Phase – 5

Data analytics with Cognos

**Project Objective:**

The primary objective of this project is to address the critical issue of customer churn within the telecom industry. Churn, or the rate at which customers terminate their services, poses a significant challenge for telecom companies. Understanding and predicting churn is vital for businesses in this sector as it allows them to proactively manage customer retention, reduce revenue loss, and enhance customer satisfaction. This project seeks to provide valuable insights and predictive models to assist a telecom company in tackling this problem effectively.

The project begins with a comprehensive exploration of the dataset, delving into customer demographics, account information, and service usage. Through data exploration, the aim is to uncover underlying patterns and correlations that may be linked to customer churn. By gaining a deeper understanding of the data, we can identify key drivers of churn and potential areas for intervention.

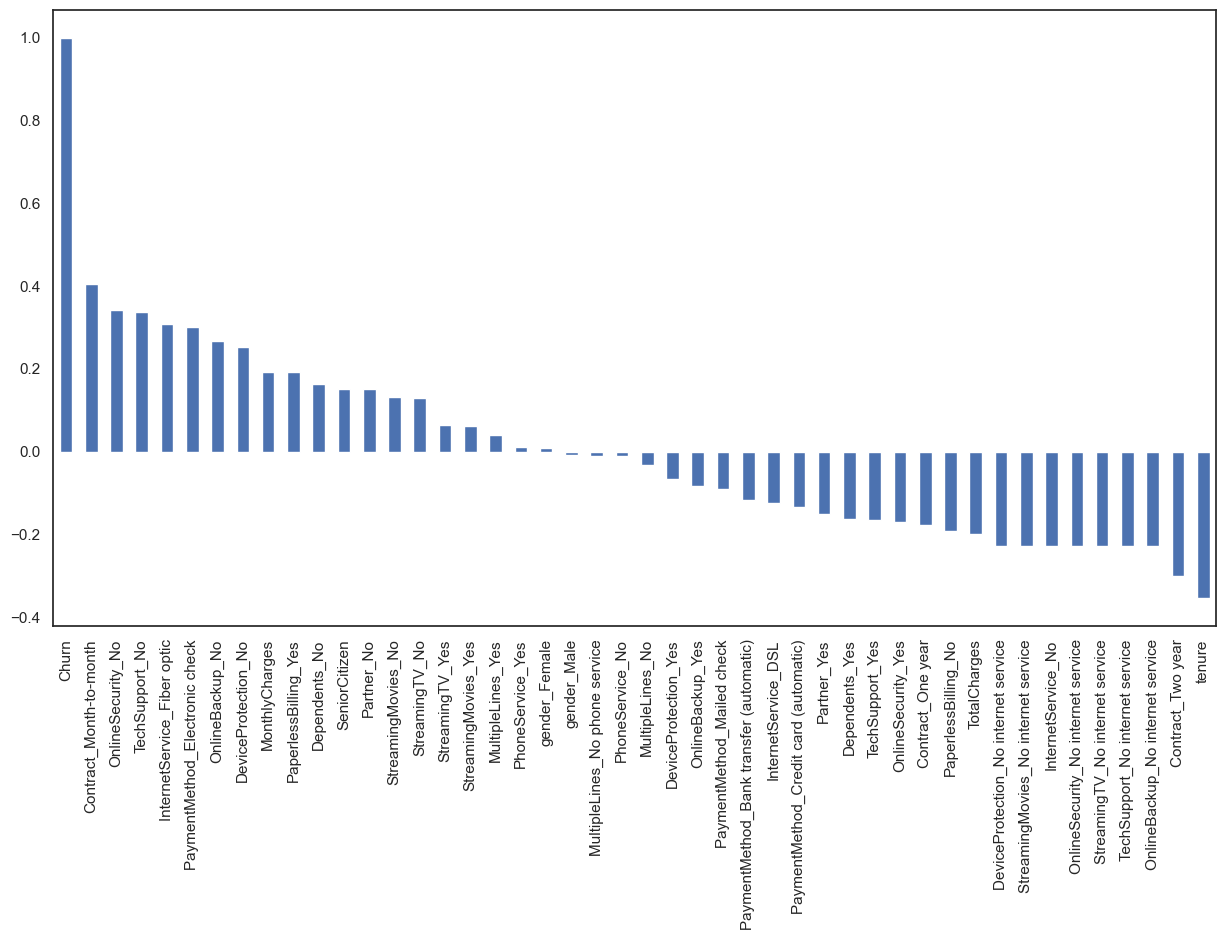
The core of the project lies in building and evaluating predictive models. Four different models, namely Logistic Regression, Random Forest, Support Vector Machine (SVM), and AdaBoost, are applied to forecast customer churn. Each model is assessed for its accuracy in predicting whether a customer will churn or stay with the telecom service. This comparative analysis helps determine which model performs best and provides the most reliable predictions.

Furthermore, the project aims to uncover valuable insights into the factors that influence customer churn. By understanding these factors, the telecom company can tailor its strategies to reduce churn rates. The project may offer recommendations and strategies for the company to retain more customers, potentially by targeting specific customer segments or improving services and offerings.

**Project Desing Thinking Ideas:**

In the context of the project aimed at analyzing and predicting customer churn in the telecom industry, the application of design thinking involves a strategic and empathetic approach to problem-solving. The initial step entails empathizing with the telecom company's perspective, comprehending the challenges associated with customer churn, and gaining insights into the industry's dynamics. Subsequently, the process moves to defining the core problem statement, which is focused on finding effective solutions to reduce churn. Ideation comes into play as creative brainstorming takes center stage, leading to the generation of diverse strategies to address churn. Prototyping involves the creation of preliminary predictive models for customer churn, while testing evaluates the models' accuracy and effectiveness. The iterative nature of design thinking ensures that feedback from testing informs ongoing adjustments, potentially sending the project back to previous stages for refinement. Although implementation of the predictive models is an optional phase, effective communication and sharing of insights throughout the project are essential to facilitate informed decision-making and drive innovative solutions to mitigate customer churn effectively. This iterative and user-centered approach aligns with the overarching goal of enhancing customer retention and optimizing business performance in the telecom industry.

**Analysis, Insights and Prediction results gained from the project :**



**Month to month contracts, absence of online security and tech support seem to be positively correlated with churn. While, tenure, two year contracts seem to be negatively correlated with churn.**

**Interestingly, services such as Online security, streaming TV, online backup, tech support, etc. without internet connection seem to be negatively related to churn.**

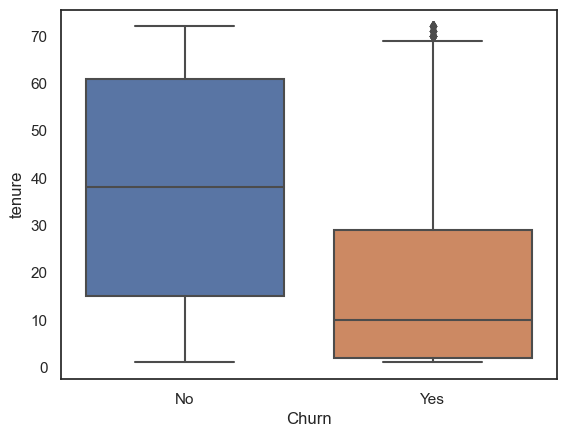
**Let's take a look at out predictor variable (Churn) and understand its interaction with other important variables as was found out in the correlation plot.**

**Lets first look at the churn rate in our data**

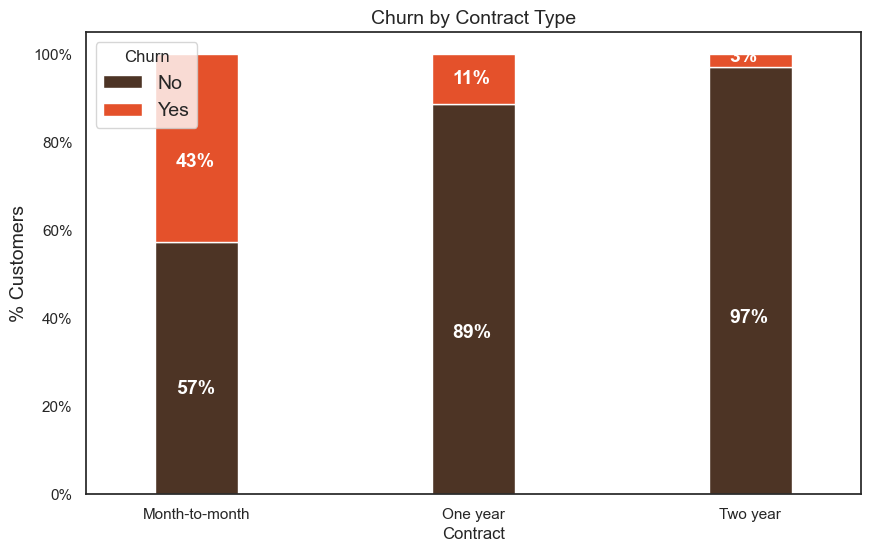
**In our data, 74% of the customers do not churn. Clearly the data is skewed as we would expect a large majority of the customers to not churn. This is important to keep in mind for our modelling as skeweness could lead to a lot of false negatives. We will see in the modelling section on how to avoid skewness in the data.**

**Lets now explore the churn rate by tenure, seniority, contract type, monthly charges and total charges to see how it varies by these variables.**

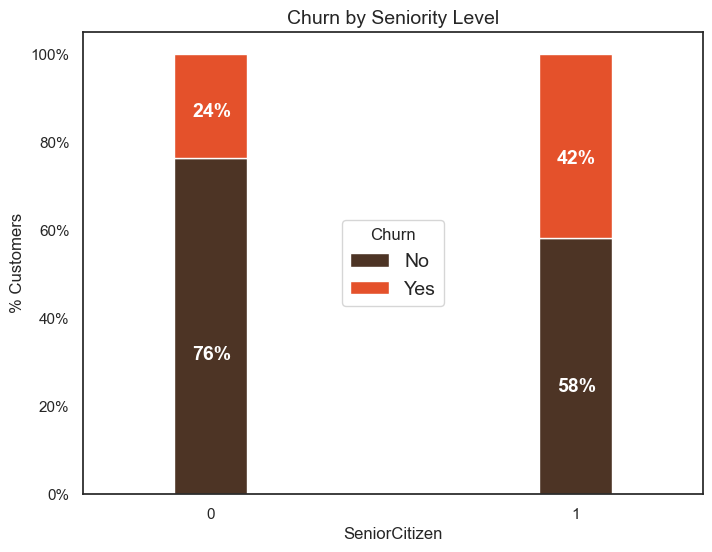
**i.) Churn vs Tenure: As we can see form the below plot, the customers who do not churn, they tend to stay for a longer tenure with the telecom company.**



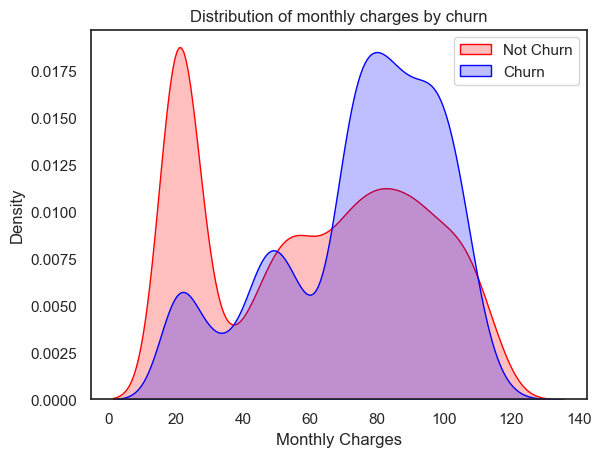
**ii.) Churn by Contract Type: Similar to what we saw in the correlation plot, the customers who have a month to month contract have a very high churn rate.**



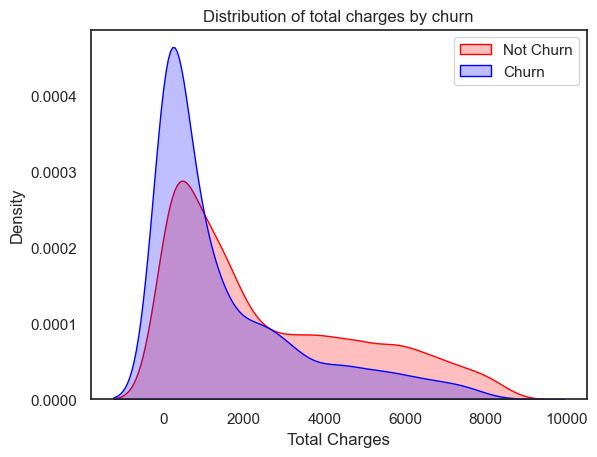
**iii.) Churn by Seniority: Senior Citizens have almost double the churn rate than younger population.**



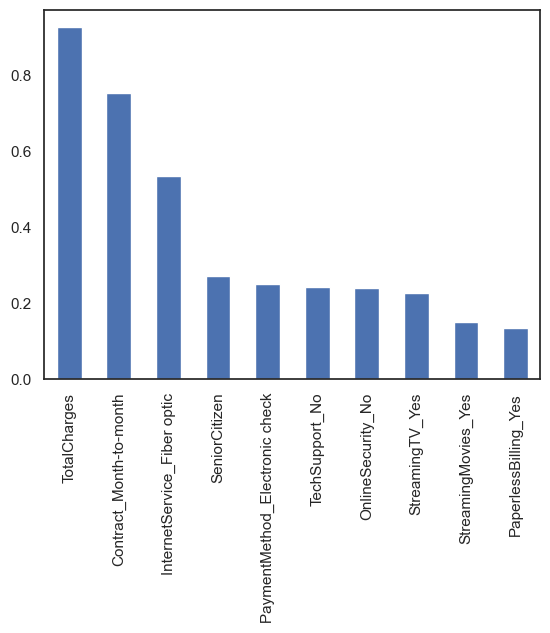
**iv.) Churn by Monthly Charges: Higher % of customers churn when the monthly charges are high.**

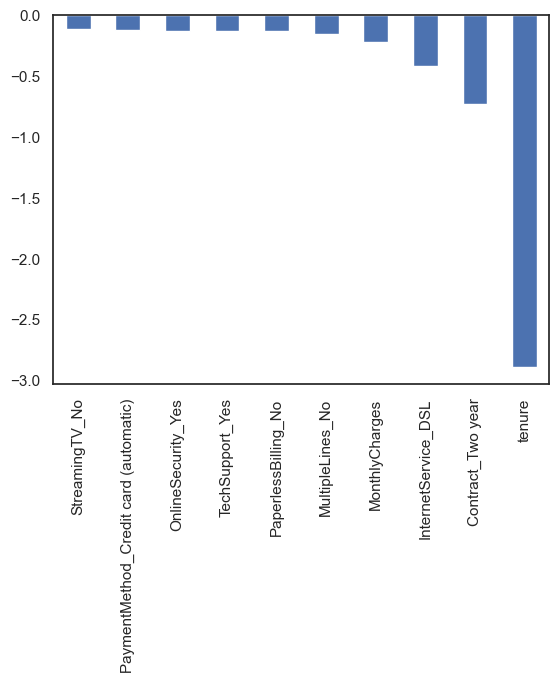


**v.) Churn by Total Charges: It seems that there is higer churn when the total charges are lower.**



**Weight of all Variables used to build the logistic Regression model**





Observations

We can see that some variables have a negative relation to our predicted variable (Churn), while some have positive relation. Negative relation means that likeliness of churn decreases with that variable. Let us summarize some of the interesting features below:

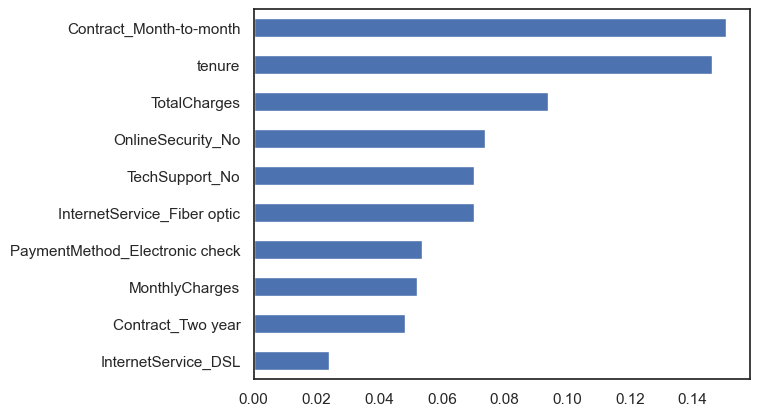
As we saw in our EDA, having a 2 month contract reduces chances of churn. 2 month contract along with tenure have the most negative relation with Churn as predicted by logistic regressions

Having DSL internet service also reduces the proability of Churn

Lastly, total charges, monthly contracts, fibre optic internet services and seniority can lead to higher churn rates. This is interesting because although fibre optic services are faster, customers are likely to churn because of it. I think we need to explore more to better understad why this is happening.

Any hypothesis on the above would be really helpful!

Variable Weights for Random Forest :



Observations:

From random forest algorithm, monthly contract, tenure and total charges are the most important predictor variables to predict churn.

The results from random forest are very similar to that of the logistic regression and in line to what we had expected from our EDA

**Final Model Accuracies :**

