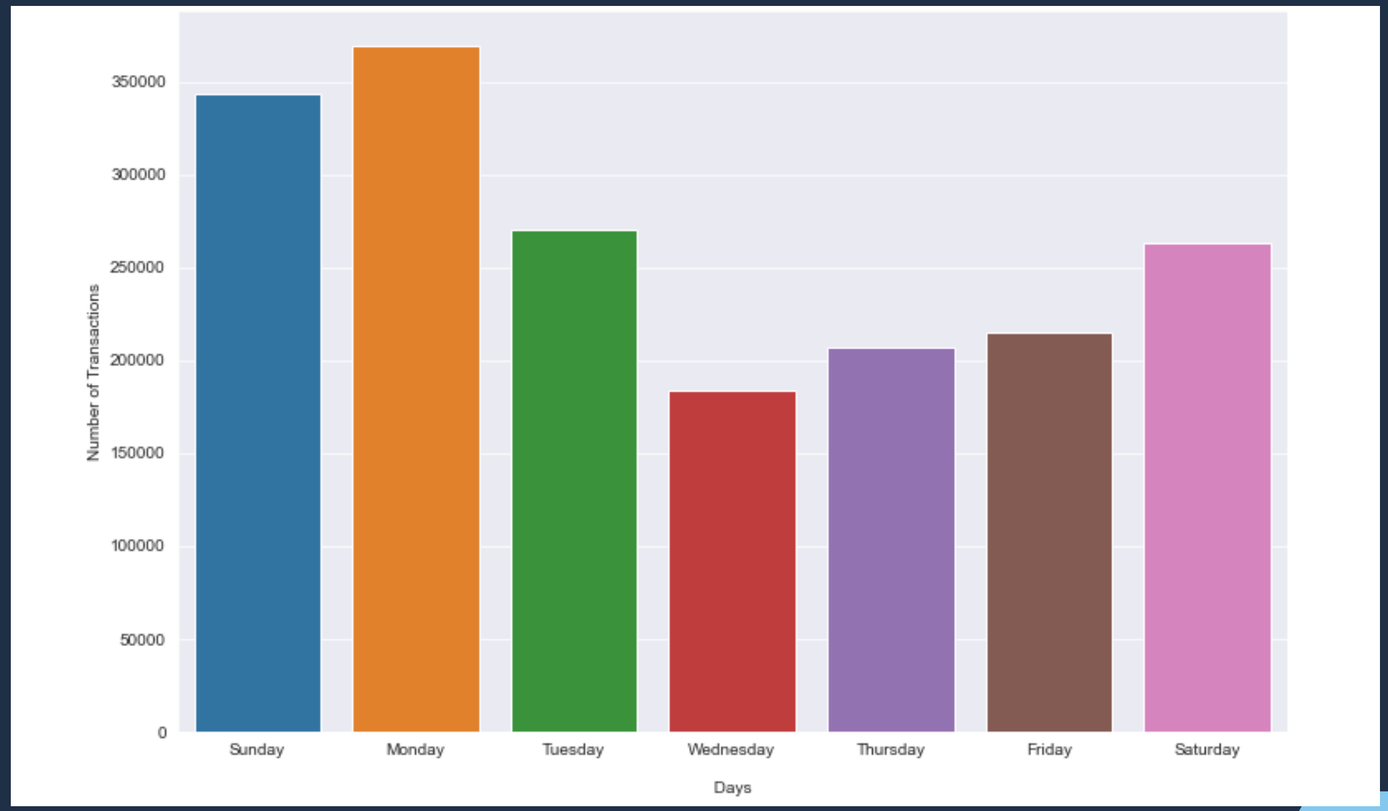
UNIVARIATE ANALYSIS

MERCHANT



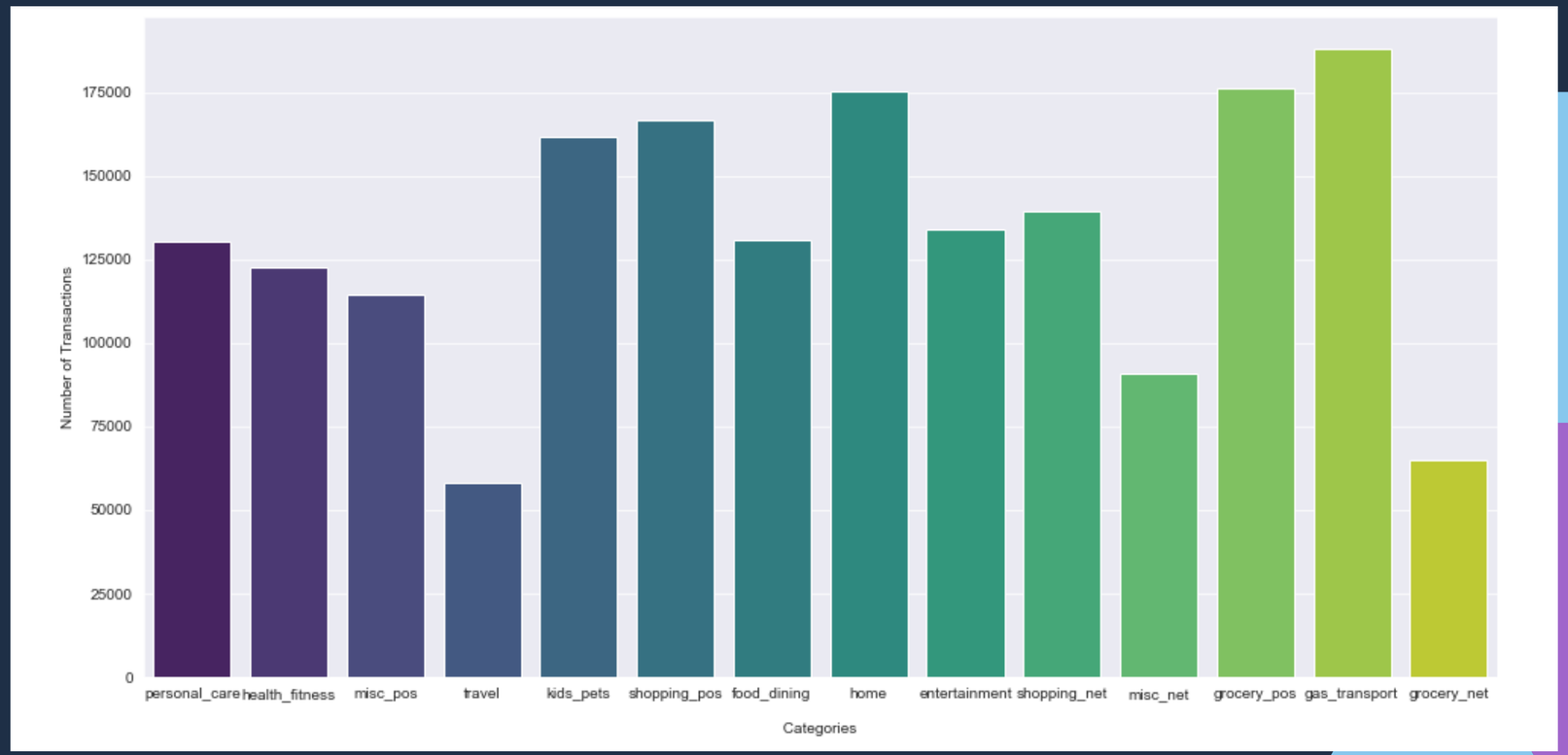
UNIVARIATE ANALYSIS

DAYS



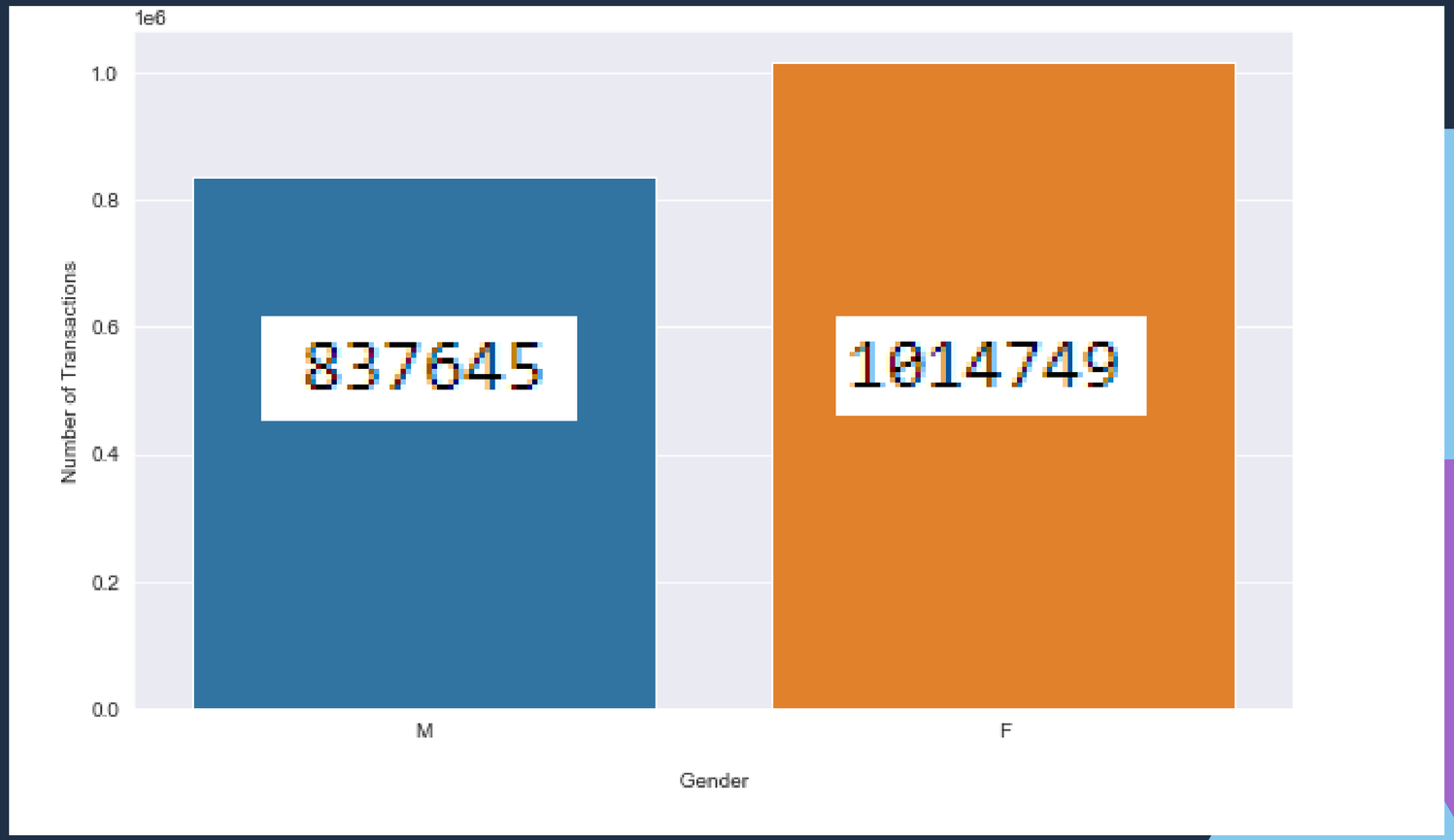
UNIVARIATE ANALYSIS

CATEGORIES



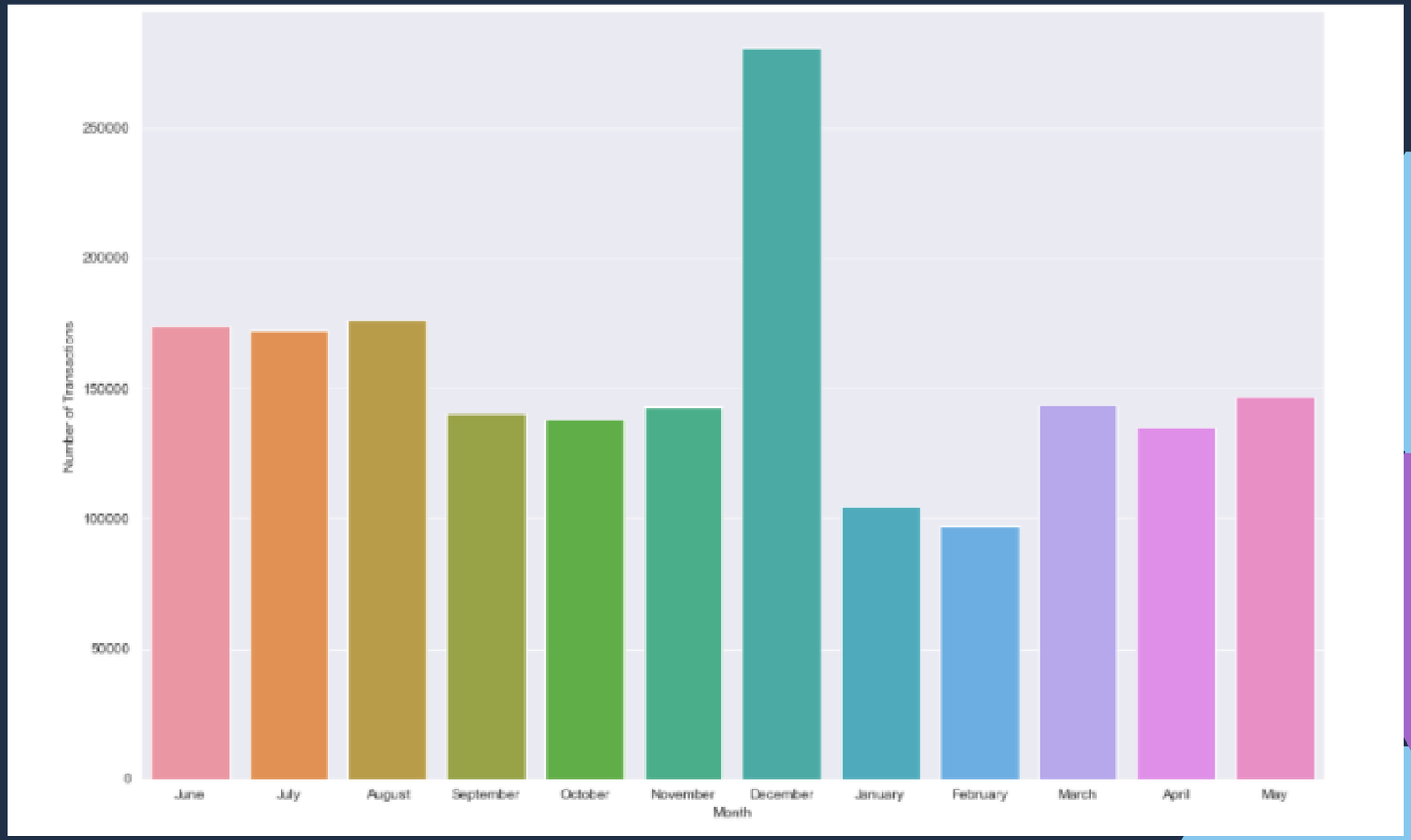
UNIVARIATE ANALYSIS

GENDER



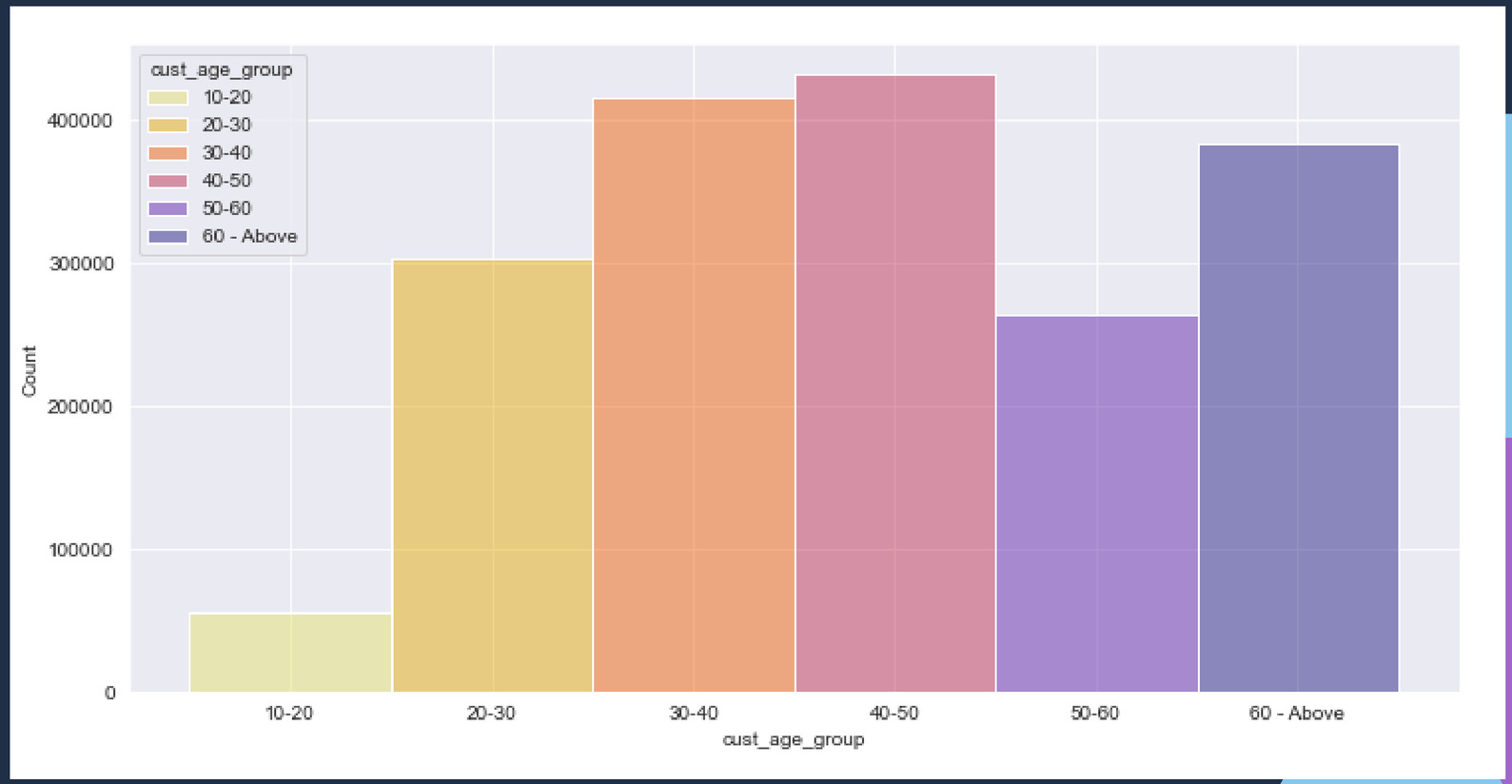
UNIVARIATE ANALYSIS

MONTH



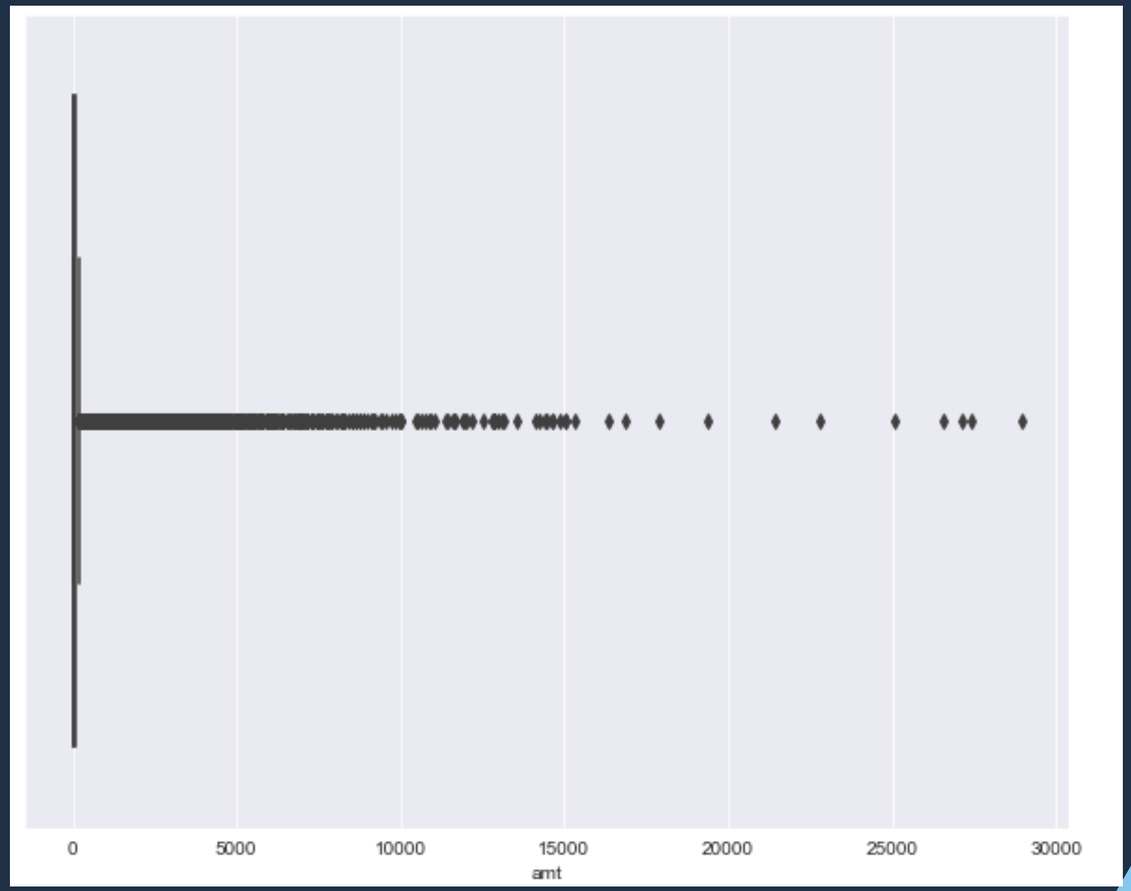
UNIVARIATE ANALYSIS

AGE GROUP



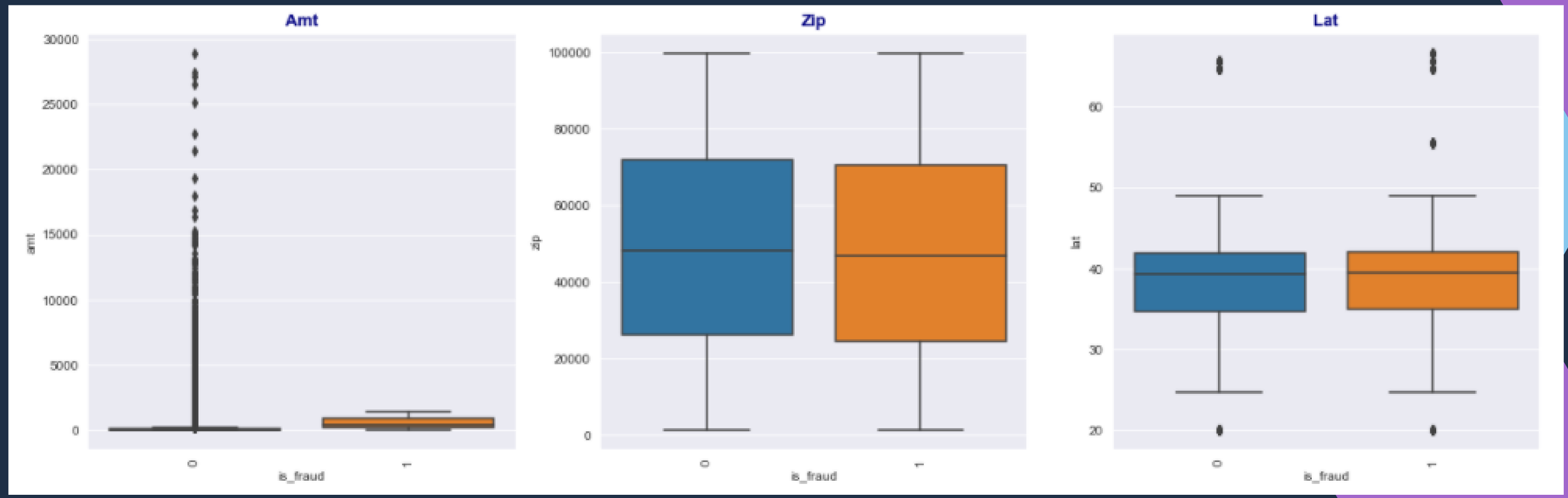
UNIVARIATE ANALYSIS

AMOUNT OF TRANSACTION



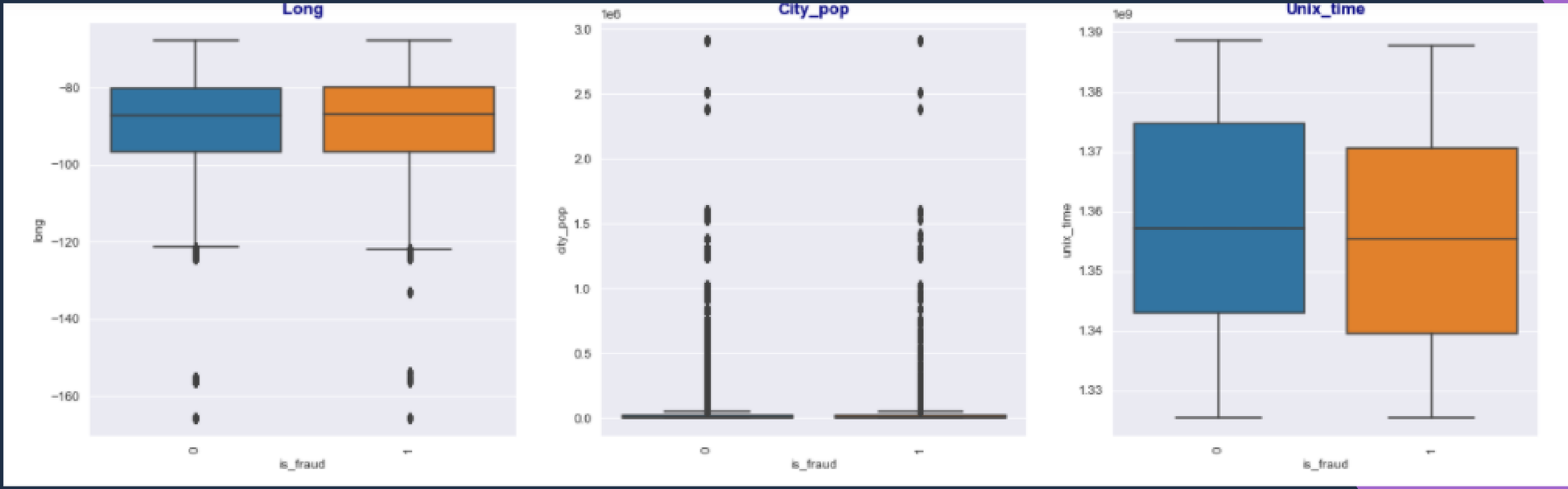
BIVARIATE ANALYSIS

NUMERICAL COLUMNS



BIVARIATE ANALYSIS

NUMERICAL COLUMNS



SUMMARY

Columns like city_pop and amt are highly skewed.

Maximum transactions were done on Sunday and Monday.

Categories like home,grocery_pos and gas_transports has highest amount of transactions above 175000.

SUMMARY

Most of the transactions are been done by female customers.

250000+ transactions were done in December month.

Majority of the transactions done by customers are of 30-50 years age group.

SUMMARY

Most of the transactions are been done by the customers having job as Film video editor, Exhibition designer and surveyor of land or geomatics.

Majority of transaction is done from the Kilback LLC merchant.

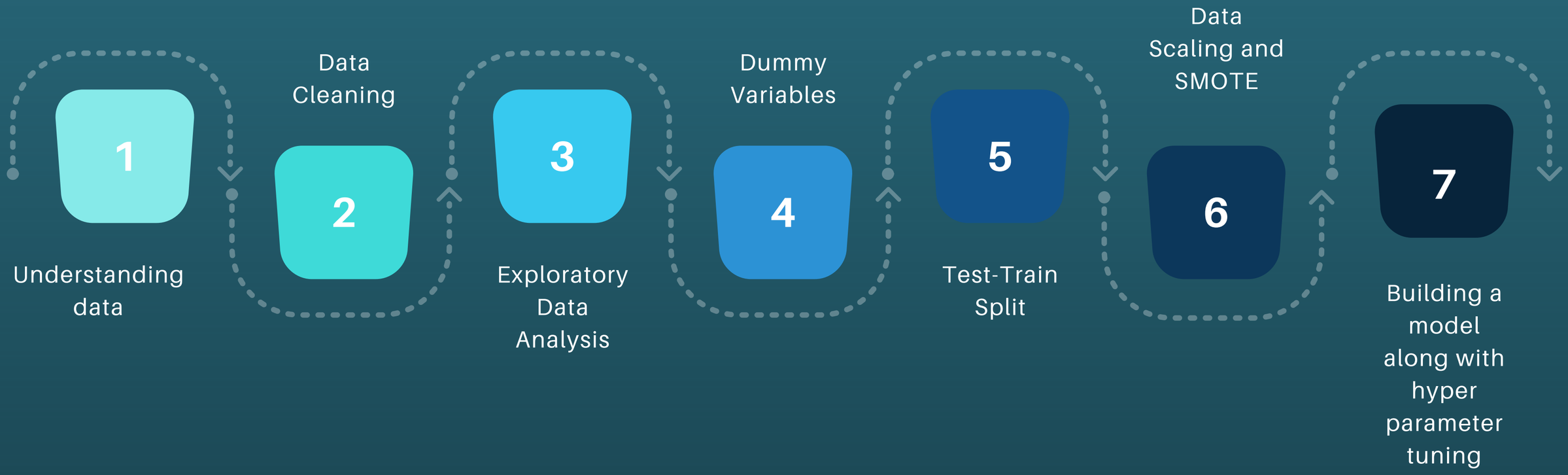
A heavy amount of non-fraudulent transactions has been done.

COST-BENEFIT ANALYSIS

- Avg no. of transaction per month: 77183.08
- Avg no. of fraudulent transactions per month: 402.12
- Avg amount per fraudulent transaction: 530.66

- An average number of transactions per month detected as fraudulent by the model: 1731.22
- An average number of transactions per month that are fraudulent but not detected by the model: 21.14


METHODOLOGY


1
Understanding data
Data Cleaning
2
3
Exploratory Data Analysis
Dummy Variables
4
5
Test-Train Split
Data Scaling and SMOTE
6
7
Building a model along with hyper parameter tuning

Oversampling is been done for adjusting the class distribution of a dataset. For that, SMOTE is used.

Different models like Logistic regression, Decison Tree and Random Forest Classifier were used.

Decision trees and random forest showed promising results with high accuracy, precision and recall. Helpful enough for the detection of fraudulent transaction.


THANK YOU