BDM 2053 Project

Model for Predicting Credit Card Customer Attrition

Presented by: Group 1

Jefford Secondes

Jovi Fez Bartolata

Maricris Resma

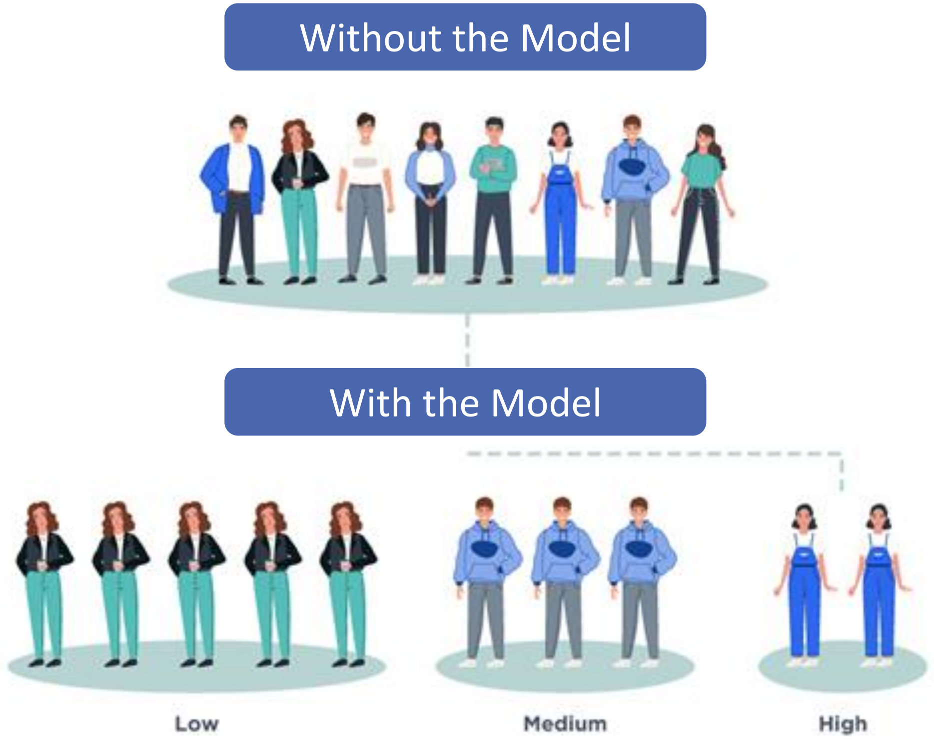
Luz Zapanta

Agenda

- Objectives
- Methodology
- Data Pre-Processing and Exploratory Data Analysis
- Modeling Techniques and Results

Objectives

- Identify early indicators of credit card attrition based on customer profile and spend behavior
- Build a predictive model to identify and segment customers based on attrition risk

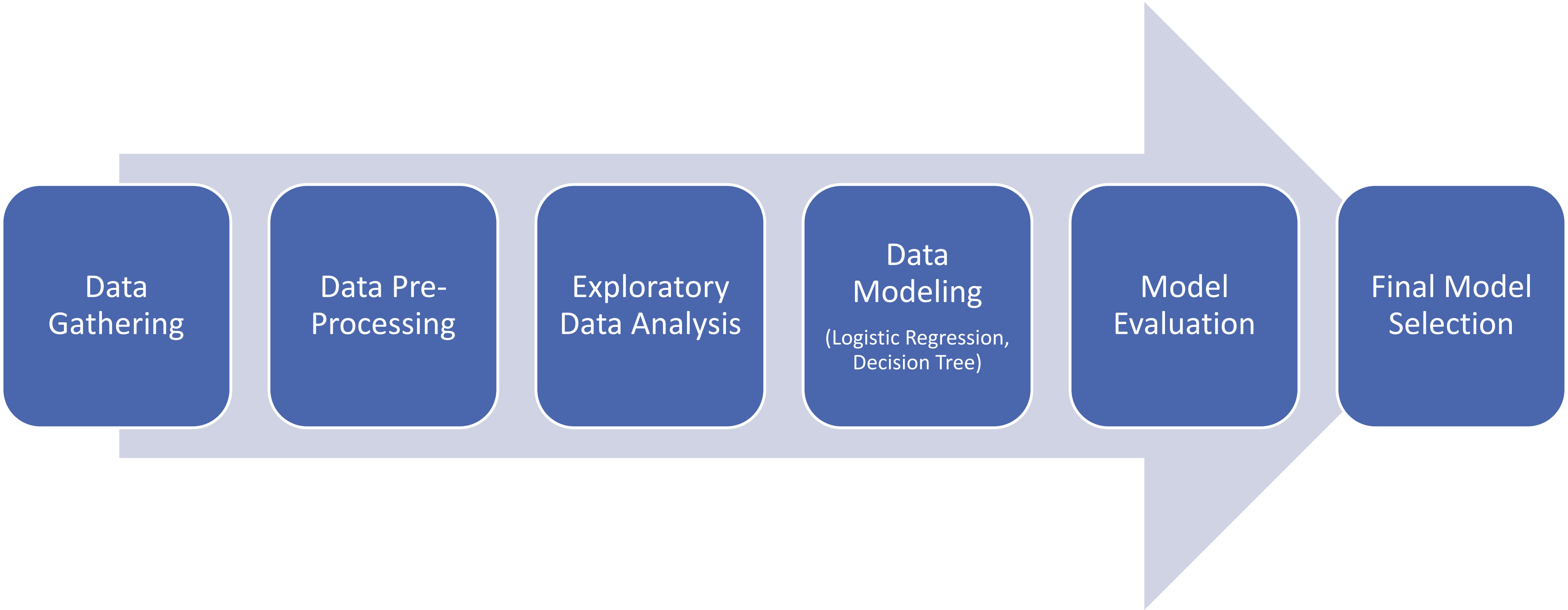


Source: https://jelvix.com/wp-content/uploads/2022/02/what-is-a-propensity-model.png

Methodology

Methodology

- Phases



Data Gathering

- Data
  - The raw data has 10,127 rows and 23 columns
  - Source: https://zenodo.org/record/4322342#.ZCM4hXbMI2x
- Dependent Variable: Attrition Flag
- Independent Variables (20)

Demographic Profile

- Age
- Gender
- Number of dependents
- Education level
- Marital status
- Income category

Customer Relationship

- Months on books
- Number of relationships with the card issuer
- Number of inactive months
- Number of contact numbers
- Type of card
- Credit limit

Spend Behavior

- Revolving balance
- Average open to buy ratio
- Transaction amount (total and Q4 to Q1 change)
- Transaction count (total and Q4 to Q1 change)
- Average utilization rate

Data Pre-Processing and Exploratory Data Analysis

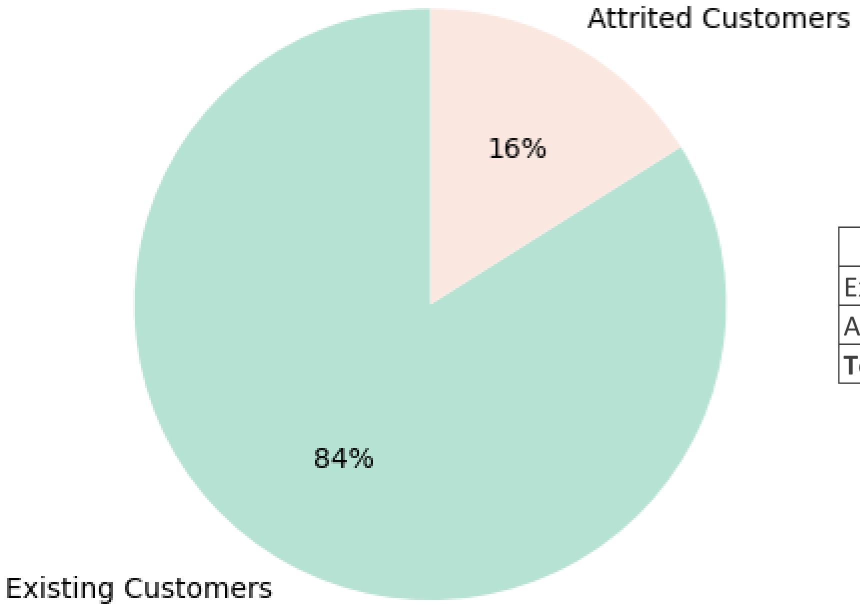
Data Pre-Processing

- Data quality checks
  - No feature has null or nan values
  - No duplicate records
  - Education_Level, Marital_Status and Income_Category have 'Unknown' data as value
- Dropped unnecessary features
- Transformed Target Variable (Attrition_Flag) to numerical value (1/0)

	column_name	data_type	count_unique_values	count_unknown	count_null	count_nan
Client_Num	Client_Num	int64	10127	0	0	0
Attrition_Flag	Attrition_Flag	object	2	0	0	0
Customer_Age	Customer_Age	int64	45	0	0	0
Gender	Gender	object	2	0	0	0
Dependent_Count	Dependent_Count	int64	6	0	0	0
Education_Level	Education_Level	object	7	1519	0	0
Marital_Status	Marital_Status	object	4	749	0	0
Income_Category	Income_Category	object	6	1112	0	0
Card_Category	Card_Category	object	4	0	0	0
Months_on_Book	Months_on_Book	int64	44	0	0	0
Total_Relationship_Count	Total_Relationship_Count	int64	6	0	0	0
Months_Inactive_12_mon	Months_Inactive_12_mon	int64	7	0	0	0
Contacts_Count_12_mon	Contacts_Count_12_mon	int64	7	0	0	0
Credit_Limit	Credit_Limit	float64	6205	0	0	0
Total_Revolving_Bal	Total_Revolving_Bal	int64	1974	0	0	0
Avg_Open_To_Buy	Avg_Open_To_Buy	float64	6813	0	0	0
Total_Amt_Chng_Q4_Q1	Total_Amt_Chng_Q4_Q1	float64	1158	0	0	0
Total_Trans_Amt	Total_Trans_Amt	int64	5033	0	0	0
Total_Trans_Ct	Total_Trans_Ct	int64	126	0	0	0
Total_Ct_Chng_Q4_Q1	Total_Ct_Chng_Q4_Q1	float64	830	0	0	0
Avg_Utilization_Ratio	Avg_Utilization_Ratio	float64	964	0	0	0
Naive_Bayes_Classifier_1	Naive_Bayes_Classifier_1	float64	1704	0	0	0
Naive_Bayes_Classifier_2	Naive_Bayes_Classifier_2	float64	640	0	0	0

Exploratory Data Analysis

- Overall attrition rate is **16%**.
- Majority of the customers did not attrite (84%), therefore, we have an imbalanced class.

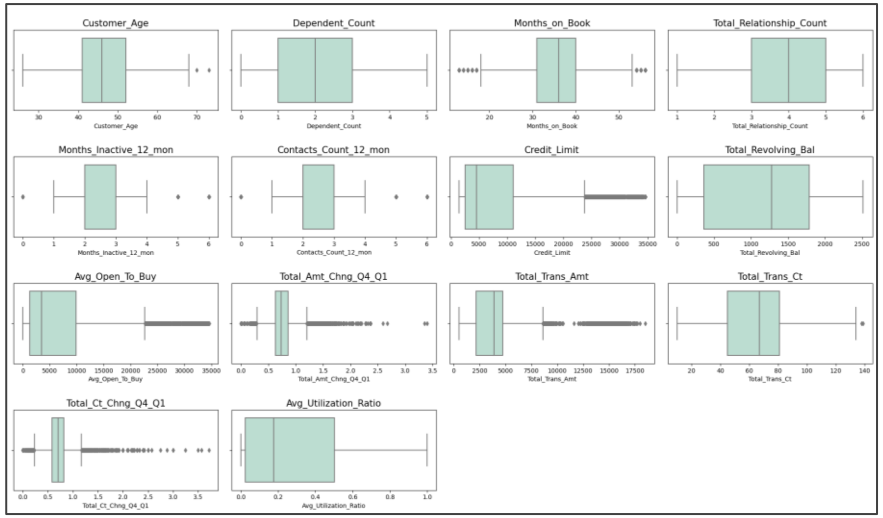


Attrition Flag	Count	%
Existing Customers: 0	8,500	84%
Attrited Customers: 1	1,627	16%
Total	**10,127**	**100%**

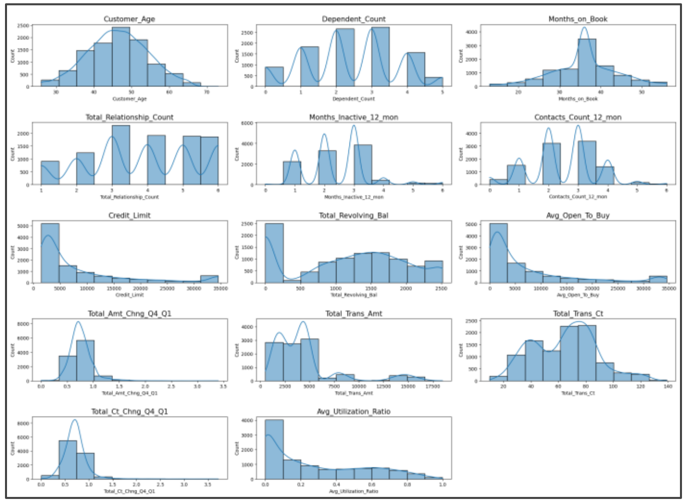
Exploratory Data Analysis

- **Gender:** Attrition is more prevalent in female customers vs male customers.
- **Education Level:** There are more Doctorate and Post-Grads customers who attrited.
- **Marital Status:** Those with relationship status = "Single" or "Unknown" recorded more attrition than those who are married or divorced.
- **Income Category:** Customers in the 60K-80K income bracket have the lowest attrition rate.
- **Card Category:** Those with premium and gold cards have more attrition than blue and silver cardholders
- Dummy variables were created based on the results of EDA

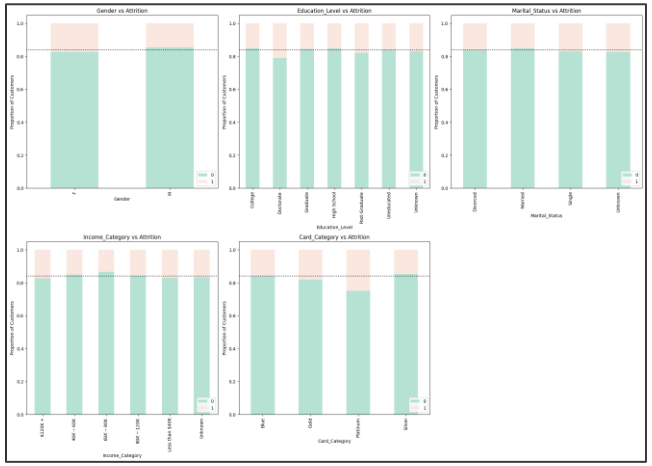
Boxplot for Numerical Features



Histogram for Numerical Features

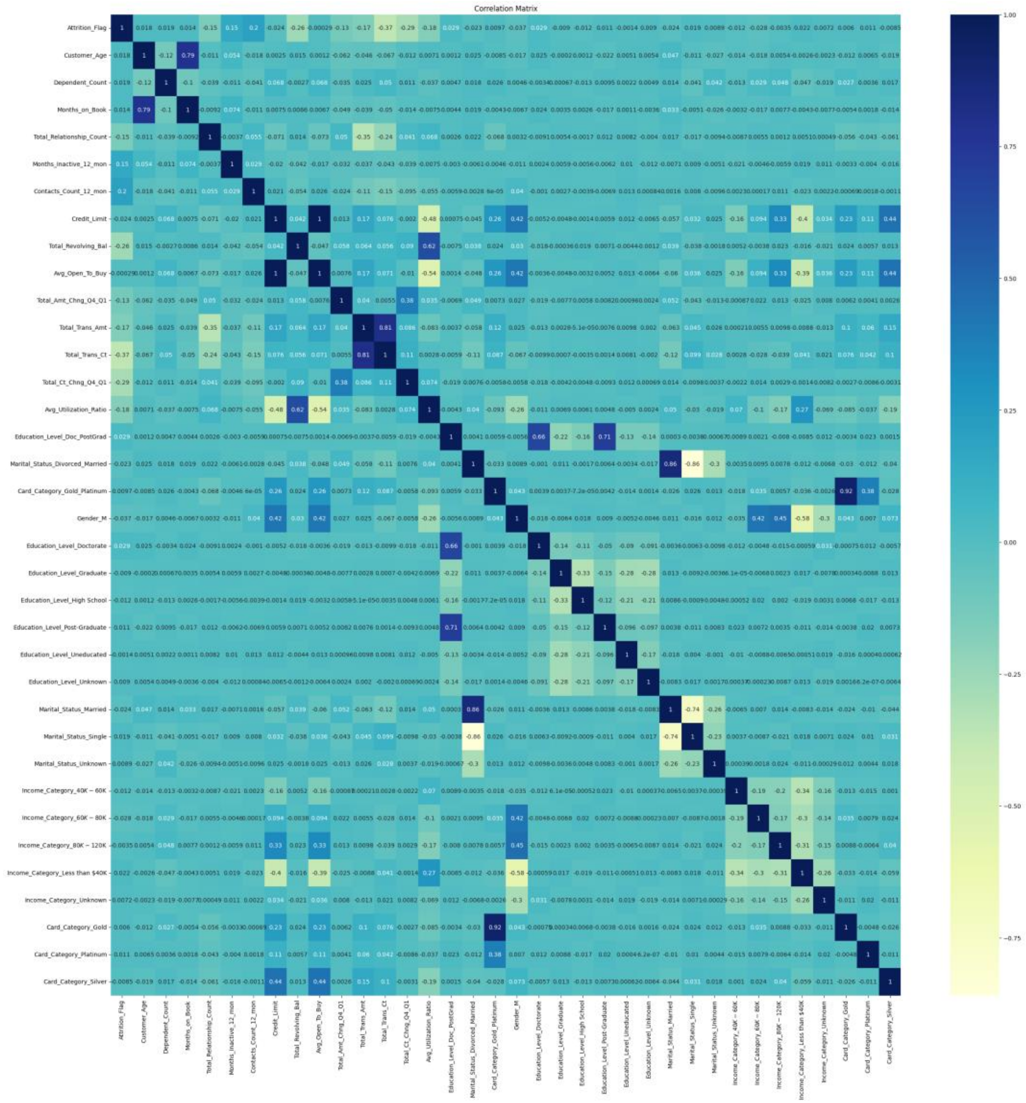


Stacked Column Chart for Categorical Features



Exploratory Data Analysis

- **Correlation Heat Map:** Some of the variables with strong correlation with Attrition are:
    - Total_Trans_Ct: total transaction count
    - Total_Ct_Chng_Q4_Q1: change in total transaction count from Q4 to Q1
    - Total_Revolving_Bal: total revolving balance



Modeling Techniques and Results

Modeling Techniques

Data Split

- Train (80%): 8,101
- Test (20%): 2,026

Modeling

- Logistic Regression
- Decision Tree

Model Evaluation

- Accuracy Score
- Precision Score
- Recall Score

Final Model Selection

Modeling Techniques

Logistic Regression

1. Standardize/Scale numeric features
2. Oversampling using SMOTE (Synthetic Minority Oversampling Technique) algorithm
3. Fit Logistic Regression
   - Recursive Feature Elimination (30 features)
   - Manual feature selection
4. Model Assumption Checks
   - p-value of predictors
   - Signs of coefficients
   - Test for multicollinearity using Variance Inflation Factor (VIF)
5. Model Evaluation

Decision Tree

1. K-Fold cross validation to select optimal max_depth (result = 4)
   - Repeated stratified k-fold (10 folds, 3 repeats)
   - Oversampling using SMOTE (Synthetic Minority Oversampling Technique) algorithm
2. Fit Decision Tree
   - max depth = 4
   - min leaf size = 500
3. Model Evaluation

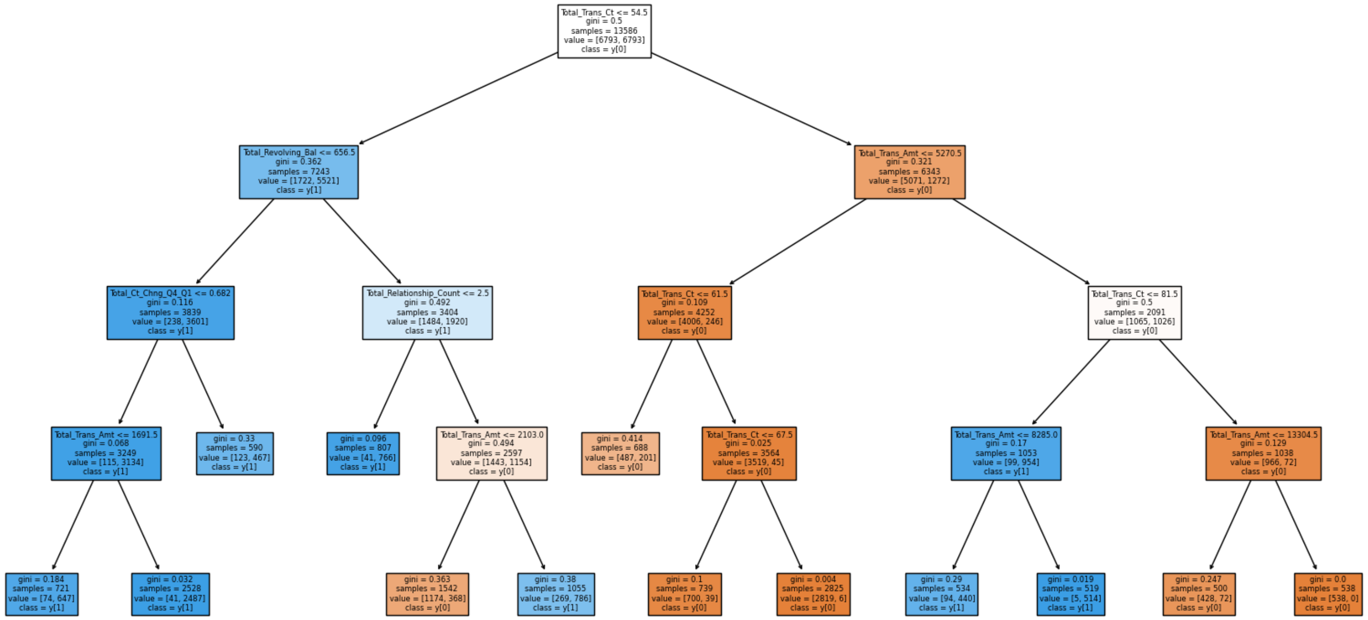
Logistic Regression Results

- Significant Features
  - ⬇ Gender (if customer is Male)
  - ⬆ Number of Dependents
  - ⬇ Total Relationship Count
  - ⬆ Number of Inactive Months
  - ⬆ Credit Limit
  - ⬇ Total Revolving Balance
  - ⬇ Total Transaction Count

Predictors	Coefficient	P-value	VIF Factor	P-value Check	VIF Check
const	-0.6303	0.0	2.141598	Passed	Passed
Gender_M	-1.0810	0.0	1.279530	Passed	Passed
Std_Dependent_Count	0.1329	0.0	1.006042	Passed	Passed
Std_Total_Relationship_Count	-0.7370	0.0	1.019864	Passed	Passed
Std_Months_Inactive_12_mon	0.5400	0.0	1.021152	Passed	Passed
Std_Credit_Limit	0.3237	0.0	1.302371	Passed	Passed
Std_Total_Revolving_Bal	-0.7454	0.0	1.035719	Passed	Passed
Std_Total_Trans_Ct	-1.8669	0.0	1.061784	Passed	Passed

Decision Tree Results

- Important Features (Highest to Lowest)
  - Total Transaction Count
  - Total Transaction Amount
  - Total Revolving Balance
  - Total Relationship Count
  - Change in Total Transaction Count from Q4 to Q1



Feature	Feature Importance
Total_Relationship_Count	0.066952
Total_Revolving_Bal	0.107659
Total_Trans_Amt	0.187765
Total_Trans_Ct	0.631242
Total_Ct_Chng_Q4_Q1	0.006382

Model Evaluation and Final Model Selection

- Final Model: **Decision Tree**
  - **Classification Accuracy:** 91% were predicted correctly
  - **Precision:** Out of all the customers that the model predicted would attrite, 65% actually did.
  - **Recall:** Out of all the customers that actually attrited, the model predicted this outcome correctly for 85% of those customers.

Set	Metric	Logistic Regression	Decision Tree
Train	Accuracy Score	82%	90%
	Precision Score	81%	90%
	Recall Score	83%	90%
Test	Accuracy Score	80%	91%
	Precision Score	43%	65%
	Recall Score	80%	85%

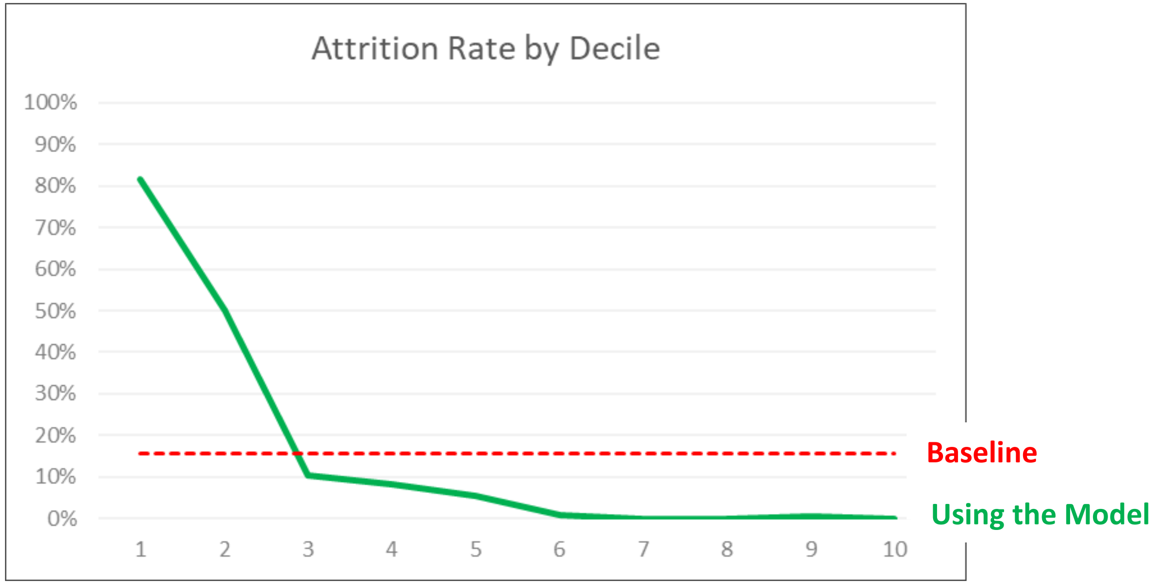
Model Use Case

- Targeting Customers for Anti-Attrition Campaign

  If we prioritize the Top 20% customers with highest probability of attrition, we have 66% chance of getting customers who will cancel their credit card account – **50 PPS higher** than our 16% baseline.

Test Set

Decile	No. of Attrited Customer	No. of Customers	% Attrited Customer	Cumulative %
1	165	202	82%	82%
2	102	203	50%	66%
3	21	202	10%	47%
4	17	203	8%	38%
5	11	203	5%	31%
6	2	202	1%	26%
7	0	203	0%	22%
8	0	202	0%	20%
9	1	203	0%	17%
10	0	203	0%	16%
Overall	**319**	**2,026**	**16%**	**16%**



Thank you.