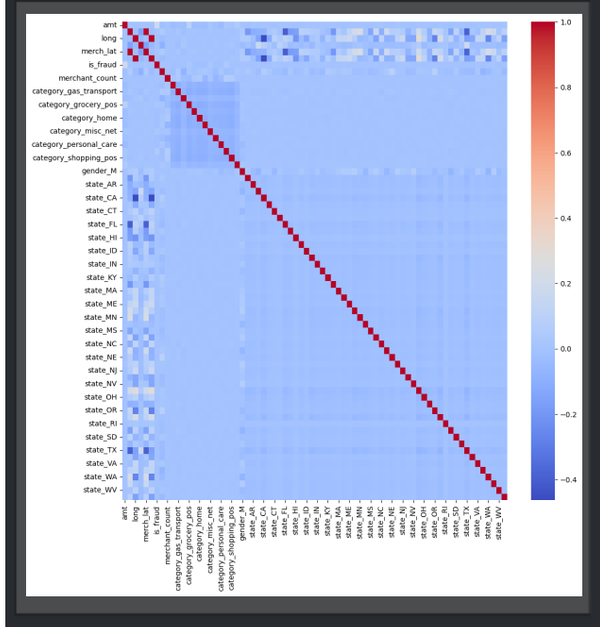
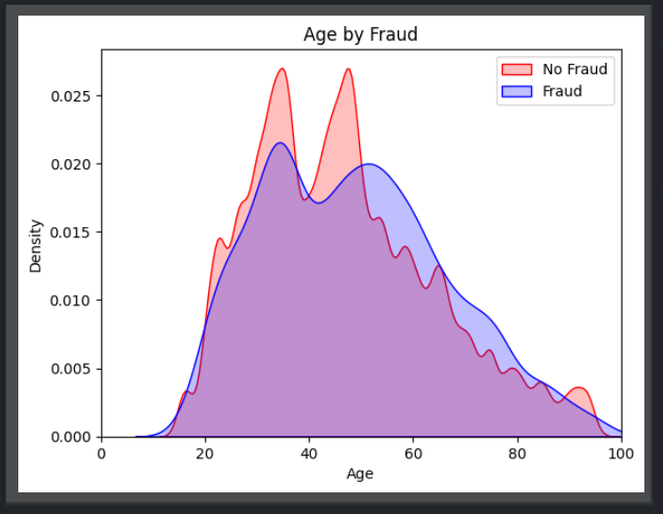
Credit Card Fraud

Khayyam Khan

This project is about finding possible variables that affect credit card fraud and creating a model that will predict it.

From this heatmap we can deduce that the main predictor of credit card fraud is the transaction amount. Location, category, merchant amount, and victim age have small predictive power.



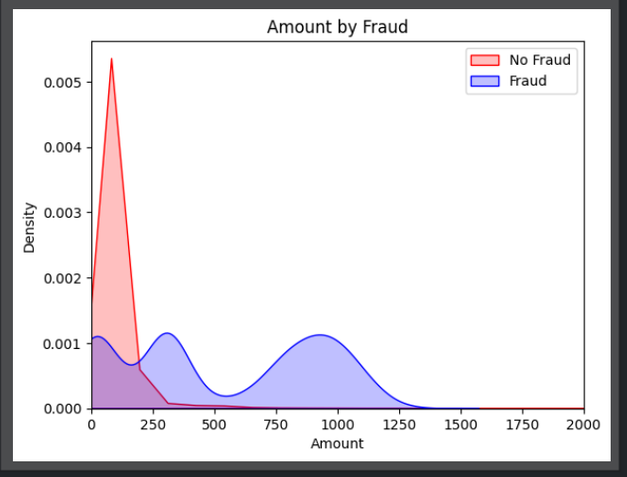


As we can see, all age ranges are subject to credit card fraud, however, it is noted that there is a higher percentage of a given transaction to be fraud if the card holders age is greater than 50.

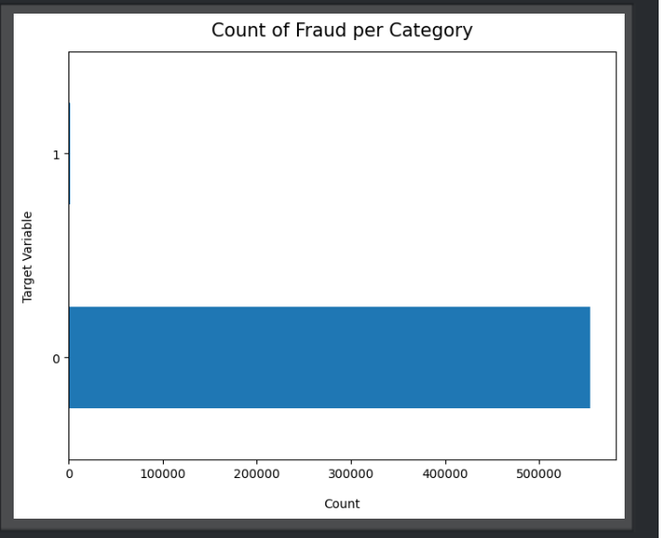
The majority of transactions occur under the $300 mark.

There is a higher chance that a given transaction will be fraudulent if the tendered amount is greater than $300.

Fraudulent transactions peak at around the $20, $300, and $1000 price.



This dataset is very imbalanced. Oversampling is necessary!



Model has 94% accuracy with great precision, recall, and f1-score after oversampling.

0.9424647381707854

	precision	recall	f1-score	support
0	0.97	0.92	0.94	106646
1	0.92	0.97	0.94	109239
accuracy			0.94	215885
macro avg	0.94	0.94	0.94	215885
weighted avg	0.94	0.94	0.94	215885