CHURN ANALYSIS

- **REPORT BY**

- ABHISHEK PAL
- SANTOSH KUMAR,
- ANCHAL GUPTA,
- DHRUV GALA,
- SALVADER RON NATHANIEL,
- KEERTHANA S K

"If Your Retention is Poor then Nothing Else Matters

" ***-Brian Balfour,Founder/CEO of Reforge.***

PROBLEM STATEMENT:

Identifying the trend in the customer churning report of a telecom company and predicting which type or segment of customer is likely to churn. Customer churn or customer attrition basically means that, when for a particular company a customer stops being a customer for that particular company. Customers might not have to be the paying-ones only. For instance, anyone who is foregoing Google or Facebook becomes a churned-out customer for that company respectively.

Since customer churning hurts the growth of a company, many companies are obsessed with reducing customer churning. Many solutions are being suggested to solve this problem. Churn Data Analytics is one among them and I strongly believe that Data Analytics can contribute a lot to this field. Data Analytics can not only solve problems in this field but it can also optimize the techniques that have been used for decades.

1. BUSINESS NEED ASSESSMENT:

- Many telecom industries know that their sales follow a certain trend and know what sort of plan (for their various services, e.g., TV Streaming, Movie Streaming) they need to stay competitive. But they do not know it accurately and most of the times they miss one thing or the other.
- Analysing their datasets and forecasting their sales will enable them to give proper plan types, the required competitive monthly charges for various services that they are giving.
- This analysis will cut down their costs and reduce the churning of customers.

2. TARGET SPECIFICATIONS AND CHARACTERIZATIONS:

The target here is to develop a model that will forecast future sales, cut down costs and reduce customer churning.

The trend recognition has to be done by a data scientist who has some knowledge about the telecom industry. Employing someone who has no knowledge about the telecom industry is not a great idea as they will not be able to give the insights of someone who knows about this field.

The model should be able to handle large volumes of data, as in a telecom industry there will be a lot of features for the model to look at and the size of data depends on the sales of a particular month or week.

We should also know in advance whether the customers need our model to forecast the sales for a week or for a month.

MAIN REASONS FOR CUSTOMER CHURN IN TELECOM
service quality
features availability
competitive advantage
negative customer service

3. EXTERNAL SEARCH (INFORMATION SOURCES/REFERENCES):

REFERENCES:

1. https://mixpanel.com/blog/churnanalytics/#:~:text=Churn%20analytics%20is%20the%20process,larger%20margins%2C%20and%20higher%20profits.

2. https://www.gadgetsnow.com/telecom/which-telecom-company-gained-more-new-mobile-subscribers-in-may-as-per-trai-data/articleshow/92989638.cms

3. https://timesofindia.indiatimes.com/business/india-business/telecom-subscribers-in-india-increased-by-2-9-million-in-may-trai/articleshow/92980021.cms

4. https://www.business-standard.com/article/companies/mobile-user-base-dips-22-mn-in-march-suffers-biggest-fall-since-april-2018-119052101593_1.html

5. https://dot.gov.in/#loaded

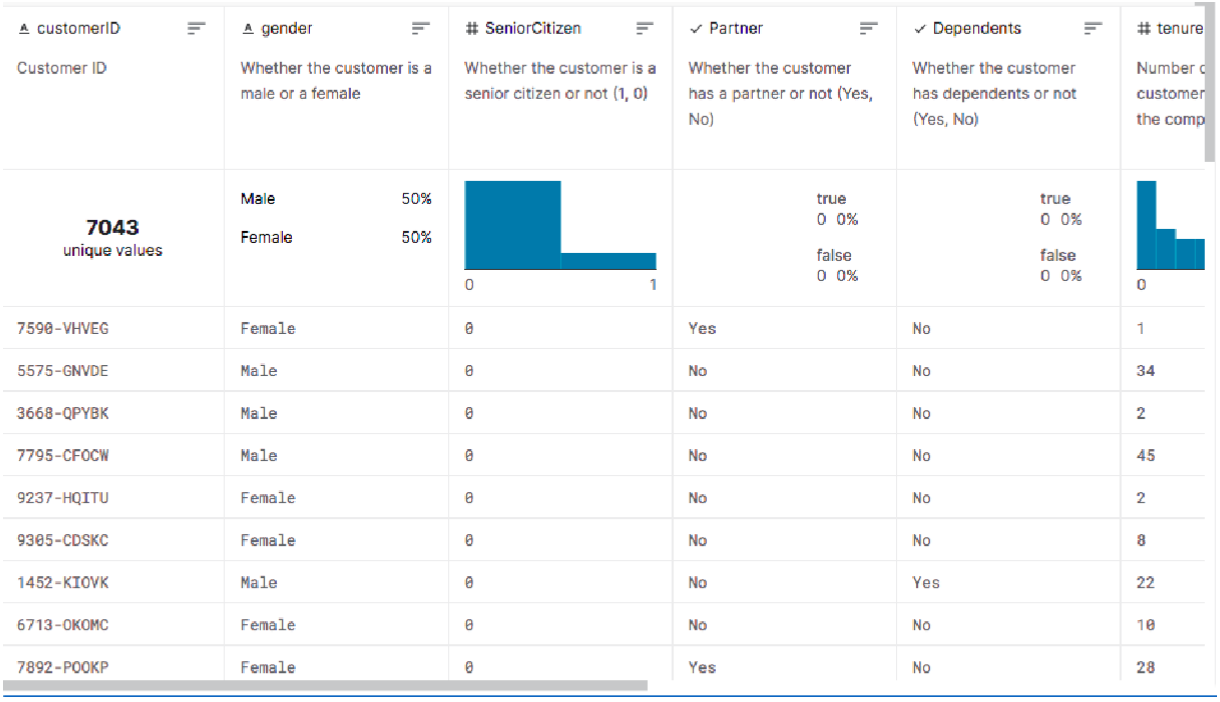
6. https://www.indmoney.com/articles/stocks/telecom-subscribers-in-india-

june- 2022

DATASETS:

1. Customer Churn Prediction|EDA|ANN

https://www.kaggle.com/code/anubhavgoyal10/customer-churn-prediction-eda-ann/data



customerID	gender	# SeniorCitizen	✓ Partner	✓ Dependents	# tenure
Customer ID	Whether the customer is a male or a female	Whether the customer is a senior citizen or not (1, 0)	Whether the customer has a partner or not (Yes, No)	Whether the customer has dependents or not (Yes, No)	Number c customer the comp
7043 unique values	Male 50% Female 50%	0 1	true 0 0% false 0 0%	true 0 0% false 0 0%	0
7590-VHVEG	Female	0	Yes	No	1
5575-GNVDE	Male	0	No	No	34
3668-QPYBK	Male	0	No	No	2
7795-CFOCW	Male	0	No	No	45
9237-HQITU	Female	0	No	No	2
9305-CDSKC	Female	0	No	No	8
1452-KIOVK	Male	0	No	Yes	22
6713-OKOMC	Female	0	No	No	10
7892-POOKP	Female	0	Yes	No	28

This dataset will show that we will be able to predict what type of customers who are likely to churn out based on the parameters stated in the columns of the above data set.

5. **Benchmarking:**

Below mention are the companies that perform churn analysis for various businesses: -

i. Churnly – Churnly's AI is powerful and impressive enough to help you deal with customer churn at every stage. Churnly provides end-to-end solutions designed for Windows. This online Customer Success system offers Account Alerts, Customer Engagement, Health Score, on boarding, Usage Tracking / Analytics at one place.

ii. Qymatix - Qymatix is a boon for B2B organizations. By using high-grade machine learning, this churn prediction software can track the customer journey, do instant analysis, and create behaviour-based modelling.

iii. Trifacta - Trifacta gathers data stored in different locations and do predictive analysis to reduce customer churn. This churn prediction software enables you to work directly with the data available and remove IT dependencies. Its churn predictions are data-driven and realistic.

iv. Data Science Studio (DSS) - Data Science Studio (DSS) is what you should get now if you want to do real-time churn prediction from available data. With its great data exploration ability, this churn prediction software ensures that not a single data goes waste

6. Applicable Patent:

i. **Churn prediction and management system -** A system and method for managing churn among the customers of a business is provided. The system and method provide for an analysis of the causes of customer churn and identifies customers who are most likely to churn in the future. Identifying likely churners allows appropriate steps to be taken to prevent customers who are likely to churn from actually churning. The system included a dedicated data mart, a population architecture, a data manipulation module, a data mining tool and an end user access module for accessing results and preparing preconfigured reports.

ii. **Managing customer loss using customer value** - Techniques are provided to determine, based on information about customers, the most valuable customers that have a high likelihood of being lost. The value of a customer may be based on the contribution of the customer to profit generated by a business enterprise.

iii. **Method for predicting churners in a telecommunications network** - Data pertaining to interactions between a plurality of customers is obtained. A graph is formed, having a plurality of nodes representing the customers and a plurality of edges representing interactions between the customers. A sub-set of the customers are denoted as previously churned customers.

7. Applicable Constraints:

The constraint that needs to be applied while performing a churn analysis is: -

i. Understanding problem and final goal
ii. Prototype Selection
iii. Data Collection
iv. Data preparation and processing
iv. Modelling and testing
v. Model Deployment and monitoring

8. Business Opportunities:

India is expected to have a digital economy of $1 trillion by 2025. Over the last seven years, the Indian Telecom Tower industry has grown significantly by 65%. The number of mobile towers increased from 400,000 in 2014 to 660,000 in 2021. It is also estimated that 5G technology will contribute approximately $450 bn to the Indian Economy in the period of 2023-2040.
Department of Telecom, DoT India, has put forth the new telecom policy, known as the National Digital Communications Policy 2018 or NDCP 2018.Through this Policy, it foresees investment worth US$ 100 billion in the telecommunications sector by 2022,

- By increasing the India’s contribution to global value chains.
- Development of Standard Essential Patents (SEPs) in the field of digital communication technologies.
- Train/ Re-skill 1 million manpower for building New Age Skills.
- Accelerate the transition to Industry 4.0

5G is the next technology frontier in the telecom sector

5G will be used in India to enable digital India, smart cities & smart Village missions for India. Telecom operators, Tower operators, policymakers, and device providers and software and hardware providers have to work together for the smooth functioning.

Tower sharing

Creating separate tower companies will help telecom companies working at lower operating costs and improve the capital structure. This has also provided an additional revenue stream.

The Indian mobile economy is growing rapidly and will contribute substantially to India's Gross Domestic Product (GDP) according to recent statistics. In 2019, India surpassed the US to become the second-largest market in terms of the number of app downloads. The Government has enabled easy market access to telecom equipment and a fair and proactive regulatory framework that has ensured the availability of telecom services to consumers at affordable prices. The deregulation of Foreign Direct Investment (FDI) norms has made the sector one of the fastest growing and the top five employment opportunity generator in the country.

The Government of India is working to digitally connect the rural and remote regions in the country and has decided a new affordable tariff structure with the principle of more you use, less you pay. The changes will soon be reflected in tariff changes by service providers in the country. The Indian mobile phone industry expects that the Government of India's boost to the production of battery chargers will result in setting up of 365 factories, thereby generating 800,000 jobs by 2025.

Growing demand in both e-commerce sector and needy of huge data among mass, policy support, increasing investment are the few among qualities that attract telecom sector in India. Nevertheless, telecom sector should identify the existing challenges like huge taxes, large invest cycles etc, should be done away with the existing policy support.

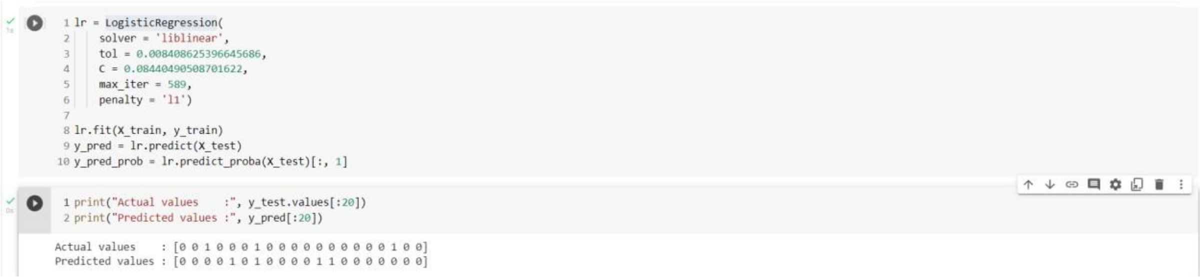
9. Concept Generation:

Every business whether small or large needs churn analysis software to know why customers are shifting from their product to other products. It helps us to known about the lack of usage of the product, poor service and better price at somewhere else. In this report we have built a prototype model for churn analysis that can be built in short period of time and is useful in long term in future, and as technology is increasing day by day and more startups are emerging churn analysis is applicable everywhere for analyzing customers and planning new strategies and step to be in the market.

For any business to be successful one should understand the customer and it needs and how the needs are changing day-to-day and how to ensure the customer trust on the company product. Company and stakeholders are investing more time and money finding out the reason and predict the type of customers that can switch to other brand and minimize it. The sudden downfall of user interaction with the company product also gives rise to perform churn analysis and finding the cause of it.

10.Concept Development:

In order to build this project, we have used a sample churn analysis dataset of telecommunication services and analysing how much customer is spending on for communication services and whether they will leave the company in future or not, for this we have downloaded sample dataset from Kaggle and a machine learning model is being created for which we have used random forest, Logistic Regression, Naïve Bayes and XGBoost algorithm, we can also use other machine learning algorithm as well. This model will give us the idea whether out of a certain customer how many will leave and how many will stay. Some predicted values will slight differ than actual values but the overall prediction is around 76% which in terms is consider a realistic and model performance is good.



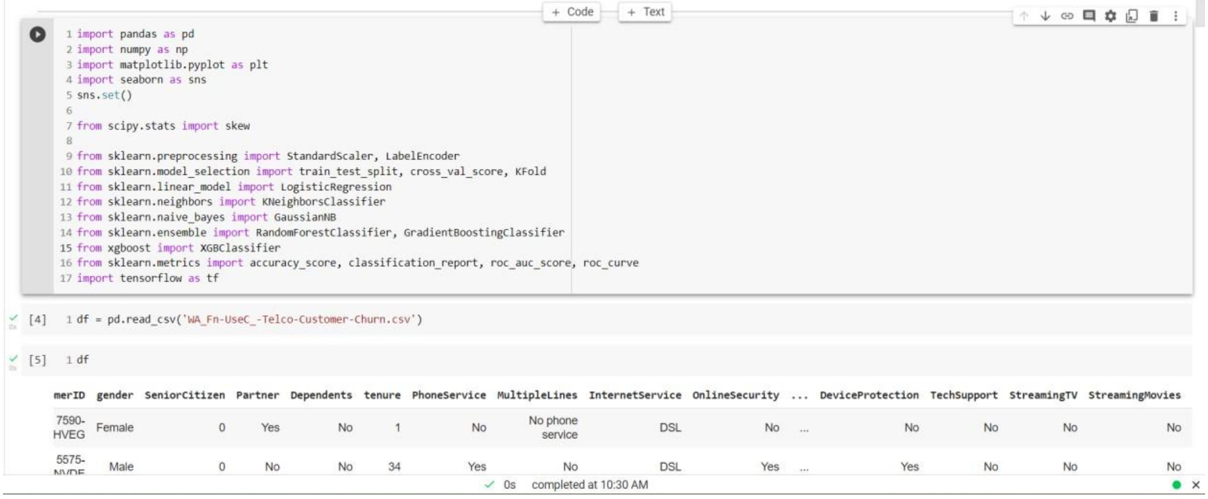
```
lr = LogisticRegression(
    solver = 'liblinear',
    tol = 0.008408625396645686,
    C = 0.08440490508701622,
    max_iter = 589,
    penalty = 'l1')

lr.fit(X_train, y_train)
y_pred = lr.predict(X_test)
y_pred_prob = lr.predict_proba(X_test)[:, 1]
```

```
print("Actual values    :", y_test.values[:20])
print("Predicted values :", y_pred[:20])
```

```
Actual values    : [0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0]
Predicted values : [0 0 0 0 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0]
```

Code Implementation

Importing required libraries and dataset



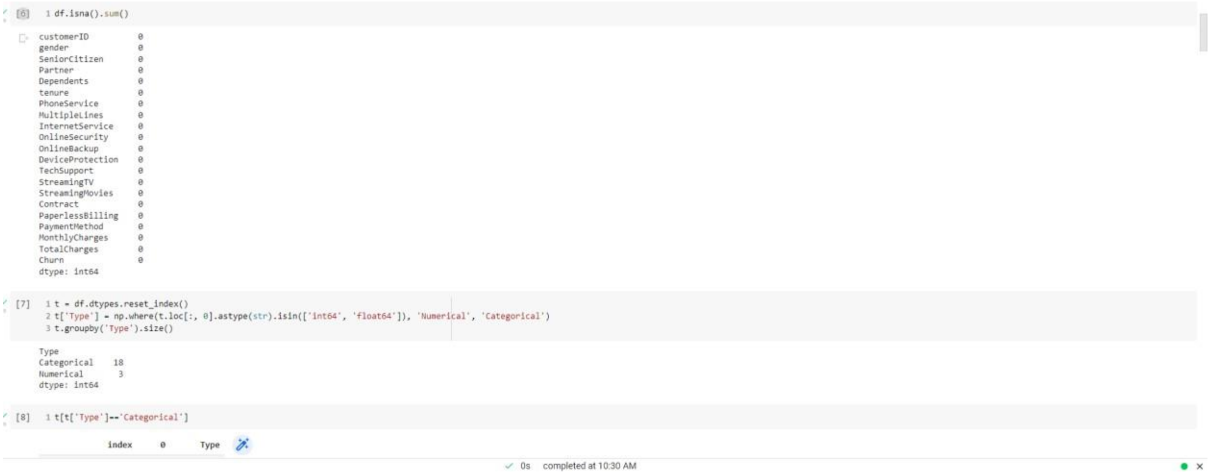
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()

from scipy.stats import skew

from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import train_test_split, cross_val_score, KFold
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report, roc_auc_score, roc_curve
import tensorflow as tf
```

```
[4] df = pd.read_csv('WA_Fn-UseC_-Telco-Customer-Churn.csv')
```

```
[5] df
```

merID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies
7590-HVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No
5575-NVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No	No

Data Pre-processing



```
df.isna().sum()
```

```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64
```

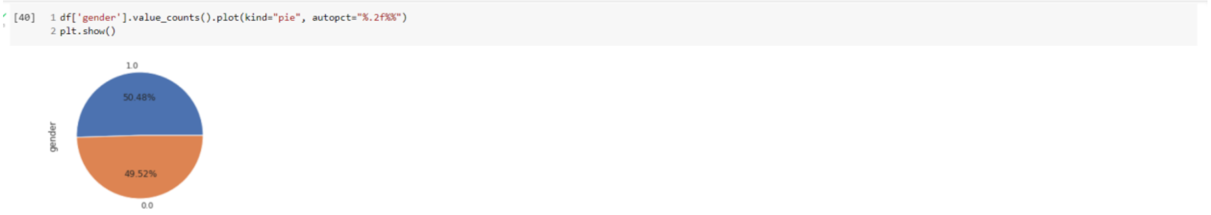
```
t = df.dtypes.reset_index()
t['Type'] = np.where(t.loc[:, 0].astype(str).isin(['int64', 'float64']), 'Numerical', 'Categorical')
t.groupby('Type').size()
```

```
Type
Categorical    18
Numerical       3
dtype: int64
```

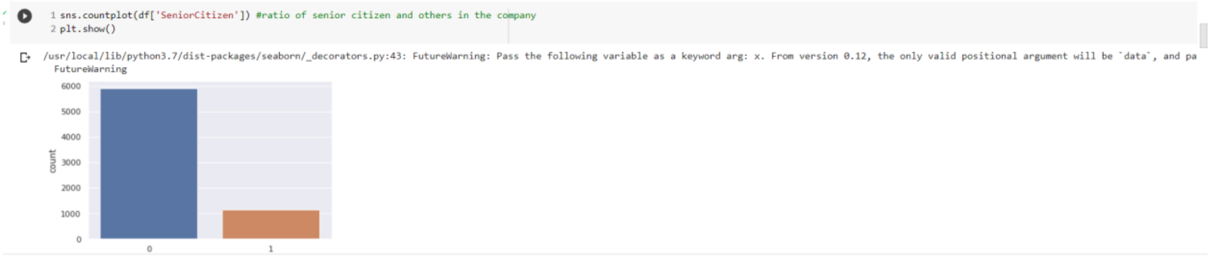
```
t[t['Type']=='Categorical']
```

```
        index    0    Type
```

EDA

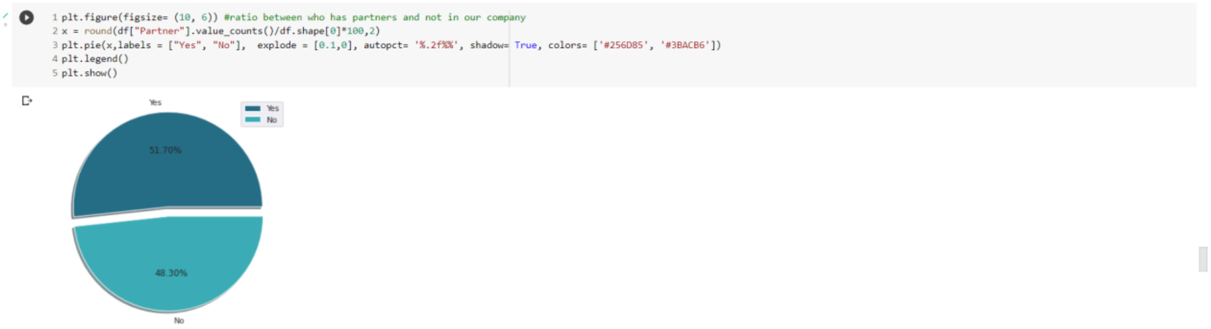


```
df['gender'].value_counts().plot(kind="pie", autopct="%.2f%%")
plt.show()
```

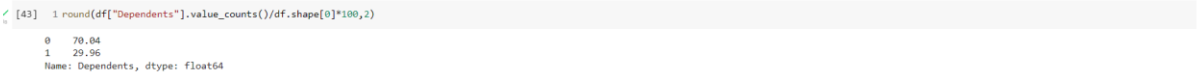


```
sns.countplot(df['SeniorCitizen']) #ratio of senior citizen and others in the company
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and pa
  FutureWarning
```

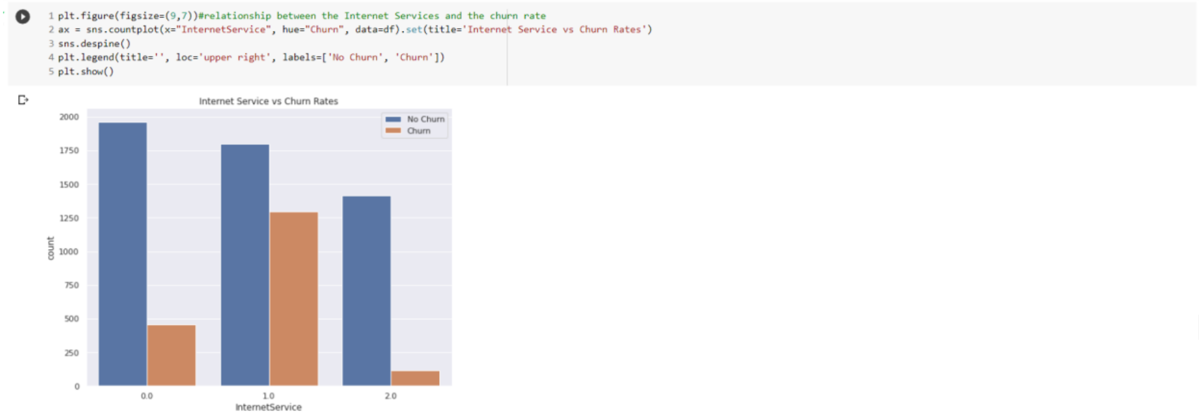


```
plt.figure(figsize= (10, 6)) #ratio between who has partners and not in our company
x = round(df["Partner"].value_counts()/df.shape[0]*100,2)
plt.pie(x,labels = ["Yes", "No"],  explode = [0.1,0], autopct= '%.2f%%', shadow= True, colors= ['#256D85', '#3BACB6'])
plt.legend()
plt.show()
```



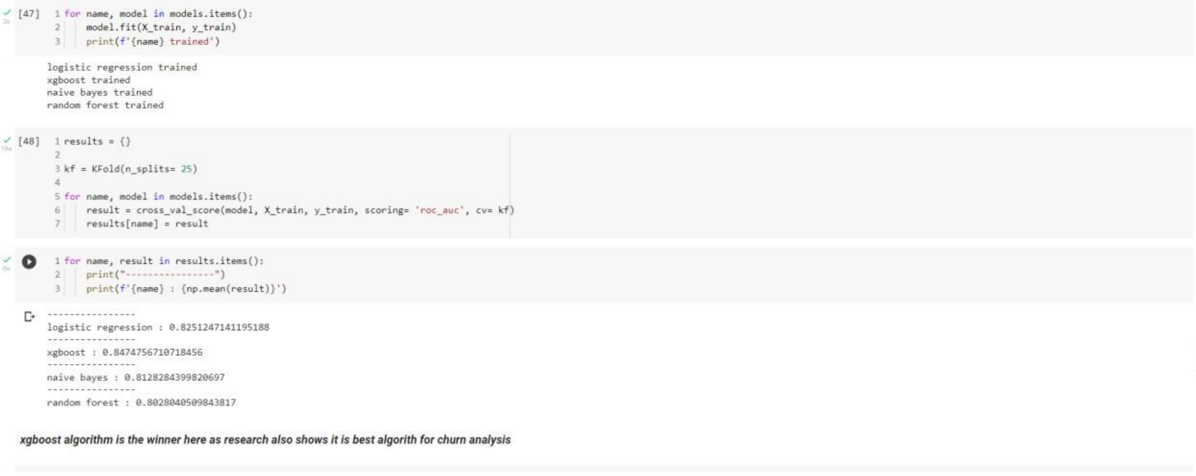
```
round(df["Dependents"].value_counts()/df.shape[0]*100,2)
```

```
0    70.04
1    29.96
Name: Dependents, dtype: float64
```



```
plt.figure(figsize=(9,7))#relationship between the Internet Services and the churn rate
ax = sns.countplot(x="InternetService", hue="Churn", data=df).set(title='Internet Service vs Churn Rates')
sns.despine()
plt.legend(title='', loc='upper right', labels=['No Churn', 'Churn'])
plt.show()
```

Model Creation



```
[47]
for name, model in models.items():
    model.fit(X_train, y_train)
    print(f'{name} trained')
```

```
logistic regression trained
xgboost trained
naive bayes trained
random forest trained
```

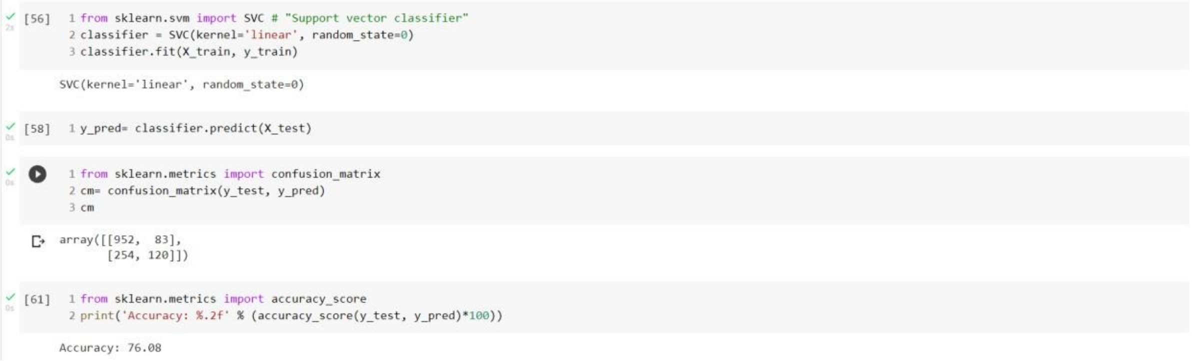
```
[48]
results = {}

kf = KFold(n_splits= 25)

for name, model in models.items():
    result = cross_val_score(model, X_train, y_train, scoring= 'roc_auc', cv= kf)
    results[name] = result
```

```
for name, result in results.items():
    print("----------------")
    print(f'{name} : {np.mean(result)}')
```

```
----------------
logistic regression : 0.8251247141195188
----------------
xgboost : 0.8474756710718456
----------------
naive bayes : 0.8128284399820697
----------------
random forest : 0.8028040509843817
```

xgboost algorithm is the winner here as research also shows it is best algorith for churn analysis

Model Evaluation



```
[56]
from sklearn.svm import SVC # "Support vector classifier"
classifier = SVC(kernel='linear', random_state=0)
classifier.fit(X_train, y_train)
```

```
SVC(kernel='linear', random_state=0)
```

```
[58]
y_pred= classifier.predict(X_test)
```

```
from sklearn.metrics import confusion_matrix
cm= confusion_matrix(y_test, y_pred)
cm
```

```
array([[952,  83],
       [254, 120]])
```

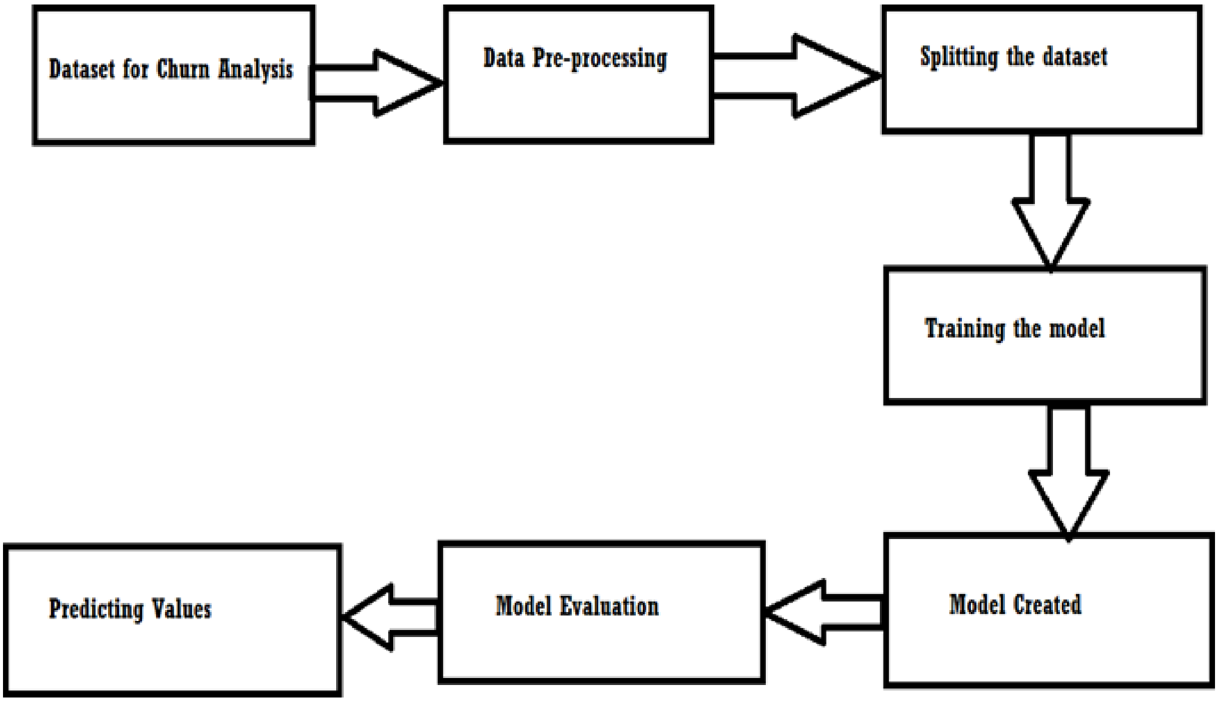
```
[61]
from sklearn.metrics import accuracy_score
print('Accuracy: %.2f' % (accuracy_score(y_test, y_pred)*100))
```

```
Accuracy: 76.08
```

11. Final Product Prototype:

Step -1: Prototype Selection

The following is the model prototype diagram for how churn analysis can be done a company’s dataset: -



Algorithm used for churn analysis:
port Vector Machine” (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. SVMs are used in applications like handwriting recognition, intrusion detection, face detection, email classification, gene classification, and in web pages. This is one of the reasons we use SVMs in machine learning. It can handle both classification and regression on linear and non-linear data.

A) **Feasibility-**It is a very feasible project because in the present time and future every sector like telecom, entertainment is facing problems of leaving customers due to heavy competition so every sector needs a churn analysis of their company and product.

B) **Viability-** As the retail industry and global company grows in India and the world, there will always be small businesses existing which can use this service to improvise on their sales and data warehousing techniques. So, it is viable to survive in the long-term future as well but improvements are necessary as new technologies emerge. So, there is huge competition between them to increase more and more customers and they have to consider also that no customer will leave them and if customers are decreasing then they have to find the problem and analyse them by using churn analysis so churn analysis is very useful for all the company.

C) **Monetization-**This service is directly monetizable as it can be directly released as a service on completion which can be used by businesses. Telcos that have the most advanced data analytics programs can focus on their core businesses and operations. For example, a challenger in Eastern Europe used an advanced analytics anti-churn model to employ a retention strategy. As a result, it has achieved a 58% churn reduction in paid TV services and 17% in mobile services. Another leading global telco provider is using customer data to improve revenue assurance and fraud management. Telco operators from the Middle East combine mobile network probe data and customer data to analyse the network's quality of service and develop new insight for personalized promotional targeting. External monetization business models range from providing customized data as a service to developing custom products and solutions for third parties. Footfall analytics that allow tracking customers' locations and using them for marketing purposes is one of the most popular monetization examples for telcos. In many cases, this allowed operators to diversify in the advertising value chain, ensuring their competitive advantage and reducing customer churn in the telecom sector. In the telecom sector churn is one of the major solutions to remain competitive. Churn is used to make a model that encompasses the customer hazard and customer survival functions accurately to obtain insights on rate of churn. Customer churn is the customer action in ending the service due to service dissatisfaction provided or other firms offering better provided within the budget of the customer. The prediction of churn is the method of recognizing the existing customers who are likely to terminate the services soon. It will be an essential influence to the revenue of an organization if it loses customers. Machine Learning algorithms have been used for an accurate prediction of churn. Machine learning is an artificial intelligence part that offers the capability to permit PC to learn the algorithm automatically without involvement of humans. ML algorithms are used for enhancing the prediction performance. Therefore, in this study the churn is proposed using SVM for predicting customer churn in the telecom sector. The proposed algorithms are used to define the best accuracy performance.

For creating the model, we have used SVM algorithm but we can also use other supervised machine learning algorithms as well. The most powerful and effective algorithm used is Support Vector Machine. It gives more accuracy than other algorithms.

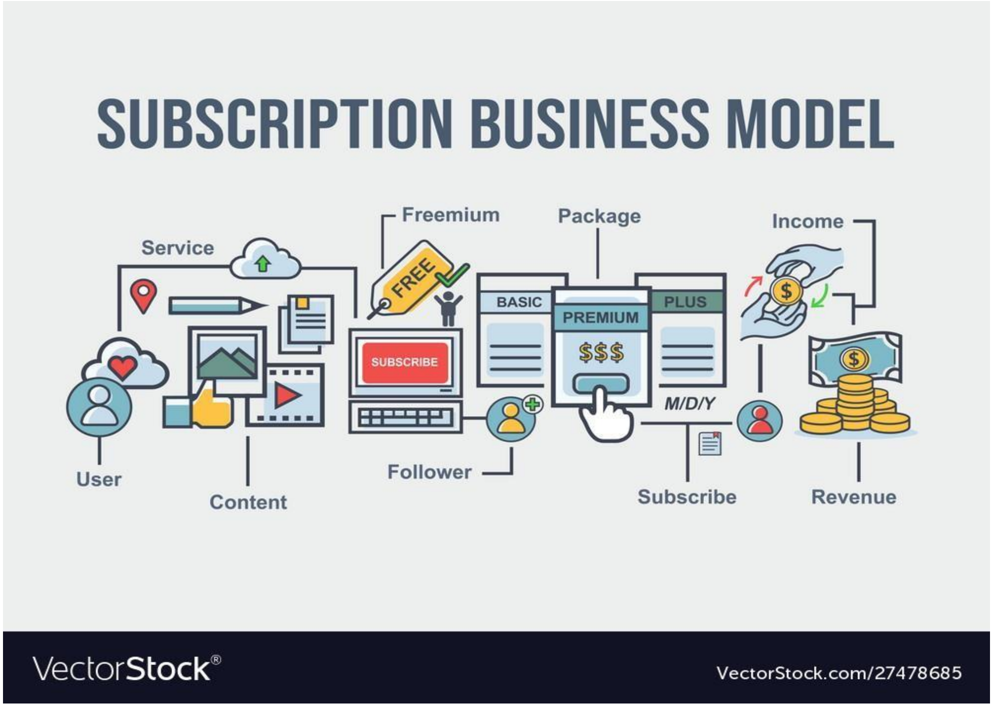
Step-2 Prototype Development:

Github link of the code - https://github.com/salvaderron/Feynn_Labs_Internship/blob/main/Churn_Analysis.ipynb

Step-3 Business Modelling:

For churn analysis and its business there are very models are available in the market but here we will discuss three model which are **Subscription based model**, **sliding box model and Learning-to-rank approach** among these three all fits at different situation but according to me subscription-based model fits in almost all condition and it is best.

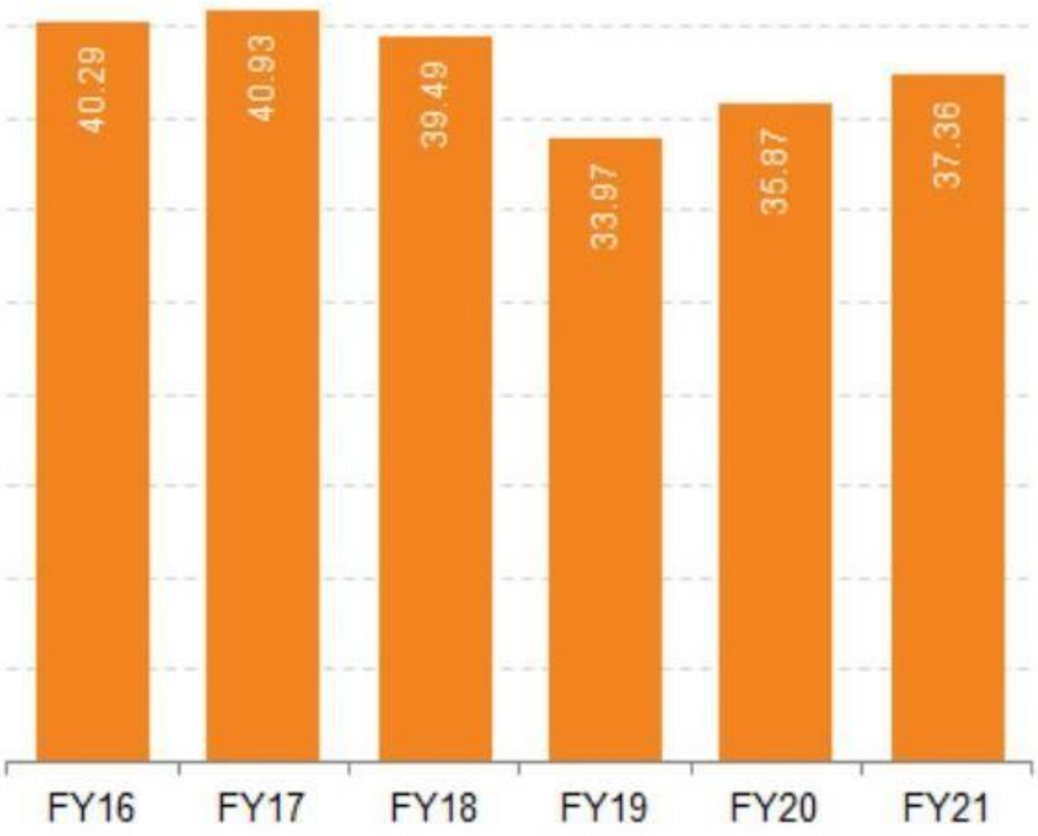
1. Subscription based model-For this service, it is beneficial to use a Subscription Based Model, where initially some features will be provided for free to engage customer retention and increase our customer count. Later it will be charged a subscription fee to use the service further for their business. In the subscription business model, customers pay a fixed amount of money on fixed time intervals to get access to the product or service provided by the company. The major problem is user conversion; how to convert the users into paid users.



2. Sliding box model-This approach avoids modelling time to event (churn) directly and focuses on predicting if an event occurred in a predefined time frame – our “box”. Deciding on how big this time frame should be is somewhat arbitrary – once again think which size best suits your business needs. This approach is fairly easy to explain and allows us to use common classification algorithms, including state-of-the-art boosting methods. It was also proven useful in situations when you have well-defined groups of customers and want to target them with specific campaigns. But this comes at a price – what you get as an output is a probability of N days without a (churn) event. Translating it into actionable “insights” may be cumbersome. Choosing the right size of the “box” is often not easy, too.

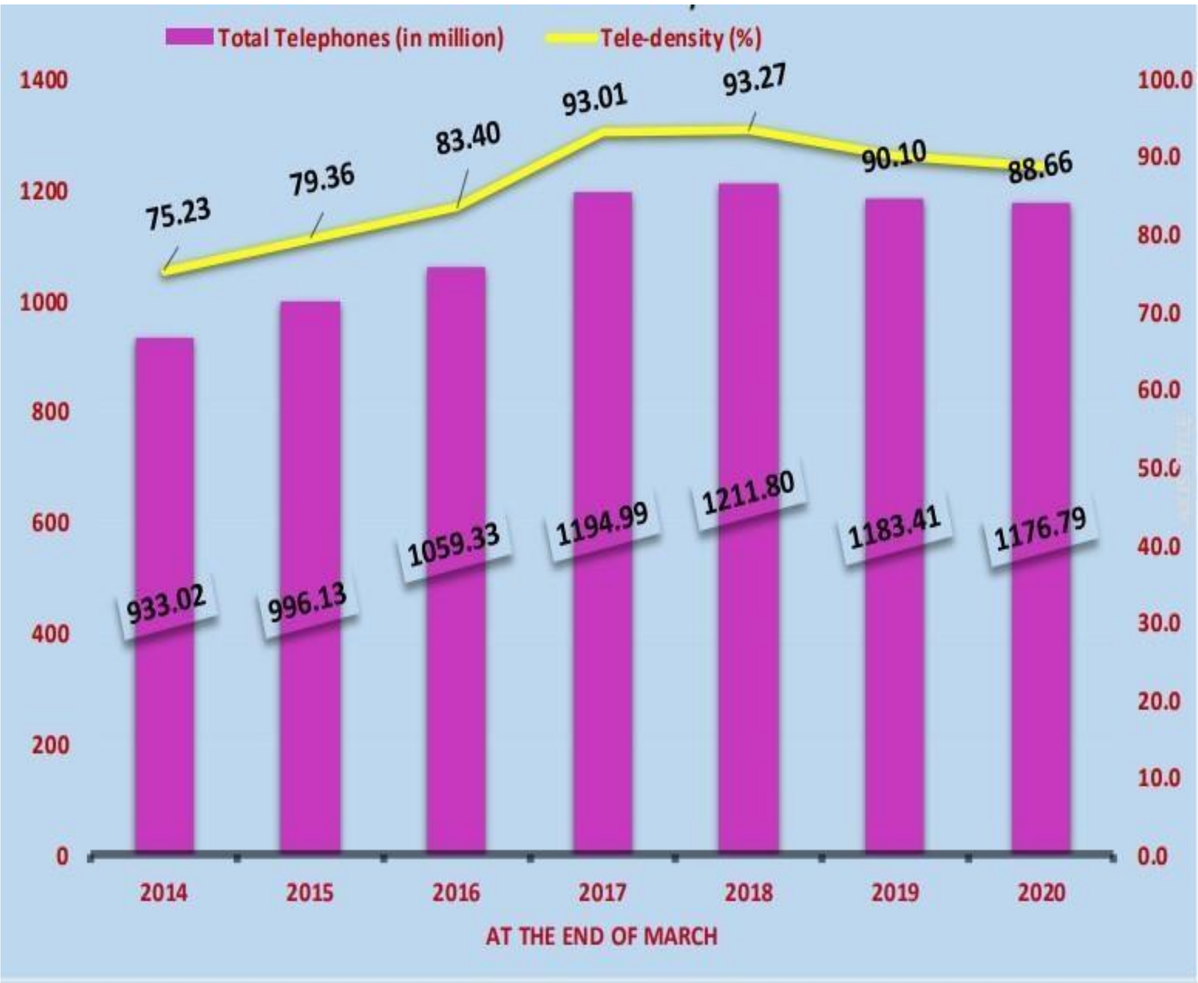
3. Learning-to-rank approach-In sliding box models, we defined churned in a binary way. However, we may want to know some “grades” here, that is, which customers are more churned than others. In such a scenario you can rank customers according to the risk of churn. This approach induces some order – let’s say we know there was at least say 5 days until an event, we can compare this to when we know that there were 3 days to an event (finally, 3<5). The customer with fewer days to the event is more “churned” than the other. In the simplest scenario, such ranking is really defined by all such pairwise comparisons. This approach also has some pitfalls – the training dataset (training time) grows quadratically because the dataset consists of pairwise combinations of all the observations (if you choose a more complex approach, it may grow even faster). Moreover, our results are somewhat “relative”. We may only be able to answer whether a customer is more churned than someone else, but answering if an individual customer is predicted as churned or not may be an entirely different issue.

Step-4 Financial Modelling:



Telecom sector gross revenue (US $ BILLION)

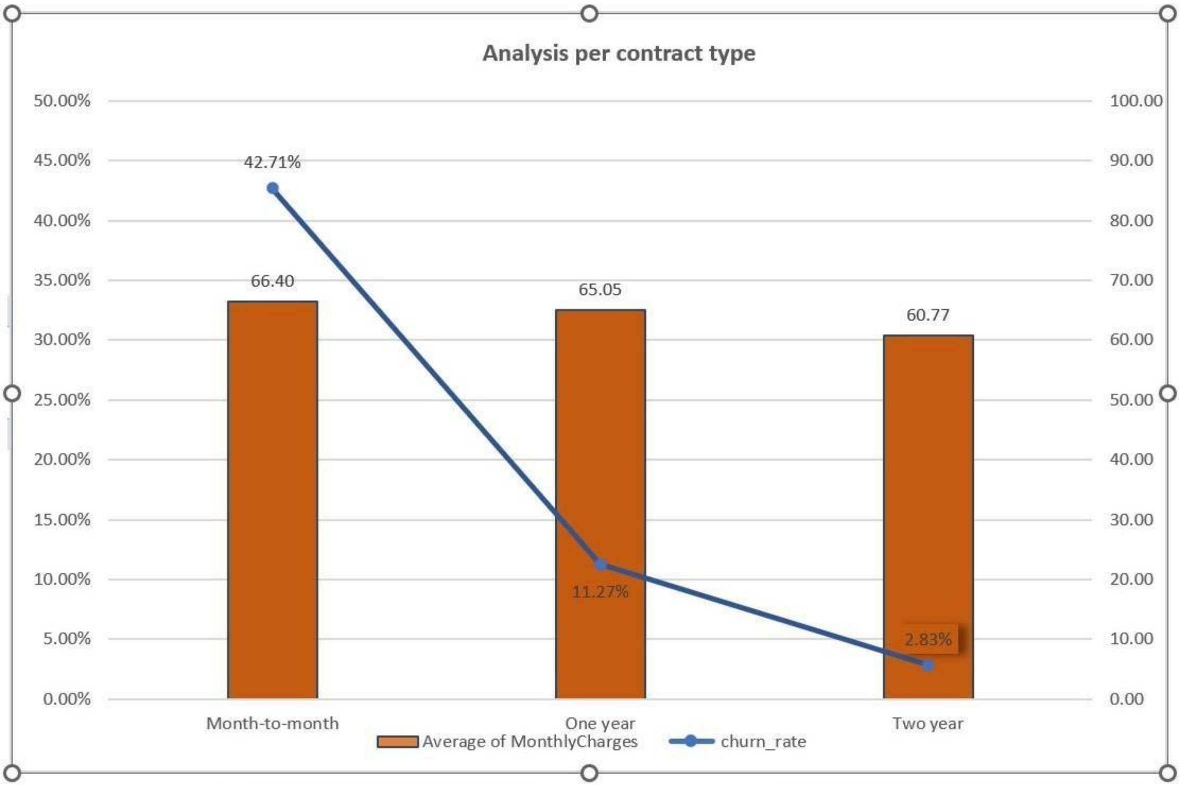
The above diagram showing how the telecom sector in India has been growing for the past five years. The Indian telecom sector is the world's second-largest telecommunications market with a subscriber base of 1.16 billion. The subscribers include base, wireless, broadband subscriptions and they have grown consistently. Gross revenue of the telecom sector stood at Rs. 64,801 crore (US$ 8.74 billion) in the first quarter of FY22.



TREND OF TELEPHONES AND TELE DENSITY IN INDIA

Teledensity: Telephone density or teledensity is the number of telephone connections for every hundred individuals living within an area.

From the above graph we can see telephone subscribers were increased consistently till 2018 then got declined in following years. One of the main reasons was operators weeded out their low revenue customers**.**



Graph of customers who churned based on the given data set.

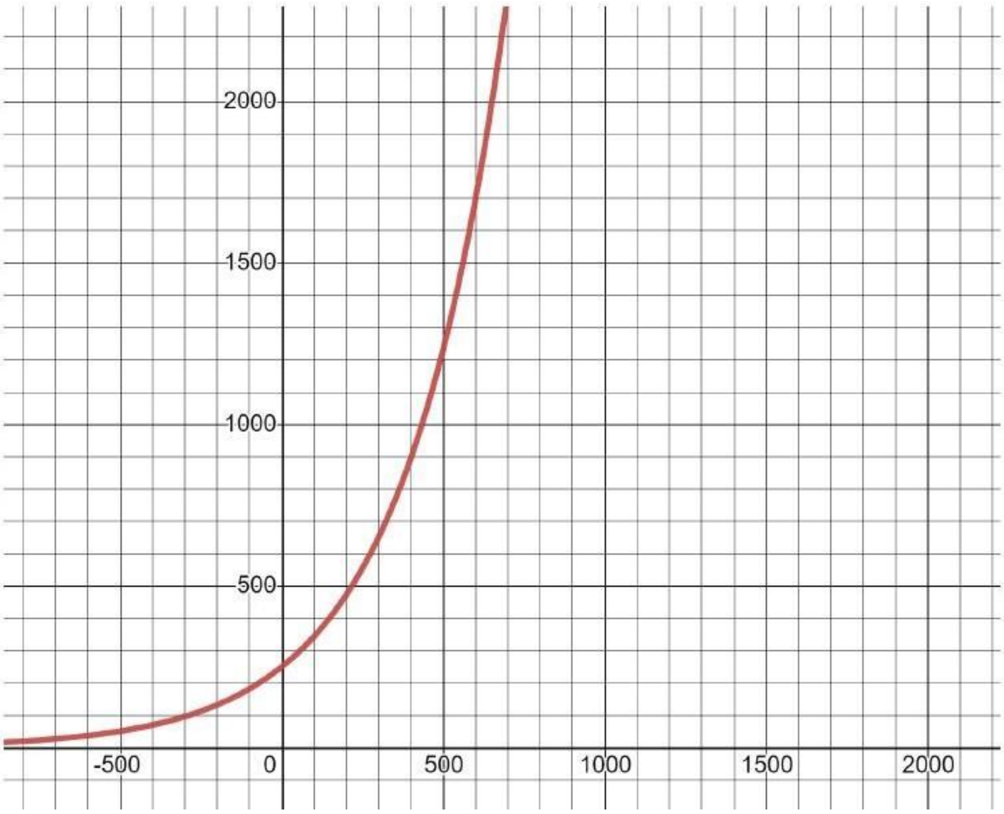
From the graph around 42.7% subscribers churned under month-to-month contract, around 11.27% and 2.83% churned under one year and two-year contract type respectively**.**

Financial equation.

As per recent statistics from Telecom Regulatory Authority of India, the number of telephone subscribers in India is around 1170.73 million subscribers having monthly growth rate of 0.78 % in the year of 2022.

For growth

Let’s consider the initial subscribers = 7043 for getting the graph



$Y = a(1+r)^t$

Y = The projected subscribers in over time, t

a= initial value or starting value of Y when t=0.

t= time interval

r=rate of growth

(1+r) = 1+0.0078 = 1.0078

For decay

$Y=a(1-r)^t$

(1-r) = 1-0.0078=0.9922

12.Conclusion:

In the competitive telecom sector standardization and public policies of mobile communication permits customers to switch over from one carrier to another carrier easily resulting in a competitive market. The prediction of churn or the task of recognizing customers who are probable to discontinue service use is alucrative and essential issue of telecom sector. Customer churn is often a critical problem for the telecom sector as customers do not delay to leave if they do not predict what they are viewing for. Customers mainly need value for money, competitive cost and greater service quality. Customer churning is associated directly to satisfaction of customer. It is a known fact that the customer acquisition cost is larger than customer retention cost that makes the retention a difficult prototype of business. There is no standard approach which resolves the churning problems of worldwide service providers of telecom industry accurately. Big data analytics with machine learning technique is used for customer churn which sets warning bells for customers before any damage could occur, providing telecom firms the chance to take precautionary steps. These techniques are used to find the churn in customers by constructing models and studying from historical information. Conducting trials with perspective of end users, collecting their views on network, normalization of data, data set pre-processing, using feature selection, removing missing values and class imbalance and changing existing variables with derived variables develops the churn prediction accuracy which supports the telecom sector to retain their customers much efficiently. It can be concluded that big data analytics with machine learning were predicted to be an effective way for recognizing churn in customers.