**Customer Loyalty Analysis**

A PROJECT REPORT

*Submitted by*

**MARKANTI RATAN REDDY BL.EN.U4CSE19084**

**M SRI CHARAN BL.EN.U4CSE19078**

**MALAPATI GIRISH REDDY BL.EN.U4CSE19080**

***for the course***

***19CSE304 – Foundations of Data Science***



**AMRITA SCHOOL OF ENGINEERING, BANGALORE**

**AMRITA VISHWA VIDHYAPEETHAM**

**BANGALORE-560 035** November-2021

**Customer Loyalty Analysis**

**Introduction:**

Customer Loyality is a metric to study the number of customers of a company that stopped using the company’s services and products. Every company tries to expand by bringing in new clients and extended its reach to newer variety of customers but it is harder to get new customers to use company’s services than to retain existing customers by satisfying their needs and meeting their expectations. Similarly, it is also a waste of energy and utility to invest on a customer who is mostly likely to be continuing with the company’s services. Thus, customer churn is a vital information to the company as it has a direct impact on the company’s success and growth.

The dataset we are working with is IBM’s customer churn data. The dataset was taken from the website:

<https://community.ibm.com/accelerators/catalog/content/Telco-customer-churn>

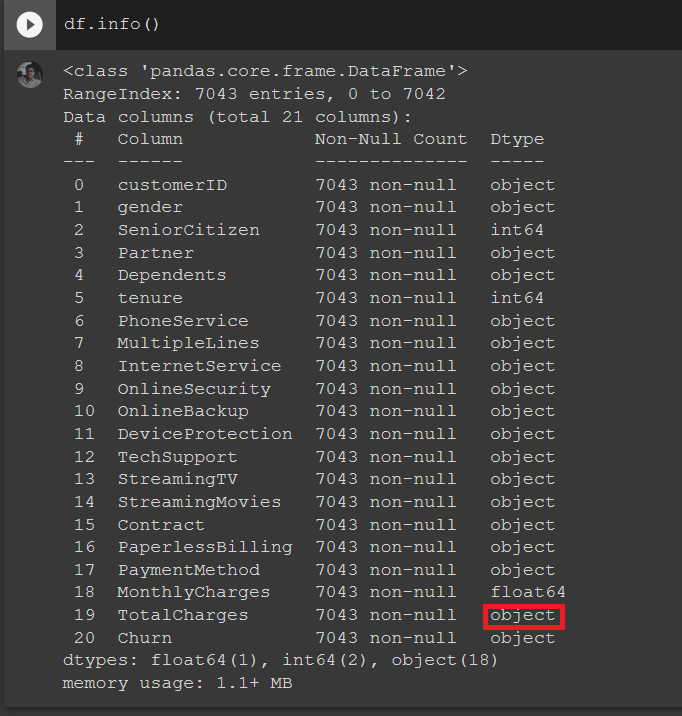
IBM has provided information about 7,043 of its churned and current users with 21 different features mentioned below.

* customerID: ID of the customer
* gender: Gender of the customer (female, male)
* SeniorCitizen: Is the customer a senior citizen? (1, 0)
* Partner: Does the customer have a partner? (Yes, No)
* Dependents: Does the customer have dependents? (Yes, No)
* tenure: Tells us the number of months the customer has stayed with the company
* PhoneService: Does the customer have a phone service? (Yes, No)
* MultipleLines: Does the customer have multiple lines? (Yes, No, No phone service)
* InternetService: How the customer’s internet service is being provided? (DSL, Fiber optic, No)
* OnlineSecurity: Does the customer have online security? (Yes, No, No internet service)
* OnlineBackup: Does the customer have online backup? (Yes, No, No internet service)
* DeviceProtection: Does the customer have device protection? (Yes, No, No internet service)
* TechSupport: Does the customer have tech support? (Yes, No, No internet service)
* StreamingTV: Did the customer opt for streaming TV service? (Yes, No, No internet service)
* StreamingMovies: Did the customer opt for streaming movies? (Yes, No, No internet service)
* Contract: The length of contract term of the customer (Month-to-month, One year, Two year)
* PaperlessBilling: Does the customer have paperless billing? (Yes, No)
* PaymentMethod: What kind of payment method is/was being used (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
* MonthlyCharges: Amount charged by the company on monthly basis (numeric)
* TotalCharges: Total amount charged by the company (object)
* Churn: Did the customer churn? (Yes or No)

We have chosen this dataset to help IBM by providing insights on its customers and an algorithm which gives the likelihood of an active customer churning.

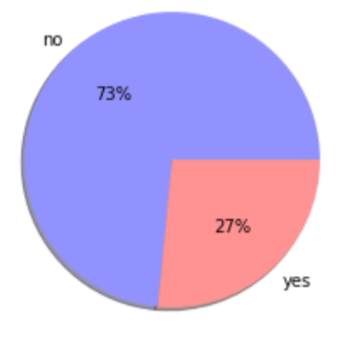
**Exploratory Data Analysis:**

Upon using .info() function on our data frame we go the following result.



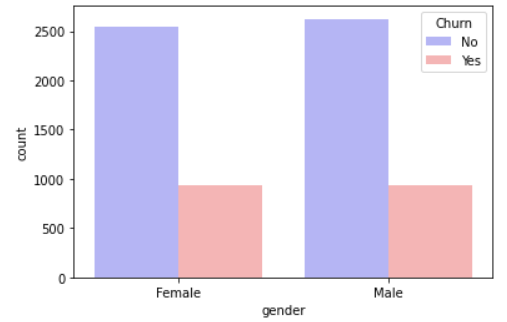
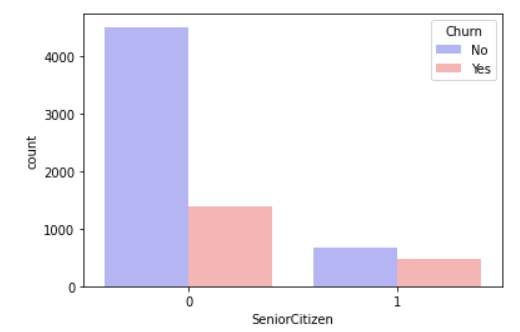
The column ‘TotalCharges’ which is supposed to be numerical is of type object. So, our first step is to convert the data present in that column to numbers.

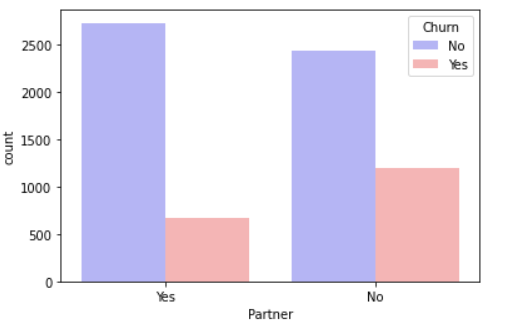
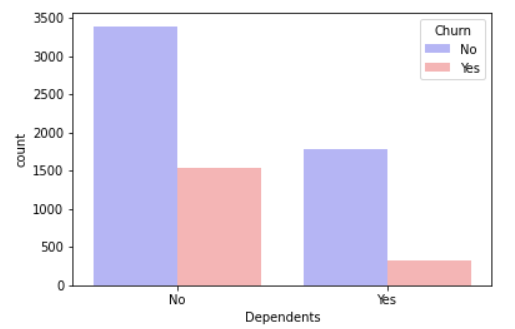
**Customer Churn Rate**

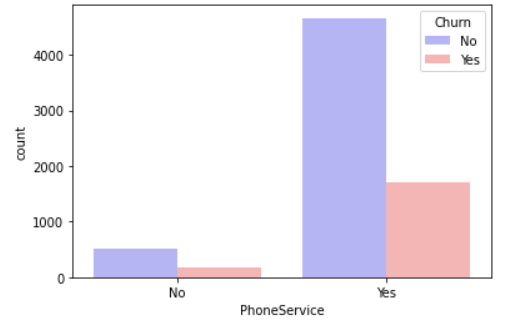
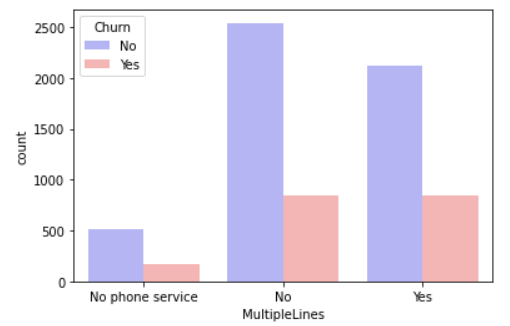


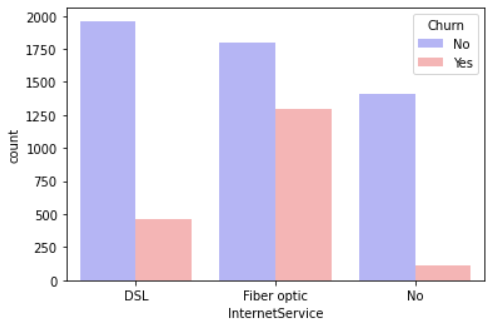
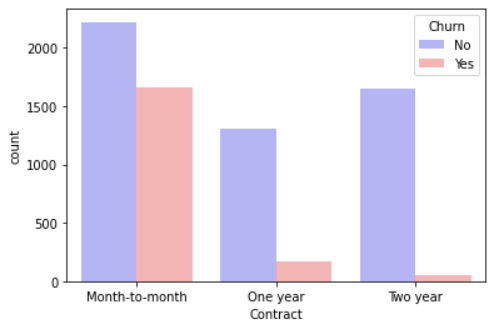
From the above pie chart, we see that the percentage of customers that have churned is 1/3rd of the customers who have not churned. The dataset is imbalanced and requires sampling during model building phase.

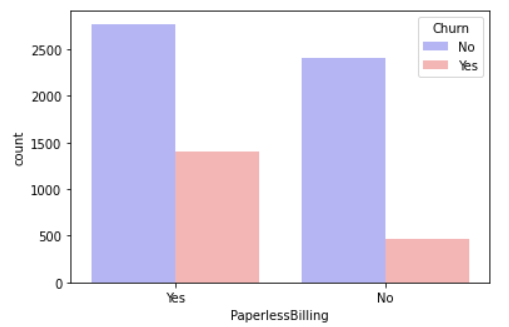
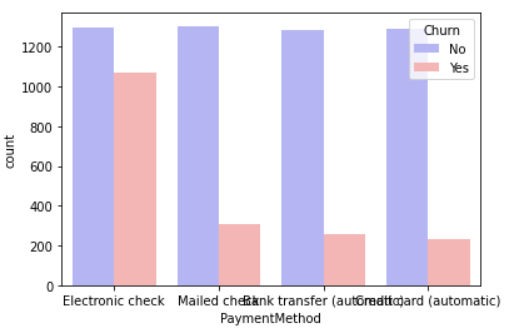
**Data analysis**

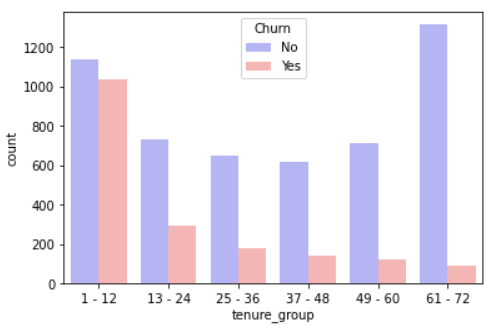
 

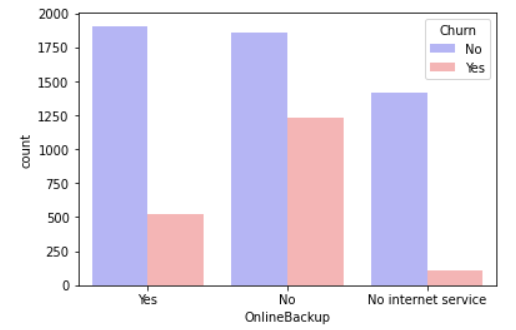
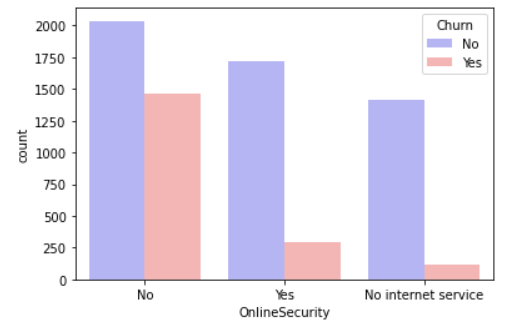
 

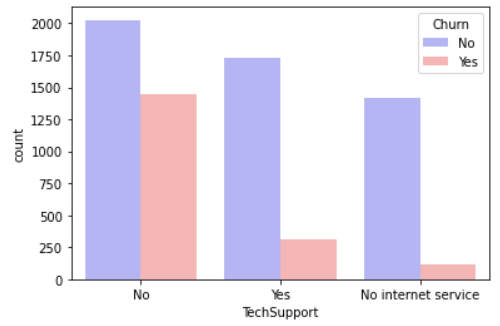
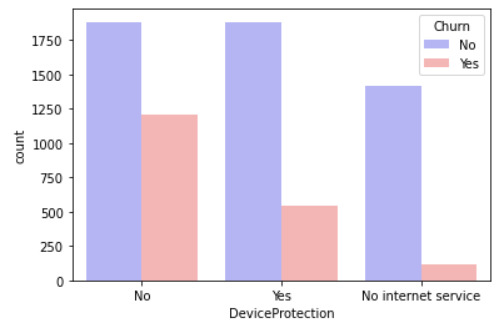
 

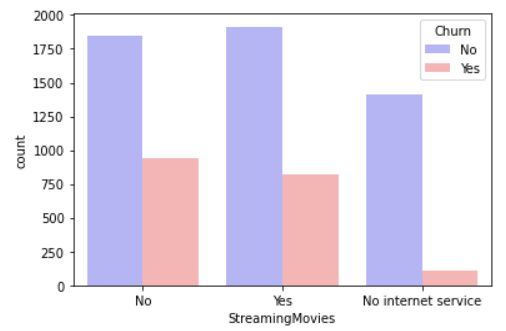
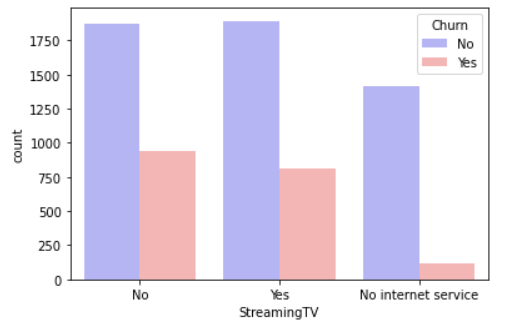


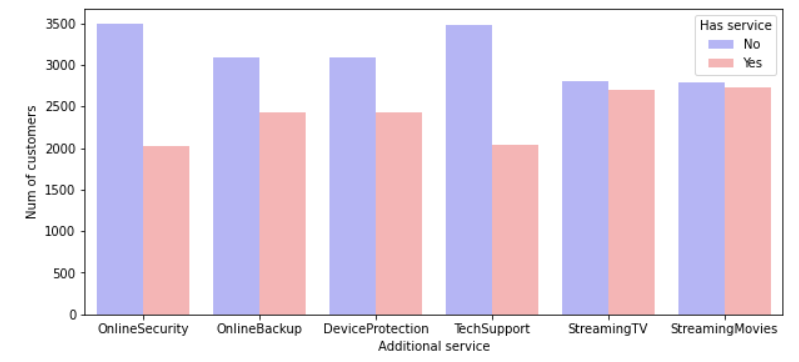
1. There are very few senior citizens who are using the company’s services and the likelihood of a senior citizen churning is much higher than that of juniors.
2. Customers who don’t have any partners, dependents are more likely to churn.
3. Customers who have opted for paperless billing show a higher percentage of churn that the customers who have not opted for the same.
4. Customers who use fibre optics for internet service, electronic check as payment method and a month-to-month contract are more likely to churn.
5. Customers who are in the tenure group of 1-12 are most likely to churn. This shows that customers become more loyal to the company as time progresses.

**Effect of additional features on customer churn**



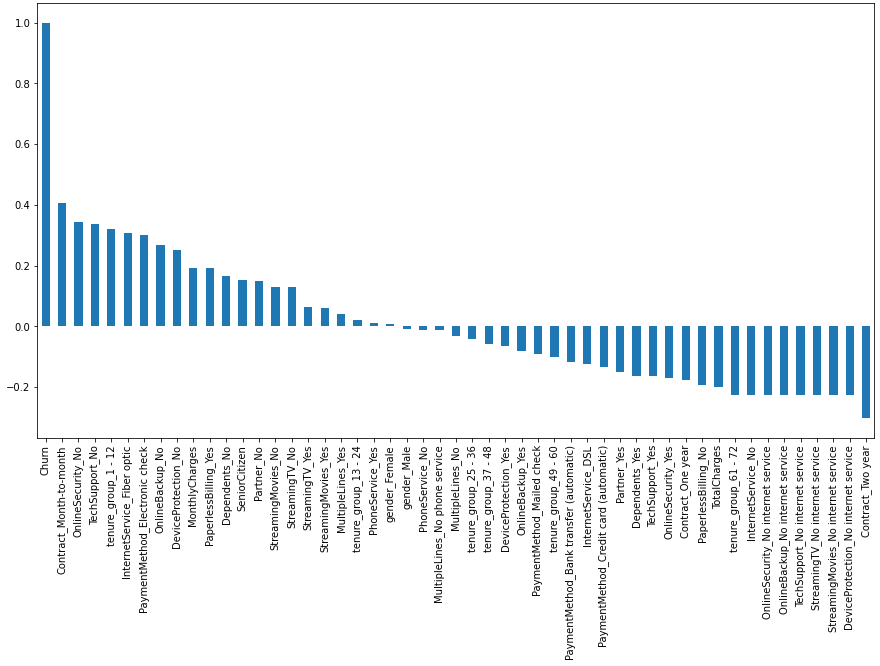






1. Having features such as online security, online backup, device protection, tech support shows lower churning rate implying that customers prefer to enjoy additional features.
2. Many users are not preferring additional internet services, but users who have opted for these services show high retention rate from the above graphs, hence the company has to focus on advertising these services to its customers to decrease the customer churn rate

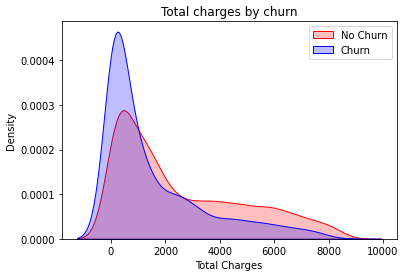
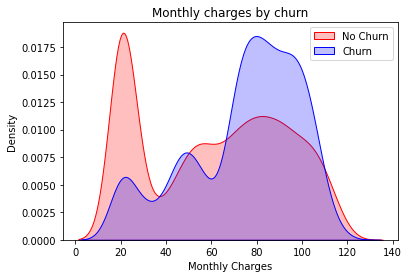
**Impact of each feature on the customer churn**

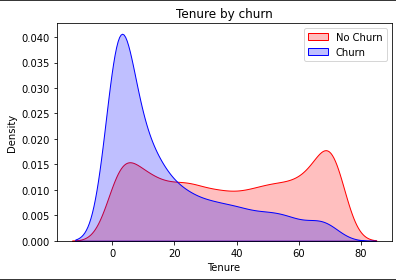


Month to month contracts, no online security and no tech support seem to be positively correlated with churn. While, tenure, two-year contracts seem to be negatively correlated with churn.

Also, services such as online security, online backup, device protection, tech support without internet connection are negatively related to churn.

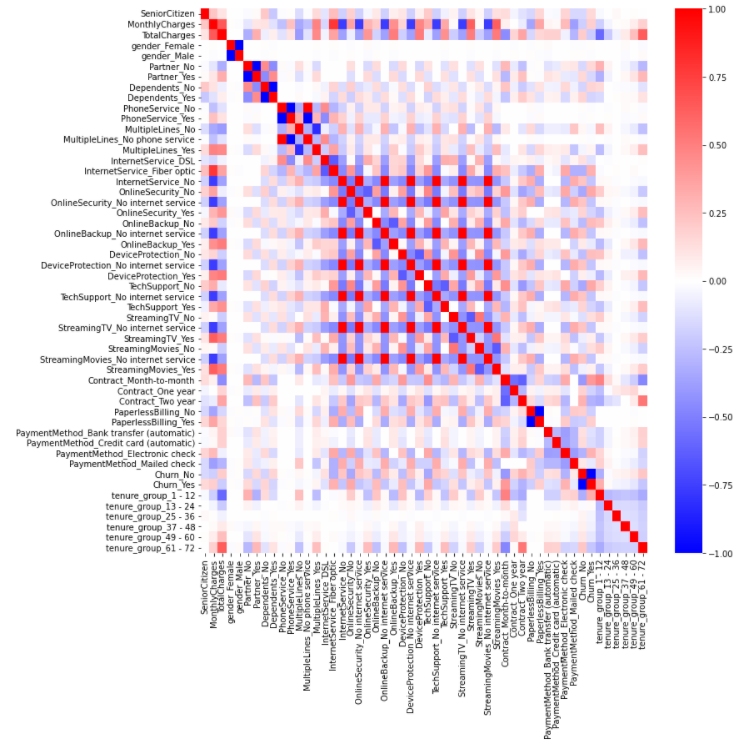
**Effect of Charges and Tenure on customer churn**



1. Customers with high monthly charges are more likely to churn.
2. Surprisingly, there is a high amount of churn at lower ‘TotalCharges’.
3. New customers are more likely to churn.
4. ‘tenure’ and ‘Monthly Charges’ are potentially important features for churn analysis.

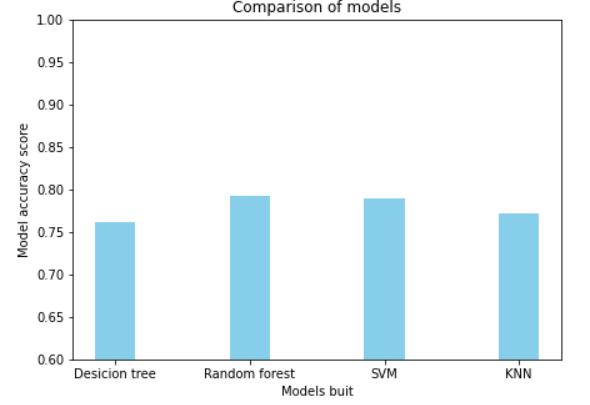
**Feature correlation Heat map**



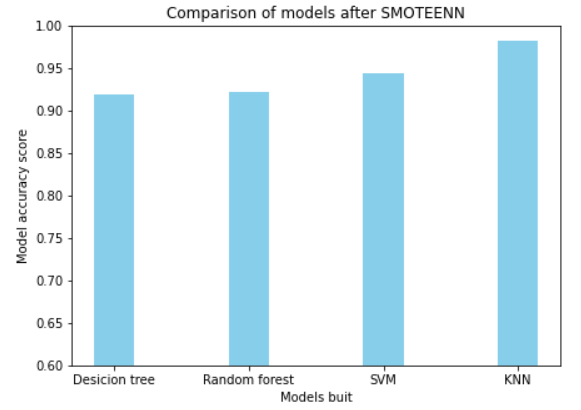
**Model Building :**

Models used:

* Decision tree classifier
* Random forest classifier
* Support vector classifier
* K Nearest neighbour classifier



From the above barograph we can observe that all the models are not performing well, that is because we have imbalanced data with respect to churners and non-churners. To overcome this problem we used sklearn.SMOTEENN which uses SMOTE oversampling technique



By the above bar graph, we can conclude that for our dataset KNN outperforms all the other models.

**Suggestions to the company:**

1. As the customers who don’t have any partners, dependents are more likely to churn, if company can come up with family packs.
2. Try to convince the customers to opt for high tenure plans.
3. Additional internet services like online security, online backup, device protection, tech support are not preferred by many users but user who have opted for these services show high retention rate from the above graphs hence the company has to

focus on advertising these services to its customers to decrease the customer churnrate.

**Links Referred ;**

<https://community.ibm.com/accelerators/catalog/content/Telco-customer-churn>

<https://www.analyticsvidhya.com/blog/2021/08/understanding-bar-plots-in-python-beginners-guide-to-data-visualization/>

<https://imbalanced-learn.org/stable/references/generated/imblearn.combine.SMOTEENN.html>

<https://scikit-learn.org/stable/modules/svm.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>