CHULA ENGINEERING
Innovation toward Sustainability | ACTNOW

INDUSTRIAL ENGINEERING
CHULALONGKORN UNIVERSITY

Customer Churn Prediction & Retention Strategies for Telecommunication Business

Presented by Group 23:

Thatchakarn	Chariyasethapong	6130239421
Teethavat	Techaphatipong	6130252521

"Business"

Obj:

(max)

Profit = Revenue – Cost

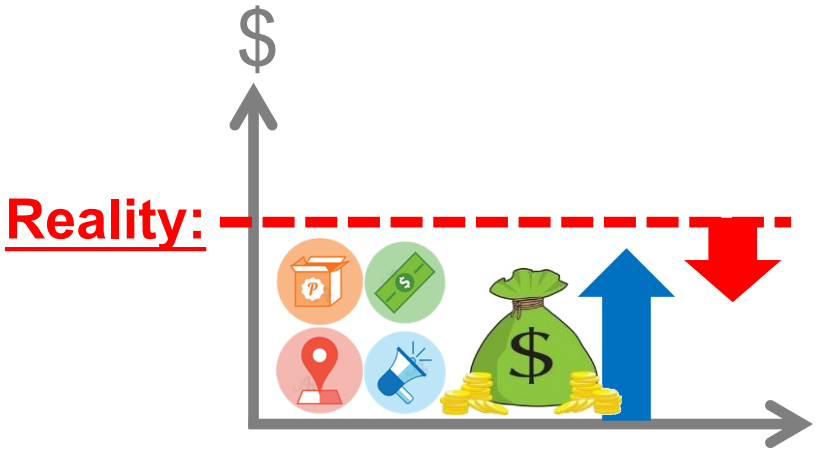
#Sales ⇧

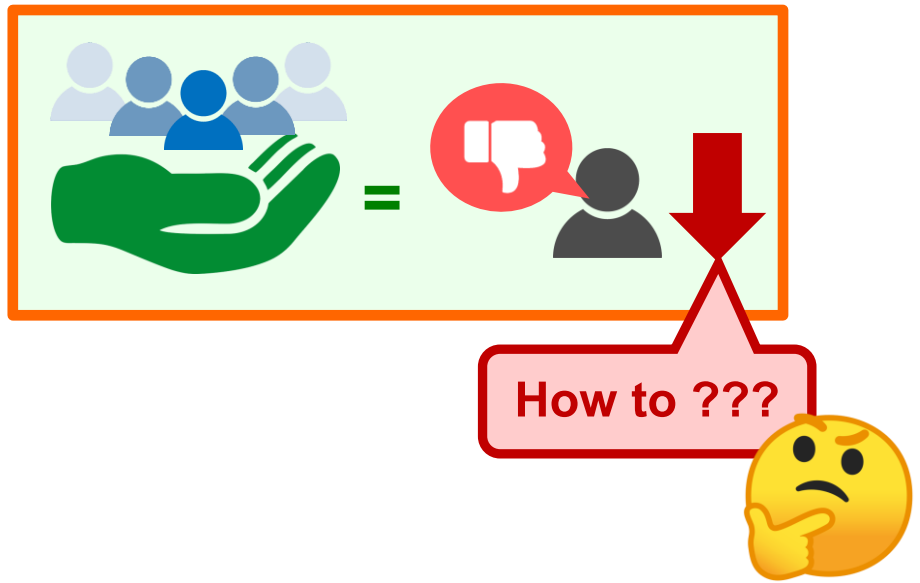
$

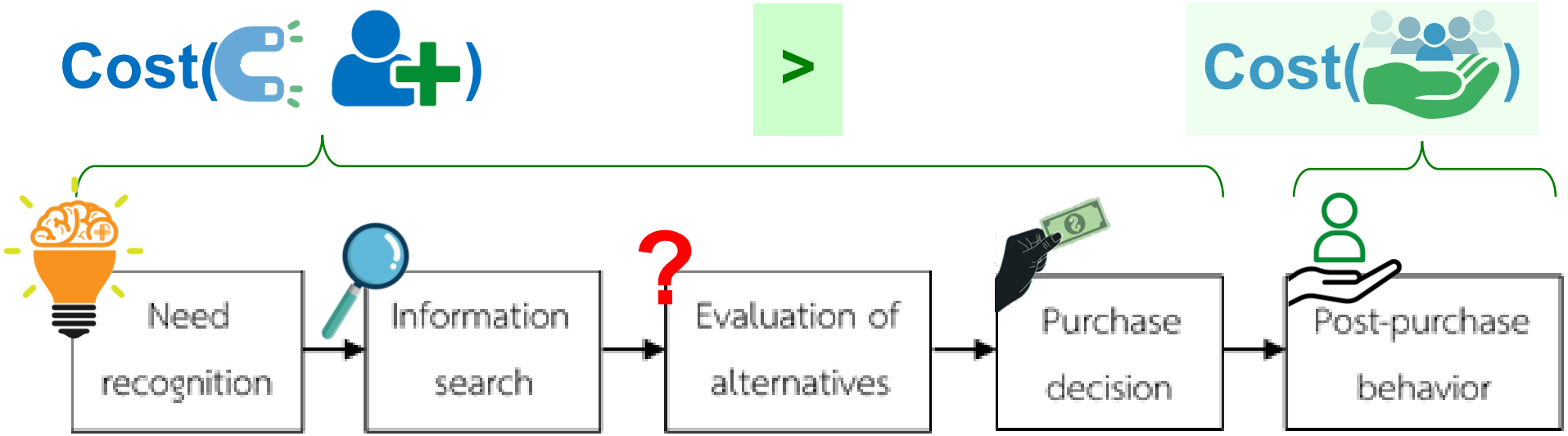
Marketing Department

Need recognition → Information search → ? Evaluation of alternatives → Purchase decision → Post-purchase behavior

Buyer Decision Process







Buyer Decision Process

Implementation of a Data Analytics Platform for a Telecom Company

Industry	Technologies	Business gains
Telecommunications	AWS, Python, Cloud	Reduced AWS computing resources by 80%

Customer

The Customer is a Texas-based telecom company participating in the federal Lifeline Support Program and providing pre-paid cell phones and service packages to low-income individuals.

Results

With ScienceSoft's big data services, the Customer was able to:

- Measure the engagement and identify the preferences of a particular user.
- Spot trends in the users' behavior.
- Make predictions about how users would behave.
- Invoice advertisers based on their calculated share.
- Benefit from insightful data analytics (for example, daily earnings, number of new users, customer service data and more).

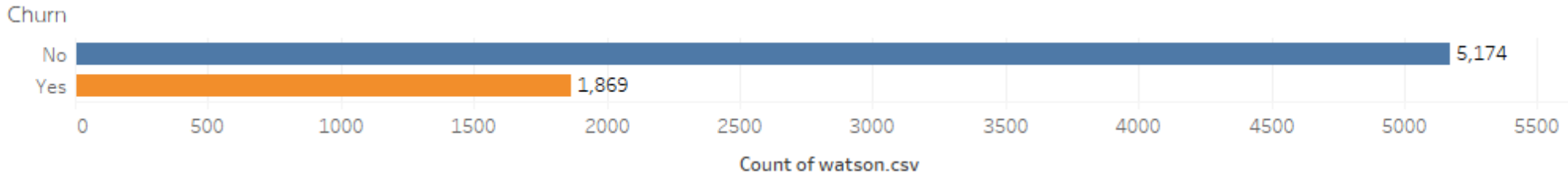
The use of Amazon Spot Instances allowed the Customer to reduce the costs of AWS computing resources by 80%.

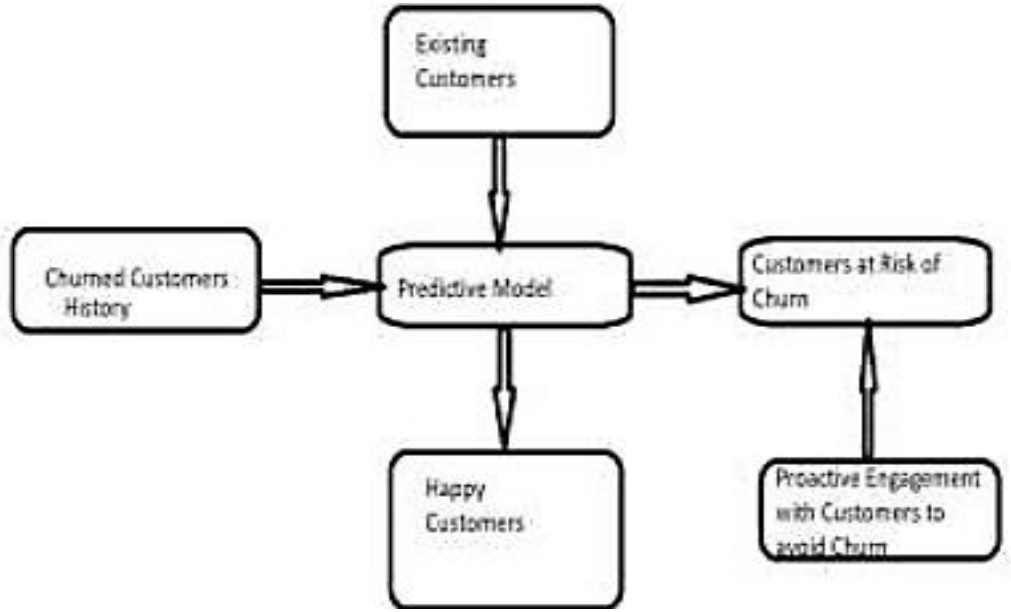
A Business Problem in Telecom company

Telecom Customer Churn Prediction

Abhijit Sundararajan
Department of MSIS, Rutgers University
E-mail: abhijit.sundararajan@rutgers.edu

Kemal Gursoy
Department of MSIS, Rutgers University,
E-mail: kgursoy@business.rutgers.edu





5. Comparison of Methods

Models	Accuracy Score
Random Forest Classifier	0.9355
SVM	0.8192
Decision Tree	0.762
Logistic Regression	0.7894

Scope & Objective

https://www.kaggle.com/blastchar/telco-customer-churn

Telco Customer Churn

Focused customer retention programs

BlastChar • updated 4 years ago (Version 1)

Data | Tasks (1) | Code (758) | Discussion (14) | Activity | Metadata | Download (978 kB) | New Notebook

Usability 8.8 | License Data files © Original Authors | Tags business

Description

Context

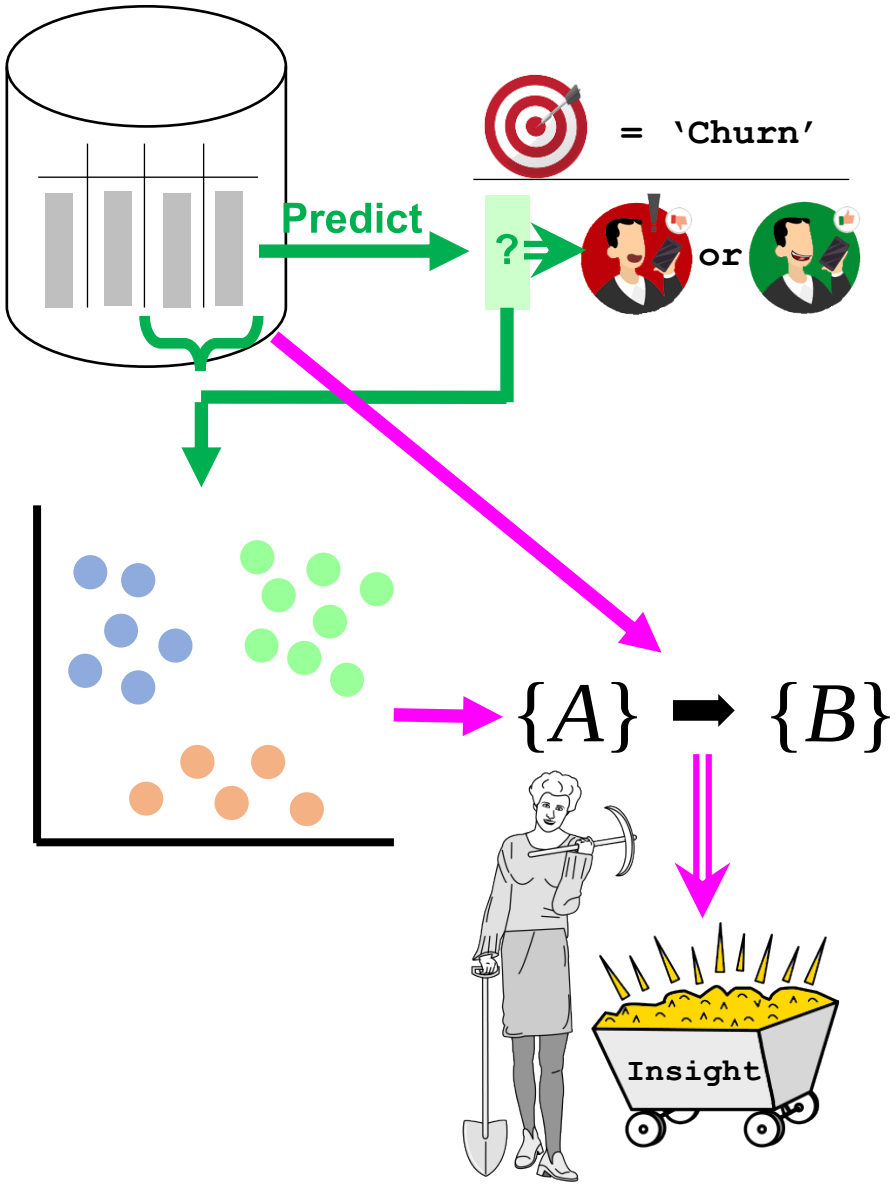
"Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs." [IBM Sample Data Sets]

Content

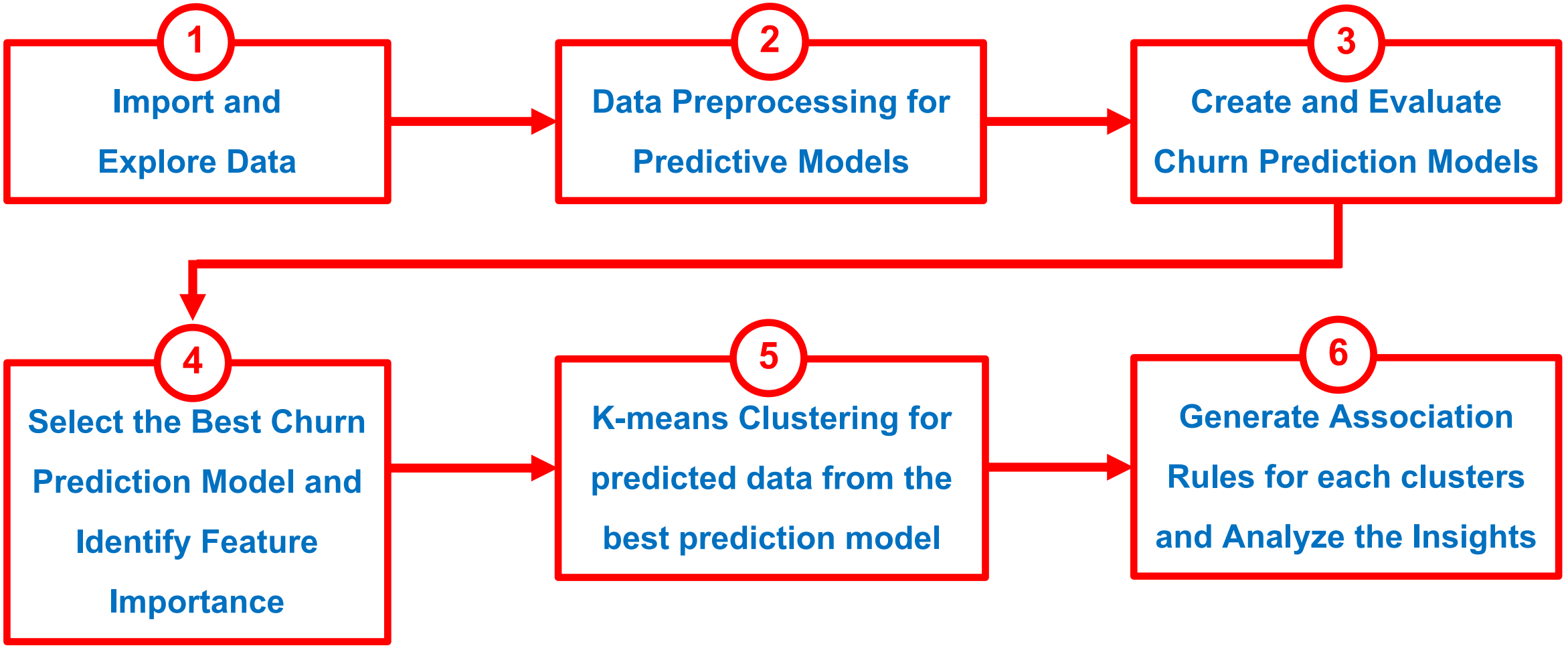
Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

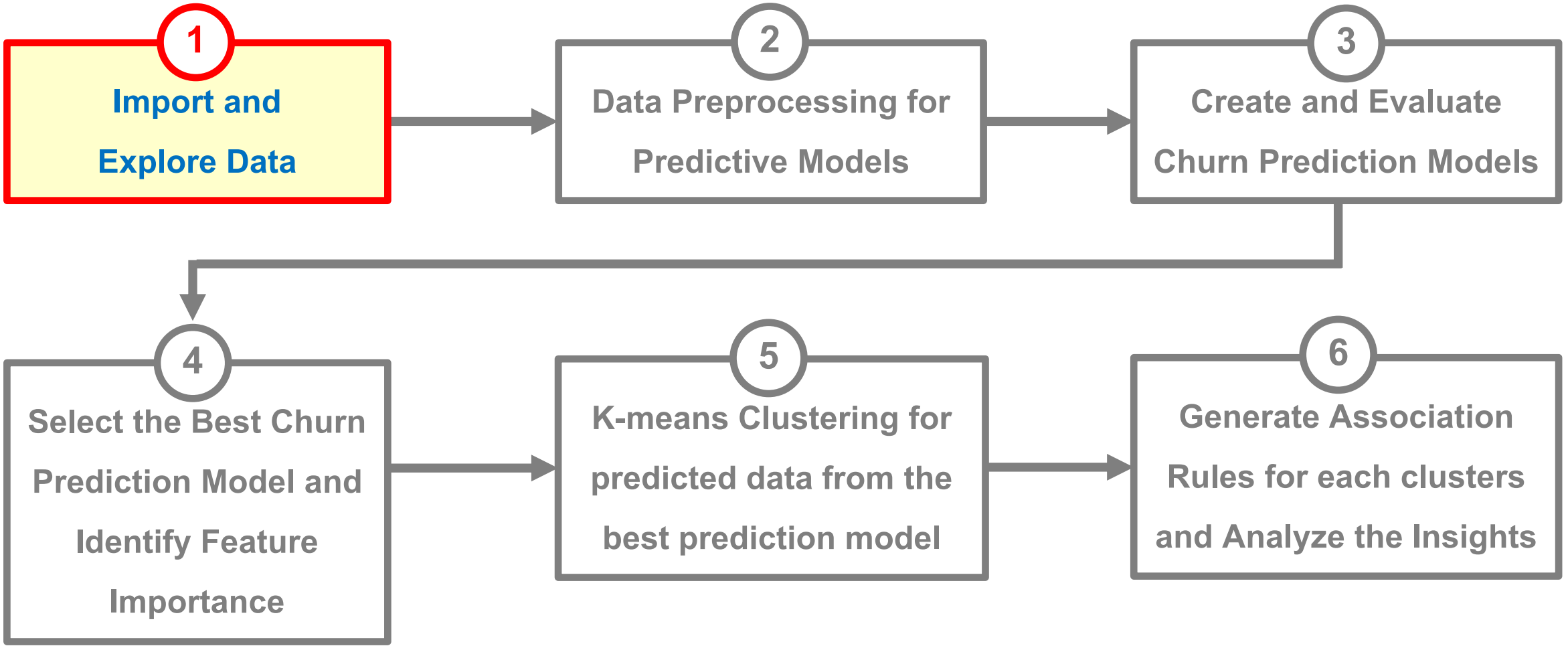
- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents



Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

1 Import and Explore Data

```
!gdown https://drive.google.com/uc?id=1qXvzqhXaclx6dqKAQKYXnHeUZEZffnLE
```

```
df = pd.read_csv('watson.csv', header=0, delimiter=",")
df
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	...	Churn
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	...	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	...	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	...	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	...	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	...	Yes
...	...	...	...	...	...	...	...	...	...	...	...
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	...	No
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	...	No
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	...	No
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	...	Yes
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	...	No

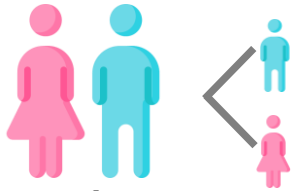
7043 rows × 21 columns

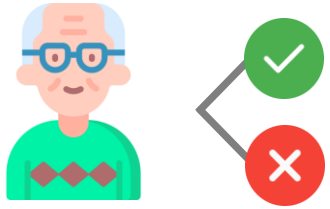
```
df.info()
```

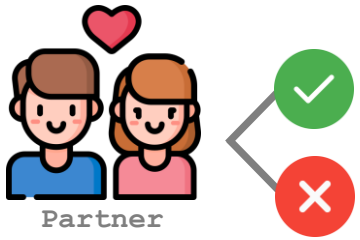
0s

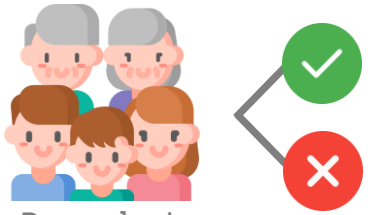
```
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7043 non-null   object 
 1   gender            7043 non-null   object 
 2   SeniorCitizen     7043 non-null   int64  
 3   Partner           7043 non-null   object 
 4   Dependents        7043 non-null   object 
 5   tenure            7043 non-null   int64  
 6   PhoneService      7043 non-null   object 
 7   MultipleLines     7043 non-null   object 
 8   InternetService   7043 non-null   object 
 9   OnlineSecurity    7043 non-null   object 
 10  OnlineBackup      7043 non-null   object 
 11  DeviceProtection  7043 non-null   object 
 12  TechSupport       7043 non-null   object 
 13  StreamingTV       7043 non-null   object 
 14  StreamingMovies   7043 non-null   object 
 15  Contract          7043 non-null   object 
 16  PaperlessBilling  7043 non-null   object 
 17  PaymentMethod     7043 non-null   object 
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object 
 20  Churn             7043 non-null   object 
dtypes: float64(1), int64(2), object(18)
```

Demographic

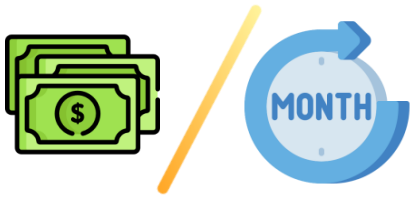

Gender


SeniorCitizen


Partner

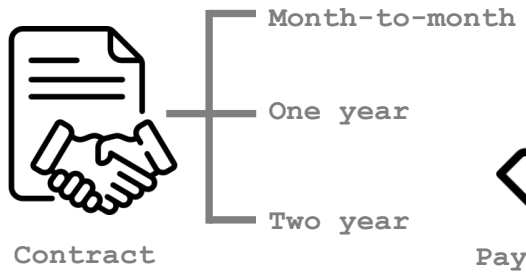

Dependents

Services


MonthlyCharges

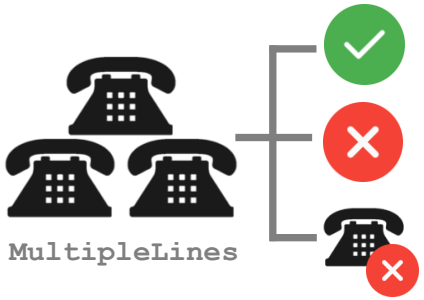

TotalCharges


Tenure

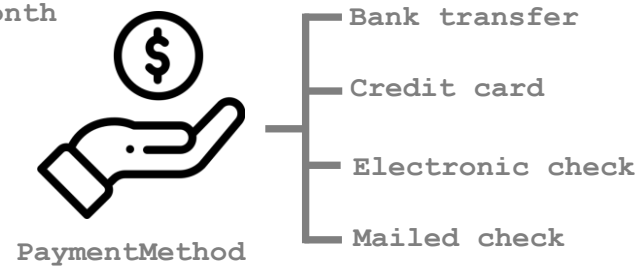


Contract

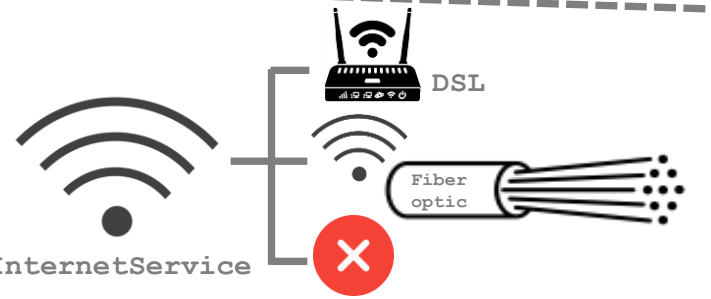

PhoneService


MultipleLines


PaperlessBilling



PaymentMethod



InternetService


OnlineSecurity

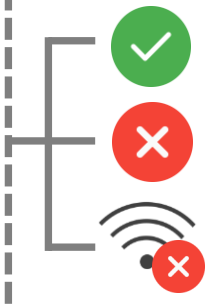

TechSupport


OnlineBackup


StreamingTV


DeviceProtection


StreamingMovies



customerID
7590-VHVEG
5575-GNVDE
3668-QPYBK
7795-CFOCW
9237-HQITU
...
6840-RESVB
2234-XADUH
4801-JZAZL
8361-LTMKD
3186-AJIEK

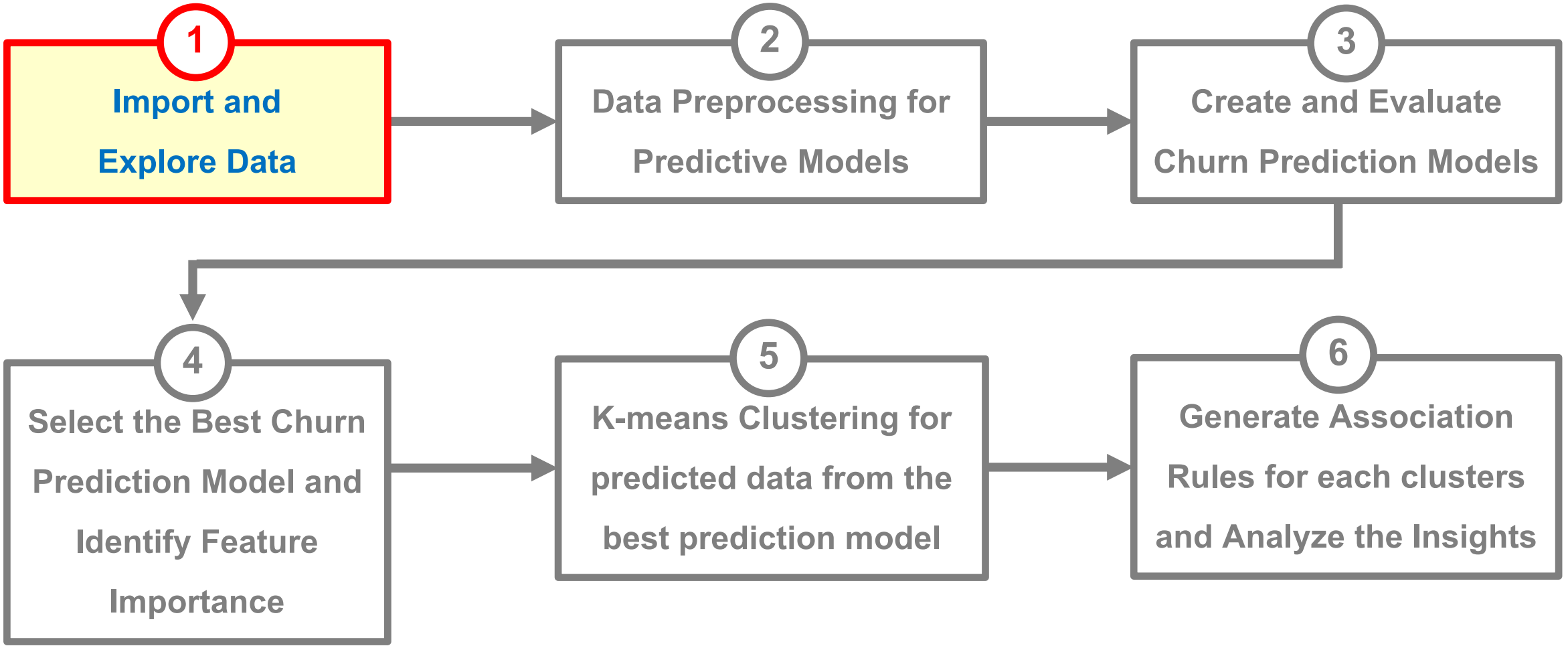
```
df.info()
```

0s

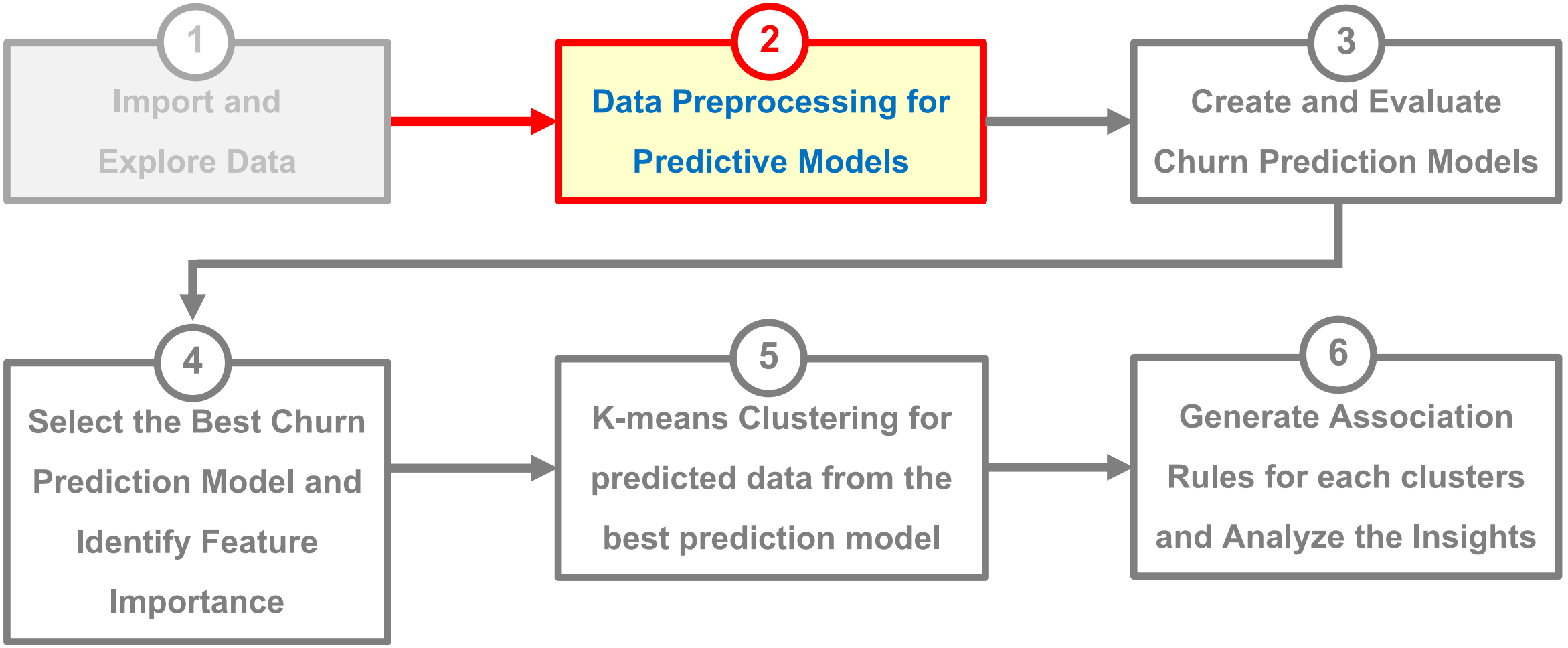
```
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
```

Continuous (float) ?

Procedure

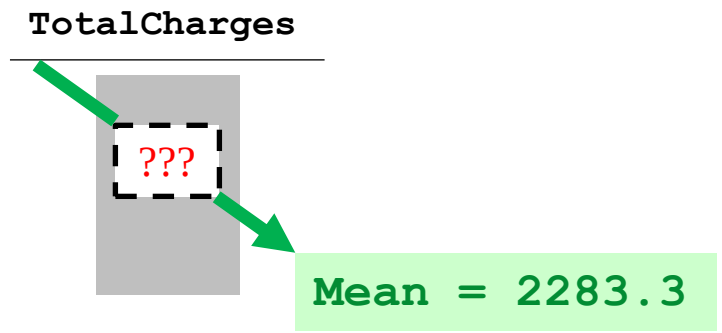

1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

2 Data Preprocessing for Predictive Models

Data Cleansing



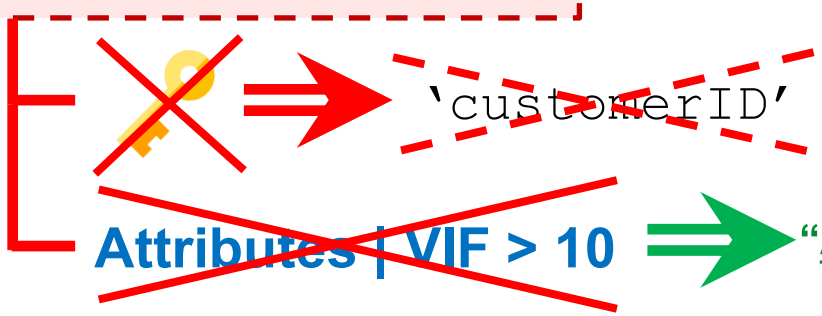
Data Transformation

- **One-hot encoding**

E.g. Contract = {Month-to-month, One year, Two year}

Contract	Contract_Month-to-month	Contract_One year	Contract_Two year
Month-to-month	1	0	0
One year	0	1	0
Two year	0	0	1

Attribute Selection



~~'customerID'~~

~~Attributes | VIF > 10~~ ⇒ "Absence of Multicollinearity between Independent Variables" ~ Logistics Regression Assumption

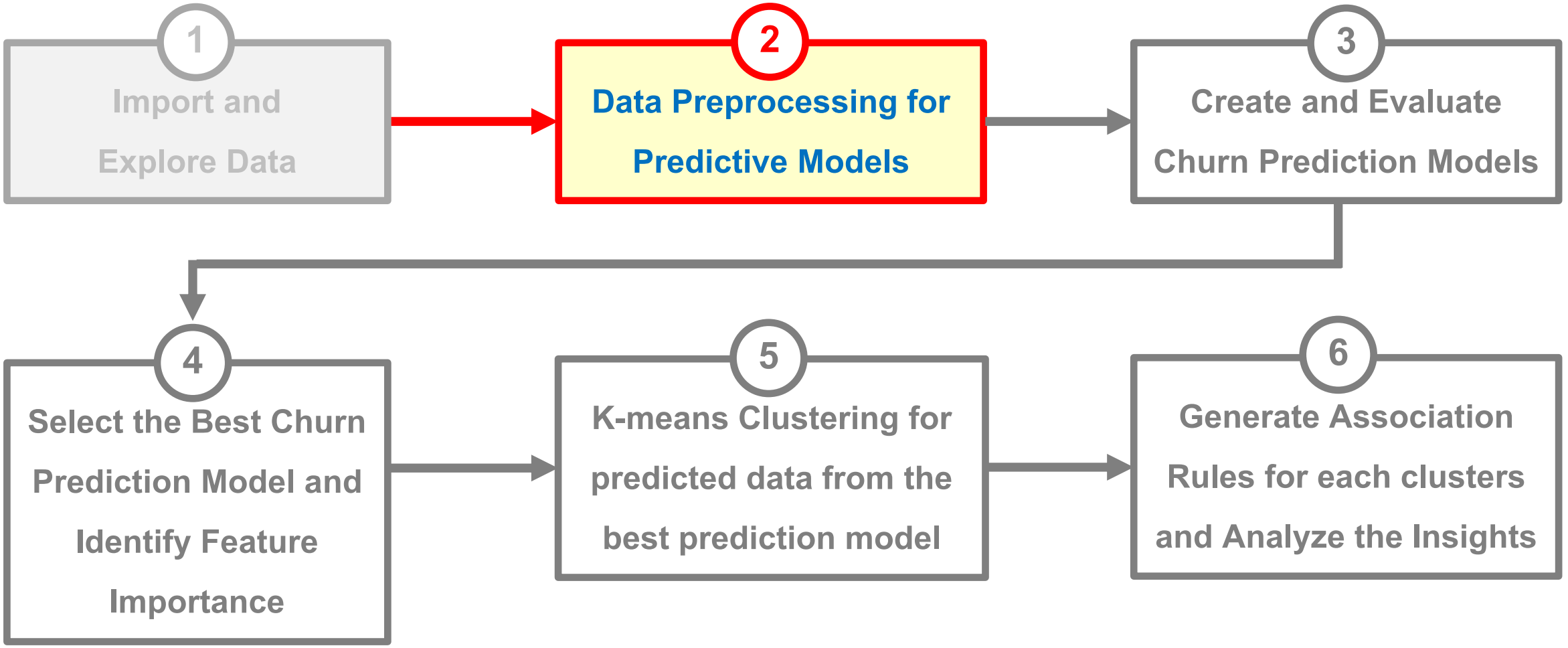
[Only for Logistics Regression Model]

```
attribute_collinear =
    ['MonthlyCharges', 'TotalCharges', 'gender_Female', 'gender_Male', 'Partner_No', 'Partner_Yes', 'Dependents_No',
    'Dependents_Yes', 'PhoneService_No', 'PhoneService_Yes','MultipleLines_No','MultipleLines_No phoneservice',
    'MultipleLines_Yes', 'InternetService_DSL', 'InternetService_Fiber optic', 'InternetService_No', 'OnlineSecurity_No',
    'OnlineSecurity_No internet service', 'OnlineSecurity_Yes', 'OnlineBackup_No', 'OnlineBackup_No internet service',
    'OnlineBackup_Yes', 'DeviceProtection_No', 'DeviceProtection_No internet service', 'DeviceProtection_Yes',
    'TechSupport_No', 'TechSupport_No internet service', 'TechSupport_Yes', 'StreamingTV_No',
    'StreamingTV_No internet service', 'StreamingTV_Yes', 'StreamingMovies_No', 'StreamingMovies_No internet service',
    'StreamingMovies_Yes', 'Contract_Month-to-month', 'Contract_One year', 'Contract_Two year', 'PaperlessBilling_No',
    'PaperlessBilling_Yes', 'PaymentMethod_Bank transfer (automatic)', 'PaymentMethod_Credit card (automatic)',
    'PaymentMethod_Electronic check', 'PaymentMethod_Mailed check']
```

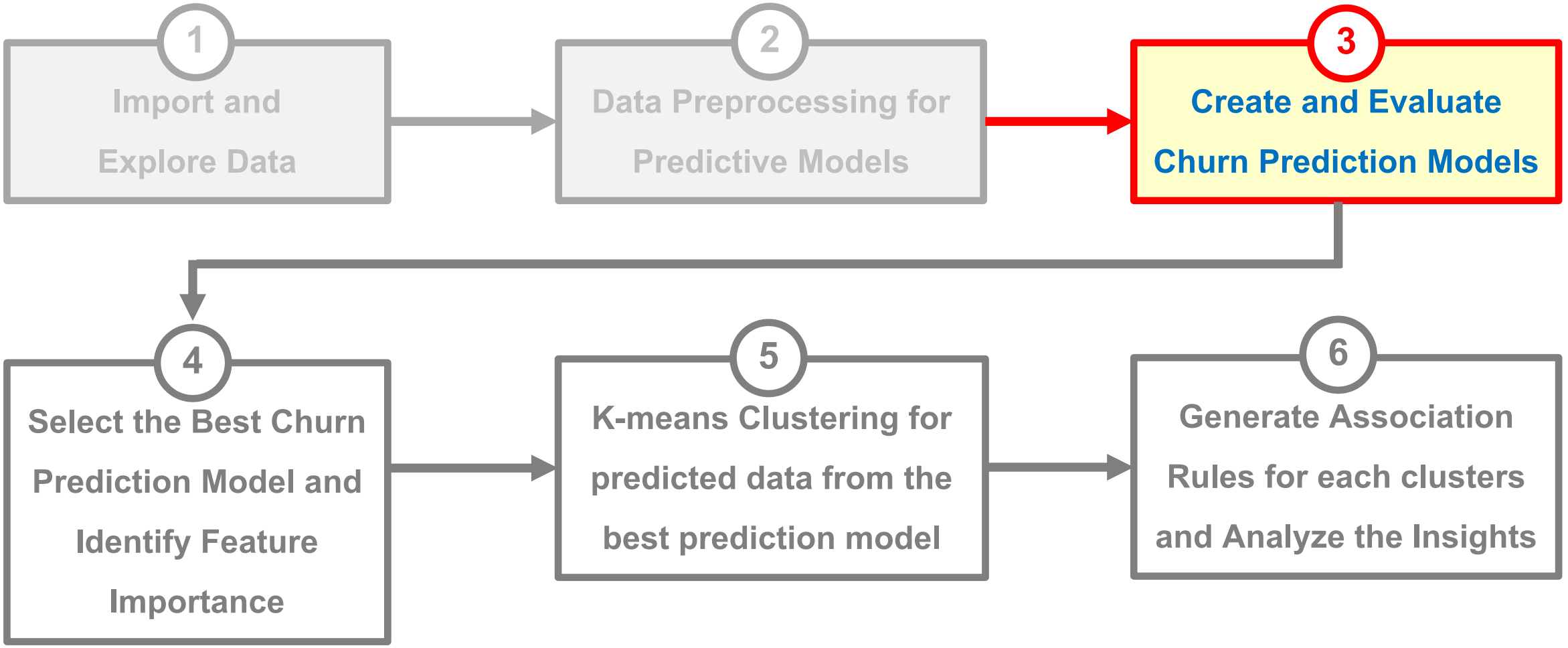
```
len(attribute_collinear)
```

```
43
```

Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

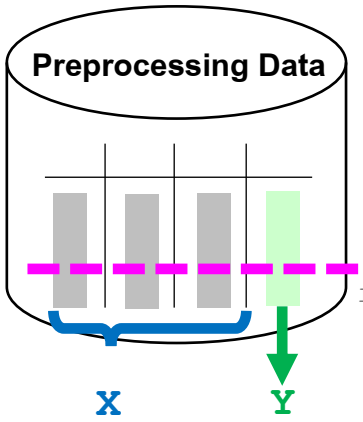
Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

3 Create and Evaluate Churn Prediction Models

Hyperparameter Tuning with

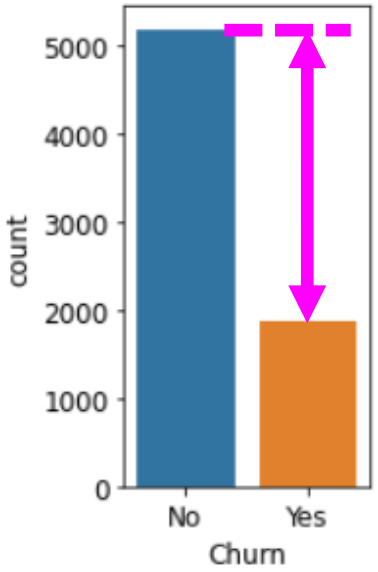
```
.GridSearchCV(estimator, param_grid,
scoring = f1_score, n_jobs = -1, cv =10)
```



Stratified Sampling

```
random_state = 2565
```

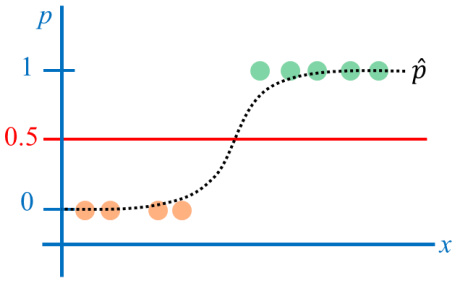
(Input)(Target)

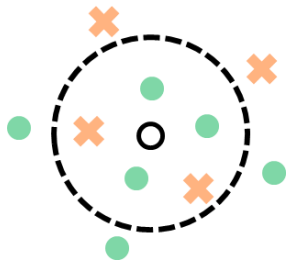


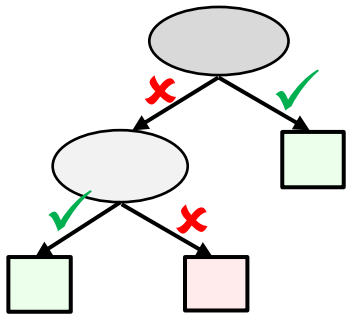
80% Training set

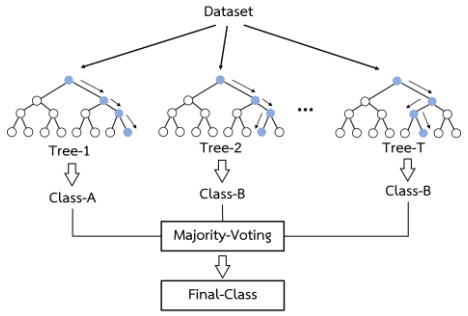
20% Testing set

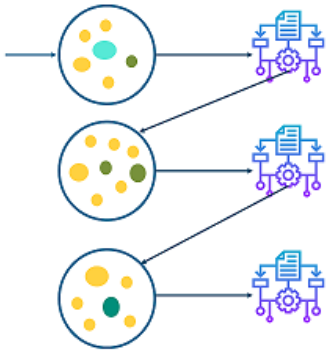
Prediction Model



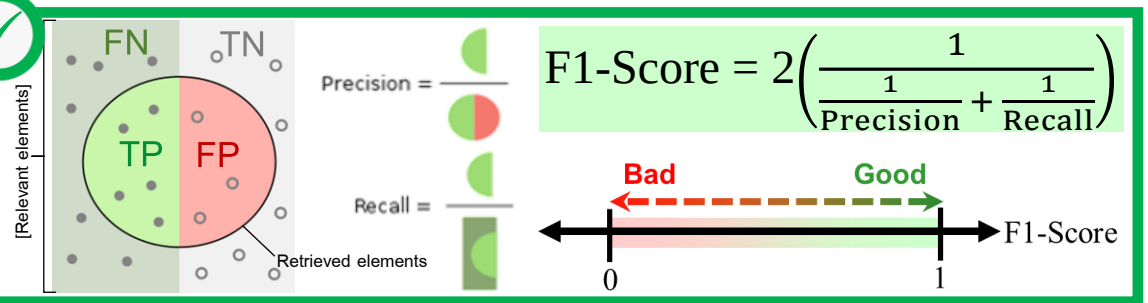








Model



F1-Score

Source: https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f

No	Actual	Predicted	Match
1	Airplane	Airplane	✓
2	Car	Boat	X
3	Car	Car	✓
4	Car	Car	✓
5	Car	Boat	X
6	Airplane	Boat	X
7	Boat	Boat	✓
8	Car	Airplane	X
9	Airplane	Airplane	✓
10	Car	Car	✓

		Predicted: Airplane	Predicted: Boat	Predicted: Car
Actual	Airplane	2	1	0
Actual	Boat	0	1	0
Actual	Car	1	2	3

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)
Airplane	2	1	1
Boat	1	3	0
Car	3	0	3

	precision	recall	f1-score	support
Aeroplane	0.67	0.67	0.67	3
Boat	0.25	1.00	0.40	1
Car	1.00	0.50	0.67	6
accuracy			0.60	10
macro avg	0.64	0.72	0.58	10
weighted avg	0.82	0.60	0.64	10

Per-Class F1 scores

Average F1 scores

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

Macro-Averaged F1 Score

$$\frac{0.67 + 0.40 + 0.67}{3} = \mathbf{0.58}$$

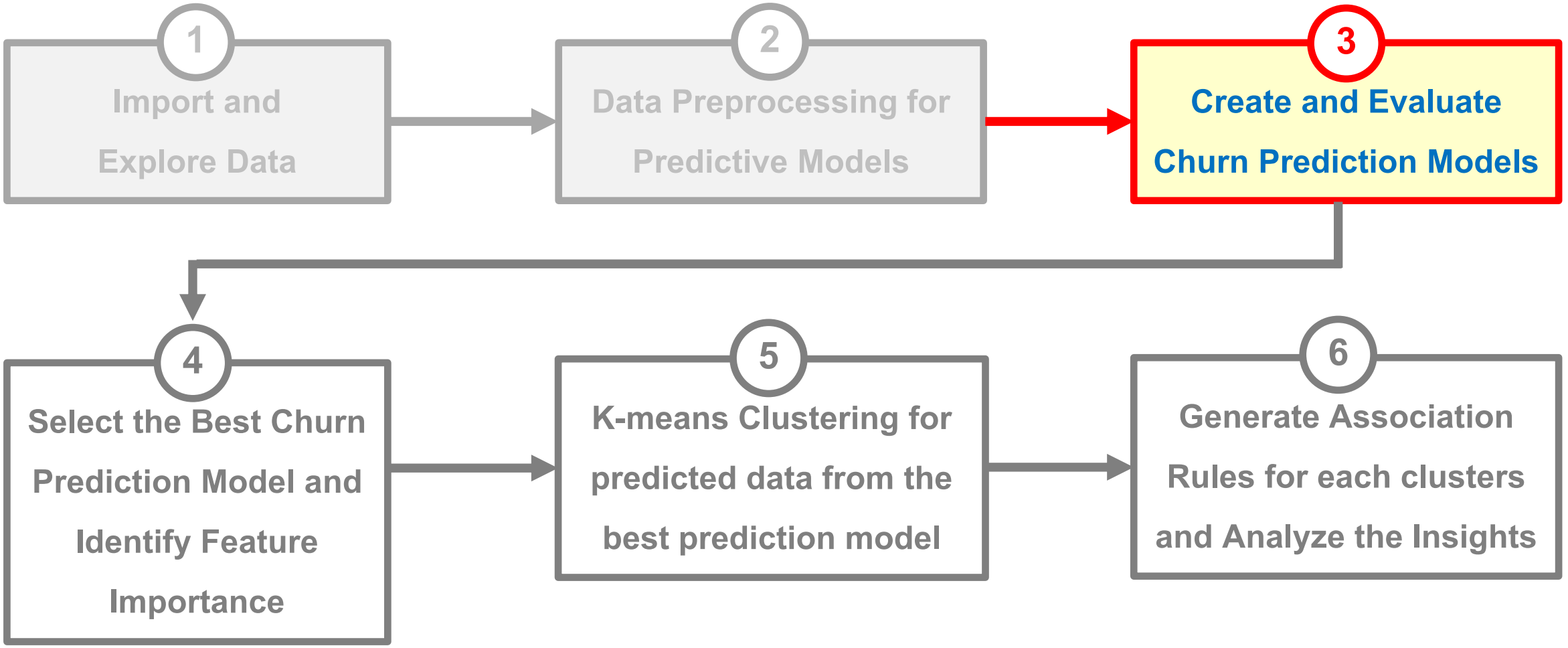
Weighted Average F1 Score

$$(0.67 * 0.3) + (0.40 * 0.1) + (0.67 * 0.6) = \mathbf{0.64}$$

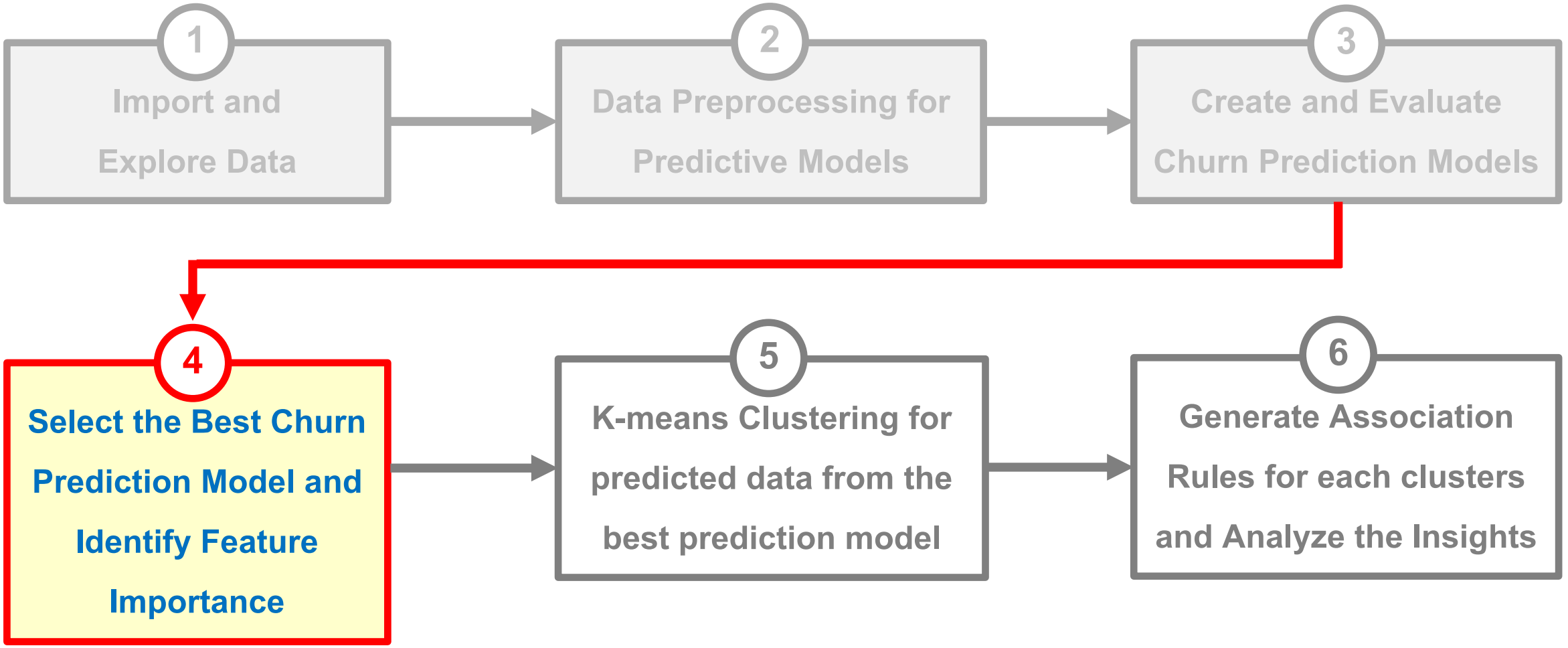
Micro-Averaged F1 Score

$$\frac{TP}{TP + \frac{1}{2}(FP+FN)} = \frac{6}{6 + \frac{1}{2}(4+4)} = \mathbf{0.60}$$

Procedure

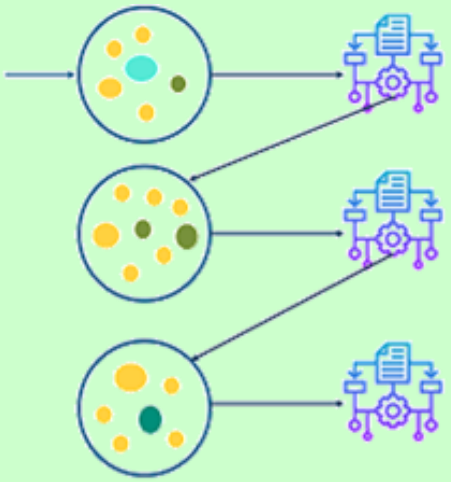

1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

4 Select the Best Churn Prediction Model and Identify Feature Importance

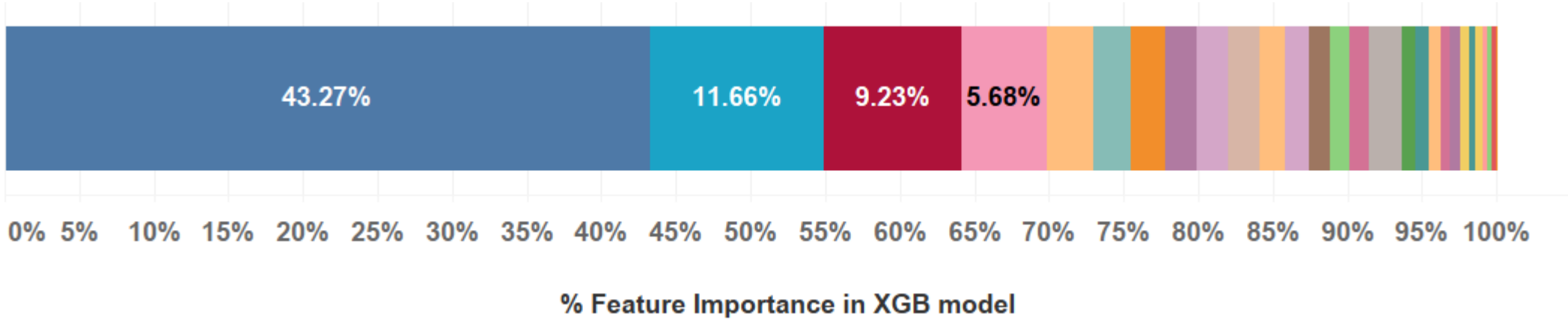
The Best Churn Prediction Model =



XGBoost

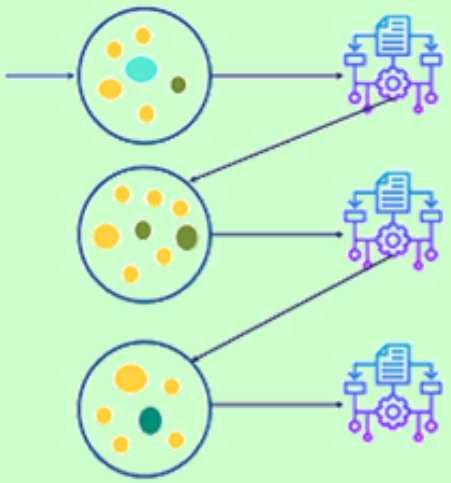
```
XGBClassifier(max_depth = 6,
              learning_rate = 0.3,
              subsample=1.0,
              tree_method='auto')
```

F1-score(macro) = 72.657%



4 Select the Best Churn Prediction Model and Identify Feature Importance

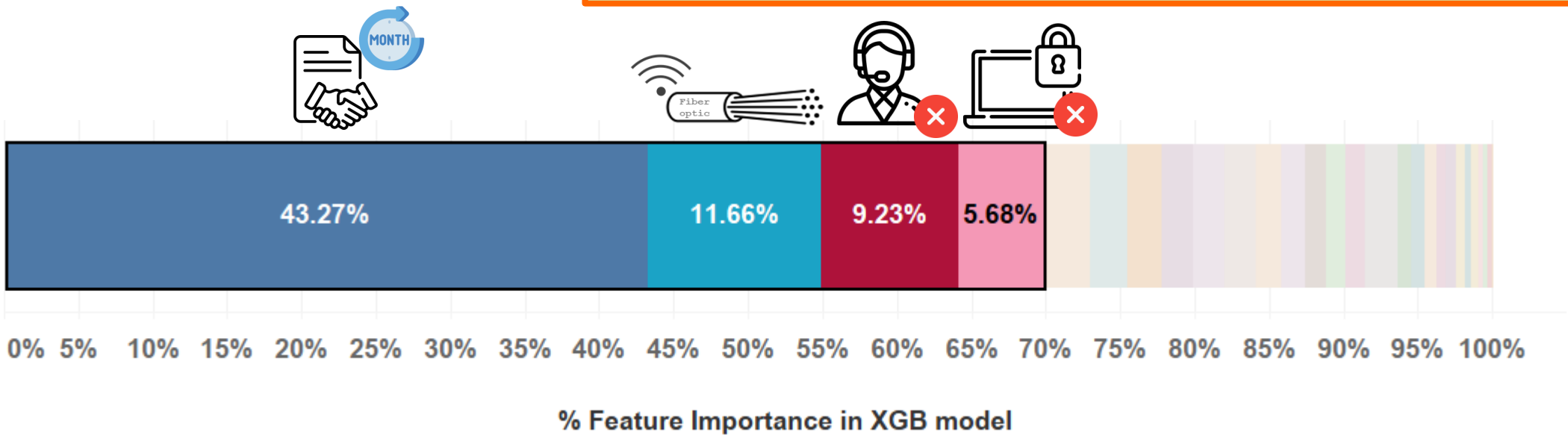
The Best Churn Prediction Model =



XGBoost

```
XGBClassifier(max_depth = 6,
              learning_rate = 0.3,
              subsample=1.0,
              tree_method='auto')
```

F1-score(macro) = 72.657%



Comparing Prediction Model Result

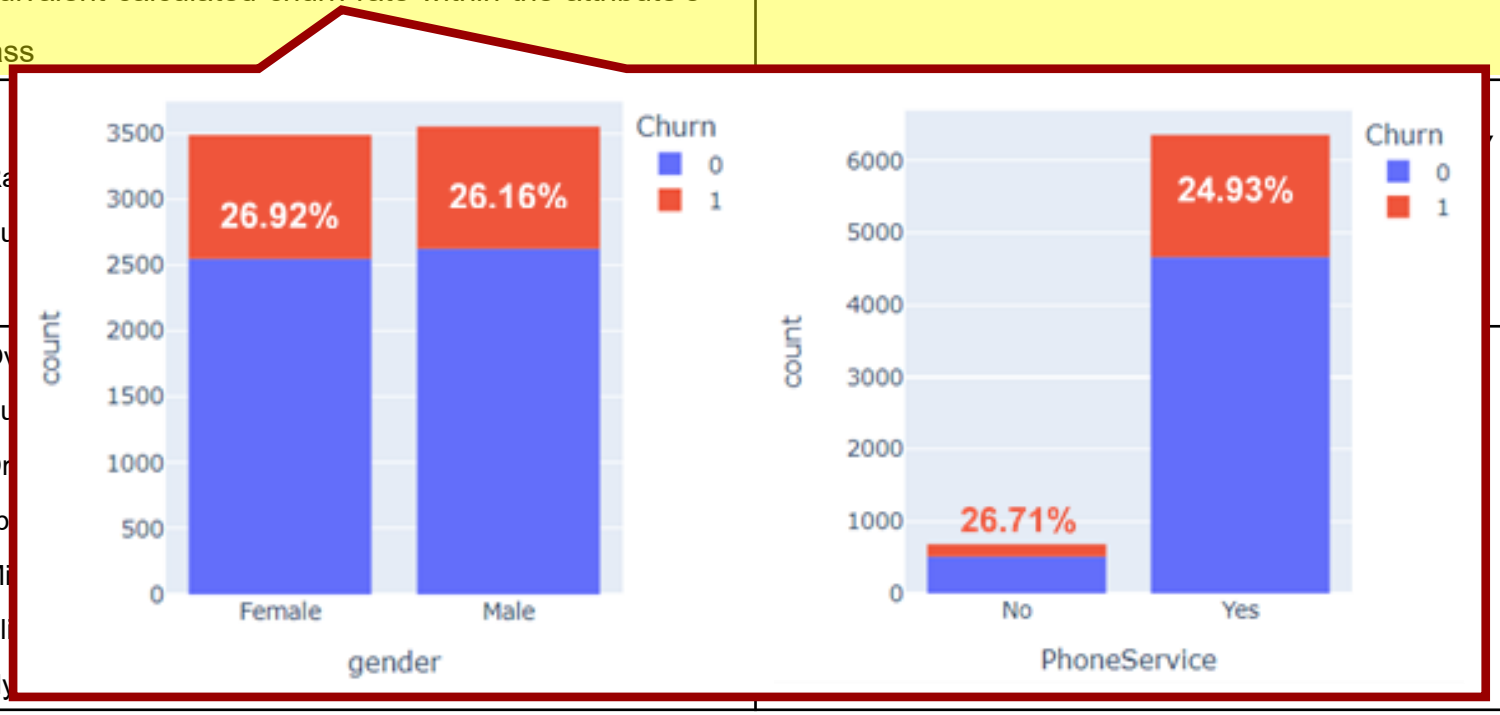
Prediction Model	Fine-tuned hyperparameter(s)	Total tuned combinations	Total runtime [seconds]	Average runtime per combination [seconds/combination]	Testing set performance	
					Accuracy	F1-score (macro)
Logistics Regression	`C=0.01, penalty='none', solver='newton-cg'`	100	47.486	0.475	74.521%	54.852%
Decision Tree	`max_depth=6, min_samples_split=40`	585	157.723	0.270	78.070%	71.359%
Random Forest	`max_depth=10, min_samples_split=10, n_estimators=1000, max_features = 'auto'`	480	4425.416	9.220	79.631%	71.396%
XGBoost	`max_depth = 6, learning_rate = 0.3, subsample=1.0, tree_method='auto'`	180	1906.808	10.593	80.199%	72.657%
K-Nearest Neighbors	`n_neighbors=21, weights='distance'`	100	120.616	1.206	77.573%	67.478%
Support Vector Machine	`'kernel'='rbf', C=100, gamma=0.0001`	64	910.459	14.226	78.211%	69.788%

Comparing Prediction Model with Kaggle user's

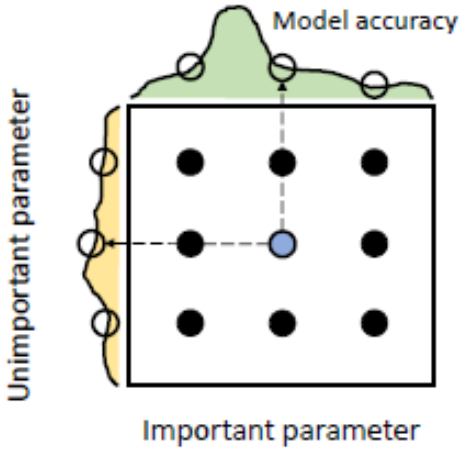
Rank	Title	Method	Algorithm	Author	Accuracy	F1-score
1	XGBoost & LightGBM & Catboost - Imbalanced Data	- Manual hyperparameters tuning - Transform the data "No internet service" to "No" in the attribute, 'TechSupport', 'StreamingTV', 'StreamingMovies', 'OnlineSecurity', 'OnlineBackup', and 'DeviceProtection' - Drop the attributes, 'Gender' and 'PhoneService' from equivalent calculated churn rate within the attribute's class	`XGBClassifier(random_state=0)`	 Kaan Boke	81.62%	75.22%
2	XGBoost Tuned With Random Search	- Random Search with `n_iter=10` - Substitute missing data in 'TotalCharges' with 0	`XGBClassifier(colsample_bytree=0.6, gamma=2, learning_rate=0.02, max_depth=5,min_child_weight=10, n_estimators=200, nthread=-1, silent=True,subsample=0.6)`	 Graeme Keleher	81.41%	74.81%
3	Customer Churn With Oversampling Techniques	- Oversampling with BorderlineSMOTE - Substitute missing data in 'TotalCharges' with median - Drop the attribute 'CustomerID' and then eliminate the duplicate datasets - Min-max normalization - Eliminate the attribute which have VIF > 10 - Hyperparameter tuning with Random search	`XGBClassifier(learning_rate_init= 0.001, max_depth= 6, n_estimators= 300, n_jobs=-1, random_state = 0, min_child_weight= 2, colsample_bytree= 0.8, gamma= 2)`	 Arezoo Dahesh	78.22%	73.14%

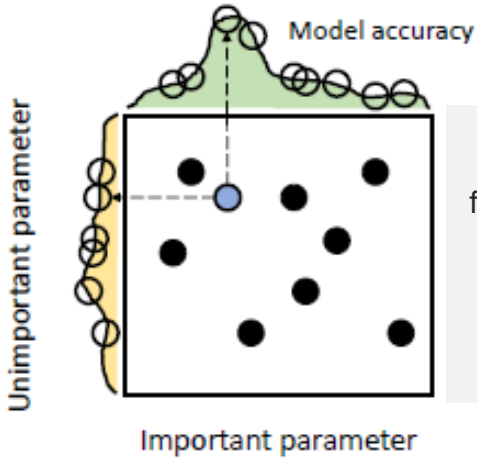
Comparing Prediction Model with Kaggle user's

Rank	Title	Method	Algorithm	Author	Accuracy	F1-score
1	XGBoost & LightGBM & Catboost - Imbalanced Data	- Manual hyperparameters tuning - Transform the data "No internet service" to "No" in the attribute, 'TechSupport', 'StreamingTV', 'StreamingMovies', 'OnlineSecurity', 'OnlineBackup', and 'DeviceProtection' - Drop the attributes, 'Gender' and 'PhoneService' from equivalent calculated churn rate within the attribute's class	XGBClassifier(random_state=0)	Kaan Boke	81.62%	75.22%
2	XGBoost Tuned With Random Search	- Ra - Su		Graeme Keleher	81.41%	74.81%
3	Customer Churn With Oversampling Techniques	- Ov - Su - Dr dup - Mi - El - Hy		Arezoo Dahesh	78.22%	73.14%



Comparing Predi





sampled with fixed numbers of iterations given by `n_iter`

- Find absolute best way for tuned hyperparameter
- Use exhaustive search (Full range of possibilities)
  = Time consuming + Chance for overfitting
- Search randomly selected combination
  = Faster run time + P(Overfitting) ⇩
- Potential for variance btw run ∵ Random

Rank	Title				Accuracy	F1-score

1	XGBoost & LightGBM & Catboost - Imbalanced Data	- Manual hyperpara
		- Transform the dat
		attribute, `'TechSu`
		`'StreamingMovi`
		`'OnlineBackup`
		- Drop the attri
		equivalent
		class				

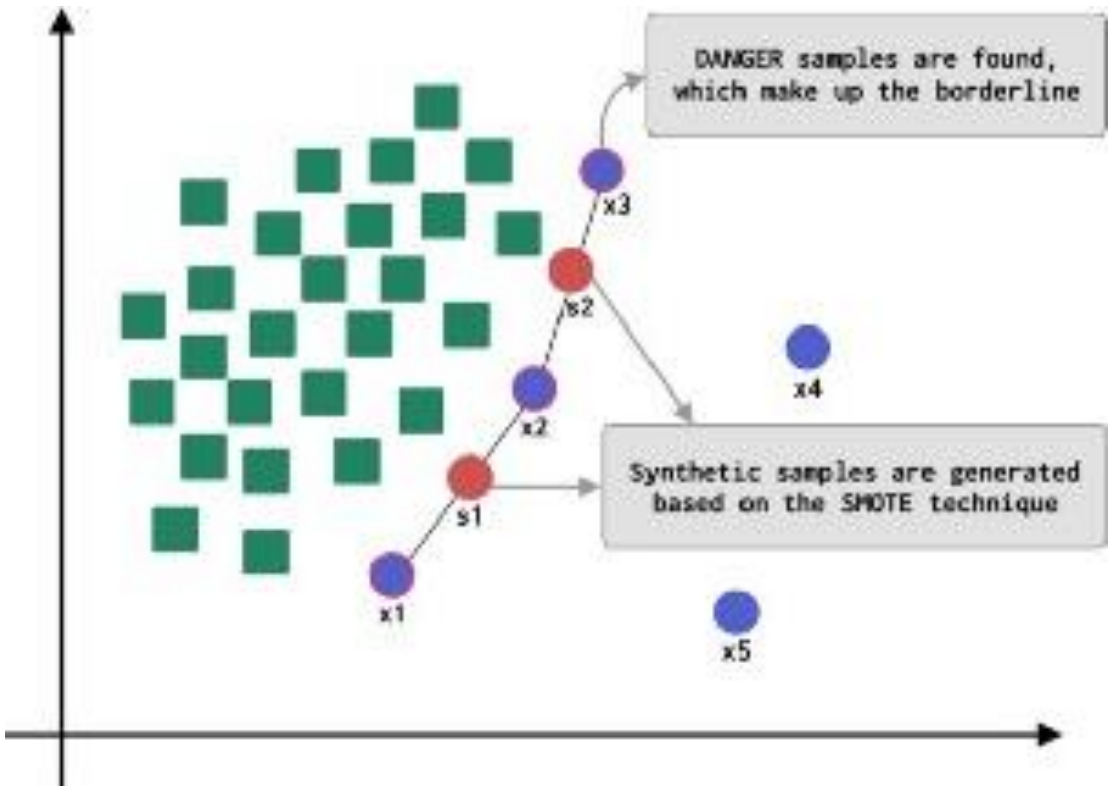
2	XGBoost Tuned With Random Search	- Random Search with `n_iter=10` - Substitute missing data in `'TotalCharges'` with 0	`XGBClassifier(colsample_bytree=0.6, gamma=2, learning_rate=0.02, max_depth=5,min_child_weight=10, n_estimators=200, nthread=-1, silent=True,subsample=0.6)`	 Graeme Keleher	81.41%	74.81%
3	Customer Churn With Oversampling Techniques	- Oversampling with BorderlineSMOTE - Substitute missing data in `'TotalCharges'` with median - Drop the attribute `'CustomerID'` and then eliminate the duplicate datasets - Min-max normalization - Eliminate the attribute which have VIF > 10 - Hyperparameter tuning with Random search	`XGBClassifier(learning_rate_init= 0.001, max_depth= 6, n_estimators= 300, n_jobs=-1, random_state = 0, min_child_weight= 2, colsample_bytree= 0.8, gamma= 2)`	 Arezoo Dahesh	78.22%	73.14%

Source:
https://betterprogramming.pub/comparing-grid-and-randomized-search-methods-in-python-cd9fe9c3572d

Comparing Predict

Rank	Title					
1	XGBoost & LightGBM & Catboost - Imbalanced Data	- Manual hyperparame - Transform the data attribute, 'TechSuppo - Drop the attributes equivalent calculate class				
2	XGBoost Tuned With Random Search	- Random Search with - Substitute missing da	rue,subsample=0.6)			
3	Customer Churn With Oversampling Techniques	- Oversampling with BorderlineSMOTE - Substitute missing data in 'TotalCharges' with median - Drop the attribute 'CustomerID' and then eliminate the duplicate datasets - Min-max normalization - Eliminate the attribute which have VIF > 10 - Hyperparameter tuning with Random search	XGBClassifier(learning_rate_init= 0.001, max_depth= 6, n_estimators= 300, n_jobs=-1, random_state = 0, min_child_weight= 2, colsample_bytree= 0.8, gamma= 2)	Arezoo Dahesh	78.22%	73.14%

Source: https://towardsdatascience.com/smote-synthetic-data-augmentation-for-tabular-data-1ce28090debc



Generate synthetic data by considering only samples that make up the border that divides one class from another

- Get higher accuracy than SMOTE based on border decision regions
- Tend to evade some important (extreme) examples in the minority class → bias in model

(Nagarajan, 2017, p.9)

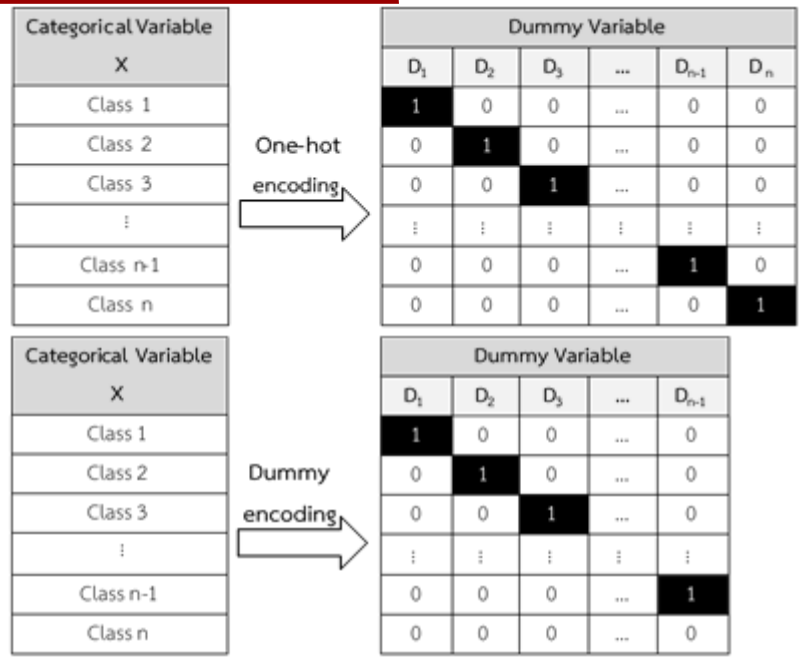
Comparing Prediction Model with Kaggle users'

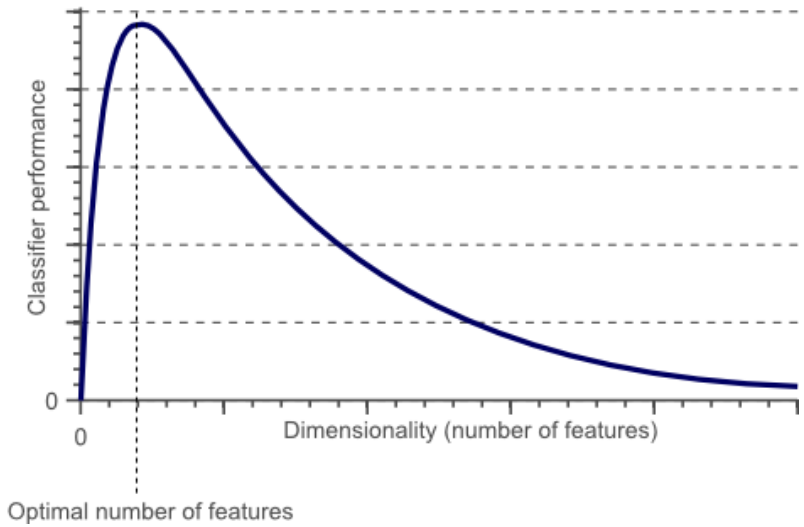
Rank	Title	Method	Algorithm	Author	Accuracy	F1-score
4	Senior Project	- One-hot encoding for categorical attributes - Hyperparameter tuning with `GridSearchCV()` - Substitute missing data in `'TotalCharges'` with mean	`XGBClassifier(max_depth = 6, learning_rate = 0.3, subsample=0.5, tree_method='hist')`	Thatchakarn, Teethavat	79.77%	72.46%
5	Telco Customer Churn Prediction	- Substitute missing data in `'TotalCharges'` with mean - Dummy encoding for categorical attributes	`RandomForestClassifier()`	 Ichıng Wang	79.13%	71.79%

6	Customer Churn Prediction Using ANN	- Eliminate the datasets having a missing value
		- Transform class `"No internet service"` to `"No"` in the attribute, `'TechSupport'`, `'StreamingTV'`, `'StreamingMovies'`, `'OnlineSecurity'`, `'OnlineBackup'`, and `'DeviceProtection'`
		- Transform class `"No Phoneservice"` to `"No"` in the attribute, `'MultipleLines'`
		- Min-max normalization
			`model = keras.Sequential([ keras.layers.Dense(27, input_shape=(27,), activation='relu'), keras.layers.Dense(15, activation='relu'), keras.layers.Dense(1, activation='sigmoid')])`
			
			`model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])`
			
			`model.fit(X_train, y_train, epochs=100)`
				 Jatin	79.25%	70.98%

Comparing Prediction Model with Kaggle users'

Rank	Title	Method	Algorithm	Author	Accuracy	F1-score
4	Senior Project	- One-hot encoding for categorical attributes - Hyperparameter tuning with `GridSearchCV()` - Substitute missing data in `'TotalCharges'` with mean	`XGBClassifier(max_depth = 6, learning_rate = 0.3, subsample=0.5, tree_method='hist')`	Thatchakarn, Teethavat	79.77%	72.46%
5	Telco Customer Churn Prediction	- Substitute missing data in `'TotalCharges'` with mean - Dummy encoding for categorical attributes	`RandomForestClassifier()`	 Iching Wang	79.13%	71.79%





Curse of Dimensionality

"As the dimensionality increases, the classifier's performance increases until the optimal number of features is reached. Further increasing the dimensionality without increasing the number of training samples results in a decrease in classifier performance."

Comparing

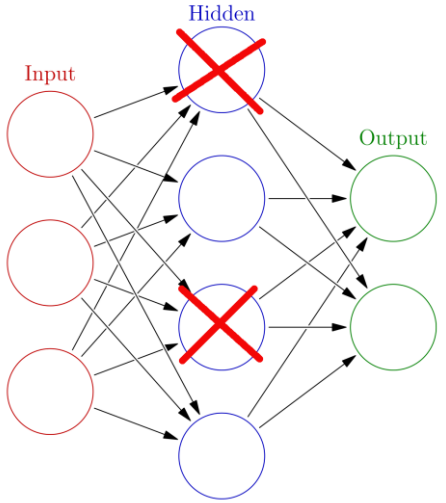
No regularization = Lack of generalization = Tend to overfit ⇧

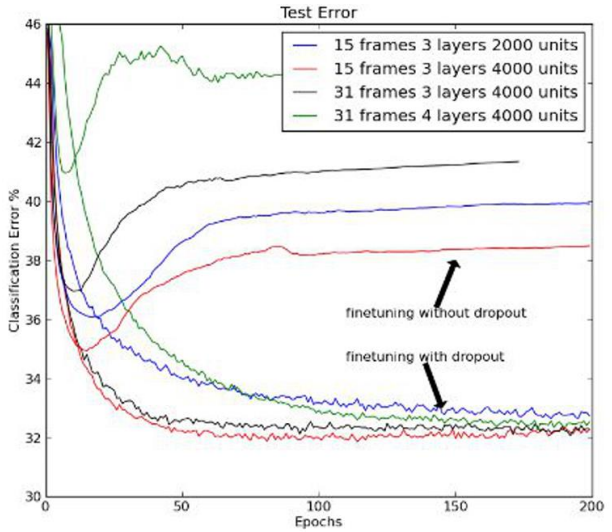
L1 regularization: Regularizing term is a sum

- **Original loss** + $C\Sigma|w_i|$

L2 regularization: Regularizing term is a sum of squares

- **Original loss** + $C\Sigma w_i^2$



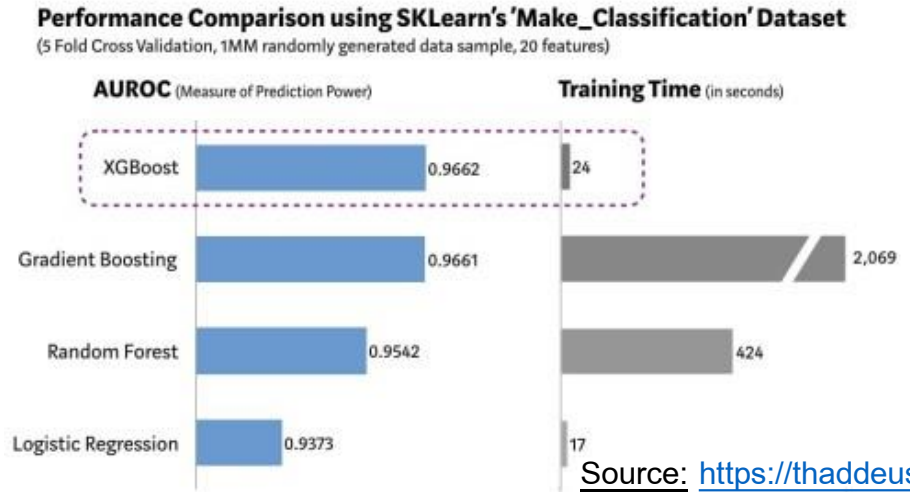


Rank	Title					
4	Senior Project	- - -				
5	Telco Customer Churn Prediction	- S - D				

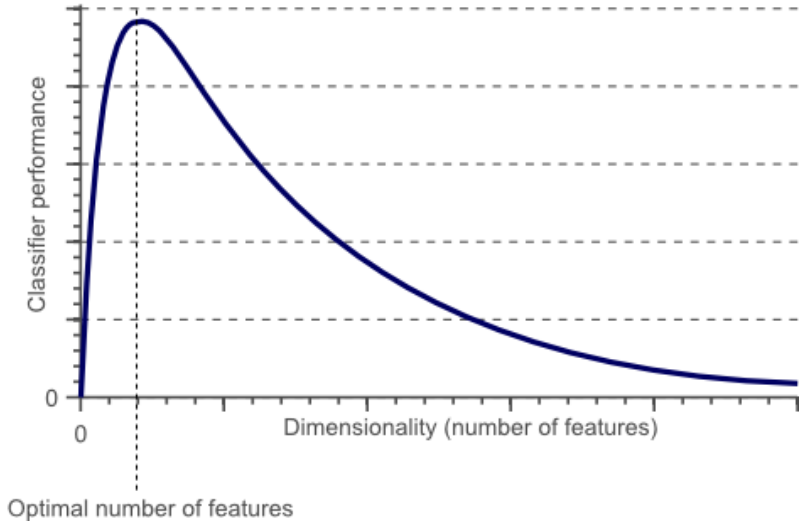
6	Customer Churn Prediction Using ANN	- Eliminate the datasets having a missing value
		- Transform class "No internet service" to "No" in the attribute, 'TechSupport', 'StreamingTV', 'StreamingMovies', 'OnlineSecurity', 'OnlineBackup', and 'DeviceProtection'
		- Transform class "No Phoneservice" to "No" in the attribute, 'MultipleLines'
		- Min-max normalization
			`model = keras.Sequential([ keras.layers.Dense(27, input_shape=(27,), activation='relu'), keras.layers.Dense(15, activation='relu'), keras.layers.Dense(1, activation='sigmoid')])`
			
			`model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])`
			
			`model.fit(X_train, y_train, epochs=100)`
				 Jatin	79.25%	70.98%

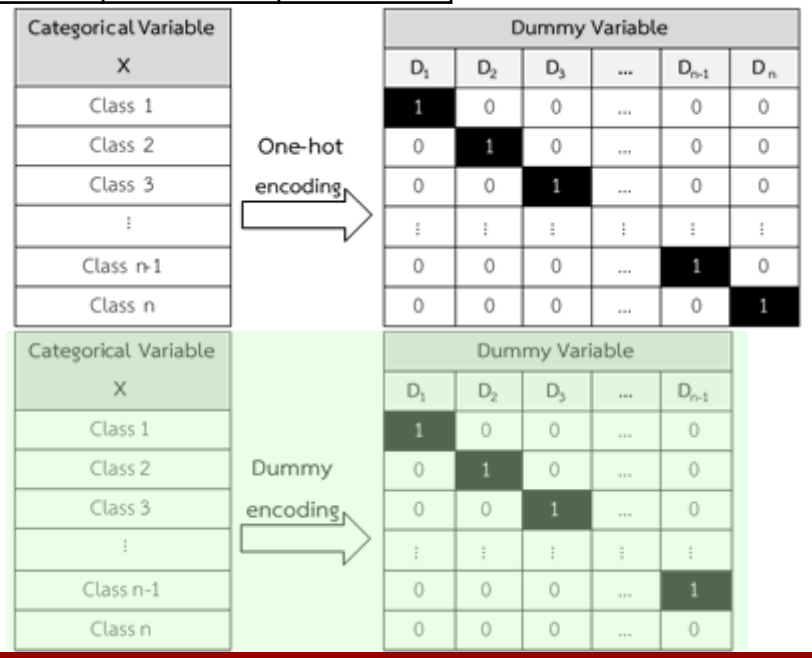
Discussing Result from model comparison with Kaggle users'

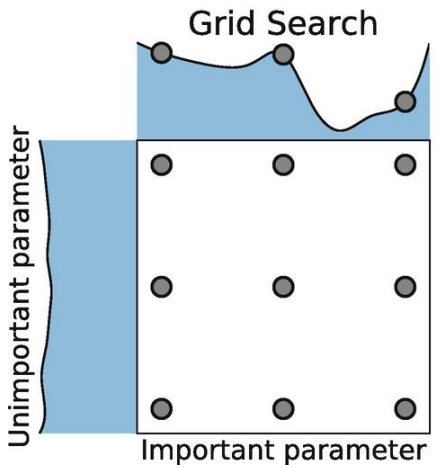
Rank	Algorithm	Author	Accuracy	F1-score
1	XGB Classifier	Kaan Boke	81.62%	75.22%
2	XGB Classifier	Graeme Keleher	81.41%	74.81%
3	XGB Classifier	Arezoo Dahesh	78.22%	73.14%
4	XGB Classifier	Thatchakarn, Teethavat	79.77%	72.46%
5	RandomForest Classifier	Iching Wang	79.13%	71.79%
6	Neural Network: Sequence Model (RNN)	Jatin	79.25%	70.98%

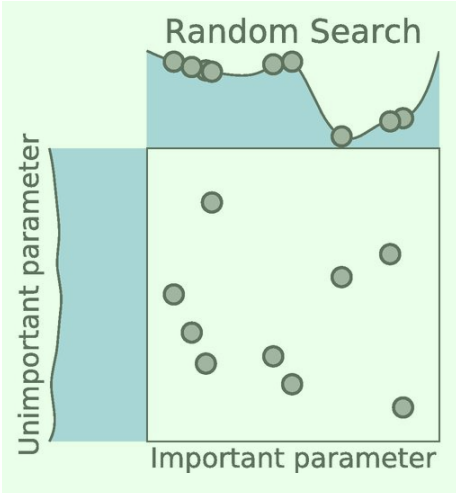


Source: https://thaddeus-segura.com/xgboost/

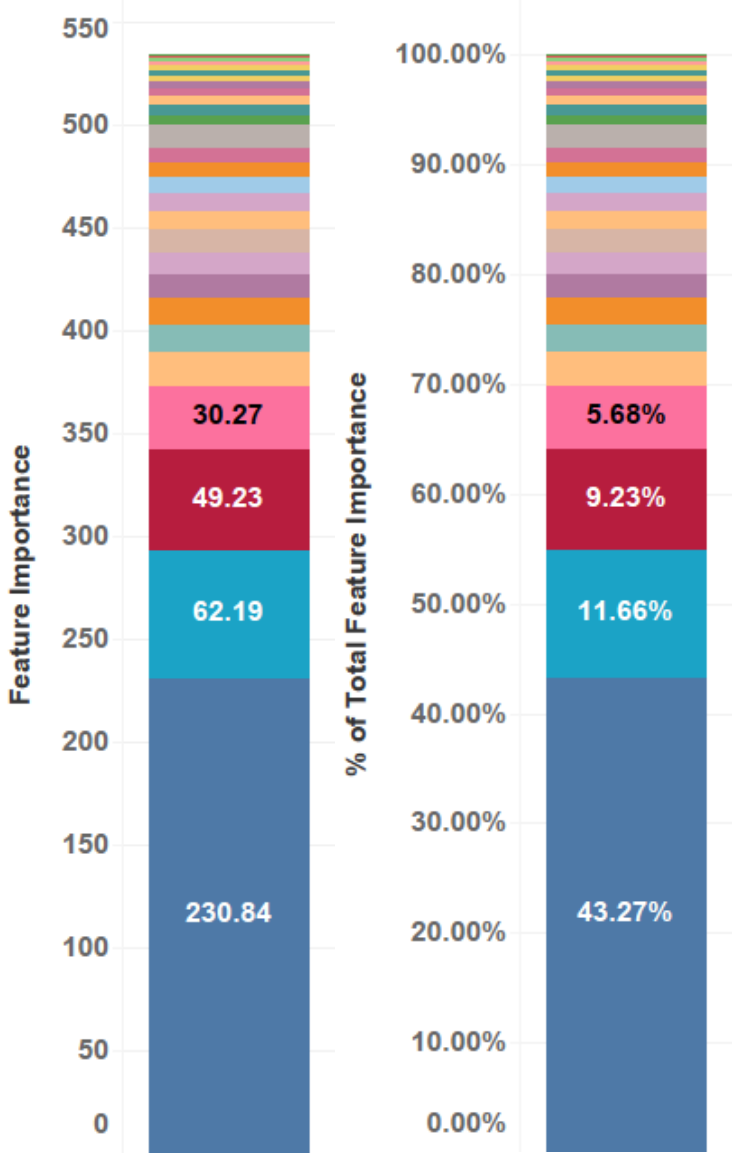




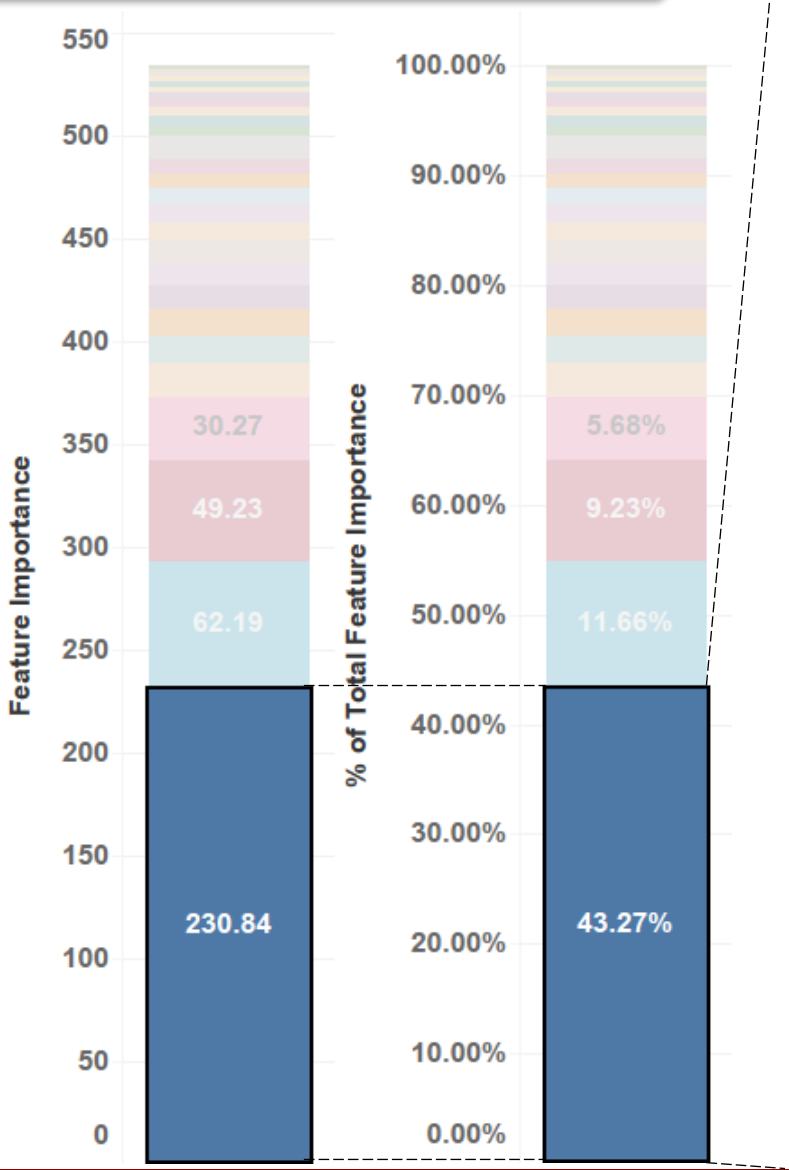


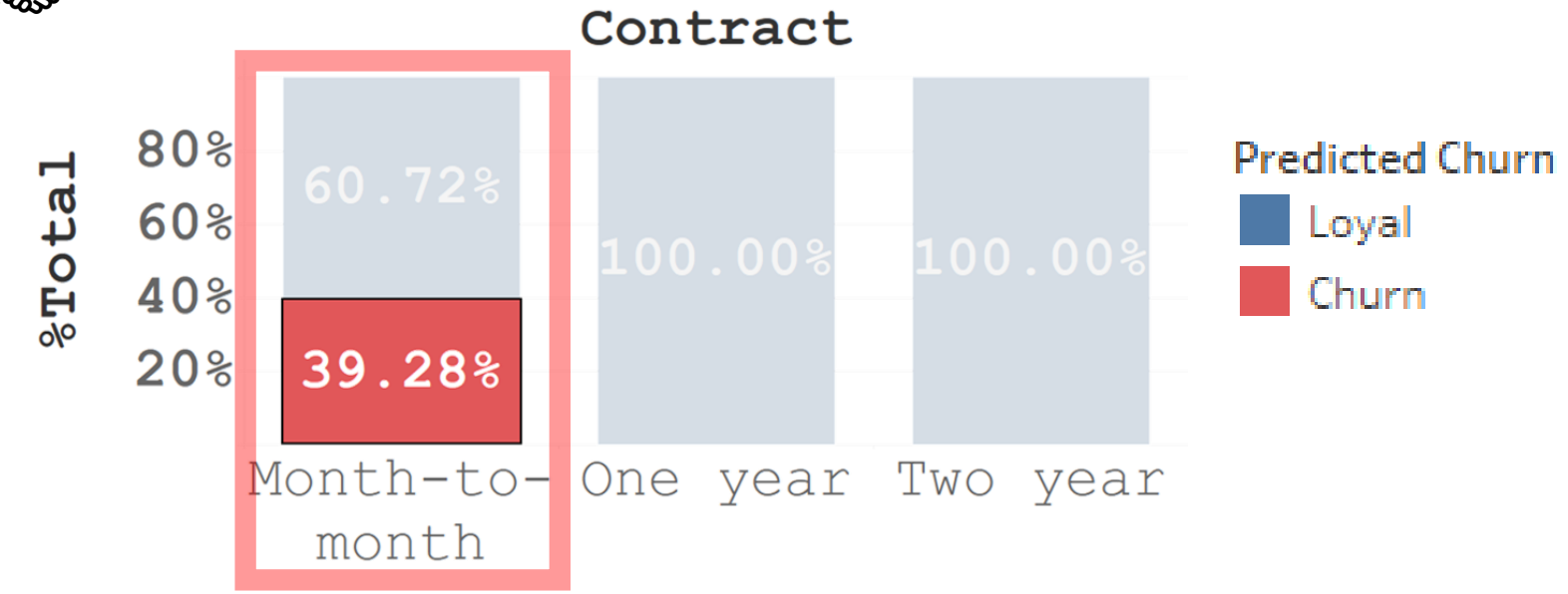


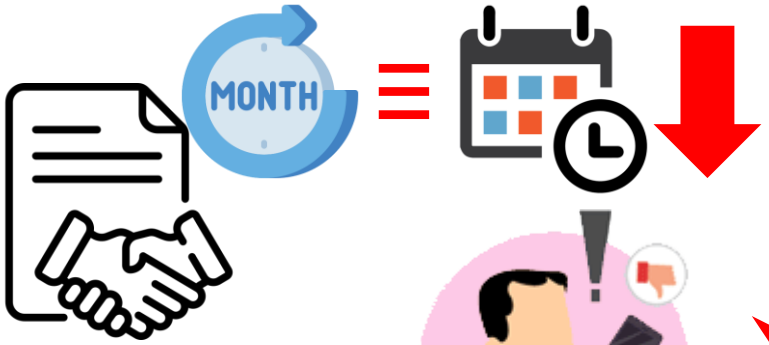
Feature Importance

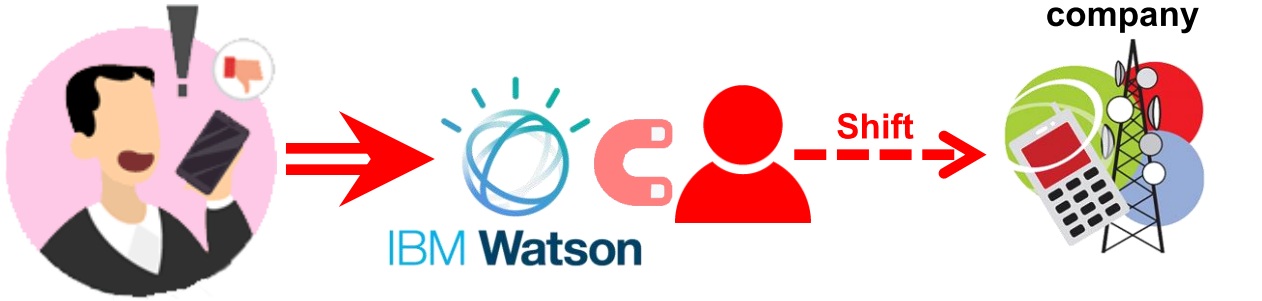


Feature Importance

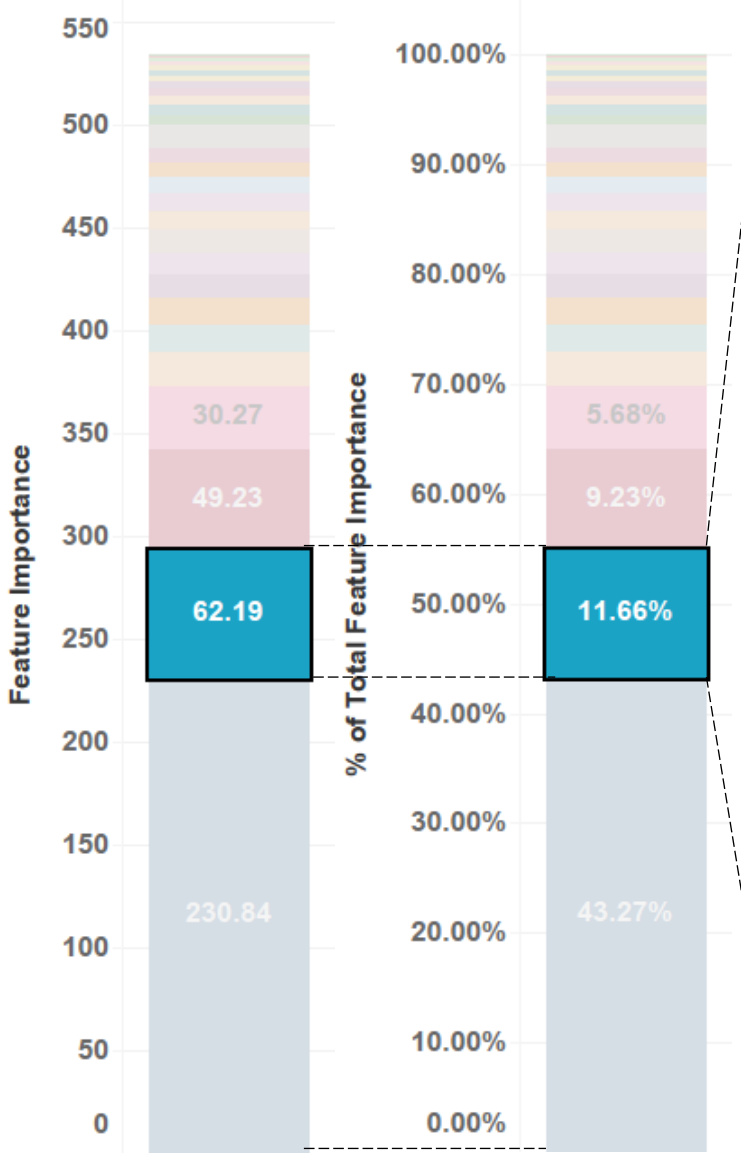


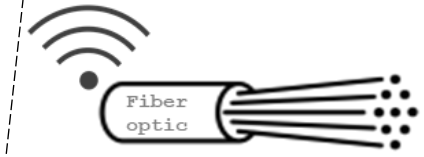


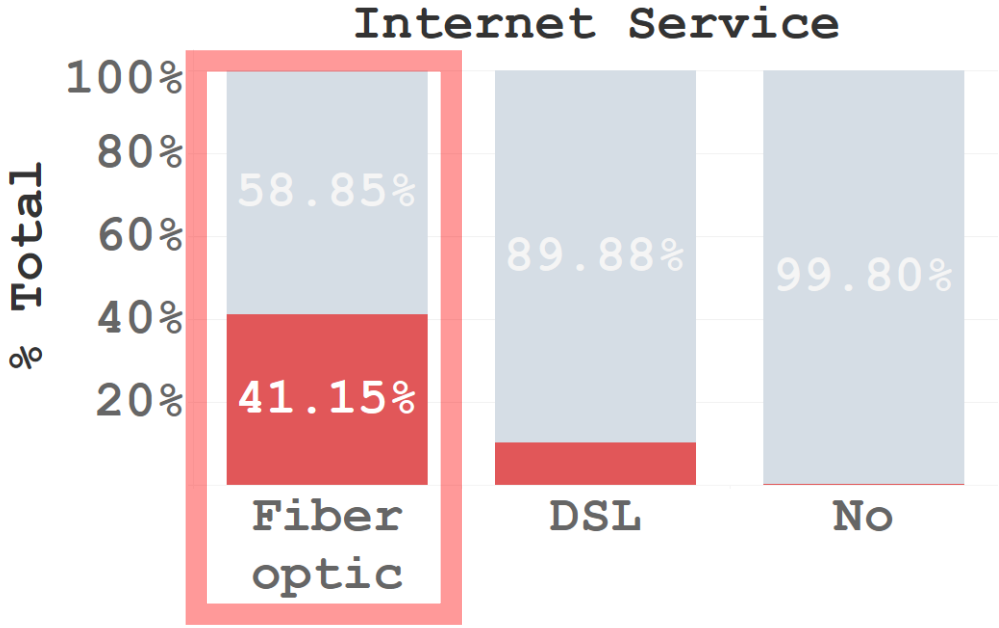


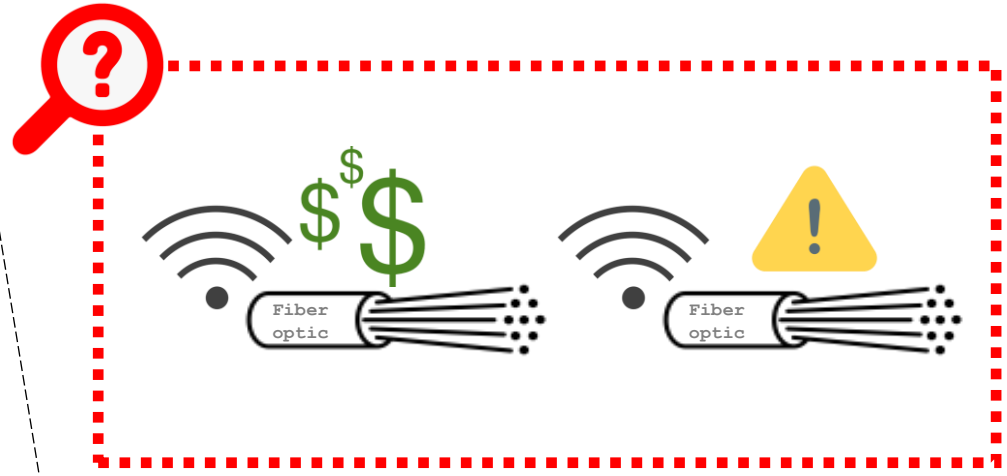


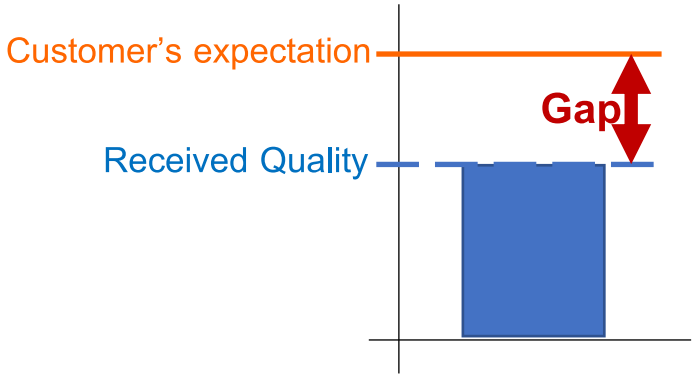
Feature Importance



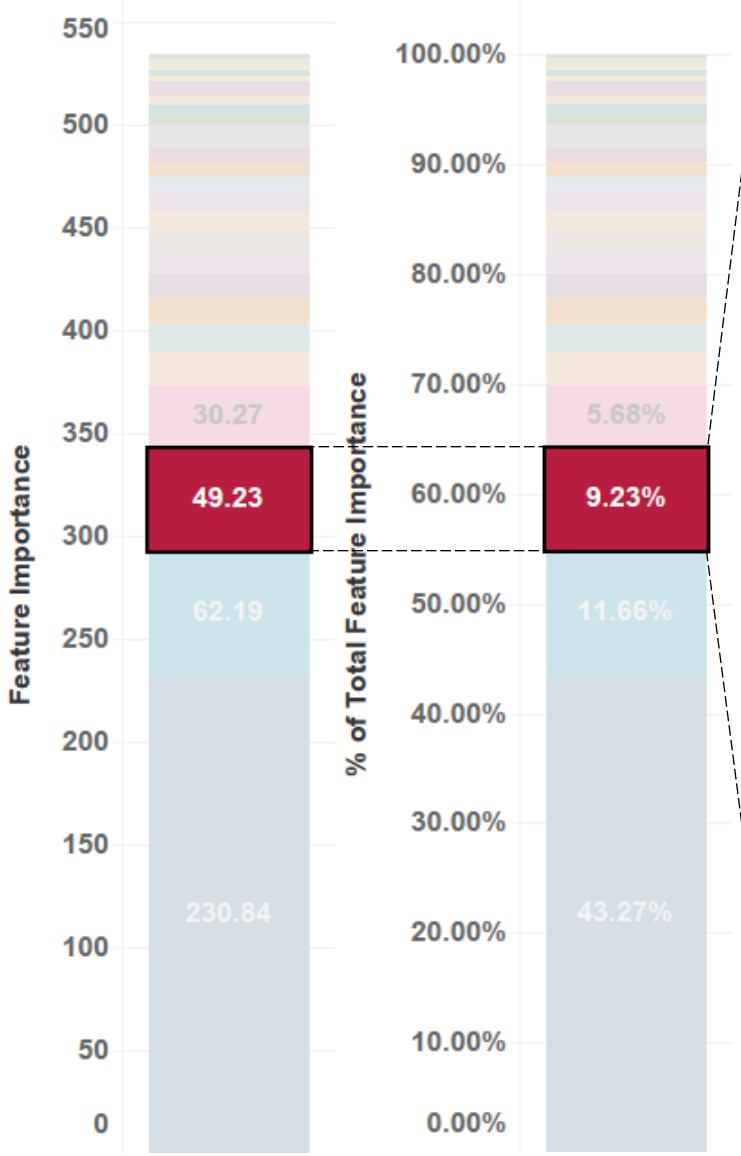


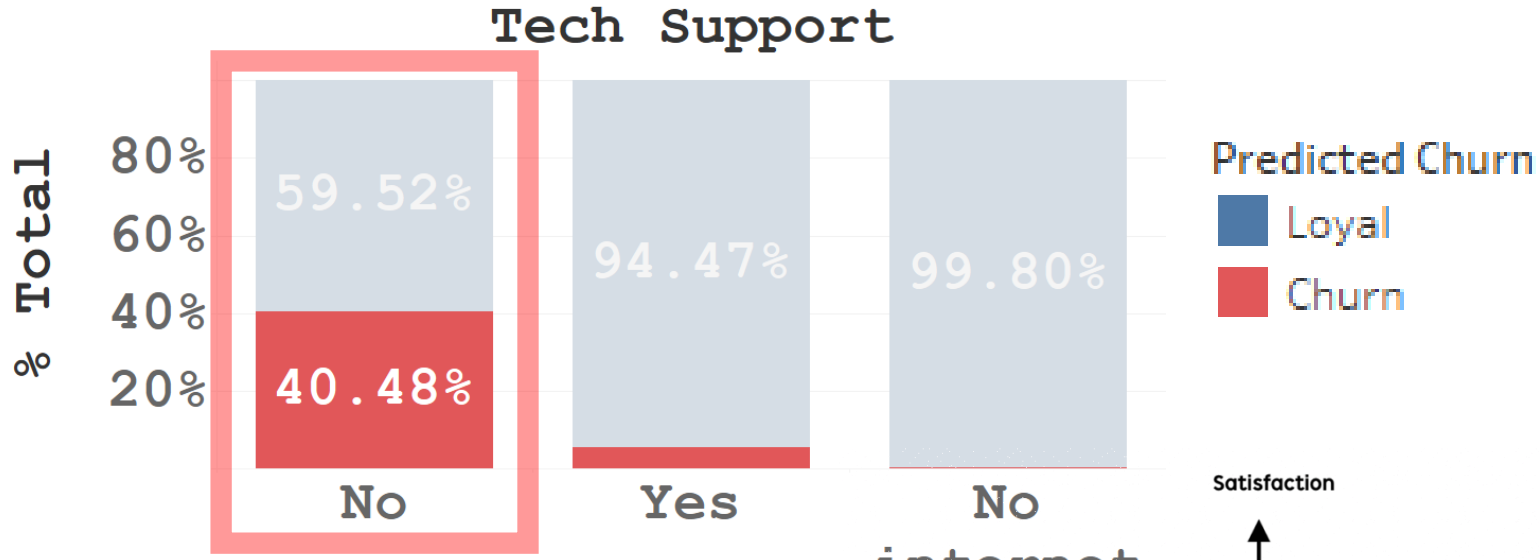


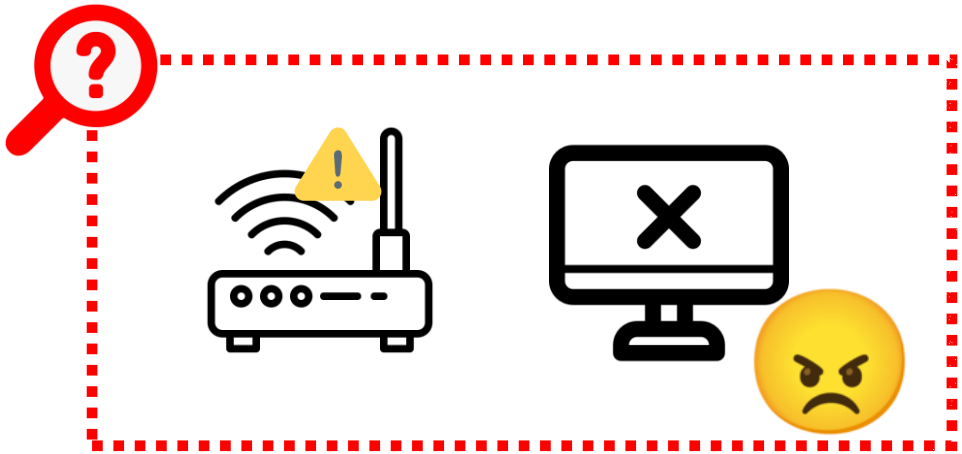


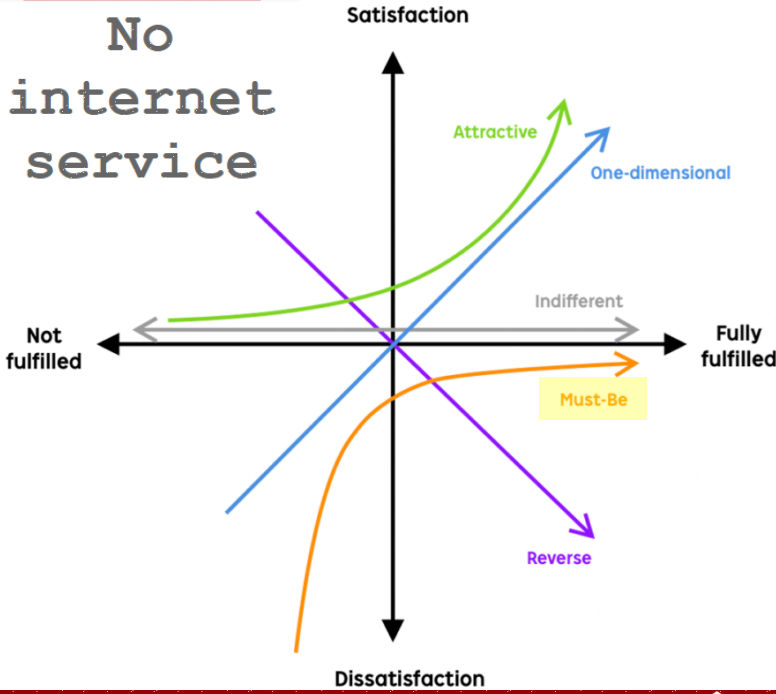


Feature Importance

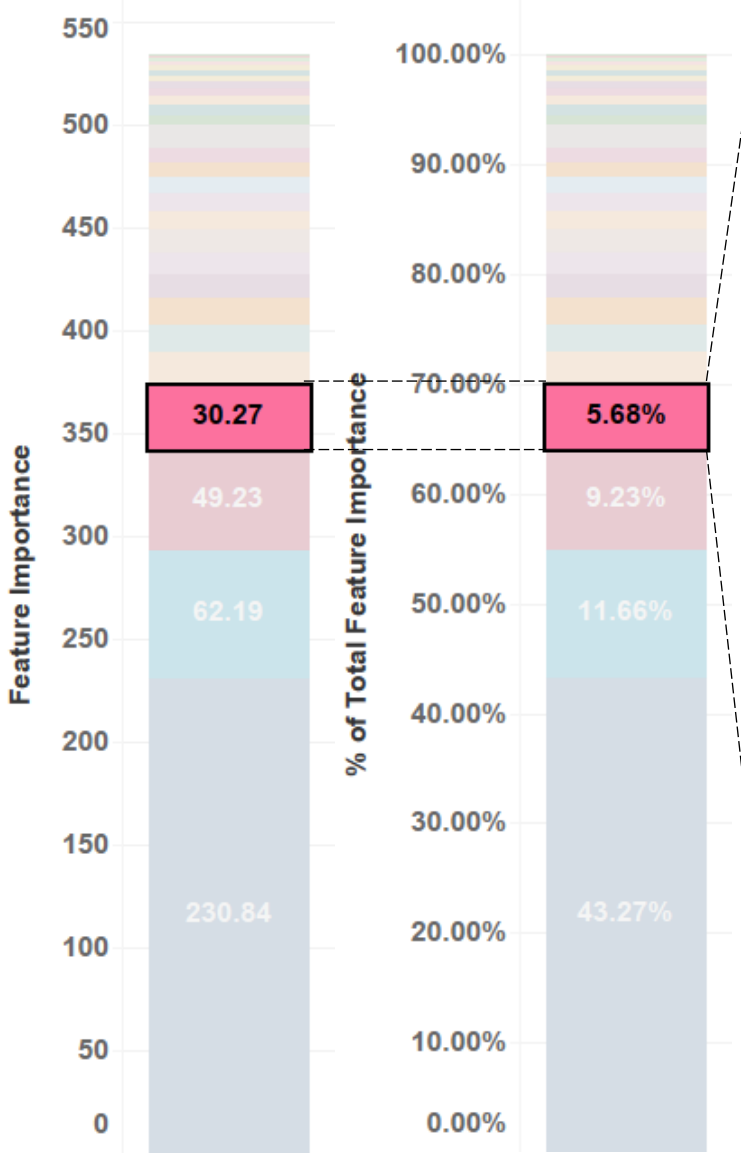


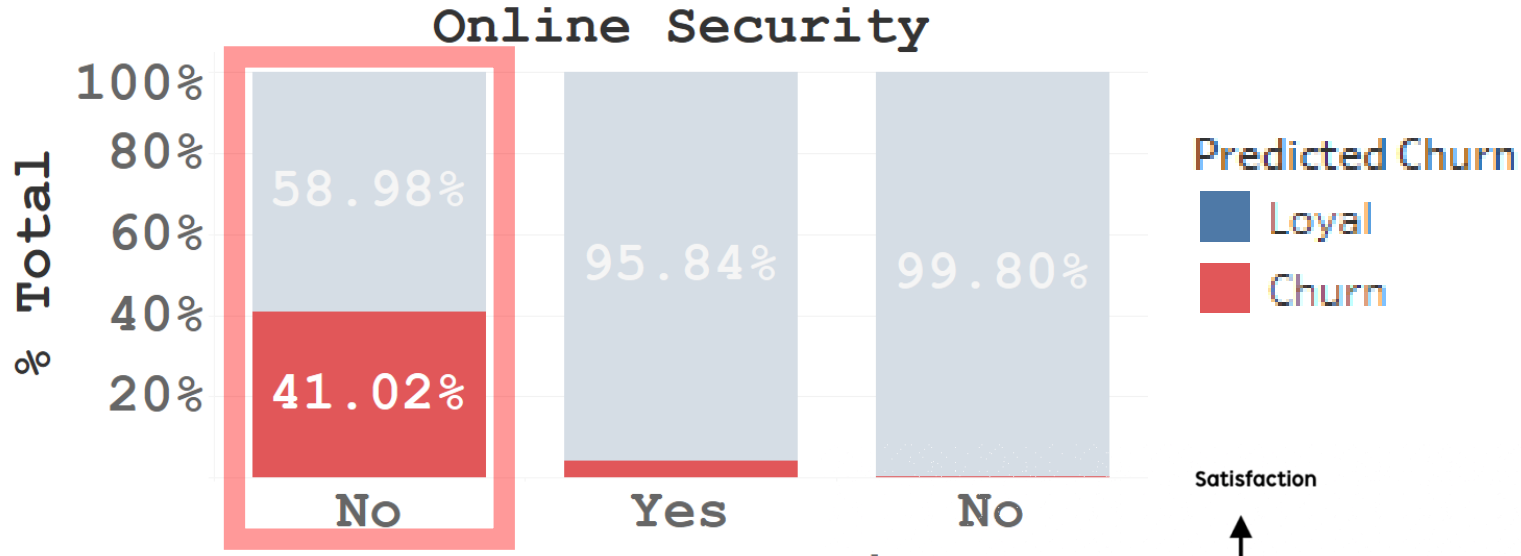


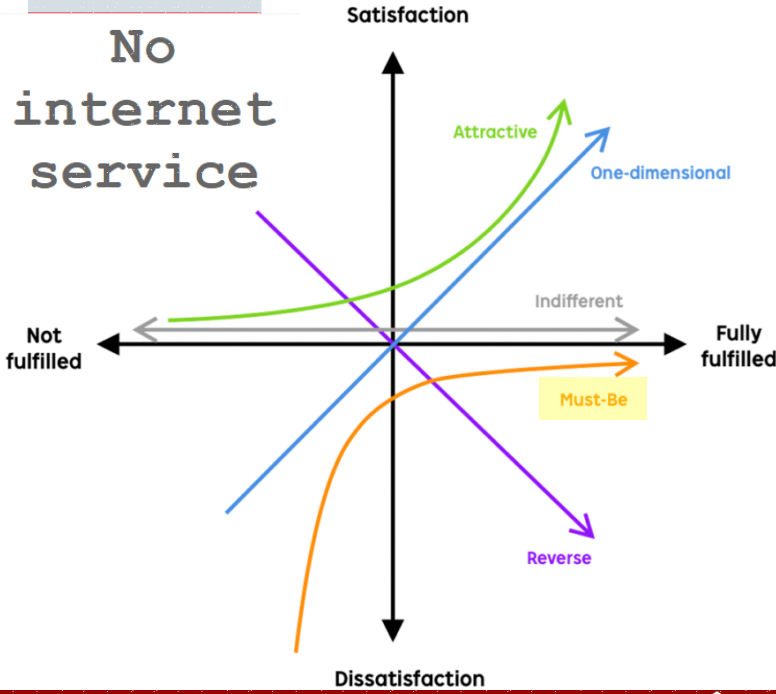




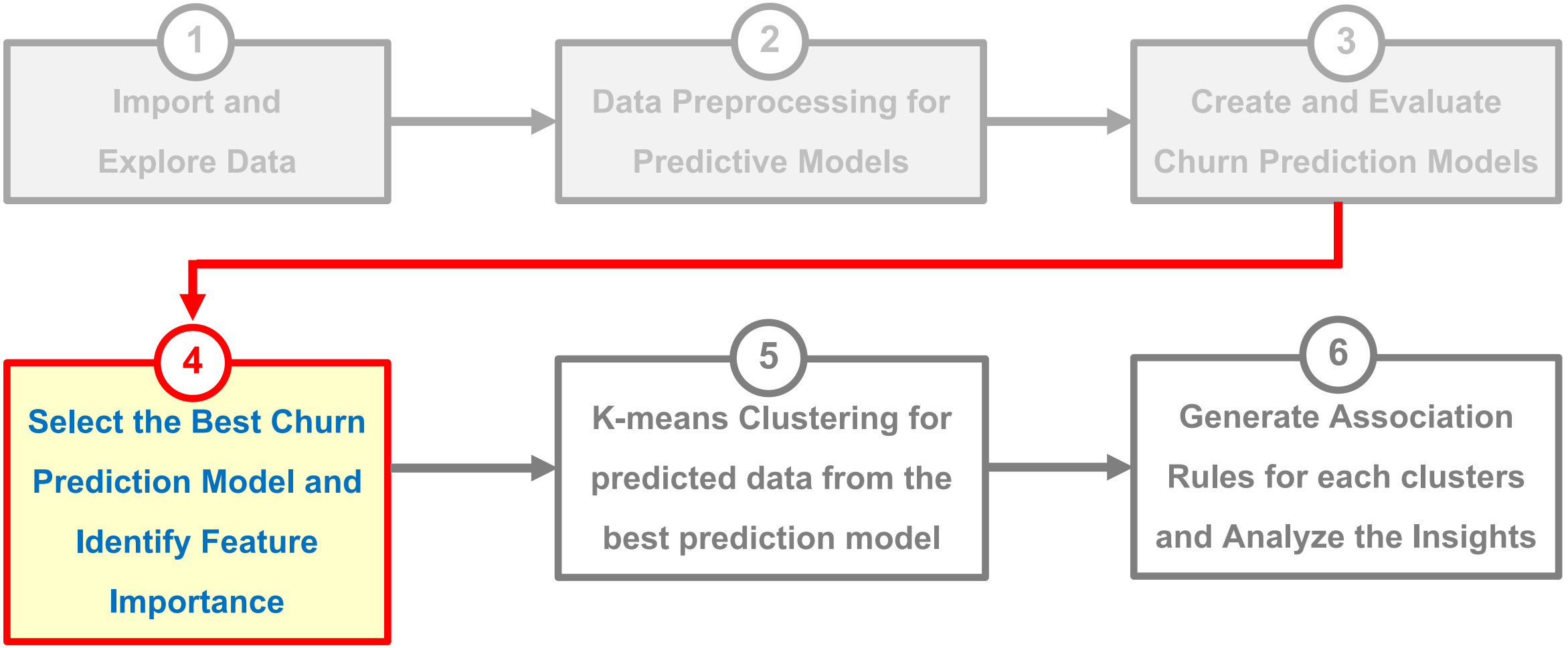
Feature Importance



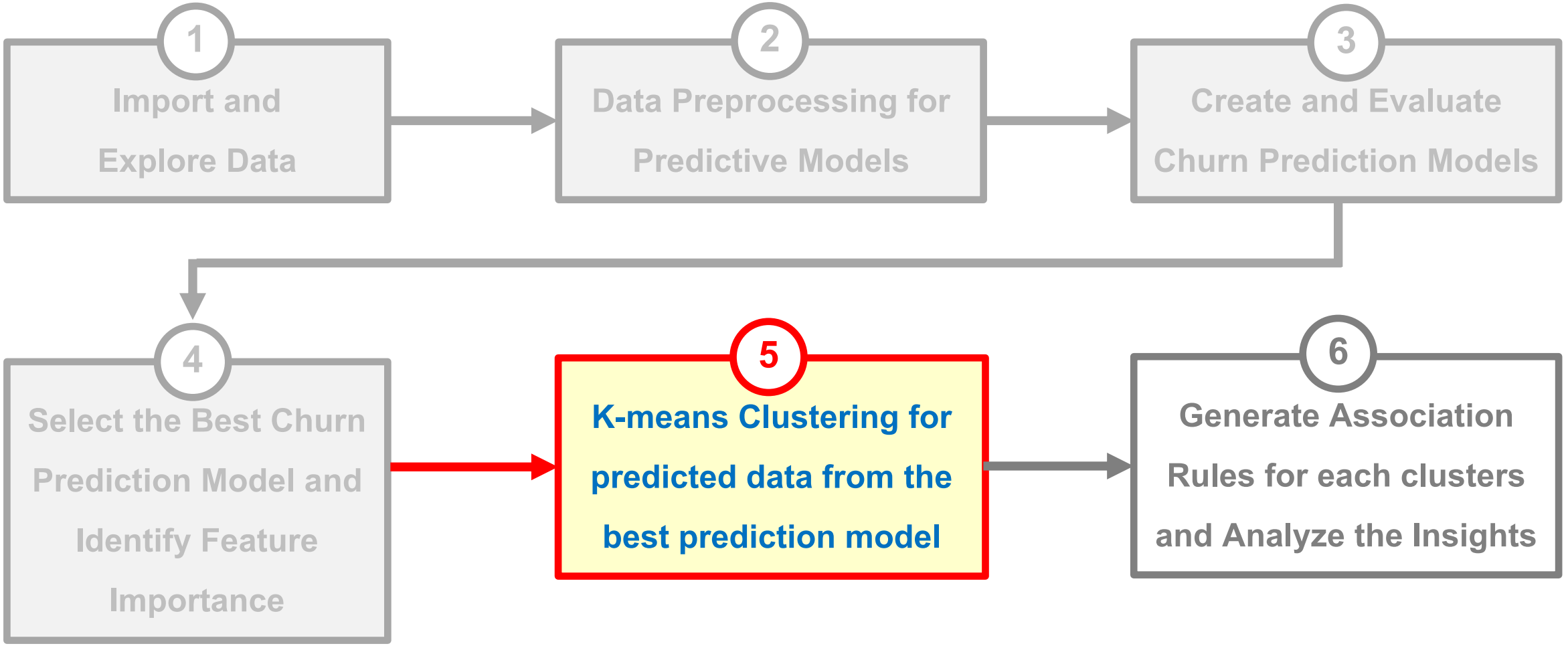




Procedure

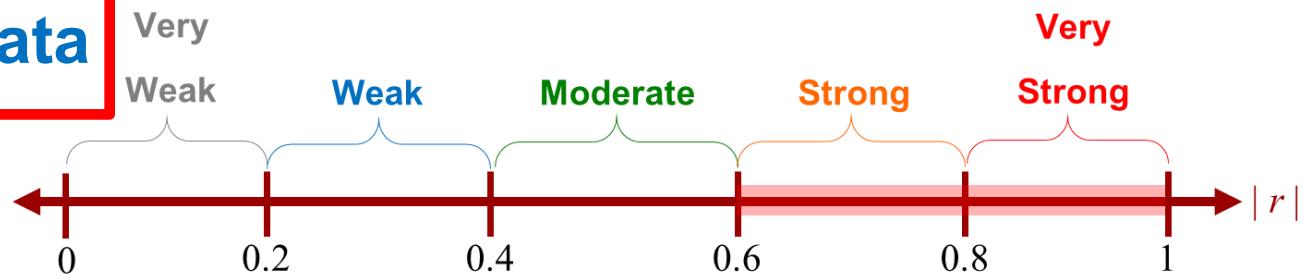

1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

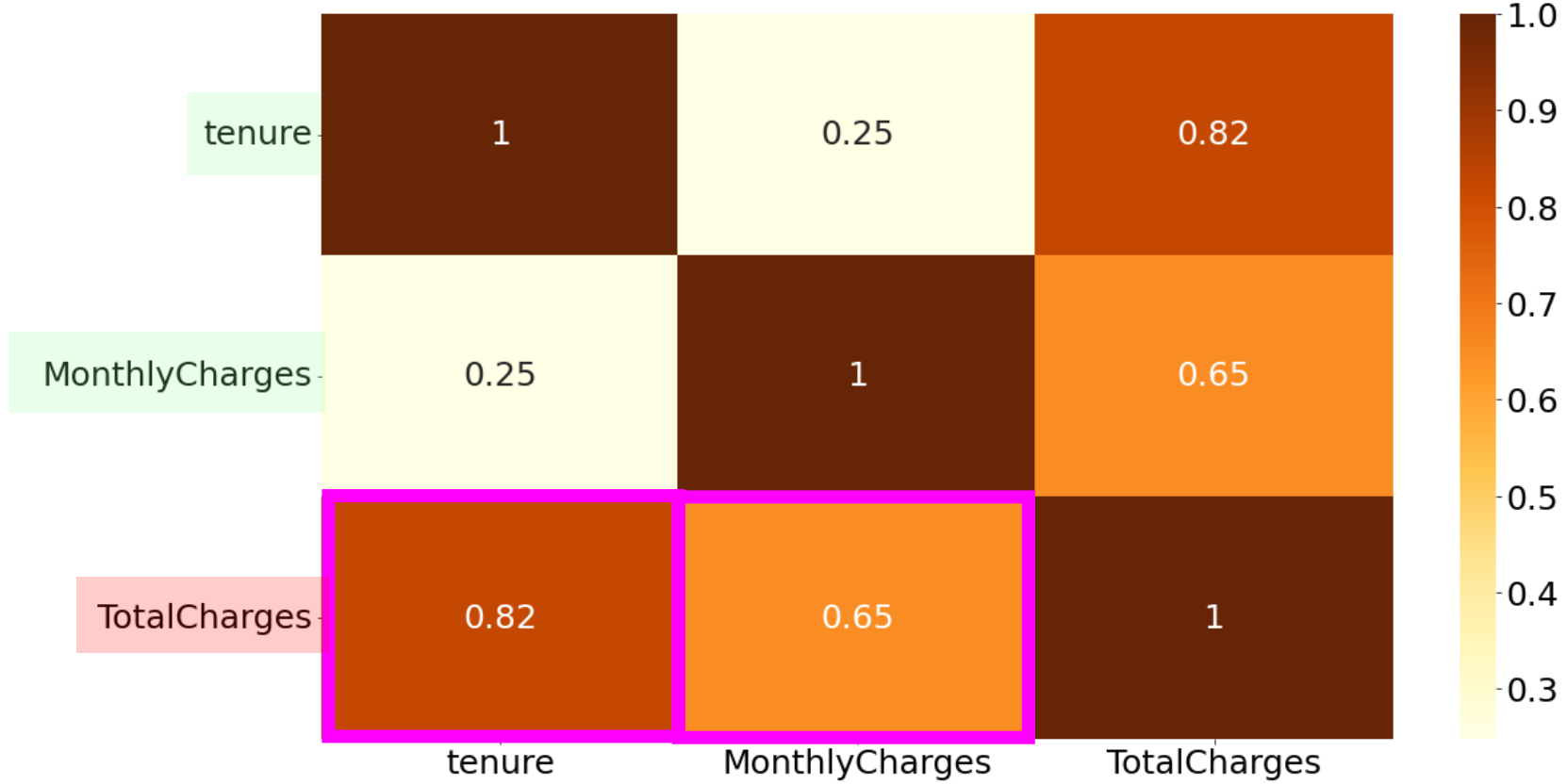
Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

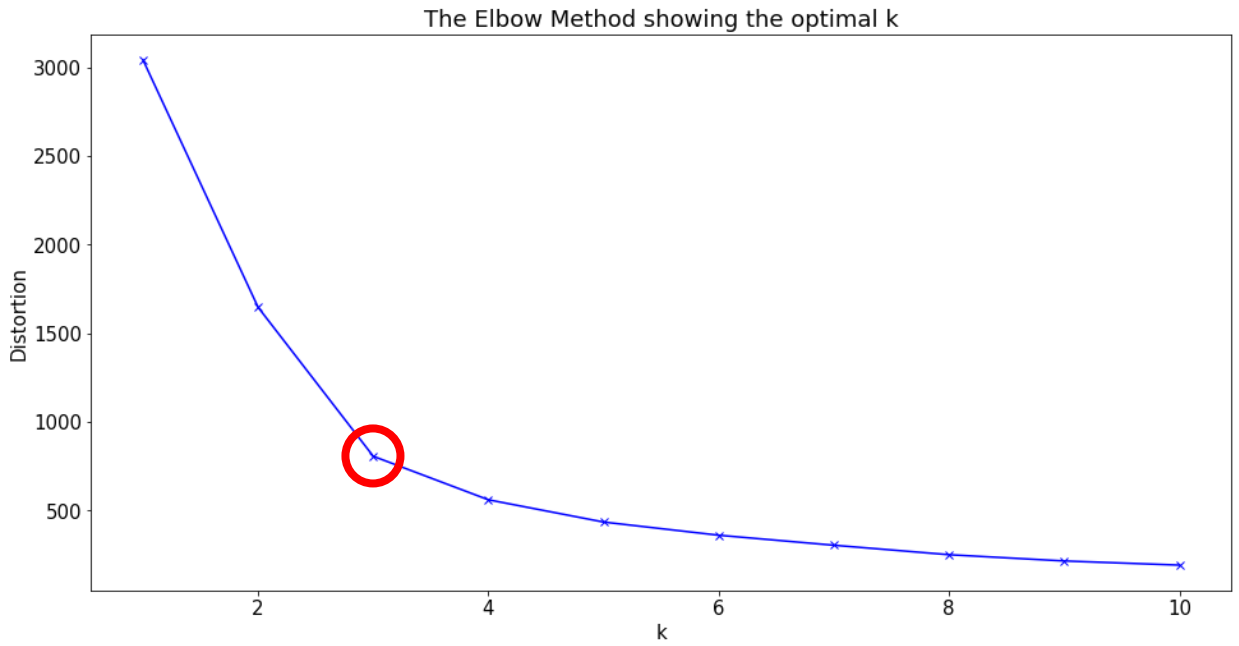
5 K-means Clustering for predicted data

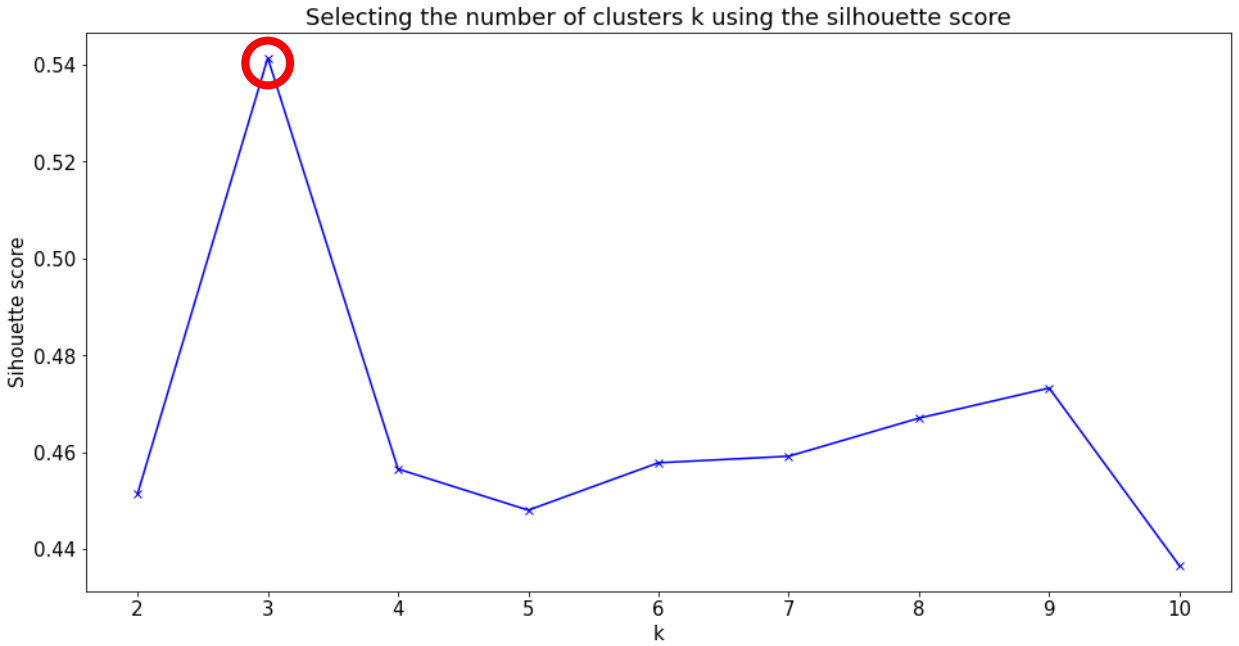
Attribute Selection





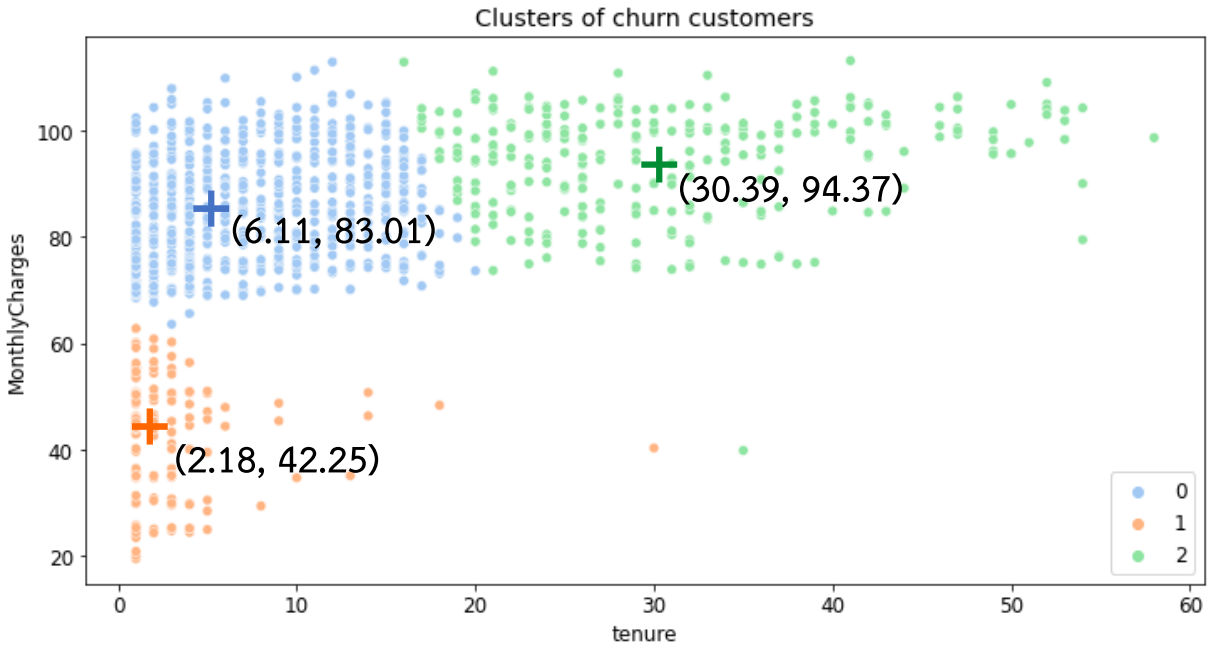
Selecting optimal number of cluster (k)

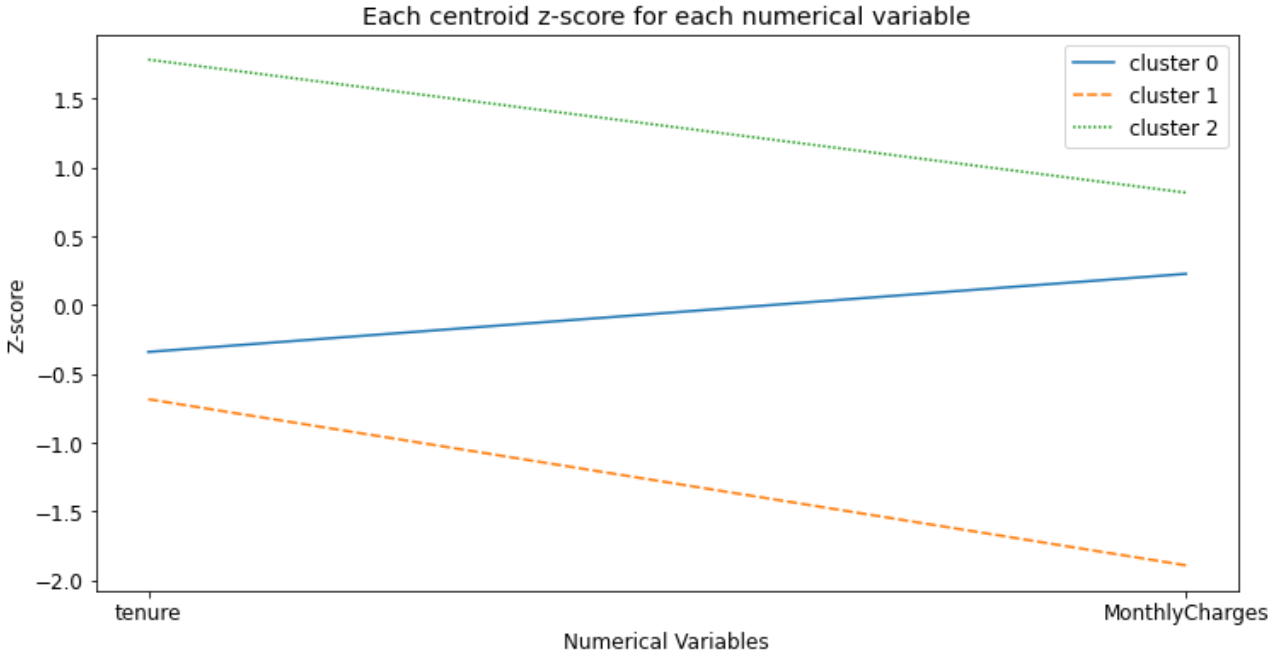


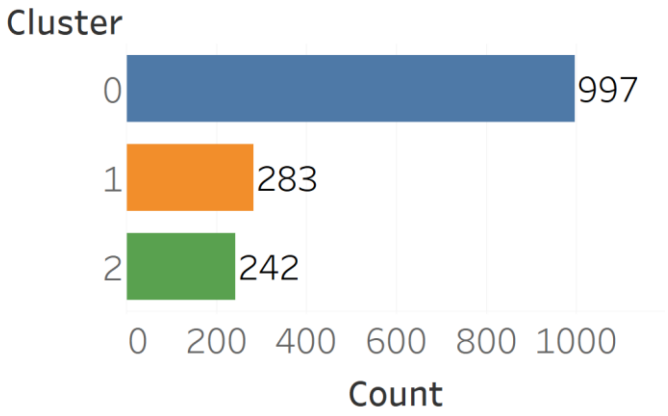


Creating Clustering Models

High intra-class similarity, low inter-class similarity







$$Z = \frac{x - \bar{x}}{S}$$

Z = standard score

x = observed value

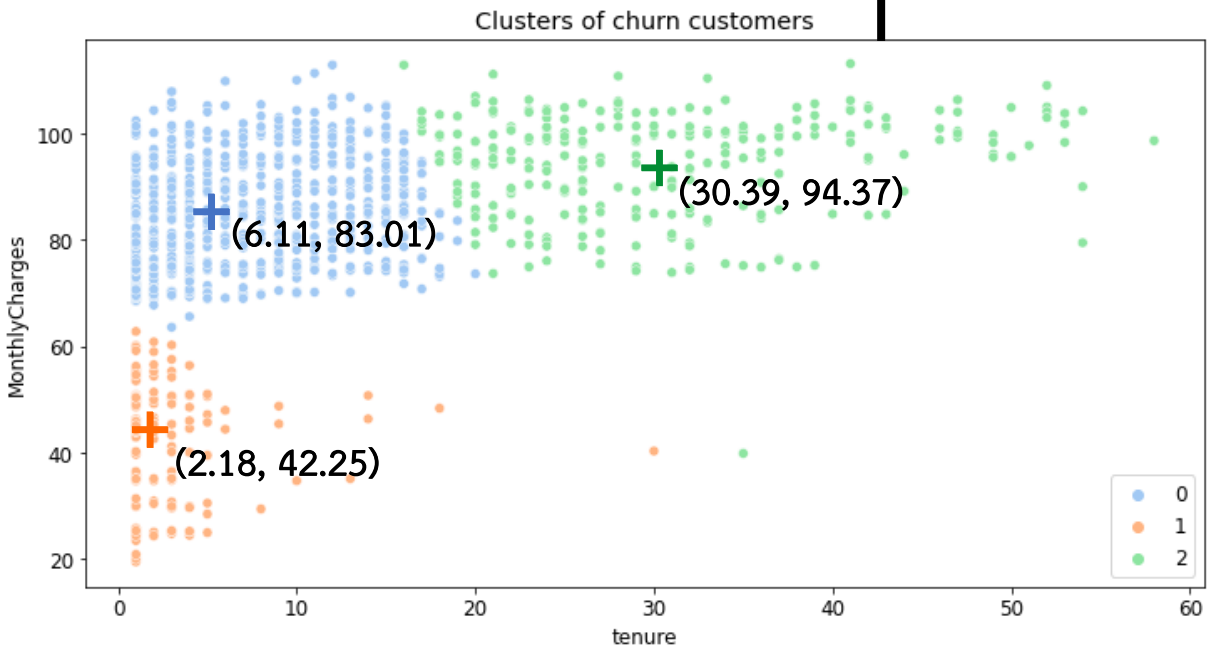
$\bar{x}$ = mean of the sample

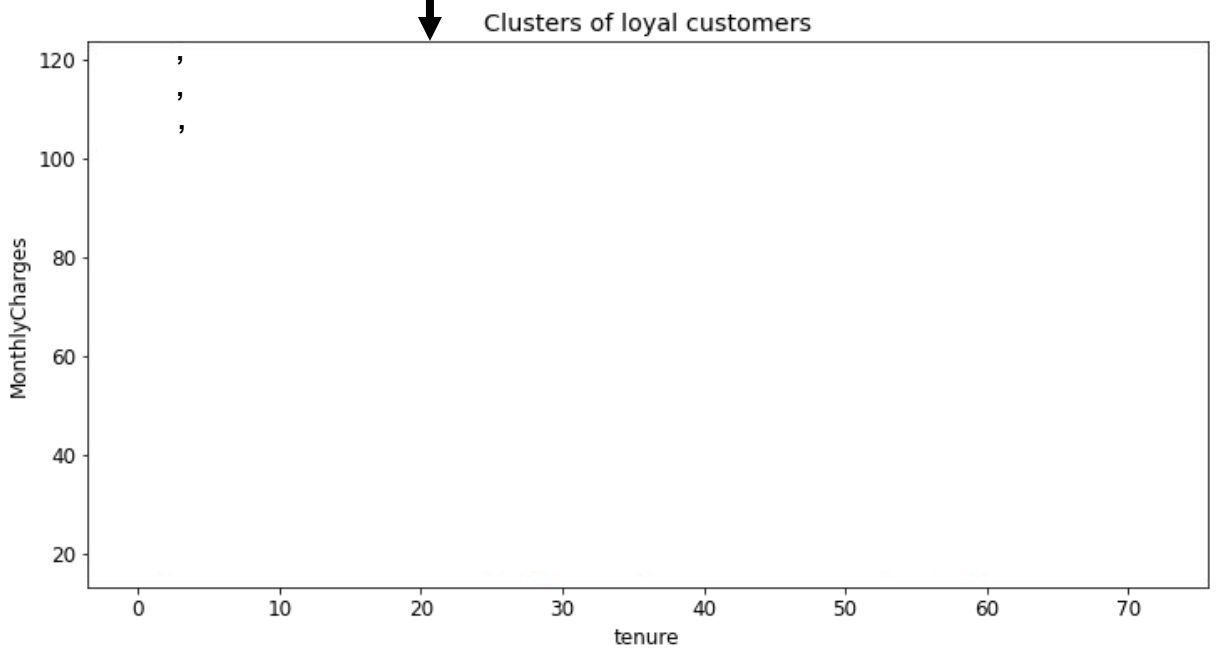
S = standard deviation of the sample

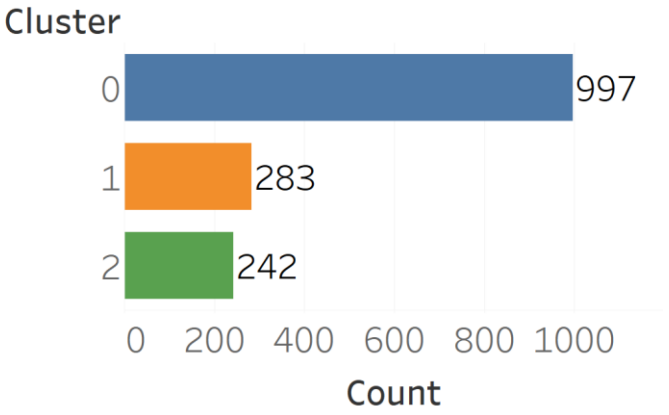
Creating Clustering Models

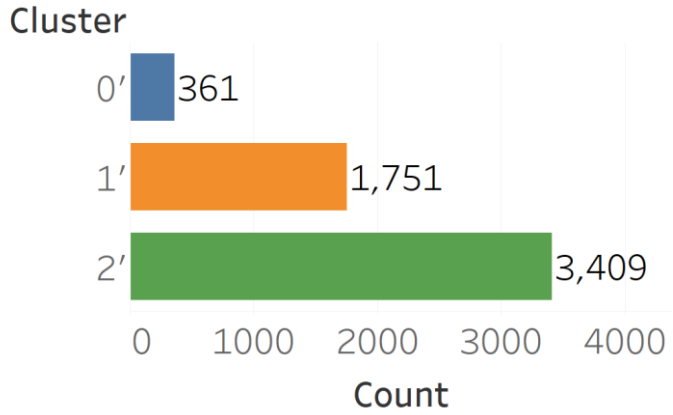
High intra-class similarity, low inter-class similarity

Use the same centroids as in Churn clusters



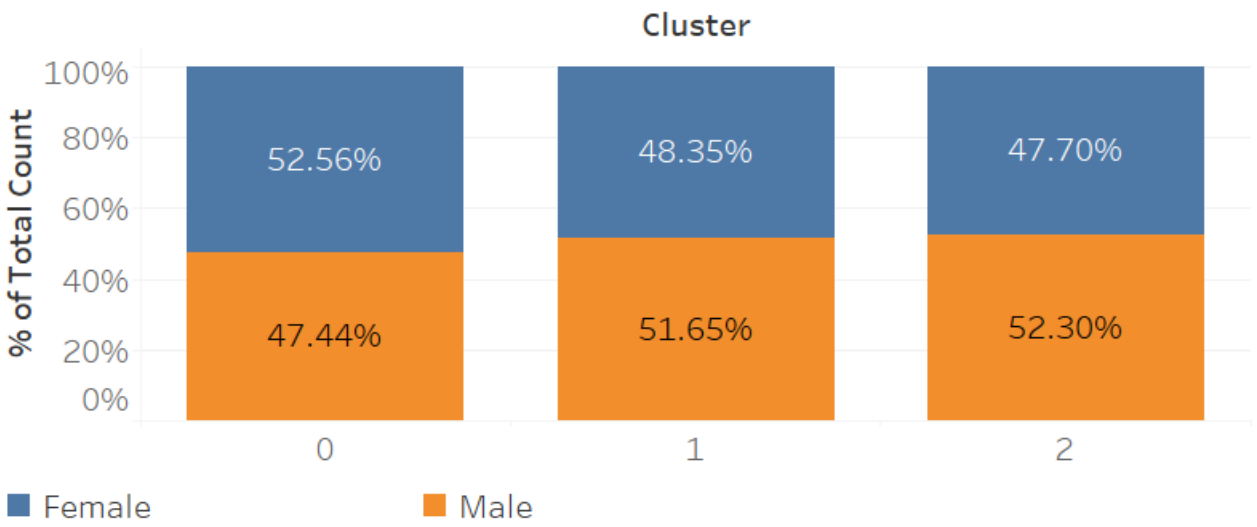




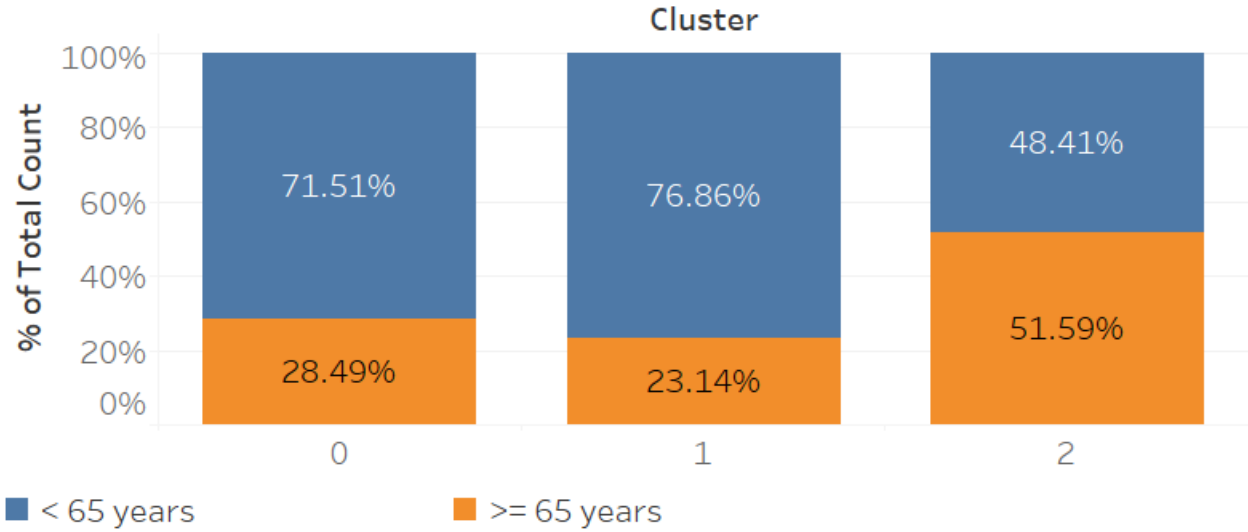


Demographics Data of Churn Cluster

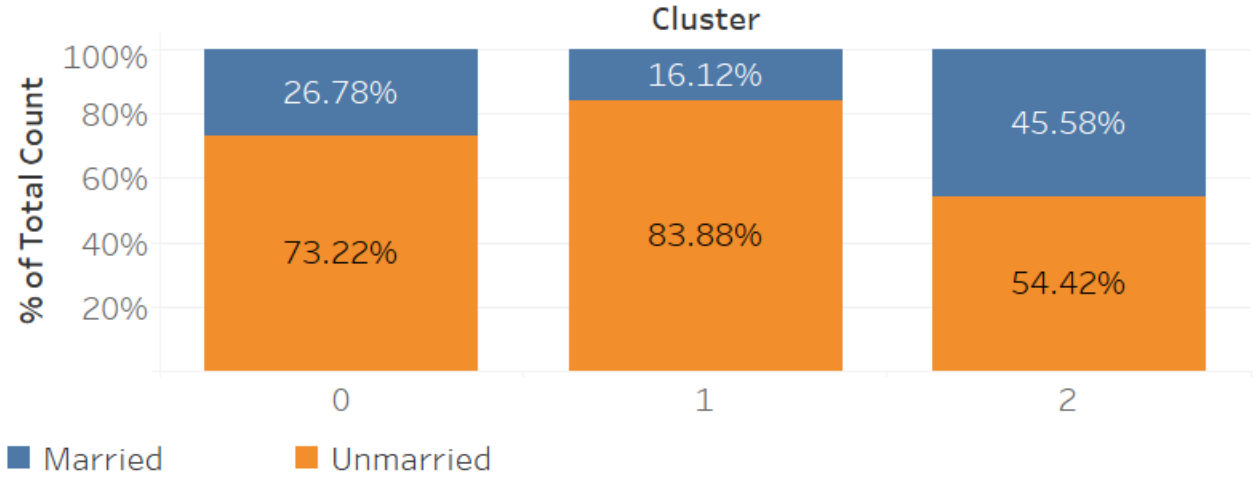
Gender



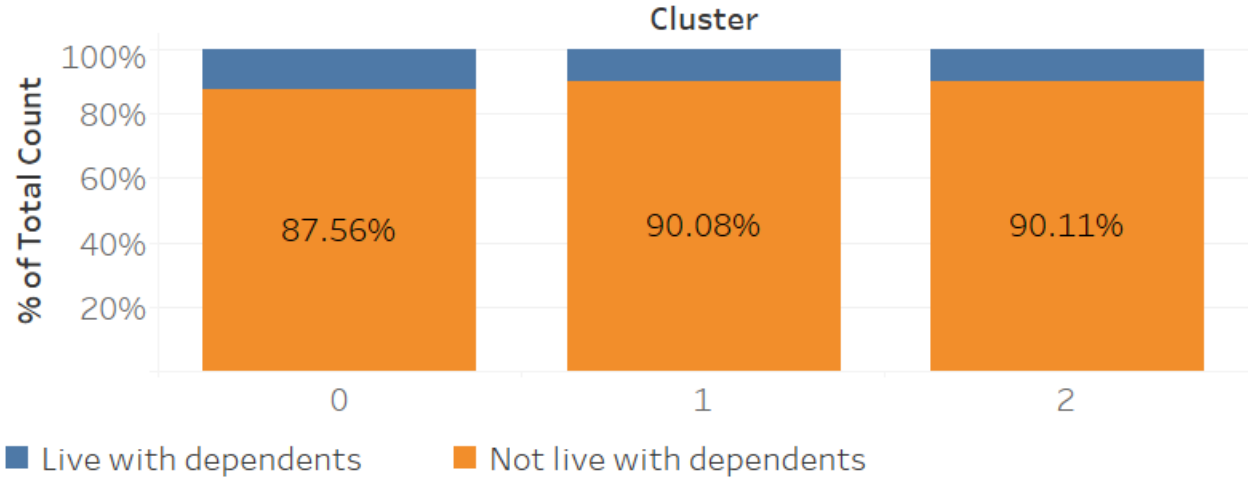
Senior Citizen



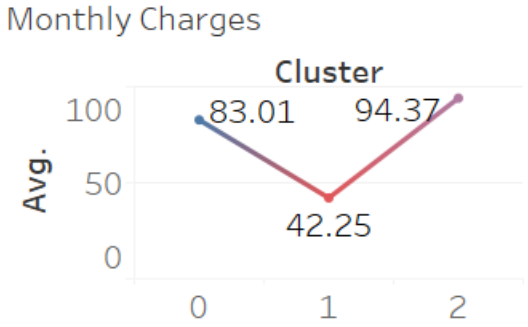
Partner

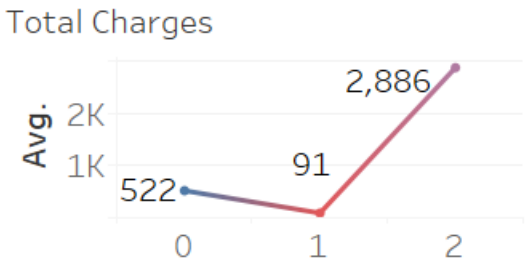


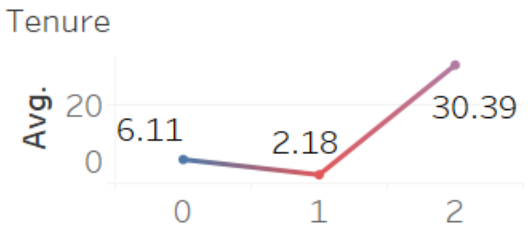
Dependent

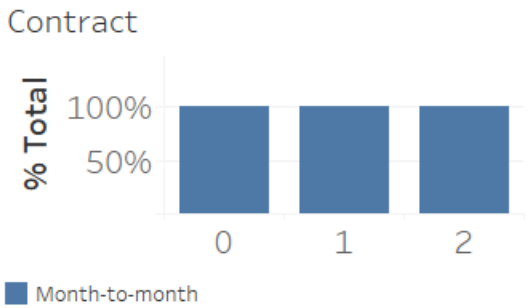


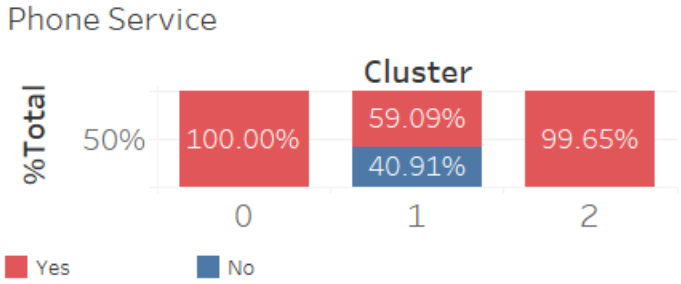
Service Data of Churn Cluster

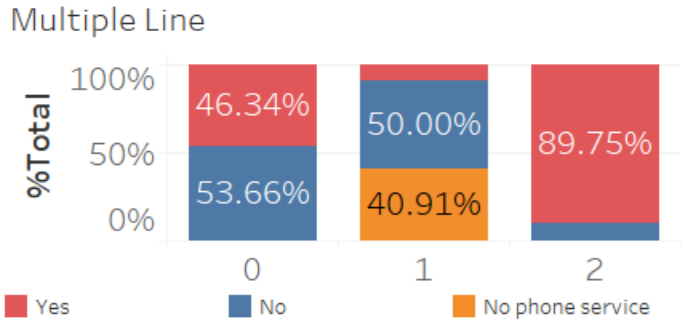


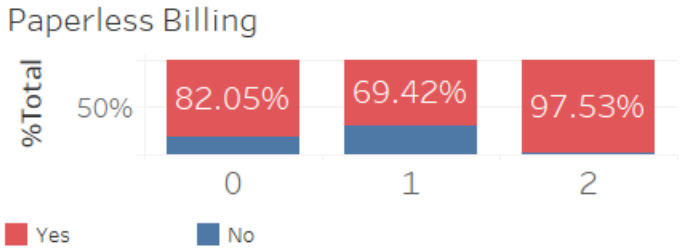


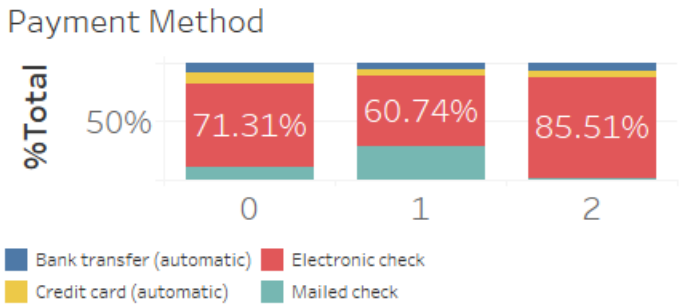


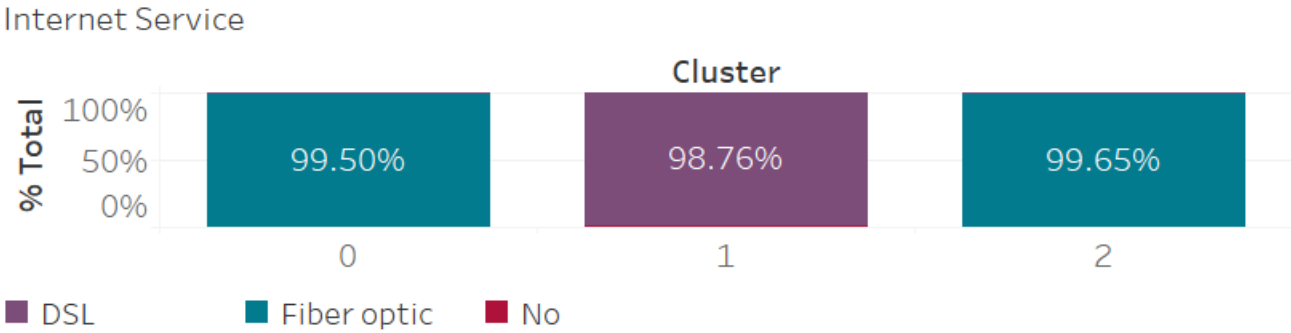


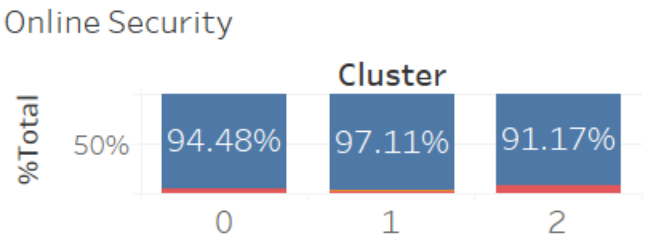


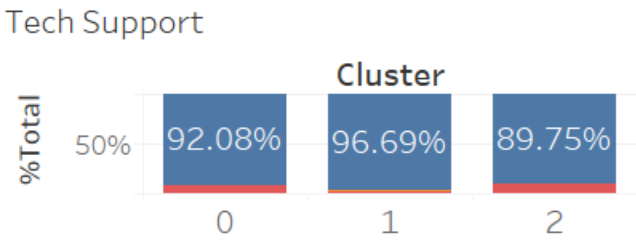


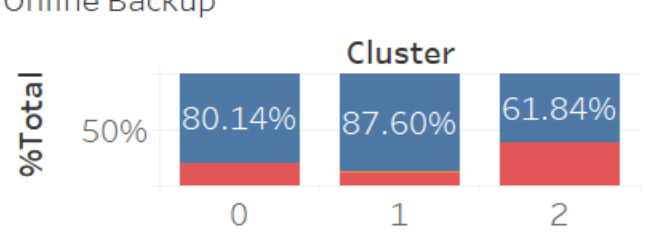


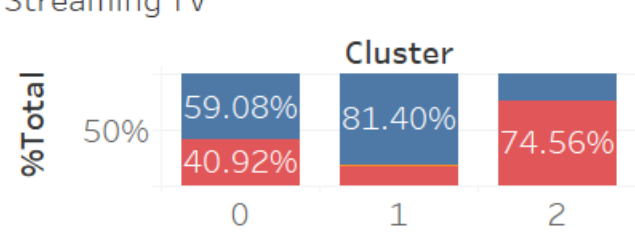


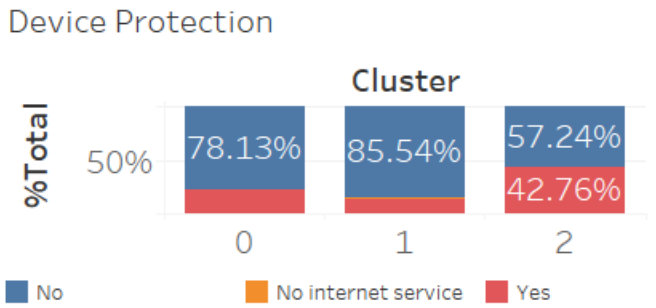


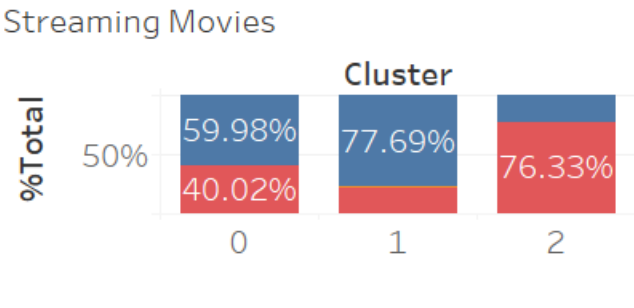






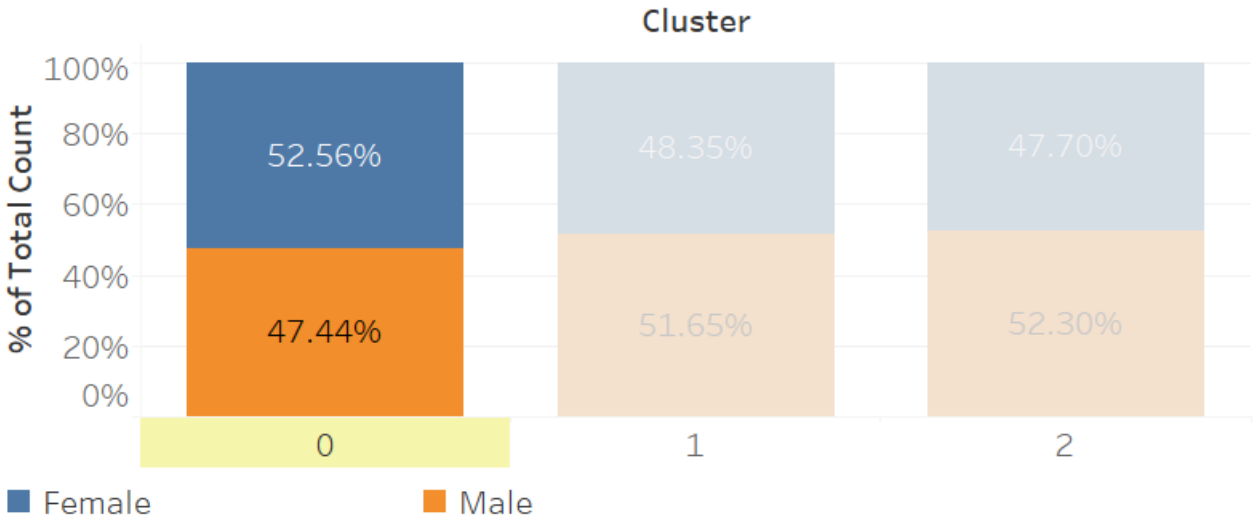


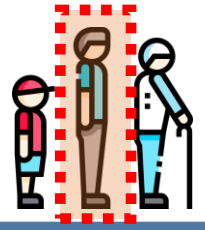




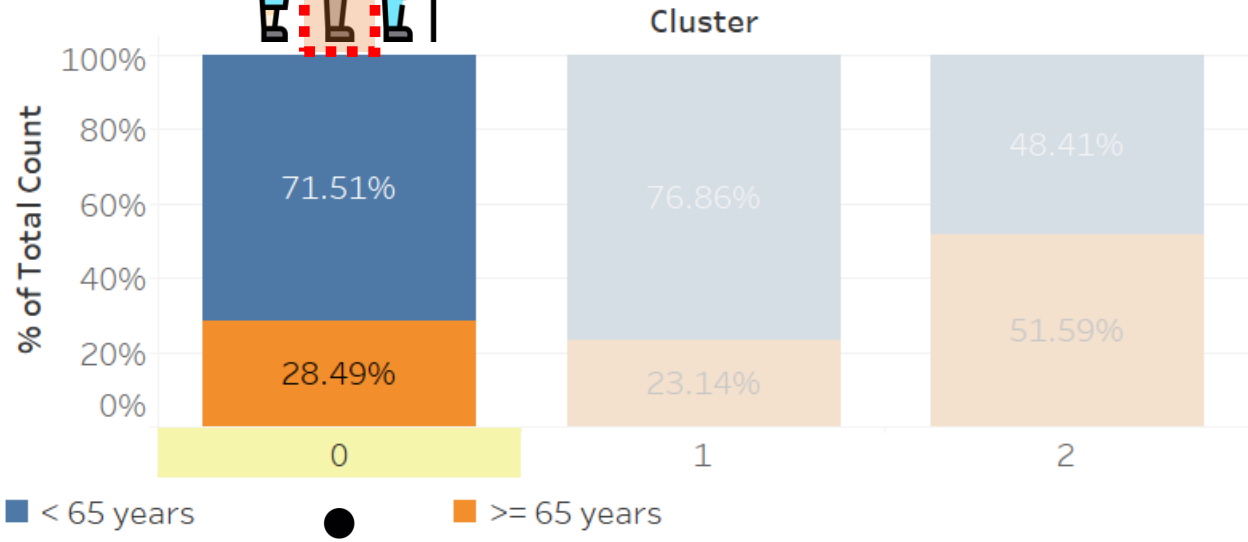
Demographics Data of Churn Cluster

Gender

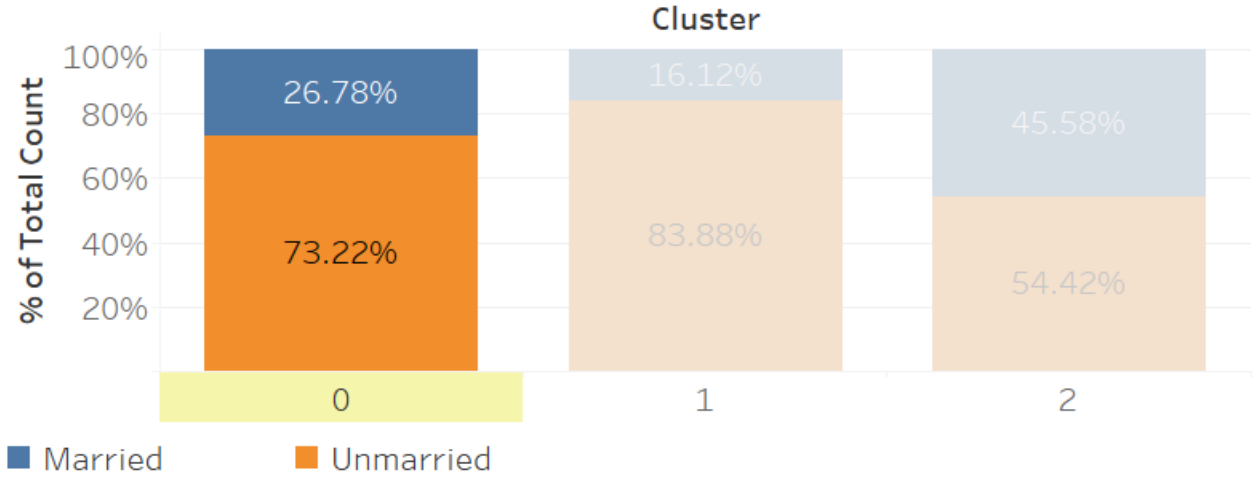




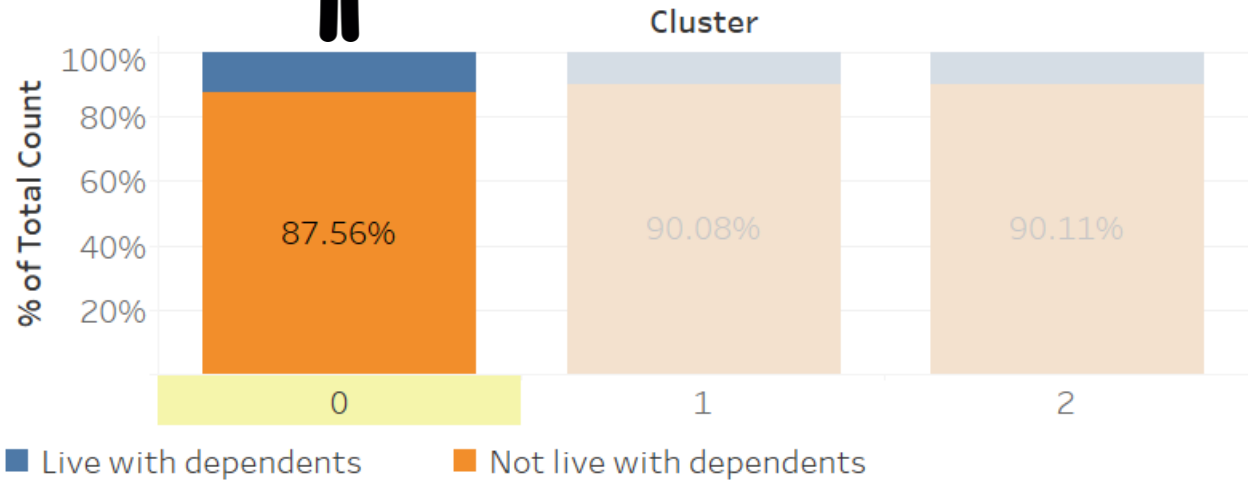
Senior Citizen



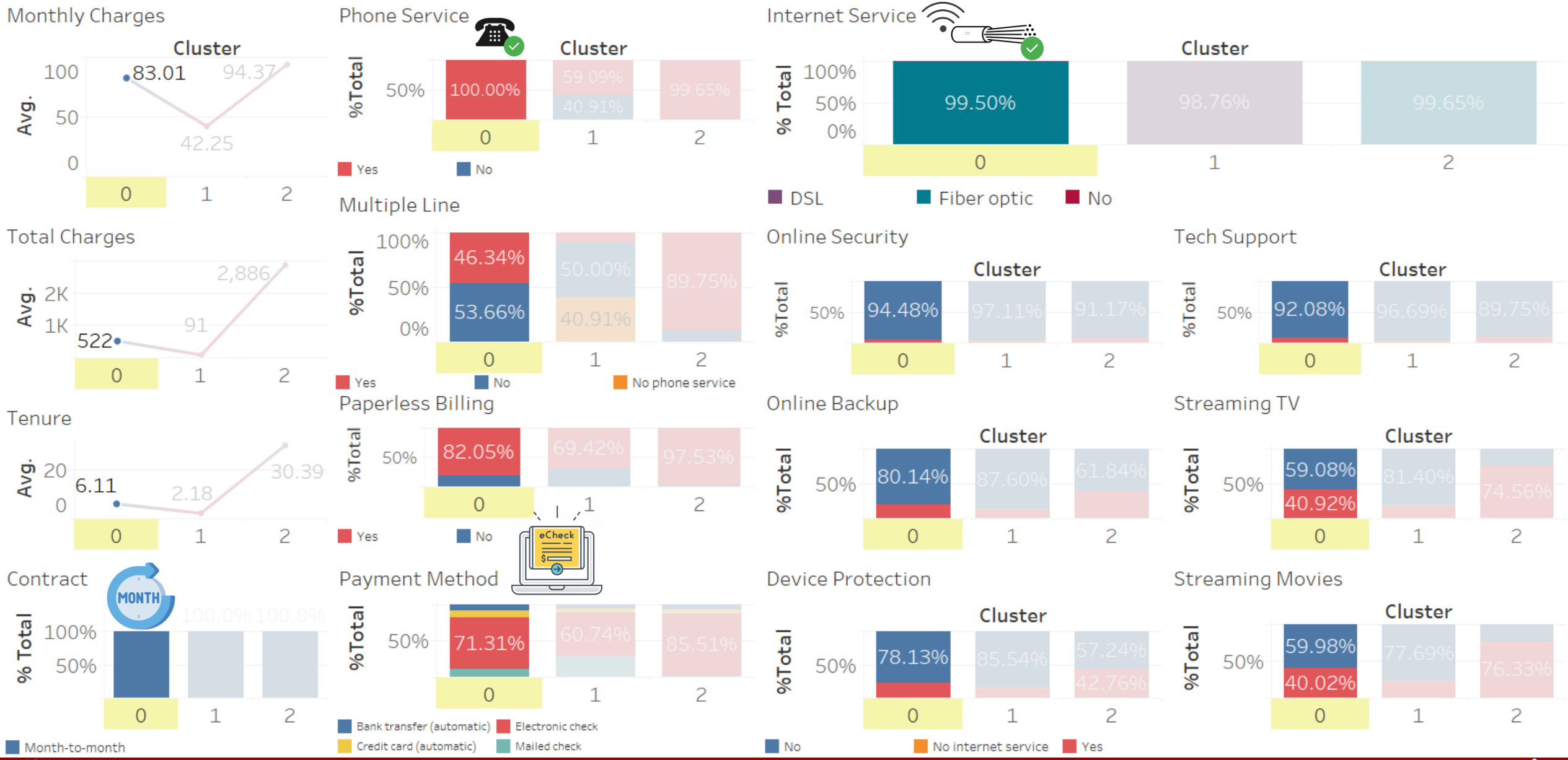
Partner



Dependent

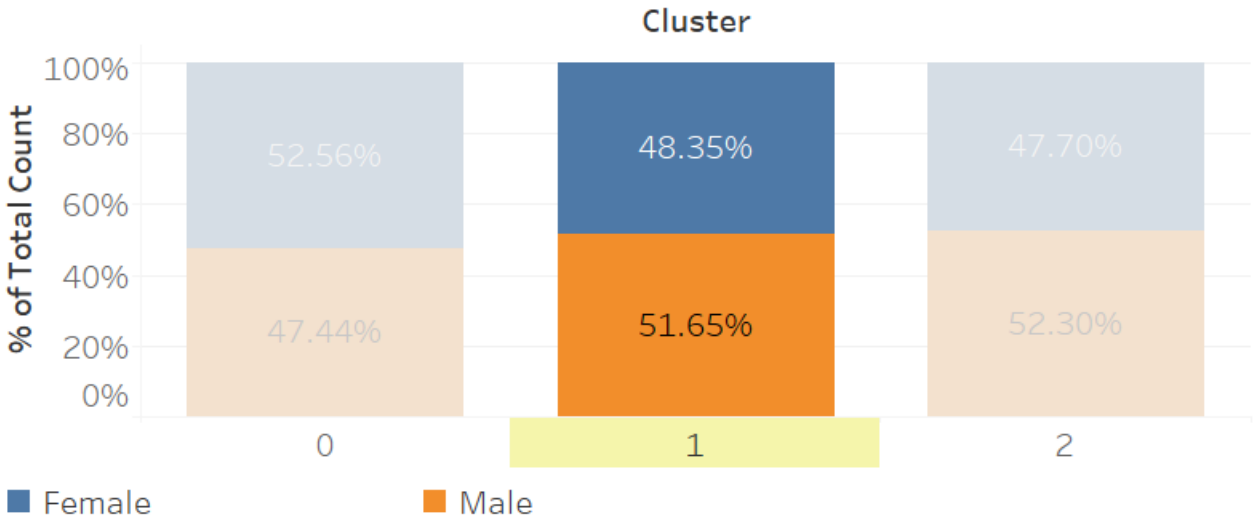


Service Data of Churn Cluster

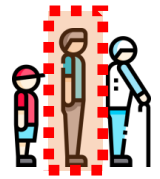


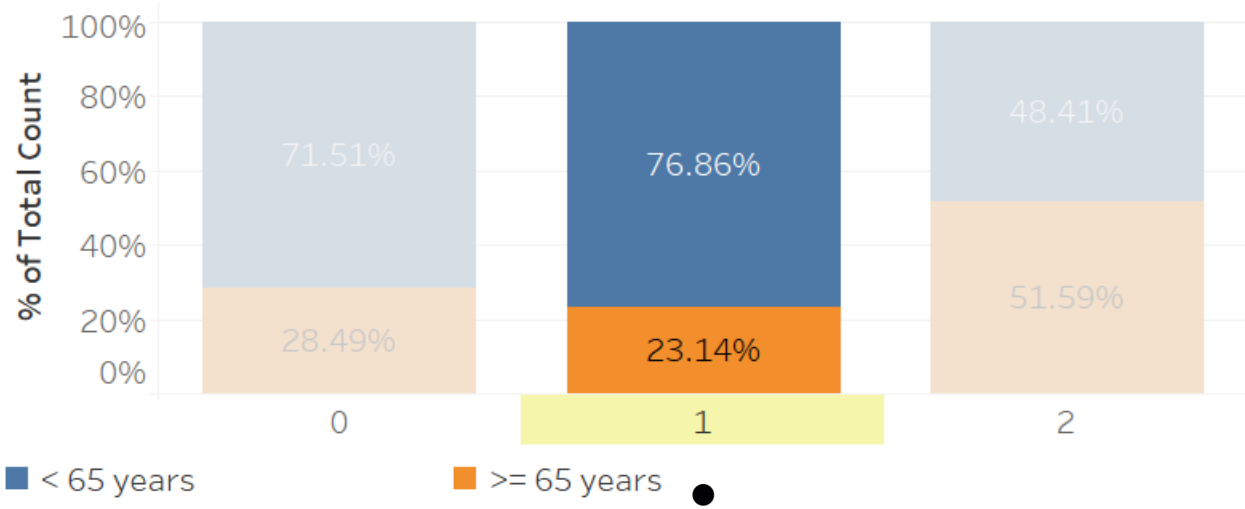
Demographics Data of Churn Cluster

Gender

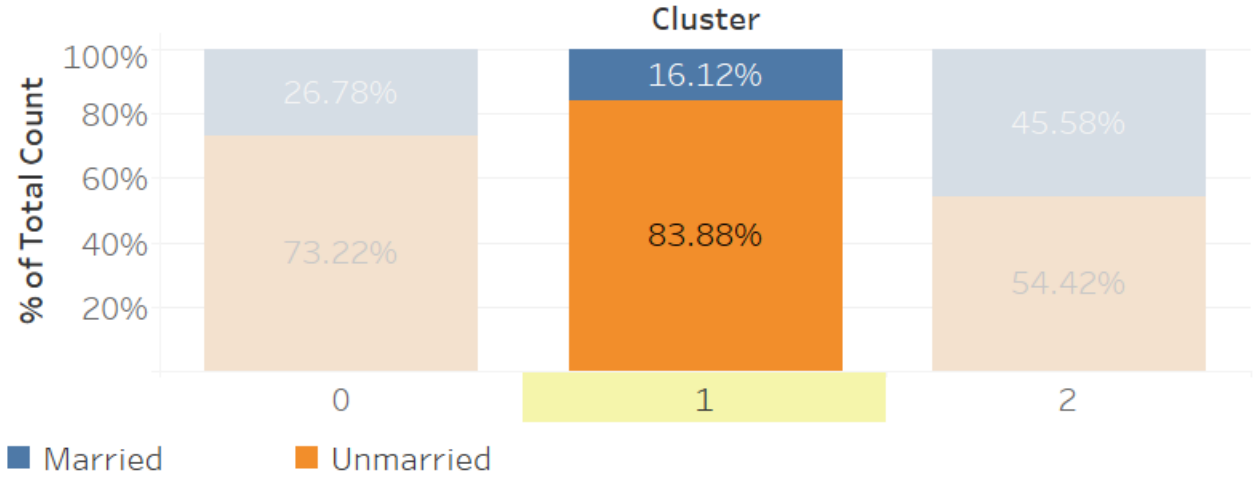


Senior Citizen

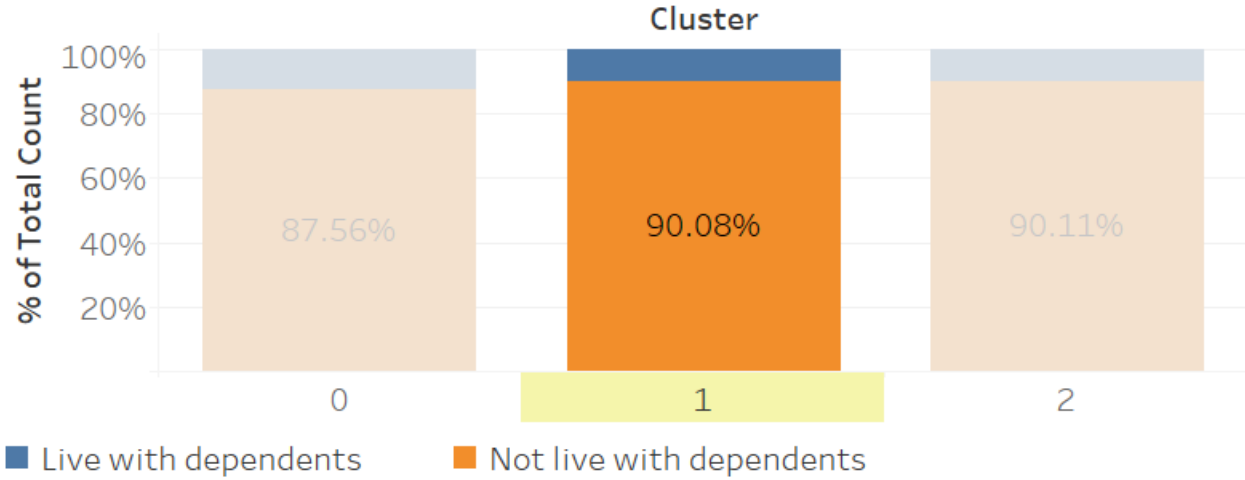




Partner

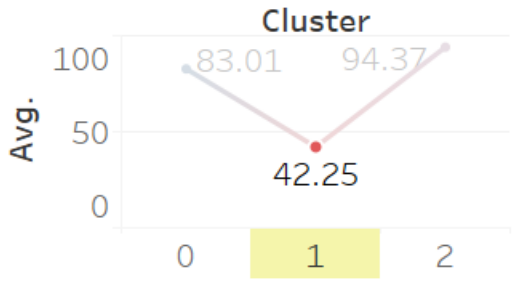


Dependent

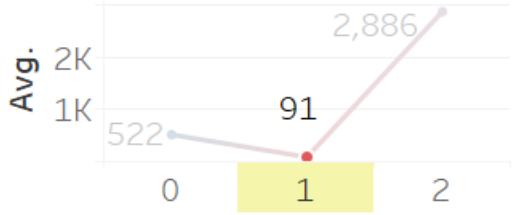


Service Data of Churn Cluster

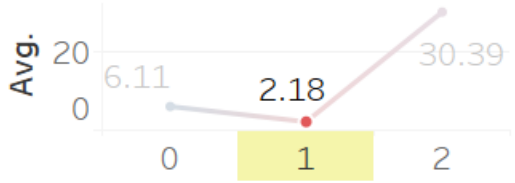
Monthly Charges



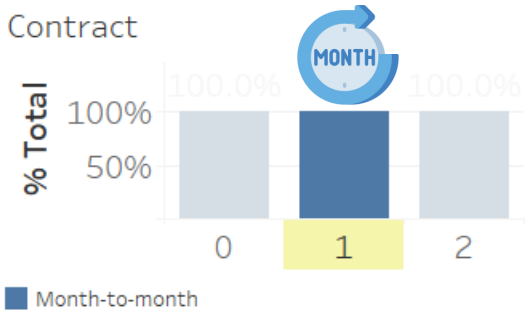
Total Charges



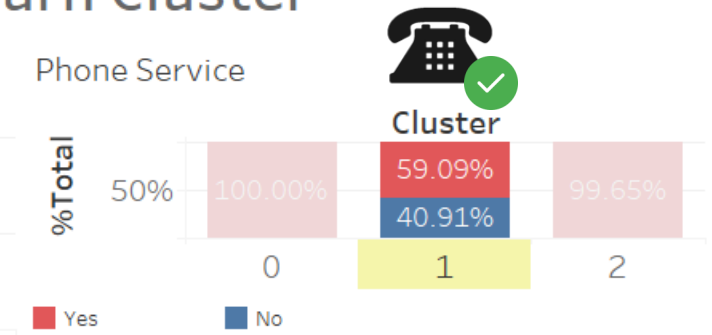
Tenure



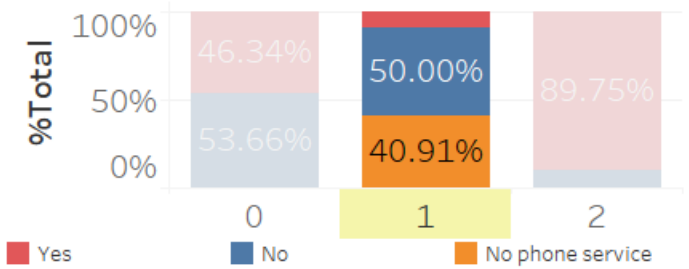
Contract



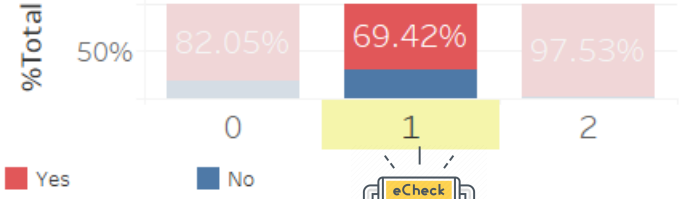
Phone Service



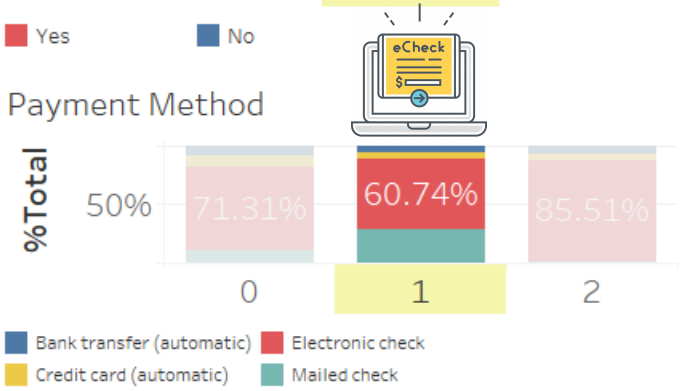
Multiple Line



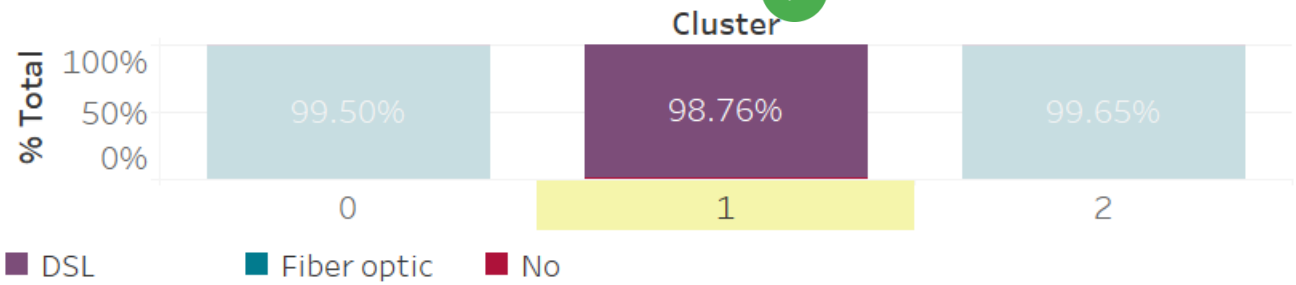
Paperless Billing



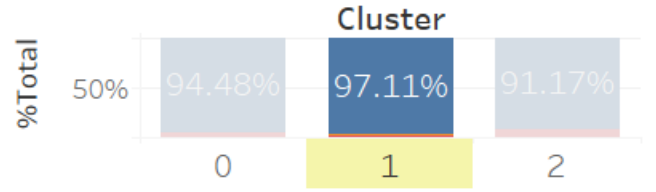
Payment Method



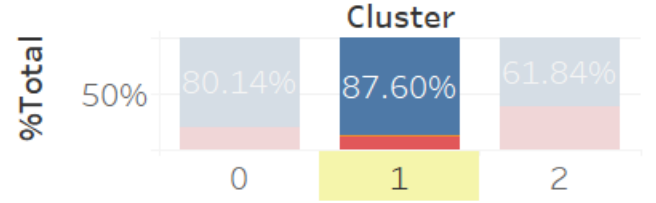
Internet Service



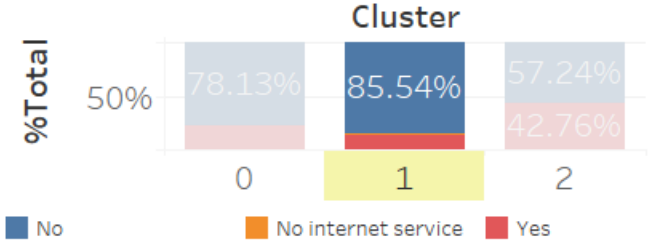
Online Security



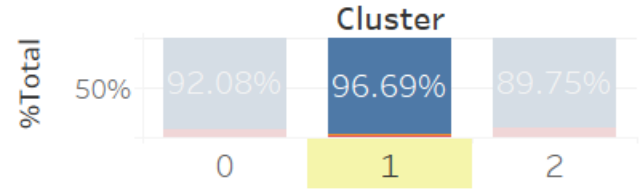
Online Backup



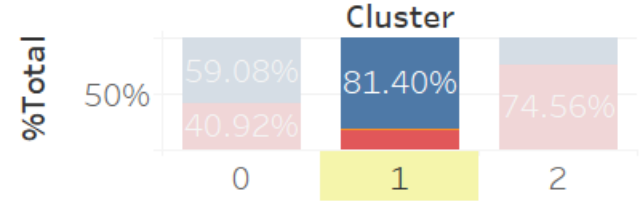
Device Protection



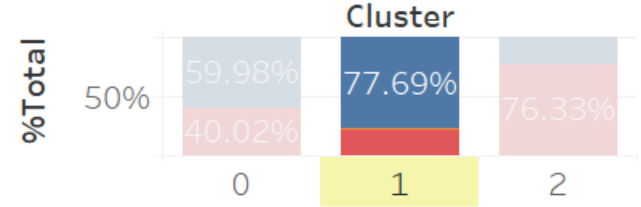
Tech Support



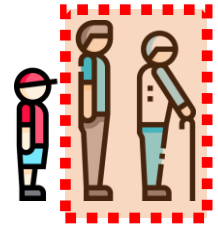
Streaming TV



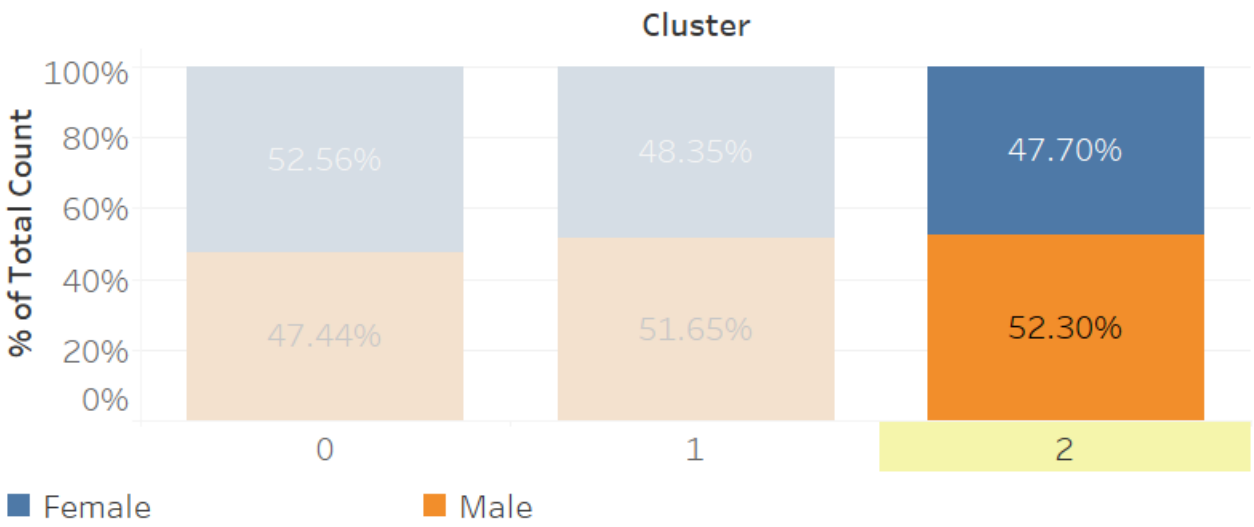
Streaming Movies



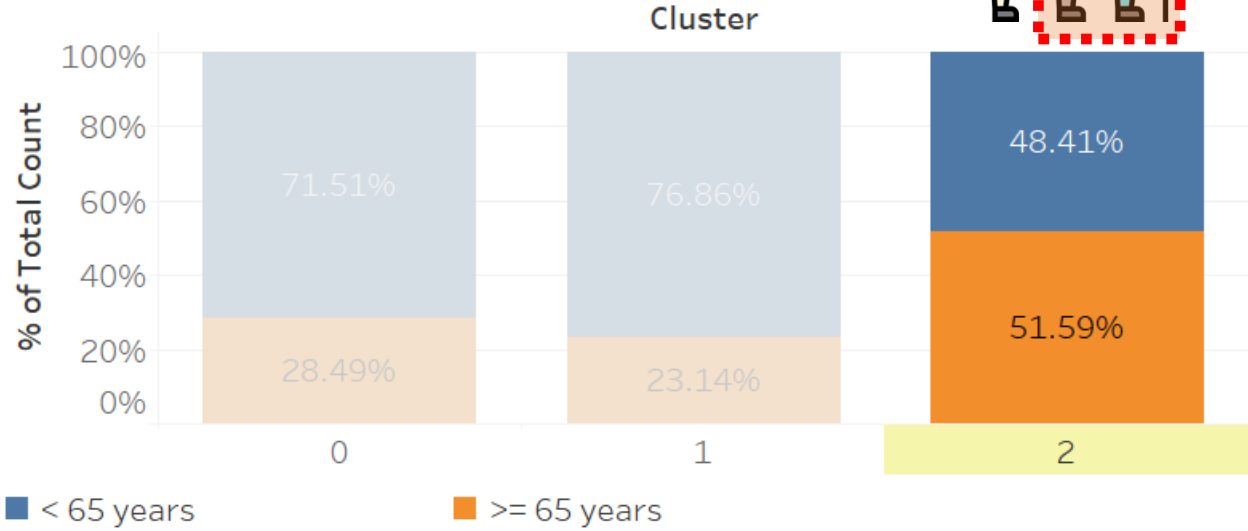
Demographics Data of Churn Cluster



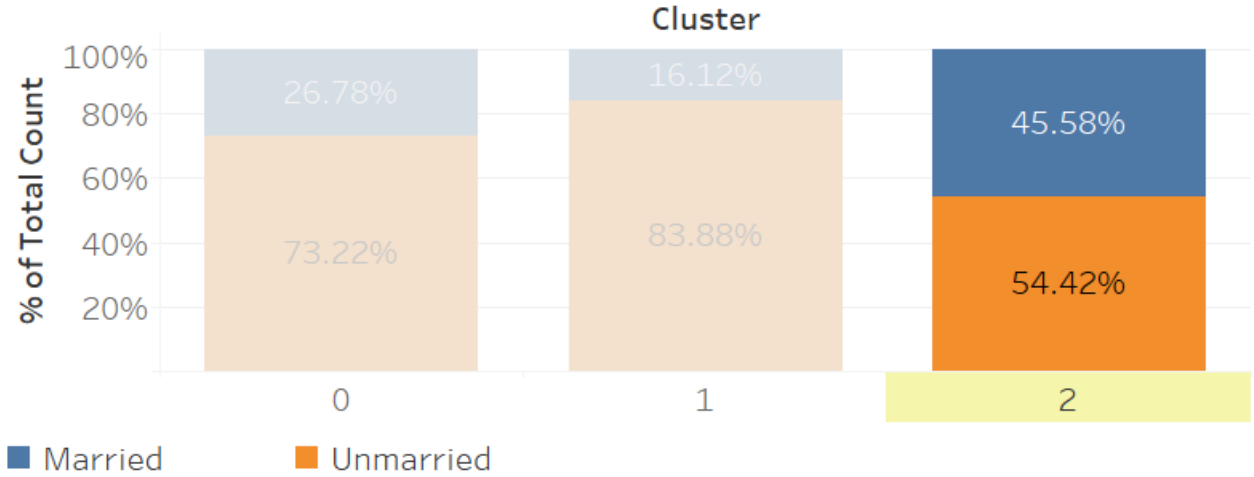
Gender



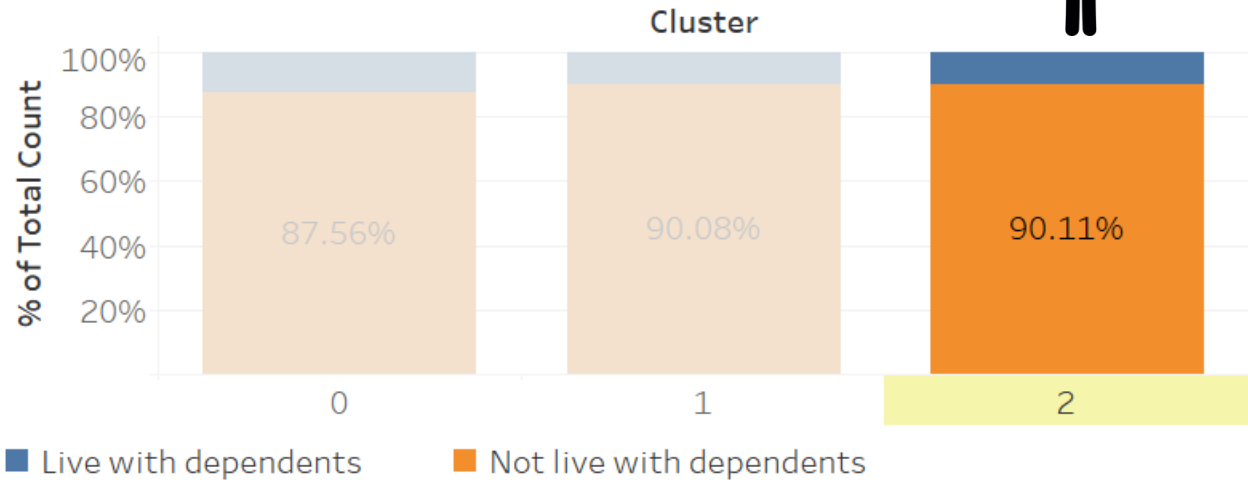
Senior Citizen



Partner

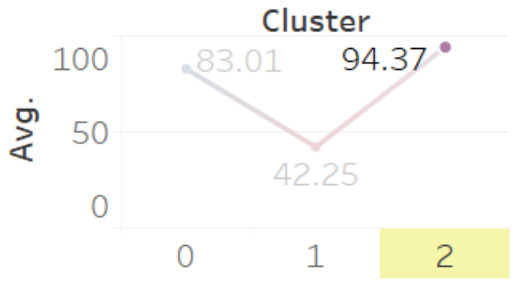


Dependent

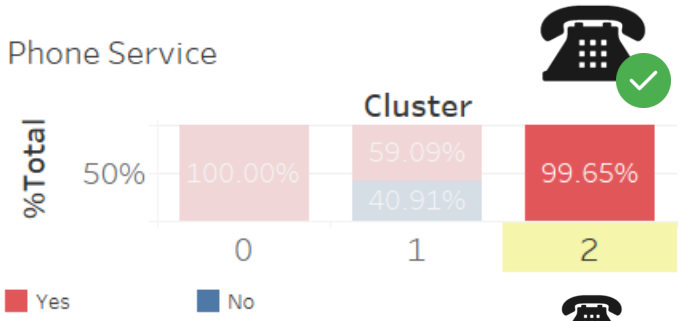


Service Data of Churn Cluster

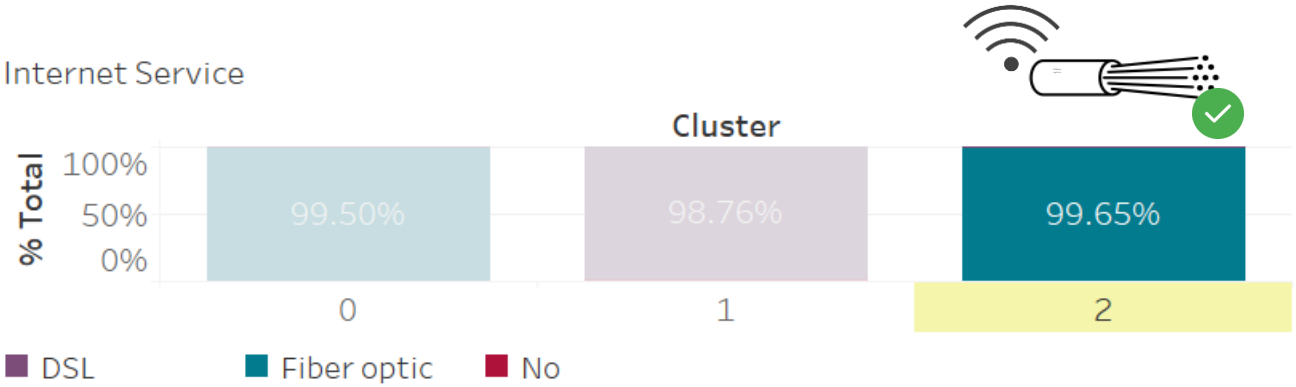
Monthly Charges



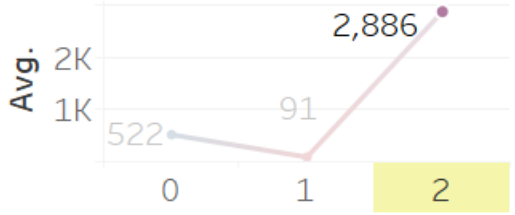
Phone Service



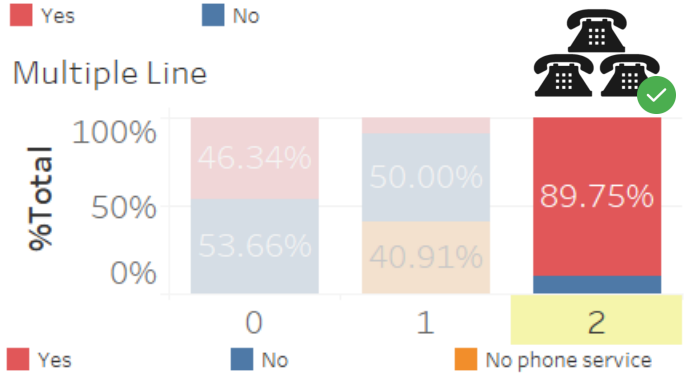
Internet Service



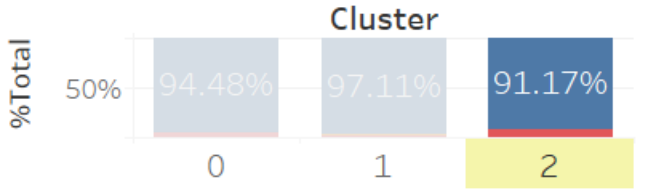
Total Charges



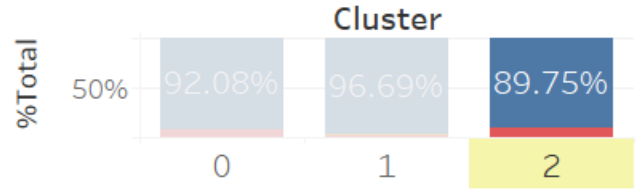
Multiple Line



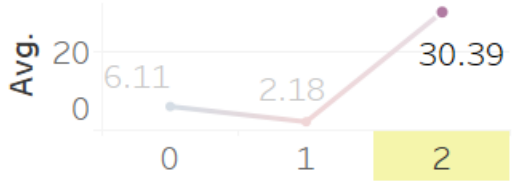
Online Security



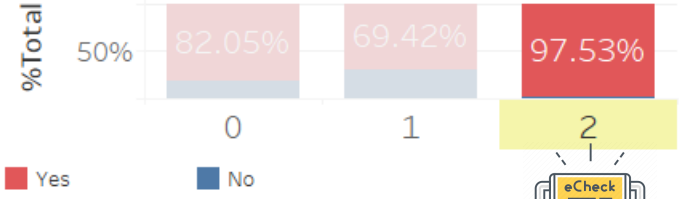
Tech Support



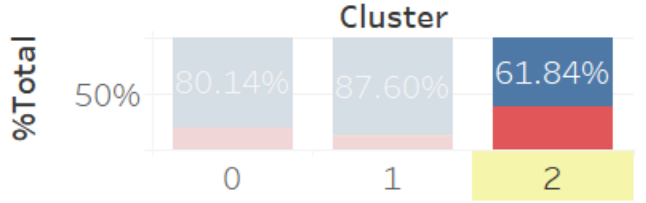
Tenure



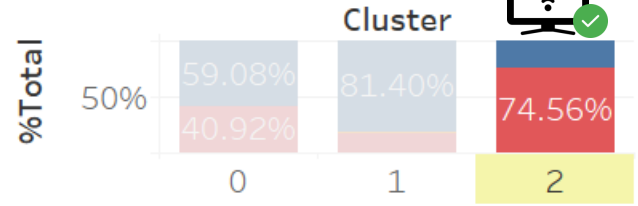
Paperless Billing



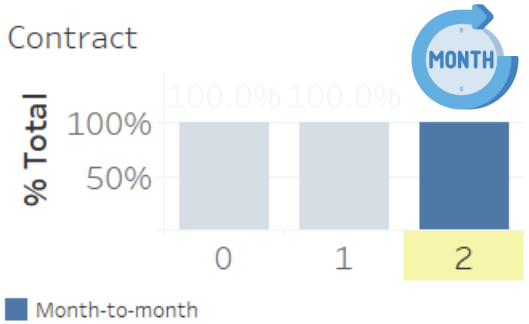
Online Backup



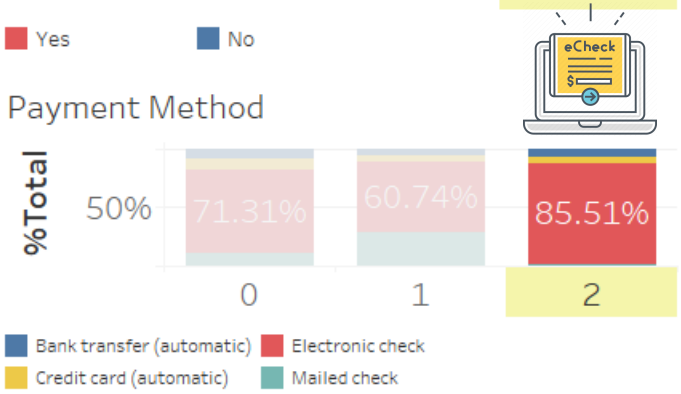
Streaming TV



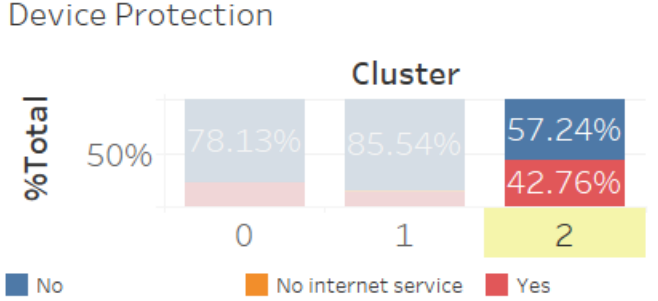
Contract



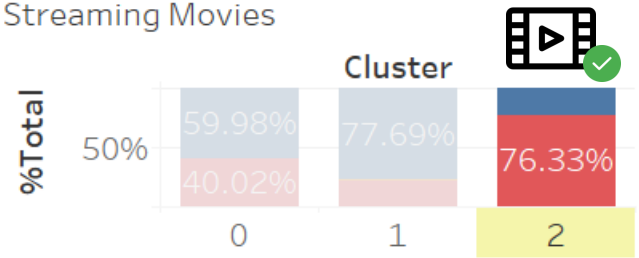
Payment Method



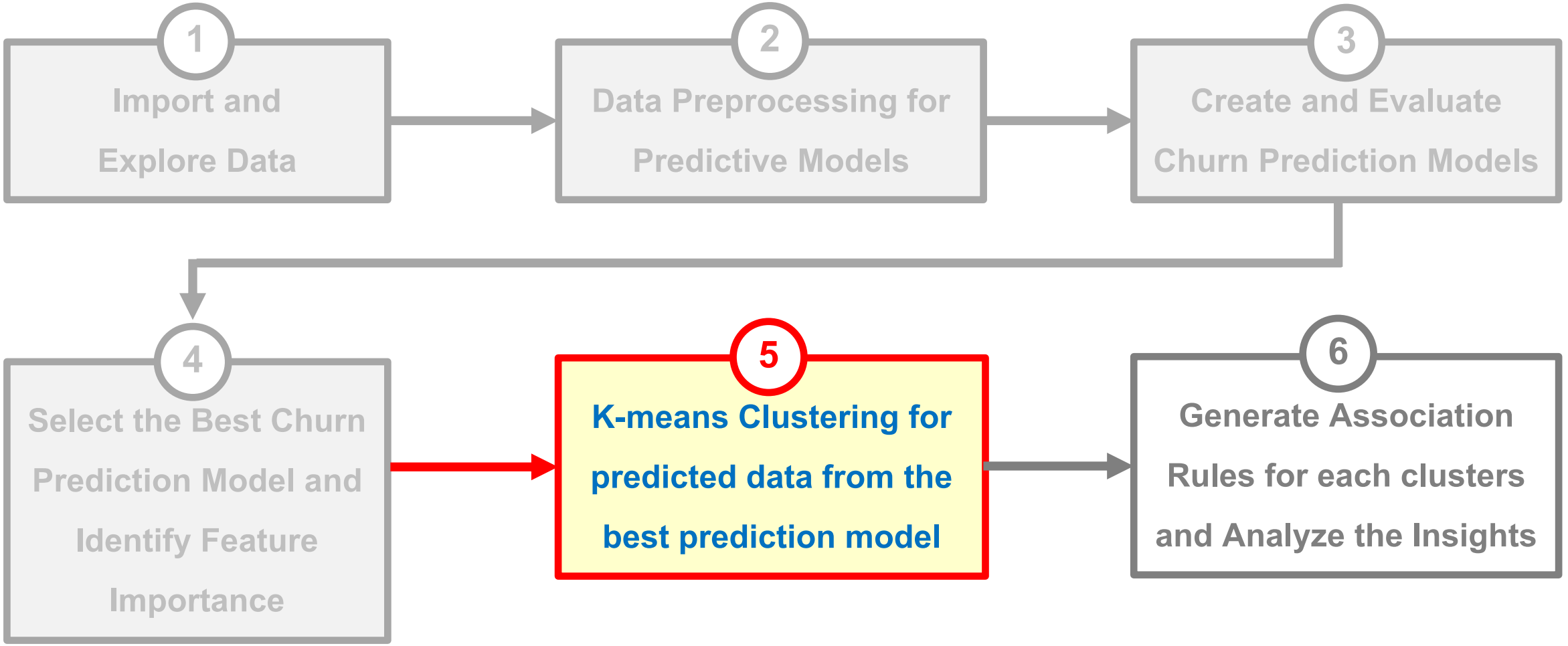
Device Protection



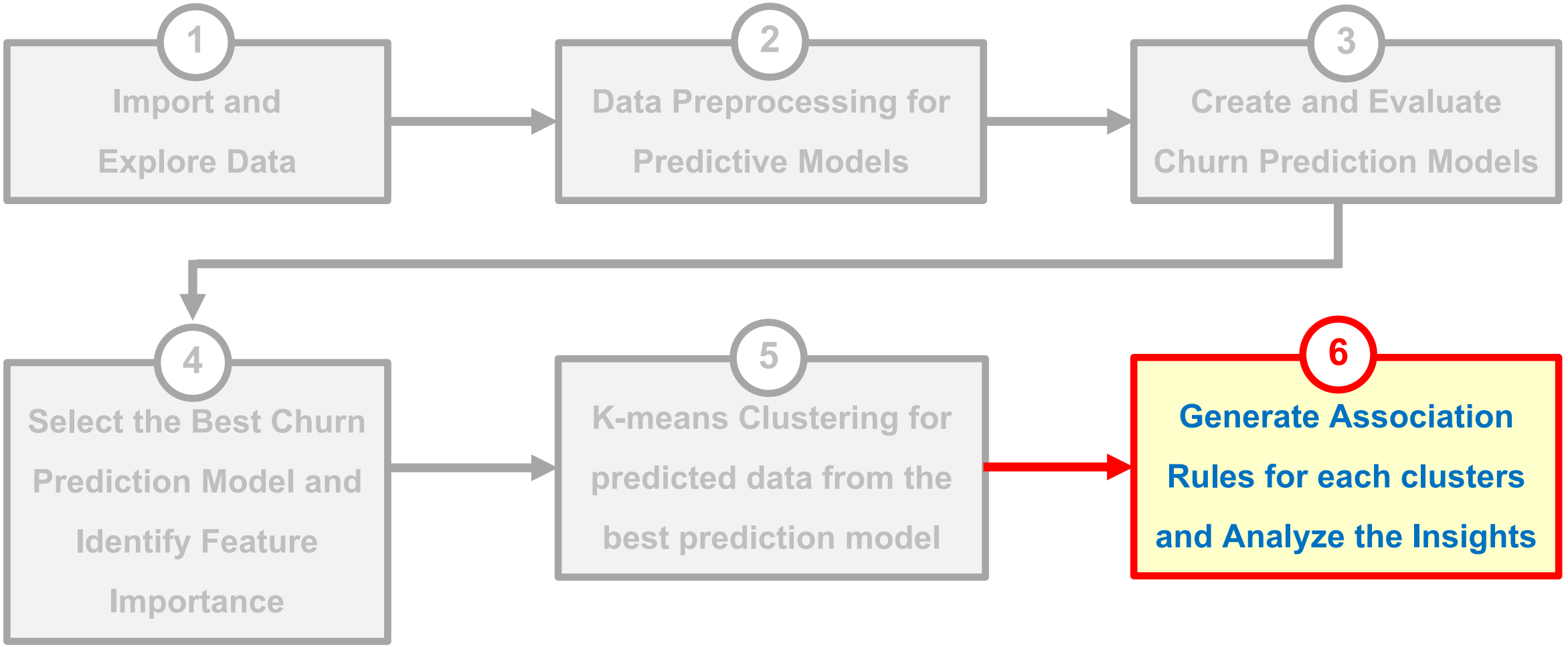
Streaming Movies



Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

Procedure


1
Import and Explore Data
2
Data Preprocessing for Predictive Models
3
Create and Evaluate Churn Prediction Models
4
Select the Best Churn Prediction Model and Identify Feature Importance
5
K-means Clustering for predicted data from the best prediction model
6
Generate Association Rules for each clusters and Analyze the Insights

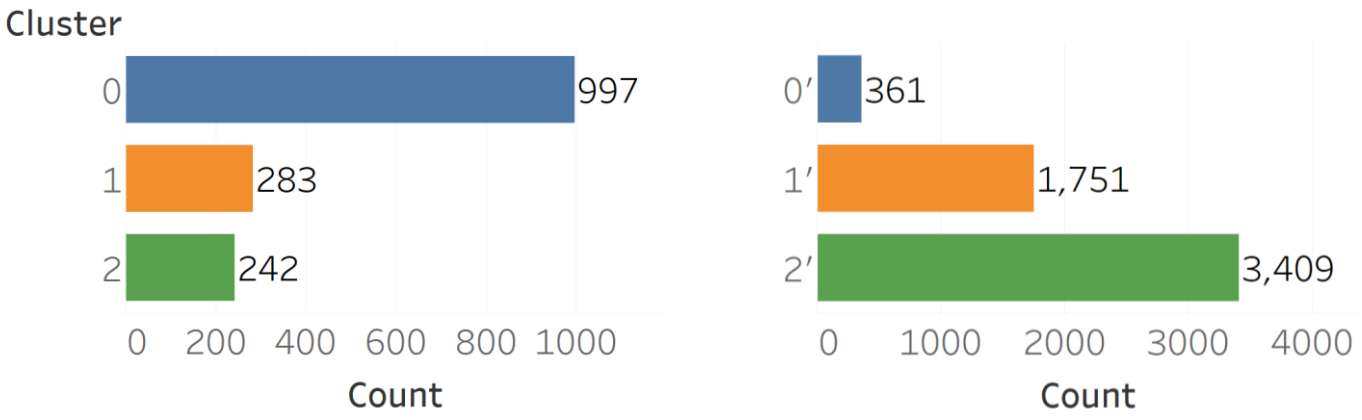
6 Generate Association Rules and Analyze the Insights

Assigning minimum thresholds

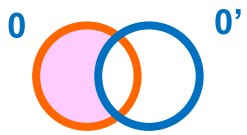
- Minimum Support:

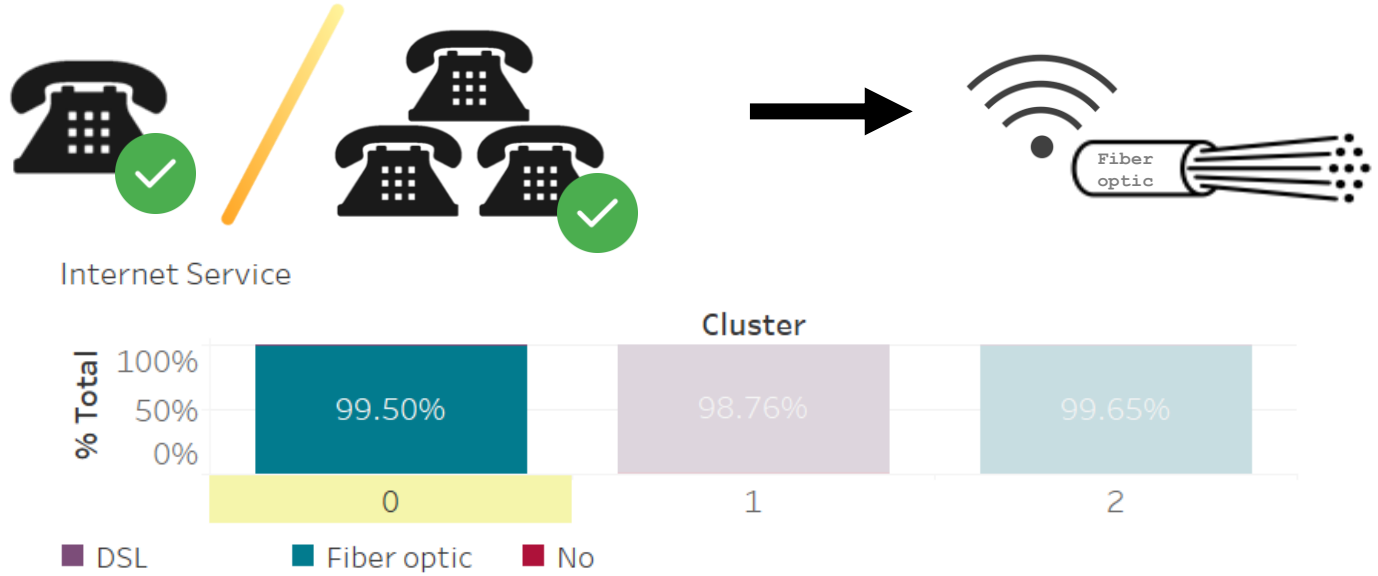
#Customers in cluster	Minimum Support
<= 150	0.9
151 – 300	0.8
301 – 450	0.7
451 – 600	0.6
601 – 750	0.5
751 – 900	0.4
> 900	0.3

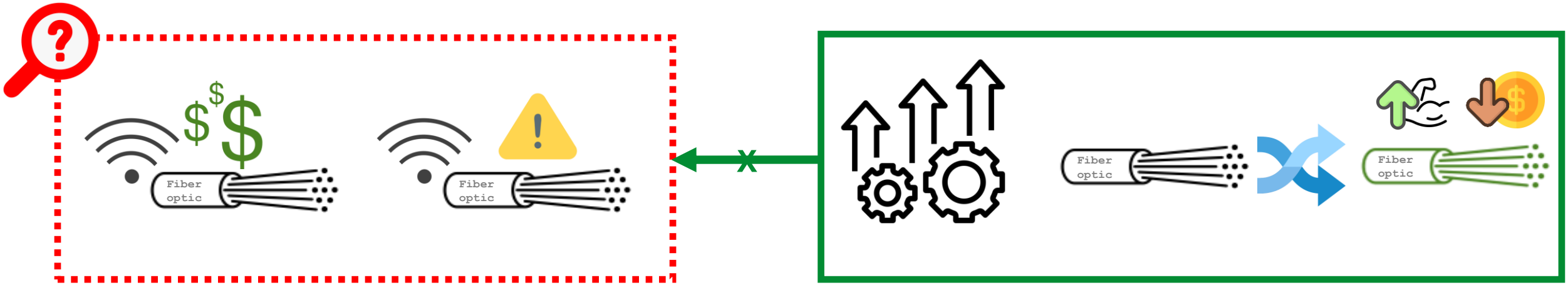
- Minimum Confidence = 0.8
- Lift > 1



Cluster 0


0
0'

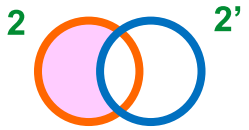

Fiber optic
Internet Service
Cluster
% Total
100%
50%
0%
99.50%
98.76%
99.65%
0
1
2
DSL
Fiber optic
No

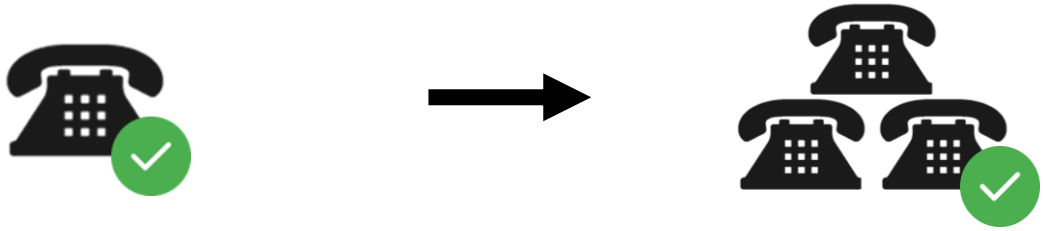

$
$
$
Fiber optic
Fiber optic
Fiber optic
Fiber optic

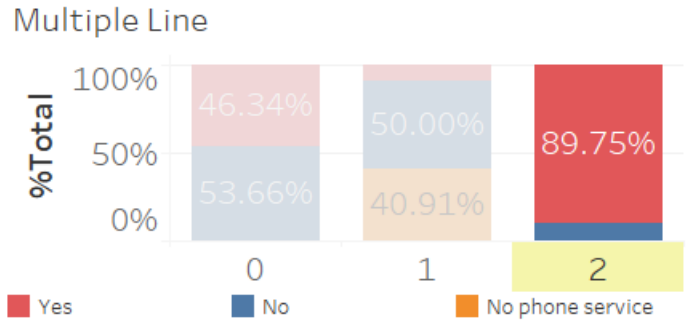
Cluster 1

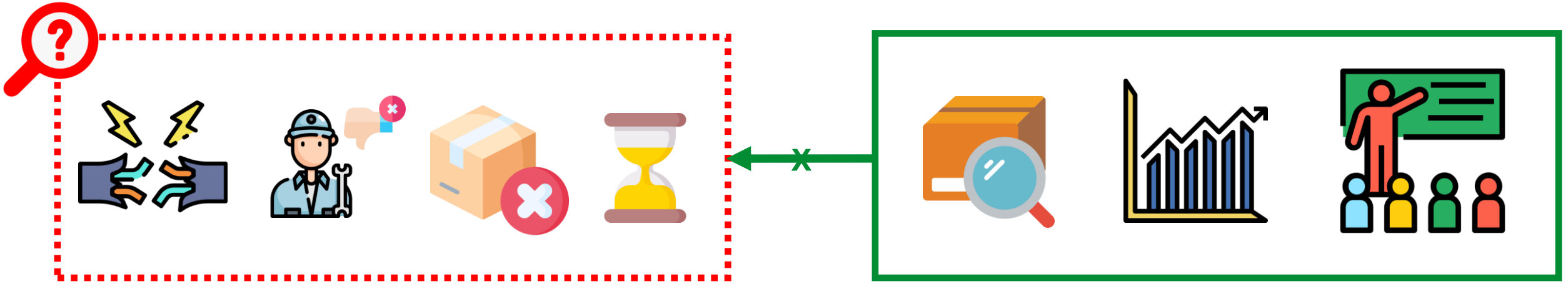
Not found any association rules

Cluster 2

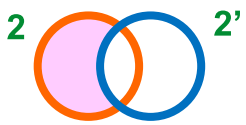

2
2'

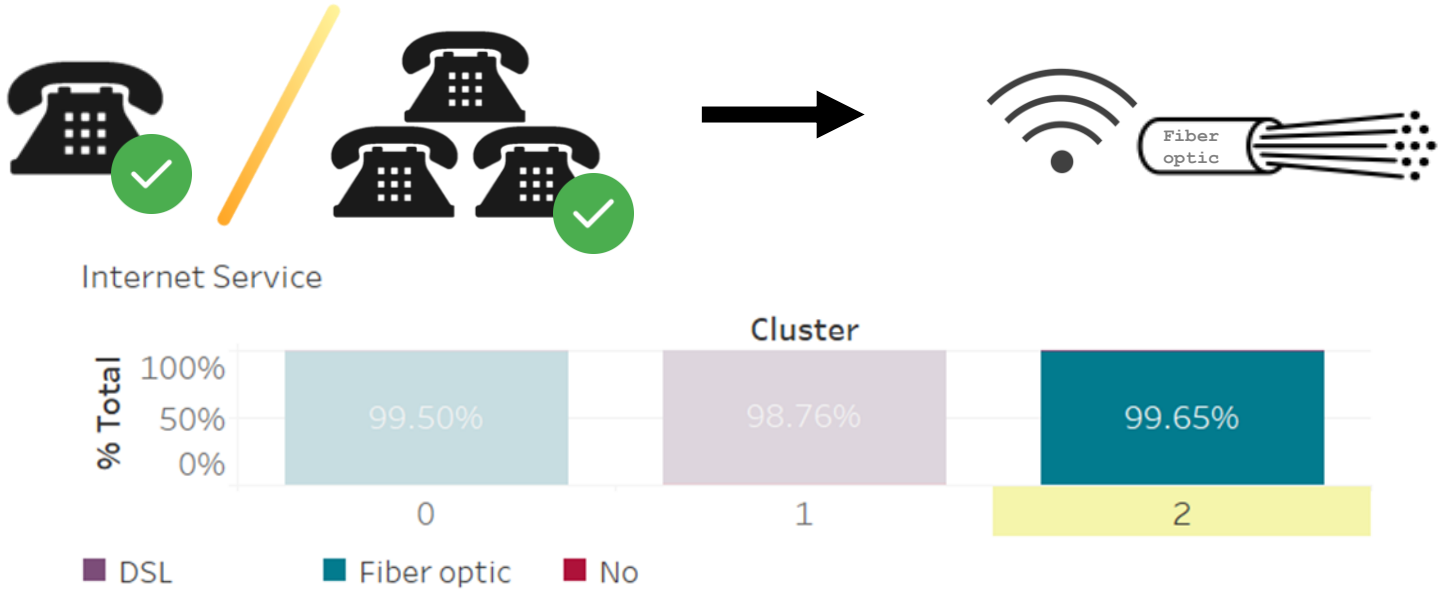


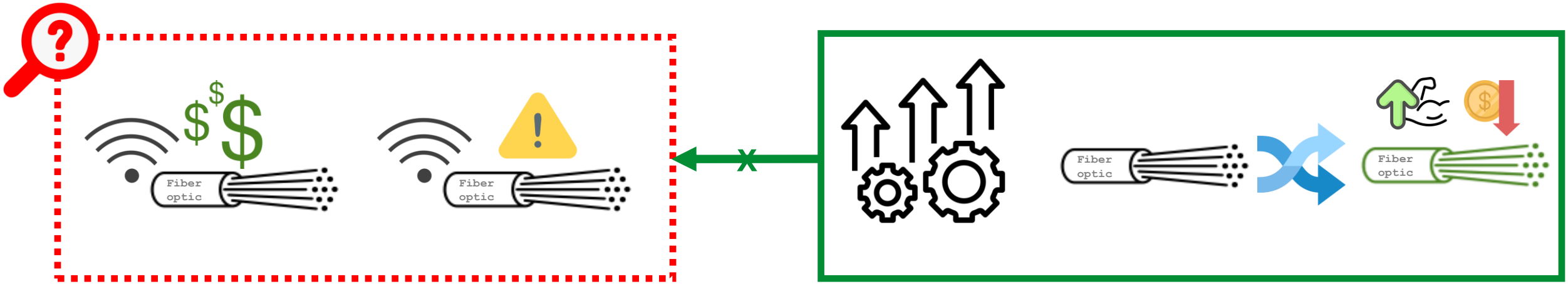

Multiple Line
%Total
100%
50%
0%
46.34%
53.66%
50.00%
40.91%
89.75%
0
1
2
Yes
No
No phone service



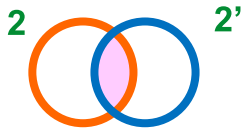
Cluster 2

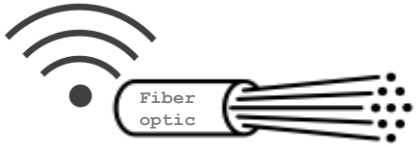

2
2'


Fiber optic
Internet Service
Cluster
% Total
100%
50%
0%
99.50%
98.76%
99.65%
0
1
2
DSL
Fiber optic
No

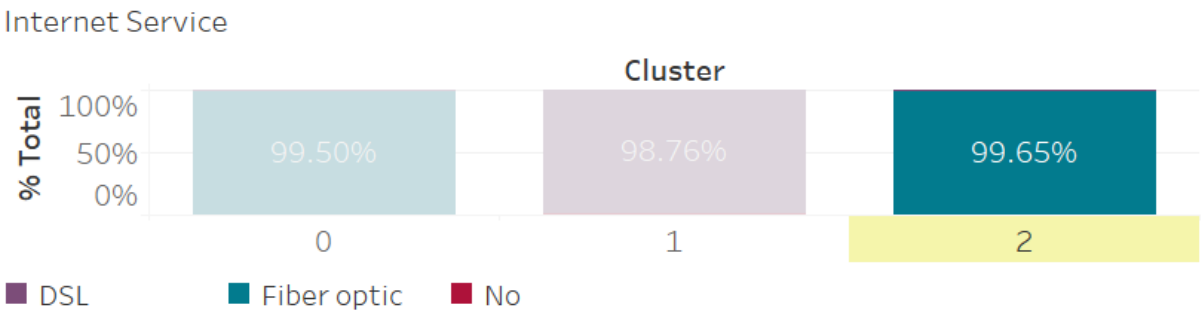

Fiber optic
Fiber optic
Fiber optic
Fiber optic

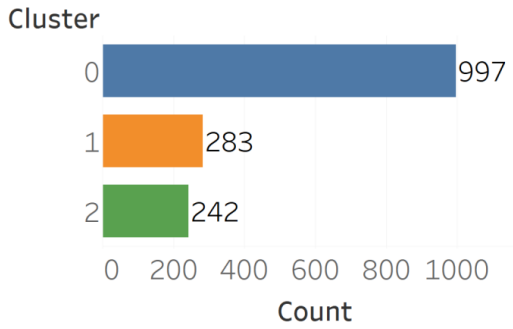
Cluster 2

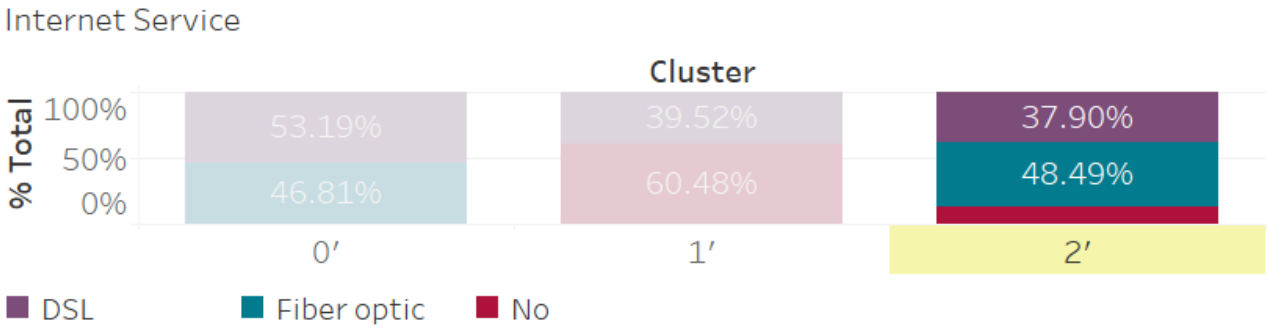

2
2'

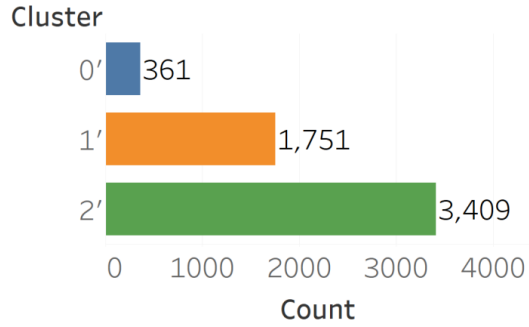

Fiber
optic


%
SALE

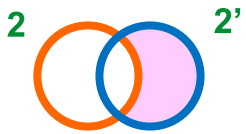

Internet Service
Cluster
% Total
100%
50%
0%
99.50%
98.76%
99.65%
0
1
2
DSL
Fiber optic
No

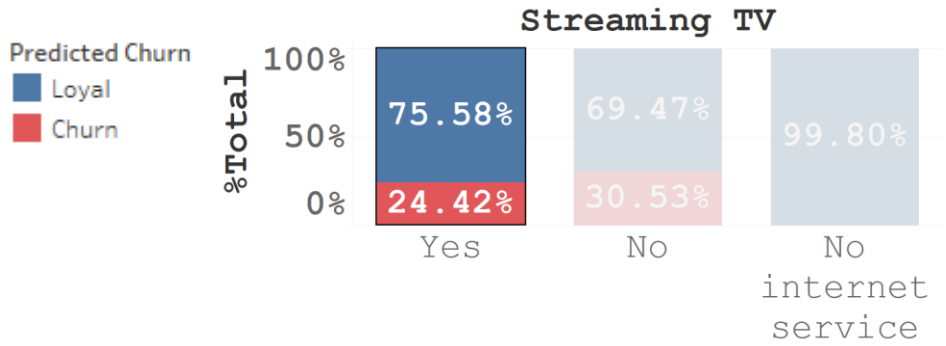

Cluster
0
1
2
997
283
242
0
200
400
600
800
1000
Count

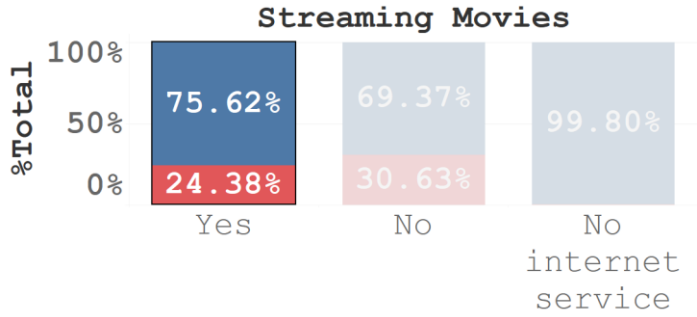

Internet Service
Cluster
% Total
100%
50%
0%
53.19%
46.81%
39.52%
60.48%
37.90%
48.49%
0'
1'
2'
DSL
Fiber optic
No

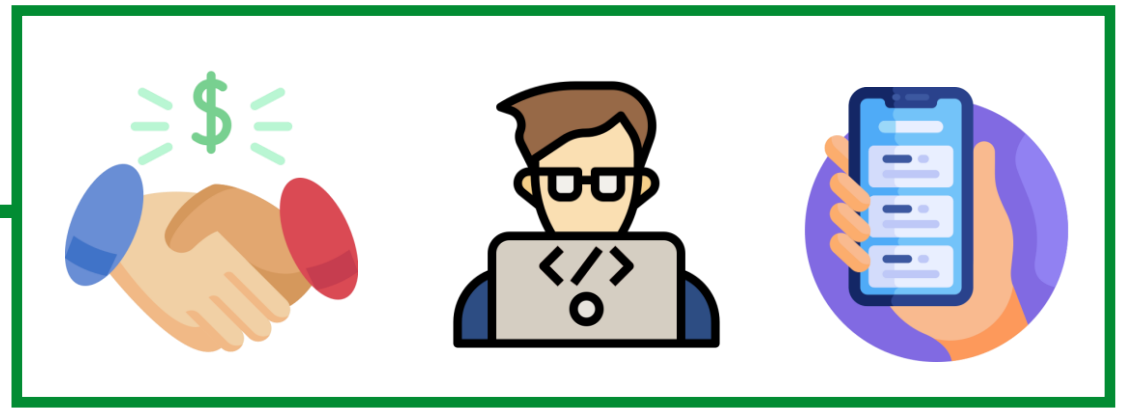

Cluster
0'
1'
2'
361
1,751
3,409
0
1000
2000
3000
4000
Count

Cluster 2

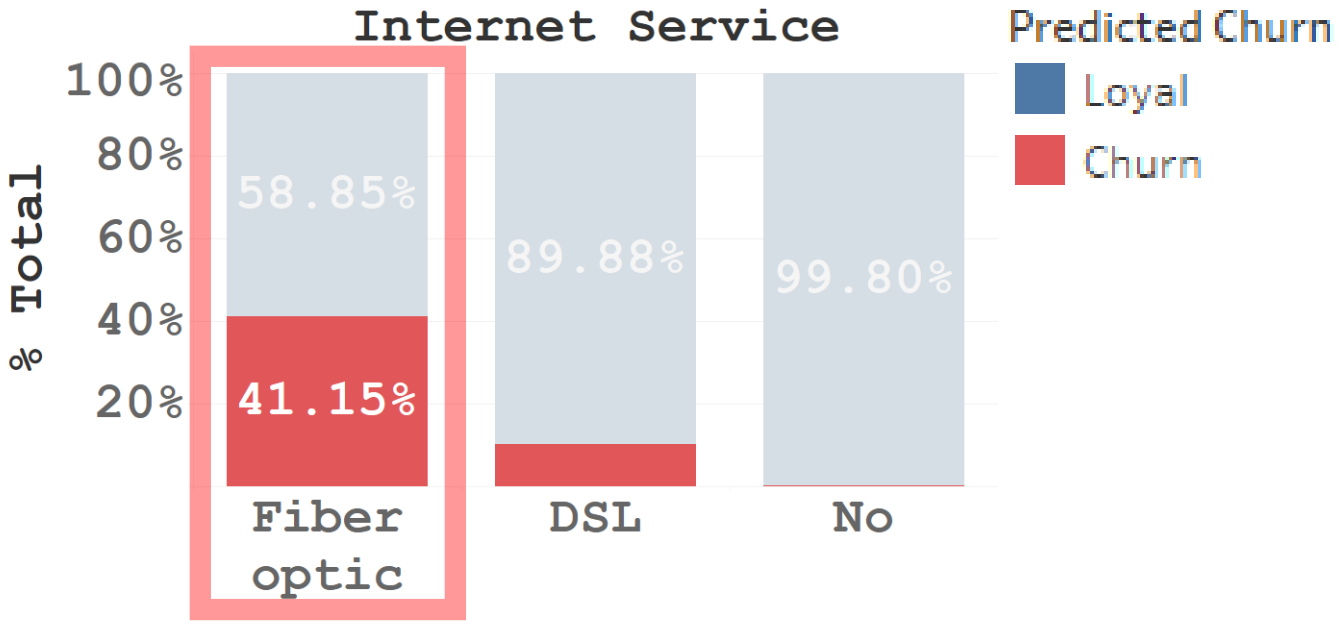

2
2'

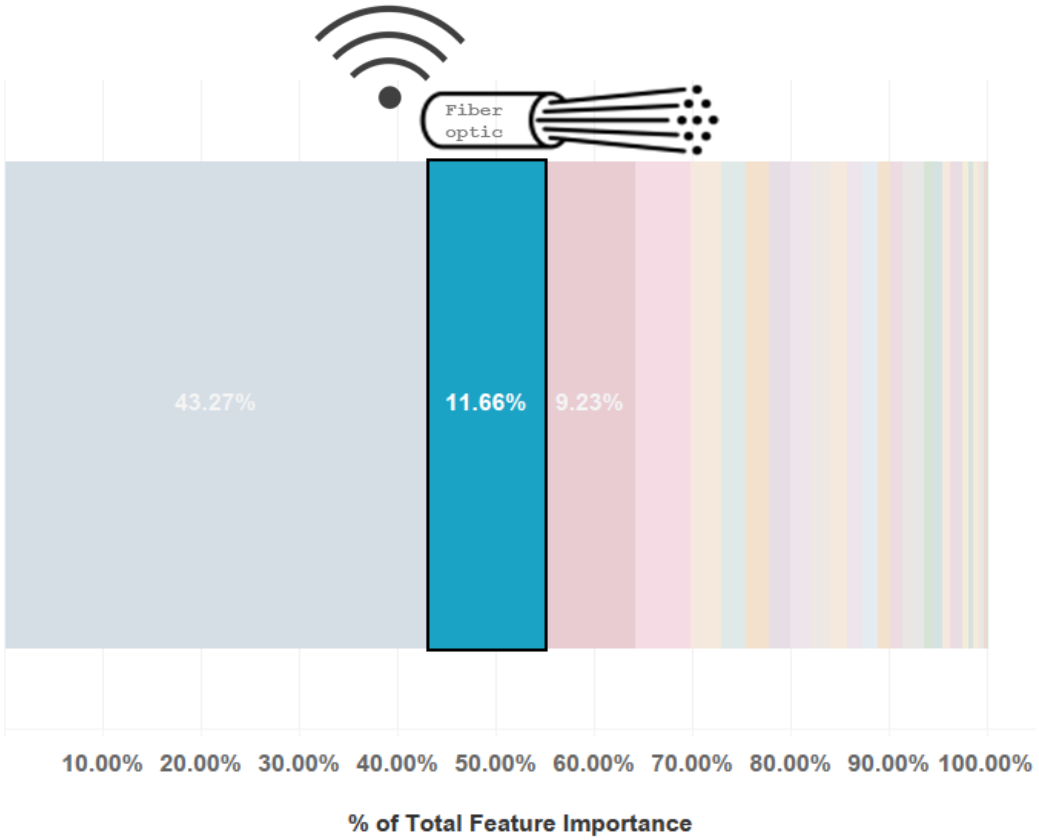

Streaming TV
Predicted Churn
Loyal
Churn
%Total
100%
50%
0%
75.58%
24.42%
69.47%
30.53%
99.80%
Yes
No
No internet service

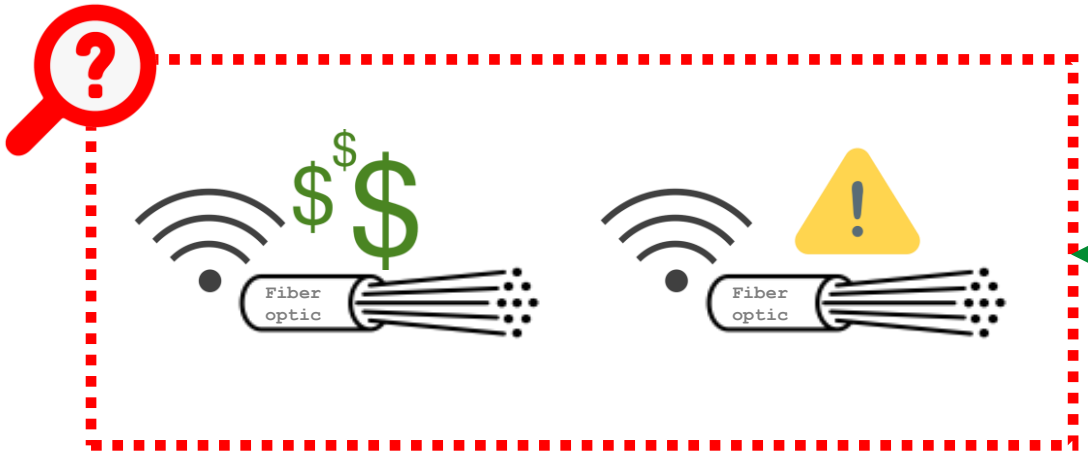

Streaming Movies
%Total
100%
50%
0%
75.62%
24.38%
69.37%
30.63%
99.80%
Yes
No
No internet service

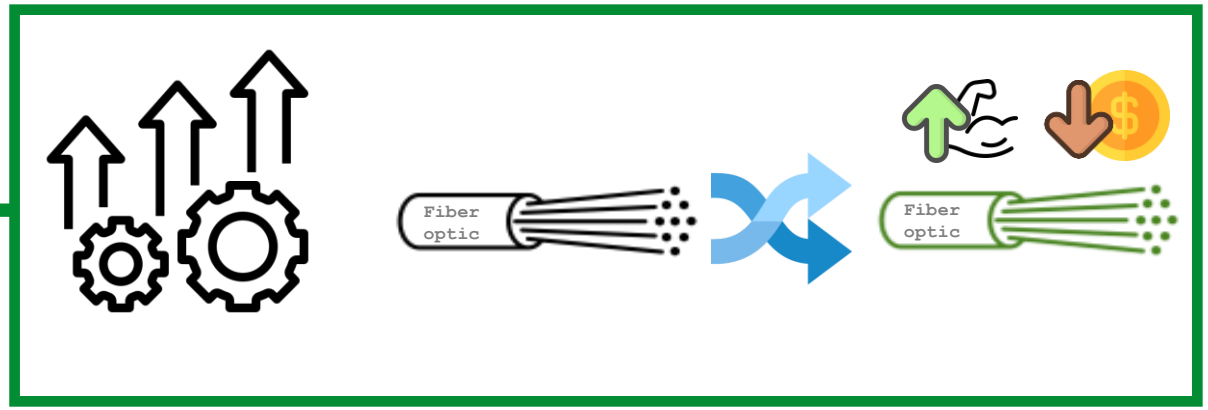


Wrap-up: Retention Strategies

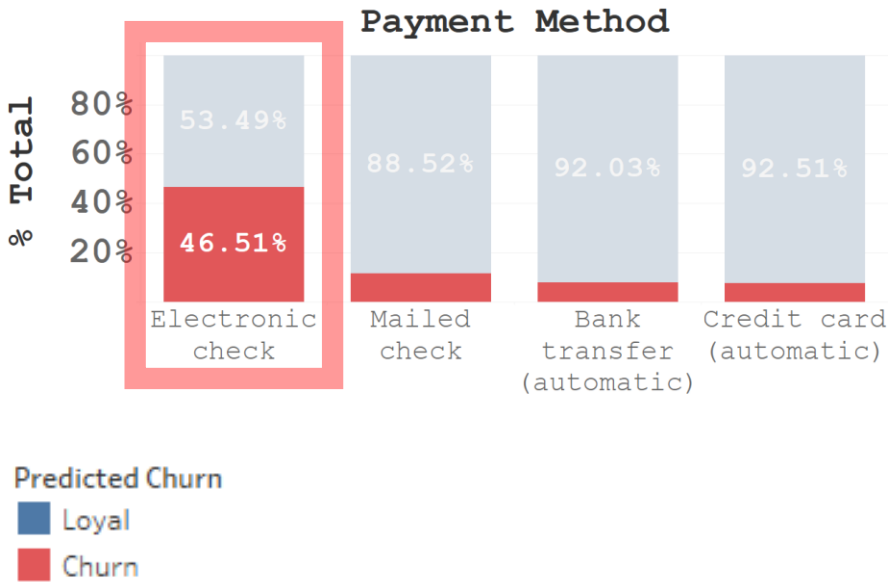




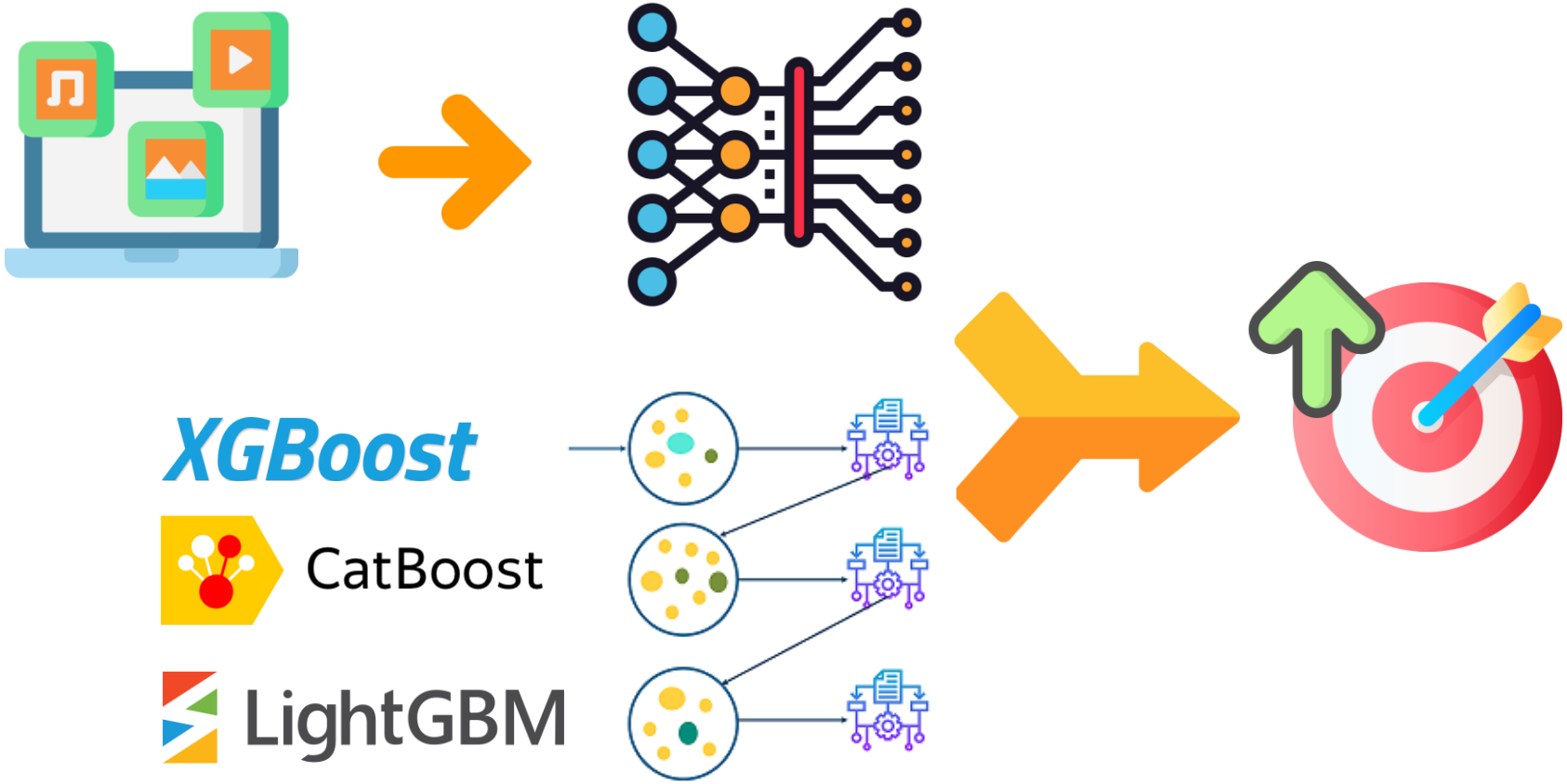




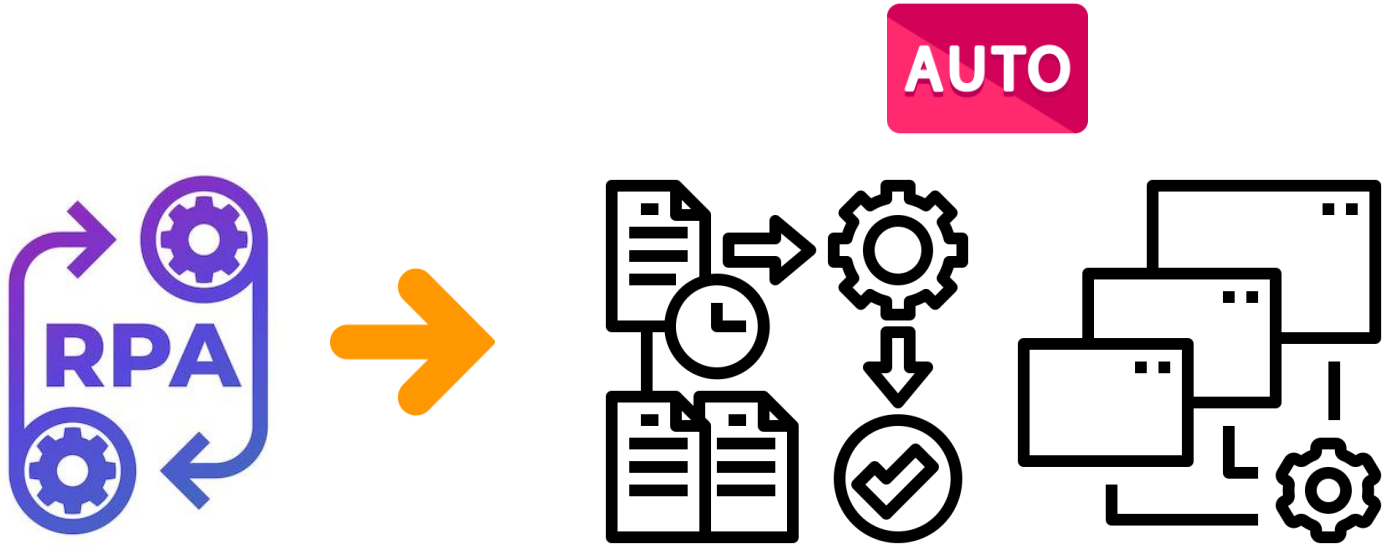
Wrap-up: Retention Strategies



Future Work



Future Work



CHULA ENGINEERING

Innovation toward Sustainability | ACTNOW