

Customer Churn Prediction Application for Telecom Companies

Feynn Labs

Isha Sudarshini

ABSTRACT

With the rapid development of telecommunication industry, the service providers are inclined more towards expansion of the subscriber base. To meet the need of surviving in the competitive environment, the retention of existing customers has become a huge challenge. In the survey done in the Telecom industry, it is stated that the cost of acquiring a new customer is far more than retaining the existing one. Therefore, by collecting knowledge from the telecom industries can help in predicting the association of the customers as whether or not they will leave the company. The required action needs to be undertaken by the telecom industries in order to initiate the acquisition of their associated customers for making their market value stagnant.

In this report, we have proposed a way to analyse customer information and predict which customers are likely to leave, allowing the telecom company to take proactive retention measures.

1. INTRODUCTION

Today is the competitive world of communication technologies. The technical progress and the increasing number of operators raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues:

1. acquire new customers
2. upsell the existing customers, and
3. increase the retention period of customers.

However, comparing these strategies by taking the value of return on investment (ROI) of each into account has shown that the third strategy is the most profitable strategy, which proves that retaining an existing customer cost much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another"

Customer Churn is the major issue that almost all the Telecommunication Industries in the world faces now. In telecommunication paradigm, churn is defined to be the activity of customers leaving the company and discarding the services offered by it due to dissatisfaction of the services and/or due to better offering from other network providers within the affordable price tag of the customer. This leads to a potential loss of revenue/profit to the company. Also, it has become a challenging task to retain the customers.

Many research confirmed that machine learning technology is highly efficient to predict this situation. This technique is applied through learning from previous data.

Background: Operators are losing share in today's competitive market



Industry and external outlook

- **Tougher Telecom Environment**
 - Economic instability and uncertainty
 - Mobile Market is saturated and dominated by a few players
 - Intense competition leading to price wars
- **Smarter & More Demanding Customers**
 - Escalating personal and business reliance on telecommunications
 - Technology explosion
 - More demanding, less loyal customers
 - Comparison shoppers



Internal Outlook

- **Tremendous Growth Potential**
 - Generation of vast quantities of data
 - Drive new revenue growth through customer centricity
 - Continue to exploit cost efficiencies
- **Key questions that clients ask around churn**
 - How can I understand my churn situation better; both at the organization (macro) & subscriber (micro) levels?
 - What are the key drivers of churn and what is influencing them?
 - What are the appropriate churn initiatives that should be launched to address the different churn drivers?

Need to
Manage Churn

Churn is a key driver of EBITDA margin and an industry-wide challenge.

A churned customer provides less revenue or zero revenue and increases competitor market share.

Increase acquisition cost for the service provider if the customer churned to competition. It costs up to 5 times as much for a Service Provider to acquire a new subscriber as to retain an existing one

2. PROBLEM STATEMENT

Telecom companies face a significant challenge in retaining customers in a highly competitive market.

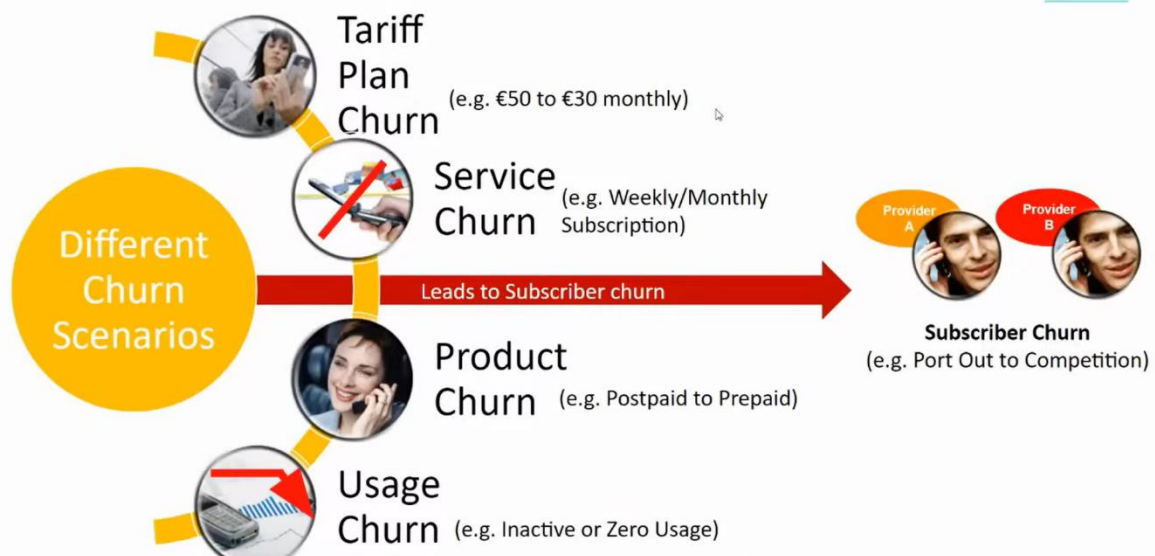
Customer churn, or the loss of subscribers, directly impacts the growth potential of these companies. Some key problems caused are:

- Revenue Loss
- Increased Marketing and Sales Cost
- Negative brand perception
- Impact on profitability

Understanding the factors that contribute to churn and accurately predicting which customers are likely to leave can help telecom companies develop targeted retention strategies, enhance customer satisfaction, and reduce churn rates.

The objective of this project is to develop a robust customer churn analysis and prediction web-app for telecom companies. The application will leverage historical customer data to identify patterns and key indicators of churn, enabling the company to proactively address potential churn issues. Churn rates can provide valuable insights into the health of a business and its customer base. By analyzing churn rates, businesses can identify trends, patterns, and potential causes for customer attrition

Subscriber Churn can be in different forms and not just exit from the base



Understanding the customer journey and its relation to churn rates

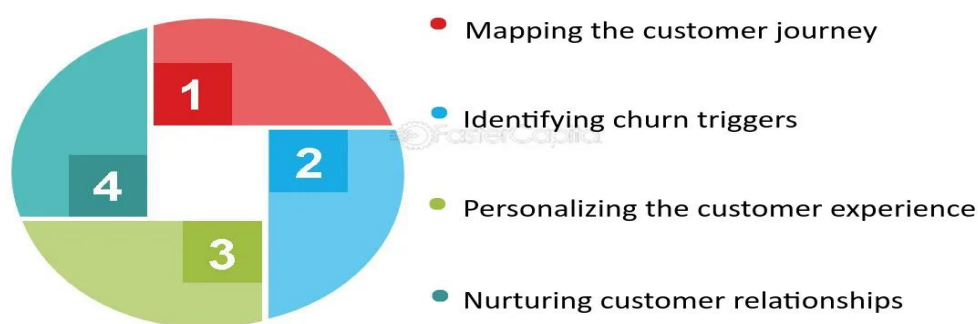
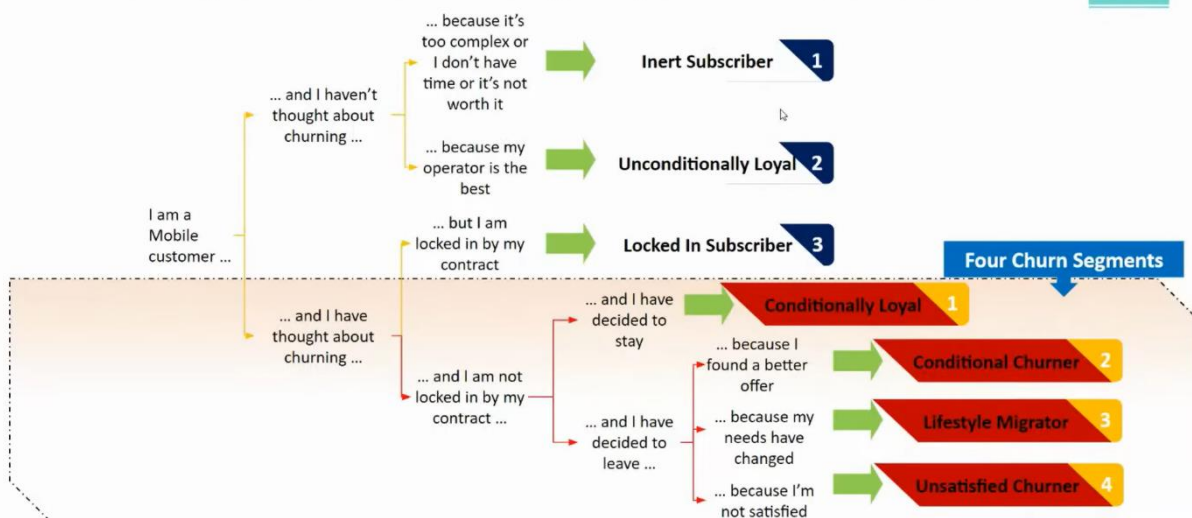


Fig 1 [3]

3. CUSTOMER NEEDS ASSESSMENT

- It is important to understand the demographics of the telecom company's customer base, including age, gender, location, and income level.
- Customers expect high-quality service across all touchpoints, including customer support, billing, and technical assistance. They demand transparent pricing structures with no hidden fees or unexpected charges.
- The application will analyse customer behaviour such as usage patterns, service preferences, and interaction history.

Decision cycle of a subscriber: Changes as per needs and/or experiences



4. BUSINESS NEEDS ASSESSMENT

- The application should align with the strategic objectives of the telecom company, such as reducing churn rate, increasing customer retention, and improving profitability.
- It is important to stay informed about industry trends, emerging technologies, and changing customer preferences in the telecommunications sector.
- The platform should be able to seamlessly integrate with existing CRM, ERP, and other business systems.

5. TARGET SPECIFICATION AND CHARACTERIZATION

The proposed system will provide real-time predictions of churn risk for individual customers, such that the predictions and insights are interpretable by business users. It will adhere to industry best practices and regulatory requirements for data protection, encryption, and access control, thus ensuring the privacy and security of customer data throughout the model lifecycle.

6. EXTERNAL SEARCH

6.1 DATASET

The dataset used for this project is an IBM sample dataset obtained from Kaggle.

<https://www.kaggle.com/datasets/blatchar/telco-customer-churn>

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

The image shows a Jupyter Notebook interface with the following code and output:

```

In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.ticker as mtick
import matplotlib.pyplot as plt

sns.set(style = 'white')

import os

In [4]: telecom_cust = pd.read_csv(r'C:\Users\ishas\Downloads\WA_Fn-UseC_-Telco-Customer-Churn.csv')

In [5]: telecom_cust.head()

Out[5]:

```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupp
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...		No
1	5576-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...		Yes
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...		No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...		Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...		No

5 rows x 21 columns

6.2 REFERENCES

- [1] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, 2015, pp. 1-6, doi: 10.1109/ICRITO.2015.7359318.
- [2] Customer Churn Prediction Using Machine Learning: Main Approaches and Models
<https://www.kdnuggets.com/2019/05/churn-prediction-machine-learning.html>
- [3] The Ripple Effect: Understanding the Impact of Churn Rates on Business
<https://fastercapital.com/content/The-Ripple-Effect--Understanding-the-Impact-of-Churn-on-Business.html>
- [4] V. Umayaparvathi, K. Iyakutti, "A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 04 | Apr-2016
- [5] Ahmad, A.K., Jafar, A. & Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. J Big Data 6, 28 (2019).
<https://doi.org/10.1186/s40537-019-0191-6>

7. BENCHMARKING

Feature/Metric	System 1: IBM SPSS Modeler	System 2: SAS Customer Intelligence 360	System 3: Microsoft Azure Machine Learning	System 4: RapidMiner Studio
Size	Software	Software	Cloud-based	Software
Weight	N/A	N/A	N/A	N/A
Cost	High	High	Medium	Low
Flexibility	High	High	High	Medium
Ease of Integration	Excellent	Excellent	Good	Good
Scalability	High	High	Excellent	Good
Algorithm Support	Wide range (Regression, Decision Trees, Neural Networks, etc.)	Wide range (Regression, Decision Trees, Neural Networks, etc.)	Wide range (Regression, Decision Trees, Neural Networks, etc.)	Good (Regression, Decision Trees, Neural Networks, etc.)
Data Handling	Excellent (Supports large datasets)	Excellent (Supports large datasets)	Excellent (Supports large datasets)	Good (Supports large datasets)
User Interface	User-friendly	User-friendly	Moderate (technical expertise required)	User-friendly
Support and Training	Extensive (online and offline)	Extensive (online and offline)	Moderate (online resources)	Moderate (online resources)
Customization	High	High	High	Medium

8.APPLICABLE REGULATIONS

While customer churn prediction in telecom doesn't directly involve health, safety, or environmental regulations, there are still standards and guidelines to consider that impact your project's development:

- Data Privacy Regulations
- Data Security Standards
- Industry Best Practices by TIA (Telecommunication Industry Association)

9. APPLICABLE CONSTRAINTS

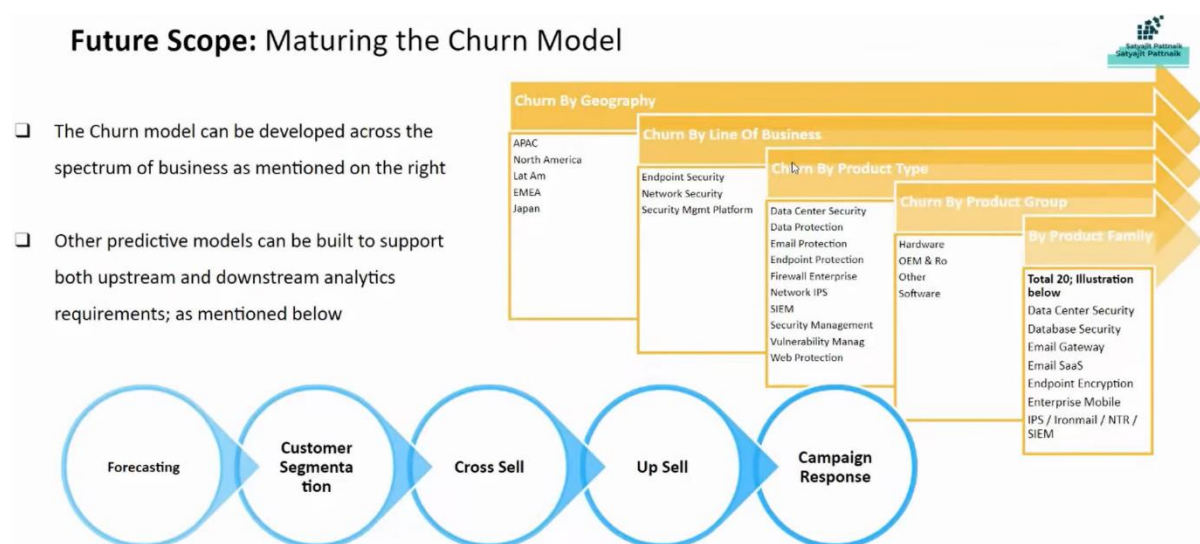
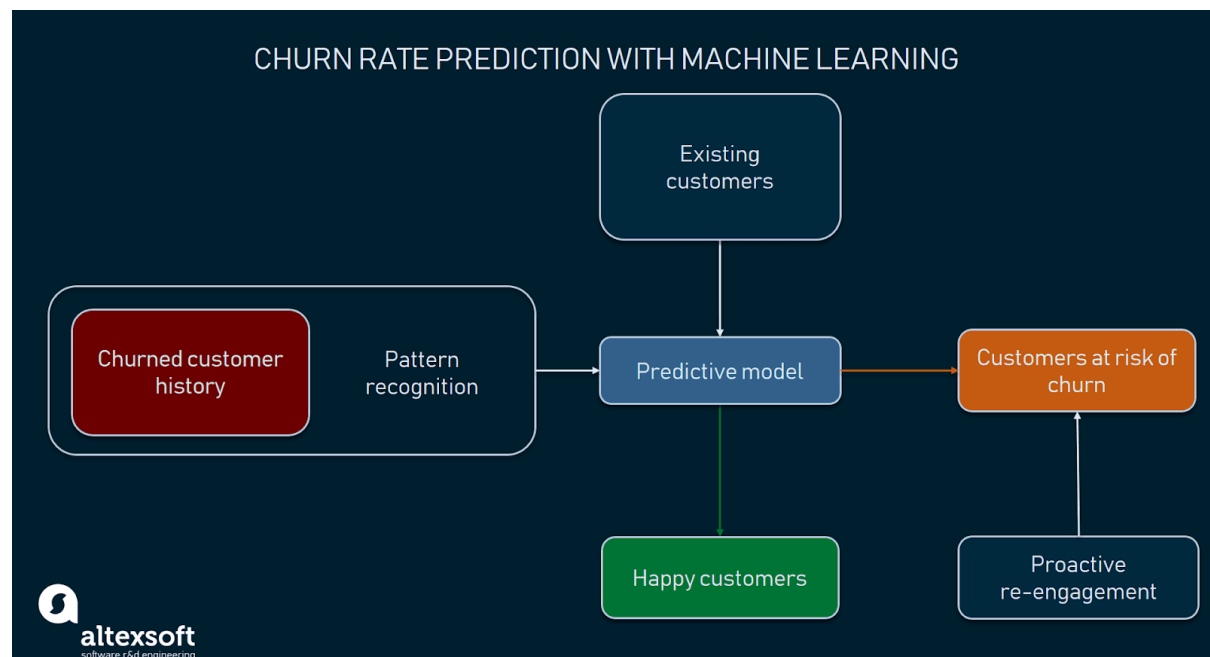
- Data quality and availability
- Data privacy and security
- Integration with existing systems

10. BUSINESS OPPURTUNITY

With increasing competition and customer expectations, telecom companies are seeking advanced solutions to retain their customer base. By addressing the critical issue of customer churn with a sophisticated, user-friendly web-app, there is a significant business opportunity

to help telecom companies enhance their customer retention efforts, optimize revenue, and stay competitive in a dynamic market.

11. CONCEPT GENERATION



The concept generation process involves several steps to ensure that a wide range of creative and innovative ideas are considered, both at the system-level and subsystem-level.

After outlining the problem of customer churn in telecom companies and the need for a predictive solution, the primary goals of the web-app, such as accurate churn prediction,

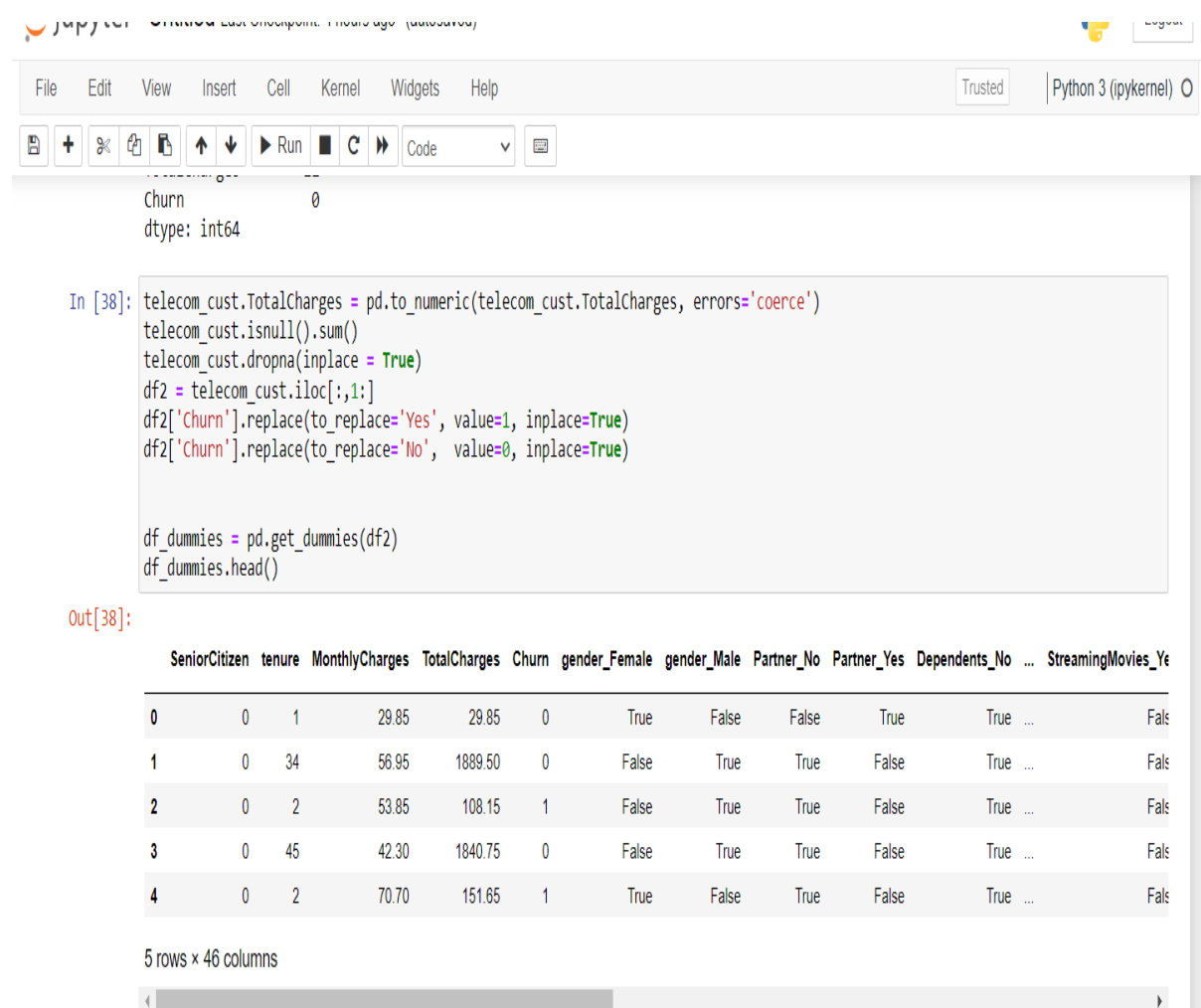
actionable insights, user-friendly interface, and seamless integration with existing systems were established.

The dataset was then obtained from Kaggle.

11.1 DATA PREPROCESSING

Data preprocessing, that includes both data cleaning and data transformation, encompasses all activities that prepare the raw data for analysis or modelling.

The missing values were handled and the unnecessary columns removed, thus correcting the inconsistencies in the data. Then data transformation was done where the categorical variables were converted to numeric and dummy variables were created.



```
Churn      0
dtype: int64

In [38]: telecom_cust.TotalCharges = pd.to_numeric(telecom_cust.TotalCharges, errors='coerce')
telecom_cust.isnull().sum()
telecom_cust.dropna(inplace = True)
df2 = telecom_cust.iloc[:,1:]
df2['churn'].replace(to_replace='Yes', value=1, inplace=True)
df2['churn'].replace(to_replace='No', value=0, inplace=True)

df_dummies = pd.get_dummies(df2)
df_dummies.head()
```

Out[38]:

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents_No	...	StreamingMovies_Yes
0	0	1	29.85	29.85	0	True	False	False	True	True	...	False
1	0	34	56.95	1889.50	0	False	True	True	False	True	...	False
2	0	2	53.85	108.15	1	False	True	True	False	True	...	False
3	0	45	42.30	1840.75	0	False	True	True	False	True	...	False
4	0	2	70.70	151.65	1	True	False	True	False	True	...	False

5 rows x 46 columns

11.2 DATA ANALYSIS

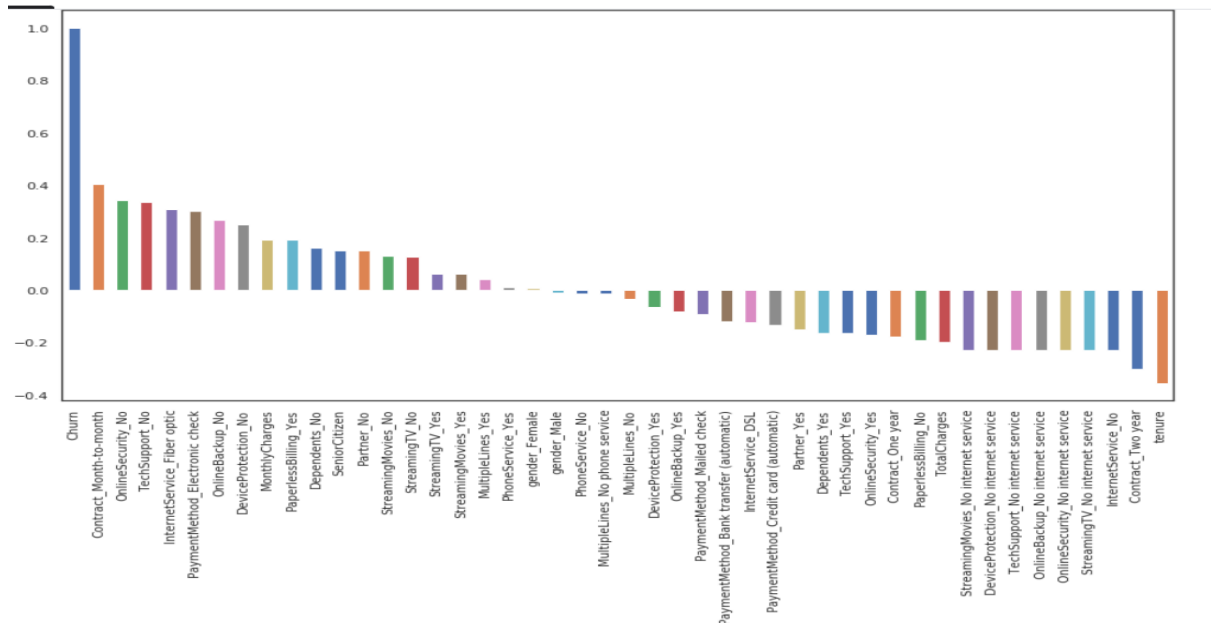
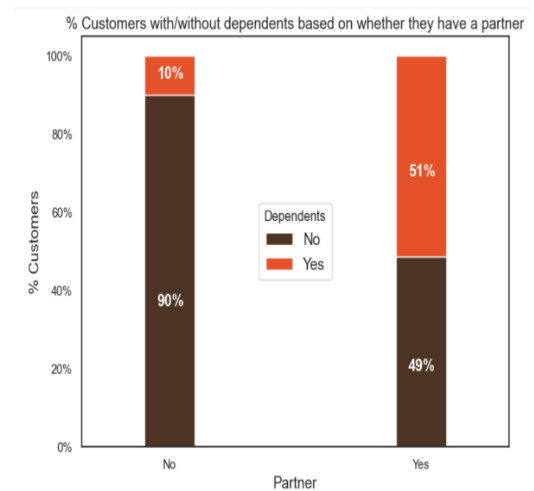
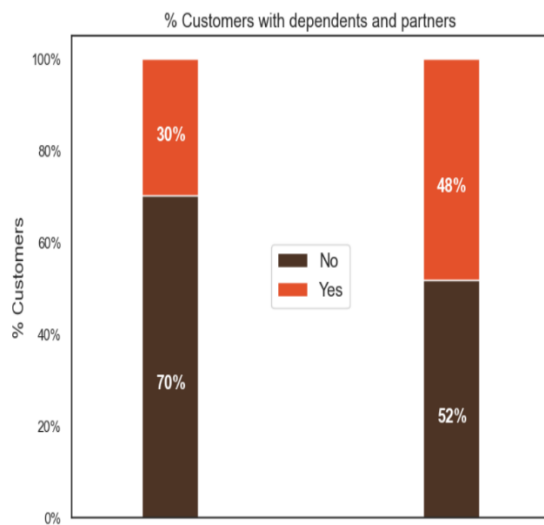
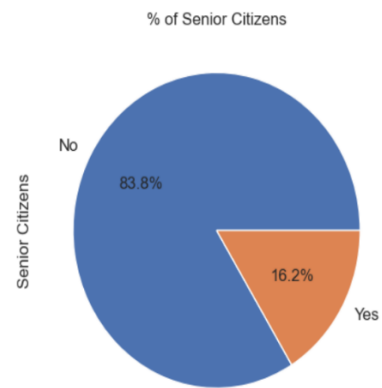
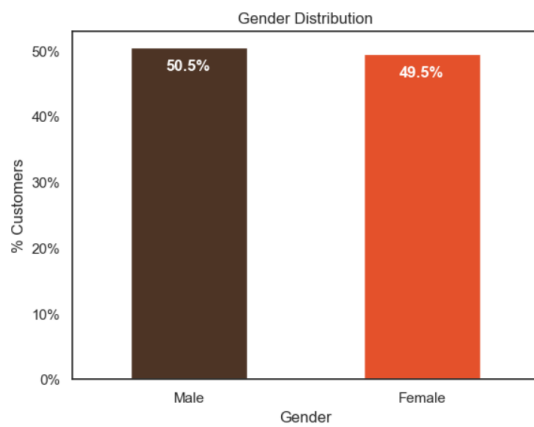
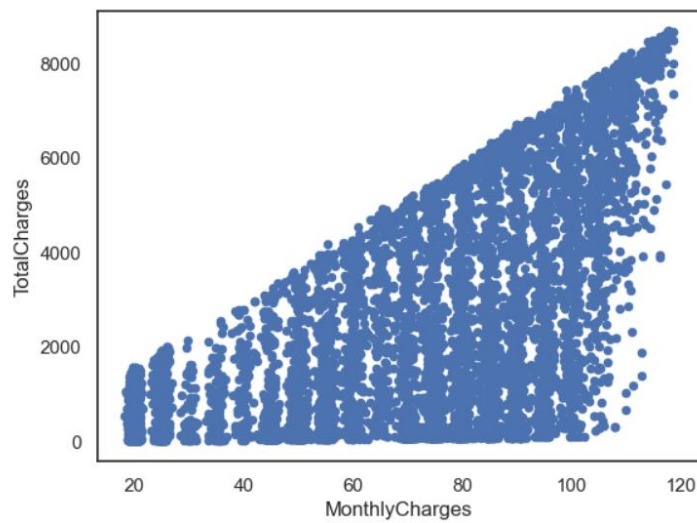
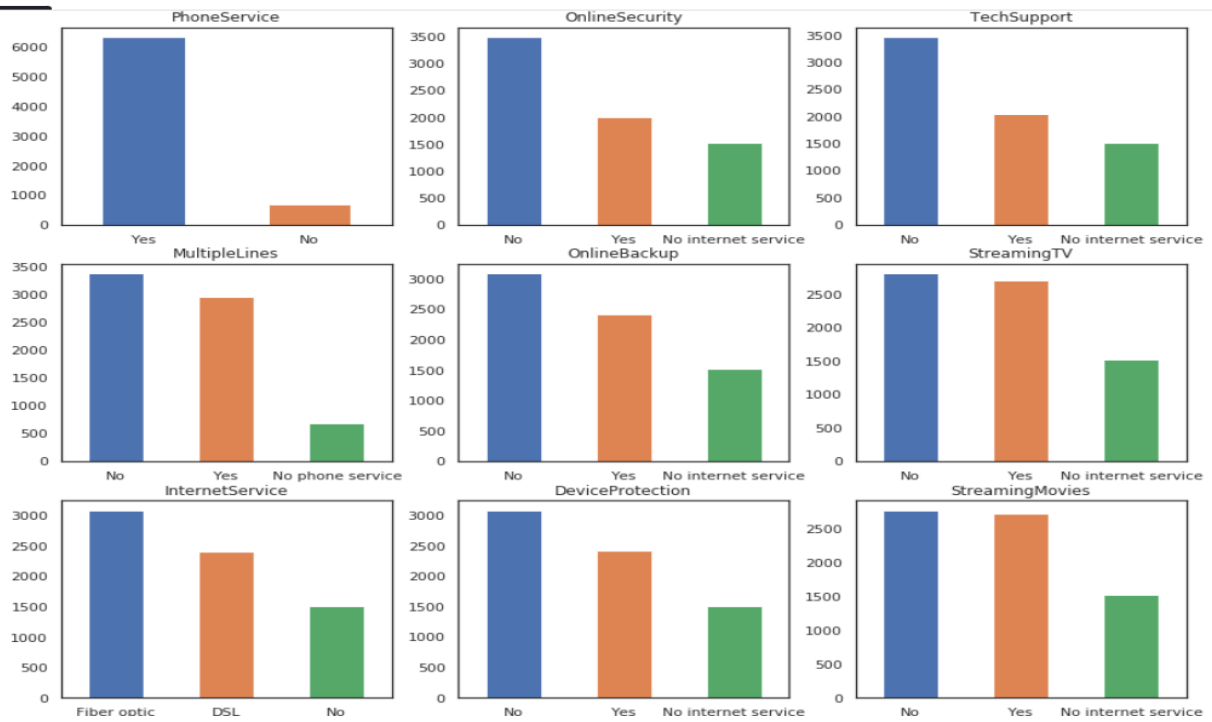
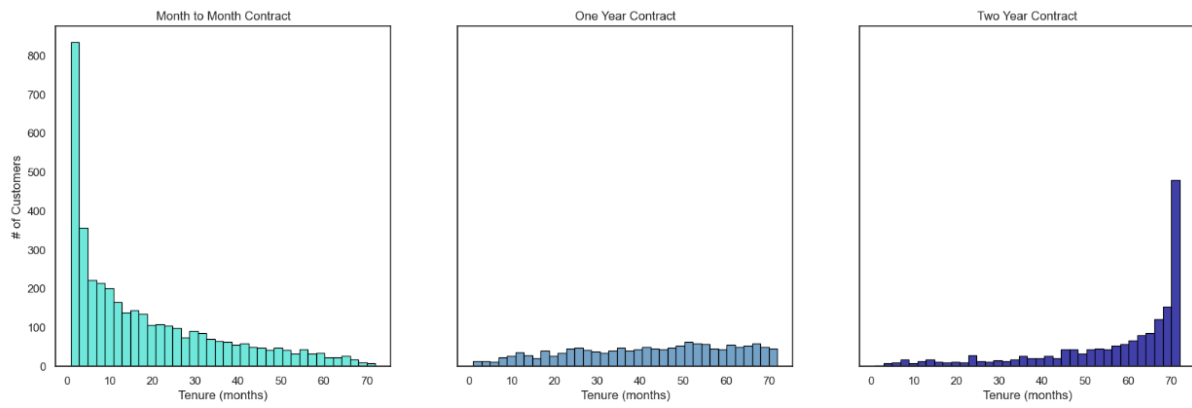


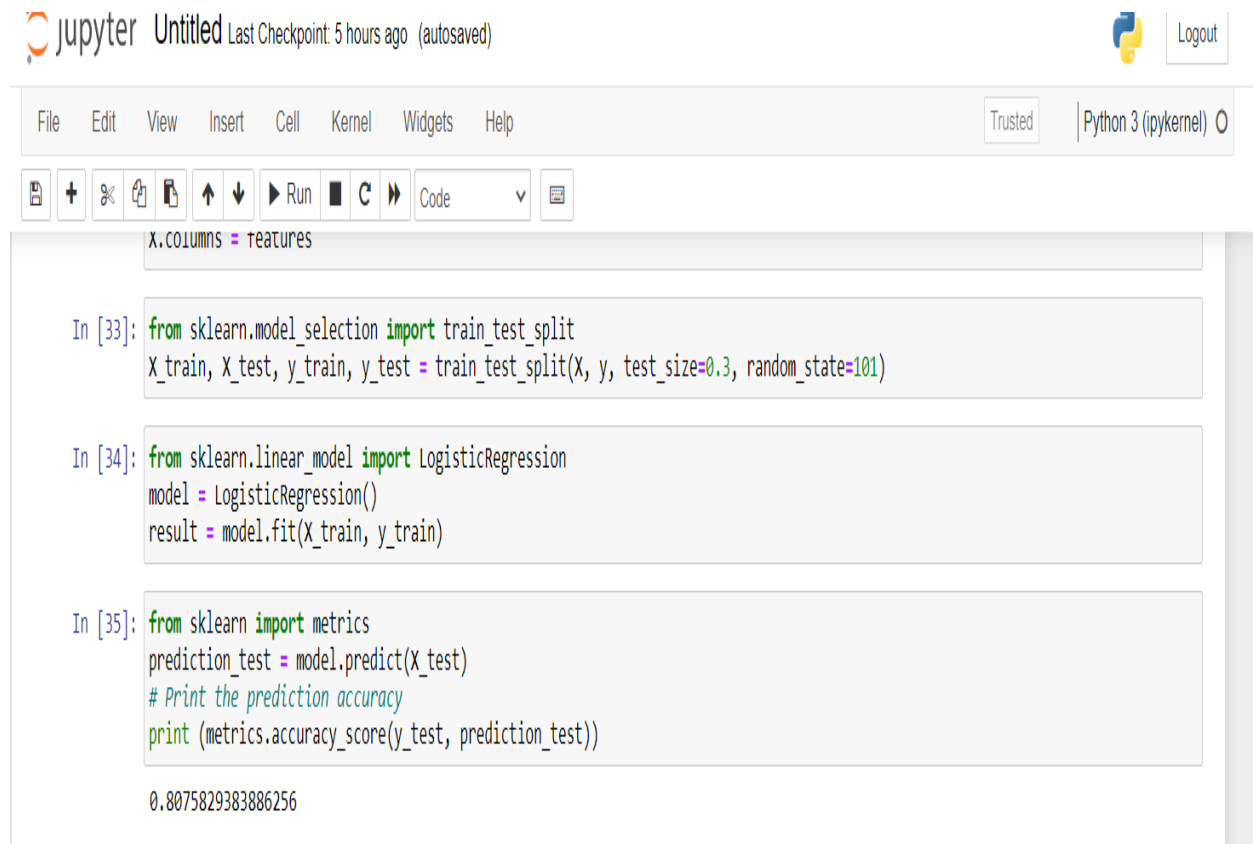
Fig: Correlation of “churn” variable with other variables





11.3 ML MODELLING

Logistic Regression algorithm is used for the model.



The screenshot shows a Jupyter Notebook titled 'Untitled' with a last checkpoint 5 hours ago. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a 'Trusted' status indicator, and a 'Python 3 (pykernel)' environment selector. Below the menu is a toolbar with icons for file operations, navigation, and execution. The notebook contains three code cells:

```
X.columns = features
```

```
In [33]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=101)
```

```
In [34]: from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
result = model.fit(X_train, y_train)
```

```
In [35]: from sklearn import metrics
prediction_test = model.predict(X_test)
# Print the prediction accuracy
print (metrics.accuracy_score(y_test, prediction_test))
```

The output of the final cell is the accuracy score: 0.8075829383886256.

12. CONCEPT DEVELOPMENT

The concept can be developed by using appropriate API and cloud services as necessary.

13. PROTOTYPE SELECTION

Based on the following criteria, the prototype is selected and made using logistic regression algorithm.

1. FEASIBILITY

Developing a customer churn prediction model is generally feasible within a 2-3 year timeframe, using the necessary data, resources, and expertise.

2. VIABILITY

The viability of a customer churn prediction model is generally high, given the growing importance of customer retention and data-driven decision-making. As businesses continue to focus on retaining existing customers and leveraging data analytics, a churn prediction model is likely to remain relevant and valuable in the long term.

3. MONETIZATION

The model can be monetized through various channels, such as SaaS offerings, consulting services, and integration with other tools, providing multiple streams of revenue

14. PROTOTYPE DEVELOPMENT

GitHub Link: <https://github.com/ishaaaa12/Customer-Churn-Prediction>

15. FINAL PROTOTYPE

The final prototype of the customer churn analysis web-app for telecom companies will encompass a comprehensive, user-friendly platform designed to predict and mitigate customer churn effectively. The features of the final prototype:

- User Interface
- Churn Prediction and Analysis
- Data integration and management

16. BUSINESS MODELLING

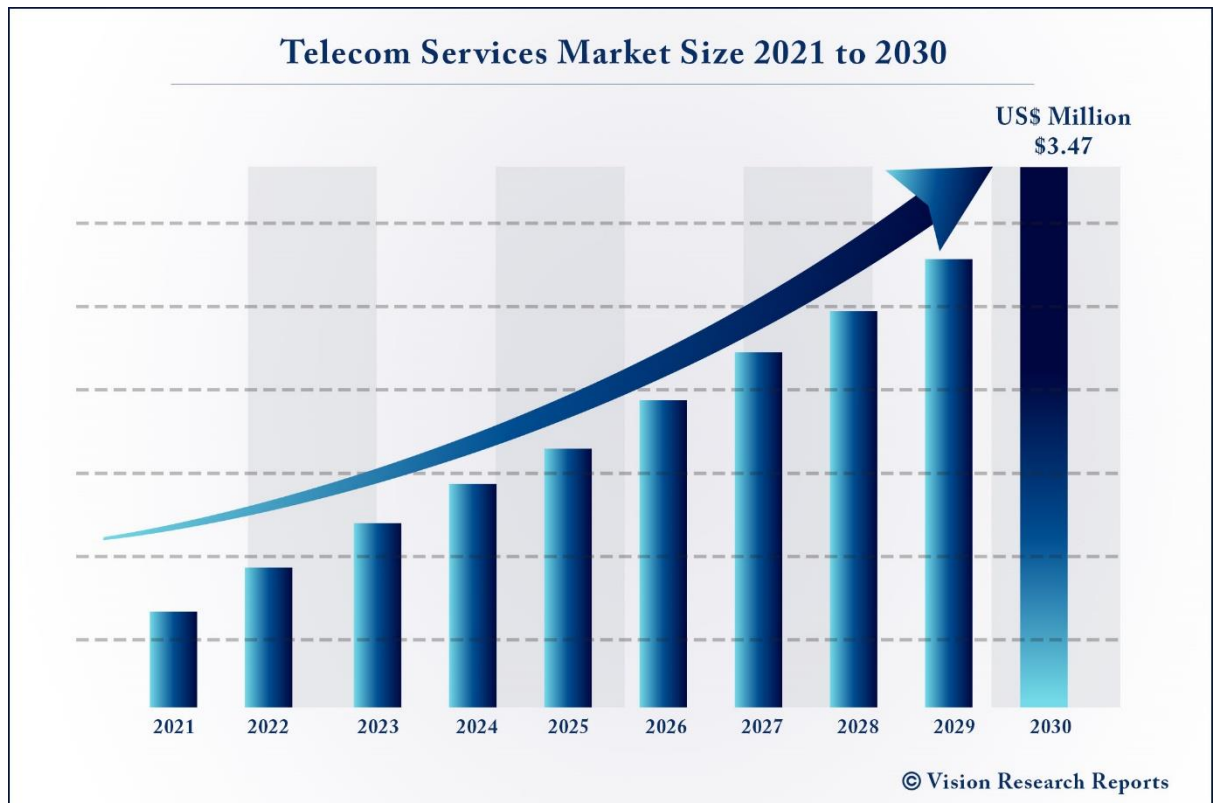
For this service, a revenue stream can be defined, with subscription fees for a SaaS model, consulting services for custom integrations, and additional data analytics services.



17. FINANCIAL MODELLING

Market:

By carefully crafting the value proposition, targeting the right market segments, and executing a well-planned go-to-market strategy, we can successfully launch and position your customer churn prediction model as a valuable tool for telecom companies seeking to enhance customer retention and drive business growth.



FINANCIAL EQUATION:

To estimate the financial performance of the customer churn prediction model as a product for businesses, we can design an equation that considers the following elements:

- Total Revenue (R): The income generated from selling the churn prediction model.
- Pricing (p): The price charged per subscription or unit of the model.
- Number of Customers (n): The total number of business clients subscribing to the model.
- Retention Rate (r): The percentage of customers retained over time, impacting future revenue.
- Total Costs (C): The combined costs of development, maintenance, customer support, and marketing.

The estimated equation for the total profit (P) can be expressed as:

$$P = R - C$$

Where the revenue RRR is:

$$R = n \cdot p \cdot (1 + r)$$

So, the equation becomes:

$$P = (n \cdot p \cdot (1 + r)) - C$$

14. CONCLUSION

The development and deployment of a customer churn analysis web-app represent a significant strategic investment for telecom companies looking to reduce customer attrition and enhance business sustainability. By leveraging machine learning algorithms, this product idea provides predictive insights that are actionable.