<u>Customer Churn Prediction – Final Report</u>

By Akhil Nair (UCF ID: 5562572)

1 Abstract:

In recent years, the telecommunications industry has undergone significant transformations, including market liberalization, the introduction of new services, and advancements in technology. These changes have intensified competition, leading to increased customer churn - a major challenge that leads to significant losses for service providers. Customer churn in the telecom industry, therefore, has become a critical issue.

To address this, the industry has turned to data mining techniques for effective churn prediction, which is crucial for customer retention strategies. Although much of the churn prediction research has focused on mobile voice services, this report explores both mobile and landline services. We employ a comprehensive analysis of various data points including customer demographics, contractual details, payment method types, carrier data, complaints, and financial transactions.

This project uses advanced data visualization to derive insights and come up with potential strategies to mitigate churn. Machine learning models like linear regression, logistic regression, decision trees, and random forests have been used for churn prediction. This approach not only highlights the problem but also provides actionable solutions to retain customers in the competitive telecommunications industry.

2 Introduction:

Customer churn significantly impacts telecommunications companies by directly affecting revenue, increasing customer acquisition costs, and reducing market share. It also harms the company's brand reputation, diminishes long-term customer relationships, and limits opportunities for upselling. Effectively managing churn is necessary for maintaining financial stability, ensuring that the company retains its competitive edge and keeps its investors' confidence.

2.1 Problem Statement:

In the telecommunications sector, predicting and managing customer churn is critical. This project seeks to analyze customer data, derive key insights, and predict churn. We want to come up with targeted strategies that improve customer retention, optimize marketing efforts, and enhance overall satisfaction, thereby driving business growth and efficiency.

2.2 Goal:

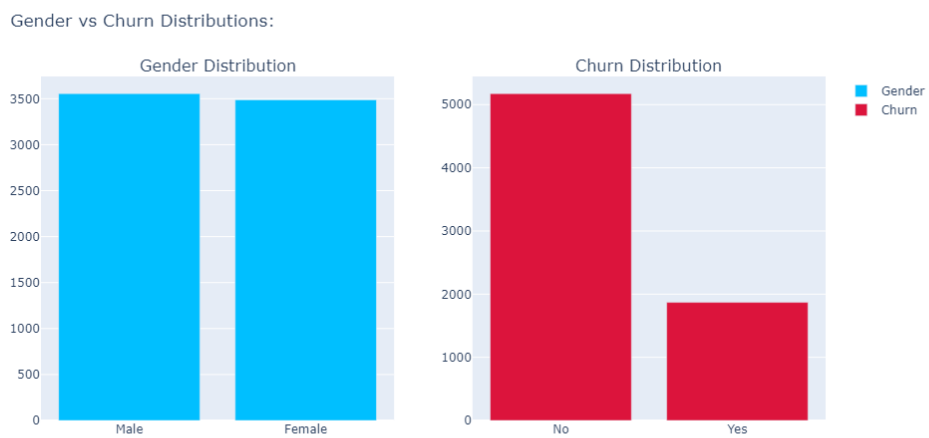
The goal of this project is to help mitigate customer churn at a telecom company. This would be accomplished by conducting extensive data analysis and data visualization to identify trends and understand customer behavior and gain insights into the factors that might be causing them to leave the company. We also make use of machine learning models to predict customer churn in this complex dataset, allowing the company to actively engage these at-risk customers, leading to better customer satisfaction and sustained business growth.

3 Dataset:

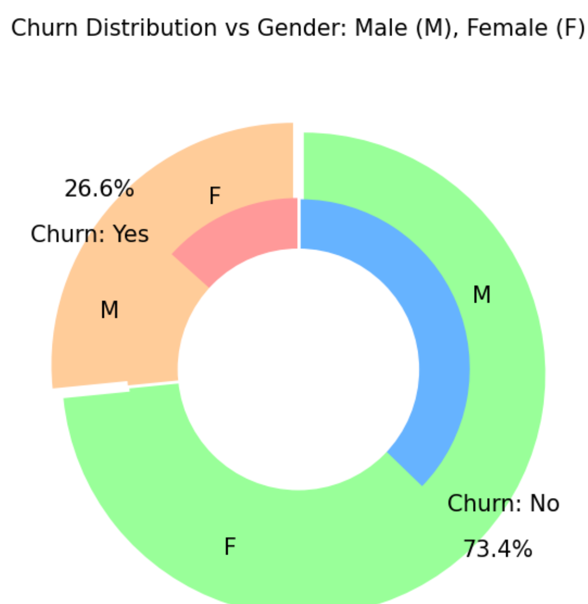
For this project we use the ‘Telco Customer Churn’ dataset (2019) from IBM Cognos Analytics, which contains 21 columns that contains the customers’ attributes and 7043 rows. This fictional dataset captures demographics, service subscriptions, billing details, and customer interaction data.

4 Data Visualization:

4.1 Gender

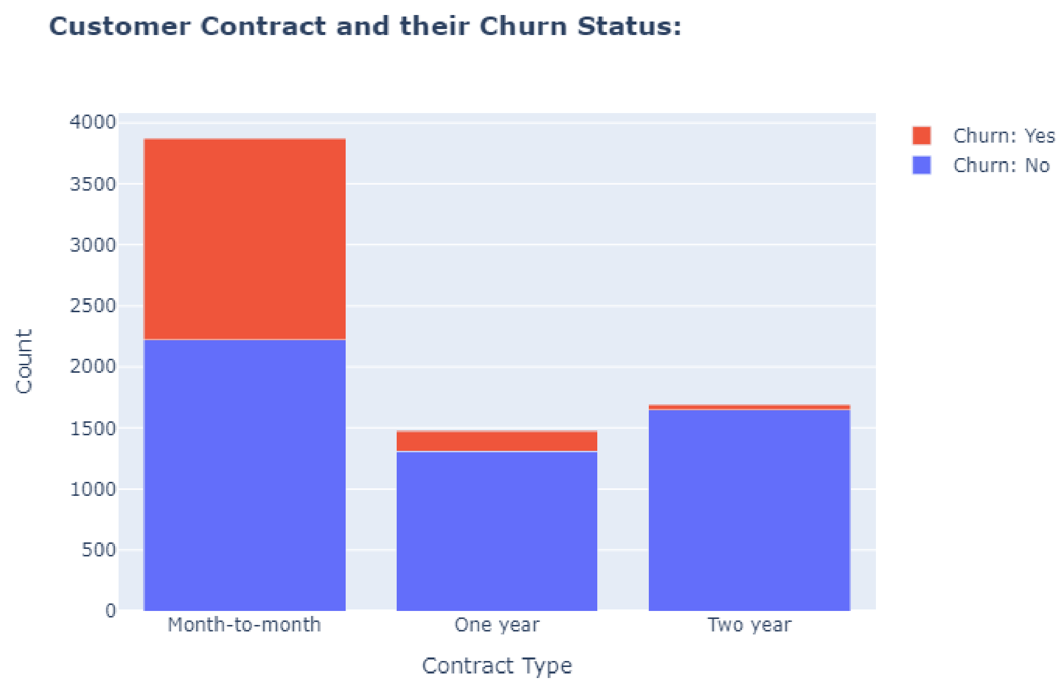


Starting with gender distribution, we find out that the genders are quite evenly distributed.

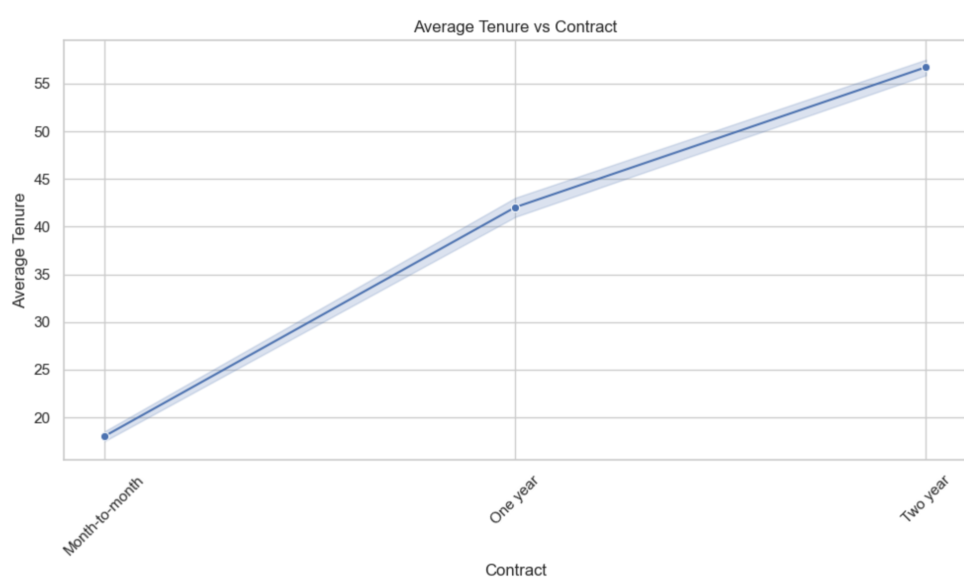


26.6% of the individuals churn, and 73.4% of them do not.

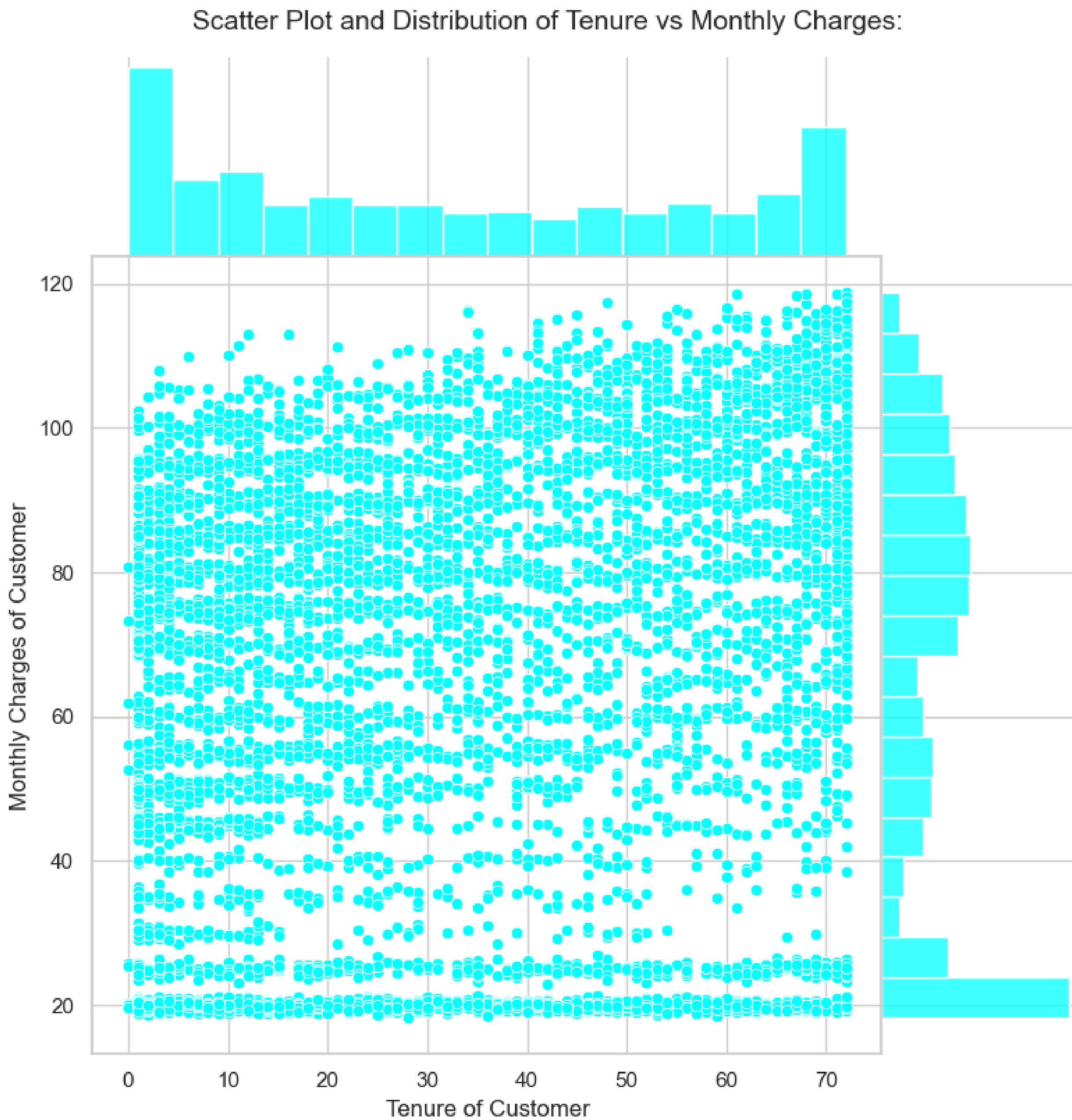
4.2 Duration of contract



We see that the customers with a month-to-month contract are much more likely to change providers, as opposed to those with yearlong or two yearlong contracts.

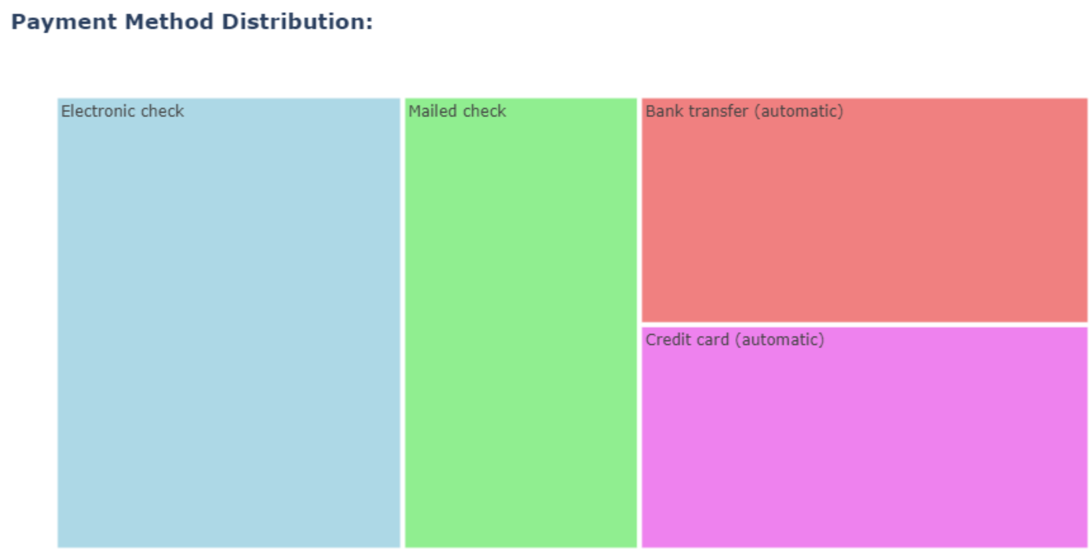


This shows us that smaller, month-to-month contracts show significantly higher churn rates than a year or two yearlong contracts.

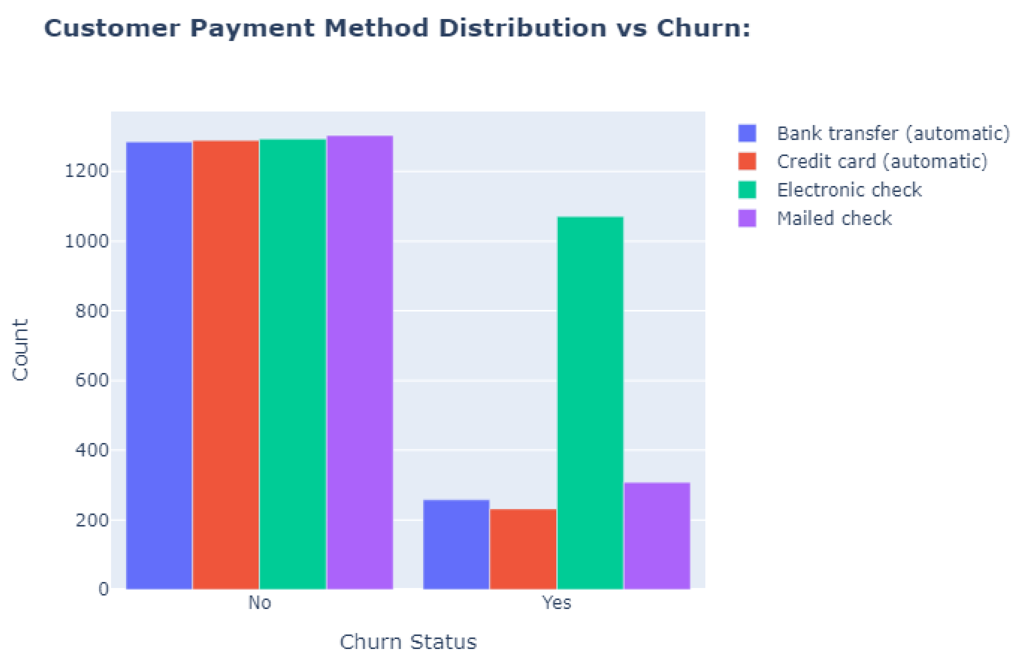


This might hint at lower monthly charges having the customers stay for longer periods. Customers with the highest monthly charges (over $100/month) have some of the highest attrition rates.

4.3 Types of Payment Methods



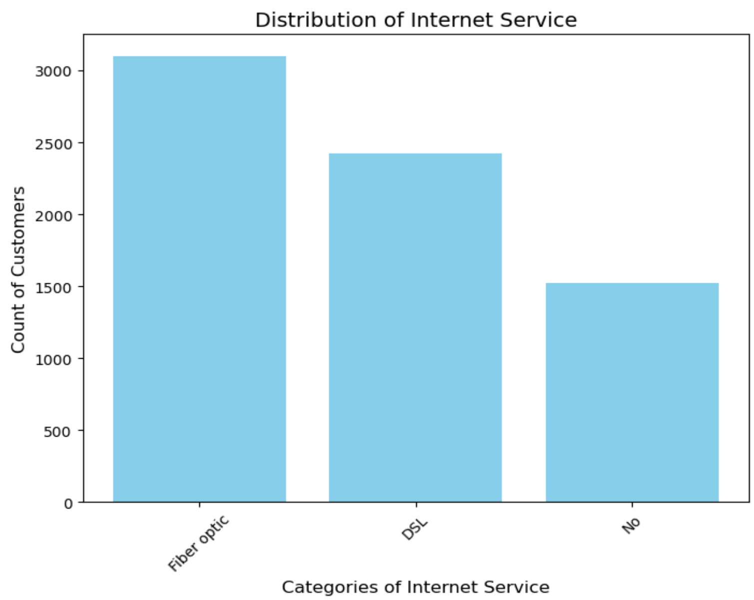
This tells us that the majority of customers pay by electronic check. Now we can try to identify the relation between these payment methods and churn.



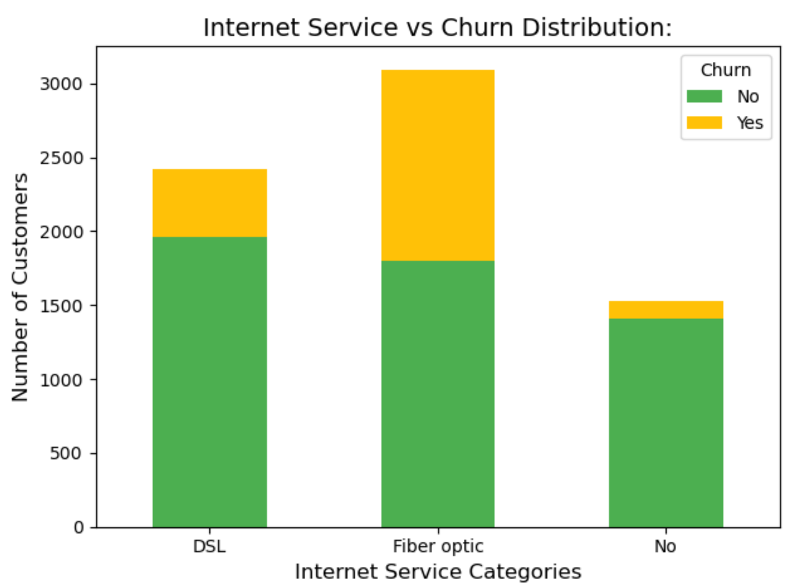
This tells us that there exists a clear relation between the customers who used Electronic checks and the ones that left the telecom company. The other payment methods have quite similar attrition rates, with mailed checks being related to more churn.

Automated bank transfer and credit card options have the lowest churn rates, which suggests having purely automated payment methods could in help reduce churn, but this might also cause the company to lose out on a large chunk of their customer base, since most customers pay by electronic check.

4.4 Types of Internet Service

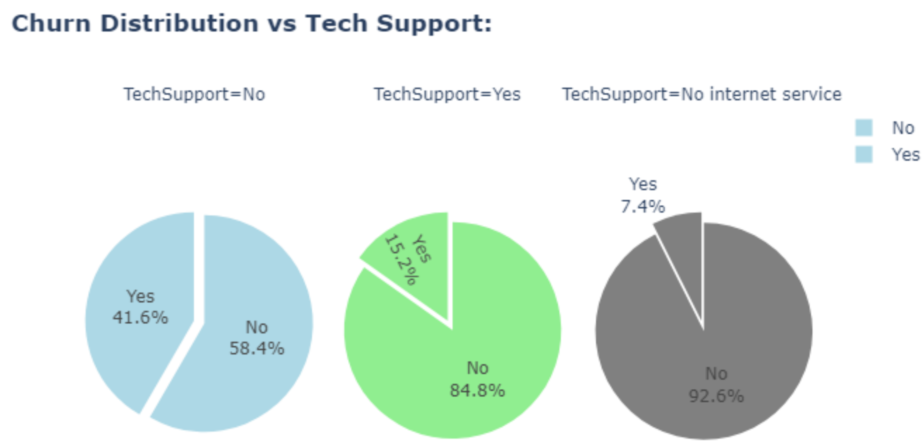


Most customers have opted for for fiber optic internet, followed by DSL, followed by no internet service at all.



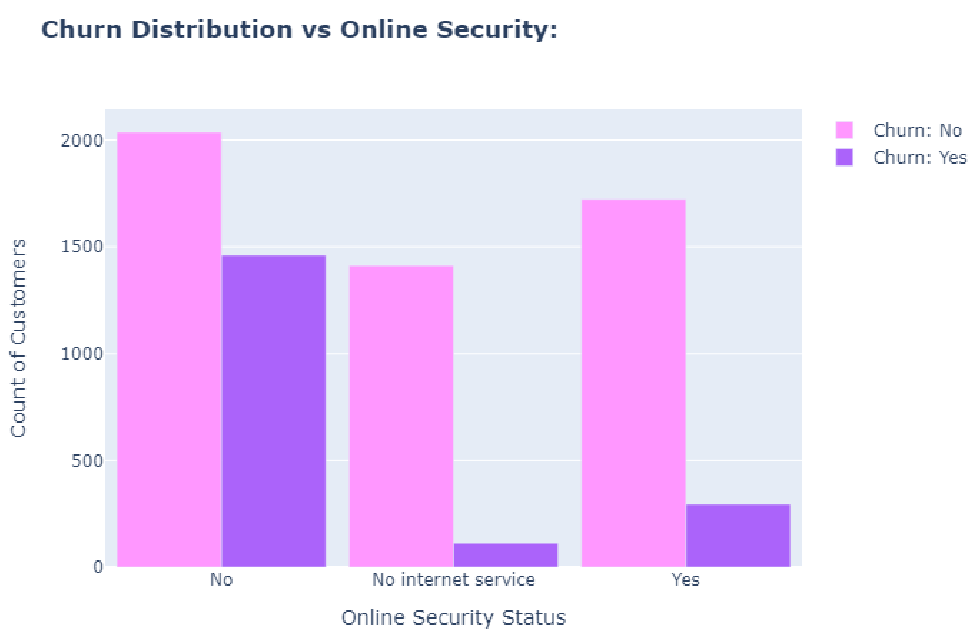
This tells us that users who opt for fiber optic internet are much more likely to churn, as opposed to those with no internet. A possible reason could be that they are most tech-savvy and could have found better deals at other telecom companies.

4.5 Tech Support offered



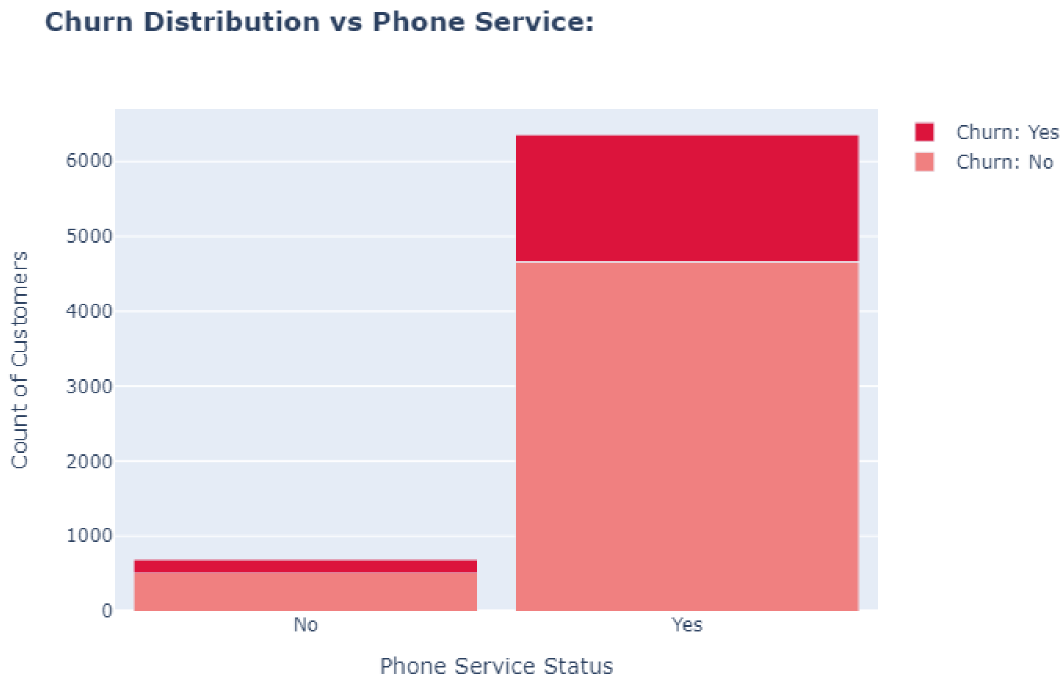
Customers who opt for tech support or are provided with it are significantly less likely to churn.

4.6 Online Security



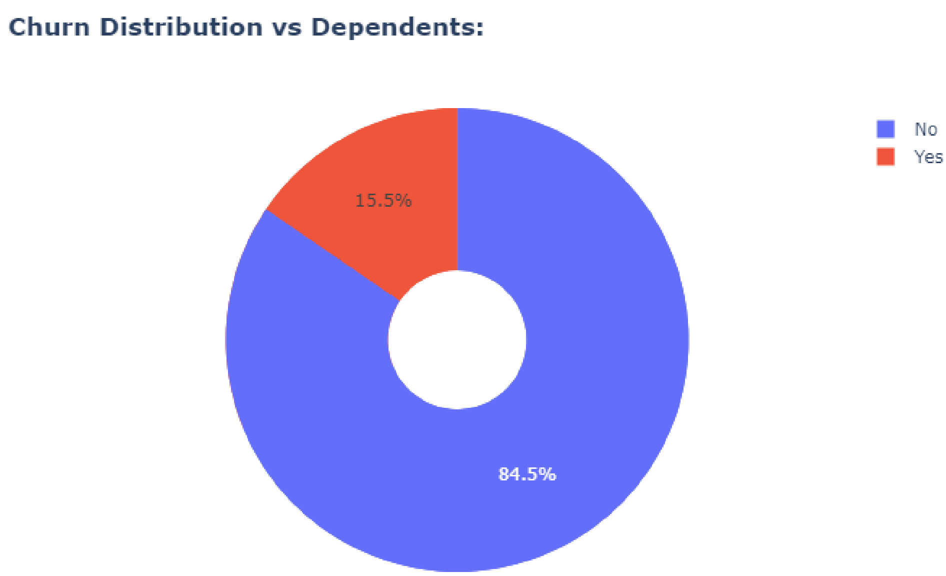
This tells us that customers with online security are far less likely to leave. Also, those with no plan for internet services are also not affected by the presence of online security.

4.7 Phone Service



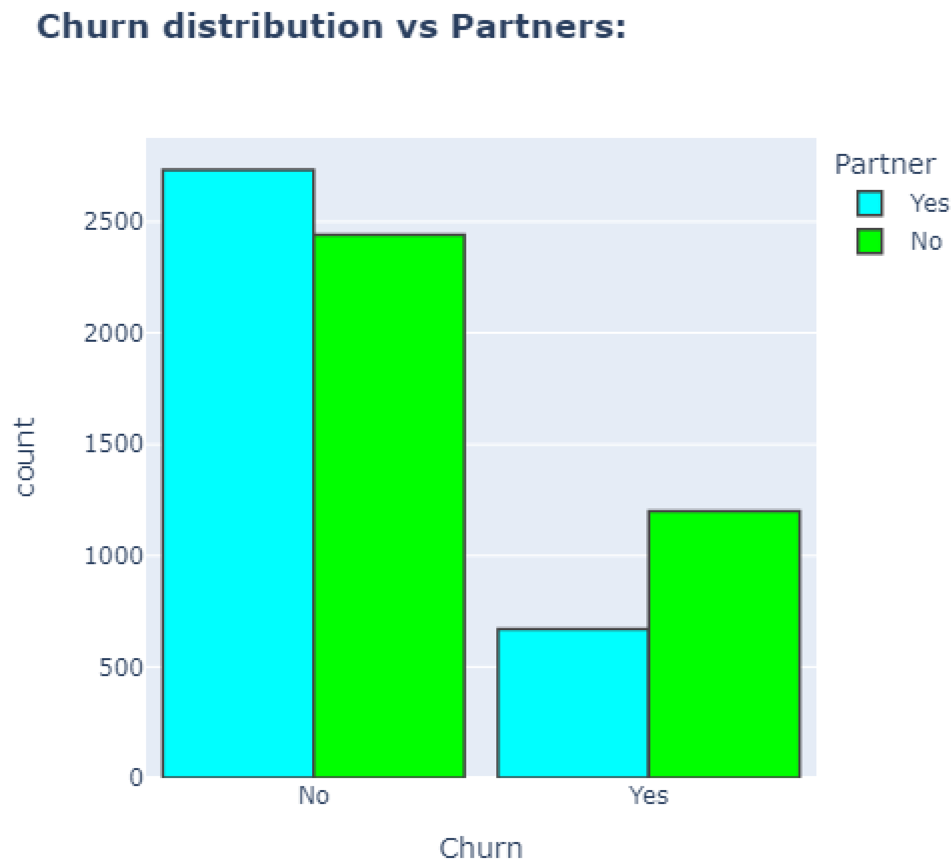
Most customers do have phone service, and about a fourth of those are likely to churn. Those with no phone plans are less likely to churn.

4.8 The presence of dependents



Individuals with dependents are much less likely to churn (84.5%) as opposed to those without any dependents.

4.9 The presence of partners



This tells us that individuals with no partners are much more likely to churn. Nearly half of those without partners end up leaving the service.

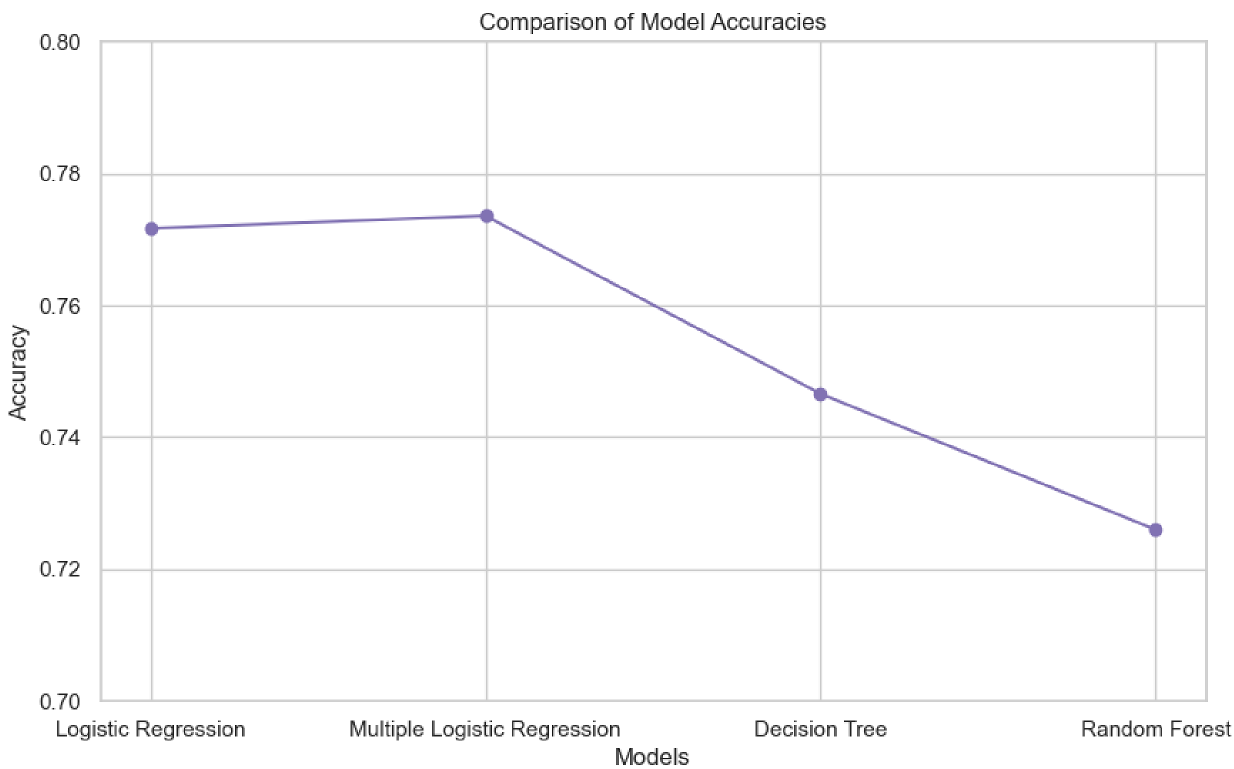
5 Model Development:

Logistic regression, multiple logistic regression, decision tree, and random forest algorithms are used to predict customer churn.

1. The Logistic Regression model uses the 'MonthlyCharges', 'Contract', 'InternetService', and 'PaymentMethod' features, which are transformed via one-hot encoding to fit categorical variables. The model is trained on 65% of the data, with the aim of predicting customer churn. Its Performance is evaluated using a confusion matrix, accuracy score, and a classification report. Its accuracy score of 0.77169 indicates a fairly high level of predictive accuracy.
2. The Multiple Logistic Regression model focuses on 'MonthlyCharges' and 'tenure' as predictive features. The model, trained on 80% of the data, aims to provide insights into how charges and customer tenure correlate with churn probability. The output also

includes a confusion matrix, classification report, and an accuracy score of 0.77359, suggesting that these two features alone provide high predictive capability.

3. The Decision Tree model uses 'tenure' as its predictive feature, and is trained on 80% of the dataset. This model gives us an intuitive understanding of decision-making through its tree-like structure. Its accuracy score of 0.7466 indicates it is slightly less effective at predicting churn compared to the logistic regression models, likely due to overfitting or the simplistic nature of using a single feature.
4. The Random Forest model incorporates both 'tenure' and 'MonthlyCharges' to predict churn, leveraging 100 decision trees to improve prediction stability and reduce the likelihood of overfitting. It is trained on 80% of the data. Its performance is detailed by a confusion matrix and an accuracy score of 0.7260. Despite its sophisticated approach, it has a lower accuracy compared to simpler models.



The Multiple logistic regression and the logistic regression models showed better performance, and the reason for this might be that there exists a linear relationship between features such as ‘tenure’ and ‘MonthlyCharges’ and the target variable ‘Churn’.

6 Results and Insights:

1. **Senior citizens:** Have a churn rate of nearly 50%. A detailed analysis by the company to identify what causes this would help prevent churn.
2. **Type of internet:** Although Fiber optic seems to be preferred, almost 50% of its users end up churning. This might be due to the demand and nature of fiber optic services. Being a modern, high speed internet service, various companies may be providing similar services for competitive rates.
3. **Technical support:** About 50% without competent technical support services experience churn. This is something that could be addressed by having a solid technical support team in place and providing the users with an easy way to contact the team.
4. **Contracts:** Most customers go for monthly contracts, but nearly 50% leave the company. Yearlong or two yearlong contracts have a much higher chance of retaining the customers. This could be indicative of monthly contracts being less attractive compared to monthly contracts offered by other companies.
5. **Payment method:** Customers opting for electronic checks as their payment method are the most likely to churn, about 4 times more likely than mailed checks or credit/debit card payers. Automated credit/debit cards seem to be the best payment method for retaining customers. However, since most customers use electronic checks (over a third), sudden changes in the payment method must be exercised with caution.
6. **Tenure:** Customers tend to churn when they just join the service, which indicates better support would help in getting them past this initial phase where they might decide to leave the telecom company.
7. **Costs:** Customers with higher charges, over $100 a month are very likely to churn. Tailored assistance would be beneficial in handling these users and offering them competitive rates and offers better than other companies might help retain these high paying individuals.

These insights and trends would help guide strategic efforts towards improving customer engagement and retention.

7 Conclusion:

In conclusion, this report shows the importance of effectively managing customer churn in the telecommunications industry. By analyzing the Telco Customer Churn dataset, we have identified key trends that shed light on the factors influencing churn behavior. Our analysis shows valuable consumer behavior trends, such as high churn rates among senior citizens and impact of people favoring service features like Fiber optic and essential support services. This could help telecom companies come up with strategies for targeted retention.

The use of data visualization techniques and machine learning models such as logistic regression, decision trees, and random forests for churn prediction would let telecom companies easily engage at-risk customers and implement effective retention initiatives. By tailoring services to meet customer needs, companies can work on their brand loyalty and enjoy sustainable business growth in today's cutthroat telecommunications market.

References:

1. https://www.kaggle.com/code/idalez/telco-customer-churn-analysis-and-prediction
2. Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. Expert Systems with Applications, 31, 515–524.
3. Ahmed, A. A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. In 2017 international conference on advanced computing and communication systems (ICACCS-2017), Jan. 06–07, 2017, Coimbatore, India
4. Naz, N. A., Shoaib, U., & Shahzad Sarfraz, M. (2018). A review on customer churn prediction data mining modeling techniques. Indian Journal of Science and Technology, 11(27), 1–27
5. Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. International Journal of Advanced Computer Science and Applications, 9(2), 273–281.
6. Kau, F. M., Masethe, H. D.. Lepota, C. K., & IAENG Member. (2017). Service provider churn prediction for telecoms company using data analytics. In Proceedings of the world congress on engineering and computer science 2017 WCECS 2017 (Vol. I), October 25–27, 2017, San Francisco, USA.
7. Gupta, Sunil, and Valarie Zeithaml. "Customer metrics and their impact on financial performance." Marketing Science 25, no. 6 (2006): 718-739.
8. Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39, no. 1 (2012): 1414-1425
9. Shaaban, Essam, Yehia Helmy, Ayman Khedr, and Mona Nasr. "A proposed churn prediction model." IJERA 2 (2012): 693-697.
10. Yihui, Qiu, and Mi Hong. "Application of Feature Extraction method in customer churn prediction based on Random Forest and Transduction." Journal of Convergence Information Technology 5, no. 3 (2010): 73-78.
11. Kirui, Clement, Li Hong, Wilson Cheruiyot, and Hillary Kirui. "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining." IJCSI International Journal of Computer Science Issues 10, no. 2 (2013): 1694-0784.
12. Kristof Coussement, Dries F. Benoit, Dirk Van denPoel, "Improved marketing decision making in a customer churn prediction context using generalized additive models", Expert Systems with Applications, Volume 37, Issue 3, 2010, Pages 2132–214.
13. Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", Expert Systems with Applications, 37, 2010,pp. 4710–4712.