Project I

By Anubhav Kumar Singh

Table of contents

01
Introduction
Problem Statement: Multiple Objectives

02
Data
Dataset explanation, Attributes, Independent and Dependent Variables

03
Methodology
How are things done? Explanation of approach and why so

04
Conclusion
Results with their meanings

Introduction

- The problem associated with this project is a real life problem
- Synthetic data from American Express

Problem Statement

Objective 1

Design a targeting campaign for credit card acquisition to target right set of prospects with right set of products with a budget of **$50,000***

Objective 2

You'll need to consider **customer preferences** and **cost and benefits** associated with acquiring a customer to create a winning strategy.

Objective 3

Calculation of **total profit** from all the acquired customers

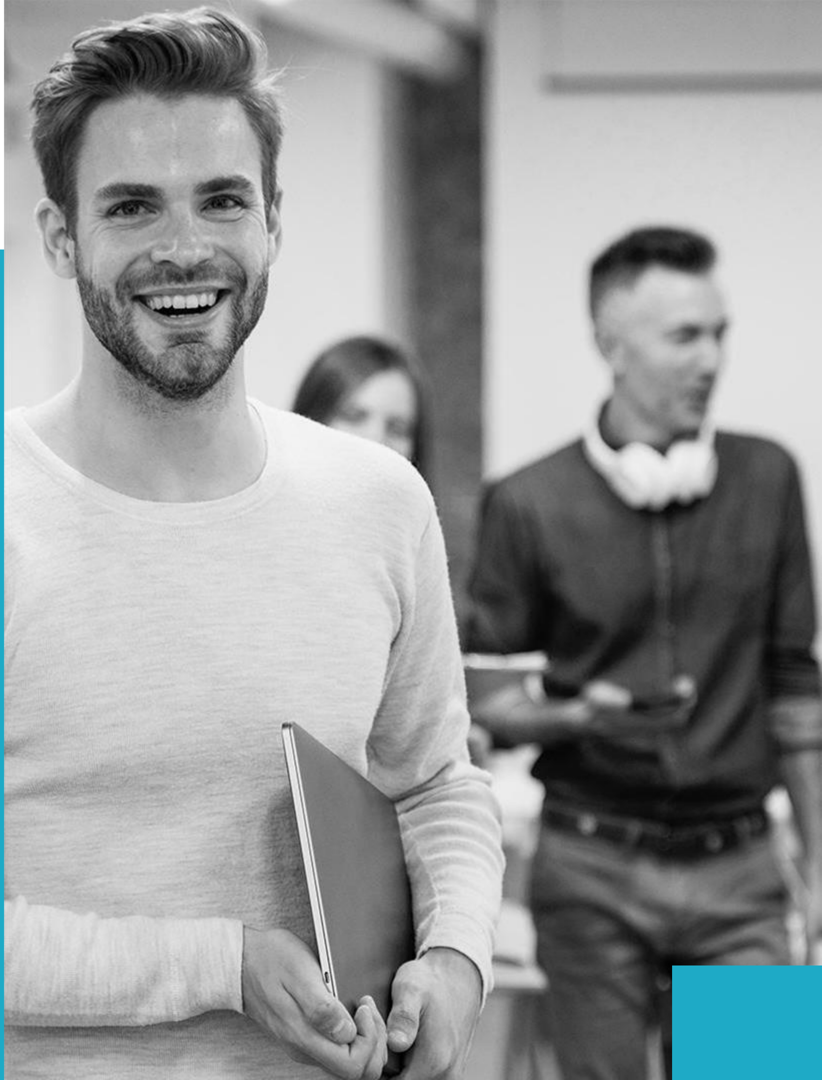
******There is penalty for underutilizing or overutilizing the budget, i.e., in case of underutilizing, there is penalty of 50% of whatever the budget is left, and in case of overutilizing there will be a penalty of the overutilized amount*

Data

In this project, there were three datasets

1. *Past Campaign Data*
2. *Performance of Customers booked from past campaign*
3. *Potential Customer Data*

Test Set – 1 and 2
Target/Validation Set – 3



Dataset Explanation

Dataset 1

List of prospects targeted in past campaign with their profiles. These prospects were randomly given products.

Prospects who liked their offer responded to the offer.

However, after underwriting, only a certain percentage were booked.

Dataset 2

List of customers acquired from past campaign with their profile at the time of acquisition, product they chose and their customer lifetime value

Dataset 3

List of prospects available for current campaign along with their profiles. This will be used as validation data to assess your solution.

Attributes with explanation

Variable	Description
ID	*Key*
long_term_credit_card_opening_propensity_score	*long term credit card opening propensity score*
creditworthiness_score	*Score on how creditworthy the customer is*
spend_potential	*Spend Potential of customer*
number_of_credit_cards	*Number of credit cards already owned by customer*
short_term_credit_card_opening_propensity_score	*Short term credit card opening propensity score*
long_term_credit_card_cancellation_propensity_score	*Long term credit card cancellation propensity score*
credit_card_utilization_score	*Aggregate utilization of credit limit across all credit cards*
number_of_attempts_to_open_credit_cards	*Number of attempts by customer to open credit card*
history_at_credit_bureau	*Number of years since first credit card or loan*
total_limit_across_credit_cards	*Average credit limit for on all credit cards*
contact_history_score	*No. of times targeted for a credit cards*
Balance Transfer Propensity Score	*Propensity of customer initiating a balance transfer*
Targeted_product	*Product for which the person was targeted*
Applied for the card	*Indicator if customer applied for the card*
Approved for the card	*Indicator if card application was approved*
Sampling Factor	*Sampling factor is used to calculate likelihood*
Customer Lifetime Value	*Present value of all cashflow expected from customer*

Acquisition Cost (in $)

Product	Targeting Cost per Contact –Incurred irrespective of response	Welcome Offer Cost per Booked Customer-Incurred only if customer is acquired
A	2	1500
B	0.7	400
C	0.5	250
D	0.2	200
E	0.4	140

Methodology

Performance of Customers booked from past campaign

Train data based on the acquired customers and implement it on the past campign data to test the results and make further modifications

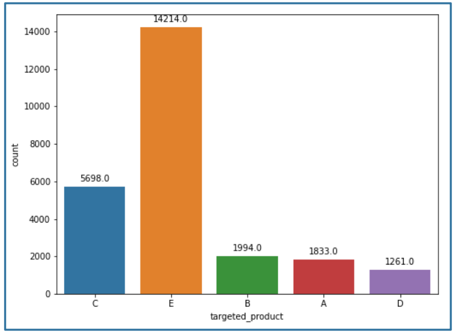
Past Campaign Data

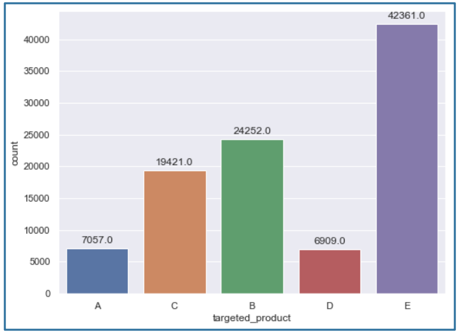
Test and validate results from the above models and make modifications in to predict credit card for the potential customers

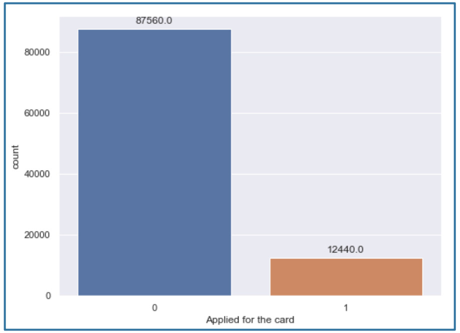
Potential Customer Data

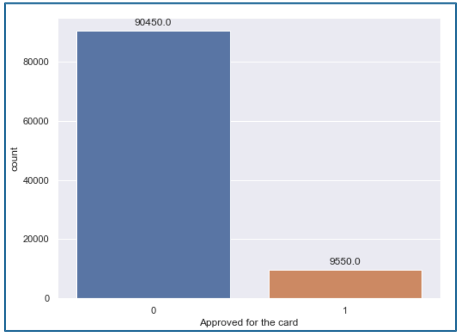
Validate results and provide output on this dataset and further calculate profit

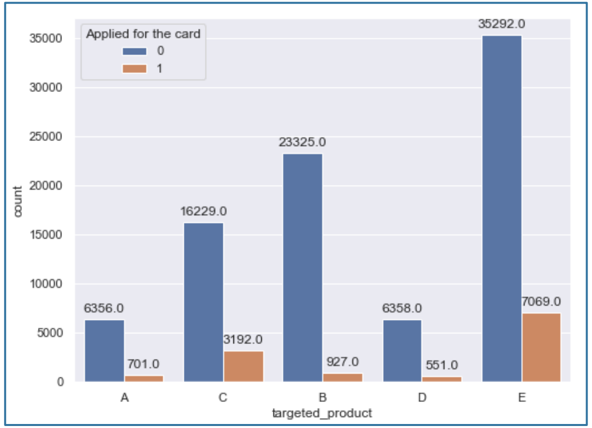
Exploratory Data Analysis

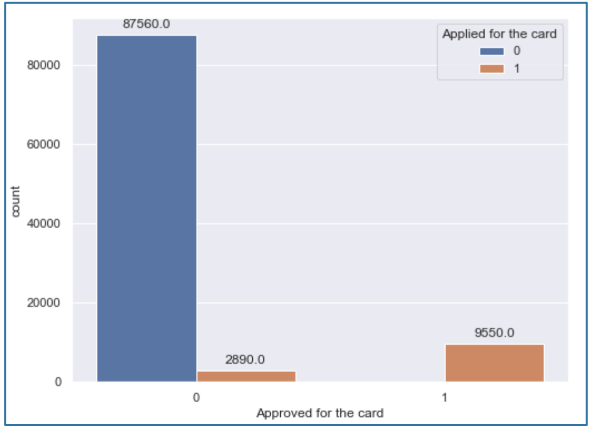


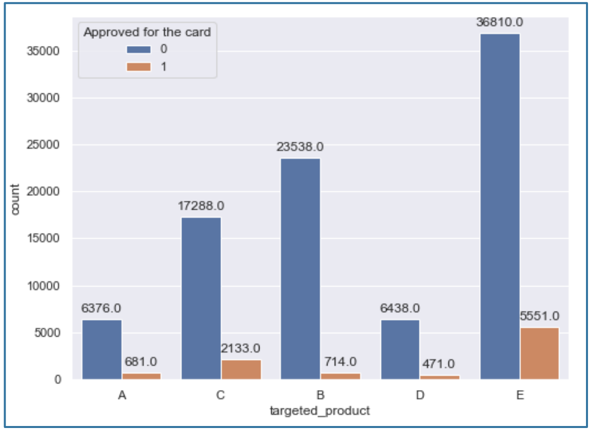




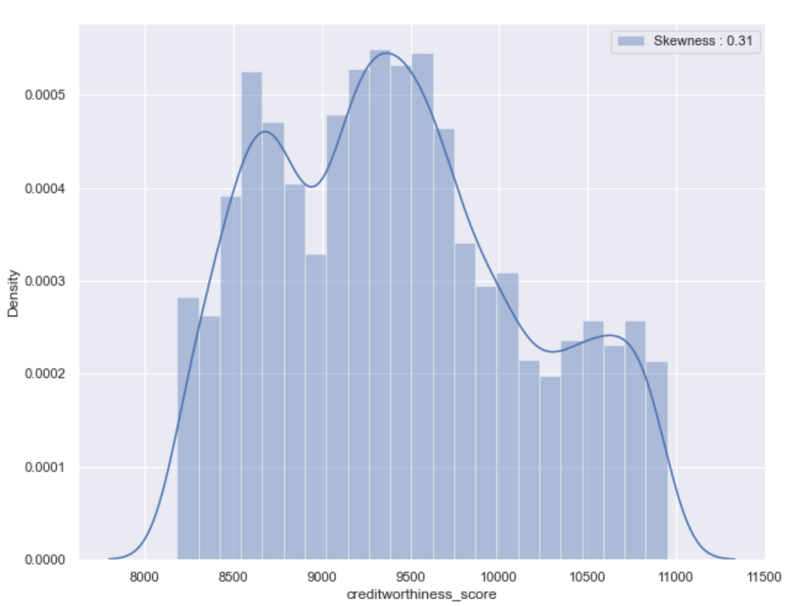


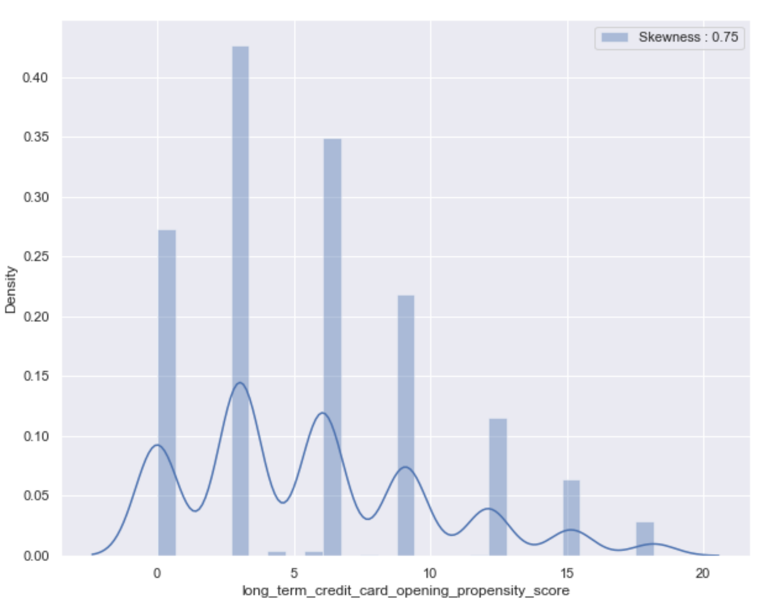




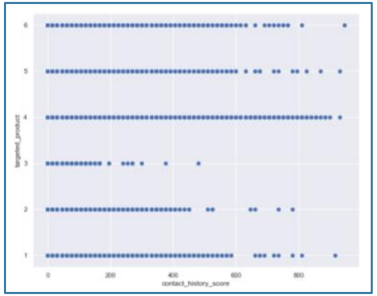


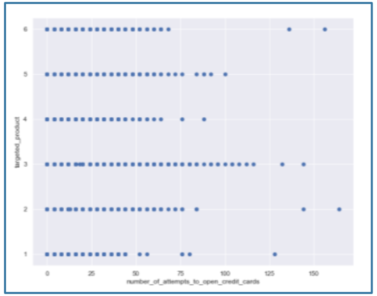
Exploratory Data Analysis

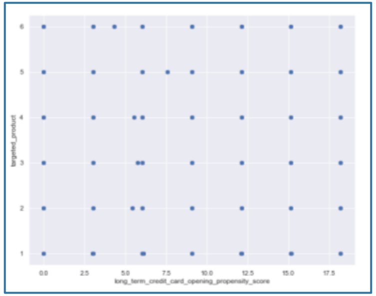


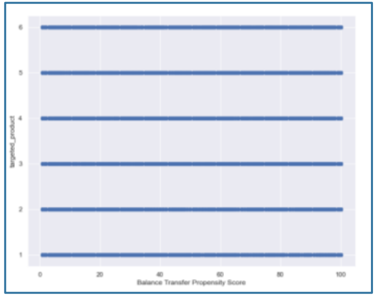


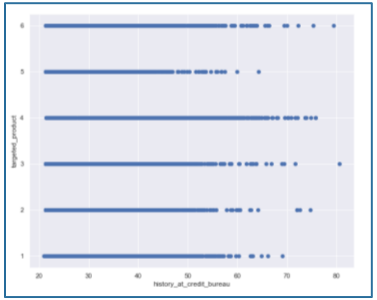
Exploratory Data Analysis

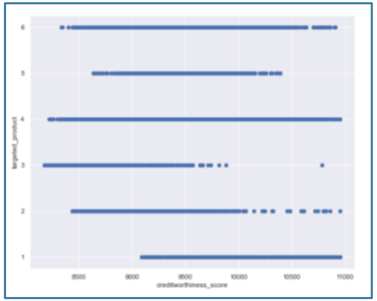


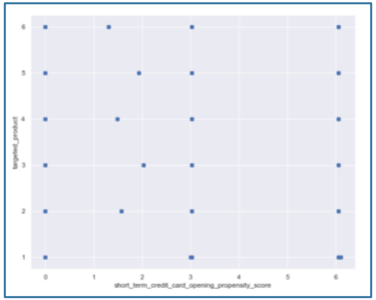


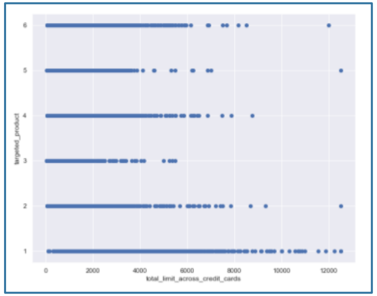


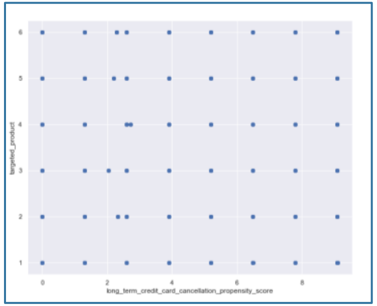


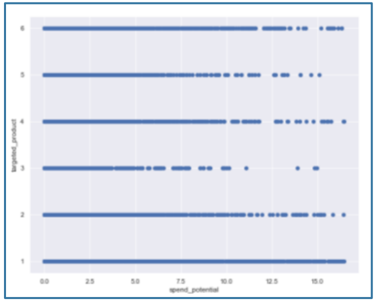


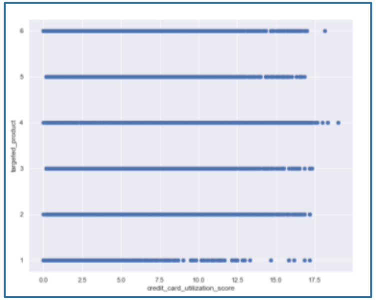


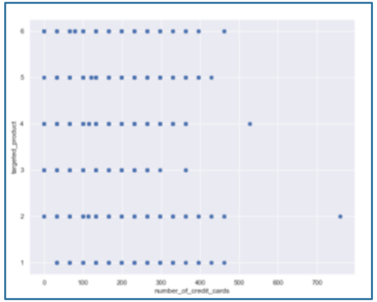




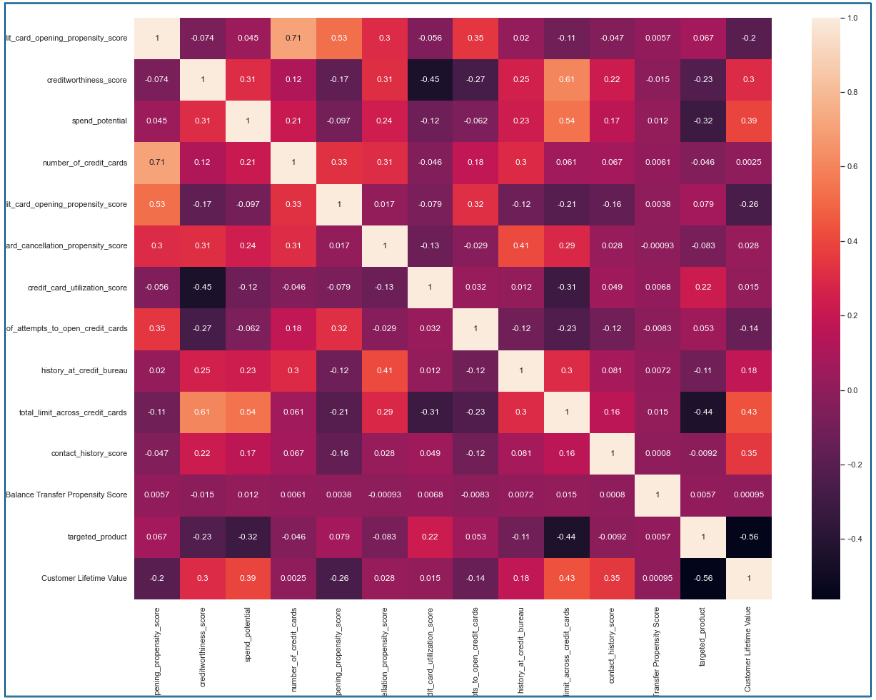


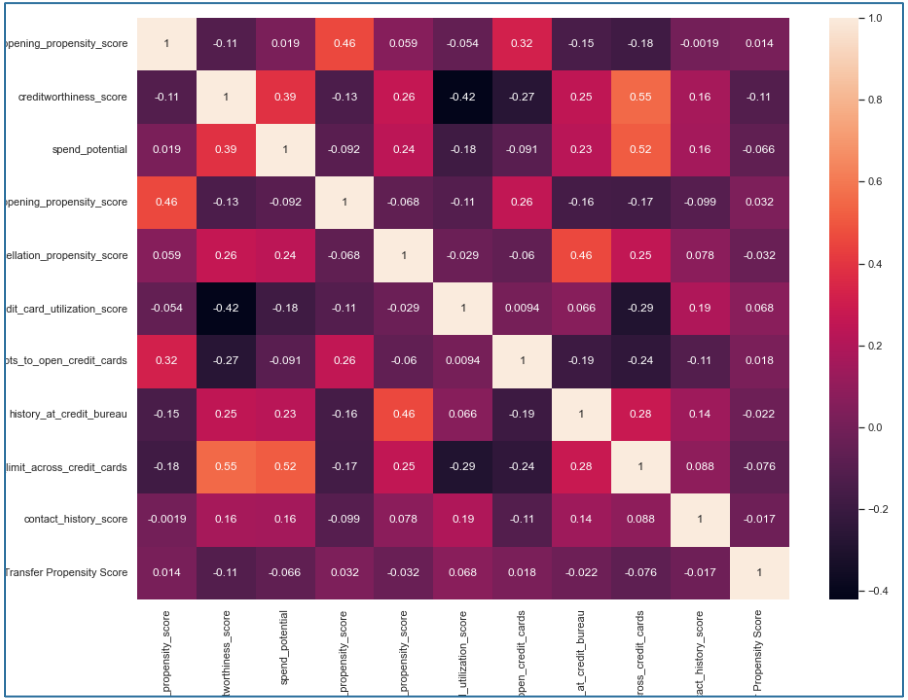






Correlation Matrices





Analysis – Dataset 1

Model Specification	Accuracy
knn credit card A	0.8813
knn credit card B	0.7850
knn credit card C	0.8504
knn credit card D	0.8077
knn credit card E	0.8109
Overall – knn	0.4568
Overall – Naïve	0.4513
Overall – SVC	0.4790
Overall – Logistic	0.4681

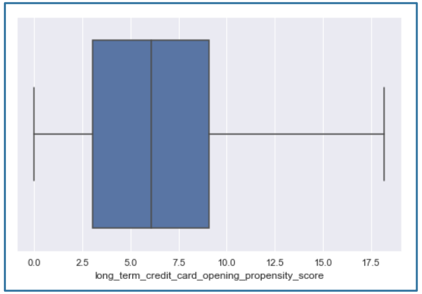
	Predicted (A)	Predicted (B)	Predicted (C)	Predicted (D)	Predicted (E)	Total
Actual (A)	340	8	1	17	81	447
Actual (B)	34	48	120	75	174	451
Actual (C)	9	23	269	34	94	429
Actual (D)	113	30	64	119	123	449
Actual (E)	43	28	38	37	278	424
Total	539	137	492	282	750	2200

- ❑ Removal of outliers, Dropping null values, as there were not much data removed after this. 21228 data points left out of 25000
- ❑ Random sampling out of the data to create a balanced dataset
- ❑ Creating multiple models and testing the accuracy of each of it with the testing dataset, i.e., 33%
- ❑ As accuracy was close for all models, so chosen all models to proceed with and check the accuracy with dataset 2
- ❑ Created a random forest regressor model to predict customer lifetime value for the potential customer data

Analysis – Dataset 2

Model Specification	Accuracy
knn credit card A	0.9274
knn credit card B	0.7843
knn credit card C	0.8544
knn credit card D	0.8973
knn credit card E	0.6463
Overall – knn	0.6256
Overall – Naïve	0.5440
Overall – SVC	0.6112
Overall – Logistic	0.5761

	Predicted (A)	Predicted (B)	Predicted (C)	Predicted (D)	Predicted (E)	Predicted (X)
Actual (A)	240	16	0	8	16	15
Actual (B)	16	163	71	9	39	28
Actual (C)	0	21	261	3	16	8
Actual (D)	18	15	12	223	19	24
Actual (E)	14	31	16	12	212	29
Actual (X)	32	38	35	44	63	51



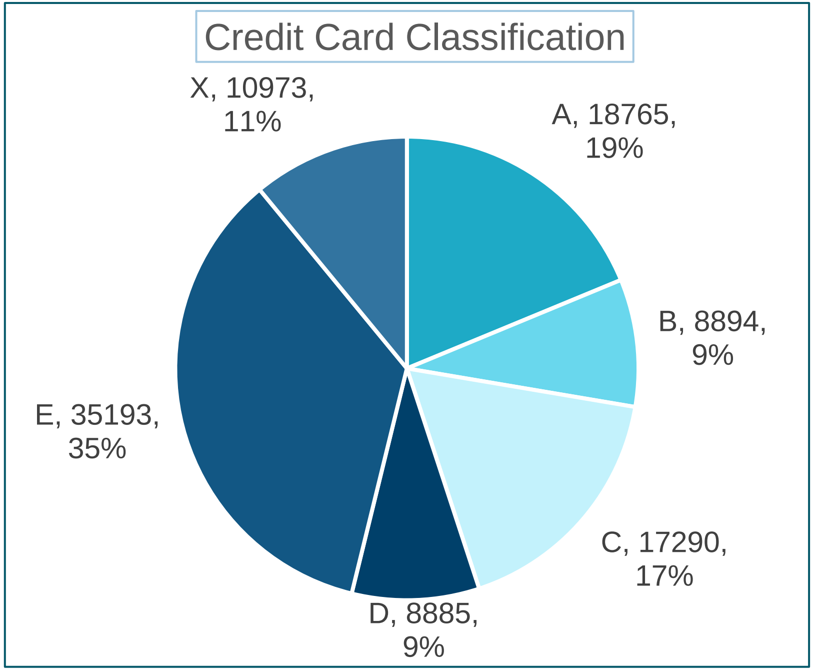
- First testing the data with knn model from performance data and doing the required modification
- No outliers left after preprocessing the data
- Created a knn model to predict the credit card for potential customer data
- It was able to classify all credit cards correctly to some extent, except the case where there is no allocation

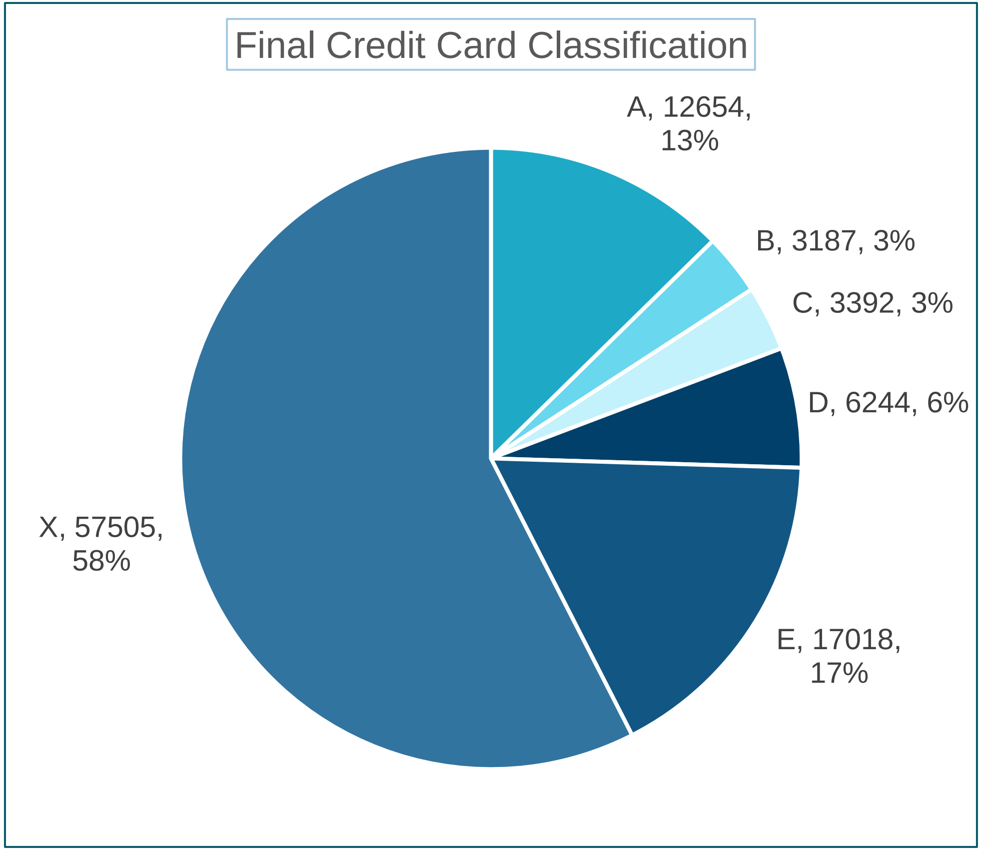
Conclusion

- ❑ Let's discuss what happens when we implement the models on potential customer data

- ❑ Results: Multiclass classification, Profit Calculation

Results


Credit Card Classification
X, 10973, 11%
A, 18765, 19%
B, 8894, 9%
C, 17290, 17%
D, 8885, 9%
E, 35193, 35%


Final Credit Card Classification
A, 12654, 13%
B, 3187, 3%
C, 3392, 3%
D, 6244, 6%
E, 17018, 17%
X, 57505, 58%

Results

Credit Card	Targeted	Response Rate	No. of applicants	Approval Rate	Approvals	Average CLV	Total CLV
A	12654	0.1101 %	14	97.1469 %	14	5441.39	76179.46
B	3187	0.0397 %	1	77.0226 %	1	4352.74	4352.74
C	3392	0.1962 %	7	66.8233 %	5	827.03	4135.15
D	6244	0.0865 %	5	85.4809 %	4	2082.53	8330.12
E	17018	0.1998 %	34	78.5259 %	27	1582.22	42719.94

- ❑ Total CLV = 76179.46 + 4352.74 + 4135.15 + 8330.12 + 42719.94 = 135717.41
- ❑ Total Budget = 50000
- ❑ Our estimated customer lifetime value is much higher than the proposed budget to acquire customers. So, this seems to be a profitable business, but we also have to look out how we have utilized the budget
- ❑ All calculations are in dollars ($)

Results

Credit Card	Targeted	Targeting Cost	Total TC	Approvals	Welcome Cost	Total WC	Total Cost
A	12654	2	25308	14	1500	21000	46308
B	3187	0.7	2230.9	1	400	400	2630.9
C	3392	0.5	1696	5	250	1250	2946
D	6244	0.2	1248.8	4	200	800	2048.8
E	17018	0.4	6807.2	27	140	3780	10587.2

- ❑ Total Cost = 46308 + 2630.9 + 2946 + 2048.8 + 10587.2 = 64520.9
- ❑ Penalty = 64520.9 – 50000 = 14520.9
- ❑ Total Cost after penalty = 64520.9 + 14520.9 = 79041.8
- ❑ Profit = 135717.41 - 79041.8 = 56675.61
- ❑ This credit classification approach seems to be profitable for the business

THANK YOU