OVERVIEW

- The Annual Data Book compiled by the Federal Trade Commission reports that Credit card fraud accounted for 393,207 of the nearly 1.4 million reports of identity theft in 2020, making credit card fraud the second most common type of identity theft!
- • Some surveys suggest that a typical organization loses 5% of their yearly revenues to fraud. As consumers increasingly switch from cash to credit card transactions, these numbers are only like to increase even more.

Combating Credit Card Fraud

- • To combat fraudulent credit card transactions, banks and other financial institutions have typically utilized algorithms that use strict rules to identify them when they occur.
- Unfortunately, these algorithms are often too rigid. As the complexity and scale of these fraudulent transactions have increased, many companies are starting to use more advanced models, such as machine learning models, to help identify and prevent credit card fraud

Goal of our Project

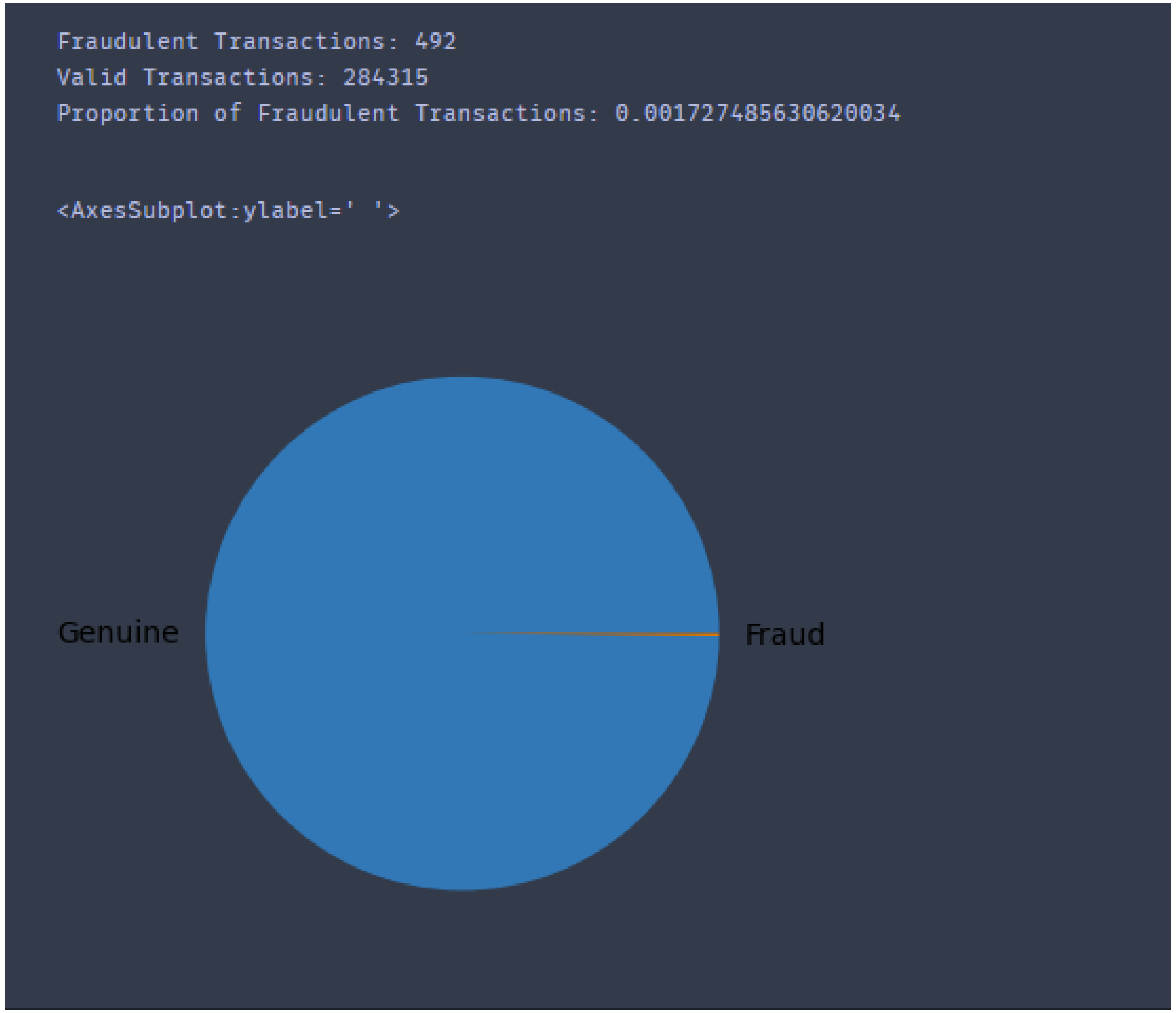
- The goal of this Credit Card Fraud Detection project is to classify a transaction as valid or fraudulent in a large dataset.
- Since we are dealing with discrete values (i.e., either fraudulent or non-fraudulent transactions), this is a binary classification problem, and we would employ the use of a supervised machine learning algorithm.

THE DATASET

The dataset used for this project has 284,807 rows of credit card transactions. Exploratory data analysis reveal as expected that we have a highly imbalanced dataset with only 0.17% of all transaction being fraud. While a large portion of the features have been anonymized with PCA, univariate and bivariate distribution plots show that the genuine transaction class has an approximately normal distribution across all features, and the fraud class was had a left skewed distribution for many of the features.

***MOVE THIS SLIDE AFTER GOAL OF OUR PROJECT

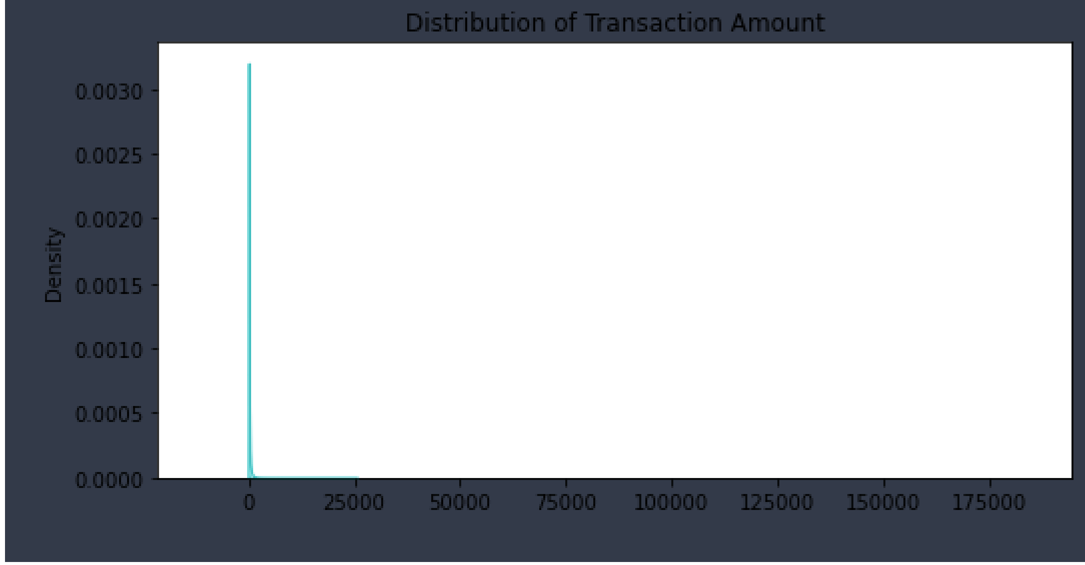
UNIVARIATE ANALYSIS



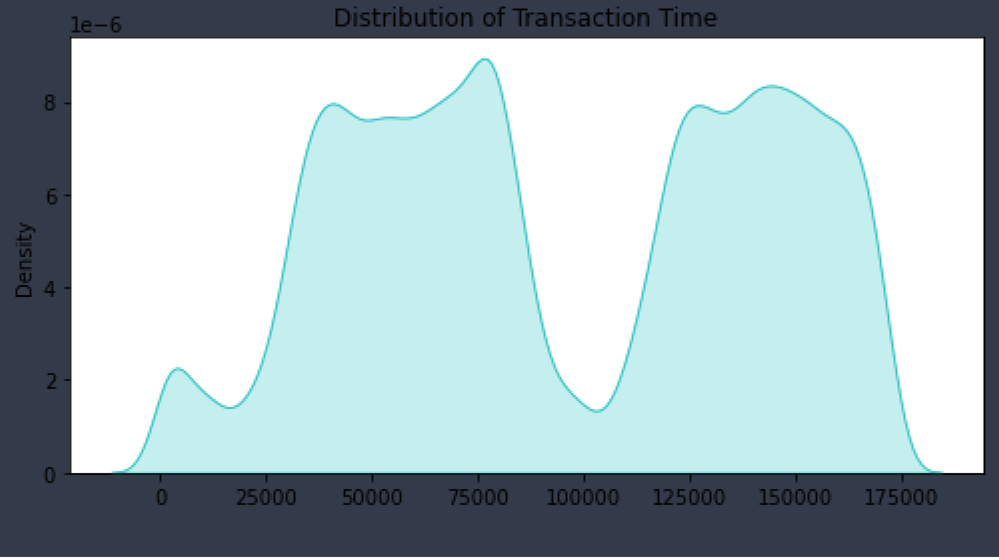
```
Fraudulent Transactions: 492
Valid Transactions: 284315
Proportion of Fraudulent Transactions: 0.001727485630620034


<AxesSubplot:ylabel=' '>
```

As you can see from the pie chart, our dataset is highly imbalanced. The percent of fraudulent transactions is less than 1% of the entire dataset.

UNIVARIATE ANALYSIS CONT.



Distribution of Transaction Amount

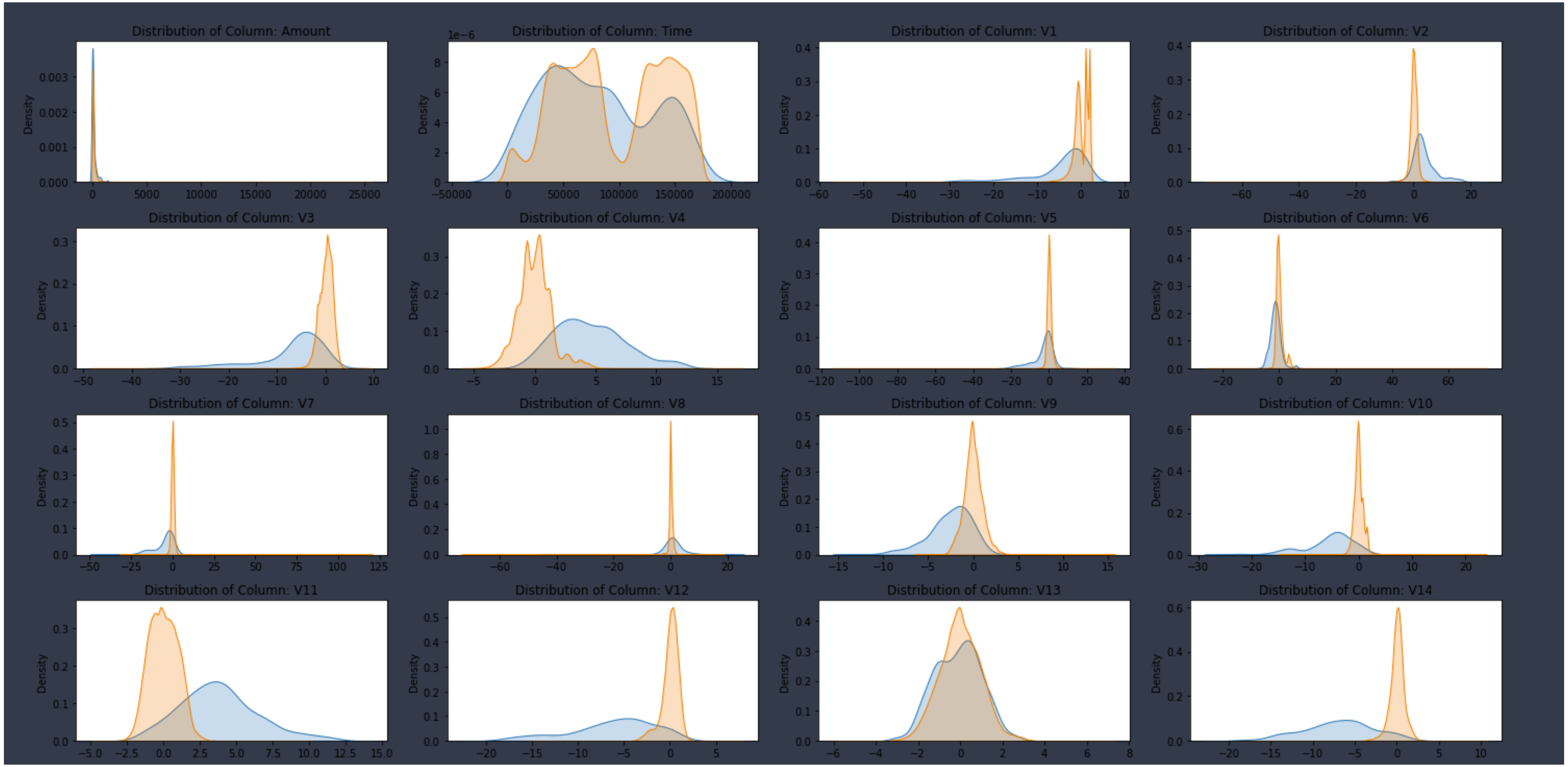


Distribution of Transaction Time

The distribution of transaction amount shows there some very large outliers in this dataset.

The distribution of the time these transactions take spans across two days.

Bivariate Analysis



Bivariate plots of all features grouped by transaction class, showed that the valid transaction class has a normal distribution shape across most of the features, conversely, the fraud class show long-tailed distribution across many of the features

- DELETE SLIDE 38. (repeat of slide 29)