

DAB 100 - Introduction to Data Analytics

Final Group Project – Group #5

Table of contents



Team introduction and contribution



Data Cleaning



Data Transformation



Data Visualization



Azure ML model



Conclusion



References

Team introduction and contribution

Kavya	-	Data cleaning
-------	---	---------------

Kavya	-	Data cleaning
-------	---	---------------

Kailash	-	Data Visualization
---------	---	--------------------

Kailash	-	ML
---------	---	----

Kailash	-	Conclusion
---------	---	------------

Data cleaning



The given Customer churn dataset contains corresponding to various customer properties and subscription.



Checked all columns for null values and found null values in “TotalCharges” variable.



Removed unwanted columns ("Dependents", "OnlineSecurity", "OnlineBackup", "DeviceProtection") from the data set.

Data transformation

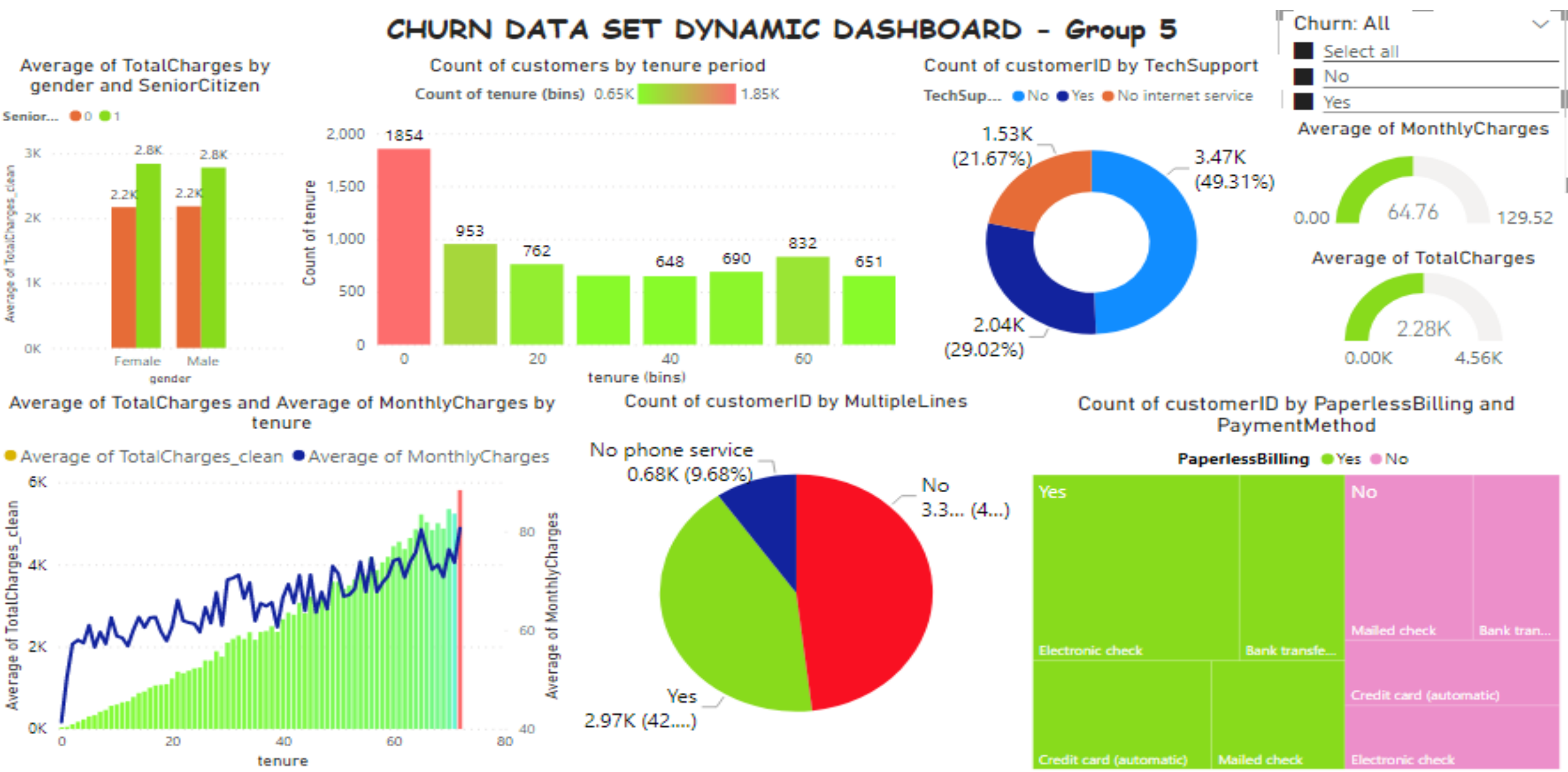
Created a new column for having TotalCharges variable filled.

All the observations with null TotalCharges values had 0 as tenure.

New column named as TotalCharges_clean is created using the below formula

- If (TotalCharges > 0 & tenure > 0) then TotalCharges else MonthlyCharges

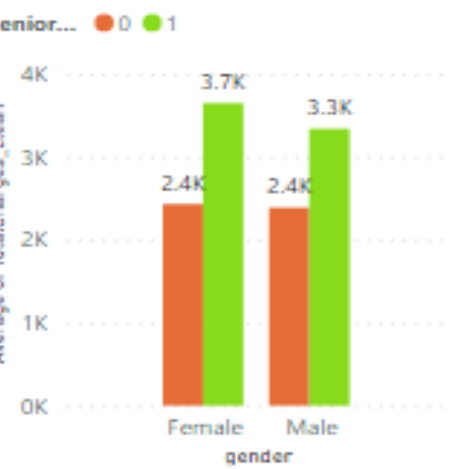
Data Visualization – All customers



Data Visualization – Non-churned customers

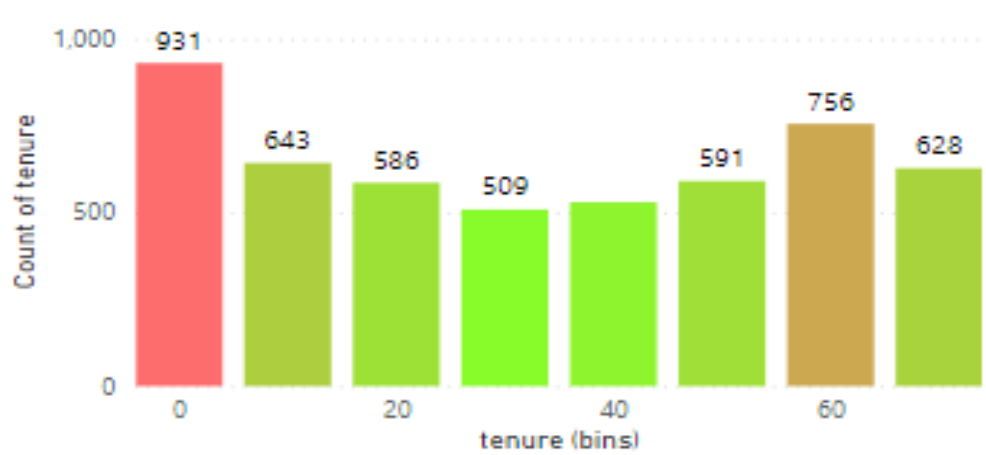
CHURN DATA SET DYNAMIC DASHBOARD - Group 5

Average of TotalCharges by gender and SeniorCitizen



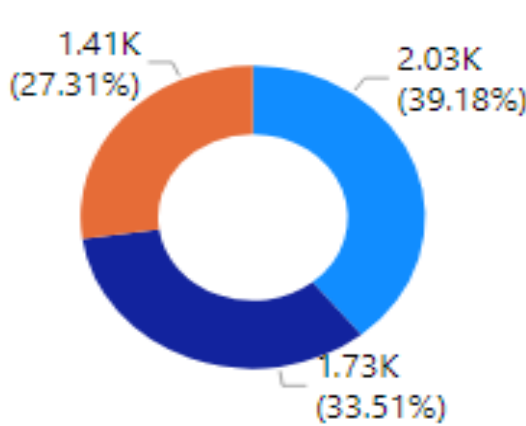
Count of customers by tenure period

Count of tenure (bins) 509 931



Count of customerID by TechSupport

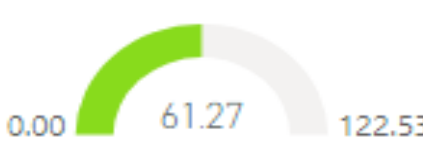
TechSup... ● No ● Yes ● No internet service



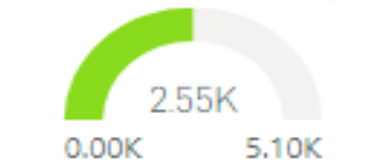
Churn: Multiple selection...

- ☐ Select all
- ☒ No
- ☐ Yes

Average of MonthlyCharges

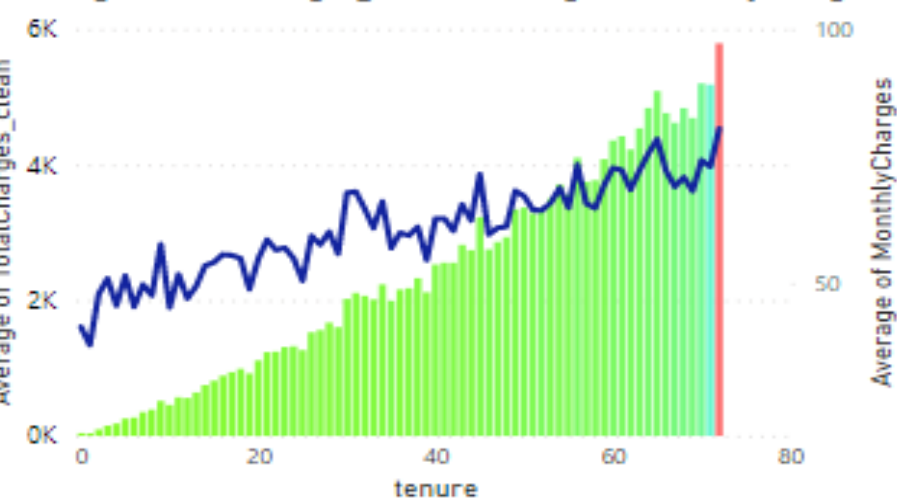


Average of TotalCharges

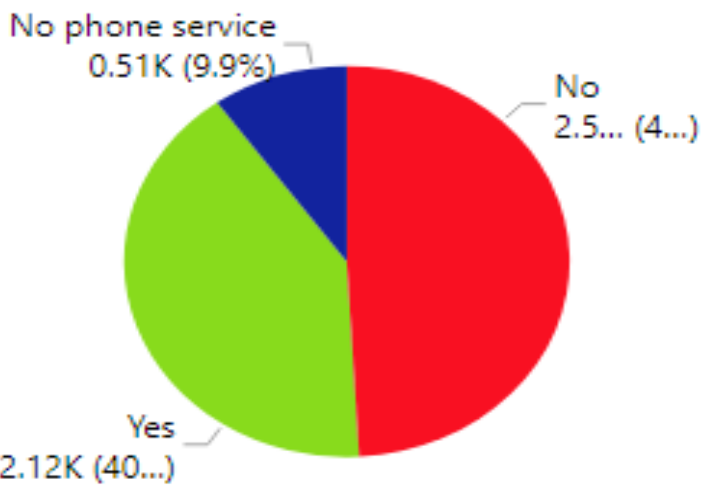


Average of TotalCharges and Average of MonthlyCharges by tenure

● Average of TotalCharges_clean ● Average of MonthlyCharges

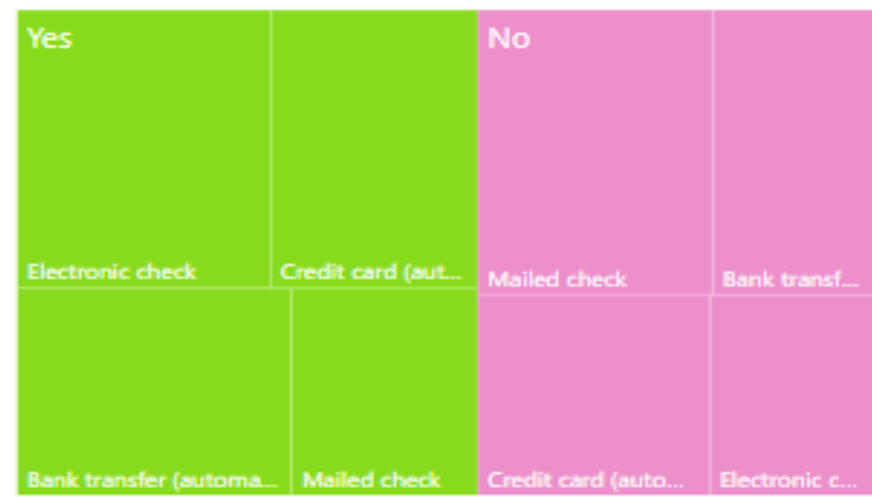


Count of customerID by MultipleLines



Count of customerID by PaperlessBilling and PaymentMethod

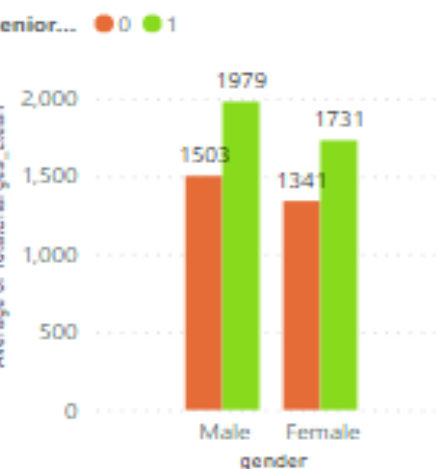
PaperlessBilling ● Yes ● No



Data Visualization – Churned customers

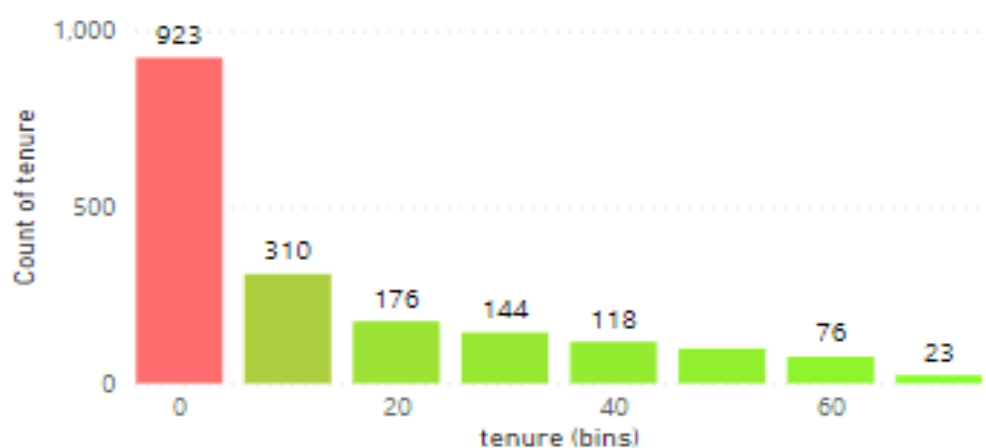
CHURN DATA SET DYNAMIC DASHBOARD - Group 5

Average of TotalCharges by gender and SeniorCitizen



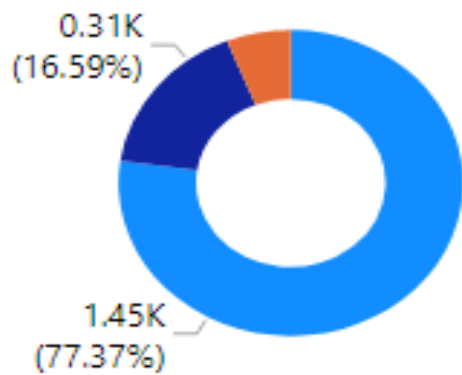
Count of customers by tenure period

Count of tenure (bins) 23 923



Count of customerID by TechSupport

TechSup... No Yes No internet service



Churn: Multiple selection...

- Select all
- No
- Yes

Average of MonthlyCharges

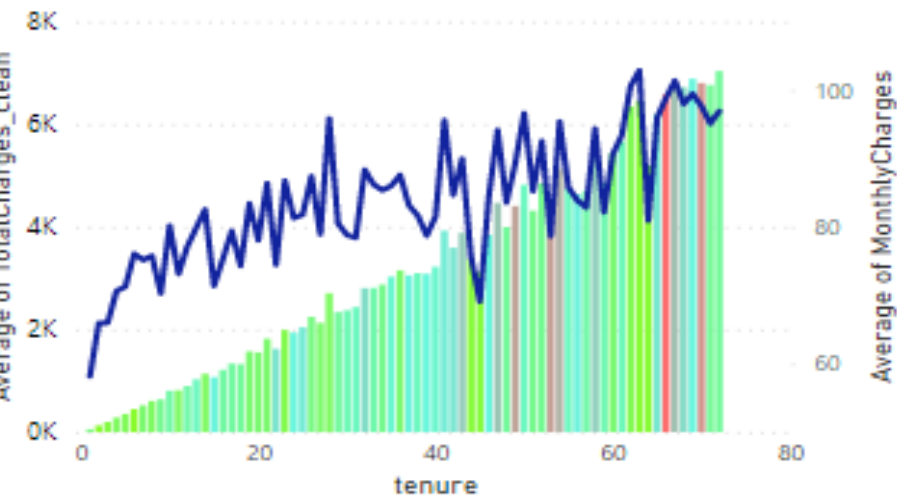


Average of TotalCharges



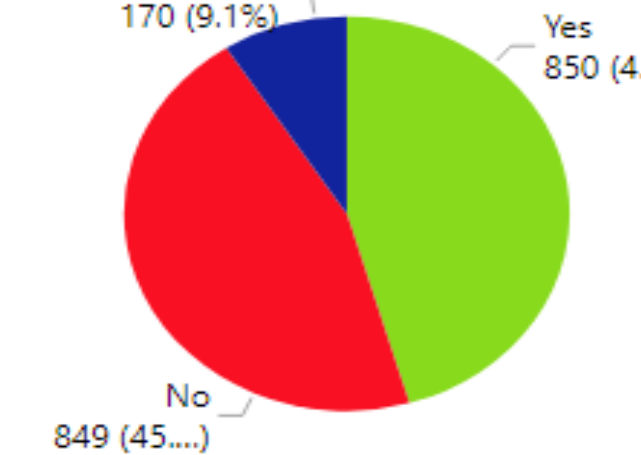
Average of TotalCharges and Average of MonthlyCharges by tenure

Average of TotalCharges_clean Average of MonthlyCharges



Count of customerID by MultipleLines

No phone service 170 (9.1%)

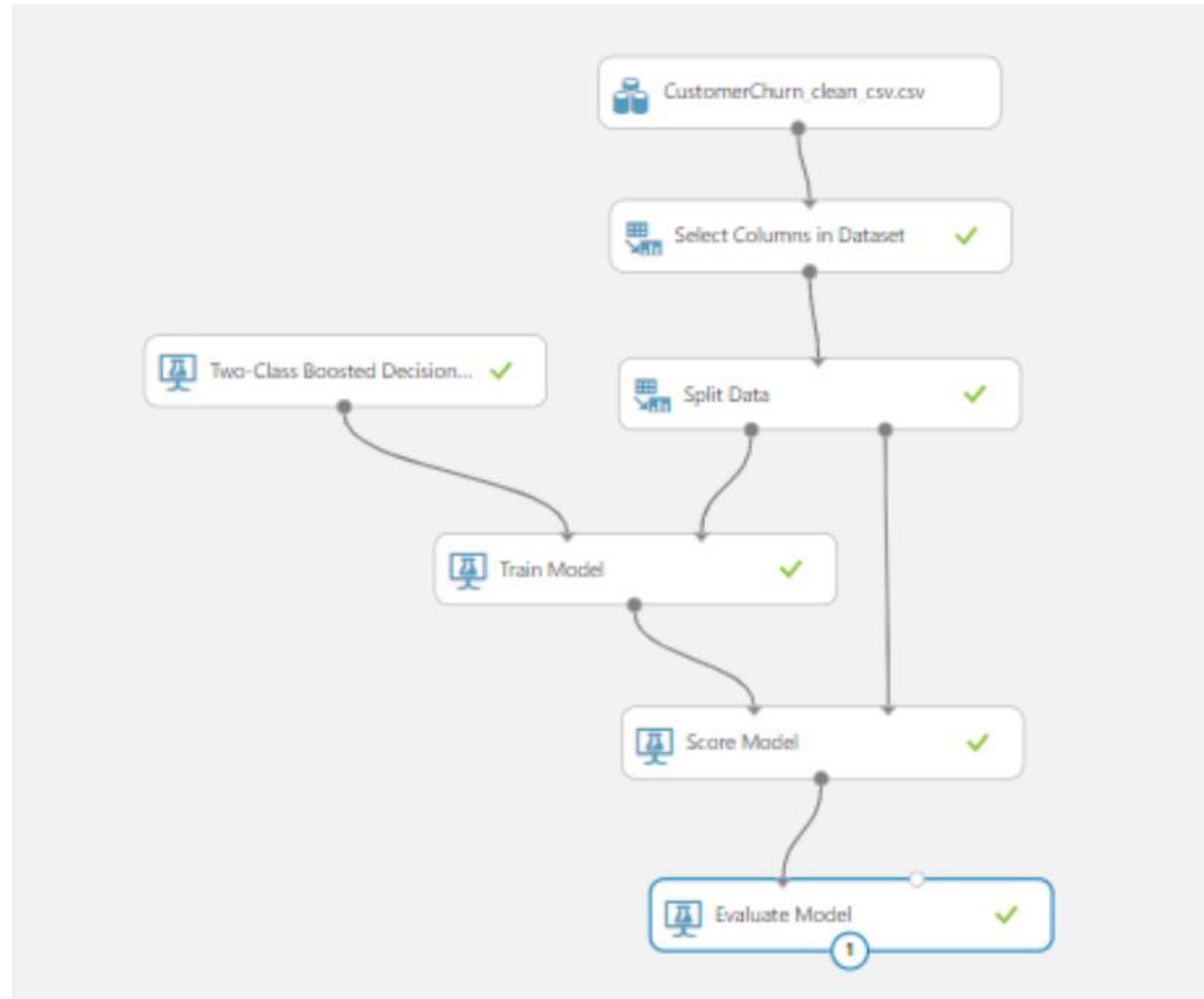


Count of customerID by PaperlessBilling and PaymentMethod

PaperlessBilling Yes No



Azure – Two Class Booster Decision Tree

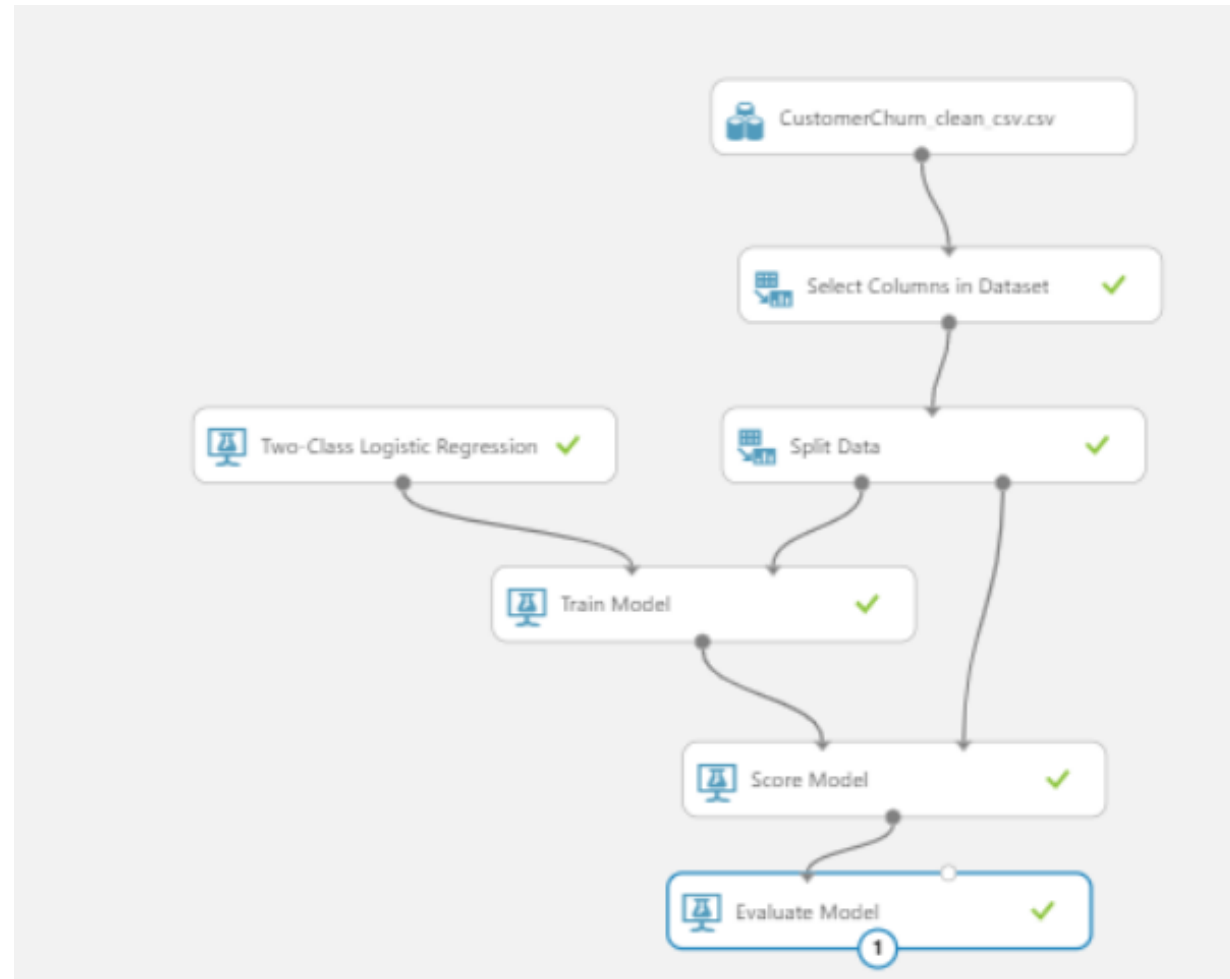


Azure – Two Class Booster Decision Tree

Observation from the Two Class Booster Decision tree with threshold value of 0.5.
The model is with 77.6% accuracy.

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
284	272	0.776	0.586	0.5	0.816
False Positive	True Negative	Recall	F1 Score		
201	1356	0.511	0.546		
Positive Label	Negative Label				
Yes	No				

Azure – Logistic Regression model



Azure – Logistic Regression model

Observation from Logistic Regression model with threshold 0.5.
The accuracy is 78.2% in predicting the churn in this model.

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
262	294	0.782	0.612	0.5	0.831
False Positive	True Negative	Recall	F1 Score		
166	1391	0.471	0.533		
Positive Label	Negative Label				
Yes	No				

Conclusion

- We have created the plot between True Positive rate and False Positive rate at varying thresholds
- The area under the curve is 0.812 and as we know that higher AUC leads to a better model because it's able to distinguish the positive values from the negative values.
- A threshold value of 0.5 implies that if the probability is more than 0.5, it is considered as likely, i.e. Positive.
- From this we're able to conclude that Logistic accuracy is more than the decision tree model.

Conclusion

- From the data and the visualizations we could conclude that most of the churned customers are the new/customers within 1 year of service tenure.
- Tech support facility has not been used by most of the customers who has churned.
- Most of the customers who churned use the electronic payment method.
- To reduce the churn the telecom service provider can provide exiting offers to increase their tenure period thereby reducing the overall churn.
- Also proper awareness is to be created among the customers to make use of tech-support effectively.

References

- <https://www.tutorialgateway.org/create-bins-in-power-bi/>