



Telco Customer Churn Study

GROUP 10:

OMAR ALTAHER

MODUANA POBEE-ORLEANS

SURYA CHAITANYA KOKKIRIMETLA

Business Case

The Telco company is facing a high customer churn rate. If not addressed, they would harm revenue and growth.

This project analyzes data from the company provided on the IBM site to comprehend the causes of churn and anticipate at-risk clients, we will assist the organization to apply focused tactics to boost retention, revenue, and profitability.

Target variable is customer churn.

The business objective is to determine why customers are churning and how to minimize it.

Data

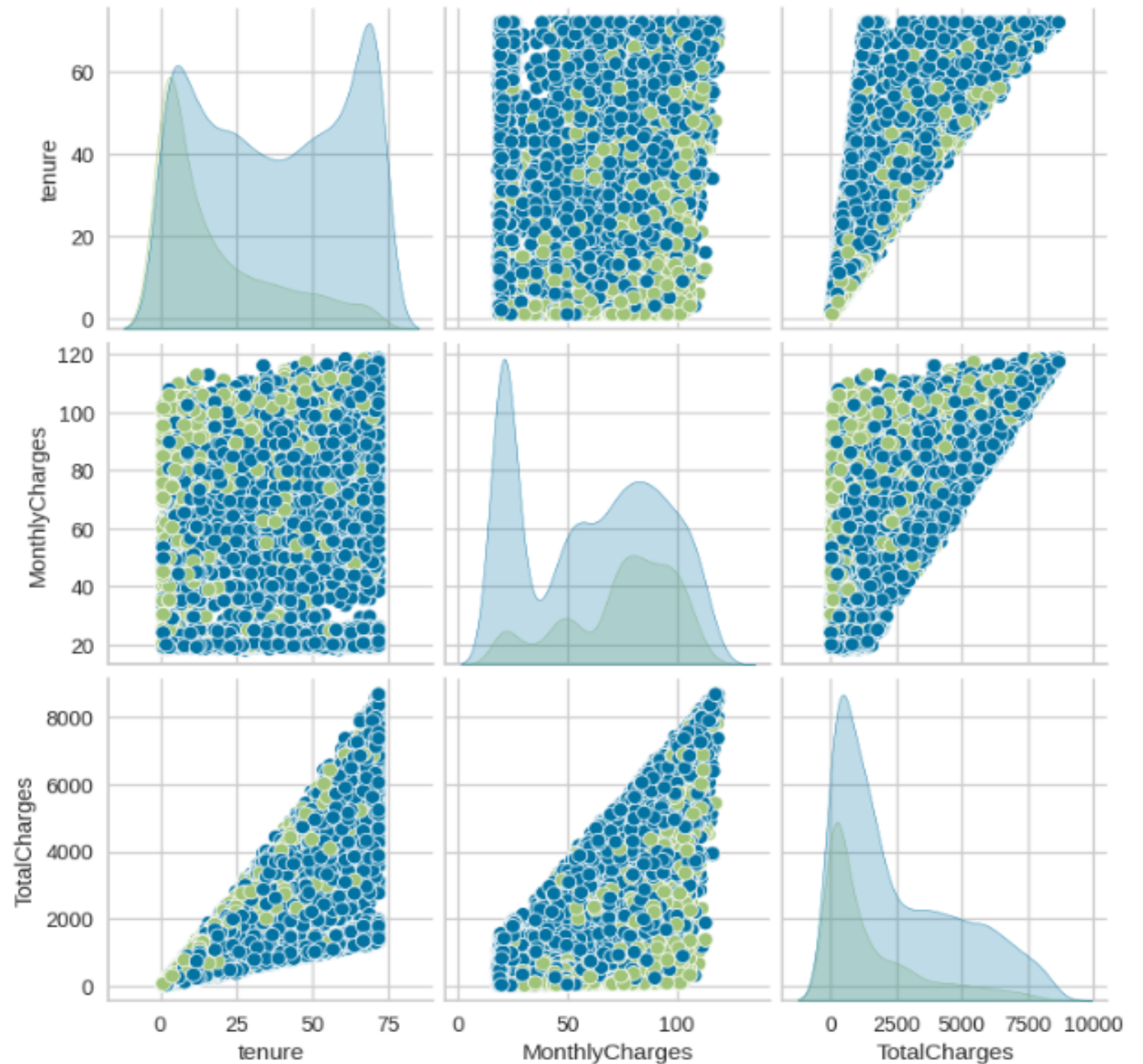
- The data includes
 - Demographic: gender, age, and dependents
 - Detailed service plans: phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
 - Customer churn
 - Contract: monthly or yearly subscription
 - Tenure etc...

```
tcc_data.dtypes
```

gender	object
SeniorCitizen	int64
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	object
Churn	object
dtype:	object

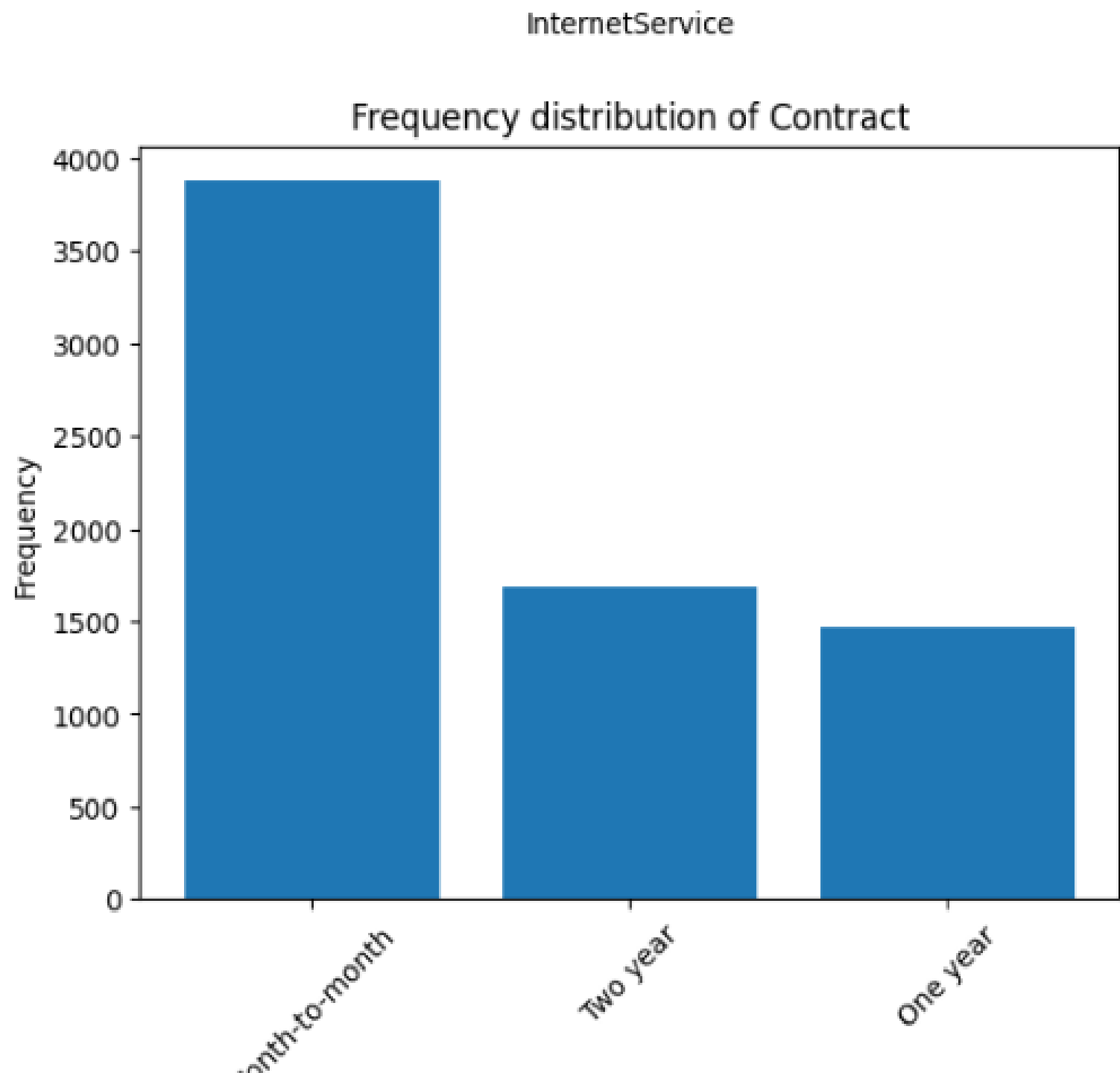
Data Variables

- Seaborn Pair plot shows the relationship between variables
- Based on the chart the customers with the highest monthly charges are more likely to churn
- The customer is less likely to churn the longer they are with the company

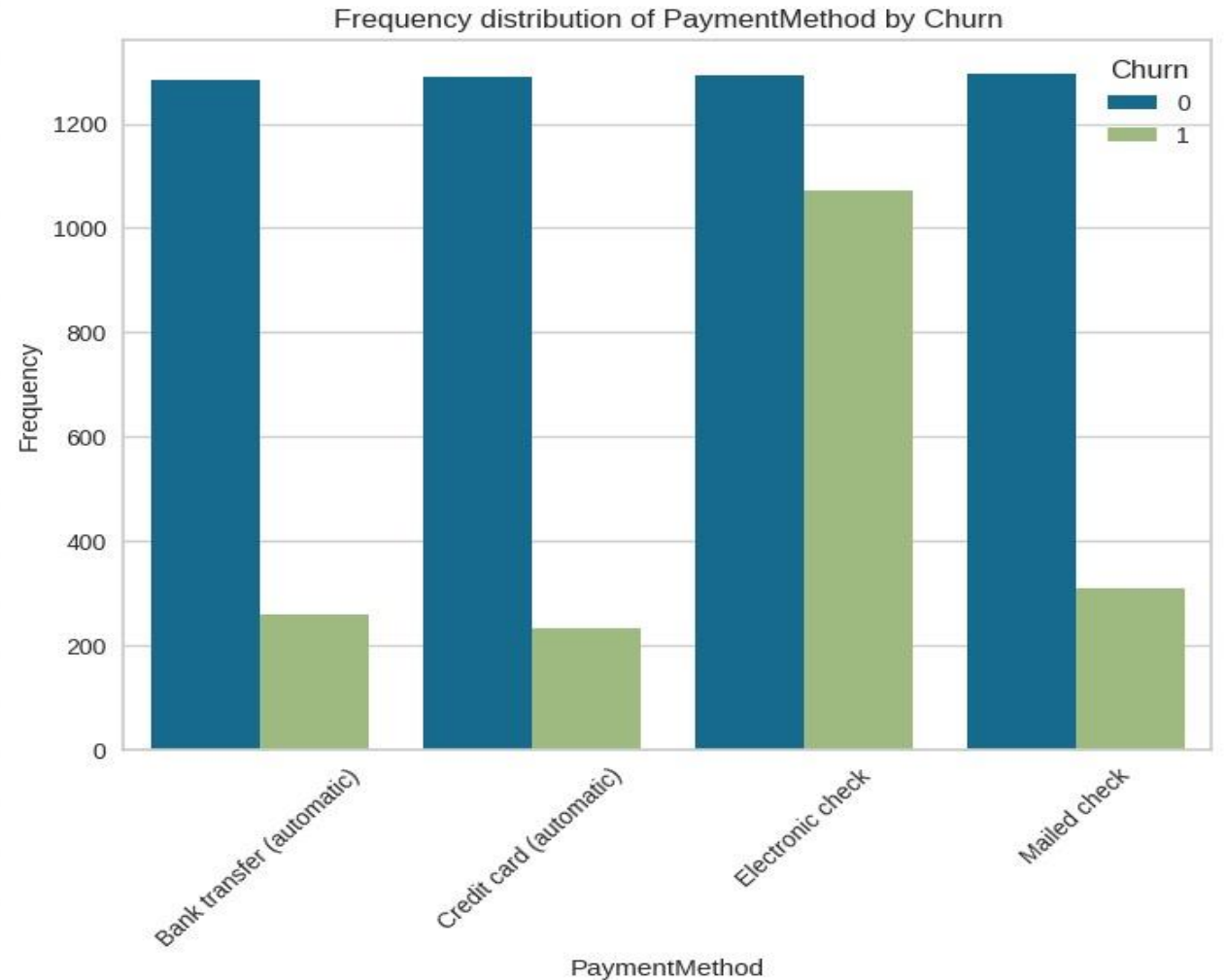
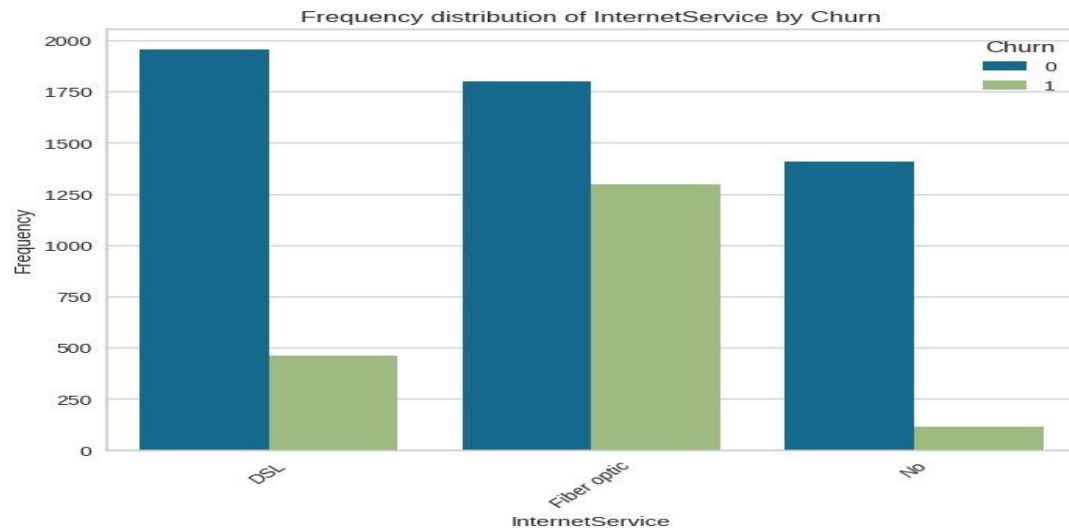
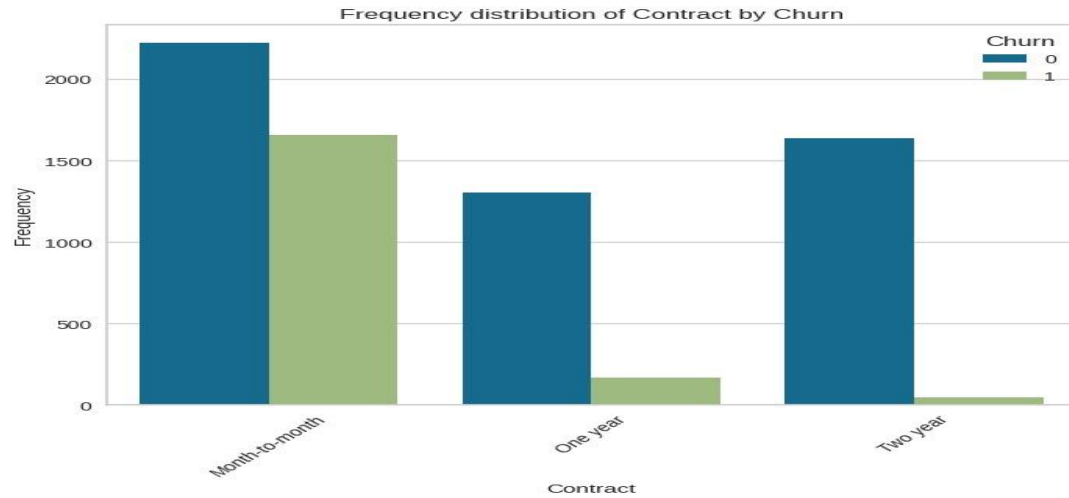


Data Variables

- The frequency distribution of contracts shows the plan Telco customers have
- Based on the chart more than half of telcos customers are on a monthly plan and the other half is split between one-to-two-year contract

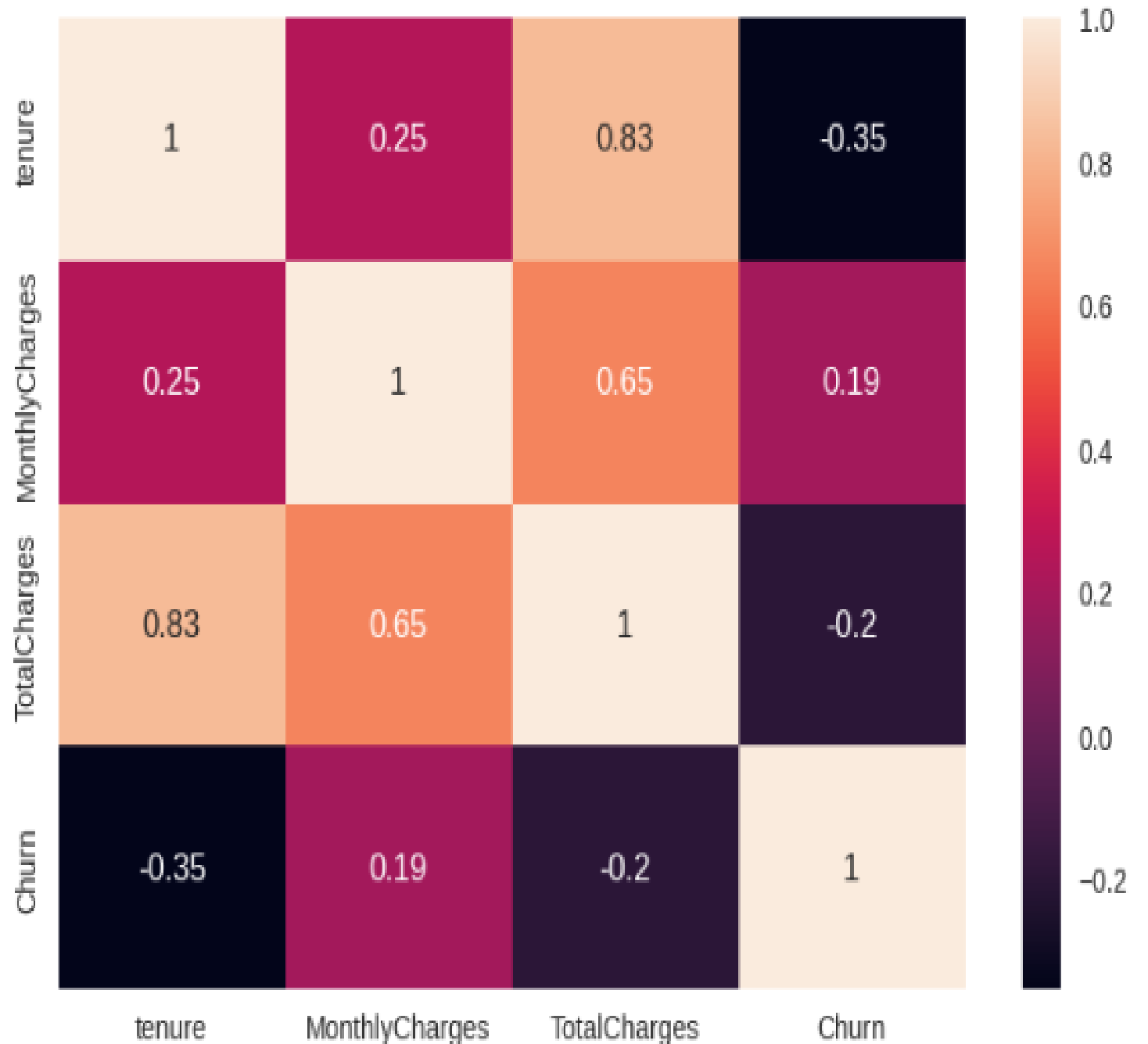


Frequency Distribution



Data Variables

- Correlation Heatmap: the heat map visualizes the correlation between the variables
- The boxes with higher numbers mean there is a stronger correlation between the variables



Models

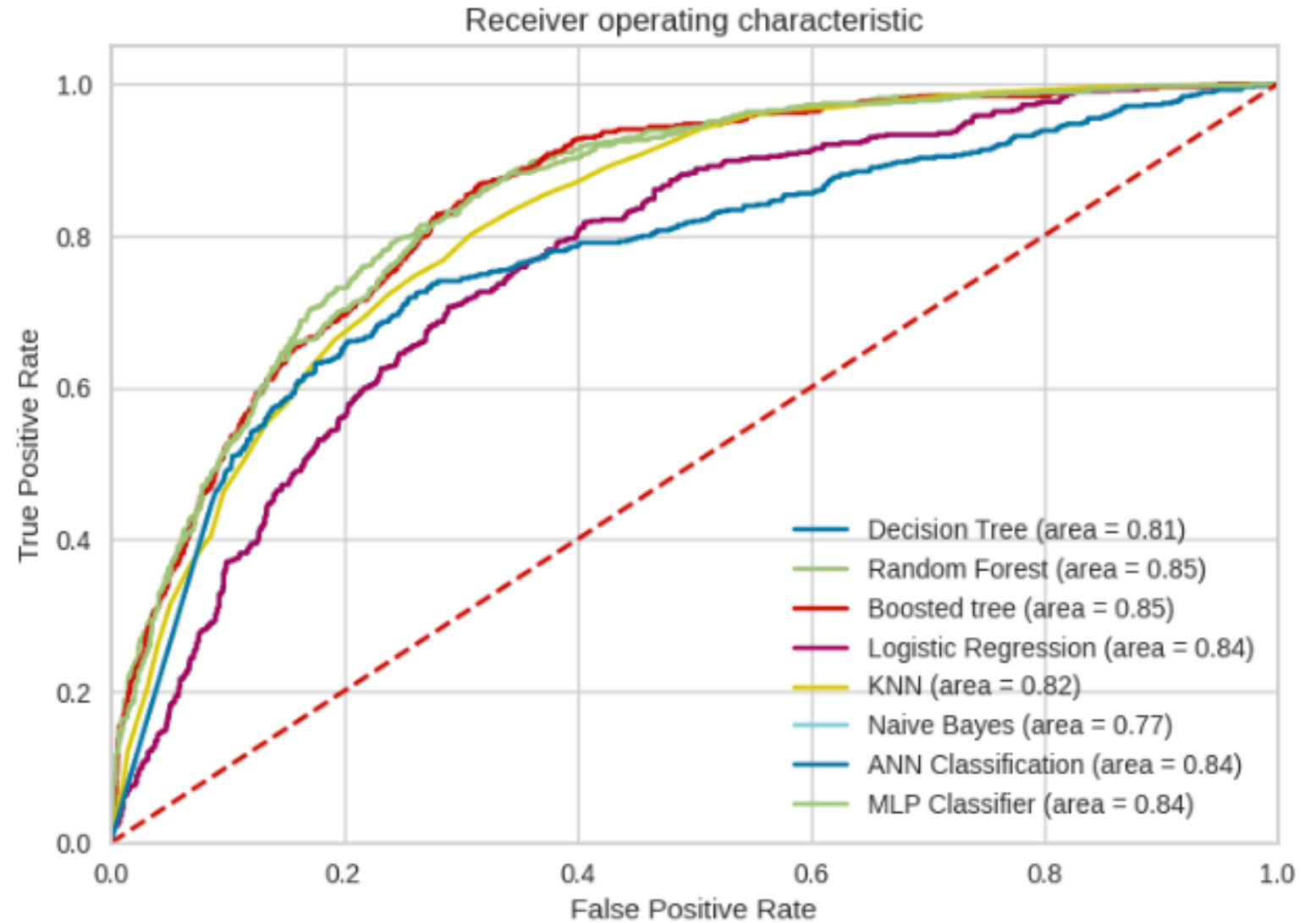


Model Approaches

	Model	Accuracy	ROC AUC Score	Precision	Recall	F1-Score
0	Logistic Regression	0.798104	0.839783	0.788657	0.798104	0.791254
1	KNN	0.783886	0.824086	0.779177	0.783886	0.781219
2	Random Forest	0.797630	0.849667	0.784259	0.797630	0.781613
3	Boosted Tree	0.800948	0.845691	0.791336	0.800948	0.793721
4	Decision Tree	0.796682	0.811354	0.783176	0.796682	0.778610
5	Naive Bayes	0.692417	0.769322	0.769627	0.692417	0.710167
6	ANN Classification	0.797156	0.844935	0.789321	0.797156	0.791972
7	MLP Classifier	0.796682	0.844935	0.789760	0.796682	0.792296

Best Model

- The best model based on the ROC AUC score is the random forest with a score of 0.849



Findings & Recommendations

■ Findings:

- The Customers are more likely to churn with low tenure
- The lower the monthly charge the less likely the customer will churn
- Customers are churning because of high monthly charges

■ Recommendations:

- Conducting surveys on understanding customer needs
- Offering customers discounts on longer subscriptions
- Focusing on customers that are less likely to churn due to customer age, contract type, partners, and dependents

Conclusion

- Comparing all the models' random forest is the best model with a ROC AUC score of 0.849
- Based on the charts Telcos customers are churning due to high monthly charges and the customers are likely to cancel their subscription if it is a month-to-month contract

Best models based on Pycaret

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ada	Ada Boost Classifier	0.8050	0.8483	0.5535	0.6596	0.6016	0.4738	0.4772	0.9180
nb	Naive Bayes	0.6963	0.8242	0.8509	0.4616	0.5984	0.3872	0.4342	0.9420
lr	Logistic Regression	0.8003	0.8475	0.5512	0.6460	0.5948	0.4633	0.4660	2.1860
gbc	Gradient Boosting Classifier	0.8045	0.8476	0.5344	0.6648	0.5923	0.4658	0.4707	2.0080
lda	Linear Discriminant Analysis	0.7964	0.8394	0.5551	0.6336	0.5916	0.4569	0.4587	0.4700
lightgbm	Light Gradient Boosting Machine	0.7987	0.8383	0.5375	0.6467	0.5865	0.4550	0.4588	0.8880
qda	Quadratic Discriminant Analysis	0.6737	0.8312	0.8638	0.4437	0.5844	0.3599	0.4177	0.4660
ridge	Ridge Classifier	0.7983	0.0000	0.5138	0.6533	0.5751	0.4453	0.4509	0.4220
rf	Random Forest Classifier	0.7974	0.8277	0.4977	0.6574	0.5663	0.4374	0.4448	2.1640
xgboost	Extreme Gradient Boosting	0.7842	0.8249	0.5122	0.6140	0.5581	0.4169	0.4202	1.8560
et	Extra Trees Classifier	0.7771	0.7987	0.4809	0.6011	0.5342	0.3901	0.3944	1.5760
knn	K Neighbors Classifier	0.7655	0.7465	0.4297	0.5799	0.4934	0.3453	0.3520	1.1780
dt	Decision Tree Classifier	0.7227	0.6474	0.4862	0.4781	0.4818	0.2926	0.2928	0.4680
svm	SVM - Linear Kernel	0.7741	0.0000	0.3242	0.6548	0.4273	0.3083	0.3401	0.4820
dummy	Dummy Classifier	0.7343	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8460

Reference

- ❑ <https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=samples-telco-customer-churn>
- ❑ https://accelerator.ca.analytics.ibm.com/bi/?perspective=authoring&pathRef=.public_folders%2FIBM%2BAccelerator%2BCatalog%2FContent%2FDAT00148&id=i9710CF25EF75468D95FFFC7D57D45204&objRef=i9710CF25EF75468D95FFFC7D57D45204&action=run&format=HTML&cmPropStr=%7B%22id%22%3A%22i9710CF25EF75468D95FFFC7D57D45204%22%2C%22type%22%3A%22reportView%22%2C%22defaultName%22%3A%22DAT00148%22%2C%22permissions%22%3A%5B%22execute%22%2C%22read%22%2C%22traverse%22%5D%7D
- ❑ <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

