HACKATHON – 2
Credit Card Consumption Prediction

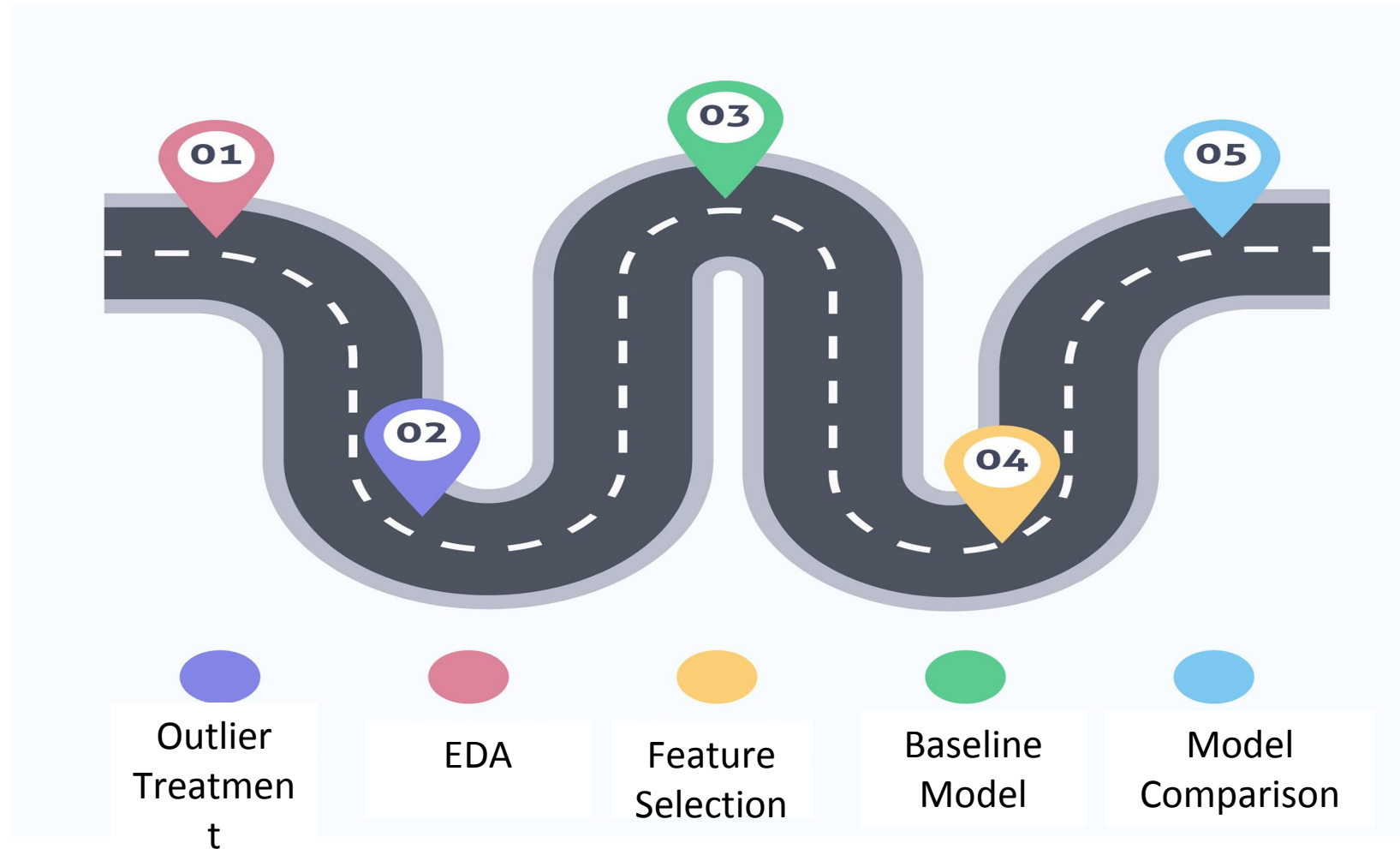## TEAM - NSVD

Dashang Makwana

Nevina Dalal

Simran Mulani

Vishakha Singh

Problem Statement

Common Man Bank wants to introduce offers and plans for the upcoming quarter based on the expenditures of their customers in the previous quarter, based on the demographics given we were asked to predict the average spend for the next three months.

Our Approach

Datasets

We were provided with train and test datasets with the following demographics:

Credentials & Personal Info

Expenditures

Loans & Investment

Average Credit Card Consumption

Preprocessing

- Removed Outliers using winsorize
- Binned age groups for better visualization
- Removed region code because it had discrete value and it showed that it doesn't affect the target variable
- Label encoded the remaining categorical variables so that regression model can take them as values
- Used Standard Scalar on the X column so that all the variable get standardized with mean=0 and standard deviation = 1

EDA
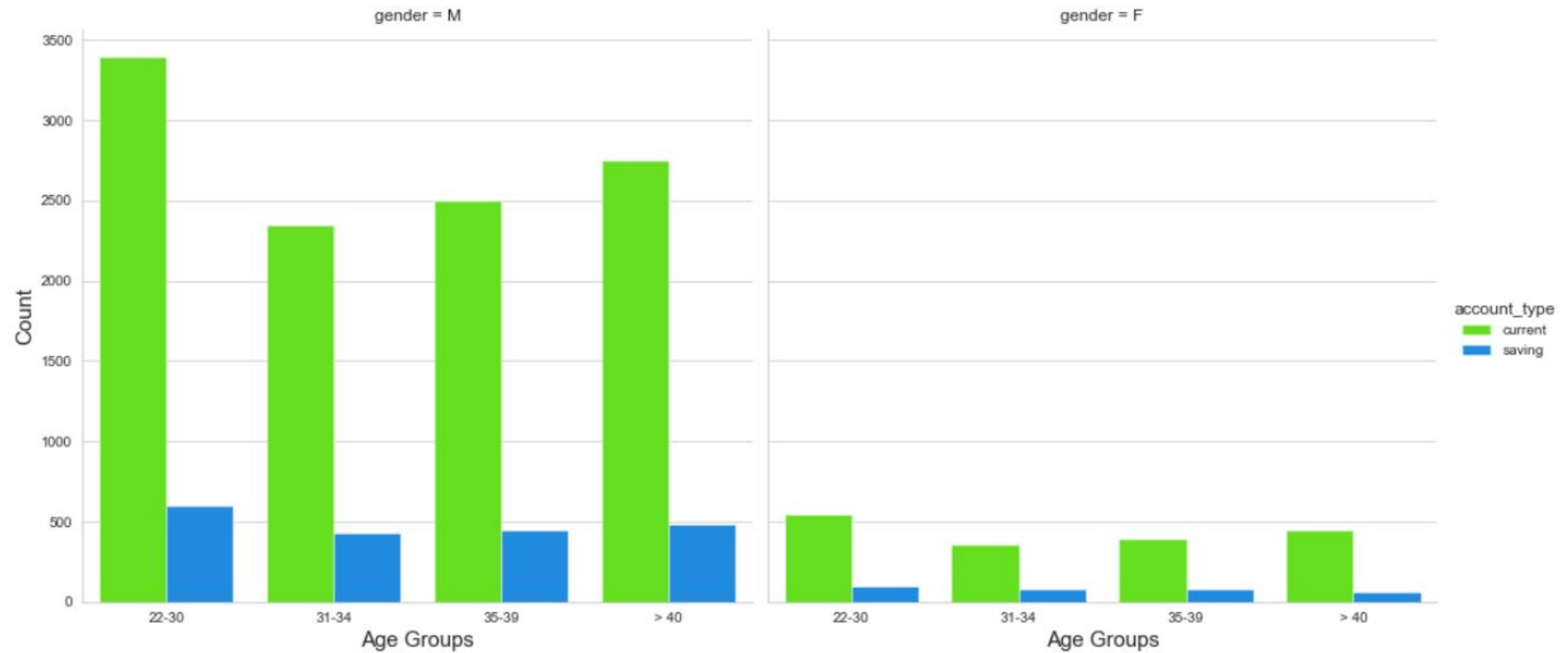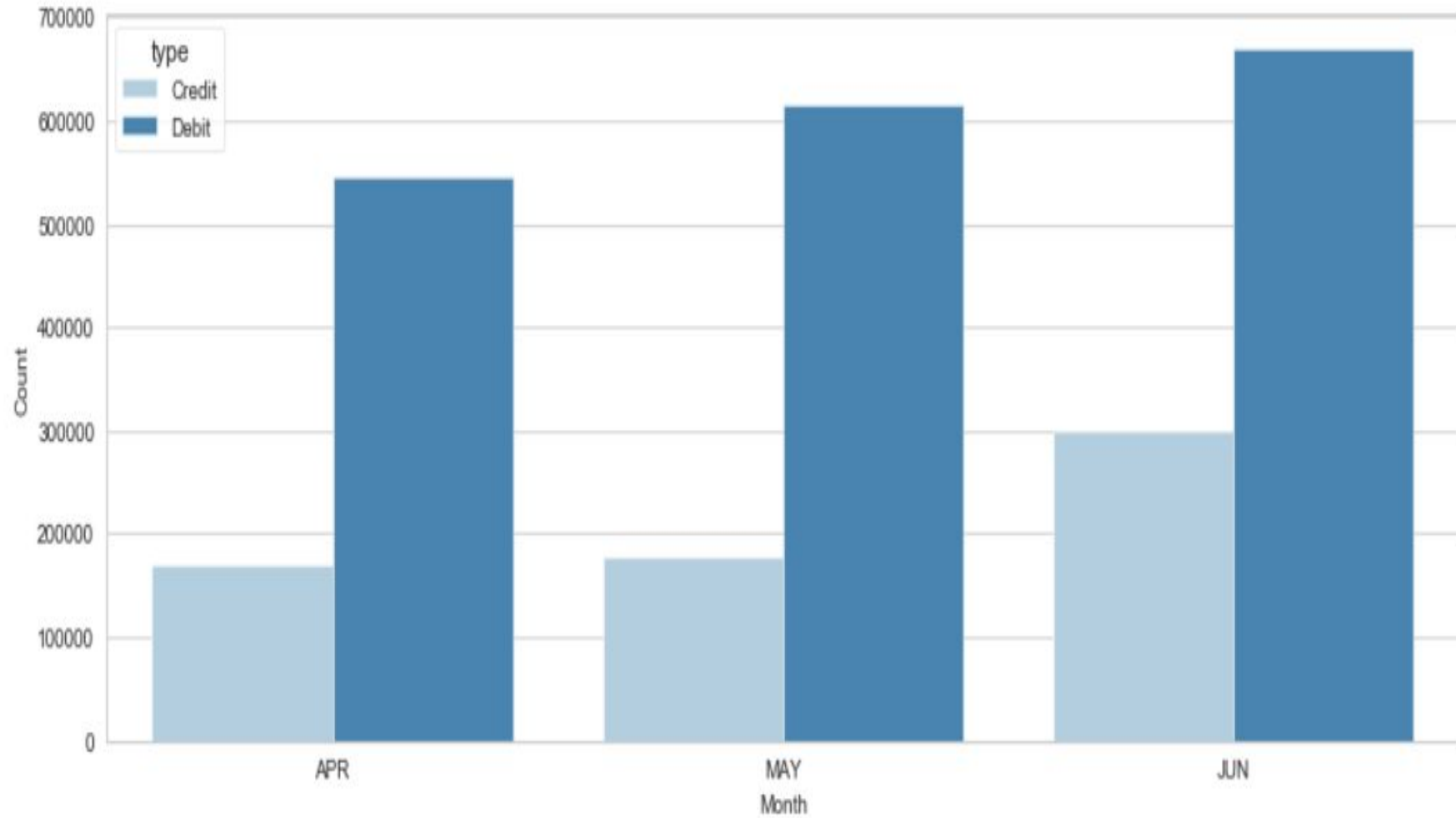

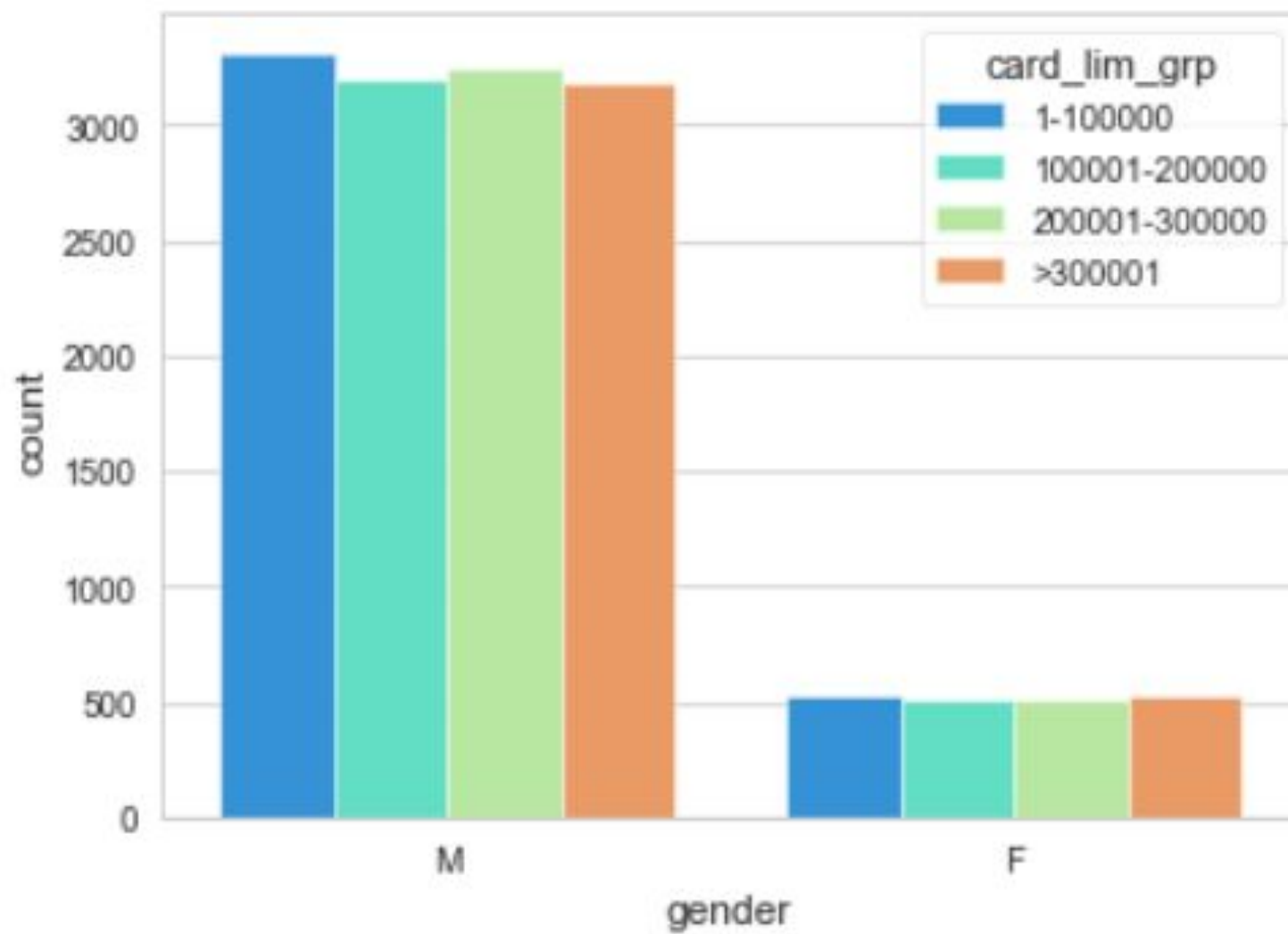Count of Age groups based on Account type and Gender

Correlation and it's Normalised Form
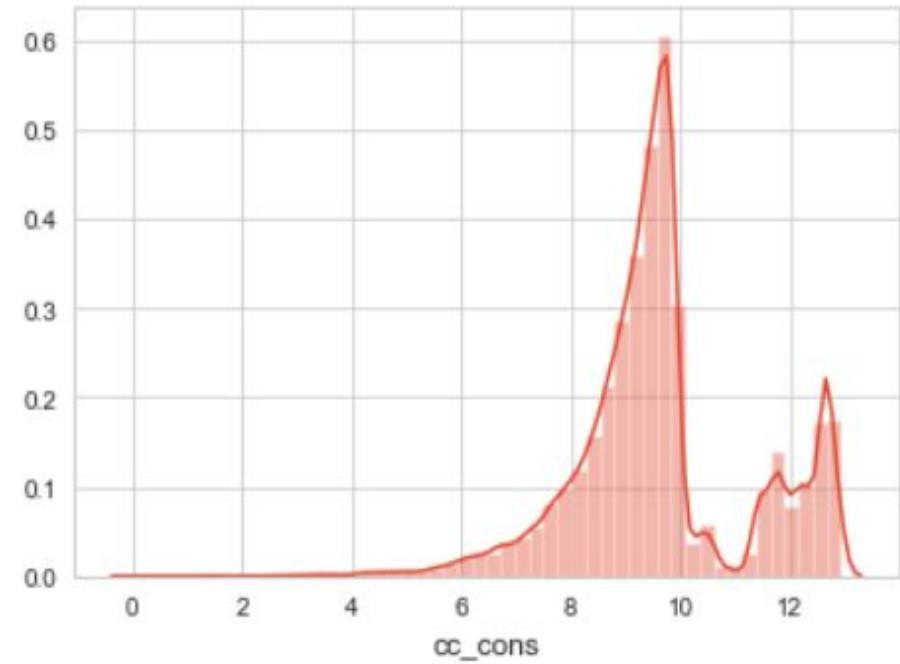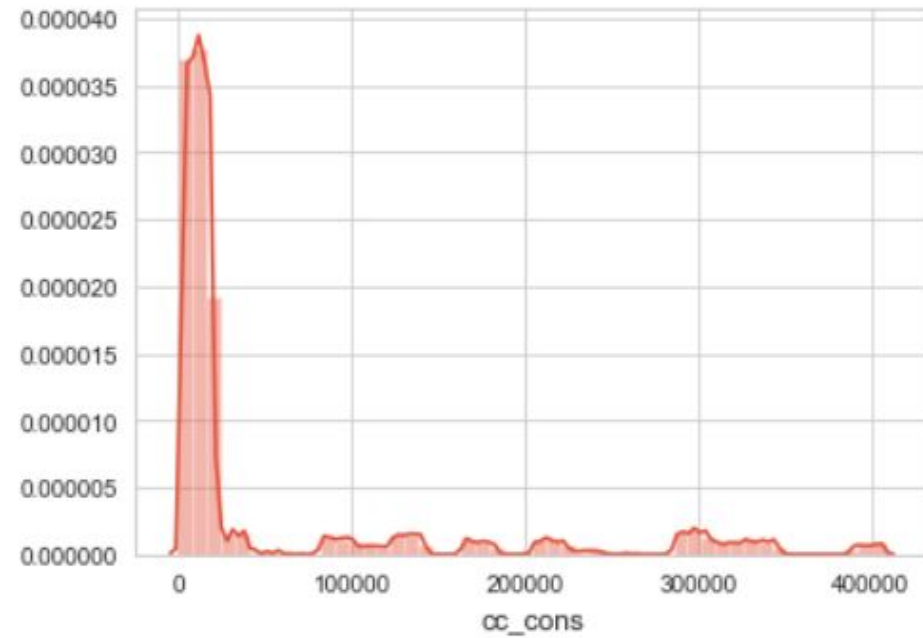
Debit and Credit Card Usage

Gender based Card Limits

Evaluation Method Used

- RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

- RMSLE (Root Mean Squared Log Error)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(x_i+1) - \log(y_i+1))^2}$$

Models Used

- Linear Regression with all variables
- Lasso Regression with all variables
- Ridge regression with all variables
- Decision Trees with all variables
- Linear Regression with feature selection
- Random Forest with feature selection

Linear Regression

1. Linear Regression is a supervised learning algorithm.

2. It is used to determine relationship between dependent variable and one or more independent variables.
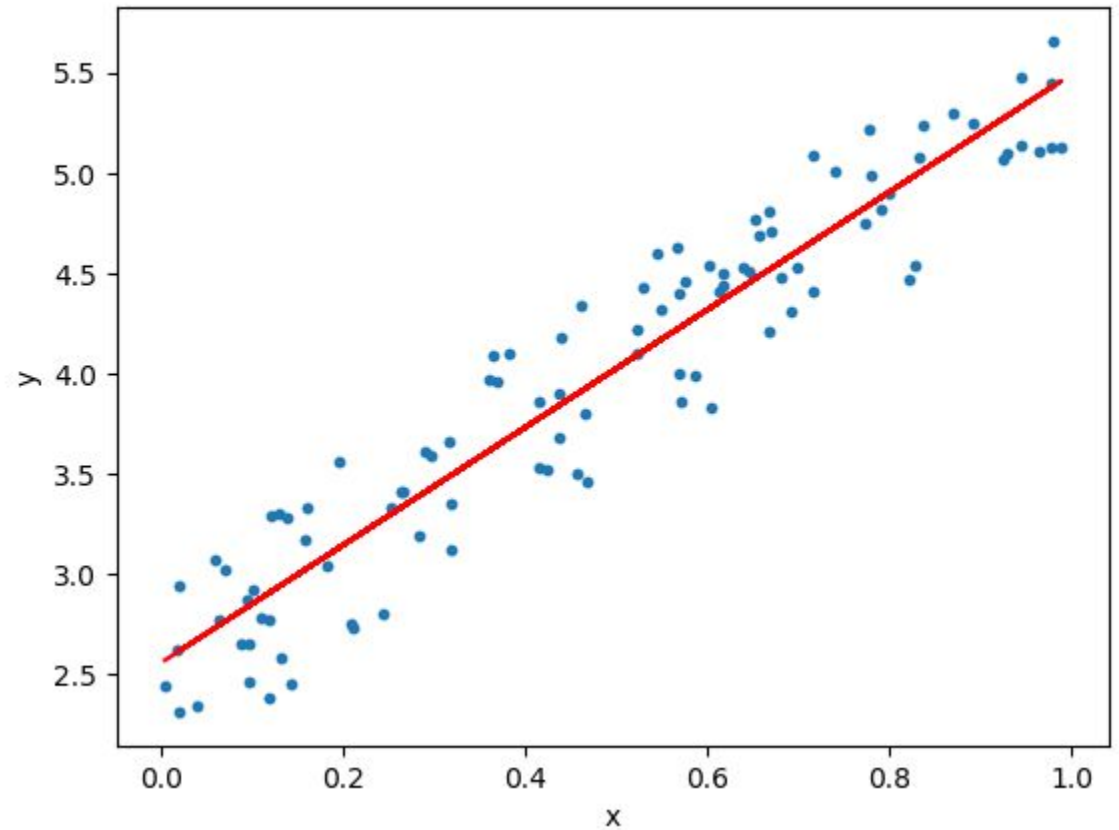
Lasso and Ridge Regression

- Ridge or lasso are forms of regularized linear regressions

- Ridge and Lasso regression are some of the simple techniques to reduce model complexity and prevent over-fitting which may result from simple linear regression. Ridge regression Ridge regression increases regression bias

- Lasso is far method of adjusting the parameters of a regression model. A method of regression that reduces the variance of these estimators.

Decision Trees

- Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results

- Decision trees regression normally use mean squared error (MSE) to decide to split a node into two or more sub-nodes.

Feature Selection

Feature selection is the process of choosing variables that are useful in predicting the response (Y).

The Recursive Feature Elimination (RFE) method works by recursively removing... number of required features as input... model to be used...

Random Forest

- Supervised Learning Algorithm works for both Regression and Classification . Part of sklearn.ensemble package.
- It is mostly preferred over Decision Trees
- It is type of Ensemble Learning
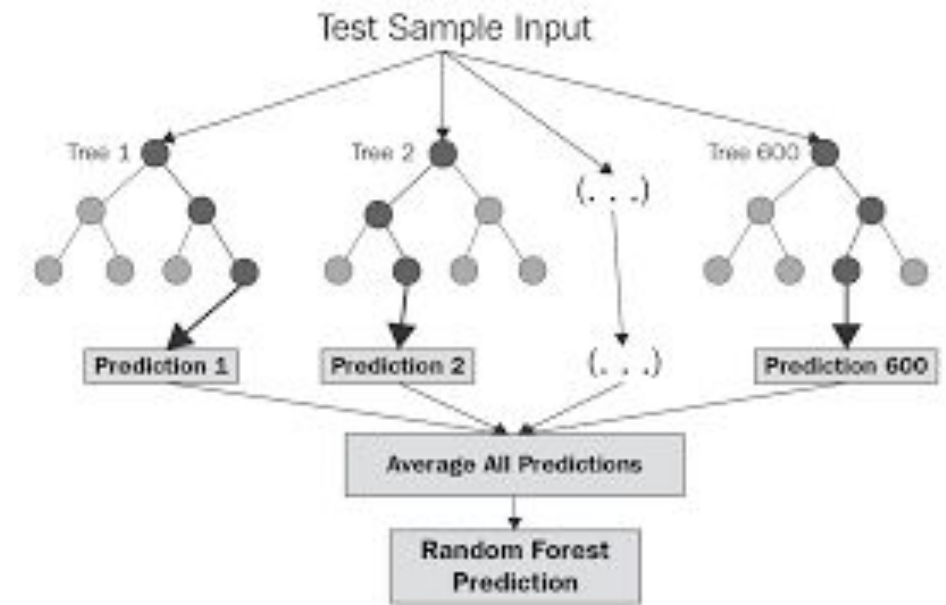
Approaches:

- Tried simple Random Forest
- Used RandomizedSearchCV and found best parameters and fine tuned out Random Forest Model

Model Comparison

Models	Train RMSE	Test RMSE	Platform Result
Multiple Linear Regression	1.6078	1.6146	1.612
Lasso Regularization	1.6014	1.6123	1.6120
Ridge Regularization	1.6077	1.6152	1.6127
Decision Tree	1.6104	1.6123	1.6104
Linear regression with Feature selection	1.6089	1.6121	1.6113
Random Forest with Feature Selection	1.6016	1.6138	1.6114

Business Insights

- People having debit cards have spend more.
  - Bank needs to provide more offers to increase Credit Card usage.
- Gender-wise Man has shown expense as compared to females.
  - As Females like to spend, Females concentric Marketing needs to be done
- Current Account sees more expenditure, but might be due to cash flows.
- In short, Marketing Campaign for Females and some special promotion for credit card to increase usage of credit card Age(20-35)

If more time was provided..

If more time was provided we would've tried another model for Lower RMSLE score, they are:

- XGBoost: XGBoost in an implementation of gradient boosted decision trees designed for speed and performance to improve the accuracy using bagging, boosting multiple models

THANK YOU ‼️