**Problem Statement**

The dataset contains transactions made by credit cards. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, the original features and more background information about the data cannot be found. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

**Data Analysis**

To analyze this data and to predict the correct results, linear regression model is used. The model was used to find highly correlated variables with the ‘Class’ and then predict on its basis the correct numbers of “fraudulent” and “non-fraudulent” transactions.

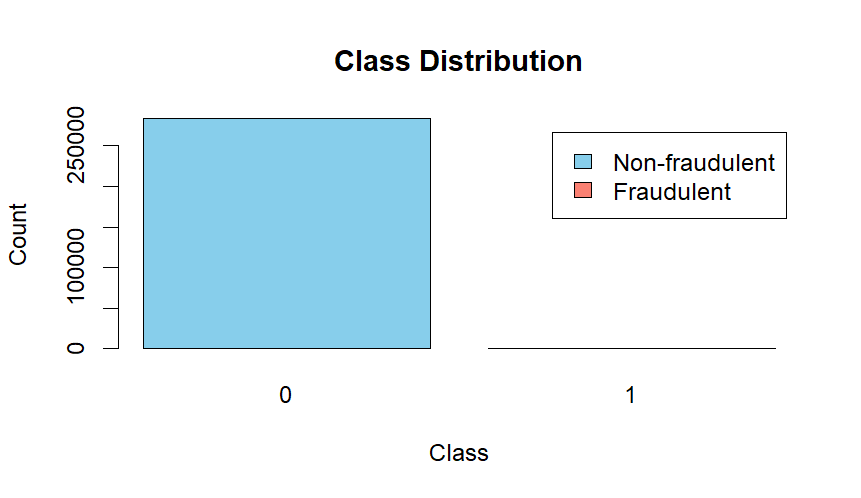


Figure 1: Exploring and Processing the Current Data

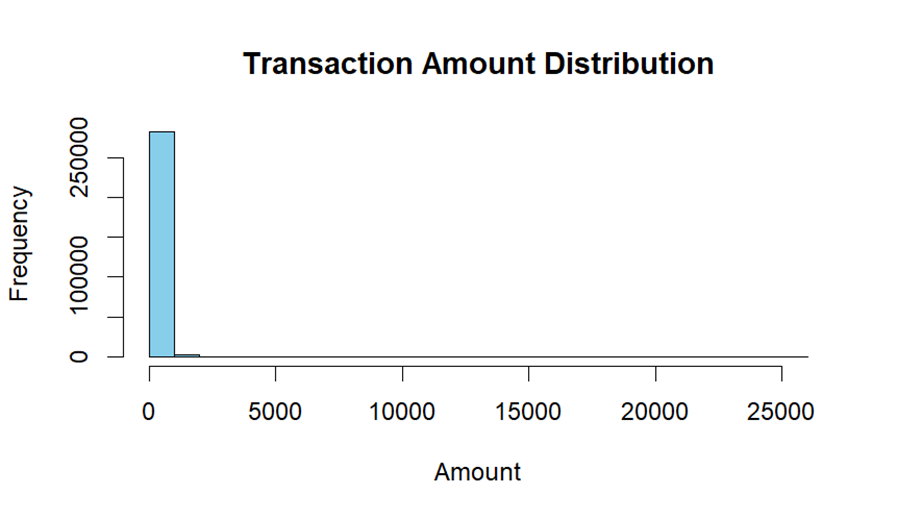


Figure 2: Current Data

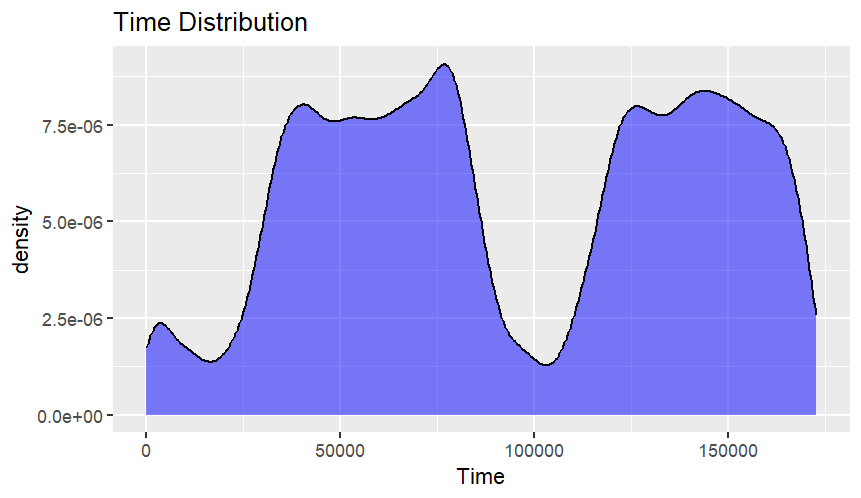


Figure 3: Density Map for Time and Frequency of Transactions

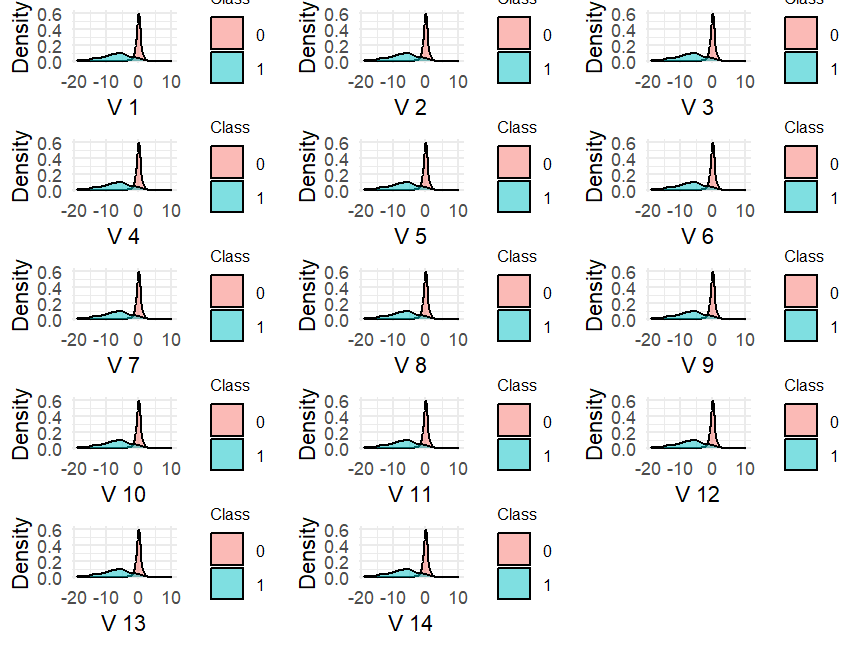


Figure 4: Density map for each variable and its correlation with Class

V17 V14 V12 V10 V16 V3

0.3264810672 0.3025436958 0.2605929249 0.2168829436 0.1965389403 0.1929608271

V7 V11 V4 V18 V1 V9

0.1872565915 0.1548756447 0.1334474862 0.1114852539 0.1013472986 0.0977326861

V5 V2 V6 V21 V19 V20

0.0949742990 0.0912886503 0.0436431607 0.0404133806 0.0347830130 0.0200903242

V8 V27 Time V24 V13 V26

0.0198751239 0.0175797282 0.0123225709 0.0072209067 0.0045697788 0.0044553975

V15 V25 V23 V22

0.0042234023 0.0033077056 0.0026851557 0.0008053175

*The Highly correlated variables with “Class”*

**Results**

Predicted

Actual 0 1

0 85276 13

1 62 92

The confusion matrix you provided shows the results of predictions made by a machine learning model on a testing dataset. Each row of the matrix represents the actual class labels, while each column represents the predicted class labels. Here's what each cell in the confusion matrix indicates:

* **True Negatives (TN):** The number of instances where the actual class is negative (0) and the model correctly predicted it as negative.
* **False Positives (FP):** The number of instances where the actual class is negative (0) but the model incorrectly predicted it as positive (1).
* **False Negatives (FN):** The number of instances where the actual class is positive (1) but the model incorrectly predicted it as negative.
* **True Positives (TP):** The number of instances where the actual class is positive (1) and the model correctly predicted it as positive.

Here's a breakdown of what each cell indicates:

* The top-left cell (85276): Indicates the number of non-fraudulent transactions (actual class 0) that were correctly classified as non-fraudulent by the model.
* The top-right cell (13): Indicates the number of non-fraudulent transactions (actual class 0) that were incorrectly classified as fraudulent by the model.
* The bottom-left cell (62): Indicates the number of fraudulent transactions (actual class 1) that were incorrectly classified as non-fraudulent by the model.
* The bottom-right cell (92): Indicates the number of fraudulent transactions (actual class 1) that were correctly classified as fraudulent by the model.

This confusion matrix provides valuable information about the performance of the model, including its accuracy, precision, recall, and other evaluation metrics. Let me know if you need further explanation or assistance!