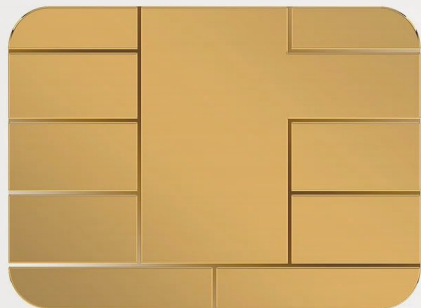


 synchrony Datathon



STAT 510'S Best Students

James/Mehmet/Nay/Yanlin

1. PREDICTION PROBLEM

Predict the customer spending for the fourth quarter of the current year. Specifically can we predict the customer spending for Q4 2025?



i. Data Exploration and Learning

ii. Solution Building and Testing

iii. Deliverable Models and Results

iv. Project Reflection

2. CLASSIFICATION PROBLEM

Classify accounts in to segments that can help identify potential accounts that would need a credit line increase because of the predicted spending.



i. Data Exploration and Learning

Predict the customer spending for the fourth quarter of the current year. Specifically can we predict the customer spending for Q4 2025?



**Transaction Data &
World Transaction Data**

Transaction	
Data/transaction_fact_20250325.csv	current_account_nbr
Counts - 493337	transaction_type
	transaction_code
	transaction_date
	transaction_amt
	payment_type
	transaction_return_cnt
	transaction_sale_cnt
	product_amt
	product_qty
	invoice_nbr
	first_purchase_ind

Important Features

We chose to aggregate these transactions by accumulating the total of type SALE for each customer by month. This makes sense given client of Synchrony credit cards with monthly billing cycles.

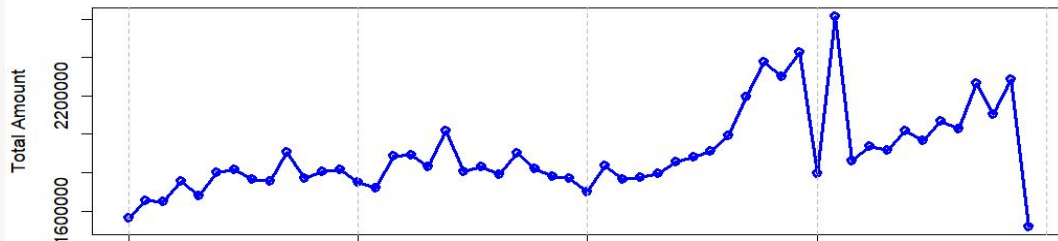
i. Data Exploration and Learning

The total amount spend with all customers aggregated and the number of transactions **increases dramatically during Q4's holiday season.**

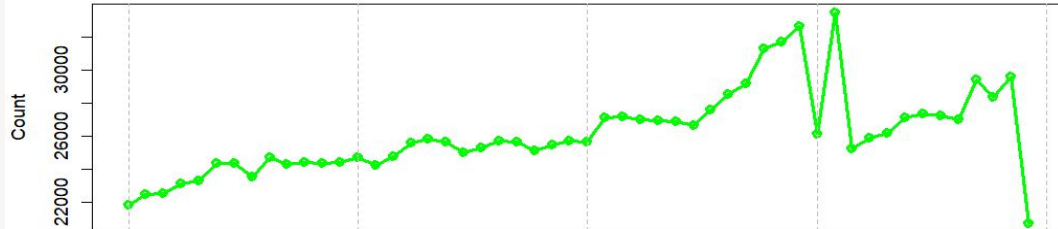
Otherwise, visually there is a slight increasing trend throughout the year in Q2 and Q3. The dip at the beginning and end is due to data cutoff.

The mean amount spent per credit card charge varies throughout the year, dipping around Q3.

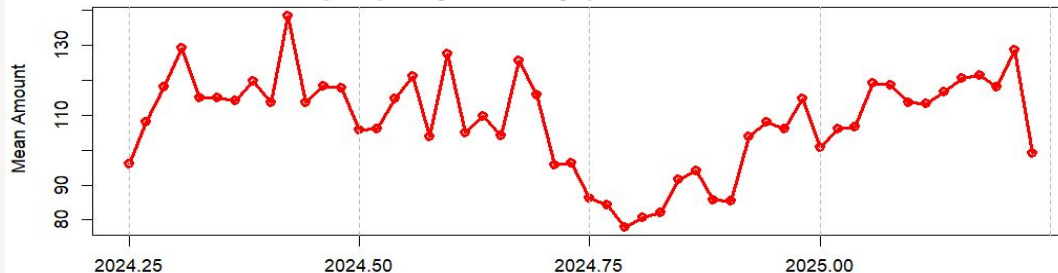
TOTAL TRANSACTION AMOUNT



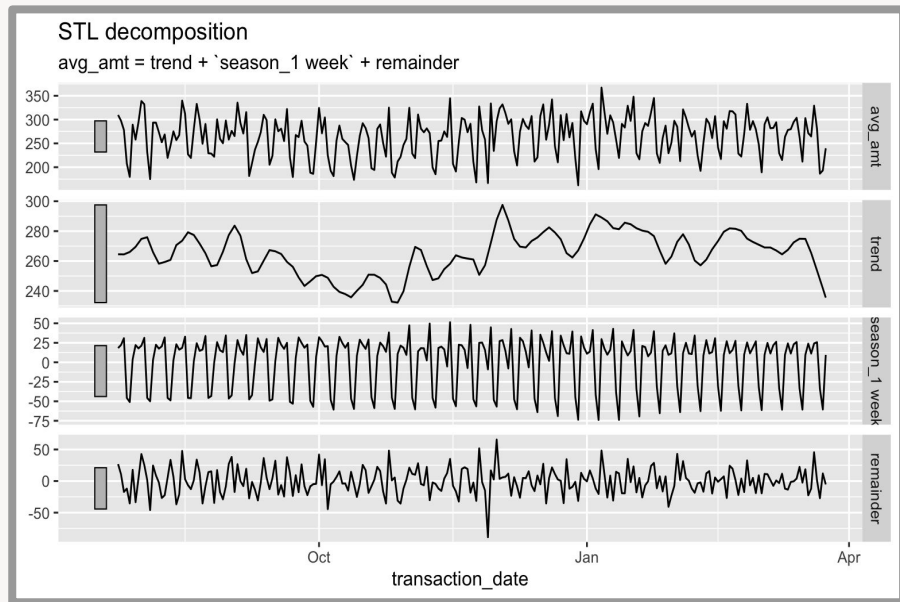
TOTAL TRANSACTION COUNT



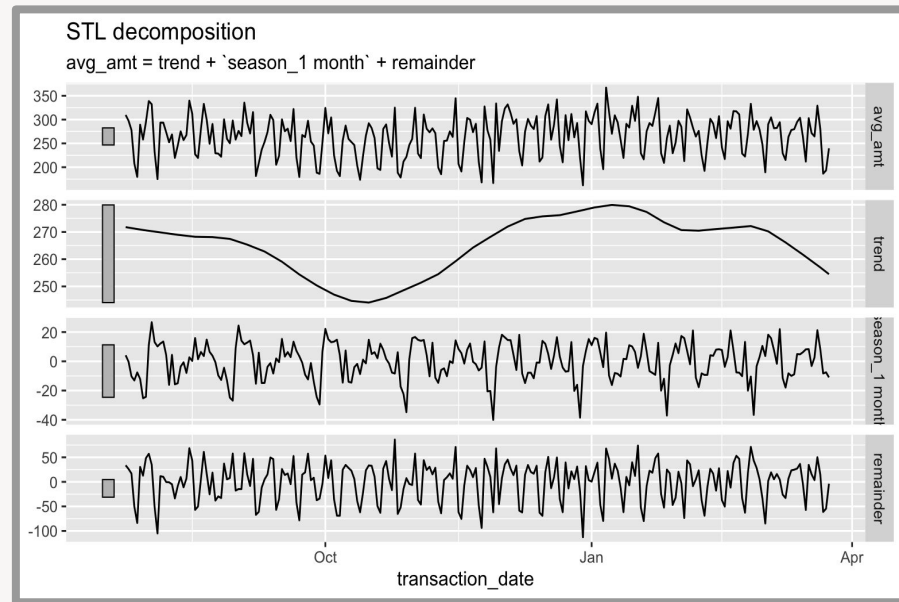
MEAN TRANSACTION AMOUNT



i. Data Exploration and Learning



Weekly Seasonality

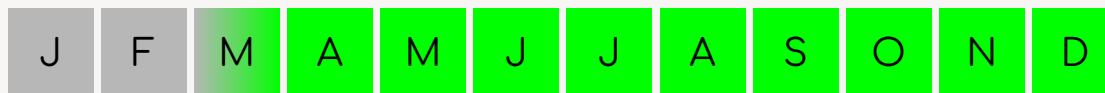


Monthly Seasonality

ii. Solution Building

ii.iv Data Limitations

2024



2025

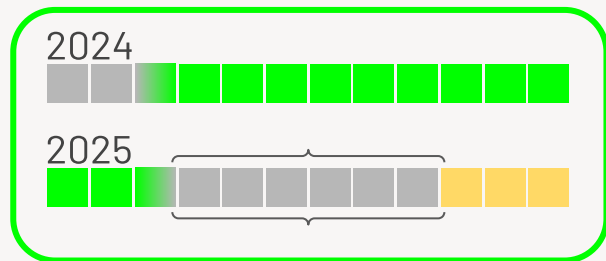


ISSUE#1. Only one year of usable transaction data

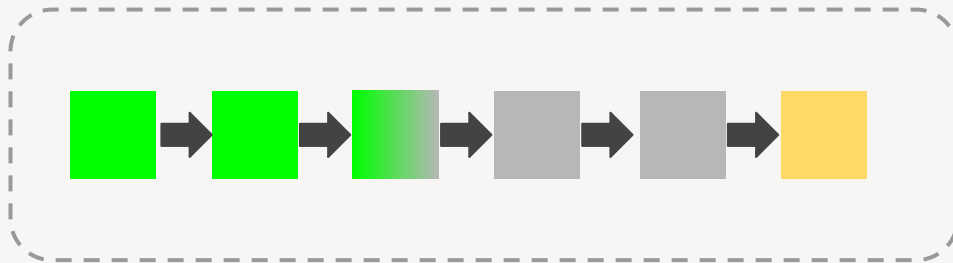
ISSUE#2. Large gap between data and 2025 Q4



ii. Solution Building



Heuristic #1: Build a model to predict month by month



Advantages: More Data for Monthly Variation

Disadvantages: Multiple Prediction Error

Blocks Use of External Macroeconomic Factors

Heuristic #2: Build a model analyzing overall trends to extrapolate



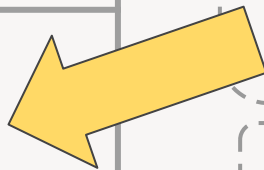
Advantages: Singular Prediction Error

Disadvantages: Lack of Prior Years



ii. Solution Building

LMM	Models fixed and random effects (-) Too Simple for TS
SARIMA	Seasonal time series model (-) Works best for Univariate Data
LSTM	Long term trends and patterns (-) Not enough yearly data to find notable trends
XGBOOST	Complex relationships

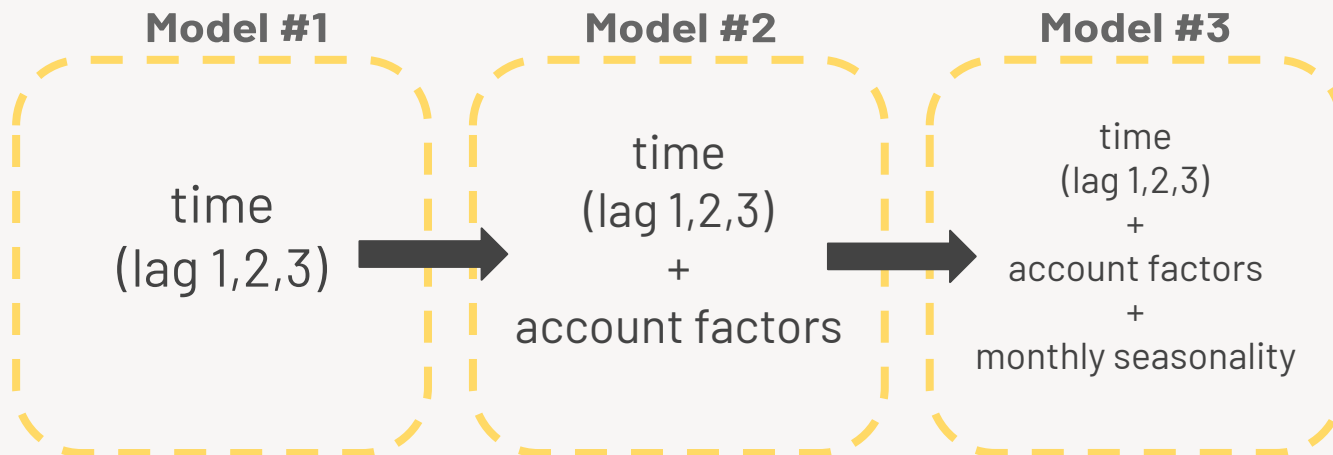


OUR SOLUTION

Follow Heuristic
#1 to learn
long-term trends
Add seasonality
and customer
based factors



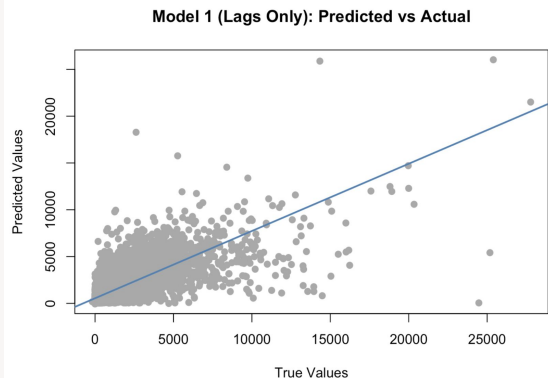
iii. Deliverable Models and Results



Account Factors:

IsHighSpender, IsDualCardHolder, IsFraudRisk, psc_ind,
ca_cash_bal_pct_crd_line, rb_new_bhv_scr cu_crd_line, cu_cur_balance,
cu_otb, ca_avg_utilz_lst_6_mnth, ca_mnth_snce_active,
ca_mnth_snce_cl_chng, ca_mob, ca_nsf_count_lst_12_mnth,
return_check_cnt_ytd, return_check_cnt_total, prev_balance

iii. Deliverable Models and Results



Call:

```
lm(formula = predictions ~ testyM)
```

Residuals:

Min	1Q	Median	3Q	Max
-18089.6	-475.3	-320.6	357.6	15854.2

Coefficients:

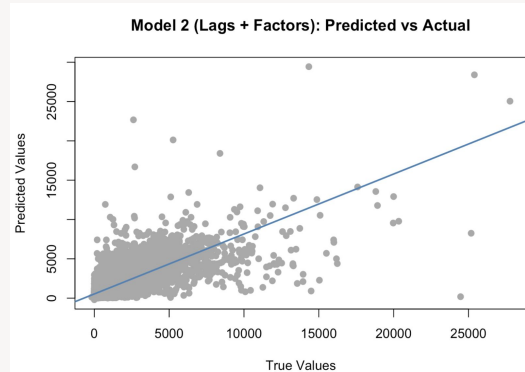
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.471e+02	1.210e+01	45.2	<2e-16 ***
testyM	7.189e-01	5.605e-03	128.3	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1082 on 11122 degrees of freedom

Multiple R-squared: 0.5966, Adjusted R-squared: 0.5966

F-statistic: 1.645e+04 on 1 and 11122 DF, p-value: < 2.2e-16



Call:

```
lm(formula = predictions ~ testFyM)
```

Residuals:

Min	1Q	Median	3Q	Max
-19020.3	-476.5	-320.4	308.8	20149.3

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.322e+02	1.283e+01	41.47	<2e-16 ***
testFyM	7.630e-01	5.943e-03	128.38	<2e-16 ***

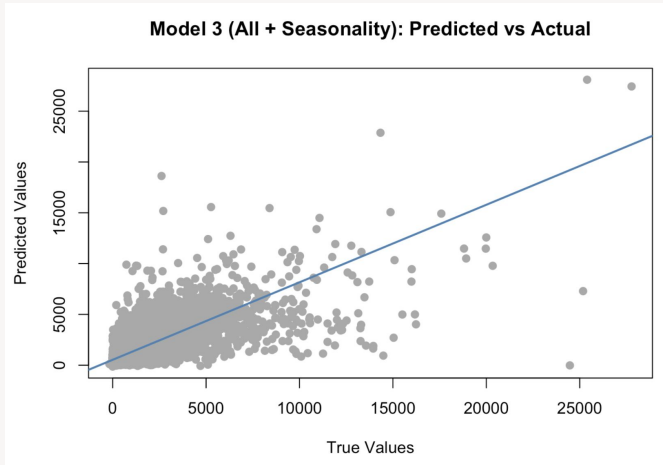
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1148 on 11122 degrees of freedom

Multiple R-squared: 0.5971, Adjusted R-squared: 0.597

F-statistic: 1.648e+04 on 1 and 11122 DF, p-value: < 2.2e-16

iii. Deliverable Models and Results



Call:

```
lm(formula = predictions ~ testFyM)
```

Residuals:

Min	1Q	Median	3Q	Max
-18138.2	-433.7	-282.3	252.0	16277.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.595e+02	1.224e+01	37.55	<2e-16 ***
testFyM	7.217e-01	5.667e-03	127.34	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1094 on 11122 degrees of freedom

Multiple R-squared: 0.5931, Adjusted R-squared: 0.5931

F-statistic: 1.621e+04 on 1 and 11122 DF, p-value: < 2.2e-16

*checks still needed to be performed on this model



i. Data Exploration and Learning

Classify accounts in to segments that can help identify potential accounts that would need a credit line increase because of the predicted spending.

1. Credit Risk Evaluation

- Account Metrics/ Behavior Scores
- Utilization Trends
- Delinquency & NSF

2. Fraud Detection

- Transaction Flags/Account Flags
- Behavioral Anomalies

3. Customer Segmentation

- Card Type
- Demographics

account_dim_20250325.csv	NEWemp_code
	IsDual
	IsFRisk
statement_fact_20250325.csv	pssc_ind
	return_check_cnt_ytd
	return_check_cnt_total
rams_batch_cur_20250325.csv	prev_balance
	ca_cash_bal_pct_crd_line
	rb_new_bhv_scr
	cu_bhv_scr
	cu_nbr_of_plastics
	cu_crd_bureau_scr
	cu_crd_line
	cu_cur_balance
	cu_otb
	ca_avg_utilz_lst_6_mnth
	ca_nsf_count_lst_12_months
	ca_mnth_since_active
	ca_mnth_since_cl_chng
	ca_mob

i. Data Exploration and Learning

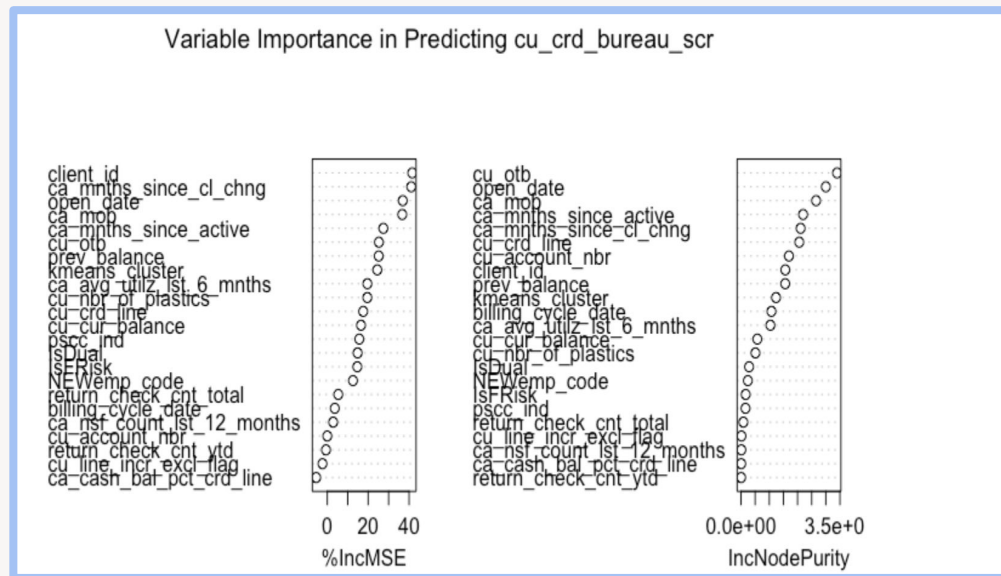
Extract Important Factors to Classify Customers

Numerous classifications possible!

Goal: achieve classification based on comprehensive factor reflecting much info on the customer

Our choice: external bureau score

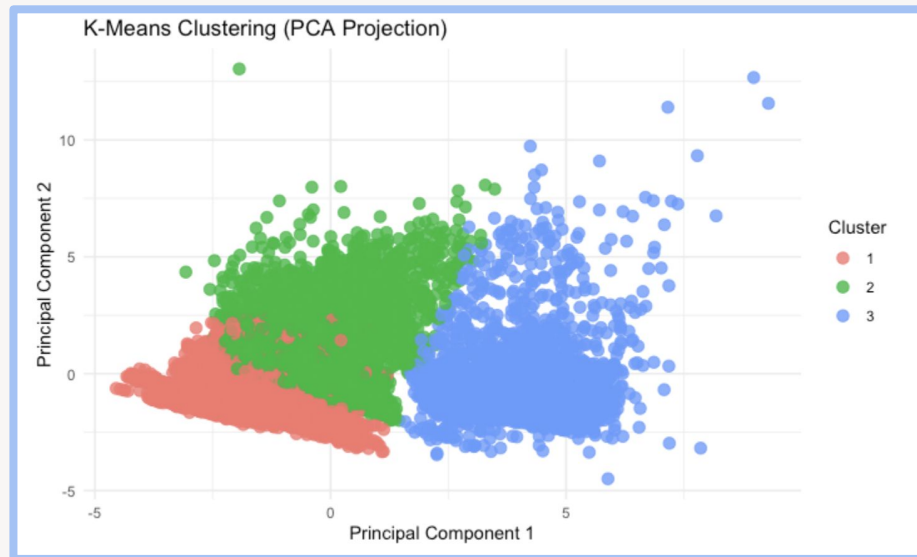
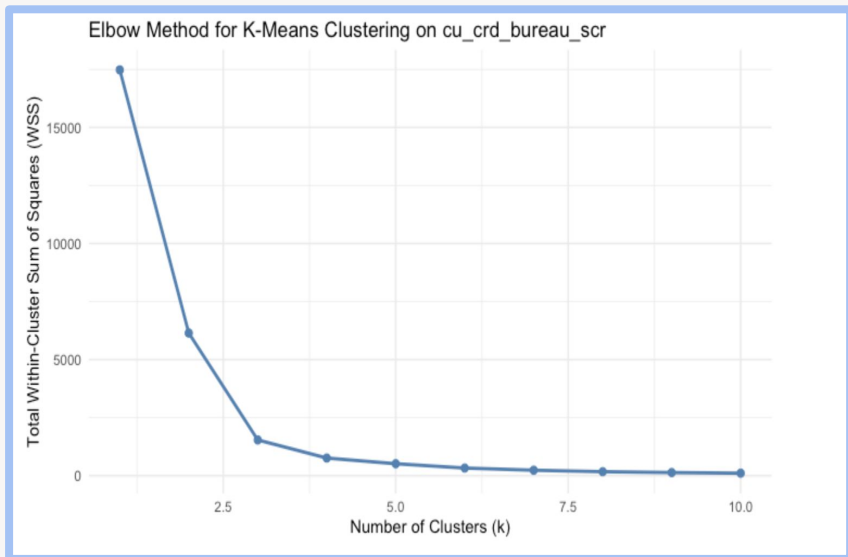
Creating variable importance rankings extracted from regression trees and random forest



ii. Solution Building

K-MEANS CLUSTERING

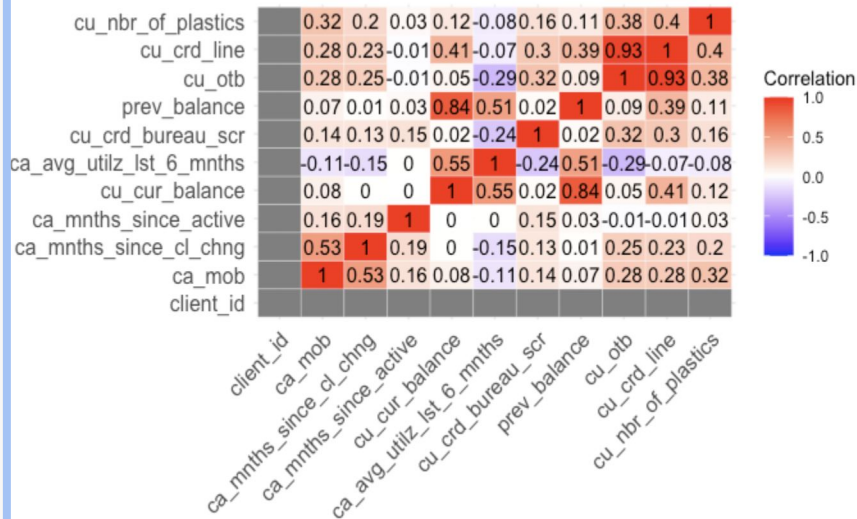
Finding the “right” number of clusters:



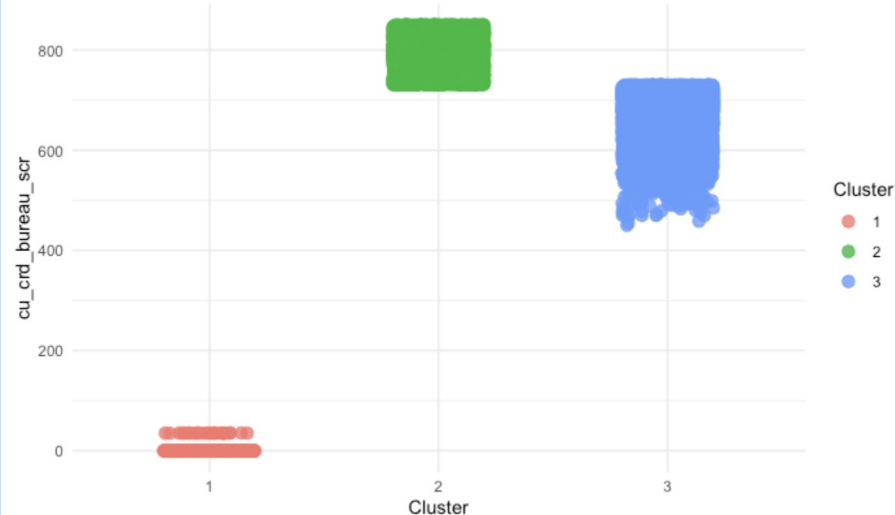
iii. Deliverable Models and Results

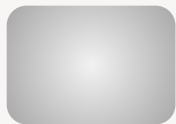
What are characteristics of each customer cluster?

Corr. Heatmap of Imp. Predictors for Ext. Bureau Score



K-Means Clustering on cu_crd_bureau_scr (3 Clusters)





iv. Project Reflection

BASED ON PREDICTION



1. Capitalize on Q4 Spending Surge
 - Increase Credit Limits Temporarily (Oct-Dec)
 - Launch Promotional Financing
2. Boost Q2-Q3 Engagement
 - Summer/Back-to-School Campaigns
 - Pre-Holiday Credit Health Checks

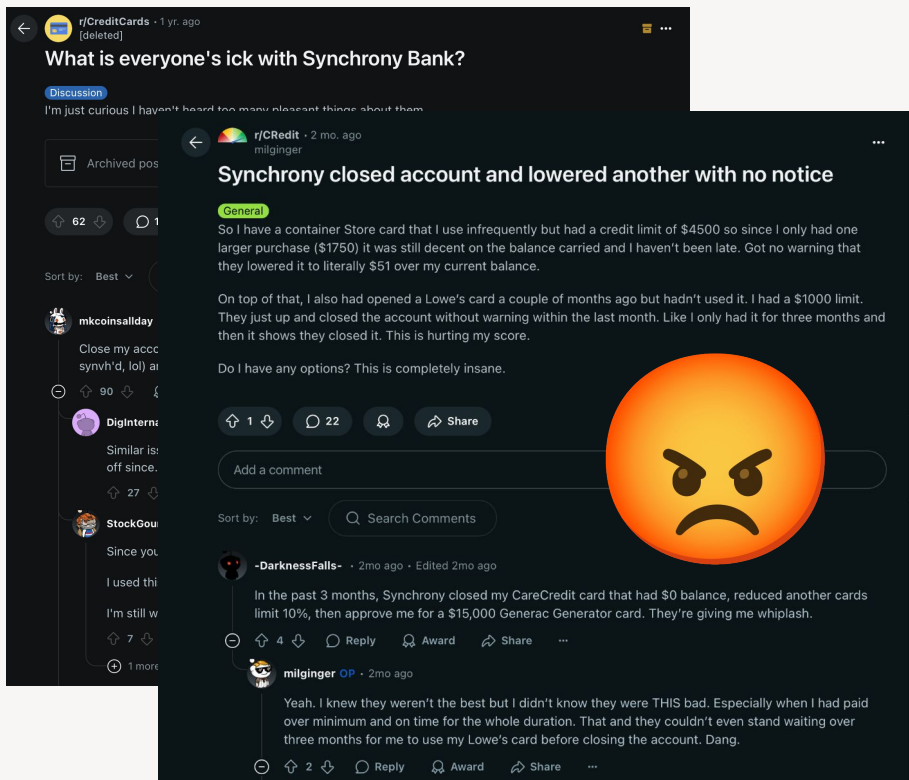
BASED ON CLASSIFICATION



1. Laser-Focus on High-Value Customer Segments
 - Bullseye Targets: Dual-Card Holders, employees
2. Weaponize Data for Smarter Limits
 - Greenlight These CLI Candidates: Open date, clients' credit

iv. Project Reflection

Other issues may exist...



Recommendations:

For Synchrony Company:

1. Enhance Transparency in Credit Decisions

- Clear Communication
- Provide a Reconsideration Process

2. Improve Risk Assessment Models

- Avoid Overly Aggressive "Balance Chasing"
- Implement gradual adjustment

For the Clients:

1. Immediate Action:

- Politely request a reconsideration

2. Document Everything:

- Ask for a written explanation (required by law).

Thank You!

03 30 2025 0700

STAT 510'S BEST STUDENTS

