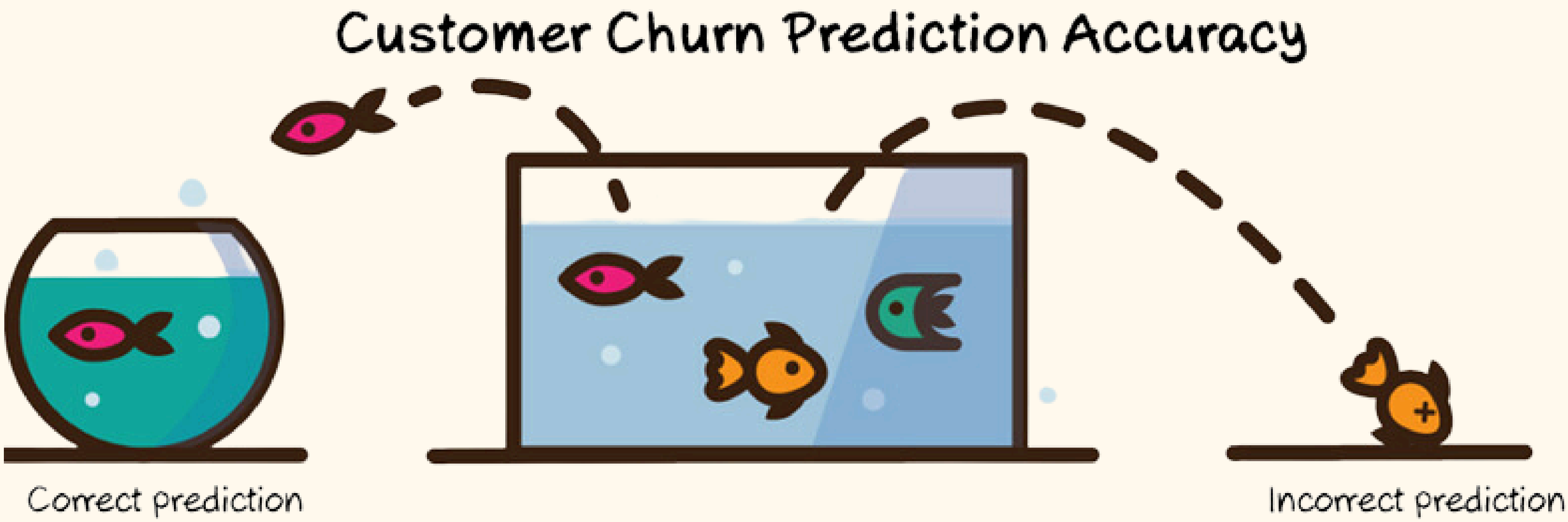
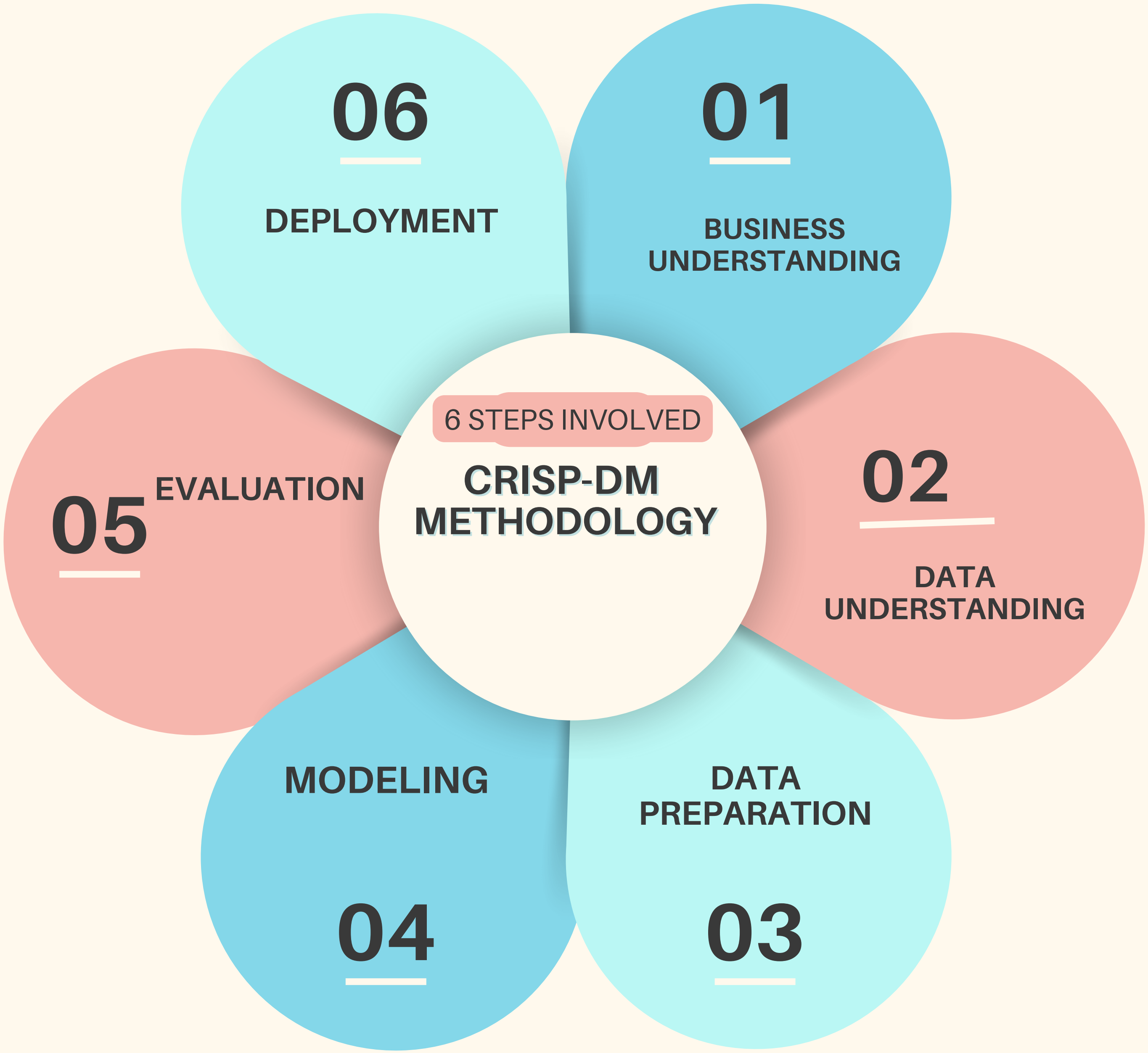
SUNBORN



LOYCE TSUMA

github : she-loycetsuma


6 STEPS INVOLVED
CRISP-DM METHODOLOGY
01
BUSINESS UNDERSTANDING
02
DATA UNDERSTANDING
03
DATA PREPARATION
04
MODELING
05
EVALUATION
06
DEPLOYMENT

OVERVIEW

02

Why?

- Understanding churn helps in implementing targeted retention strategies, ultimately improving customer loyalty and revenue.

01

Business Problem

- Predict customer churn in rental yacht services to enhance retention strategies for luxury real estate and maritime hospitality developers.

THE DATA

CustomerID
Unique identifier for each customer

Location

State in the USA

Name

Client name

Subscription Length_Months

Duration of yacht usage

CHURN
1 = YES
0 = NO

Age

Client age (minimum 18 years).

Monthly_Bill

Monthly rent charged

Gender

Male or Female

Total_Usage_GB

Total amount in GBP

INFERENTIAL STATICTICS

01

- Data cleaning
- Checking for duplicate entries
- Checking for missing values
- Checking data Structure

02

- Inferential Statistics
- Univariate Analysis
- Bivariate Analysis

03

- Checking Correlation of features
- Log Transformations of features

04

- Data Scaling
- Converting word values to numbers(OneHotEncoding)

Preparation	Decriptive statistics	Confirming Linearity	Data Normalization

Findings:

No linearity or correlation among features

AGE

- Average client age clientele is 44 years

gender

- Most clients are female at 50.2%

month

- Average monthly bill is GBP 65

CHURN

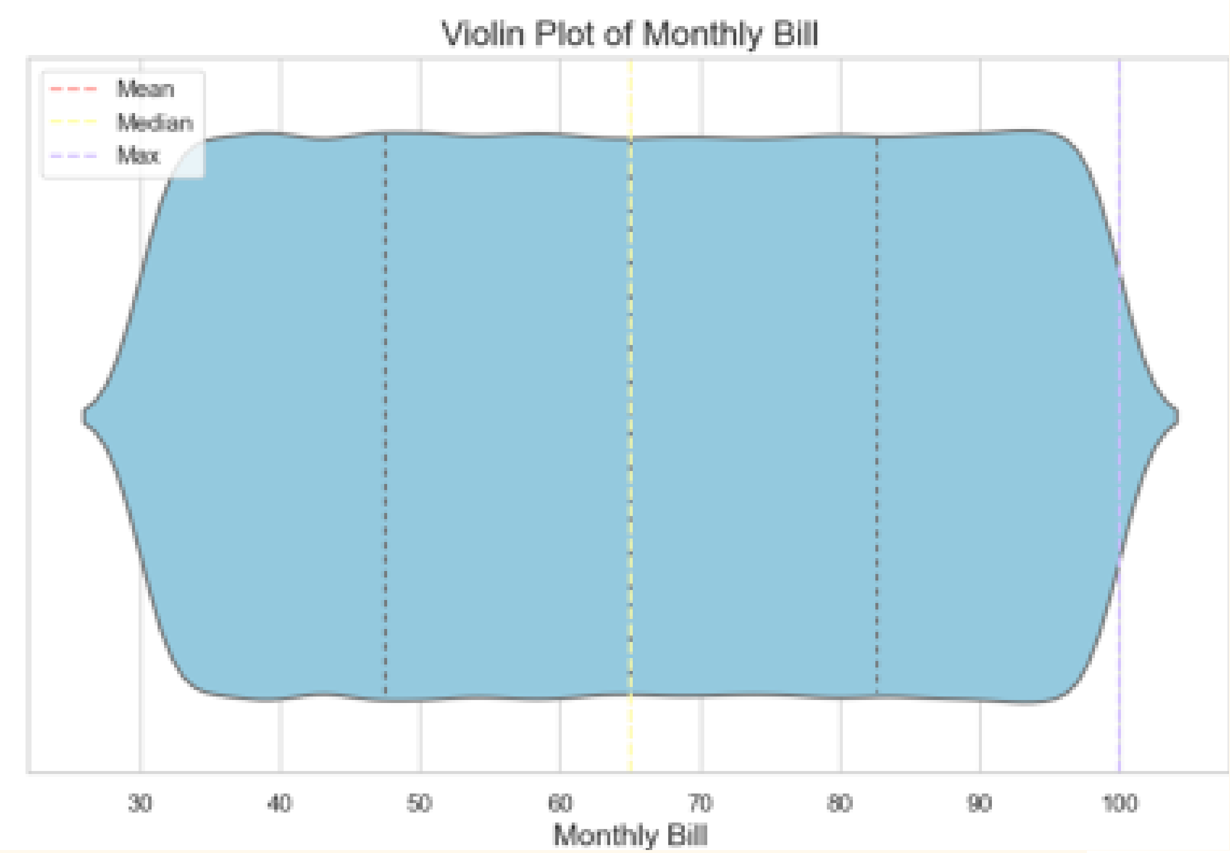
- (non-churn = 50,221 churn 49,779)- Sunborn looses almost 50% of their acquired clients

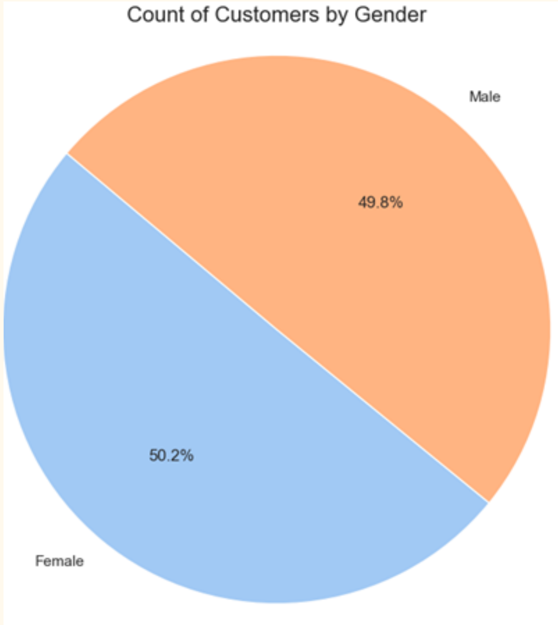
LOCATION

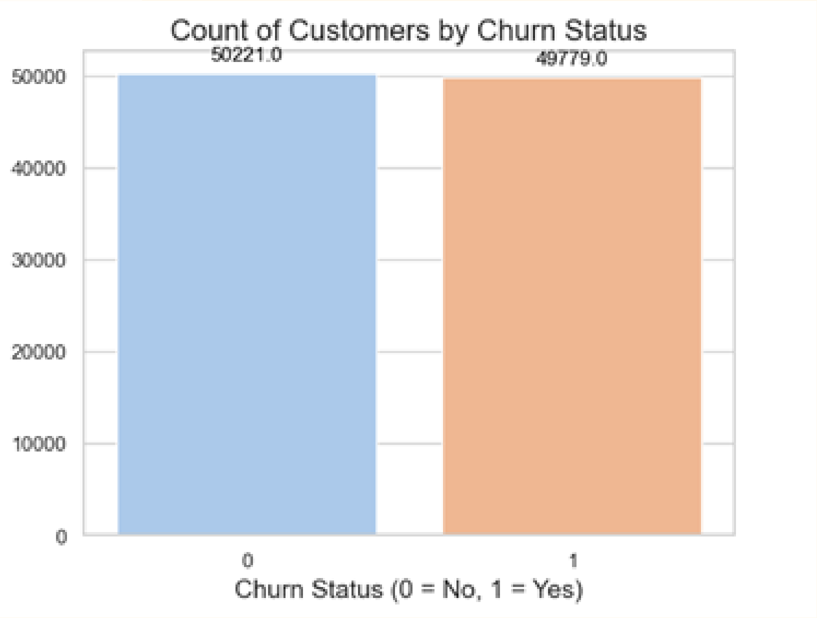
Most of the companies Clientele comes from Houston State and Los Angeles

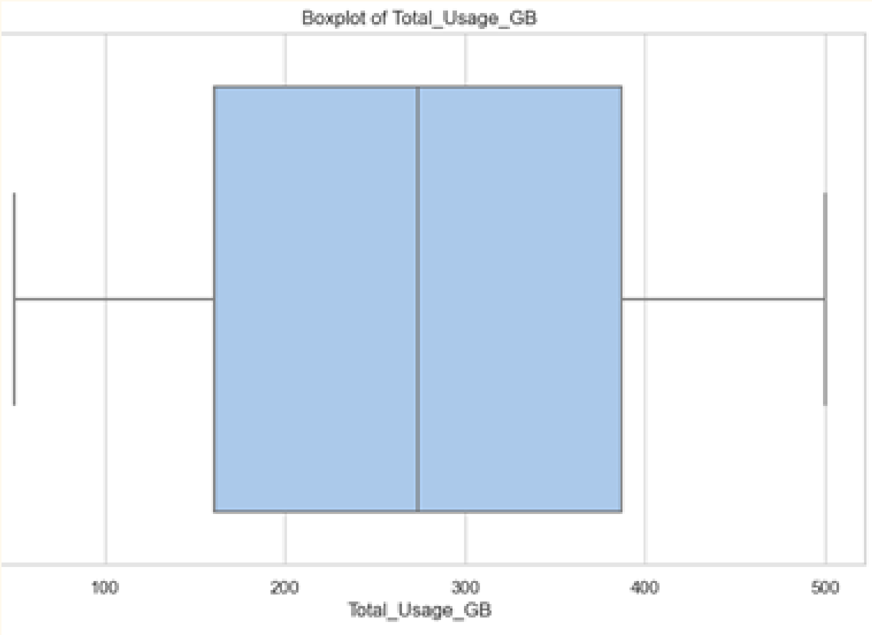
STAY

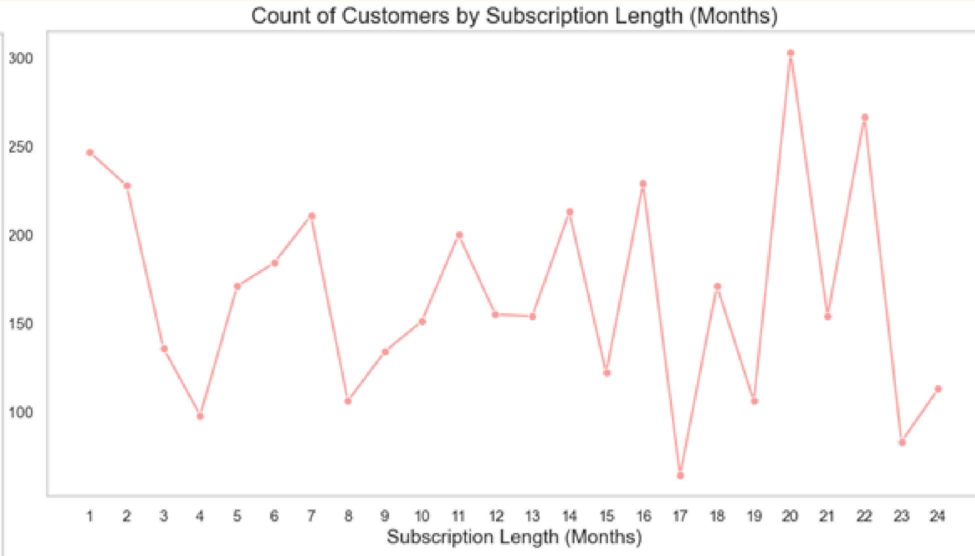
- Average client usage is 12 months max being 24 months

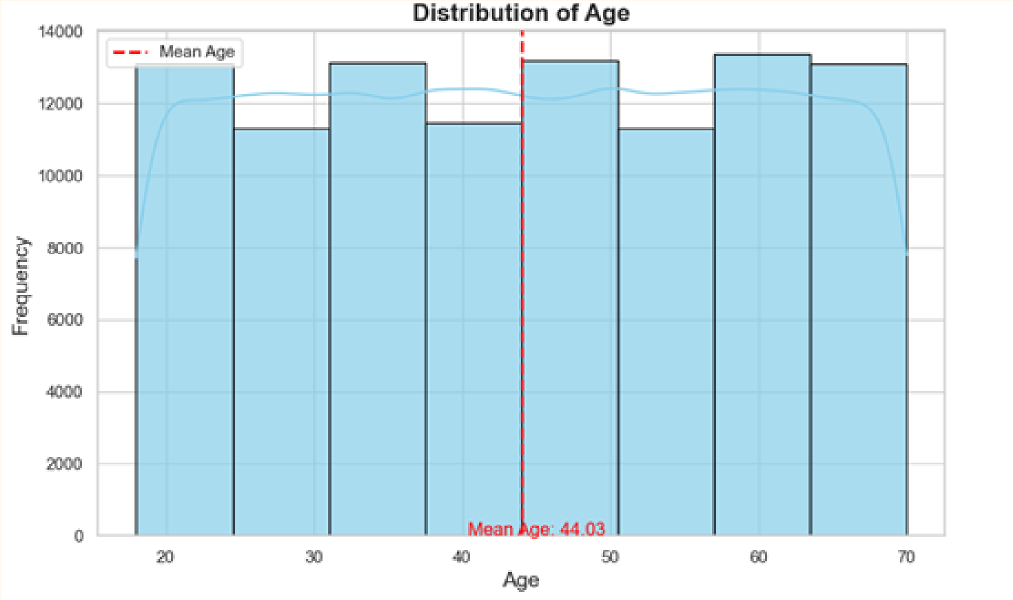

Violin Plot of Monthly Bill
Mean
Median
Max
30
40
50
60
70
80
90
100
Monthly Bill


Count of Customers by Gender
Male
49.8%
50.2%
Female

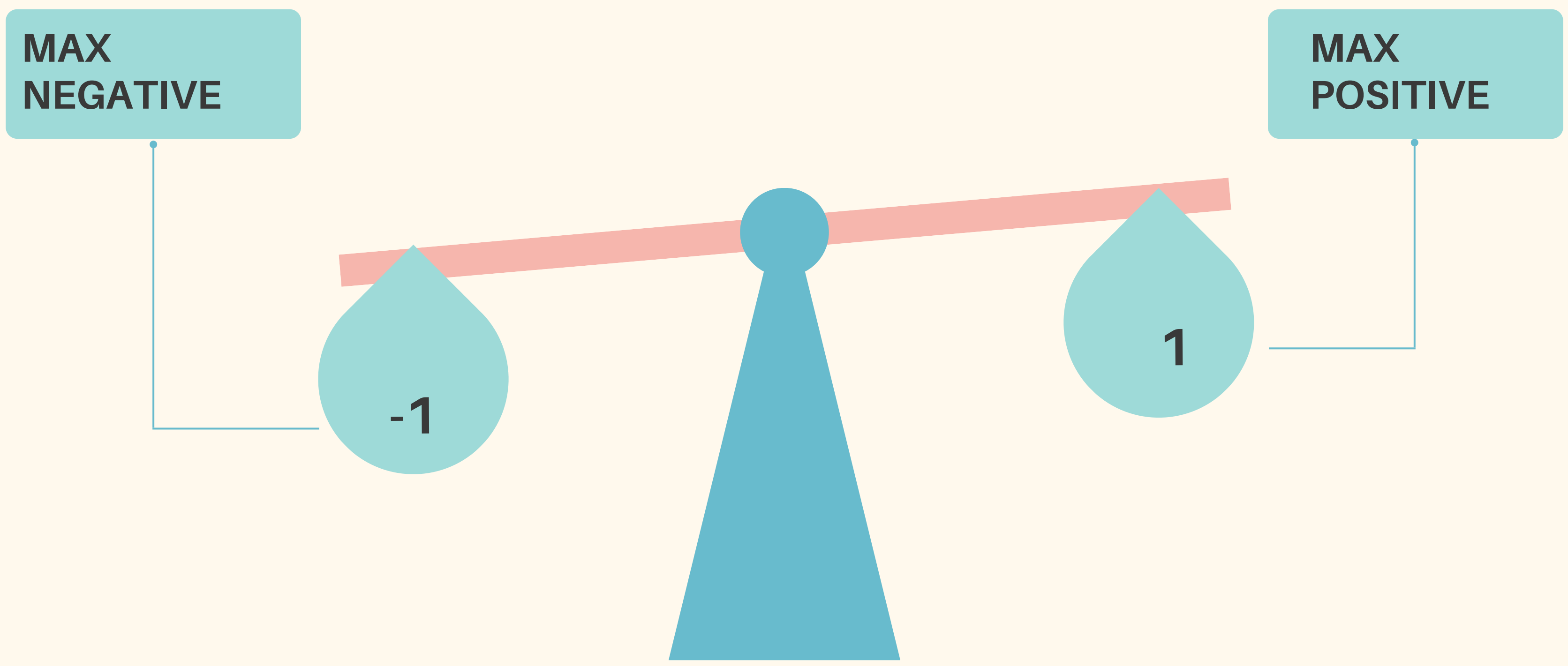

Count of Customers by Churn Status
50221.0
49779.0
50000
40000
30000
20000
10000
0
0
1
Churn Status (0 = No, 1 = Yes)


Boxplot of Total_Usage_GB
100
200
300
400
500
Total_Usage_GB


Count of Customers by Subscription Length (Months)
300
250
200
150
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Subscription Length (Months)


Distribution of Age
Mean Age
Mean Age: 44.03
14000
12000
10000
8000
6000
4000
2000
0
Frequency
20
30
40
50
60
70
Age

FEATURE CORRELATION



NO SIGNIFICANT CORRELATION NOTED

Age	0.001559
Subscription_Length	0.002328
Monthly_Bill	-0.000211
Total_Usage_GB	-0.002842
Gender_Female	-0.002121
Gender_Male	0.002121
Location_Houston	-0.006728
Location_New York	0.005835


1
Split data into Train and Test groups
2
Fit Baseline Model
3
Evaluate Model
4
Interprete Metrics
5
Hyperparameter Tuning
6
Create Final model
DECISIONTREECLASSIFIER

ACCURACY

The proportion of correct predictions

PRECISION

The ratio of true positive predictions to the total predicted positives

RECALL

Sensitivity, it measures the proportion of true positive predictions among all actual positives

F1-SCORE

The harmonic mean of precision and recall, providing a balance between the two.

BASELINE MODEL METRICS BEFORE PARAMETER TUNING

F1-SCORE

TRAIN = 50%

TEST = 50%

PRECISION

TRAIN = 100%

TEST = 50%

RECALL

TRAIN = 100%

TEST = 51%

ACCURACY

TRAIN = 100%

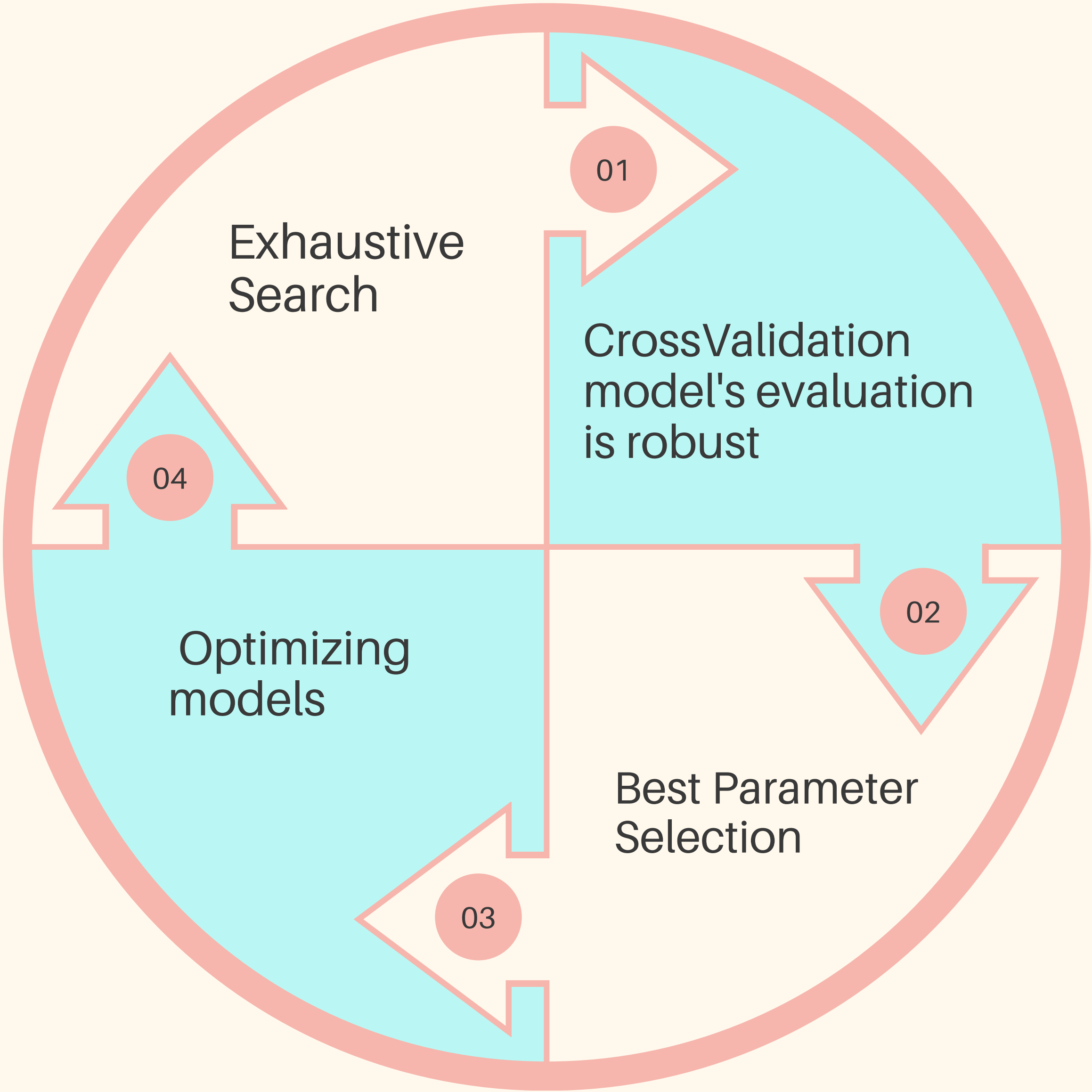
TEST = 50%

PREDICTION ON TEST DATA BEFORE PARAMETER HYPERTUNING

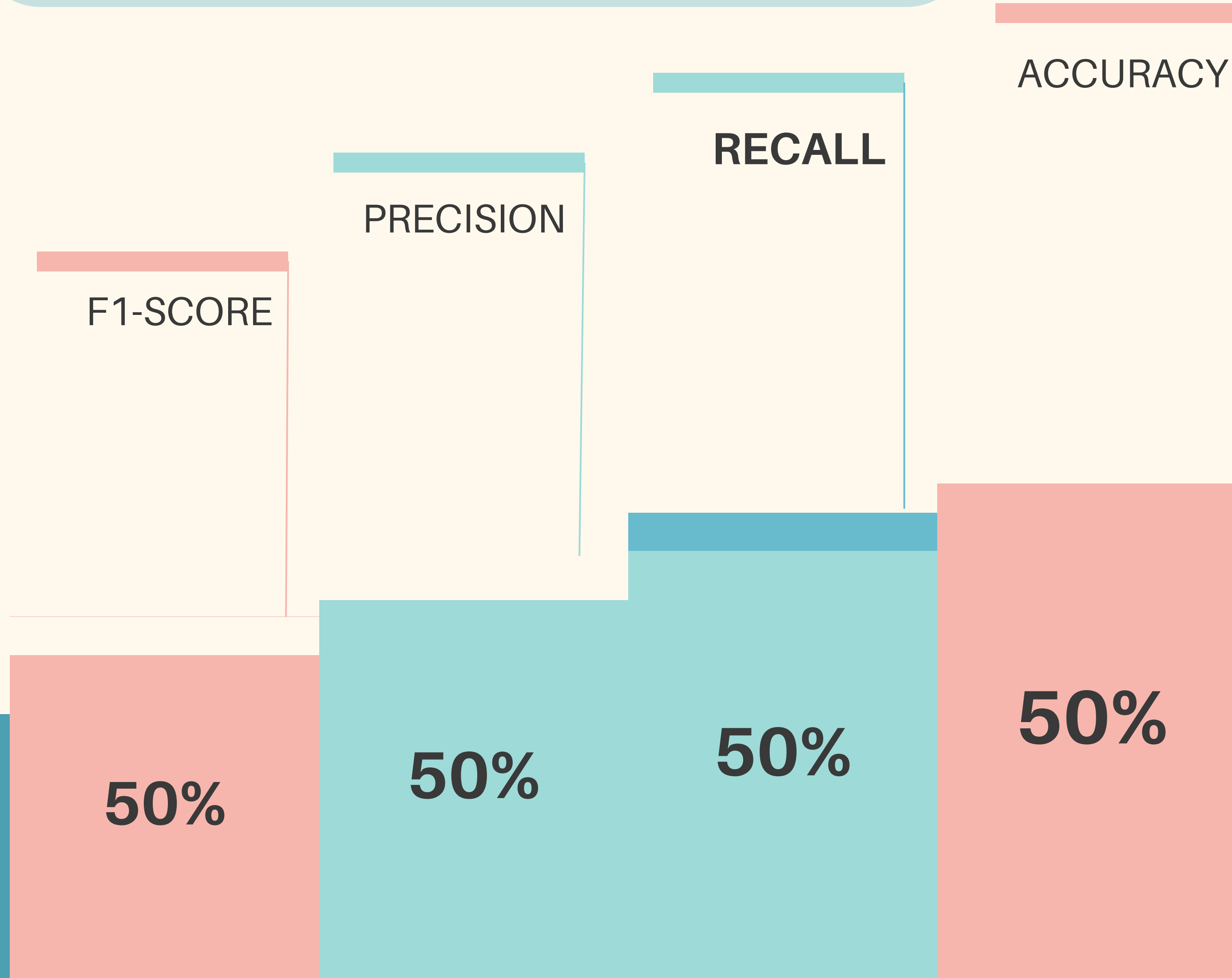
TRUE NEGATIVES	01	**5031 (No churn)**
TRUE POSITIVES	02	**5025 (churn)**
FALSE NEGATIVES	03	4896 (No churn incorrectly)
FALSE POSITIVES	**04**	5048 (churn incorrectly)

GRIDSEARCHCV

Hyperparameter tuning in machine learning models. It automates the process of searching through a specified grid of hyperparameter combinations to find the optimal parameters for a given model.



FINAL MODEL METRICS



MODEL Evaluation

Current model lacks accuracy (0.50) and insights into churn predictors; consider advanced models like Random Forests or Gradient Boosting for better feature identification.

The model's performance is akin to random guessing, necessitating exploration of different algorithms for improved accuracy

Current model limits actionable insights; enhancing predictive power is essential for reliable retention recommendations.

A more accurate model is needed to explore these relationships effectively.

RECOMMENDATIONS

01

Algorithm Selection

We should consider trying more complex models like Random Forests, Gradient Boosting, or Neural Networks.

02

Data Augmentation

Collect More Data: Increase the dataset size to improve model training.

CONCLUSION

A more sophisticated algorithm is necessary to enhance predictive accuracy and derive actionable insights for customer retention strategies.

Thank you