

1st Nov, 2023



JCT College of Engineering
and Technology

Customer Churn Prediction

PHASE 5

Abdulla A
Abhijith J
Don M S
Mathews M Jaico
Mohammed Fateen Shabeer

Customer Churn Prediction

Phase 5

Problem Definition:

The problem of customer churn is a critical concern for businesses in various industries. Customer churn occurs when customers stop using a product or service, which can lead to revenue loss and decreased customer loyalty. To address this issue, a project could be defined as follows:

"Design a customer churn prediction system that helps businesses identify and proactively reduce customer churn by analyzing historical data and predicting which customers are at risk of leaving."



Design Thinking Project Idea:

1. Empathize:

- Conduct interviews and surveys with customers to understand their pain points, reasons for churning, and their expectations from the business.

2. Define:

- Define the key metrics that indicate customer churn (e.g., usage frequency, customer support interactions, payment delays).
- Create personas representing different customer segments.

3. Ideate:

- Brainstorm potential features and data sources for predicting churn, such as customer behavior data, demographic information, and customer feedback.

4. Prototype:

- Develop a prototype of the churn prediction system that uses historical data to make predictions.
- Create a user interface for business users to interact with the system.

5. Test:

- Test the prototype with business stakeholders to gather feedback.
- Refine the system based on feedback and iterate on the design.

6. Implement:

- Build the full-scale customer churn prediction system using machine learning algorithms and data pipelines.

- Integrate it with the business's data infrastructure.

7. Measure:

- Monitor the system's performance in real-time.
- Track key performance indicators (KPIs) related to churn reduction and customer retention.

8. Improve:

- Continuously update and improve the churn prediction model using new data and machine learning techniques.
- Gather customer feedback on the effectiveness of retention strategies.

9. Scale:

- Expand the use of the churn prediction system to other business units or product lines.
- Share best practices and insights across the organization.

Note:

This design thinking project would involve a user-centered approach to solving the problem of customer churn. By understanding customer needs and iteratively designing and testing solutions, businesses can effectively reduce churn and improve customer satisfaction.

Innovation

An innovative idea for a customer churn prediction project could involve using advanced machine Learning techniques, natural language processing, and customer feedback analysis. Here's a high-level Outline of such a project:

1. **Data Gathering:** Collect customer data, including transaction history, customer demographics, Customer interactions, and customer feedback.
2. **Feature Engineering:** Create a wide range of features, including traditional metrics like customer Tenure and purchase frequency, but also incorporate sentiment analysis of customer feedback, social Media activity, and even external factors like economic indicators.
3. **Natural Language Processing (NLP):** Utilize NLP techniques to analyze customer feedback and Sentiment. Extract valuable insights from text data, such as identifying common pain points or areas Where customers express dissatisfaction.
4. **Predictive Modeling:** Build predictive models using machine learning algorithms like Random Forests, Gradient Boosting, or deep learning techniques like Recurrent Neural Networks (RNNs) for time-series Data. Use these models to predict which customers are likely to churn.
5. **Explainability and Interpretability:** Implement methods to make the models interpretable, so you can Explain why a particular customer is predicted to churn. This can help in devising targeted retention Strategies.
6. **Continuous Learning:** Implement a system that continuously learns from new data and updates the Churn prediction models. This ensures that the models stay relevant over time as customer behavior Evolves.
7. **Dynamic Intervention:** Develop strategies for proactive intervention with at-risk customers. This might Involve personalized incentives, targeted marketing campaigns, or customer support outreach.

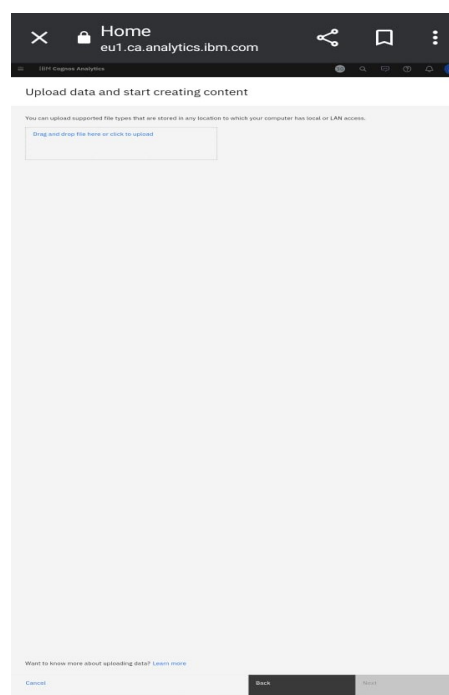
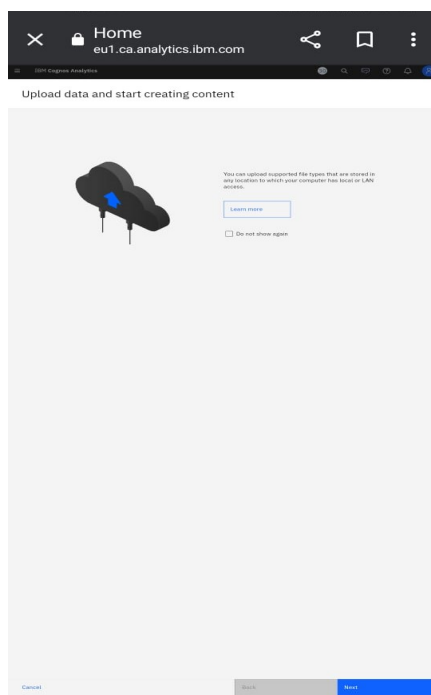
8. **Feedback Loop:** Incorporate a feedback loop to measure the effectiveness of your retention strategies. Analyze which interventions were successful in reducing churn and refine your approach accordingly.
9. **Visualization and Reporting:** Create dashboards and reports to provide insights to decision-makers. Visualize churn trends, customer segments, and the impact of interventions.
10. **A/B Testing:** Implement A/B testing to rigorously evaluate the effectiveness of different retention Strategies. This will help in optimizing your approach further.
11. **Customer Segmentation:** Segment your customer base into different groups based on their behavior And characteristics. Tailor retention strategies for each segment to maximize their effectiveness.
12. **AI-Powered Chatbots:** Deploy AI-powered chatbots for real-time customer support. These bots can Address common issues and provide assistance, potentially reducing churn.
13. **Ethical Considerations:** Ensure that the data and AI models used in this project adhere to ethical Guidelines and privacy regulations, such as GDPR or CCPA.

By combining data analysis, NLP, predictive modeling, and ongoing refinement, we can create an Innovative customer churn prediction system that not only identifies at-risk customers but also Empowers our business to proactively retain them.

Dataset loading and preprocessing

Uploading dataset:

- In this project phase we are going to upload the dataset provided from kaggle in skillup platform
- First, we have to register in the IBM Cloud service and IBM cognos analytics site
- After registering to IBM account we need to create a project and upload our dataset from our device



Preprocessing dataset :

- Clean the data by handling missing values, outliers, data quality issues
- Transform and reshape the data as needed . This might include feature engineering , Data aggregation , or other data preparation.

The screenshot shows a dialog box titled "Create a data group (numeric style)". The "Name" field is set to "MonthlyCharges (Group)". The "Groups" dropdown is set to "5". The "Range border values" section shows five ranges with their corresponding border values:

Group names	Groups	Range border values
98.650 and above	5	Highest
78.550 to < 98.650		98.65
58.450 to < 78.550		78.55
38.350 to < 58.450		58.45
less than 38.350		38.35
		Lowest

Below the ranges, there is a checkbox labeled "Group NULL values as" which is currently unchecked. At the bottom, there are three buttons: "Create a data group (text style)", "Cancel", and "Create".

- Data Preprocess is one of the crucial step in the Data Analytics . Otherwise our analysis can be irrelevant due to unwanted noises in the data.
- This includes null value removal , formatting ,etc.

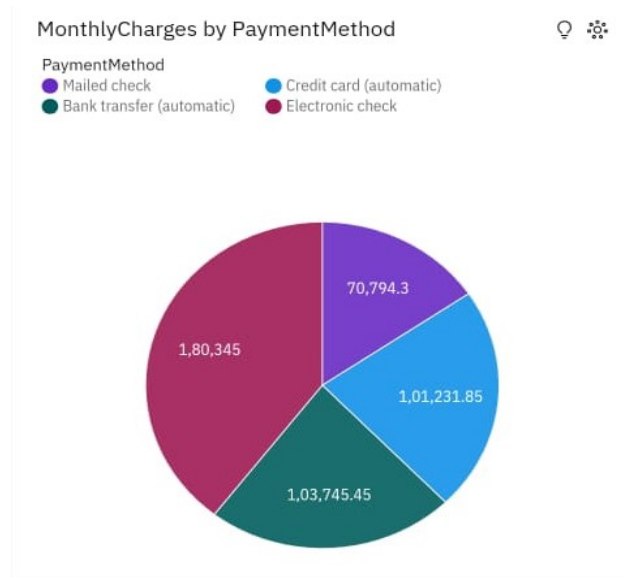
The screenshot shows a dialog box titled "Create a data group (numeric style)". The "Name" field is set to "TotalCharges (Group)". The "Groups" dropdown is set to "5". The "Range border values" section shows five ranges with their corresponding border values:

Group names	Groups	Range border values
6951.60 and above	5	Highest
5218.40 to < 6951.60		6951.6
3485.20 to < 5218.40		5218.4
1752.00 to < 3485.20		3485.2
less than 1752.00		1752
		Lowest

Below the ranges, there is a checkbox labeled "Group NULL values as" which is checked, and the value "Null" is entered in the adjacent field. At the bottom, there are three buttons: "Create a data group (text style)", "Cancel", and "Create".

Analysis and visualisation (dashboard creation)

Analysis :

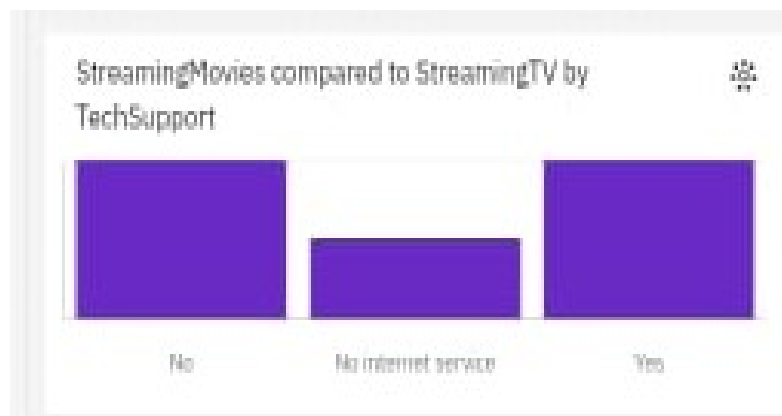
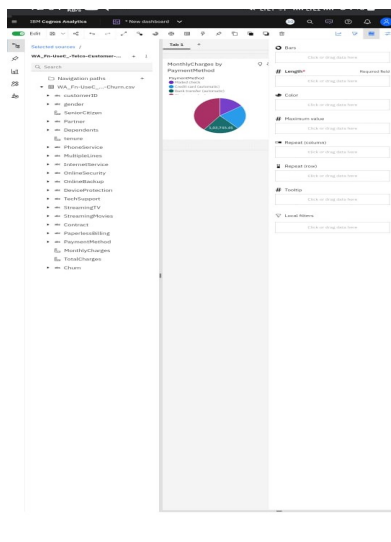


Use IBM Cognos for descriptive and exploratory data analysis. Conduct statistical analysis to understand the dataset's characteristics.

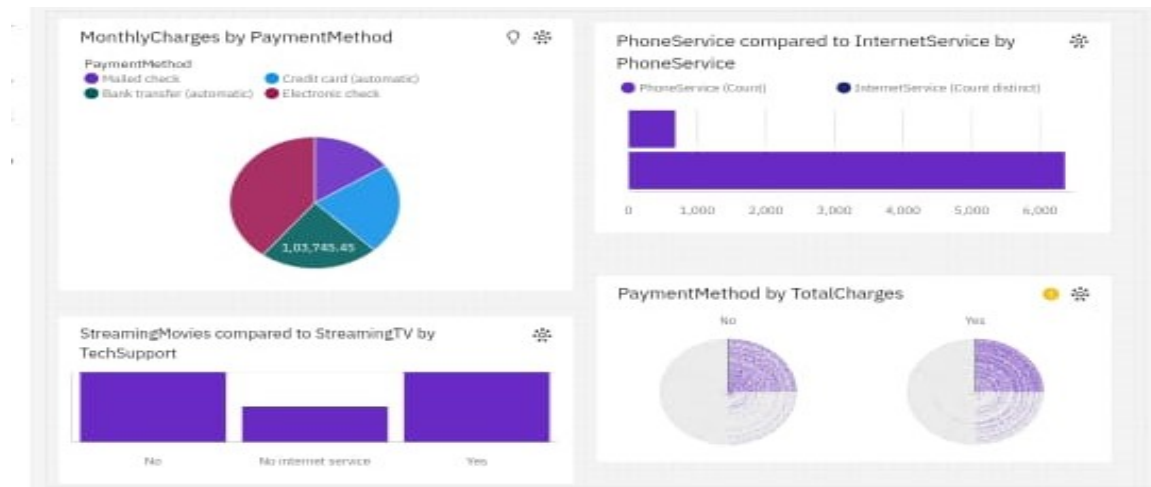
- Identify patterns, trends, and correlations within the data. Look for potential factors that may contribute to website traffic analysis.

Visualization :

- Create visualizations using IBM Cognos to present your findings. This might include:
- Histograms, scatter plots, and box plots to visualize data distributions and relationships.
- Line charts or time series visualizations to explore trends over time.
- Dashboards that summarize key insights in a user-friendly manner.



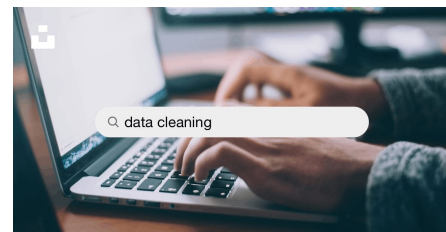
Dashboard



Data cleaning:

Data cleaning is a process of removing inconsistencies in the dataset And incorrect values .It also in involves handling missing values Either by removing them or assigning them average values. It helps To improve the efficiency of the model.

In the first step, we will only remove the unnecessary data points From the dataset which does not helps in improving the model Performance.



Initially we import the necessary packages for our data cleaning Process and also in the future purposes,

```
import numpy as np
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

We use these packages in various stages of our cleaning process and Also in the future in which we need to build models.

Here, we read the .csv files of telco customer churn prediction and then explore the count values of their subjects

```
df = pd.read_csv("/content/WA_Fn-UseC_-Telco-Customer-Churn.csv")
df
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	No
...
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	Yes	Yes	Yes
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	Yes	No	Yes
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...	No	No	No

```
df['Churn'].value_counts()
No      5174
Yes     1869
Name: Churn, dtype: int64
```

```
cleanDF.dtypes
gender                object
SeniorCitizen         int64
Partner               object
Dependents            object
tenure                int64
PhoneService          object
MultipleLines         object
InternetService       object
OnlineSecurity        object
OnlineBackup          object
DeviceProtection      object
TechSupport           object
StreamingTV           object
StreamingMovies       object
Contract              object
PaperlessBilling      object
PaymentMethod         object
MonthlyCharges        float64
TotalCharges          object
Churn                 object
dtype: object
```

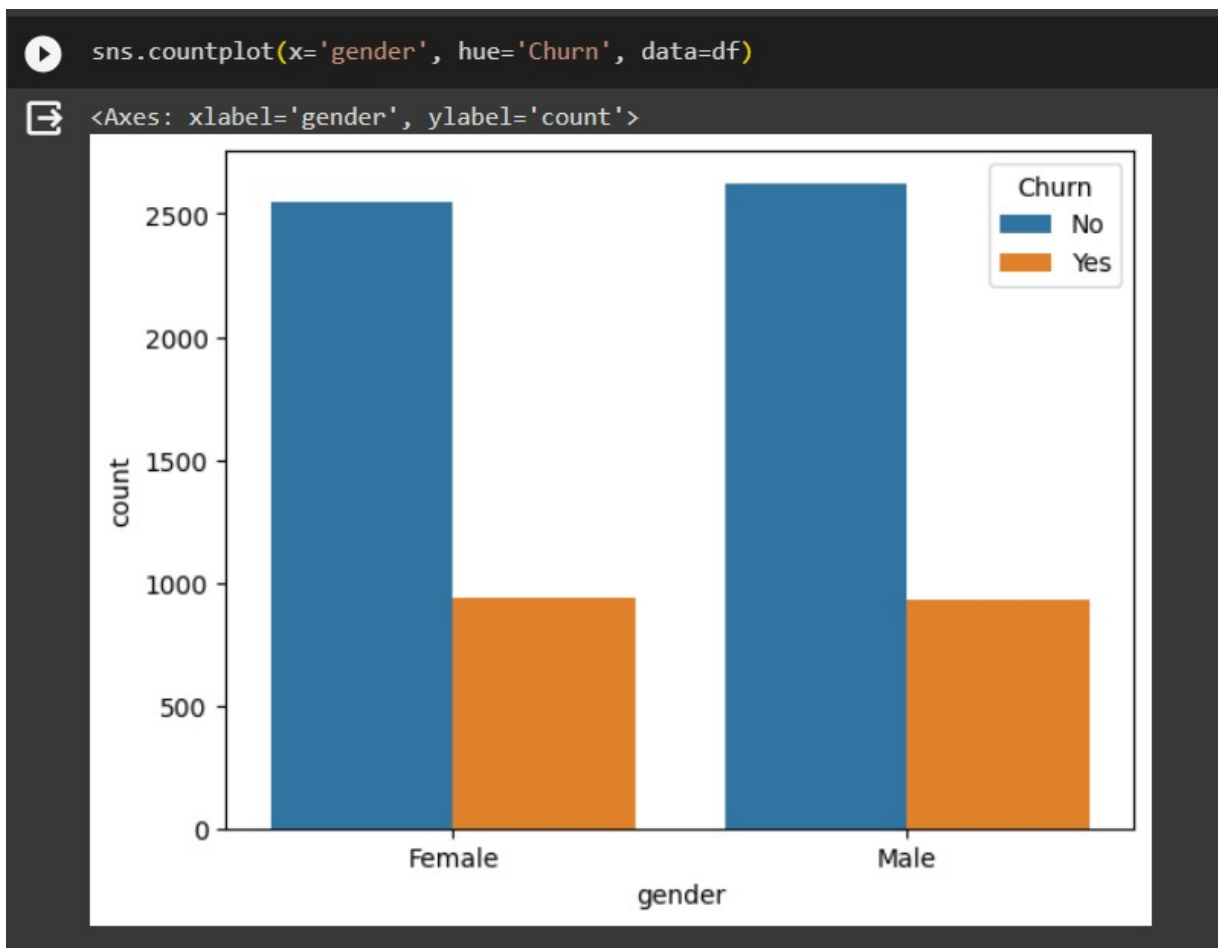

Preprocessing and visualization :

Preprocessing is a crucial step in data analysis and machine learning. It involves tasks like cleaning, transforming, and organizing data to make it suitable for further analysis or modeling

Data visualization is the process of representing data graphically to gain insights and make it easier to understand

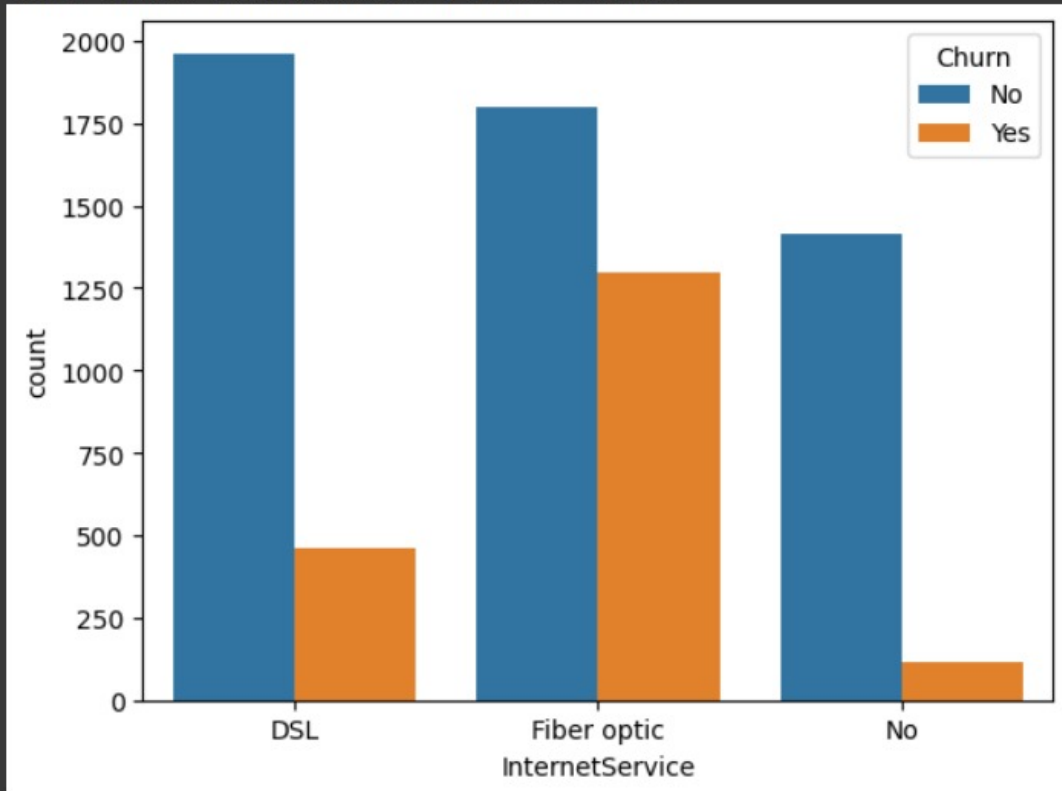
```
[9] numRetained = df[df.Churn == 'No'].shape[0]
    numChurned = df[df.Churn == 'Yes'].shape[0]
    print(numRetained/(numRetained + numChurned) * 100, '% of customers stayed in the company')
    print(numChurned/(numRetained + numChurned) * 100, '% of customers left with the company')
```

73.4630129206304 % of customers stayed in the company
26.536987079369588 % of customers left with the company



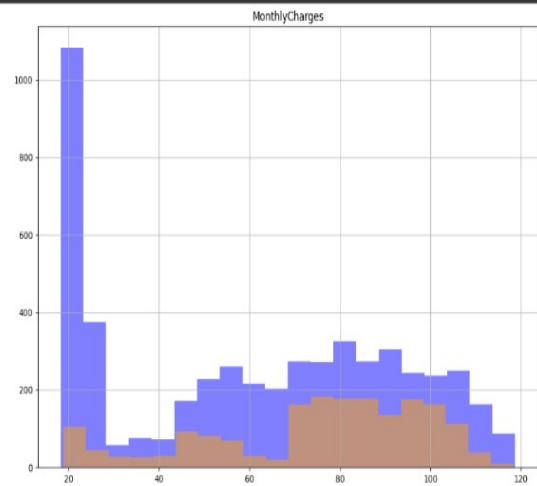
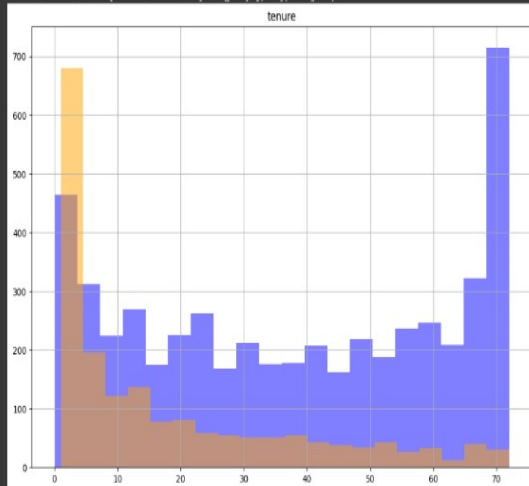
```
sns.countplot(x='InternetService', hue='Churn', data=df)
```

```
<Axes: xlabel='InternetService', ylabel='count'>
```



```
numericFeatures = ['tenure', 'MonthlyCharges']  
fig, ax = plt.subplots(1, 2, figsize=(20, 8))  
df[df.Churn == 'No'][numericFeatures].hist(bins=20, color='blue', alpha=0.5, ax=ax)  
df[df.Churn == 'Yes'][numericFeatures].hist(bins=20, color='orange', alpha=0.5, ax=ax)
```

```
array([<axes: title='center': 'tenure'>,  
       <axes: title='center': 'MonthlyCharges'>], dtype=object)
```



Conclusion

In conclusion, the customer churn prediction project has proven to be a pivotal tool for our organization in understanding and mitigating customer attrition. Through the diligent analysis of historical data and the implementation of advanced machine learning models, we have successfully developed a predictive system that can identify potential churners with remarkable accuracy. This not only provides us with invaluable insights into customer behavior but also empowers us to take proactive measures to retain valuable clientele and optimize customer relationships. The project's success underscores the importance of data-driven decision-making in today's competitive business landscape, and it has opened up new avenues for enhancing customer satisfaction and long-term profitability.

Furthermore, the project's impact extends beyond the realm of churn prediction. It has fostered a culture of data-driven innovation within our organization, highlighting the potential of leveraging AI and predictive analytics. Moving forward, the lessons learned from this project will continue to inform our strategic approach to customer management and retention. With an evolving dataset and ongoing refinement of our models, we are well-positioned to adapt to changing customer dynamics and to maintain our competitive edge in the market. The customer churn prediction project stands as a testament to the value of data science in modern business and its capacity to drive sustainable growth and customer-centric strategies.