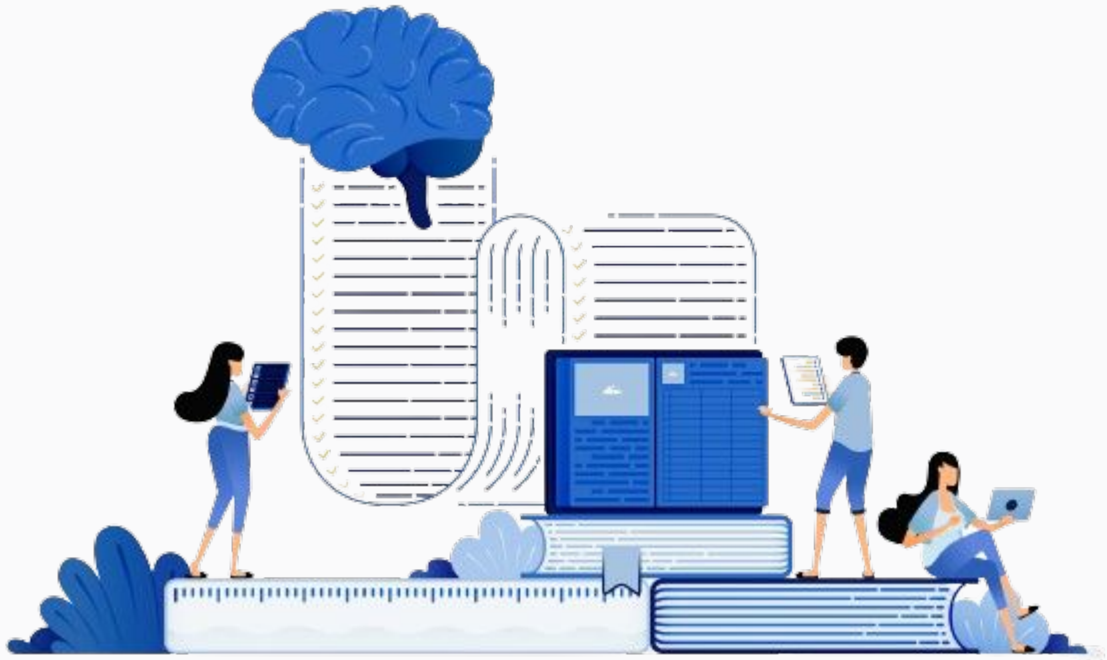
SocBiz Winter Analytics

Data Analytics and Machine Learning Model



Made by- Aryan Singh

Executive Summary

In the competitive landscape of the telecommunications sector, Zela is actively addressing the challenge of customer retention. The company has identified a noteworthy increase in customer churn, signifying a growing rate of service discontinuation.

The Solution: Data-driven insights. Zela is diving deep into customer data, analyzing usage, costs, and interactions to understand why customers are leaving. It is actively making changes in the company to take its problems

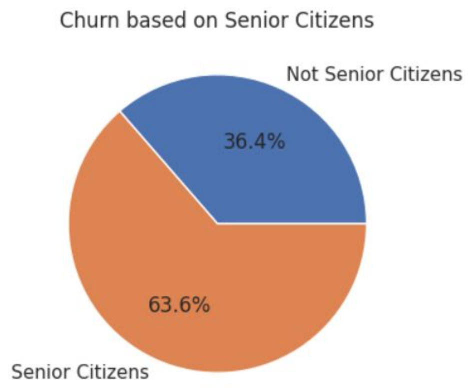
Key Strategies:

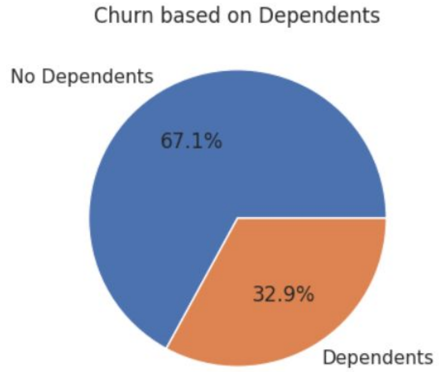
- Pricing Tweaks: Adjust pricing models to better fit customer needs and budgets.
- Sweeten the Deal: Enhance service features with offerings that add value and stand out from the competition.
- Customer First: Up the customer service game by making interactions smoother, faster, and more helpful.

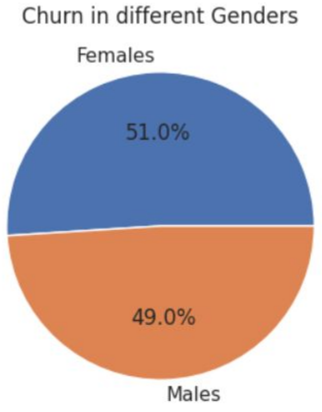
Why it Matters: By understanding and addressing the reasons behind churn, Zela can optimize its services, gain a competitive edge, and build a customer base that stays.

Zela wants to stop the exodus, turn happy customers into brand champions, and ensure its own long-term success.

Distribution of Churn based on dynamics of Customers



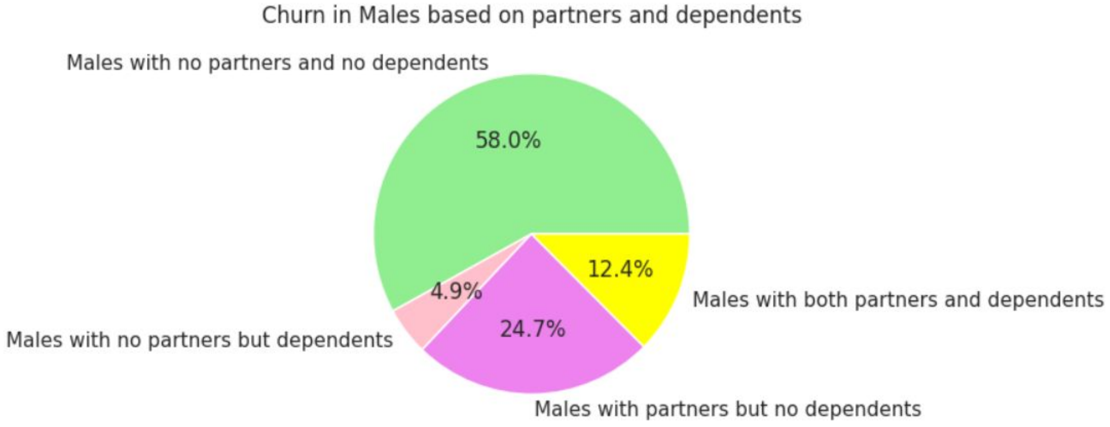


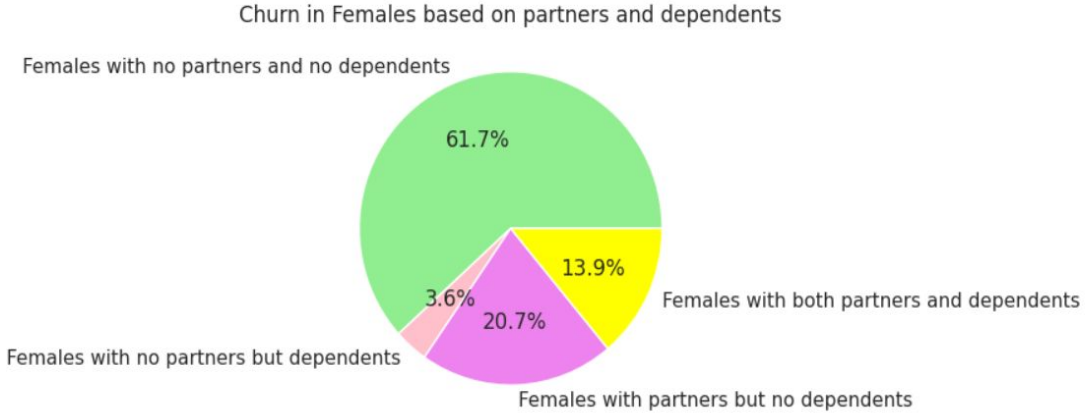


Customer churn is a significant problem in Zela with

- 63.8% of them being senior citizens.
- 67.8% of them are dependents.

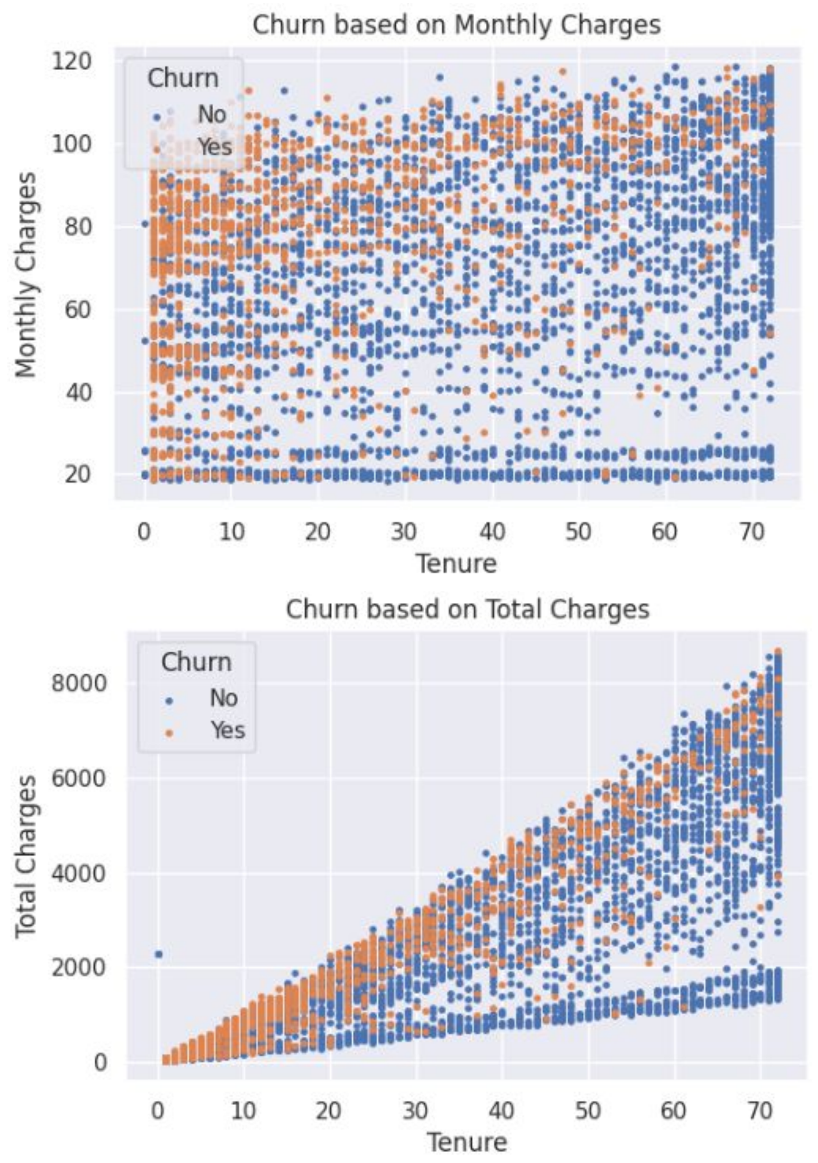
- Both the genders have almost equal churn rates.
- In males, those with no partners and no dependents are most likely to discontinue their service (58%).
- In females too, those with no partners and no dependents are most likely to discontinue their service (61.7%).



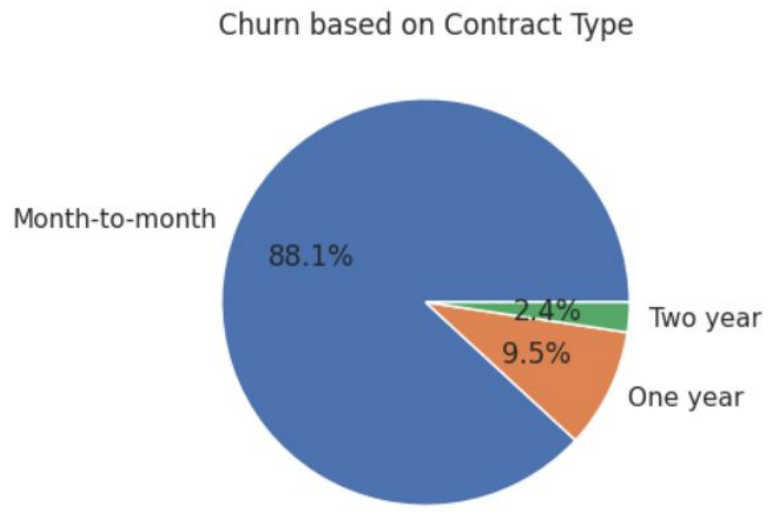


Effect of Monthly and Total Charges, Contract Type and Billing Methods

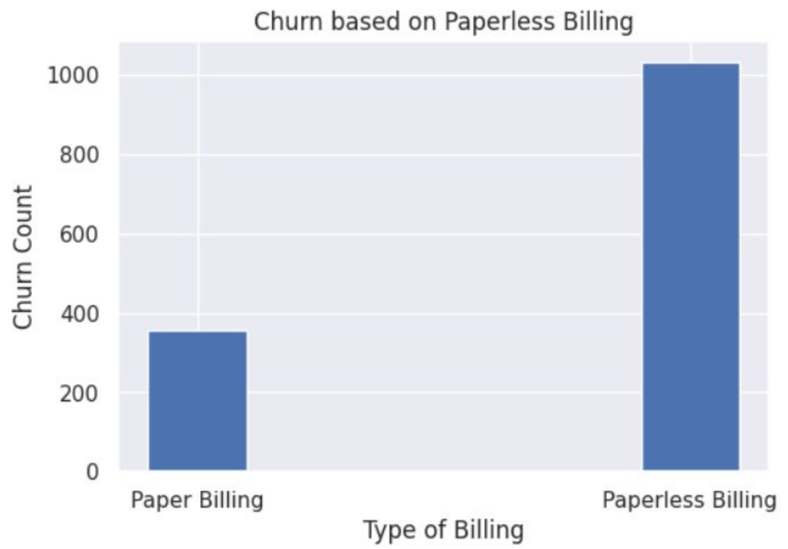
- Churn rates are directly proportional to monthly charges but inversely proportional to total charges which is caused due to low customer retention.
- High churn in month-to-month contract type further indicates that customers are not willing to commit to the company for a long time as they do not trust the company.
- Customers preferring paperless billing are more likely to discontinue their service.



Churn being proportional to high monthly charges but low total charges are a result of low tenure in customers.

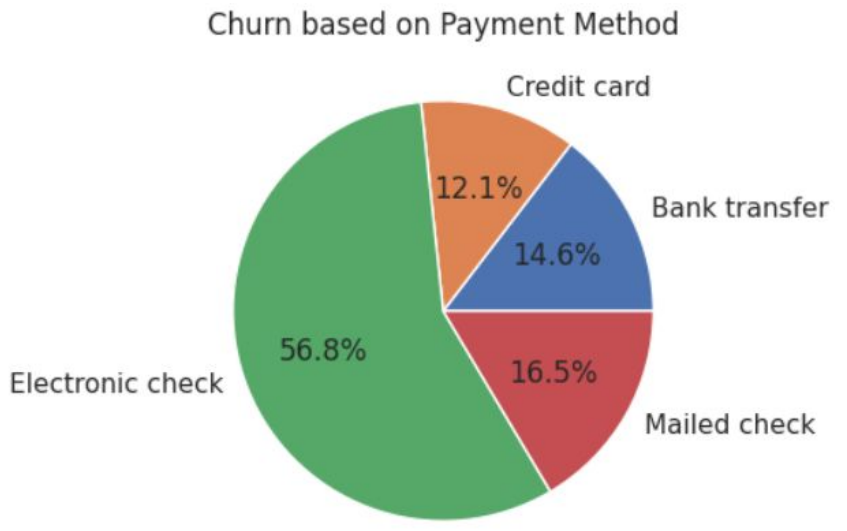


Month-to-month contract has significant percentage of churn rates.



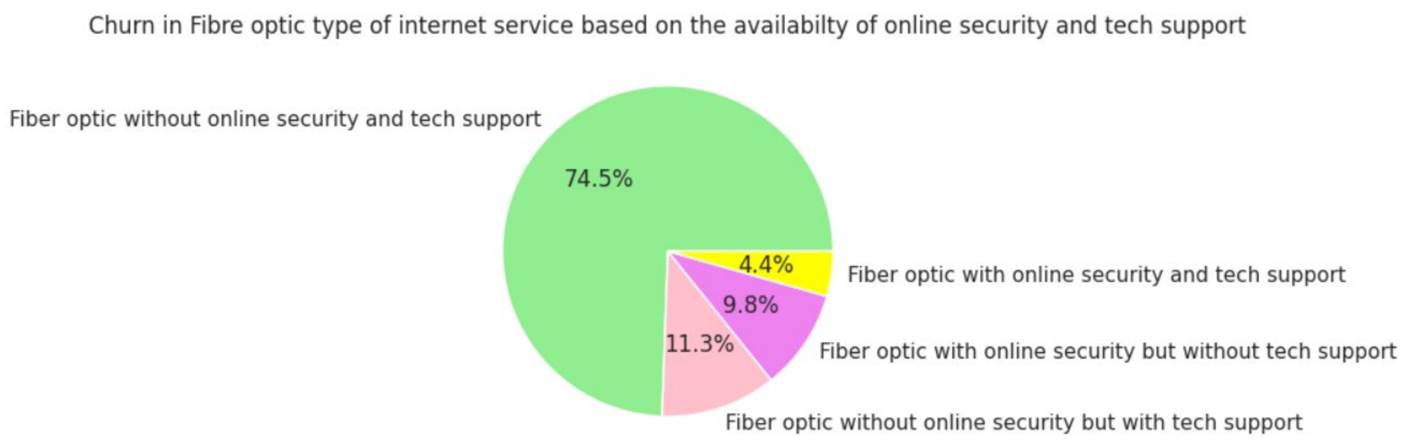
Paperless Billing accounts for significantly high churn than paper billing.

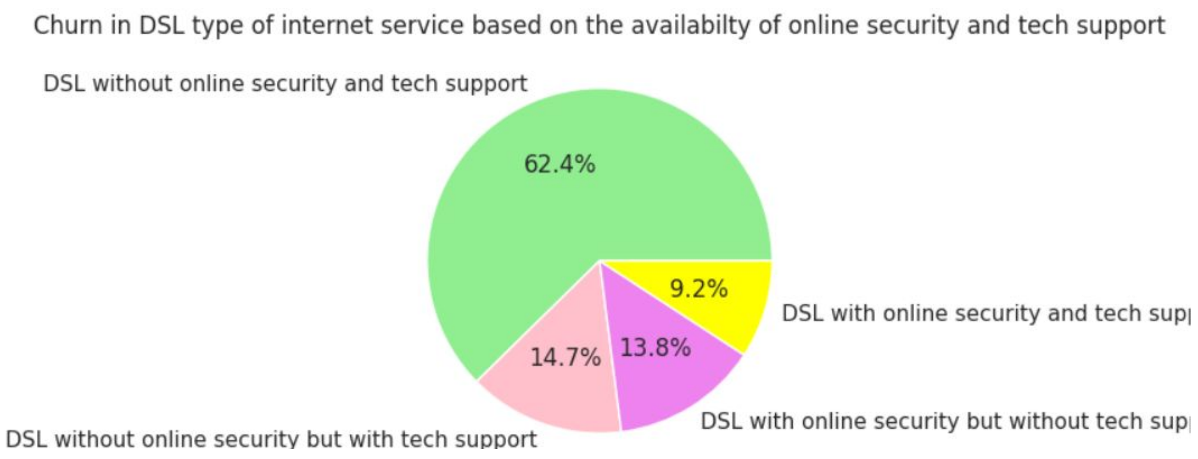
Effect of Payment Method and Type of Internet Service



Different payment methods account for churn, with Electronic check amounting to more than half of it (56.8%).

- Churn is significantly high in customers who opt for Electronic check (56.8%) as their payment method.
- Fiber optic internet service, lacking features like online security and tech support, contributes to increased churn rates due to customer dissatisfaction with the quality of service.





Churn due to the absence of features like online security and tech support is high in Fiber optic type of internet service (74.5%) as compared to DSL (62.4%) .

Strategies to be implemented

Inefficient Fiber Optic connections:

1. Enhanced Security Measures.
2. 24/7 Technical Support.
3. Service Health Checks like troubleshooting methods and feedback mechanisms.
4. Flexible Service Plans.
5. Competitive Pricing and Promotions.
6. Transparency and Communication.
7. Add App support.

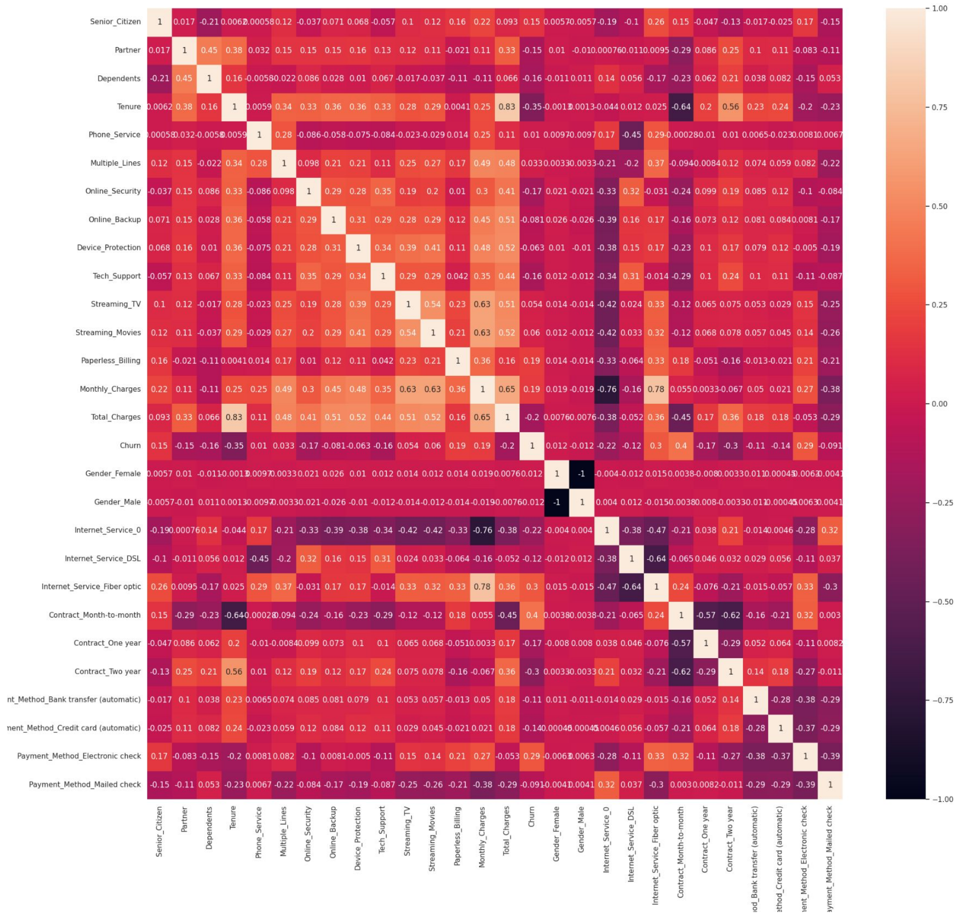
High Senior Citizen Churn Rate

1. Easy to understand terms and condition.
2. Smooth transactions.
3. Multilingual support.
4. Video tutorials for FAQs.

Low customer retention

1. Value-Added Services:
   a. Provide additional services without extra cost.
   b. Offer premium services to long-term subscribers.
2. Discounts for Commitments, like annual subscriptions.
3. Promotional Periods, offer additional services like OTT Subscription.
4. Transparent Billing show all charges clearly, ensure a smooth transaction flow.
5. Introduce family plans, to enhance customer loyalty.

Using Machine Learning Models on Data



Heatmap of Correlation among features shows that Churn is affected majorly by:

- Tenure
- Monthly and Total Charges
- Paperless Billing
- Types of Internet Service
- Contract Type
- Payment Methods

```
# Encoding the values in the dataset

data1.replace("No internet service","No",inplace=True)
data1.replace("No phone service","No",inplace=True)

data1 = data1.replace({ 'Yes':1,'No':0})

data1 = pd.get_dummies(data1, columns=['Gender','Internet_Service', 'Contract', 'Payment_Method'])
```

Encoding all columns of the dataset to convert strings to numeric categories to perform numerical operations on them

```
from sklearn.preprocessing import MinMaxScaler
numeric_cols = data1.select_dtypes(include=['int', 'float']).columns.tolist()
scaler = MinMaxScaler()
data1[numeric_cols] = scaler.fit_transform(data1[numeric_cols])
data1.head()
```

Normalising the data so that it can be fed to the machine learning model to train it.

Checking Accuracy and Predicting Churn

```
X = data1.drop('Churn', axis=1)
y = data1['Churn']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Splitting the given dataset into training data and testing data with testing data being 20% of the total dataset.

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report, \
ConfusionMatrixDisplay, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
```

Trying different machine learning models on the data to find the most appropriate for the given test data.

```
Logistic Regression
Model performance for Training set
- Accuracy: 0.8009
- F1 score: 0.7937
- Precision: 0.6523
- Recall: 0.5278
- Roc Auc Score: 0.7134
----------------------------------
Model performance for Test set
- Accuracy: 0.8127
- F1 score: 0.8033
- Precision: 0.6845
- Recall: 0.5146
- Roc Auc Score: 0.7158
===================================
```

```
X_test = data2

clf = LogisticRegression()

X_train = data1.drop('Churn', axis=1)
y_train = data1['Churn']
clf.fit(X_train, y_train)

▾ LogisticRegression
LogisticRegression()

predicted_churn = clf.predict(X_test)
print(predicted_churn)

[1. 0. 0. ... 0. 1. 0.]
```

Logistic Regression model gives us the best accuracy (81.2%). So we use this model to predict churn on the given test data.

THANK YOU