Data Analytics

Project

Created By:

Maryam Aliakbari

Dataset: CustomerChurn.csv

Cleaning & Transforming Data

Remove "CustomerID"	In this project we do not need to observe the data of a particular customer, so we just remove it.
Rephrase Values of "MultipleLine"	We find out the relation between "PhoneService" and "MultipleLines", that "No phone service" in "MultipleLines" is because of "No" Phone services. So, we re-phrase the values in this way: "No" to "SingleLine" , "Yes" to "MultipleLine" and "No phone service" to "No Phone".
Remove "PhoneService"	Based on the previous transformation step, we've used "PhoneService" values in "MultipleLines". So, we removed this variable.
Rephrase values of "SeniorCitizen" and Change Data Type.	Rephrasing the status of customer's seniority from "1" to "Yes" and "0" to "No" after we change its data type to "Text"
Create new Column "MultiMember_Family"	We created a new column based on the concept of status of the family in two columns "Partner" and "Dependents". The value of this column is "Single" if the customer has "No" partner and "No" dependent. Otherwise, it is "Multi Member"
Remove two columns "Partner" and "Dependants"	Based on previous steps , as we used the most important part of the information of these two columns in a new column, we remove these two columns.

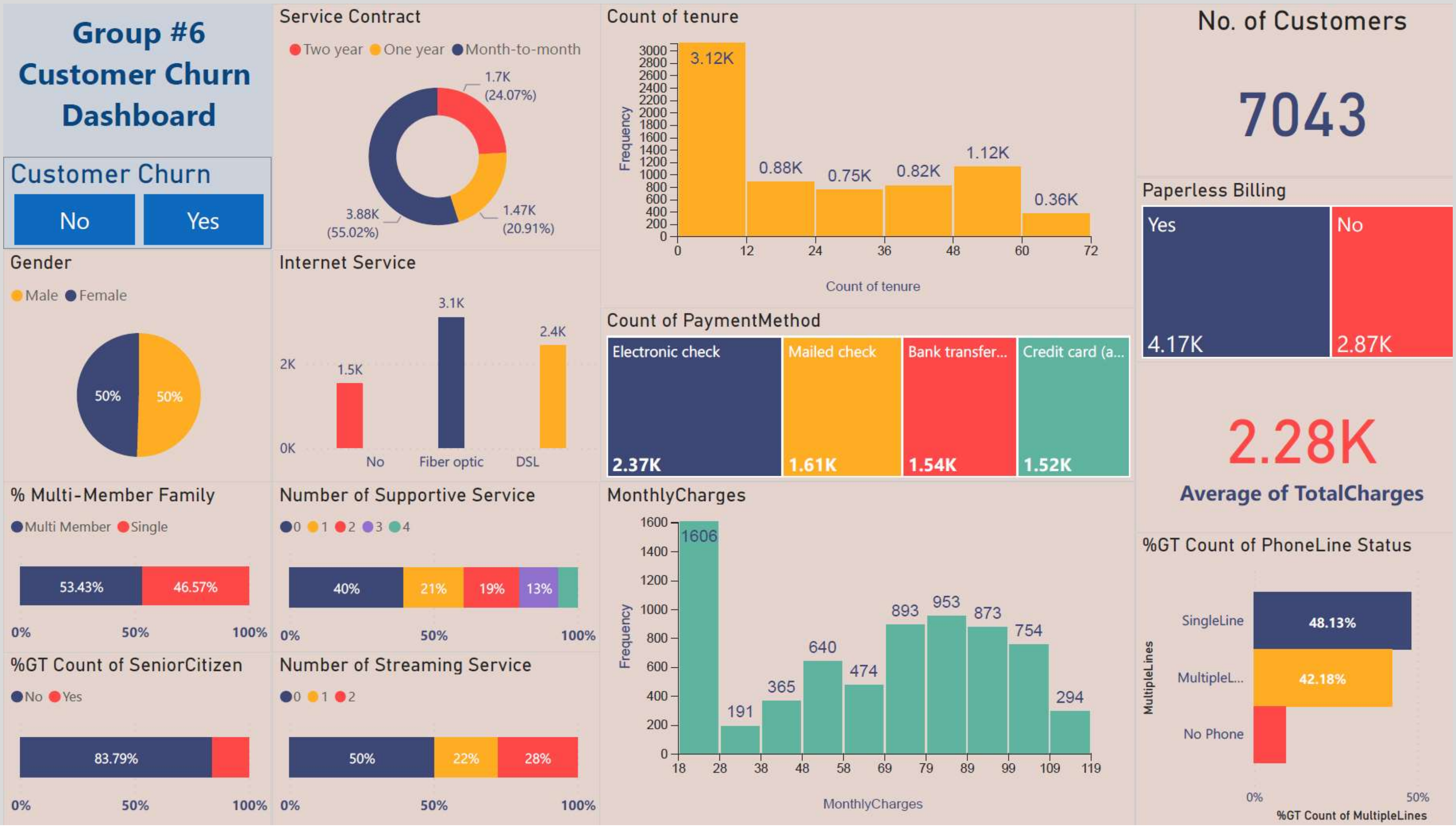
Cleaning & Transforming Data

Create "No_StreamingService"	Based on some initial EDA we found out that similar pattern of distribution between "StreamingMovies" and "StreamingTV". So, we decided to combine these two in a new column "No_StreamingService" that shows how many streaming services a customer uses.
Remove two columns "StreamingMovies" & "StreamingTV"	Based on the previous step, as we used the most important part of the information of these two columns in a new column, we removed these two columns.
Create "No_OtherService"	Based on some initial EDA we found out similar pattern of distribution among "OnlineSecurity" , "OnlineBackup", "DeviceProtection" & "TechSupport". So, we decided to merge these Supporting services status in a new column "No_OtherService" that shows how many other supportive services a customer uses.
Remove columns "OnlineSecurity" "OnlineBackup", "DeviceProtection" & "TechSupport"	These columns have been removed based on the previous step, as we've used the summary of these columns value in "No_OtherService" column.
Fill the missing Data in "TotalCharges"	We found out that there are some missing data in "TotalCharges", we also found out that these particular observations have "0" value in their "tenure". So that means they have not been charged yet. We fill these observations missing value in "TotalCharges" with the value of "0"

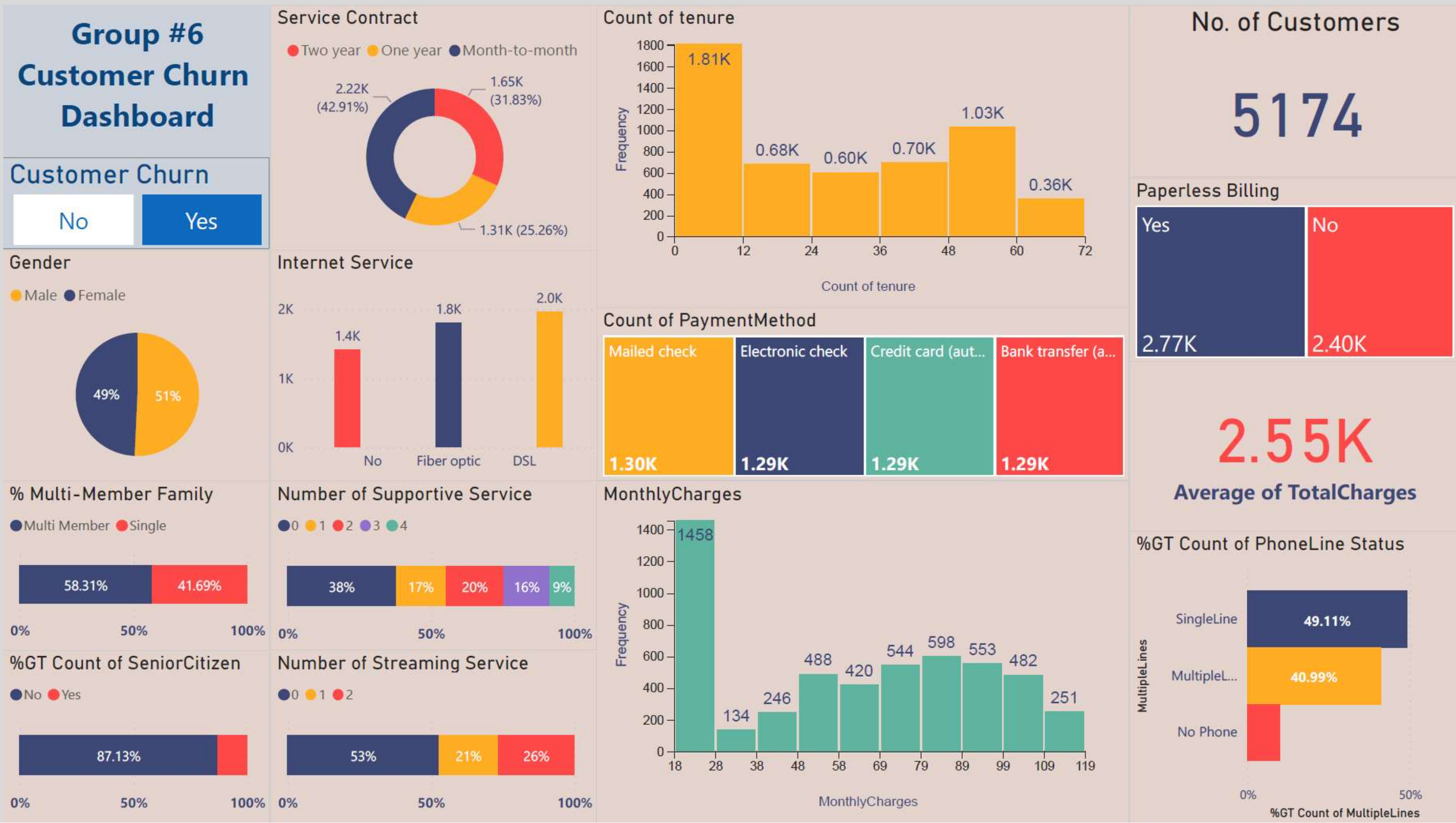
About the Dashboard

Churn Slicer	To control the dashboard components for all, churned and not-churned clients
Variations	Illustrates the variations in demographic, behavioral and contract characteristics of clients
Focus	Enables to focus on each measure
Comprehensiveness	A comprehensive glimpse of all measure at once.
Compare	Enables to compare changes in several measures

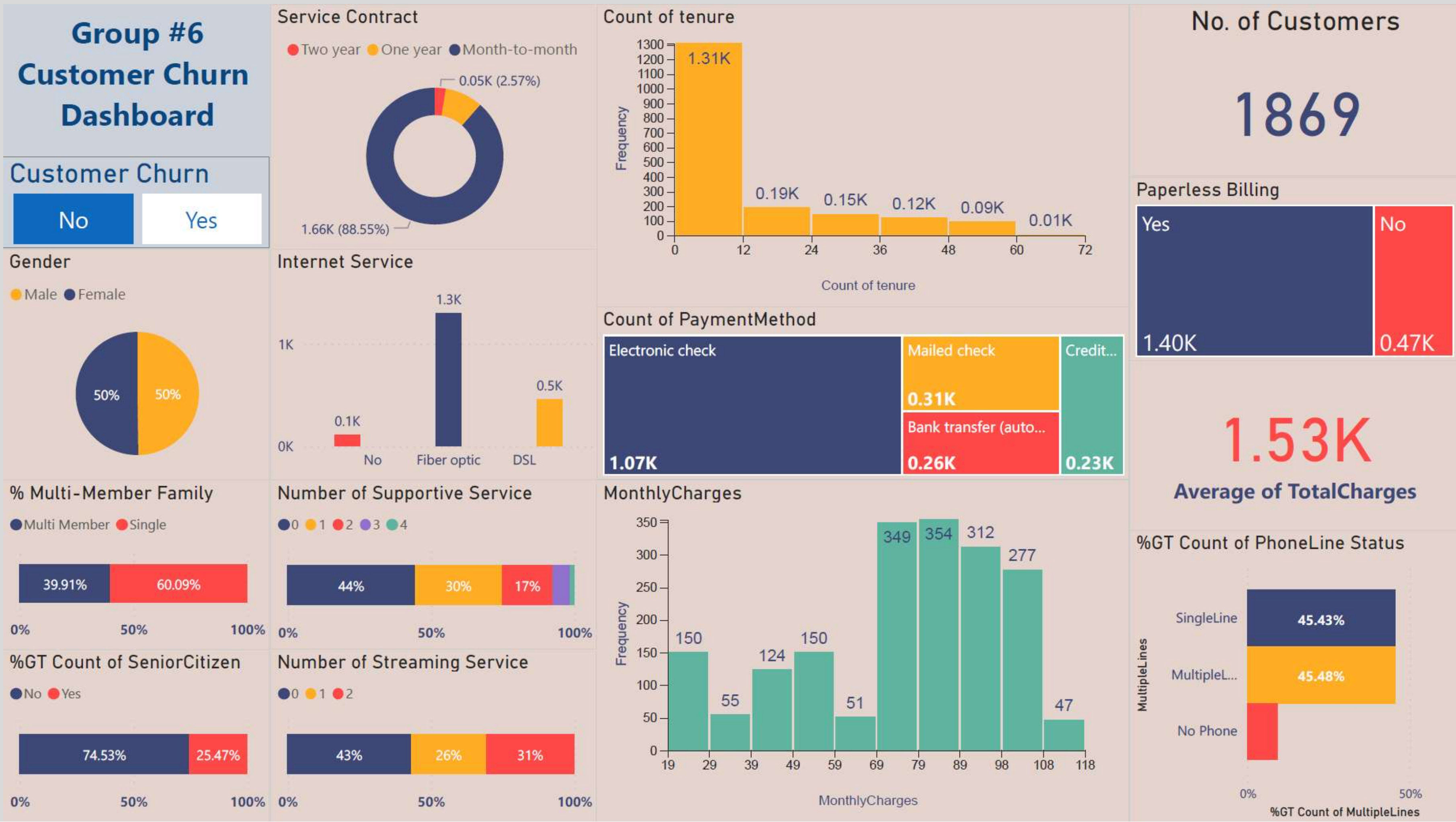
Dashboard - Full Customers View



Dashboard - Not Churned Customers View



Dashboard - Churned Customers View



Results & Recommendations

"Churned" Customer Characteristics compared to "No-Churned Customers"	
Gender	Female
Multi-member family	Lower
Senior citizen	More
Service contract	More short-term
Phone services	More multiple-lines
No. of supportive services	Fewer
No. of streaming services	More
Internet service	More Fiber Optic
Tenure	Less
Billing type	More paperless
Payment method	More e-checks
Avg. total charges	Less
Avg. monthly charges	More

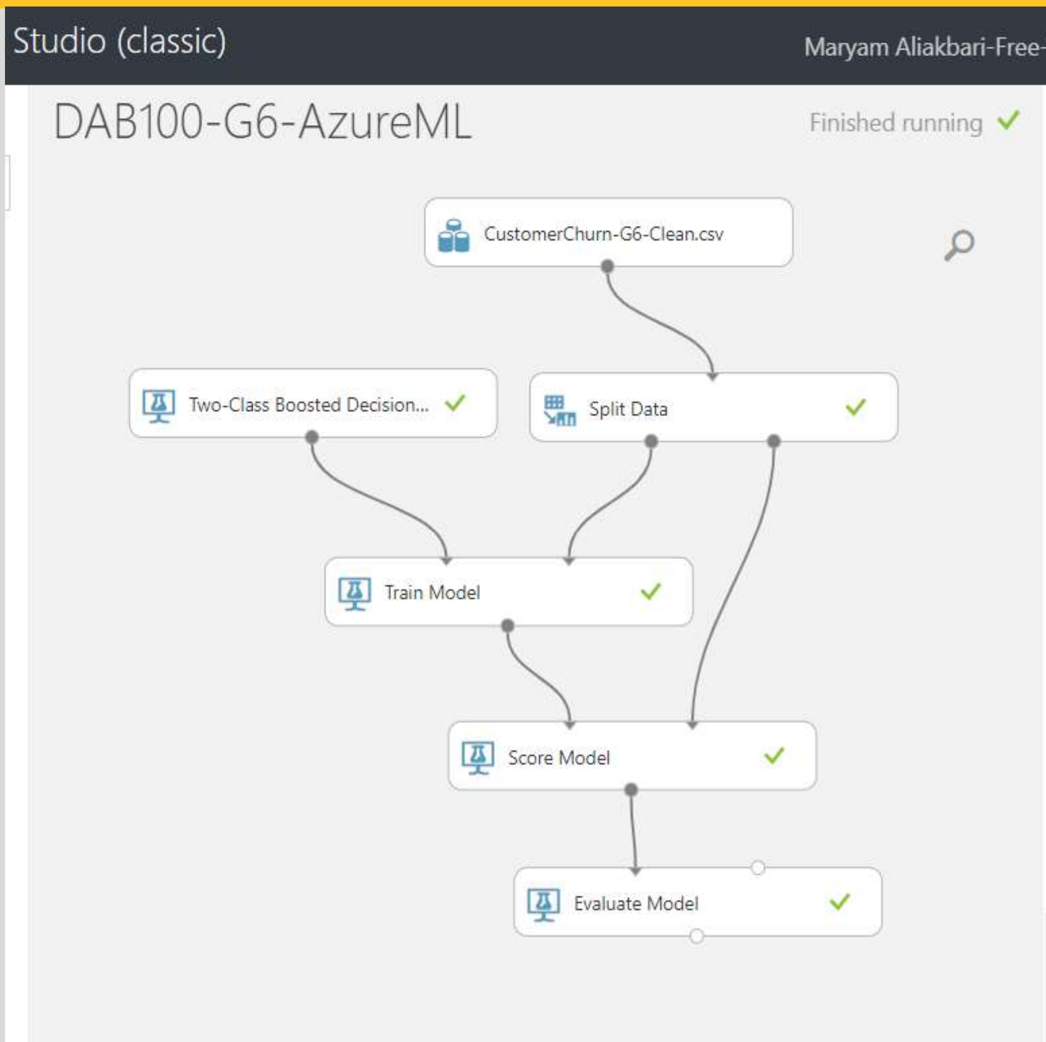
Recommendations to Decrease Churn Rate

- Promotion to increase dependents/partners
- Special offers for long-term contracts
- Offering bundle services
- Offering loyalty rewards

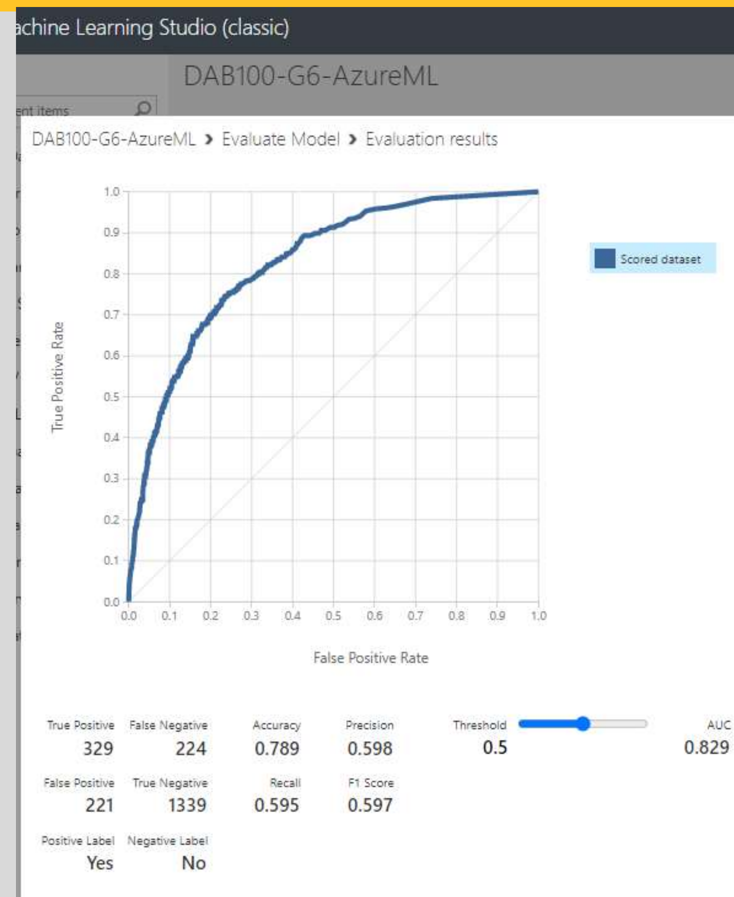
About the Models

Problem Statement	Build a model to Predict "Churned Customers" in a telecommunication company
Target Variable	Churn
Selected Columns	All Columns
Fraction of Data for Modeling	70%
Classification Model	Two-Class Boosted Decision Tree & Two-Class Logistic Regression Model

Decision Tree Model

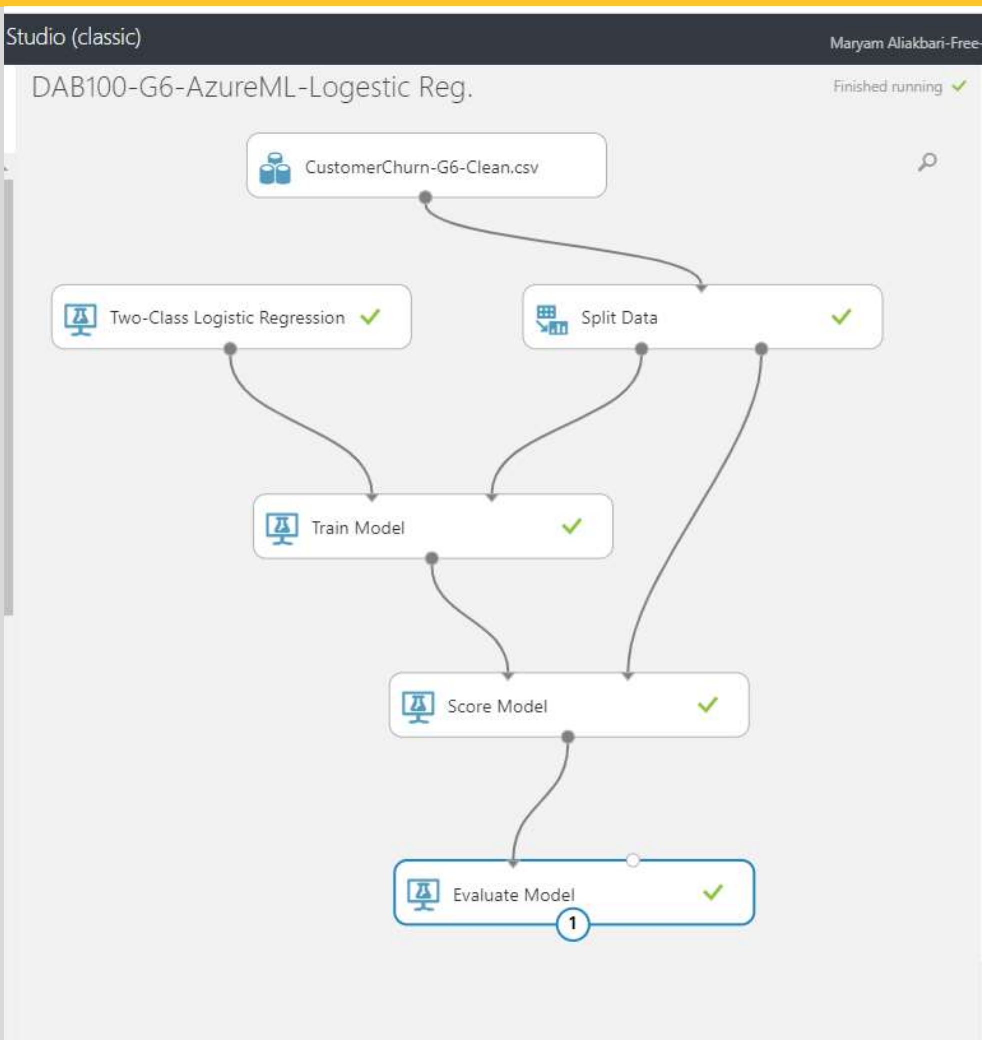


Result



Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score
(0.900,1.000]	172	64	0.112	0.789	0.436
(0.800,0.900]	64	45	0.163	0.798	0.526
(0.700,0.800]	43	39	0.202	0.800	0.569
(0.600,0.700]	29	42	0.236	0.794	0.586

Logistic Regression Model



Result

Compare the Results of the Models

Logistic Regression Result

True Positive	False Negative
309	224
162	1398
False Positive	True Negative

Accuracy	0.808
Precision	0.656
Recall	0.559
F1 Score	0.604
AUC	0.839

Decision Tree Result

True Positive	False Negative
329	224
221	1339
False Positive	True Negative

Accuracy	0.789
Precision	0.598
Recall	0.595
F1 Score	0.597
AUC	0.829

Color Guide

- Better Result
- Worse Result
- Equal

Positive Label
Yes

In overall the result of Logistic Regression Model is better than the result of Decision Tree Model