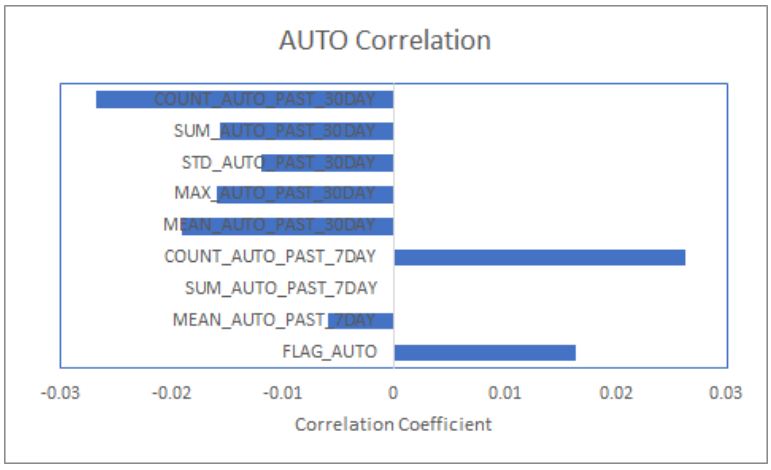
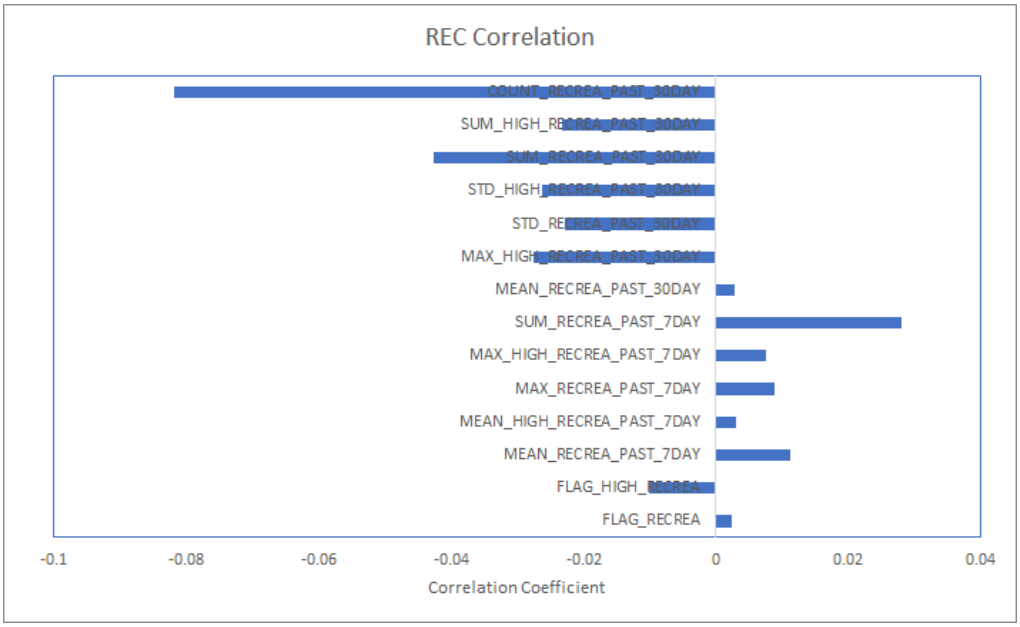
How to prevent fraud more effectively without creating operation overhead?
Which transactions should we decline to reduce fraud?
What are some key attributes that help to make the decline decision?
How to minimize the negative impact to customer experience while preventing fraud?
Can you make any long-term suggestions

We developed a transaction fraud detection model using a generalized linear model. This generates a conventional linear regression model for a continuous response variable given continuous or categorical predictors. In the case of our data, we generated correlation data from the numerous categories given in the dataset, and determined the likelihood of fraud. We separated our correlation data into categories such as AUTO, REC, and more. As seen with the figures below, we graphed each category's correlation data with histograms to visualize the correlation between a specific category and likelihood of fraud.





We also separated our correlation data into categories based on whether they were FLAGS, MEANS, SUMS,

Beyond our work with the generalized linear model, we attempted to use the K-Nearest Neighbours (KNN) algorithm to analyze the data. This would classify the new data points based on the similarity measure of the earlier stored data points, and allow for more versatility and higher accuracy for predictions given our dataset. Unfortunately, we attempted to use the KNN model and it took too long to get any results before the competition's deadline.

To improve our transaction fraud detection model, using a decision tree would potentially be more efficient. As well, having more data would have made it possible to create a better generalized model.