Customer Churn Analysis

Identifying Key Drivers of Customer Churn:
An Analytical Study of SpeedyCall's
Customer Data

Case Study

IS 3005

Tharindu Darshana

s16341

Contents

Introduction

Customer churn is a critical challenge faced by businesses in the telecommunication industry, directly impacting revenue and growth. Identifying factors and customer behaviours that influence customer churn is crucial as it helps the company to focus on those areas and address them to retain customers. This study analyzes customer data from SpeedyCall, a U.S.-based telecommunication company, to identify key factors contributing to customer churn. The analysis aims to provide actionable insights to help the company retain customers and improve their overall experience by exploring patterns in demographics, service usage, and subscription types.

<u>Objectives:</u>

- To identify significant factors and behavioral patterns associated with customer churn in SpeedyCall.
    - To analyze how demographic factors, such as seniority and gender, influence customer churn at SpeedyCall.
    - To determine whether there is a significant difference in monthly spending between customers who churn and those who do not.
    - To examine if customers on monthly plans churn at a higher rate compared to those on yearly contracts.
    - To explore the relationship between customer tenure and churn.
    - To investigate how the presence or absence of additional services, such as tech support, impacts churn.
- To provide data-driven recommendations for improving customer satisfaction and reducing churn.

Data

This dataset contains data from 7043 customers and 21 features. For this analysis, we use a subset of those features to perform exploratory data analysis. Following are the features we are using in this study.

- **SeniorCitizen** - Whether the customer is a senior citizen or not (1, 0)
- **Partner** - Whether the customer has a partner or not (Yes, No)
- **Dependents** - Whether the customer has dependents or not (Yes, No)
- **Tenure** - Number of months the customer has stayed with the company
- **InternetService** - Customer’s internet service provider (DSL, Fiber optic, No)
- **Contract** - The contract term of the customer (Month-to-month, One year, Two year)
- **TechSupport** - Whether the customer has tech support or not (Yes, No, No internet service)
- **PaymentMethod** - The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- **MonthlyCharges** - The amount charged to the customer monthly
- **Churn** - Whether the customer churned or not (Yes or No)

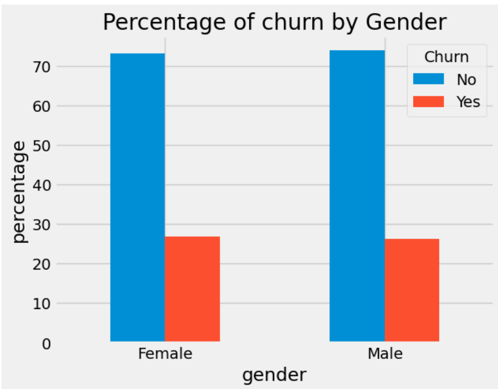
Exploratory Data Analysis

The dataset used in this study contains detailed information on 7,043 SpeedyCall customers, including their demographics, subscription details, and service usage patterns. Among these customers, 1,869 (approximately 26.5%) have churned, indicating a significant segment is at risk of leaving the service. Therefore it's a priority to analyse and find out what causes the high churn rate.

The analysis focuses on exploring customer attributes in three ways. That is analyse customer demographic variables, service usage patterns and subscription patterns to identify trends and patterns associated with churn.

Demographic Variables Analysis

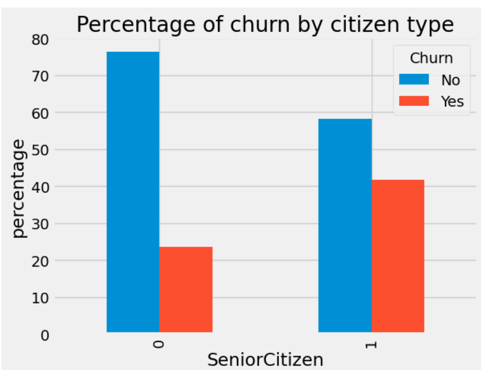
Gender



Churn	No	Yes
gender		
Female	73.08%	26.92%
Male	73.84%	26.16%

Here we can see both genders have the same percentage of churning meaning that gender is not influencing the churn

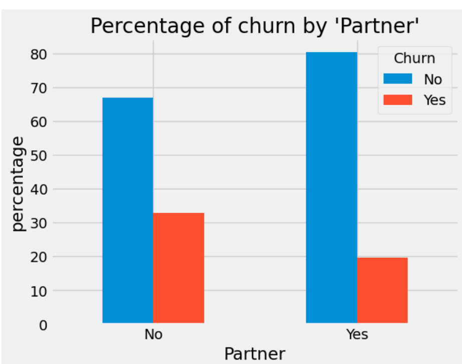
Senior Citizen

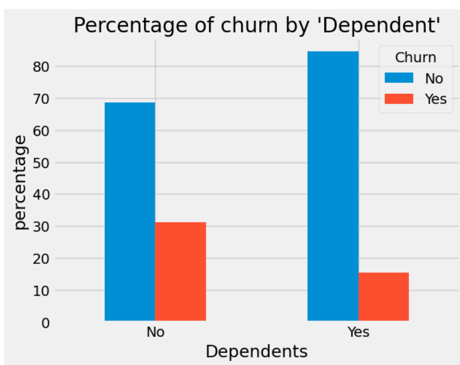


Churn	No	Yes
SeniorCitizen		
0	76.4%	23.6%
1	58.3%	41.7%

It is observed that 41.7% of senior citizens leave the service. This is a considerably high percentage. To identify the reasons behind the high percentage of churn among senior citizens, further analysis will be conducted.

Partner and Dependent

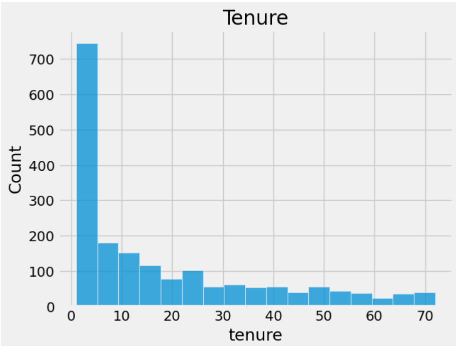




Here we can observe that customers with no partners and dependents have more tendency to churn. It might be an indicator of young individuals' tendency to leave the service.

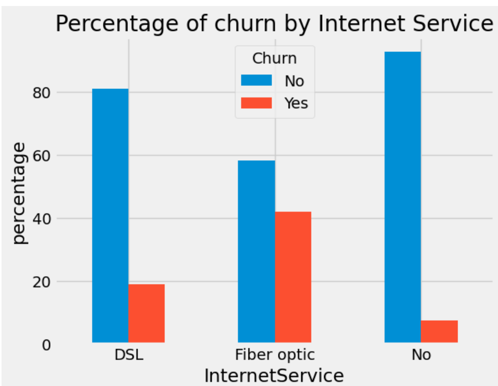
Service Usage Analysis

Tenure



This histogram depicts the tenure (number of months the customer has stayed with the company) of churned customers. We can observe a clear majority of churned customers have a tenure of less than 5 months. This is strong evidence that new customers tend to leave the company compared to long-standing customers.

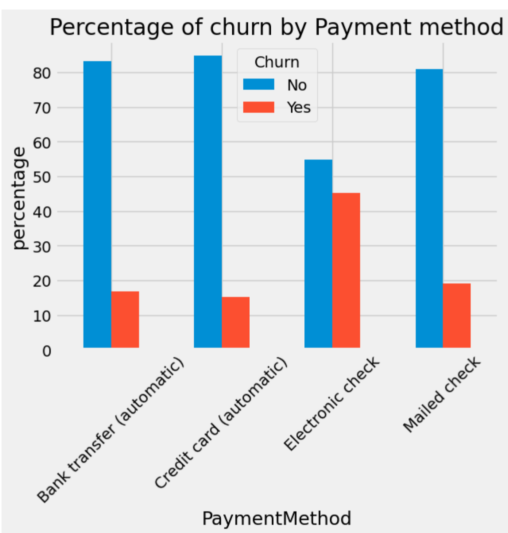
Internet Service



Churn	No	Yes
InternetService		
DSL	81.04%	18.96%
Fiber optic	58.11%	41.89%
No	92.60%	7.40%

Here we can observe that 41.9% of customers who use fiber optic have churned. It's a notably high percentage. This might be an indicator of customer dissatisfaction with fiber optic service. This can be because of high monthly charges for fiber optic service or any other technical issue with that service.

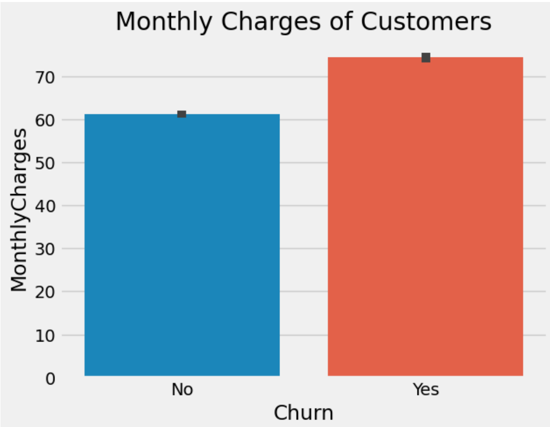
Payment Method



Churn	No	Yes
PaymentMethod		
Bank transfer (automatic)	83.29%	16.71%
Credit card (automatic)	84.76%	15.24%
Electronic check	54.71%	45.29%
Mailed check	80.89%	19.11%

Here we can observe customers who use 'Electronic check' as their payment method have a high churn percentage. In this case, 45.3% which is considerably high. This could suggest potential issues with the 'Electronic Check' payment method or indicate the influence of other underlying factors.

Monthly Charges

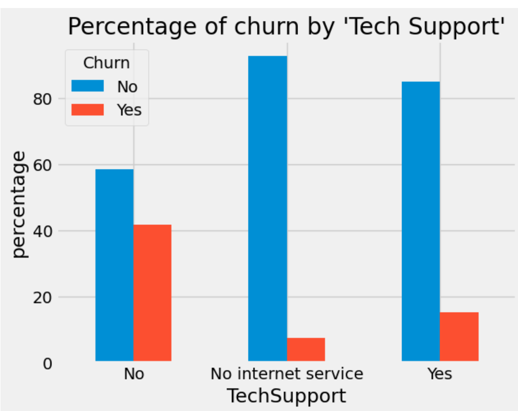



```
	Welch Two Sample t-test

data:  df_yes$MonthlyCharges and df_no$MonthlyCharges
t = 18.408, df = 4135.8, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 11.99855      Inf
sample estimates:
mean of x mean of y 
 74.44133  61.26512 
```

Here we can see the average monthly charges of churned customers are higher than those who did not. We can confirm it using the two-sample t-test as it rejects the null hypothesis of 'True difference in means is equal to 0' (since the p-value is less than 0.05). This can be evidence that high monthly charges lead the customers to churn.

Tech Support

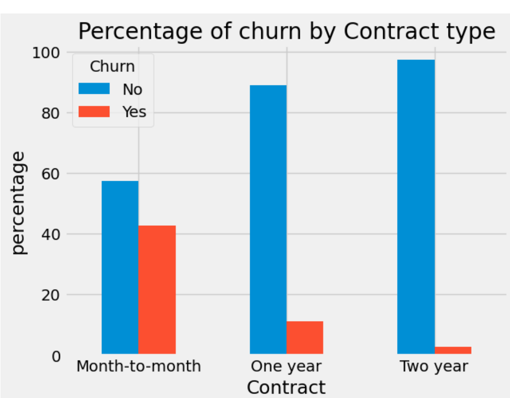


Churn	No	Yes
TechSupport		
No	58.36%	41.64%
No internet service	92.60%	7.40%
Yes	84.83%	15.17%

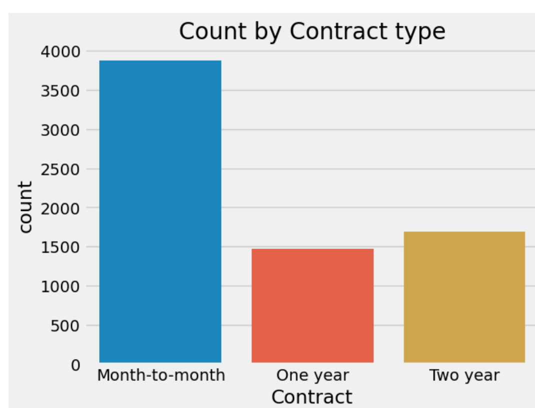
We can observe that customers without tech support have more tendency to churn. Out of those who do not have tech support, 41.6% have churned. This suggests that a lack of awareness about the service may ultimately contribute to customers discontinuing it.

Subscription Type Analysis

Contract



Churn	No	Yes
Contract		
Month-to-month	57.29%	42.71%
One year	88.73%	11.27%
Two year	97.17%	2.83%



We can observe that the highest percentage of churn was reported in customers who use the month-to-month contract type. In the 2nd bar graph, we can see that the majority of SpeedyCall's customers use the month-to-month contract type. Out of those 42.7% has churned. This suggests that the company should pay attention to the month-to-month contractors to retain them in the service.

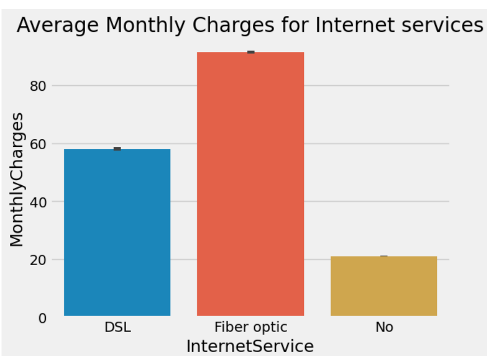
Integrated Analysis

This section combines the findings from the demographic analysis, service usage patterns, and subscription types to provide a comprehensive understanding of the factors driving customer churn. This approach not only highlights the key drivers of churn but also lays the foundation for targeted strategies to improve customer retention.

If we are to find out the underlying factors associated with more senior citizens leaving the service, we can see that the majority of senior citizens use 'Electronic Check' as their payment method which has a high percentage of churn. If we further analyse it, we can see out of churned senior citizens 82.6% used fiber optic as their internet service.

SeniorCitizen	0	1
InternetService		
DSL	36.64%	22.68%
Fiber optic	38.38%	72.77%
No	24.98%	4.55%

SeniorCitizen	0		1	
Churn	No	Yes	No	Yes
InternetService				
DSL	39.51%	27.35%	27.18%	16.39%
Fiber optic	30.19%	64.90%	65.77%	82.56%
No	30.30%	7.75%	7.06%	1.05%



From this, we can conclude that because the majority of senior citizens use fiber optic as their internet service, we observe a high percentage of churn among senior citizens. To find out the reasons for dissatisfaction among fiber optic customers, let's look at the average monthly charges for that service. Here we can clearly see that the average monthly charges for fiber optic service is much higher than for other services. This can be a reason for customer dissatisfaction with fiber optic service.

Now let's analyse what are the underlying factors associated with high churn percentage among the month-to-month contractors.

Contract	Month-to-month	One year	Two year
PaymentMethod			
Bank transfer (automatic)	15.20%	26.54%	33.27%
Credit card (automatic)	14.01%	27.02%	34.28%
Electronic check	47.74%	23.56%	9.91%
Mailed check	23.05%	22.88%	22.54%

Here we can see that 47.7% of month-to-month contractors are 'Electronic Check' payment method users. This may cause to high churn percentage among month-to-month contractors as there is a high churn percentage among the customers who use 'Electronic Check' as their payment method. Or it can be the other way around also. Further details and analysis need to be done on what is really going on here.

PaymentMethod \ InternetService	DSL	Fiber optic	No
Bank transfer (automatic)	36.66%	41.84%	21.50%
Credit card (automatic)	39.03%	39.22%	21.75%
Electronic check	27.40%	67.44%	5.16%
Mailed check	38.03%	16.00%	45.97%

If we take a look at the association between Internet service and Payment method, we can see out of those who use the 'Electronic Check' payment method, 67.4% of them use fiber optic internet service, which has a high churn percentage.

InternetService \ Contract	Month-to-month	One year	Two year
DSL	31.56%	38.70%	37.05%
Fiber optic	54.92%	36.59%	25.31%
No	13.52%	24.71%	37.64%

Also, if we check the association between Internet service and Contract type, 54.9% of fiber optic users are month-to-month contractors. Therefore it seems like all these three factors are associated with each other, influencing one another. The reason for the high churn percentage can be one or more of these factors. It can be due to an issue in 'Electronic Check' payment method that causes dissatisfaction among customers. The company should check if there are continuous customer complaints about the 'Electronic Check' payment method and any drawbacks in this payment method compared to other payment methods.

Also, the company should pay attention to fiber optic service. The high churn percentage among fiber optic users can be due to high monthly charges or any other issues with that service.

InternetService	DSL		Fiber optic		No	
Churn	No	Yes	No	Yes	No	Yes
TechSupport						
No	45.77%	75.16%	62.76%	84.89%	0.00%	0.00%
No internet service	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
Yes	54.23%	24.84%	37.24%	15.11%	0.00%	0.00%

It is observed that 84.9% of churned fiber optic users did not have tech support. This may indicate a lack of awareness about the service, which could ultimately lead to churn. Therefore the company should pay attention to expanding their tech support to customers.

Conclusion of EDA

- Gender has no influence on the churn
- It is observed that there's a high churn percentage among senior citizens and we found that it may be due to the majority of senior citizens using fiber optic service which has a high churn percentage.
- Customer dissatisfaction with fiber optic service can be due to its high monthly charges or it can be other issues related to that service. The company should pay attention to addressing these issues to improve customer satisfaction.
- Also found that among the customers who do not have tech support, 41.6% have churned. Suggesting that a lack of awareness about the service could result in customers leaving the service. The company should pay attention to expanding its tech support to customers to retain them in the service.
- The majority of customers use month-to-month contracts. And we discovered that out of them 42.7% has churned.

- Addressing the underlying factors influencing high churn percentage among month-to-month contractors, we found that contract type, internet service type and payment method are associated with each other.
- Analysing these associations it is suspected that there are potential issues related to the 'Electronic Check' payment method. It is recommended that the company should check if there are customer complaints about the 'Electronic Check' payment method or any drawbacks related to this payment method and if there are, address them accordingly. If no such issues are found with this payment method, it can be the high churn percentage of fiber optic service that influences the high churn percentage of this payment method.

Advanced Analysis

This section focuses on building a predictive model to predict customer churn based on customer data. For that, multiple machine learning models will be trained and evaluated their performance and choose the model that gives the highest accuracy as the best model.

First, the dataset was divided into the training set and the testing set. Since there's an imbalance in the response variable (Churn), the SMOTE technique was used to tackle the imbalance of the training set so that models can learn from both classes without bias. After resampling the training set, several machine learning algorithms were trained and evaluated using metrics.

	model	score
0	LogisticRegression(max_iter=1000)	0.76
1	DecisionTreeClassifier()	0.70
2	(DecisionTreeClassifier(max_features='sqrt', r...	0.77
3	SVC()	0.75
4	KNeighborsClassifier()	0.71
5	XGBRFClassifier(base_score=None, booster=None,...	0.76

Logistic regression, Random forest, and XGBoost classifier models were highly accurate, so hyperparameter tuning was carried out on them.

After performing hyperparameter tuning, their performances were evaluated using classification report.

Classification report for the XG boost model

	precision	recall	f1-score	support
0	0.88	0.78	0.83	1549
1	0.54	0.71	0.61	561
accuracy			0.76	2110
macro avg	0.71	0.74	0.72	2110
weighted avg	0.79	0.76	0.77	2110

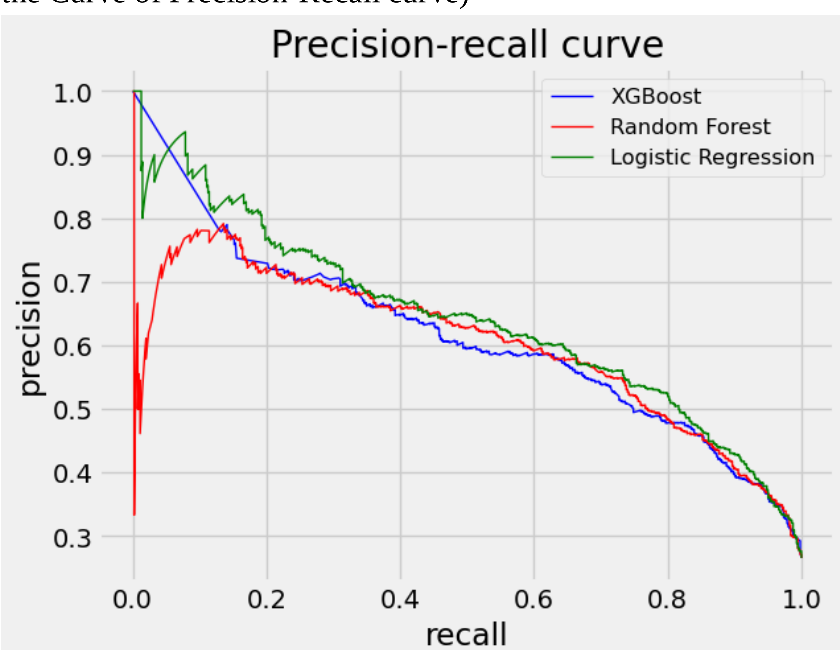
Classification report for the Random Forest model

	precision	recall	f1-score	support
0	0.88	0.81	0.84	1549
1	0.56	0.69	0.62	561
accuracy			0.77	2110
macro avg	0.72	0.75	0.73	2110
weighted avg	0.79	0.77	0.78	2110

Classification report for the Logistic regression model

	precision	recall	f1-score	support
0	0.90	0.75	0.82	1549
1	0.53	0.77	0.63	561
accuracy			0.76	2110
macro avg	0.72	0.76	0.73	2110
weighted avg	0.80	0.76	0.77	2110

The accuracy of these three models is almost equal. But the precision and recall for the positive class are different. To select the best model, the precision-recall curve and AUC-PR values were obtained. (Area Under the Curve of Precision-Recall curve)




```
AUC-PR of XGBoost:  0.6157677942866377
AUC-PR of Ranfom Forest:  0.600404670582017
AUC-PR of Logistic Regression:  0.6411398718376378
```

Since the highest AUC-PR value was obtained from the Logistic regression model, it was selected as the best model.

After that, the Logistic regression model was retrained with class weights since there is a class imbalance in the response variable and evaluated its performance.

	precision	recall	f1-score	support
0	0.91	0.72	0.80	1549
1	0.51	0.80	0.62	561
accuracy			0.74	2110
macro avg	0.71	0.76	0.71	2110
weighted avg	0.80	0.74	0.76	2110

Classification report of logistic regression after applying class weight

It was observed that although the recall of the positive class increased, both precision and overall accuracy decreased. Therefore, threshold tuning was performed to find the best threshold that balances precision and recall. The obtained best threshold was 0.56. Again, the model performance was measured for this new threshold value.

	precision	recall	f1-score	support
0	0.90	0.77	0.83	1549
1	0.54	0.77	0.64	561
accuracy			0.77	2110
macro avg	0.72	0.77	0.73	2110
weighted avg	0.81	0.77	0.78	2110

Classification report after adjusting the threshold of the Logistic regression model

After adjusting the threshold, the accuracy and precision of the positive class increased. Also, the recall value is high. However, since the recall is much higher than the precision for the positive class, the model produces more false positives in its predictions. Nevertheless, our main objective is to detect customers who are likely to churn. In this context, a higher recall is more desirable, as it ensures that most potential churners are identified, even at the cost of mistakenly flagging some customers who would not churn. This trade-off is acceptable in scenarios where the cost of missing a churner is higher than the cost of a false alarm. Therefore, this model was selected as the best model and proceeded to be deployed.

Model Deployment

After selecting the best model, we deploy it so users can use it to predict diamond prices based on their input features in real-time. For that, we create a web application using Streamlit. Streamlit is an open-source Python framework for data scientists and AI/ML engineers to deliver dynamic data apps. The web API was created as follows

- Uploaded the Python script and related files to a GitHub repository. (tdarshana01/my_projects)
- Connect that GitHub repository to Streamlit and deploy it in the Streamlit community could

Now the app is up and running in the Streamlit cloud, and anyone with the link can access it

Link: Customer Churn Prediction streamlit app

The created API works as follows

- Users can fill out a form on the web page and click the Submit button to generate a prediction.
- When the button is clicked, the input values are sent to the Streamlit backend, where the app processes them in real-time.
- The server then uses a trained machine learning model (such as logistic regression) to calculate the predicted value based on the user's input.
- The predicted result is immediately displayed back on the web page — no page reloads required.
- This interaction happens in a single script, with live re-execution triggered by the user input.

Apart from filling the form and get the prediction for a single customer, users can upload a CSV file containing customer details and get the prediction for all of the customers at once. Also, the percentage of customers likely to churn out of the customers in the uploaded file will appear after the individual predictions of those customers.

==================================End of the Project======================================