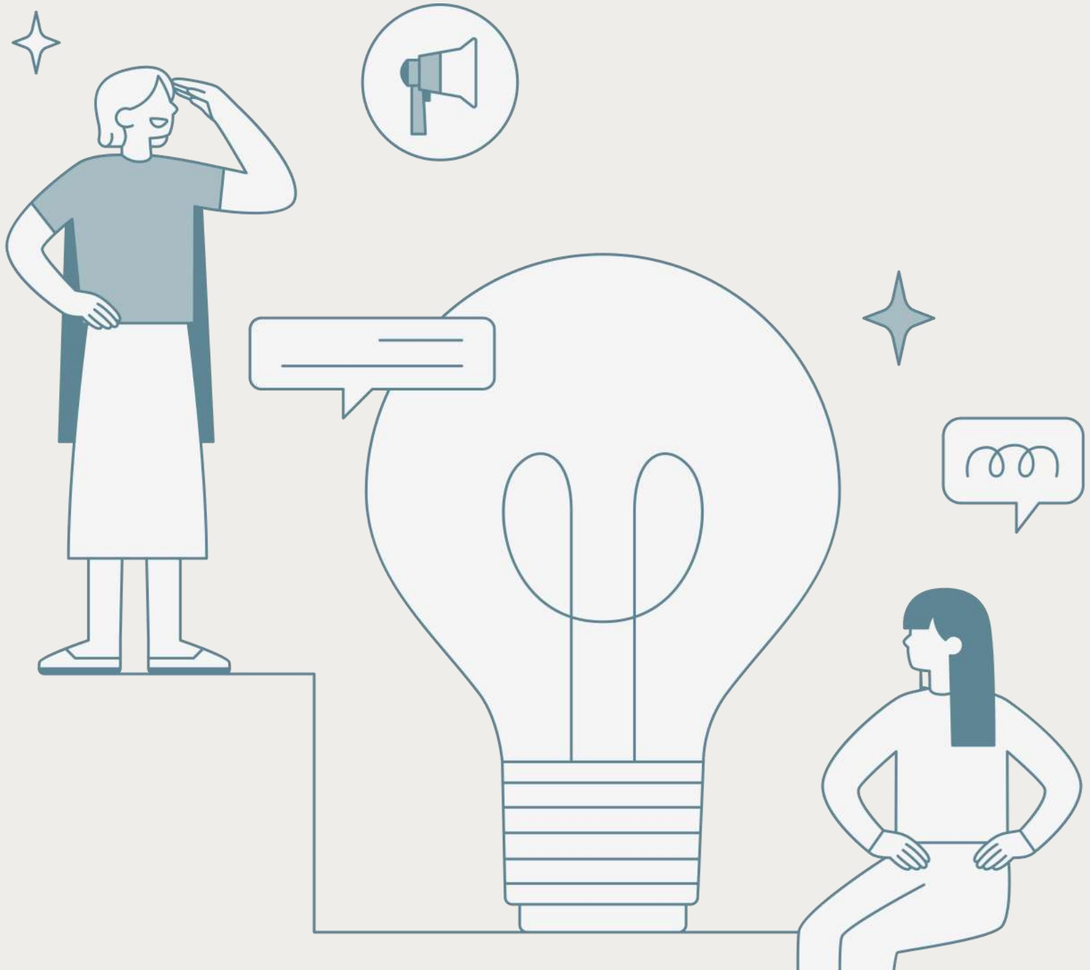
Customer Retention in the Telco Industry



Aman Agarwal

Low Ji Xiong

Dhreeti Shah

Lin, Shih-Kuan

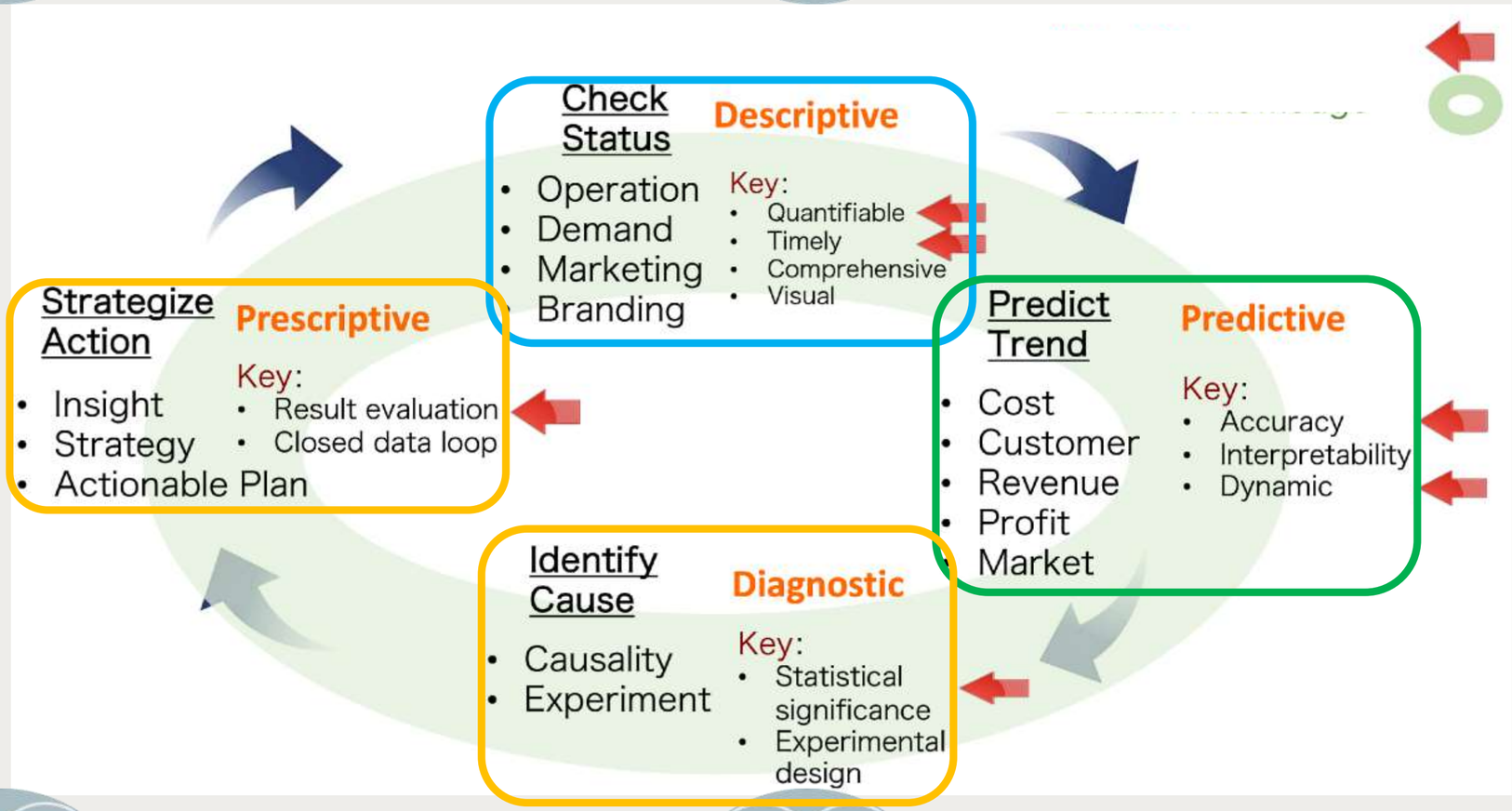
Ilansurya Ilanchezhiyan

Introduction

Descriptive Analysis

Business Problems

SCOPE



Data Science Techniques

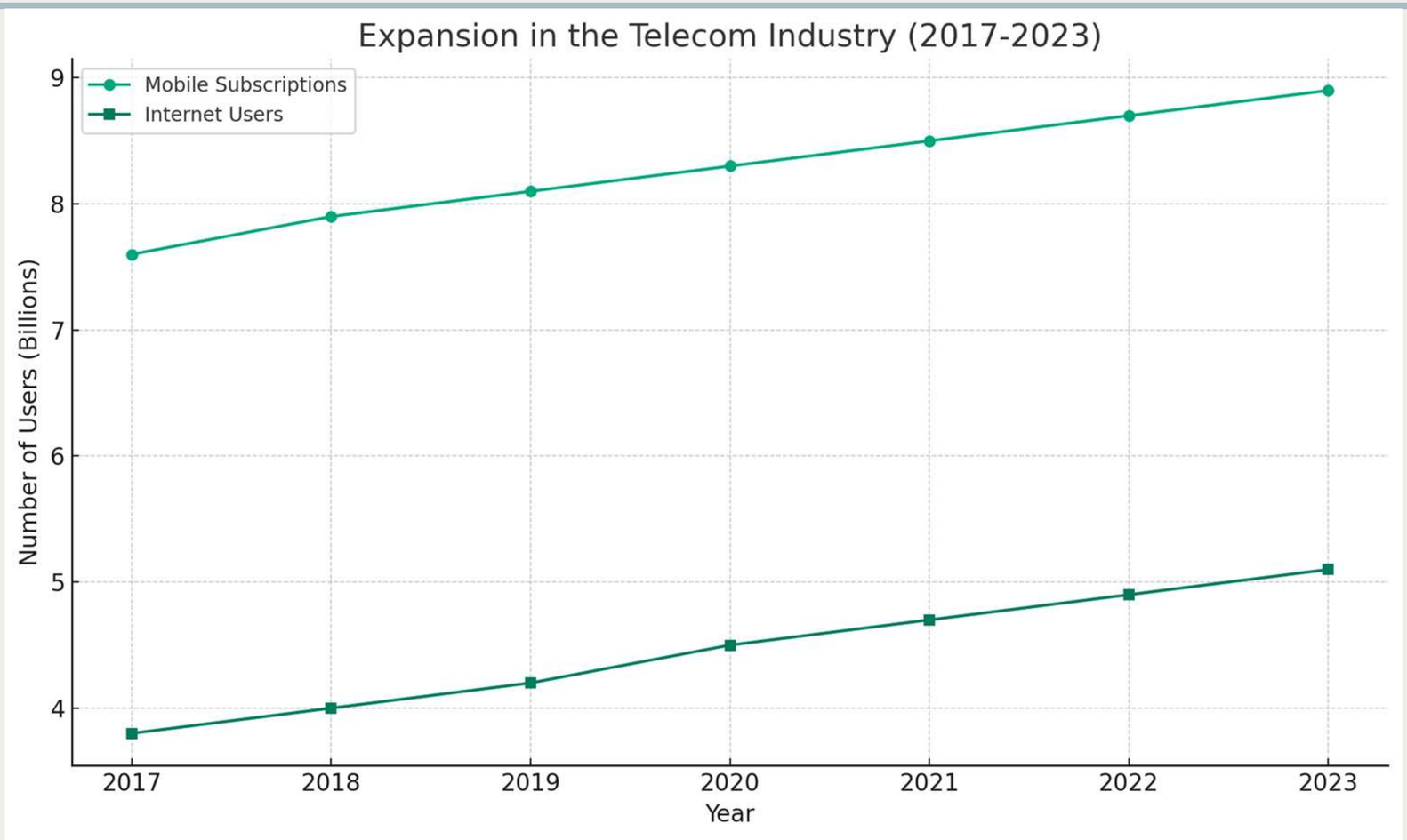
Recommendation

Way Ahead

Why is the Telco Industry important ?

➔The telecom market is witnessing an ***exponential increase in mobile data usage,*** projected to reach 483 exabytes (EB) per month by 2028

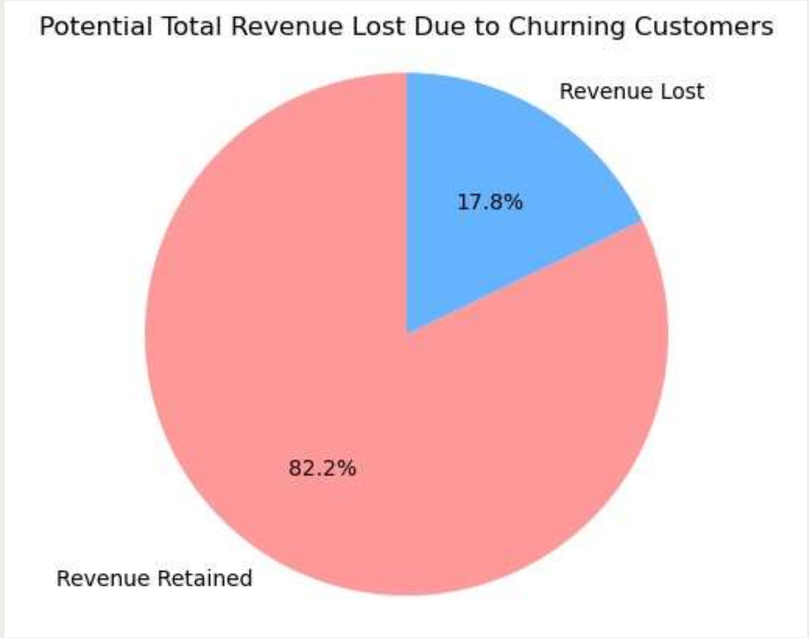
➔***Predictions for the telecom industry*** include the addition of more than 400 million new mobile subscribers by 2025.



This graph showing the ***expansion*** in the telecom industry from 2017 to 2023, highlighting both the growth in ***global mobile subscriptions and internet users over the years***. This illustrates the industry's ***significant upward trajectory***, reflecting its importance and the ***increasing*** global reliance on telecom services

Descriptive Analysis

- **Dataset**
  - 3 key segments of variables
    - *Customer demographics, Subscription, Billing*
  - 20 Variables, 7k customers
  - Generate over $16 million in revenue
    - *Avg of $2.3k per customer*



- **Churn Statistics**
  - Churn customers: 1.9k (26.6% of customer base)
  - Potential loss of revenue: $2.8 million (17.8% of total)

Understanding Challenges: Business Problem

A Problem at Hand

1. **High Customer Churn:** Company experiencing a sudden significant loss of customers
2. **Declining Market Share:** Noticeable decline in market share, signaling a weakening position in the competitive landscape

B Core Challenges Identified

1. **Understanding Customer Churn:** Critical need to identify predictive factors for customer churn
2. **Regaining Market Share :** Need for guidance in optimizing of marketing efforts and resources to regain market share

Understanding Challenges: Designing the Solution

C Business Assumptions

1. **Customer Acquisition Cost:** New customer acquisition cost significantly higher, and less preferred since characteristics and preferences are also less known
2. **Customer Preferences:** Customer needs & preferences towards services will maintain over the mid term for marketing strategies to be implementable

D Solution Strategy

1. **Customer Segmentation & Tiering:** Segment and tier customers by likelihood of churn while also prioritizing by profitability to guide targeted re-acquisition and retention strategies
2. **Targeted Promotion bundles:** Determining targeted promotion bundles to re-attract and retain high churn likelihood customers

Analysis: Techniques

Predictive (Supervised)- Decision Tree

Identify factors that relate to & predict customer churn

Predictive (Unsupervised)- K-means Clustering

Segment customers based on data features

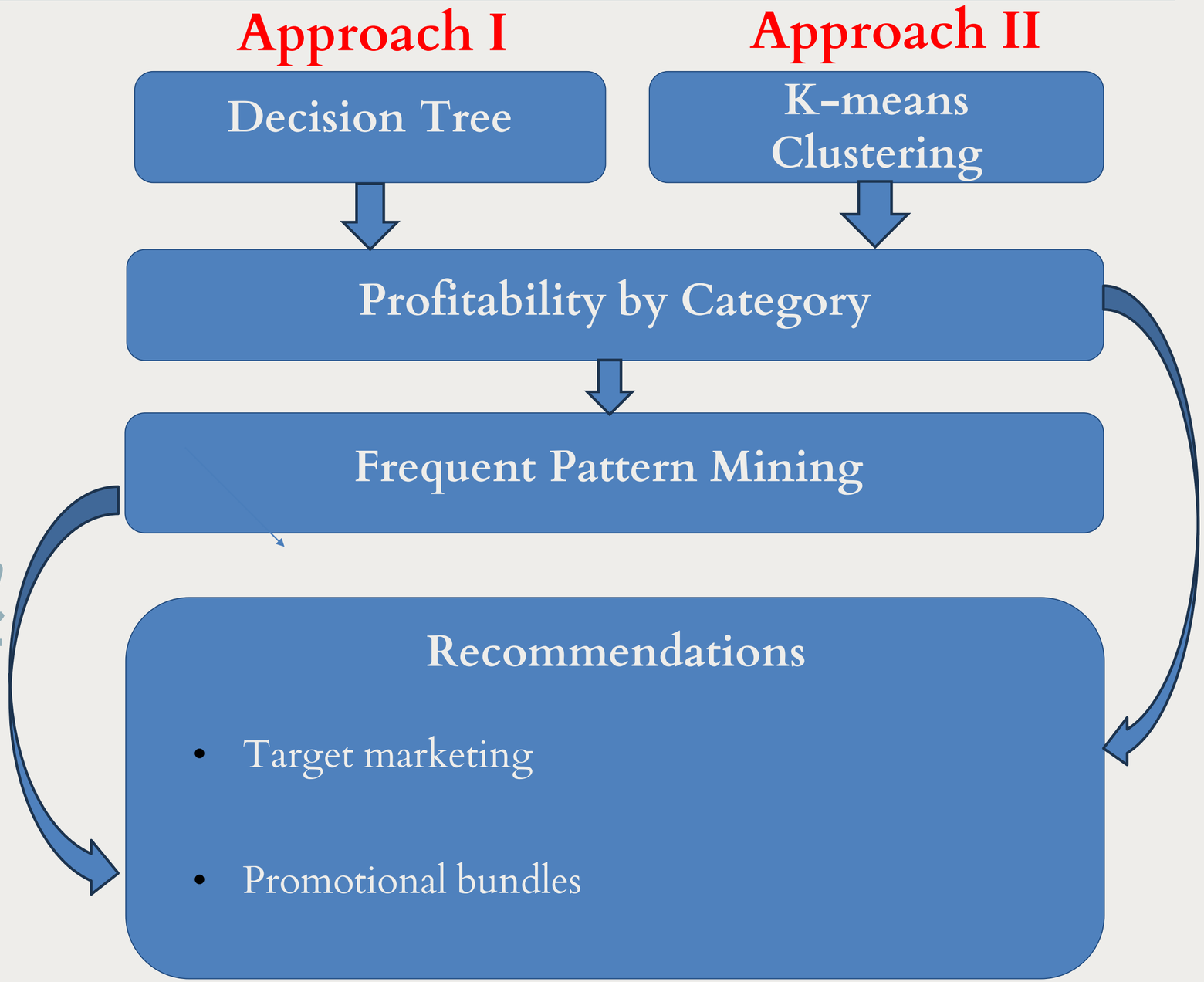
Summary & Descriptive Statistics

Determine profitability by categories

Frequent Pattern Mining

To find out customer preference across telco's services

Analysis: Application



Objective (I): Decision Tree
Identifying factors that relate to and predict customer churn

Step 1 : Data preparation and Transformation

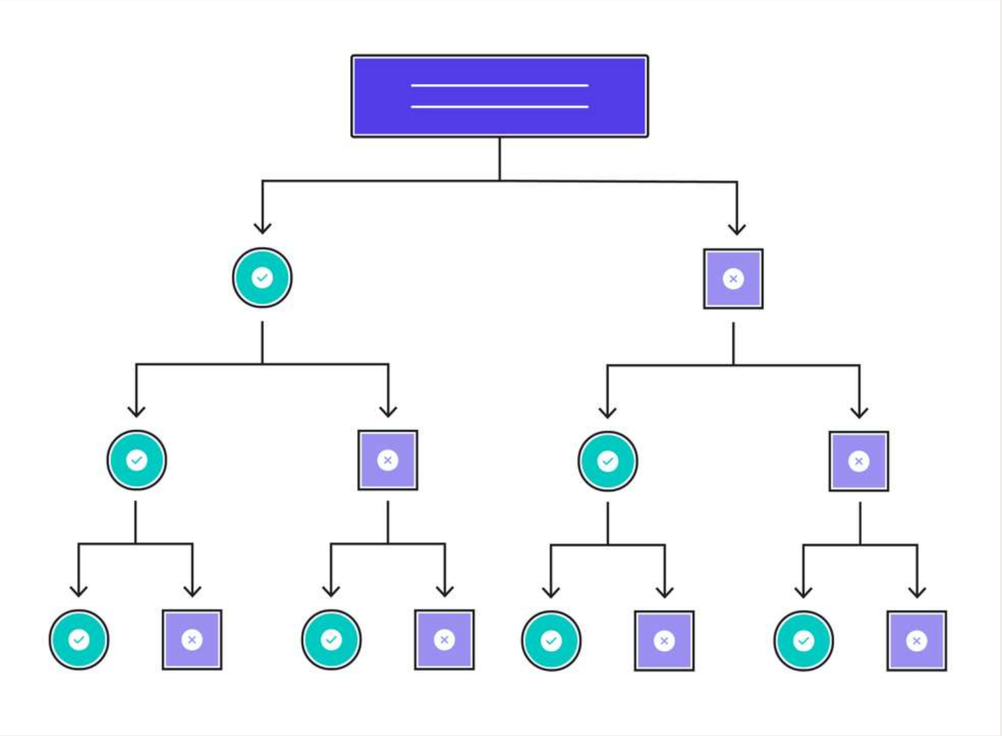
Step 2 : Correlation Matrix analysis

Step 3 : Decision Tree analysis

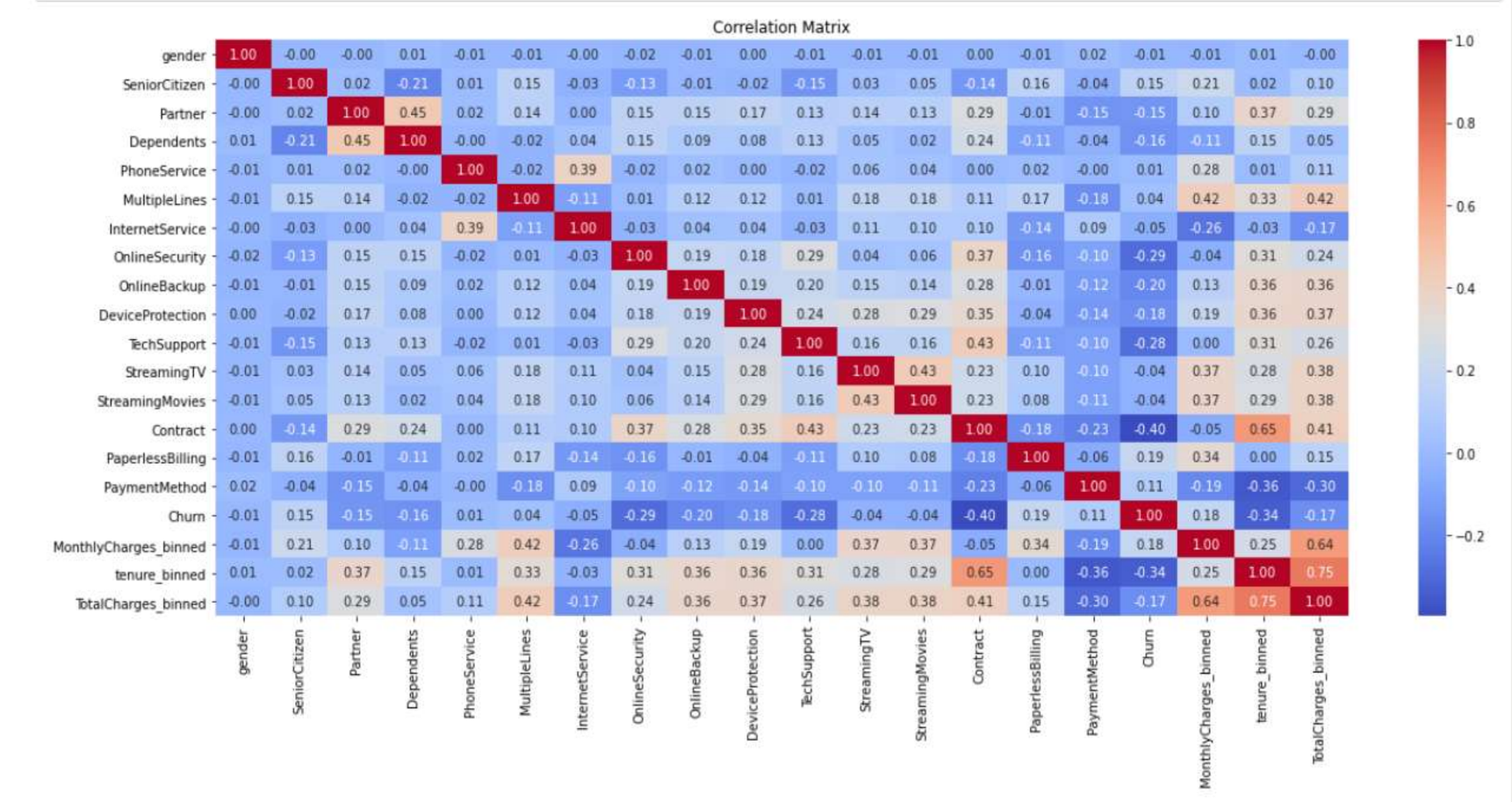
Step 4 : Extracting Important features

Step 5 : Customer Distribution in important features.

Step 6 : Model Performance



Objective (I): Decision Tree
Identifying factors that relate to and predict customer churn



Dropping totalcharges_binned from the data-frame due to high correlation

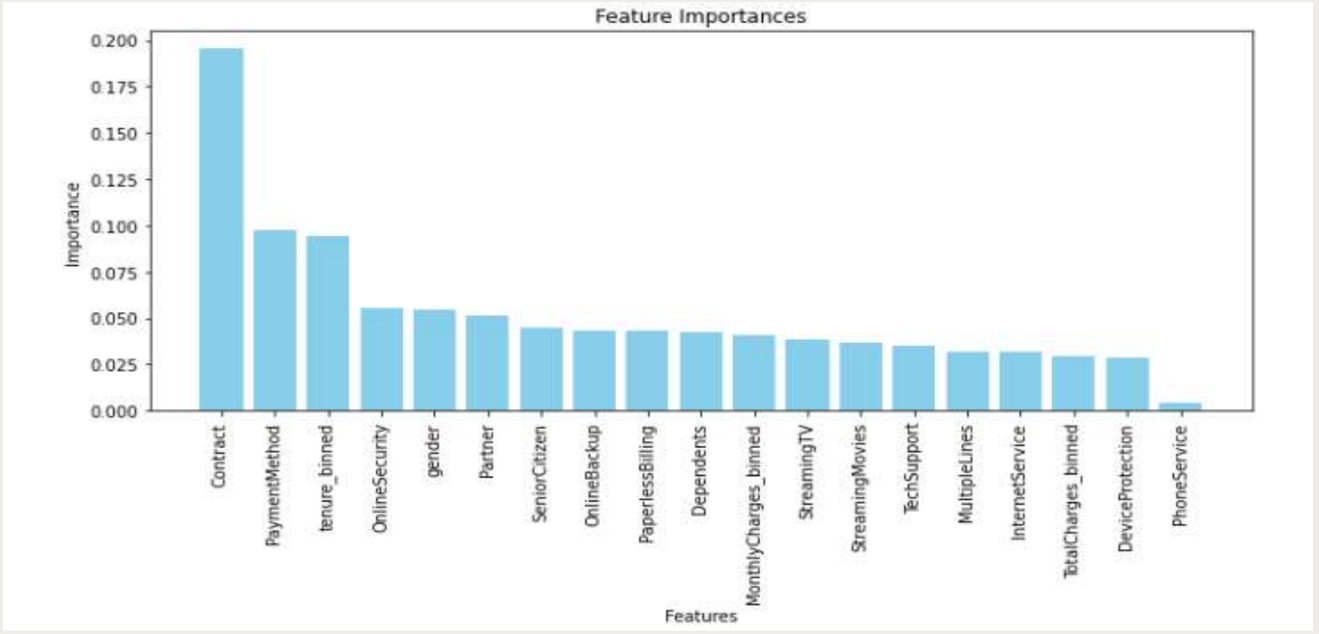
Objective (I): Decision Tree
Identifying factors that relate to and predict customer churn

Feature Importance :

Contract : ~20%

Payment method : ~9%

Tenure : ~11%



Objective (I): Decision Tree
Identifying factors that relate to and predict customer churn

Contract	Churn	No of customers	% of customers
Month to Month(0)	0	2220	31.52065881
	1	1655	23.49850916
One year (1)	0	1307	18.55743291
	1	166	2.356950163
Two year(2)	0	1647	23.3849212
	1	48	0.681527758

Tenure	Churn	No of customers	% of customers
0-15 months (0)	0	1272	18.06048559
	1	1099	15.60414596
16-29 months (1)	0	831	11.79894931
	1	295	4.188556013
30-43 months (2)	0	774	10.9896351
	1	215	3.052676416
44-57 months (3)	0	805	11.42978844
	1	142	2.016186284
58+ months (4)	0	1492	21.18415448
	1	118	1.675422405

Payment Method	Churn	No of Customers	% of customers
Bank Transfer(0)	0	1286	18.25926452
	1	258	3.6632117
Credit Card (1)	0	1290	18.3160585
	1	232	3.294050831
Electronic Check (2)	0	1294	18.37285248
	1	1071	15.2065881
Mailed Check (3)	0	1304	18.51483743
	1	308	4.373136448

This customer distribution gives us an overview for performing summary statistics and frequent pattern mining

Objective (I): Decision Tree
Identifying factors that relate to and predict customer churn

Accuracy, Confusion Matrix, and Classification Report:

Metric	Value
Accuracy	0.72

	Predicted No Churn	Predicted Churn
True Negative	2079	530
True Positive	448	465

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.80	0.81	2626
1	0.45	0.47	0.46	896
accuracy			0.72	3522
macro avg	0.63	0.64	0.64	3522
weighted avg	0.72	0.72	0.72	3522

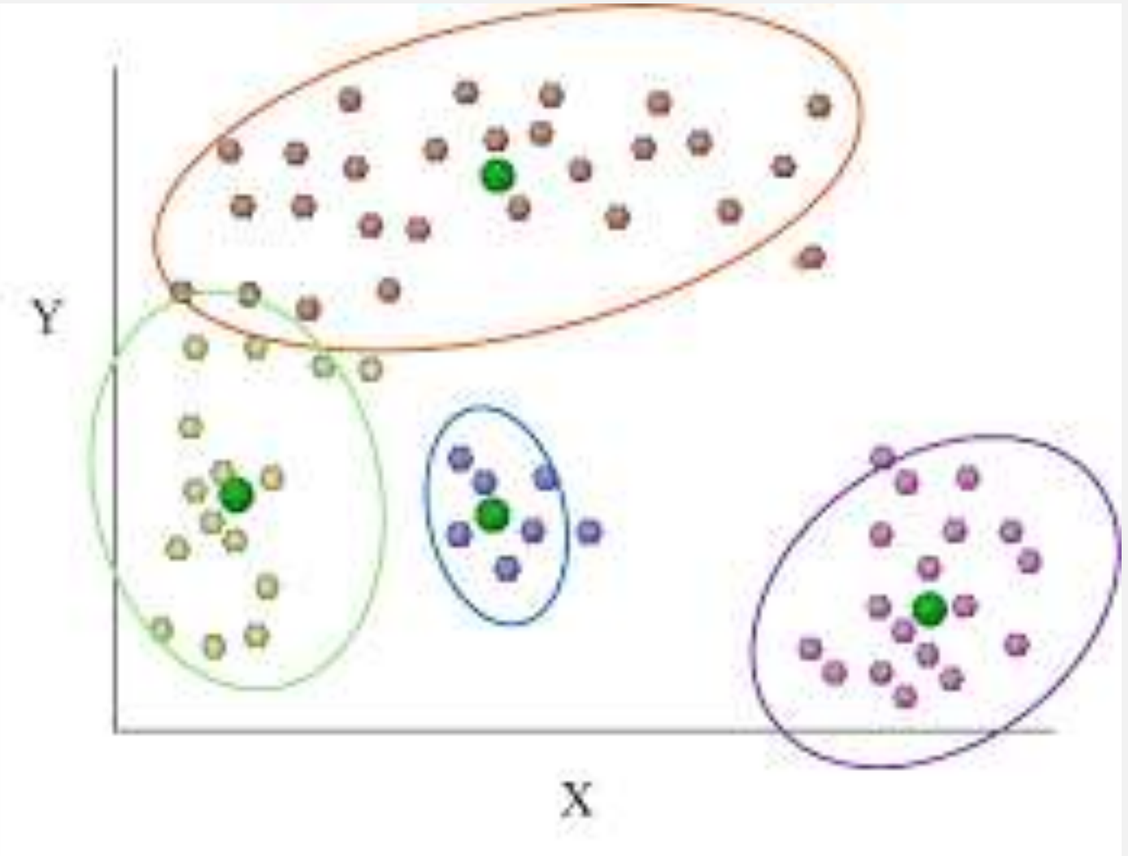
The accuracy of the decision tree analysis is 72%

Objective (II):
Identifying customer segments for targeting

Step 1: determine ideal cluster count

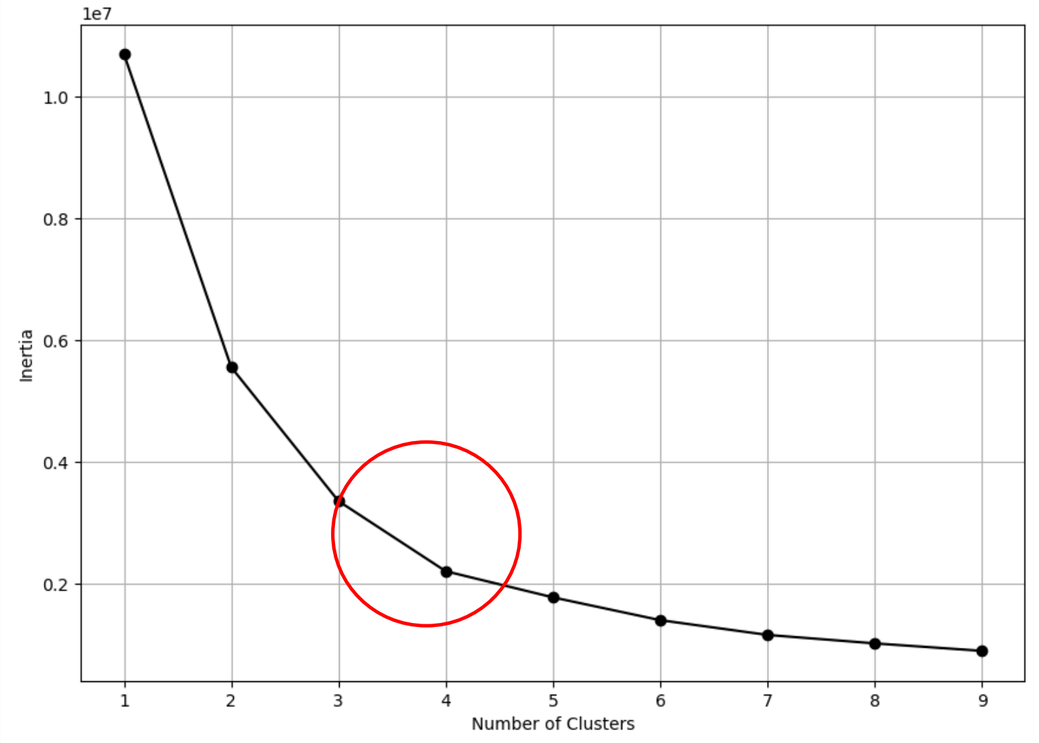
Step 2: clustering

Step 3: determining distribution of within clusters

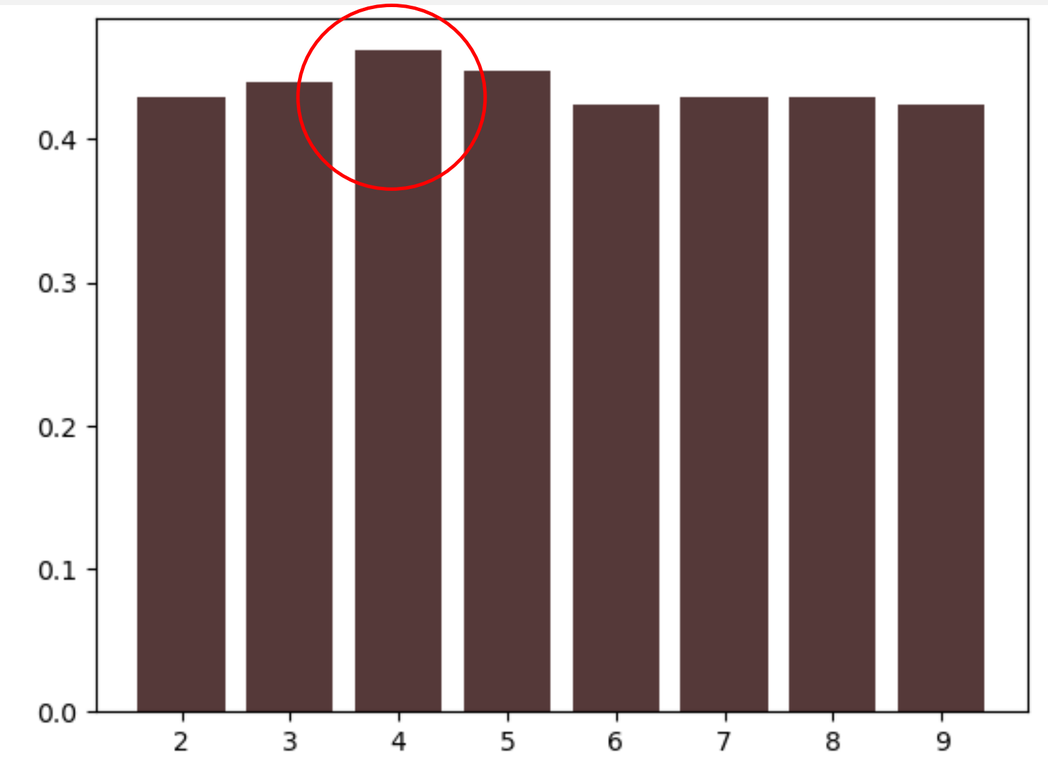


Objective (II):
Identifying customer segments for targeting

K-means elbow method

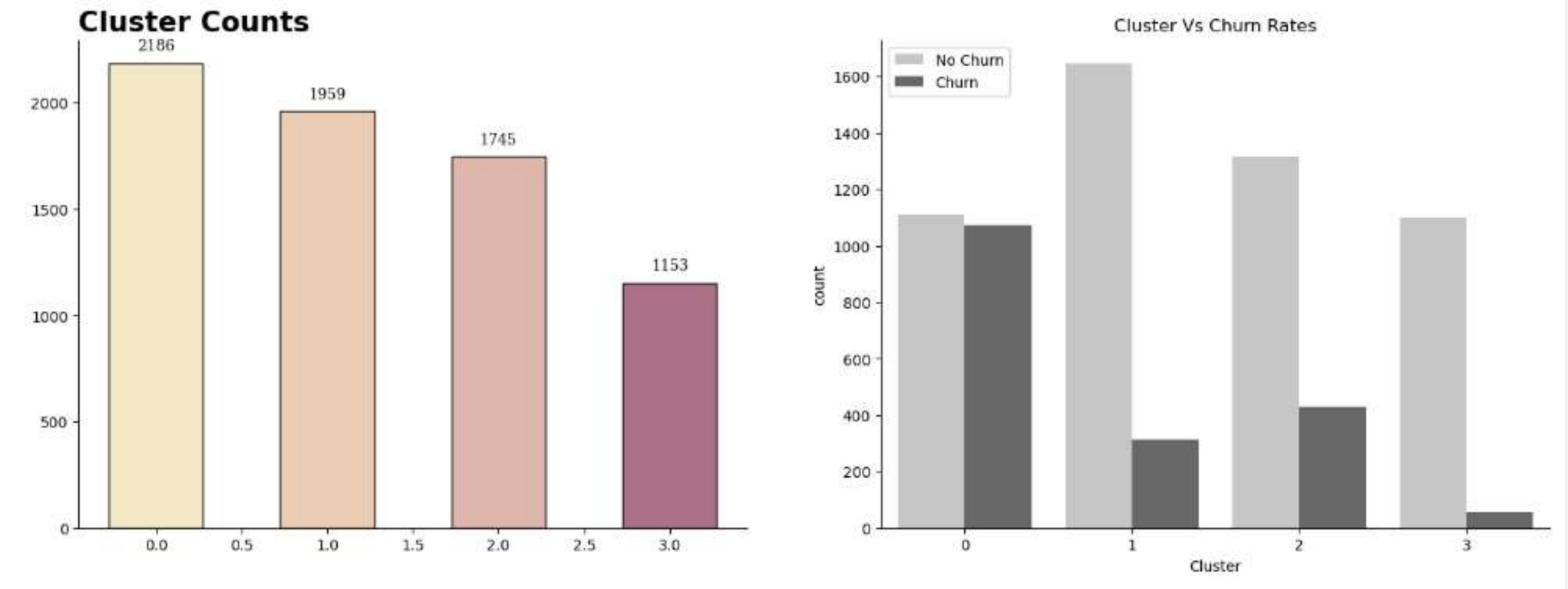


Silhouette score



For this objective, segmentation is performed. We applied "K-means elbow method and silhouette score to identify the optimal number of clusters.

Objective (II):
Identifying customer segments for targeting



After the segmentation/clustering is done, ***2 metrics*** *are used to evaluate the 4 clusters.*

Since ***cluster 0 has a relatively highest churn rate,*** *we focus on analysing cluster 0.*

Also consider ***Cluster 1 as the 2nd biggest cluster****.*

Objective (II):
Identifying customer segments for targeting

Detailed statistics of cluster 0

gender	Counts	Percentage
Female	1106	50.594694
Male	1080	49.405306

Contract	Counts	Percentage
Month-to-month	1898	86.825252
One year	239	10.933211
Two year	49	2.241537

SeniorCitizen	Counts	Percentage
0	1685	77.081427
1	501	22.918573

InternetService	Counts	Percentage
DSL	476	21.774931
Fiber optic	1710	78.225069

OnlineSecurity	Counts	Percentage
No	1700	77.767612
Yes	486	22.232388

Partner	Counts	Percentage
No	1396	63.860933
Yes	790	36.139067

Dependents	Counts	Percentage
No	1752	80.146386
Yes	434	19.853614

PhoneService	Counts	Percentage
No	14	0.640439
Yes	2172	99.359561

MultipleLines	Counts	Percentage
No	1158	52.973468
No phone service	14	0.640439
Yes	1014	46.386093

OnlineBackup	Counts	Percentage
No	1532	70.082342
Yes	654	29.917658

DeviceProtection	Counts	Percentage
No	1504	68.801464
Yes	682	31.198536

TechSupport	Counts	Percentage
No	1660	75.937786
Yes	526	24.062214

PaymentMethod	Counts	Percentage
Bank transfer (automatic)	346	15.827996
Credit card (automatic)	312	14.272644
Electronic check	1198	54.803294
Mailed check	330	15.096066

The above graphs show the detailed characteristics of cluster 0.

Objective (II):
Identifying customer segments for targeting

Detailed statistics of cluster 1

gender	Counts	Percentage
Female	975	49.770291
Male	984	50.229709

Contract	Counts	Percentage
Month-to-month	533	27.207759
One year	606	30.934150
Two year	820	41.858091

SeniorCitizen	Counts	Percentage
0	1532	78.203165
1	427	21.796835

InternetService	Counts	Percentage
DSL	573	29.249617
Fiber optic	1386	70.750383

OnlineSecurity	Counts	Percentage
No	906	46.248086
Yes	1053	53.751914

Partner	Counts	Percentage
No	624	31.852986
Yes	1335	68.147014

Dependents	Counts	Percentage
No	1315	67.126085
Yes	644	32.873915

PhoneService	Counts	Percentage
No	23	1.174068
Yes	1936	98.825932

MultipleLines	Counts	Percentage
No	445	22.715671
No phone service	23	1.174068
Yes	1491	76.110260

OnlineBackup	Counts	Percentage
No	632	32.261358
Yes	1327	67.738642

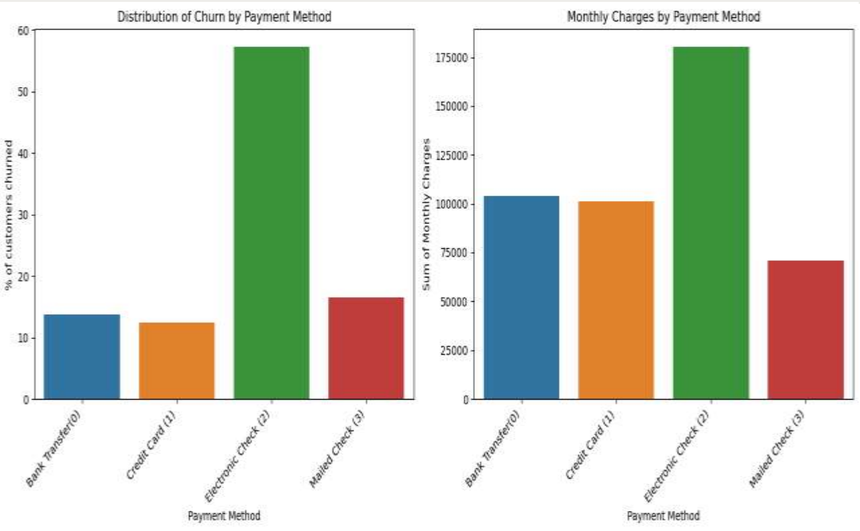
DeviceProtection	Counts	Percentage
No	612	31.240429
Yes	1347	68.759571

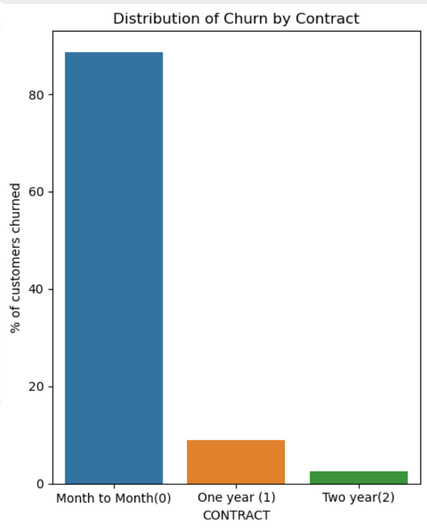
TechSupport	Counts	Percentage
No	875	44.665646
Yes	1084	55.334354

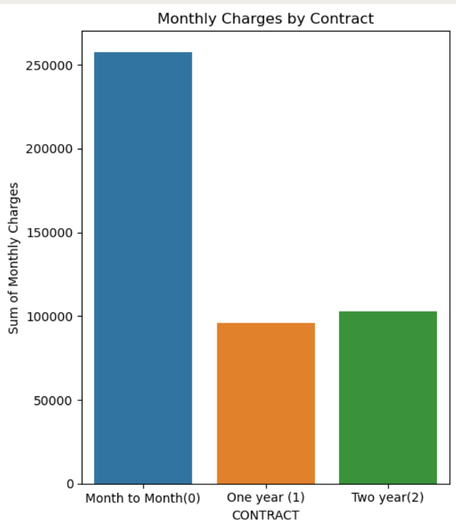
PaymentMethod	Counts	Percentage
Bank transfer (automatic)	615	31.393568
Credit card (automatic)	606	30.934150
Electronic check	598	30.525778
Mailed check	140	7.146503

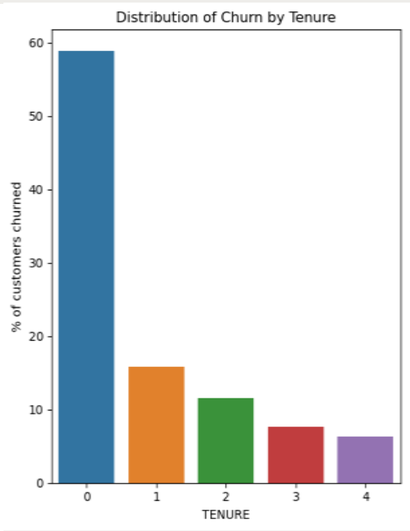
The above graphs show the detailed characteristics of cluster 1.

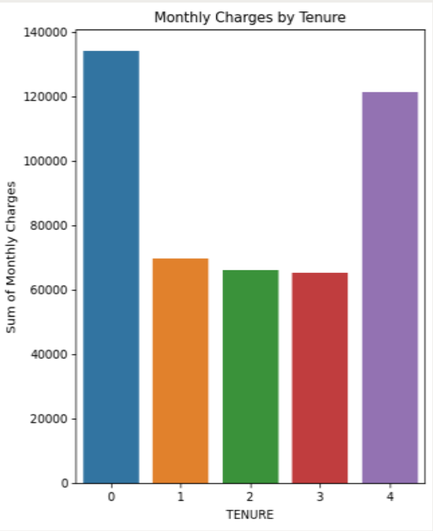
Objective (III):
Determining the most profitable categories







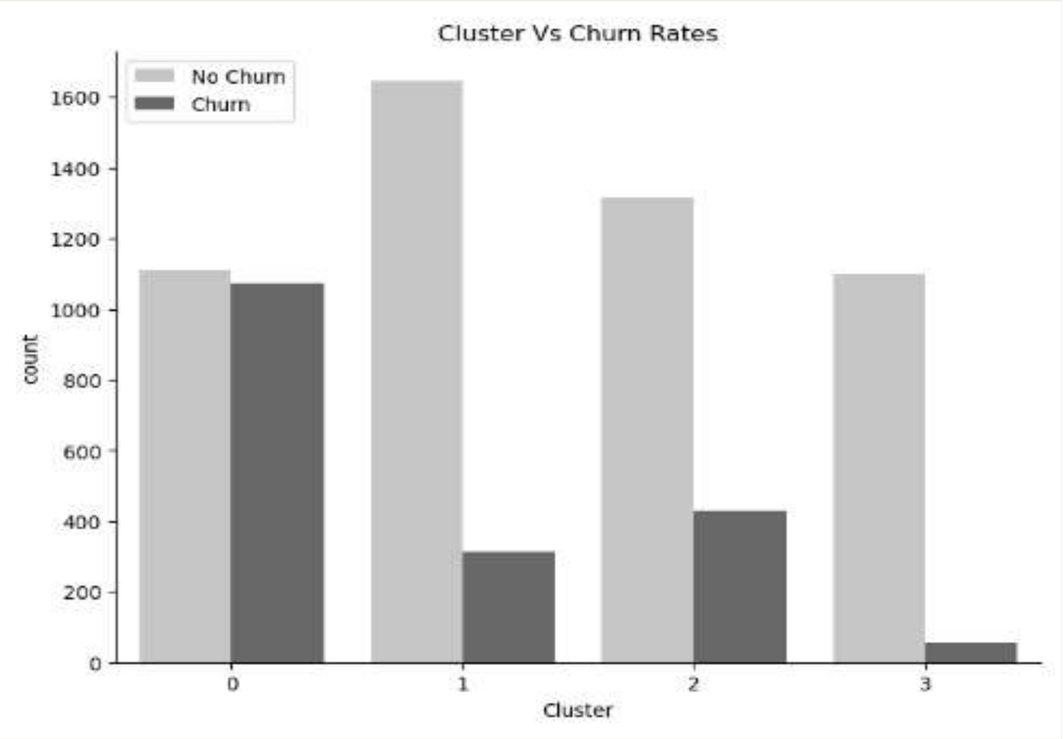


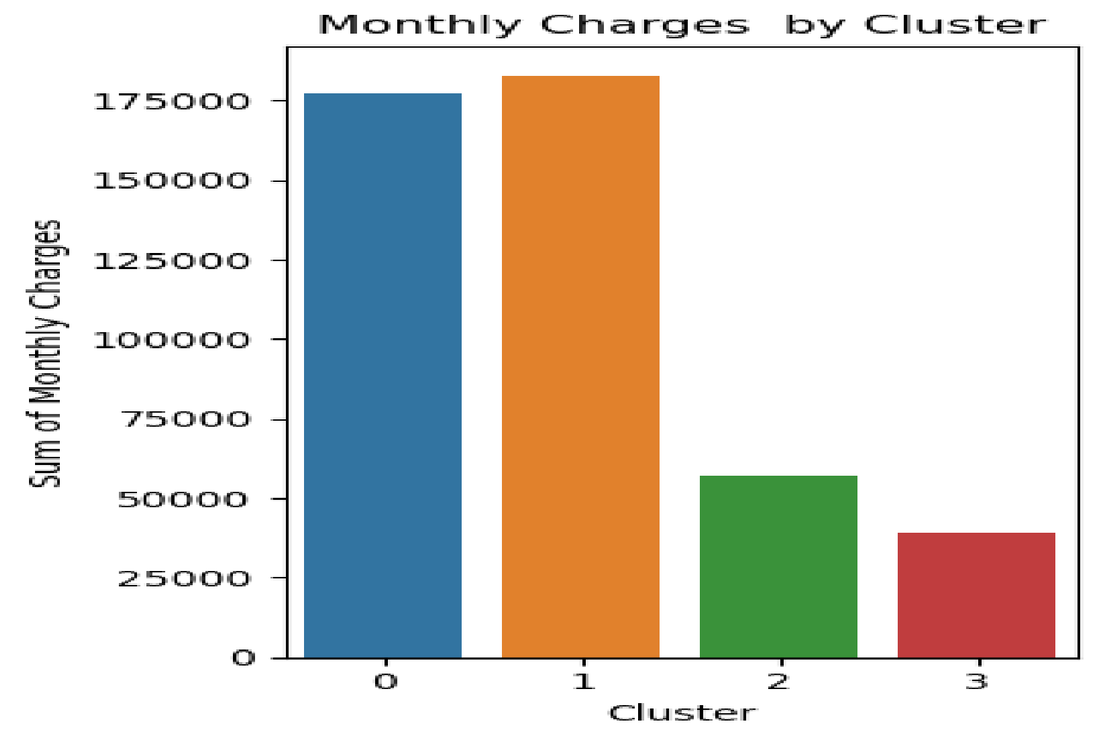


From the Decision Tree Analysis, the top three categories which affected churn were identified and were juxtaposed against revenue per month .

- ***For Payment Method –Leveraging Electronic Checks for Profitability***
- ***For Contract-Optimizing Month to Month Contracts***
- ***For Tenure- Targeting Tenure Segments(BIN 0 (0-15) and BIN 4(58+)) for Retention and Revenue***

Objective (III):
Determining the most profitable categories





Cluster 0: Unveiling the Hub of Churn Dynamics.

Strategic Approach to Customer Segmentation: Balancing Churn Reduction and Revenue Maximization by Focusing on both Cluster 0 and 1

Objective (IV):
Finding out customer preference of subscriptions

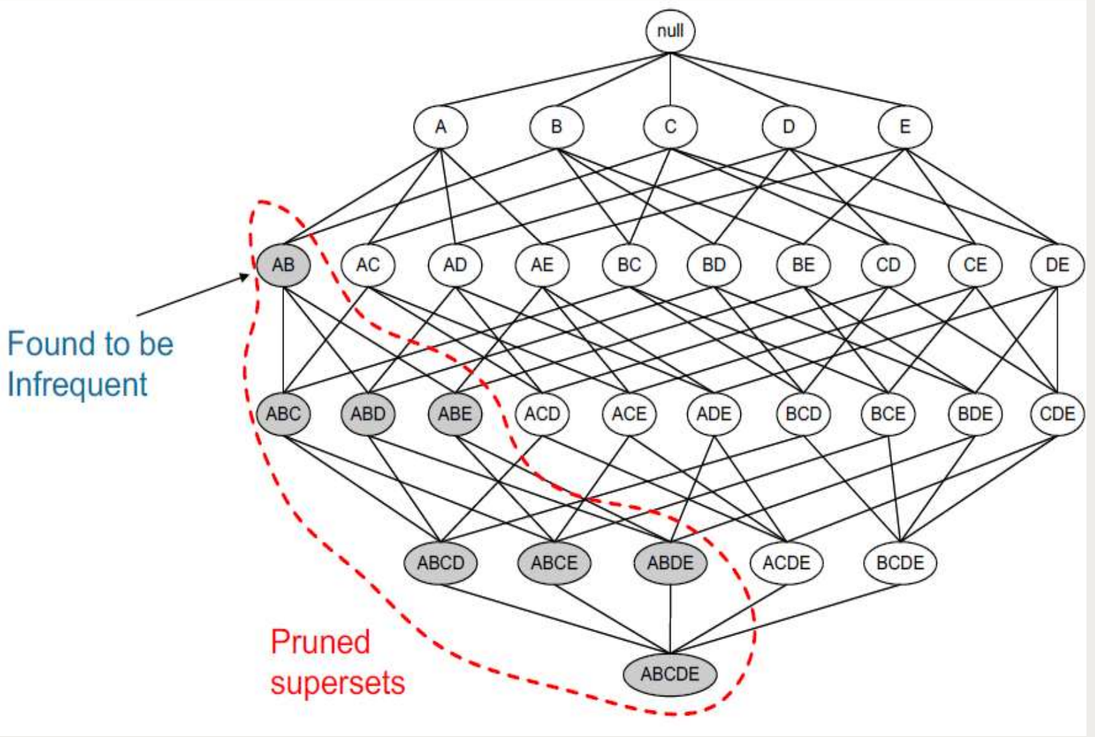
Step 1 : *Consolidate data with categories from Obj 1-3*

Step 2 : *Data Preparation for binary values*

Step 3 : *Prioritising with apriori algorithm, define functions to populate frozensets with list of frequent itemsets*

Step 4 : *Input category and min support to generate result for frequent pattern itemsets*

Step 5 : *Starting with global min support of 0.4, adjust upwards accordingly to retain minimum 1 itemset for k=3 if possible*



Objective (IV):
Finding out customer preference of subscriptions

S/N	Category of Importance	Max k value for itemset with min support of 0.4	Itemsets at Max k value and Support	Stipulated Min support
1	Cluster 0	2	'PhoneService', 'MultipleLine' (0.46)	0.4
2	Cluster 1	3	'StreamingMovies', 'PhoneService', 'MultipleLine'(0.57) **'PhoneService', 'MultipleLine', 'Fiber Optic' (0.57)** 'PhoneService', 'StreamingTV', 'MultipleLine' (0.56)	0.56
3	Month-to-Month Contract	1	'Fiber Optic' (0.55)	0.4
4	Tenure of 0-15 months	1	'Fiber Optic' (0.42)	0.4
5	Tenure of >58 months	3	'StreamingMovies', 'PhoneService', 'MultipleLine' (0.43) 'PhoneService', 'OnlineBackup', 'MultipleLine' (0.43) 'PhoneService', 'DeviceProtection', 'MultipleLine' (0.42)	0.42
6	PaymentMethod of Echeck	3	**'PhoneService', 'MultipleLine', 'Fiber Optic' (0.40)**	0.4
7	All Churn Customers	3	**'PhoneService', 'MultipleLine', 'Fiber Optic' (0.41)**	0.4
8	All Customers	2	'PhoneService', 'MultipleLine' (0.42)	0.4

- ***Results in S/N 1-7 above offer possible options for target marketing promotion bundles***
- ***'PhoneService', 'MultipleLine', 'Fiber Optic' most frequently appearing in the possible target segments, while the corresponding support for entire customer base is only 0.17, making it a viable option for the telco to consider***
- ***Next viable itemset could be 'StreamingMovies', 'PhoneService', 'MultipleLine'***

Recommendations:
Summary of Proposed Marketing Strategies

Target Marketing

Approach 1 :

- Contract: Month to Month contract customers
- Tenure : 0 to 15 and more than 58
- Payment method: E- check payments

Approach 2:

- Cluster 0 most important, by churn { slide 18}
- Cluster 1 most important, by profitability {slide 19}

Promotional Bundle

- Multiple phone line with fibre optic, most pervasive across target segments
- Multiple phone line with streaming movies

Recommendations: Identifying Causality

- **Additional Sampling Required for Causality**

➢ Existing dataset collected insufficient variables to suggest causal factors

- Earlier recommendations and analysis based on existing dataset only provides prediction based on statistical correlation which can form hypotheses at best
- Require additional data collection (sampling) such as in the form of surveys required for hypothesis testing
- E.g. End-of service surveys could be done to collect data on reasons for churn
- E.g. To consider the causal factor for churn customers having phone lines, a survey sample of customer location and connectivity could be used to determine if remote locations having poor connectivity could be a related/causal factor
- E.g. Data on period of cancellation could be collected to determine if a concentration of high service disruptions or other events could also be a related/causal factor

- **Additional Statistical Analysis**

➢ With additional sample data, further data science techniques for hypotheses testing like Anova/T-test/Z-test/F-test can be employed to determine statistical significance of sampling evidence to support certain conclusions

Way Ahead:
Performance Measurement and Strategy Calibration

- **Performance Metrics**
  - Monitoring Customer Rating (1-2 weeks after the campaign)
  - Marketing ROI (At every quarter of the campaign timeline or after 2-4 weeks whichever later)
  - CAR(Customer Acquisition Rate) (4-8 weeks after the campaign)
  - Adoption rate of promotional bundles (6 months after implementation)

- **Response**
  - Decision can be made by Telco based on above performance metrics to calibrate the strategies and analysis
    - Positive result (E.g. growth in target customer segment): indicates that the marketing strategy likely successful, re-evaluate for new issues in subsequent descriptive analysis
    - Negative result (E.g. Shrink/stagnate in target customer segment): suggests potential gaps in the initial predictions & recommendations. Assuming sampling and causal factors have been resolved, re-evaluation of the descriptive and predictive analysis may be necessary to identify areas for improvement

Data governance in business and Bias

Issues	Solution
Data itself is biased	Regular inspection on the dataset to make sure the data collected is complete in terms of the business customer base.
Outdated data	Updating the data every 6 months or whenever there is a shift in the business plan as an internal data governance in the business.
Security	Implement robust security measures including encryption , access controls and server separation to prevent data theft from the company servers.
Privacy	Can be achieved by data minimization, anonymizing customer data, password protection and performing regular audits to prevent the data from being misused , since it can lead to financial losses and reputational damage in severe cases.
Ethics	Transparency in the collection and usage of data and obtaining informed consent from individuals

References

- ***https://tridenstechnology.com/telecommunications-industry-statistics/***
- ***https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/telecommunications-industry-outlook.html***