CREDIT CARD ANALYSIS

ANTHONY FOLEY · ERIN LEE · CHRISTIAN MURESAN

FINAL PROJECT ~ INTRO TO DATA SCIENCE

Overview

Image Source: https://www.handsomefrank.com/illustrators/abbey-lossing

Predictive Models

1. Logistic Regression
   - ❖ Predict customers who either default or make their payment on time (predicts either 0 or 1)
2. Linear Regression
   - ❖ Predict the credit limit that the bank should assign to each customer (continuous predictions)

Clustering Model

3. HAC
   - ❖ Look for any distinct groups or patterns of credit card users within the data

Dimensionality Reduction

4. Lasso
   - ❖ Reduce any effects of overfitting while maintaining model complexity
   - ❖ Use lasso for variable selection or ridge for reducing noise

Key Variables

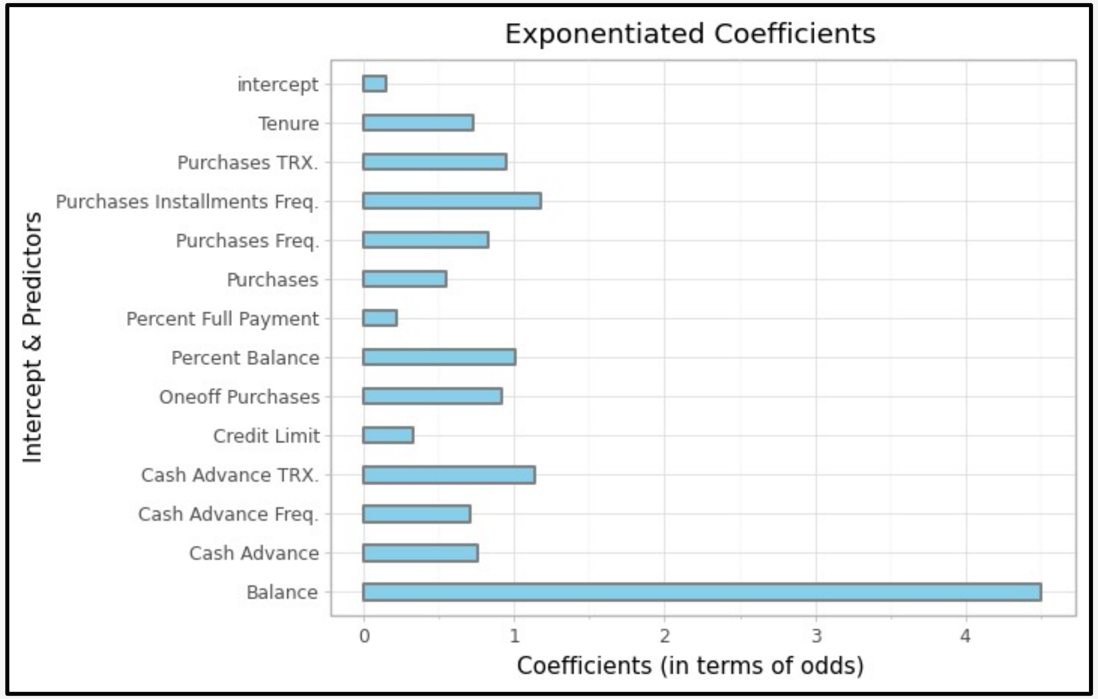
- Balance
  - Total amount of purchases still owed to bank
- Purchases
  - $ amount of purchases made from the account
- One-off Purchases
  - Amount of one-off purchases
- Installments Purchases
  - Amount of purchase done in installment
- Cash Advance
  - Cash in advance given by the user
- One-off Purchases Frequency
  - How frequently the purchases are occurring in one-go
- Purchase Installments Frequency
  - How frequently purchases in installments are being done
- Cash Advance Frequency
  - How frequently the cash in advance being paid
- Cash Advance Transactions
  - Number of transactions using Cash Advance
- Purchase Transactions
  - Number of purchase transactions made
- Credit Limit
  - Available line of credit extended to customer
- Payments
  - Payment amount by customer
- Minimum Payments
  - Minimum, required amount of money customer needs to pay
- Percent Full Payment
  - Percent of full payment paid by user
- Default
  - If a customer misses a payment or doesn’t pay the required min. amount
- Tenure
  - How many years has the individual been a customer of the bank

Image Source: https://www.quantamagazine.org/the-future-fabric-of-data-analysis-20131009/

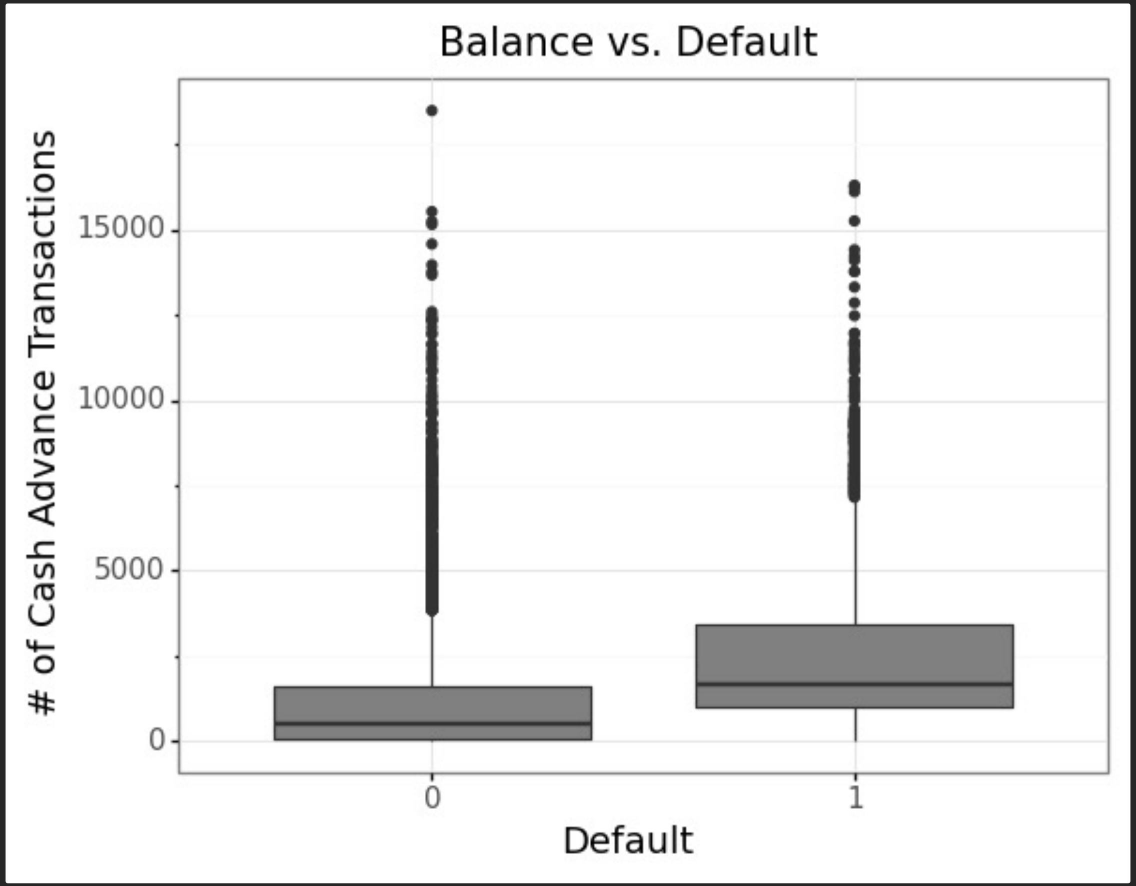
Which factors contribute the most to customers not being able to meet the minimum payments? Are these customers repeat offenders, or are they outliers?

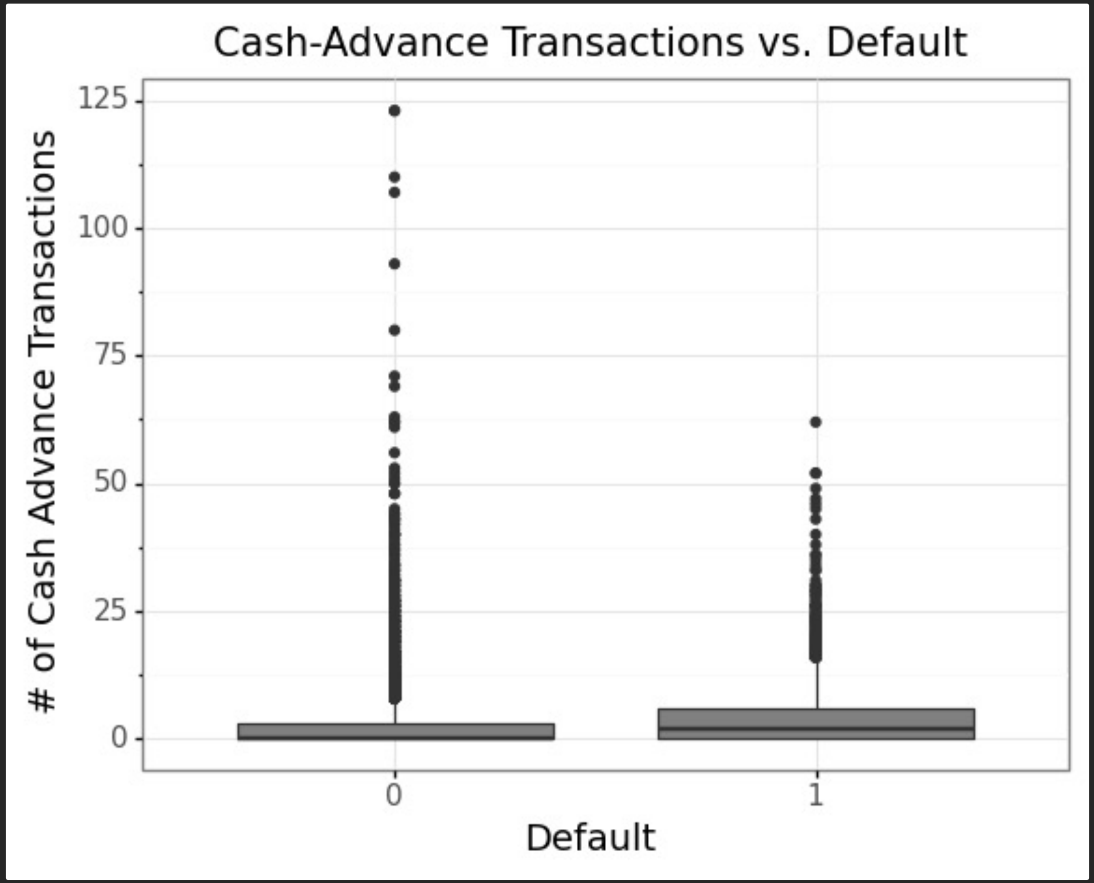
According to the odds coefficient results: *(logistic regression model)*

- ❖ "Balance" had the greatest impact on the model
- ❖ Second: Purchases Installments Frequency
- ❖ Third: Cash Advance TRX.

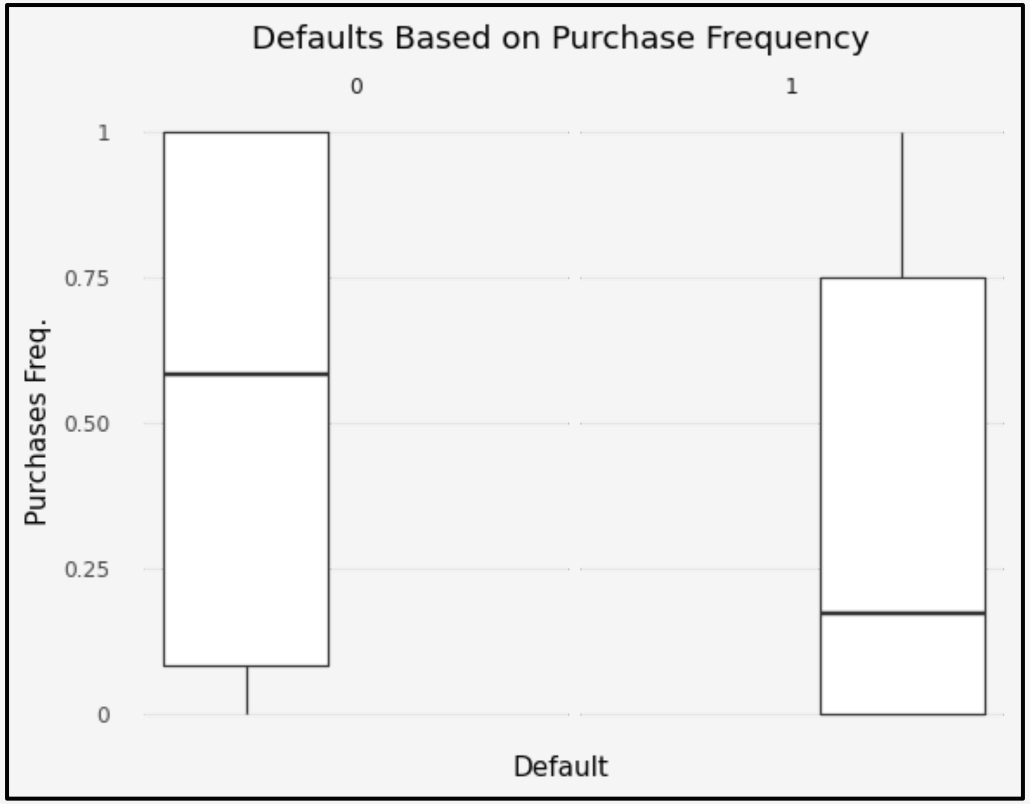


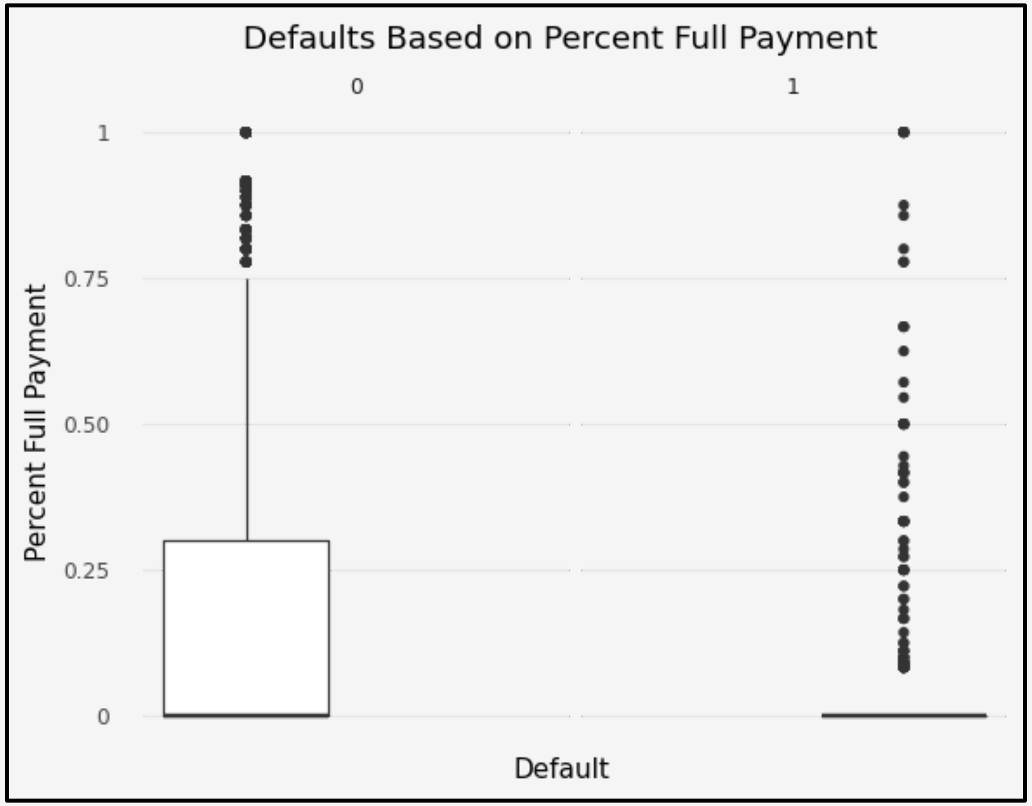
Important Variables and their Relationship to Default



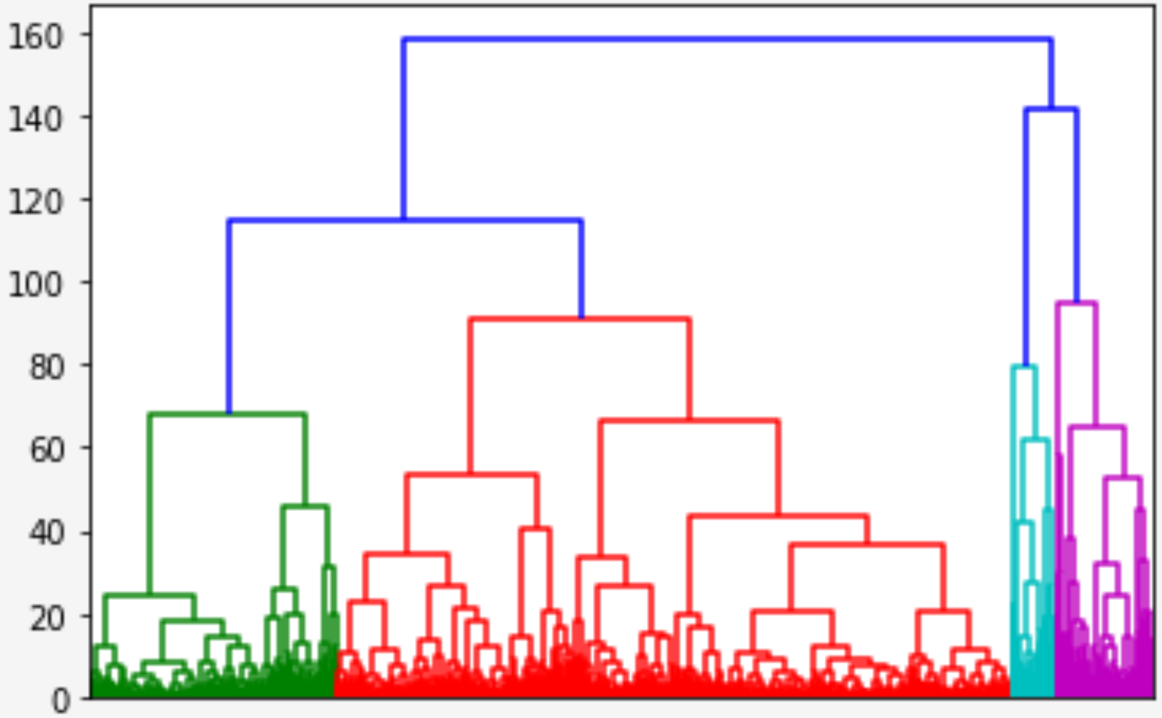


Do people who make frequent purchases or have a low percent of their full balance paid off also tend to miss their minimum payment (default)?





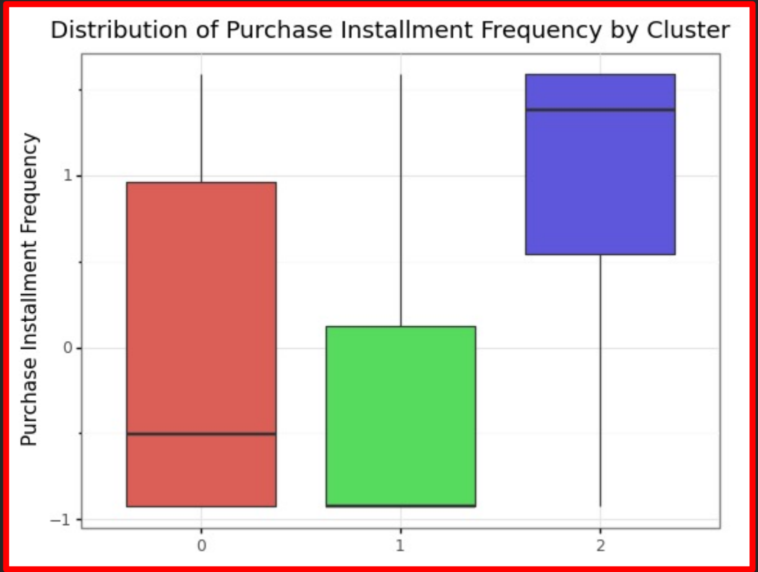
Is there a distinct group of users who frequently use a cash advance and use purchase installments frequently?

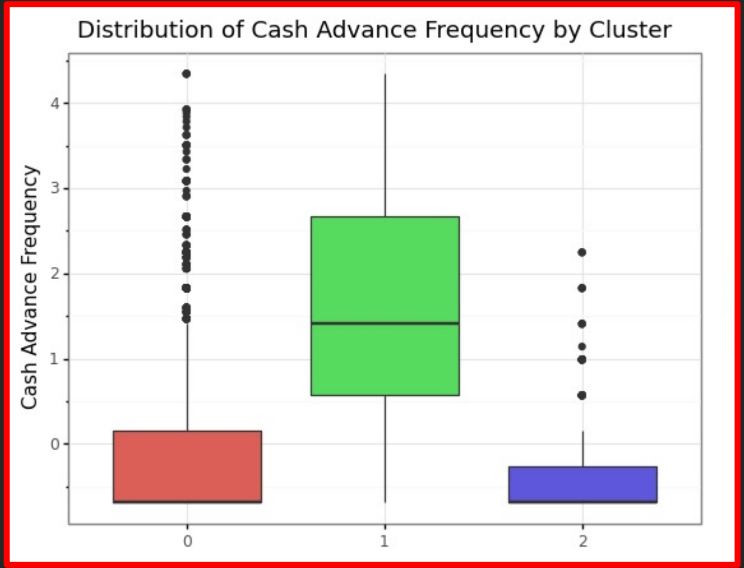


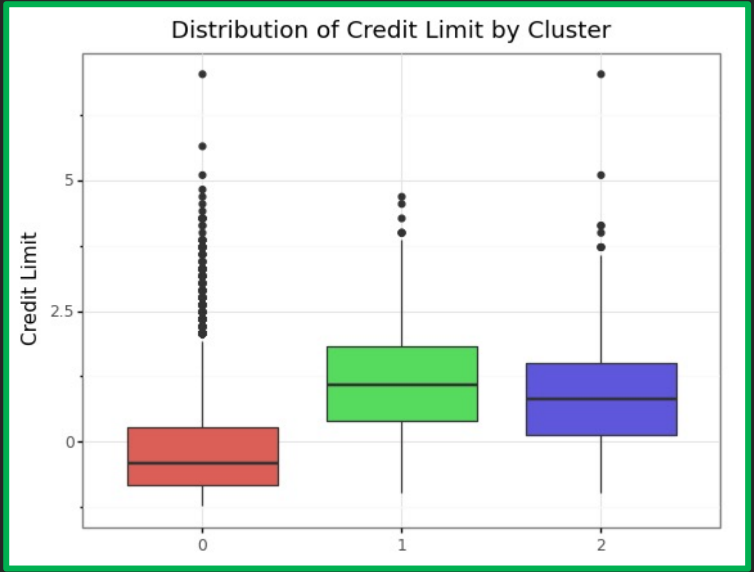
- Using a clustering method, we can identify different groups of customers and ask ourselves…
  - What are their spending patterns?
  - How much debt do they owe?
  - How high are their payments?

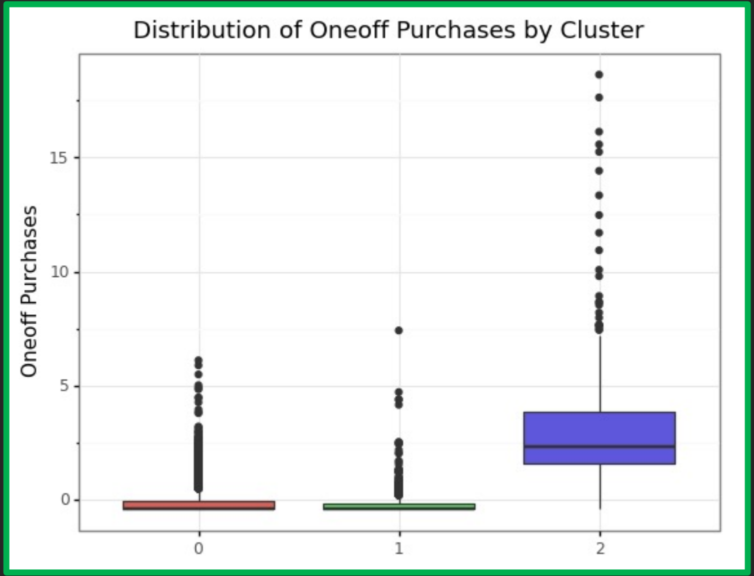
    …and so on

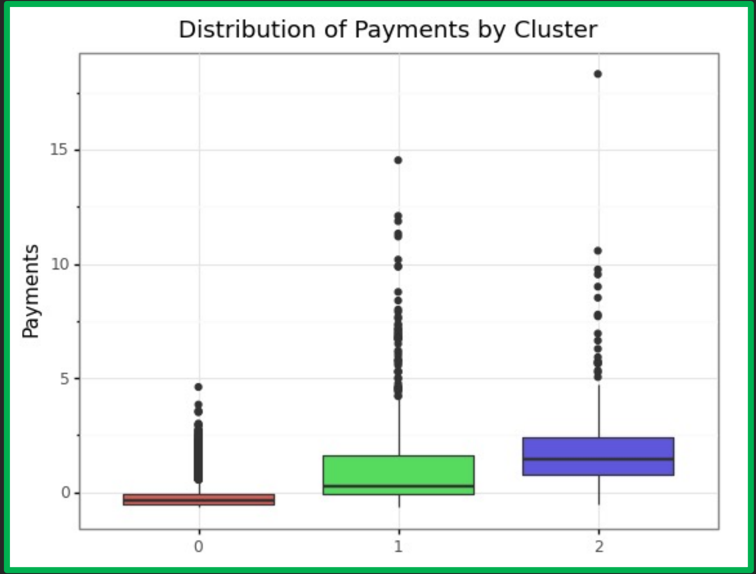
HAC Clustering Results





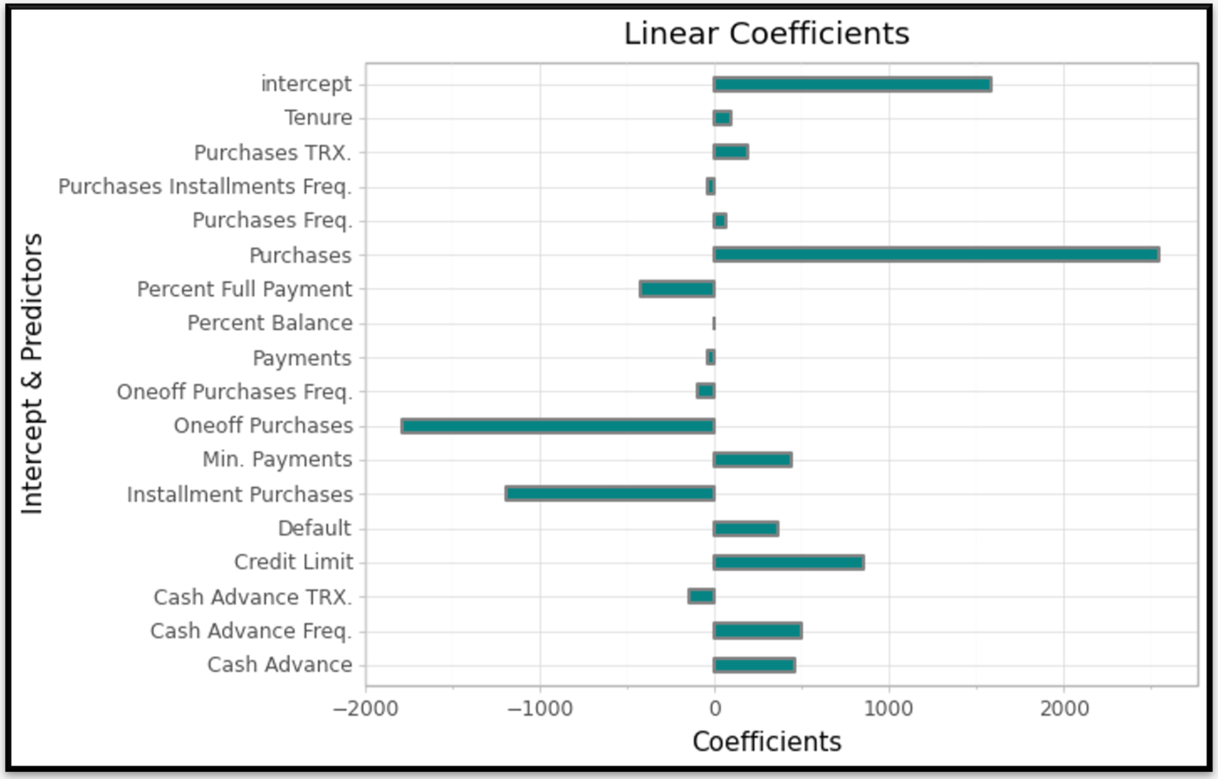




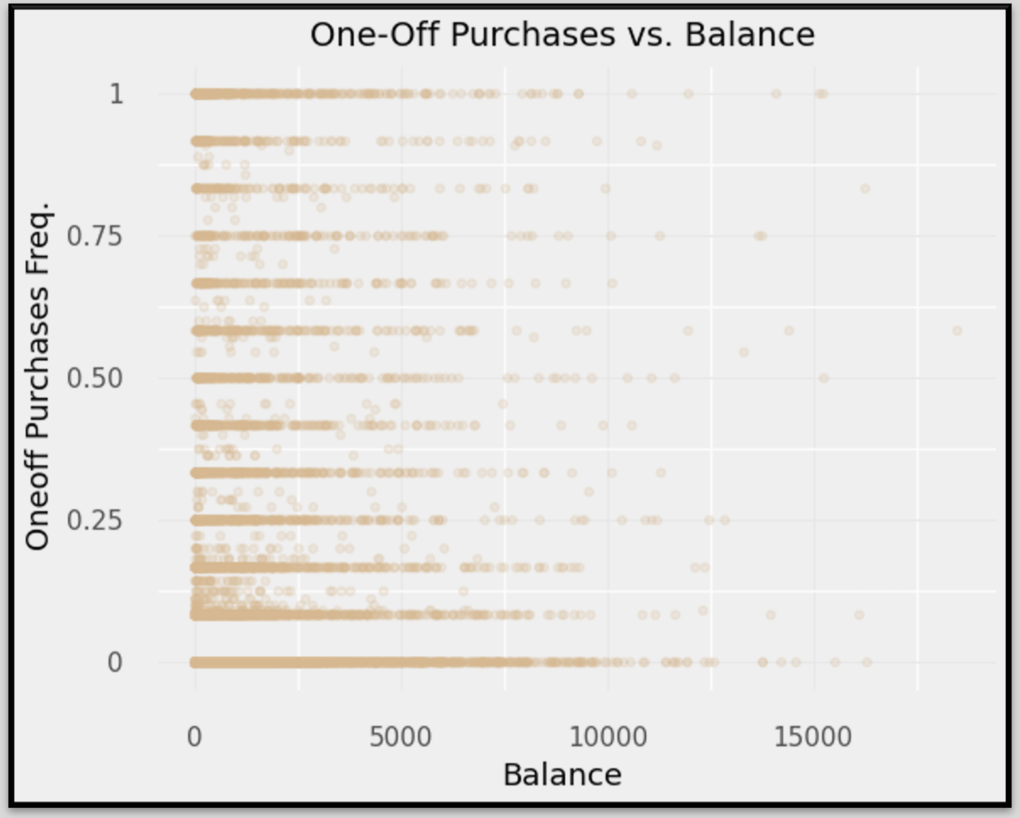


Are customers who frequently make one-off purchases more likely to have a higher balance on their credit card?

- Linear regression results:
  - Train R2: 62.33%
  - Test R2: 62.00%
  - With a coefficient of -101.5, our model had a high frequency of one-off purchases lowering predicted balance
  - Large $ amount of one-off purchases also negatively effected the amount of balance



One-Off Purchases and Balance



- Several possible ways to interpret the negative correlation
- Revolving balance
- Stricter credit terms for customers who spend too much, too often
- Big spenders who simply pay off items immediately

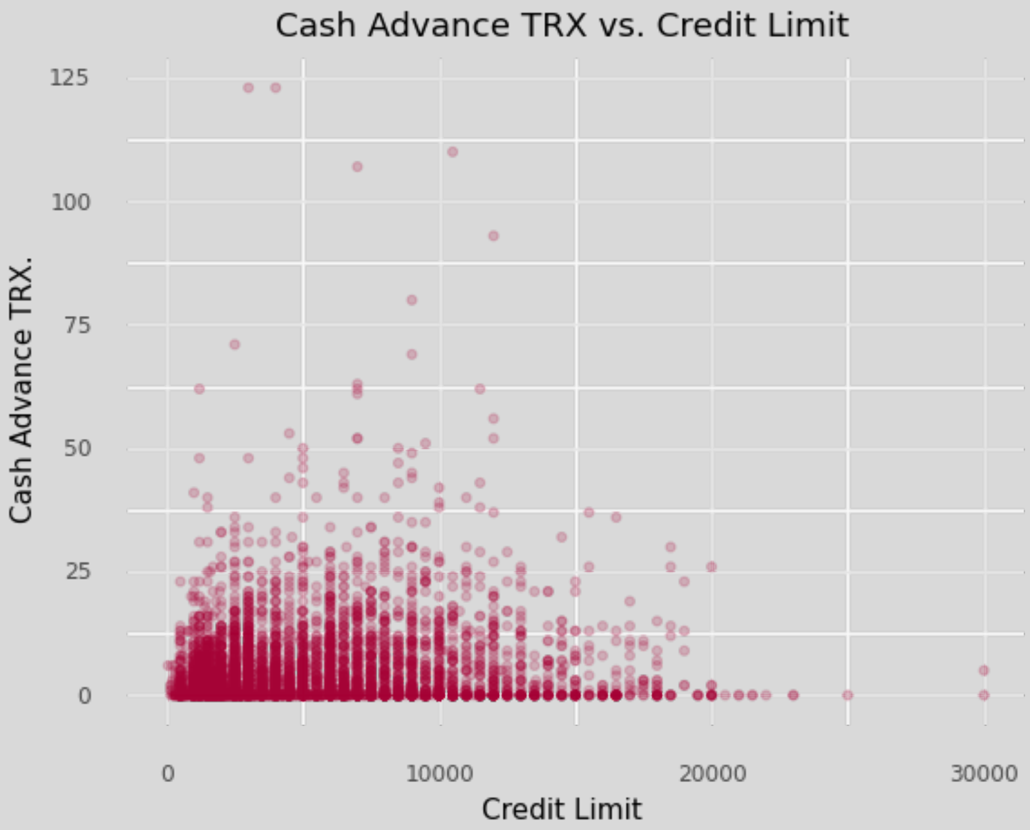
What credit limits would you recommend to the new customers based on their prior credit card statements?

- Customer 1
  - Average Balance: $3,000
  - Min. Payment: $1,800
  - Cash Advance Transactions: 10
  - Number of Transactions: 40
  - Min. Payment % : 60%

- Customer 2
  - Average Balance: $2,700
  - Min. Payment: $815
  - Cash Advance Transactions: 0
  - Number of Transactions: 22
  - Min. Payment %: 30%

Customer 1 Credit Limit: $2,112.93

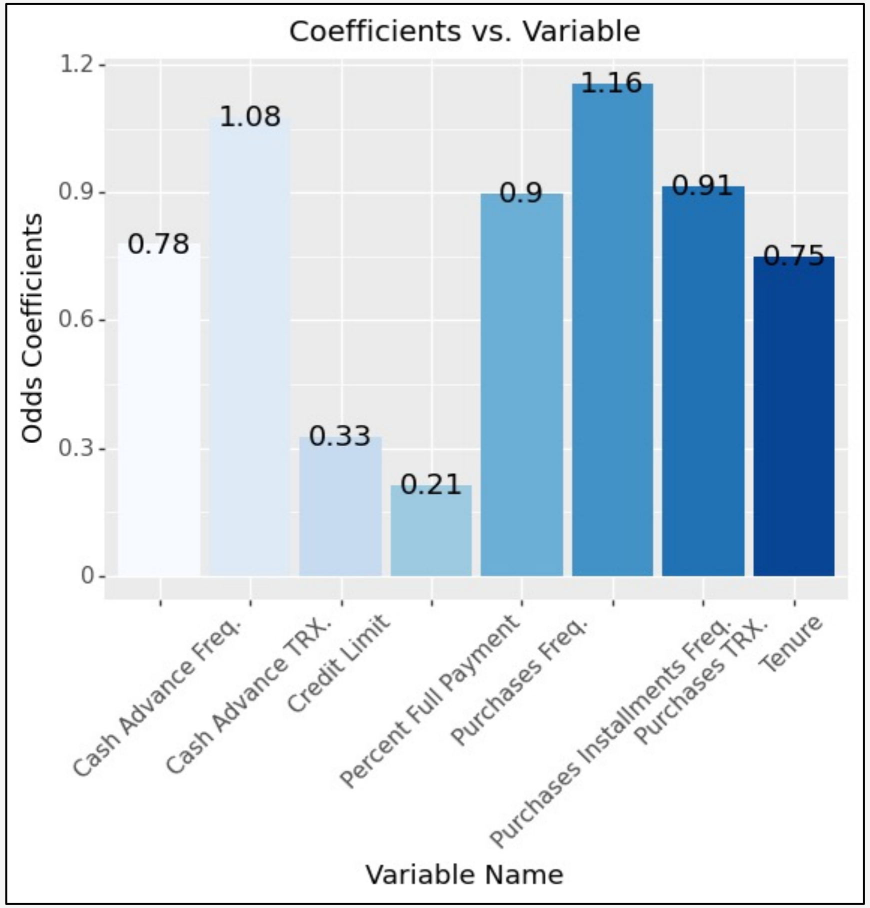
Customer 2 Credit Limit: $2,365.52



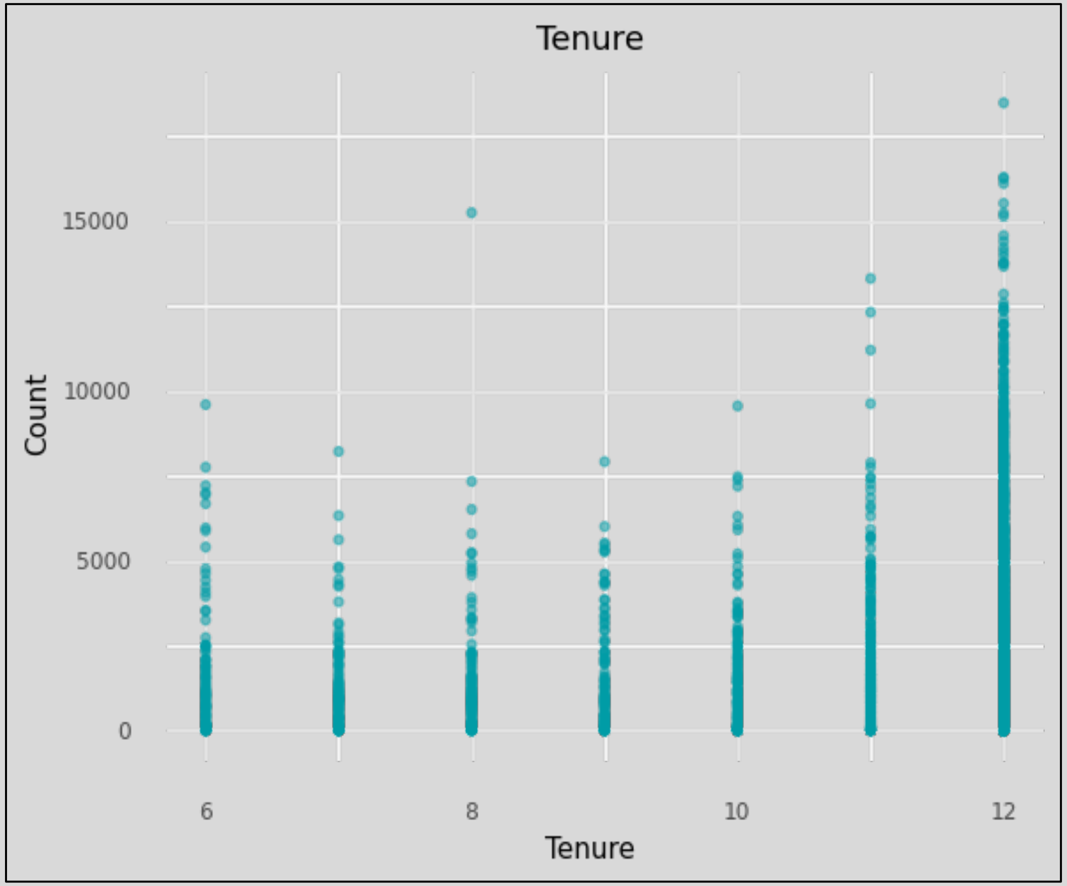
- Credit limits are predicting using our linear regression model
  - Cash advances and lots of purchases penalize credit limit
- Customer 1 clearly has the poorer spending habits on his/her card
  - Love to use cash advance, spend a lot, and have a high minimum payment
- Graph visualizes the cash advance habits of customers across various credit limits

Are customers with longer tenure more likely to meet minimum payments?

- Logistic model to predict a Default
- Logistic regression results:
  - Accuracy score of 77.3%
  - Coefficients (in Odds)
    - Purchase Installments Frequency: 1.16
    - Cash Advance Transactions: 1.08
    - Tenure: 0.75

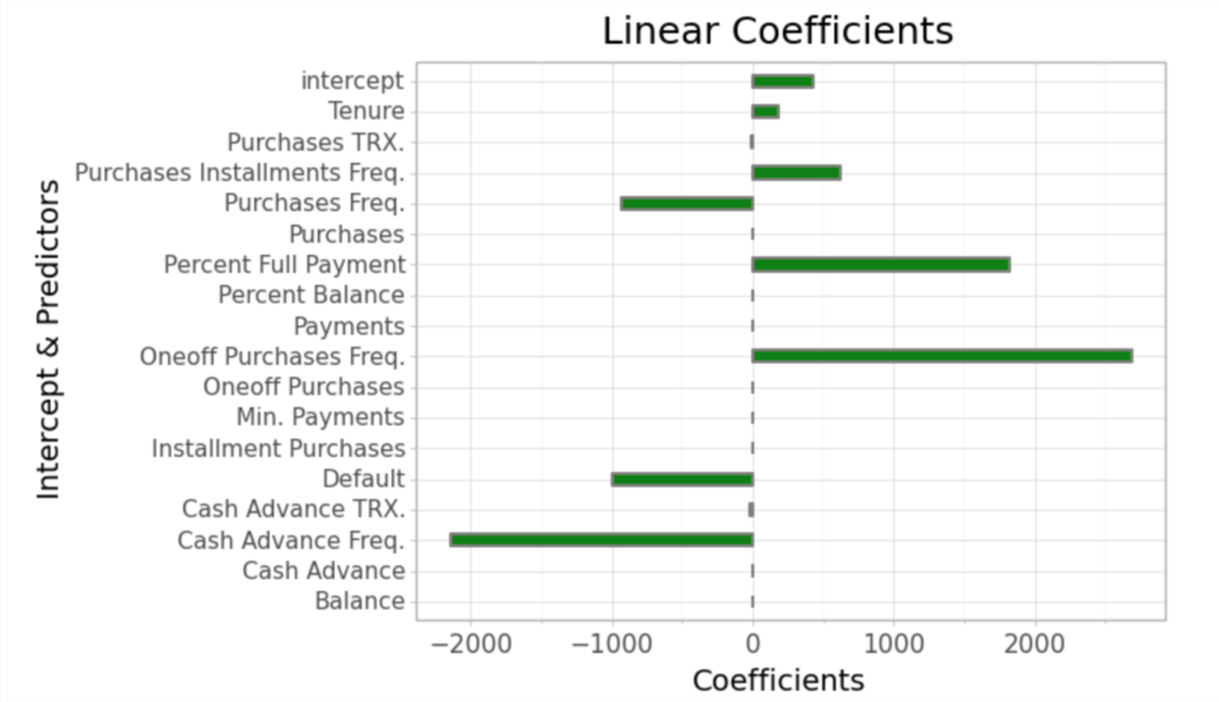


Tenure and Minimum Payments



- Longer tenure customers do not have an advantage
- Defaults are more likely to occur as time goes on
- Customer demographics could play a role
- Graph shows that most of data entries are long-term customers

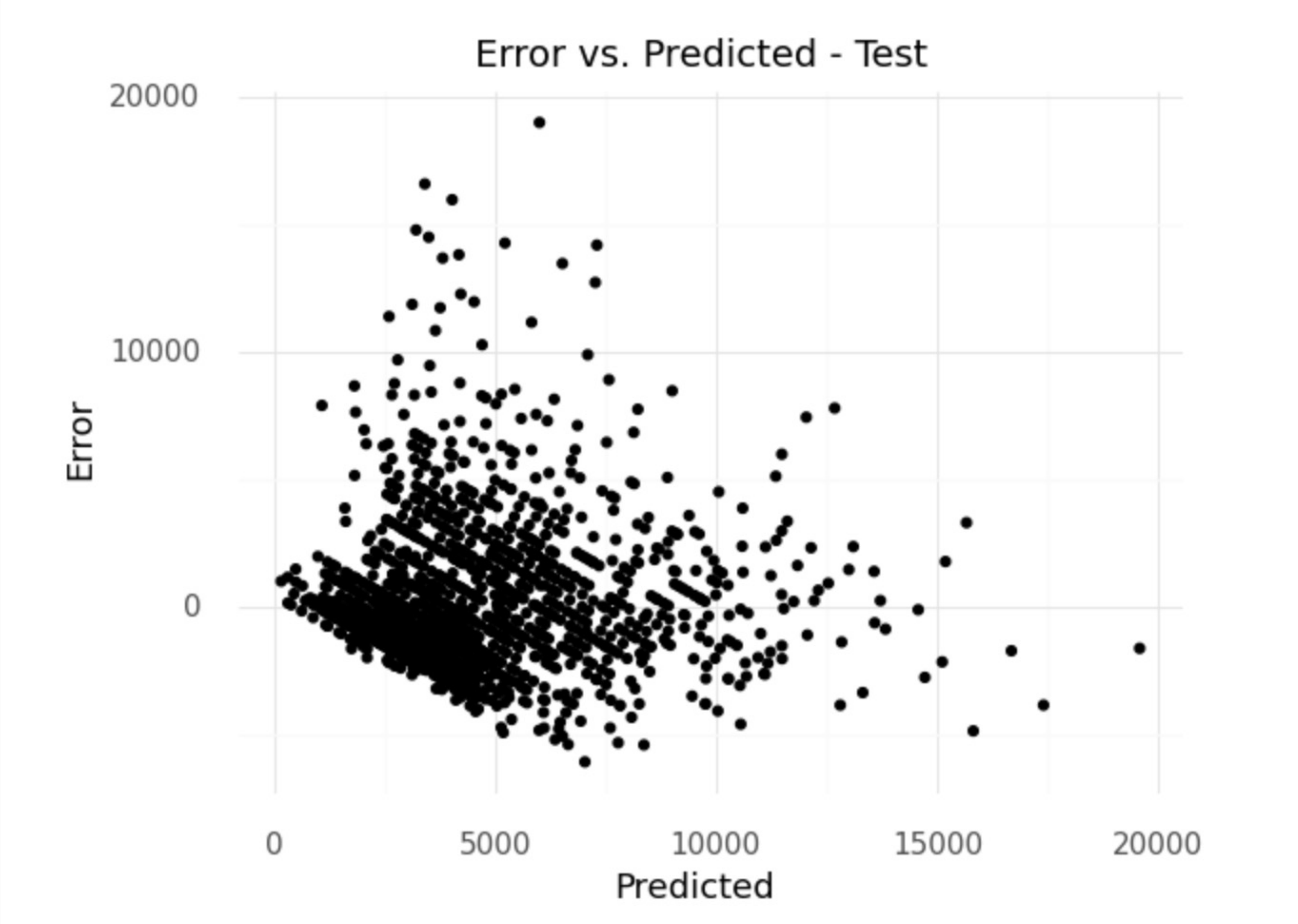
DO PEOPLE WITH A HIGHER TENURE TEND TO HAVE HIGHER CREDIT LIMITS?

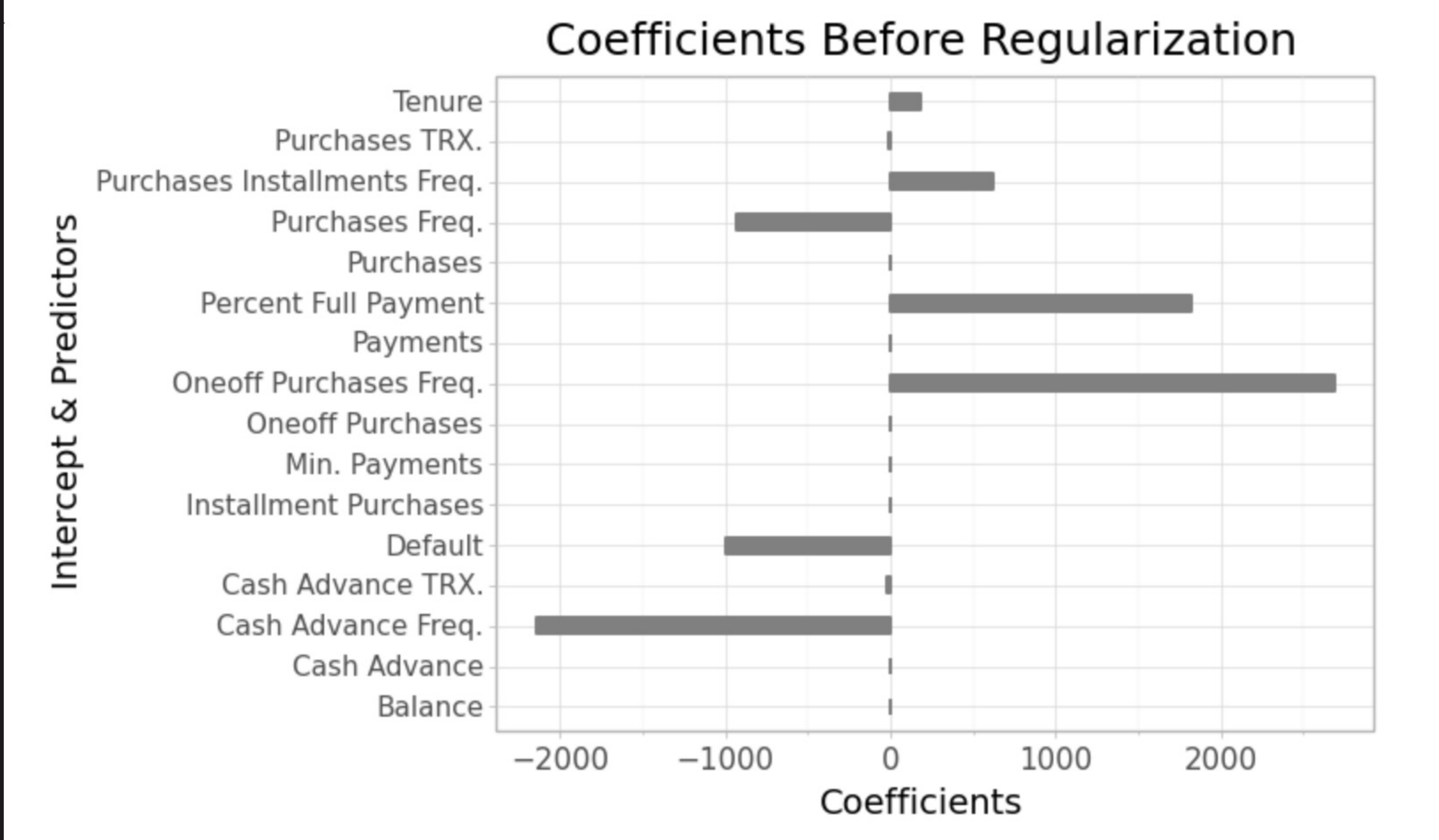


Linear Regression Results

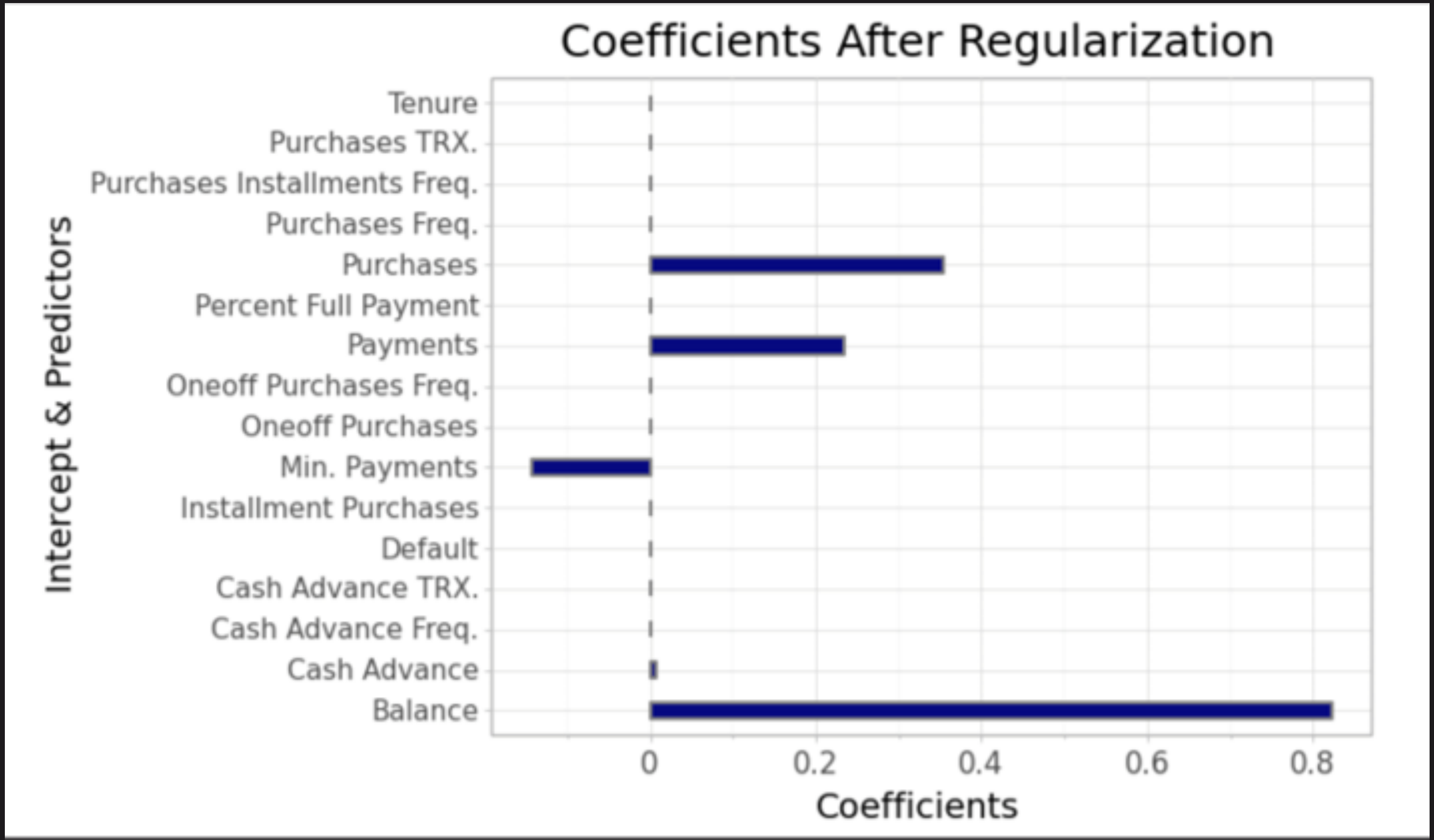
Outcome variable	Credit limit
Training R2	47%
Test R2	43%
Coefficient for Tenure	161.76

Image Source: https://www.ft.com/content/e25ce66e-b639-408b-b7ef-f524d664a30d


Error vs. Predicted - Test
20000
10000
0
Error
0
5000
10000
15000
20000
Predicted



Coefficients Before Regularization
Tenure
Purchases TRX.
Purchases Installments Freq.
Purchases Freq.
Purchases
Percent Full Payment
Payments
Oneoff Purchases Freq.
Oneoff Purchases
Min. Payments
Installment Purchases
Default
Cash Advance TRX.
Cash Advance Freq.
Cash Advance
Balance
Intercept & Predictors
−2000
−1000
0
1000
2000
Coefficients



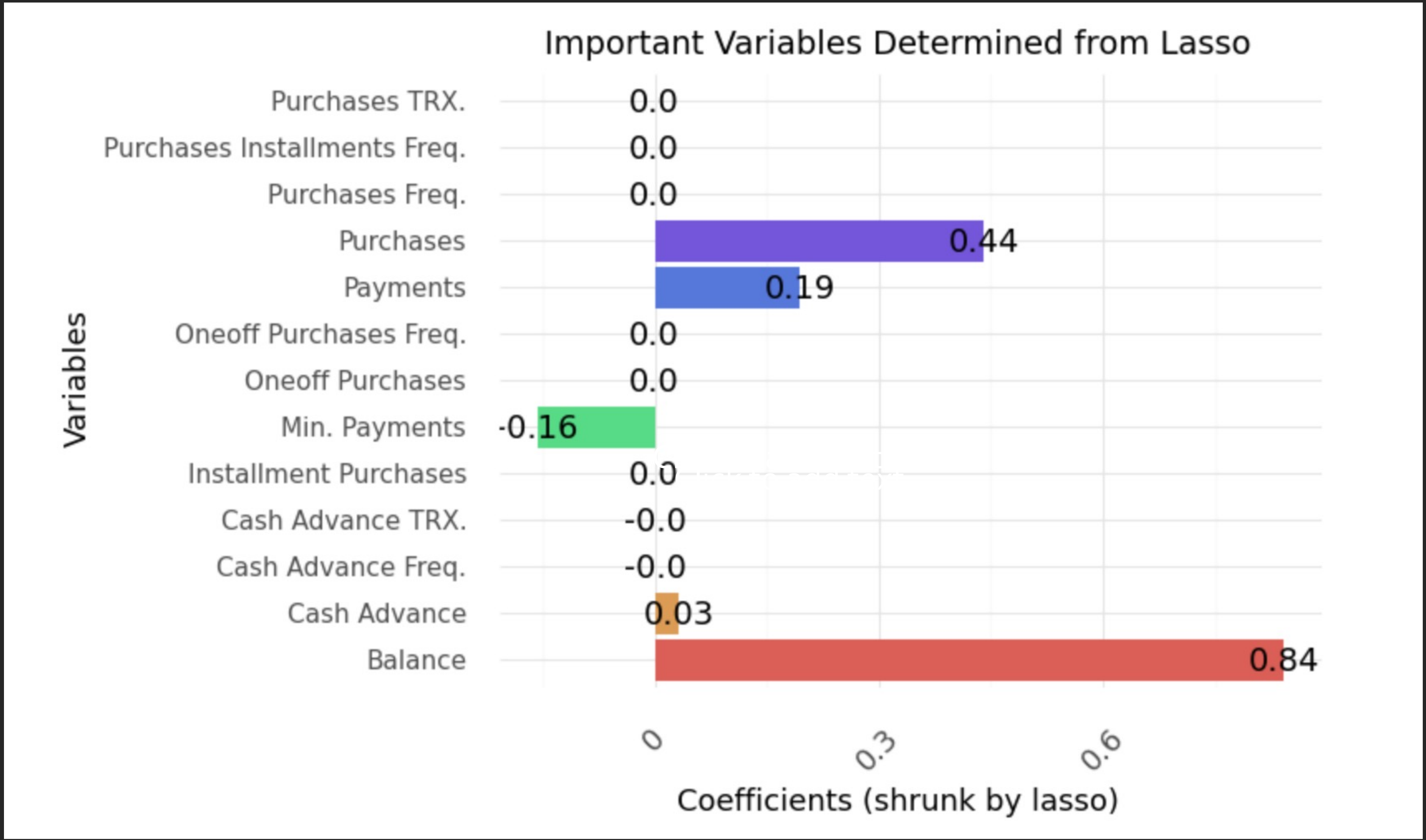
MODEL VALIDATION

Before Lasso :

```
from sklearn.model_selection import train_test_split # simple TT split validation
```

After Lasso:

```
from sklearn.linear_model import RidgeCV, LassoCV  # cross-validation
```



VARIABLES OF INTEREST:

- Balance
- Purchases
- Payments
- Min. Payments
- Cash Advance