SpeakX Assignment Task :- Data Science Role

Data set – Telecom Data

What is Customer churn?

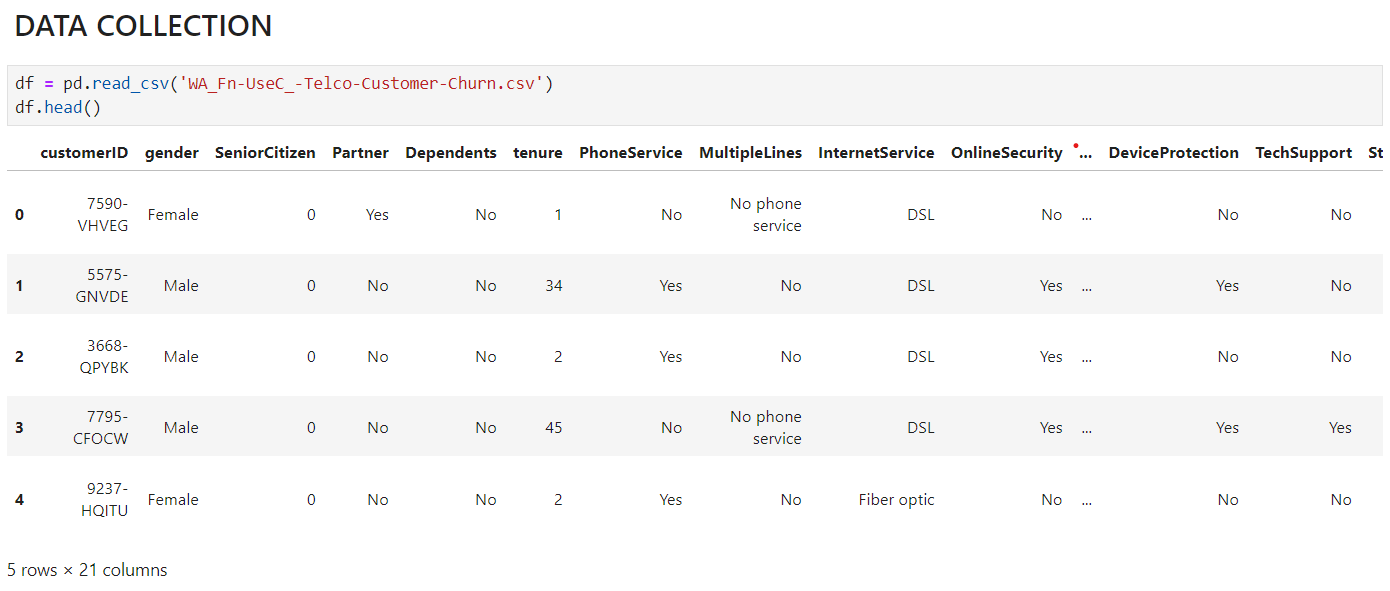
Customer churn in the telecom industry refers to the phenomenon where customers stop using a company's services and switch to a competitor or cease using telecom services altogether. This is a critical metric for telecom companies because retaining existing customers is generally more cost-effective than acquiring new ones. Churn rate, or the percentage of customers who leave over a given period, is a key indicator of a company's customer satisfaction and competitive standing.

**Causes of Customer Churn**

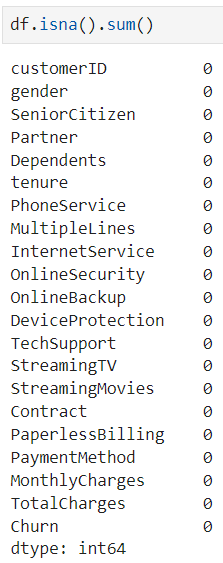
1. **Poor Customer Service**: Unresolved issues, long wait times, and unhelpful customer support can drive customers away.
2. **High Prices**: Customers might leave for more affordable options, especially if they perceive they are not getting good value for their money.
3. **Better Offers Elsewhere**: Competitors' promotions and better deals can attract customers.
4. **Network Quality**: Issues with coverage, call quality, and data speeds can lead to dissatisfaction.
5. **Billing Issues**: Complicated or incorrect billing can frustrate customers.
6. **Lack of Personalization**: Services that do not meet individual needs can prompt customers to find providers that offer more tailored solutions.

Steps involved in Project:

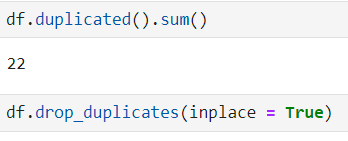
1. Data Collection



1. Data Cleaning

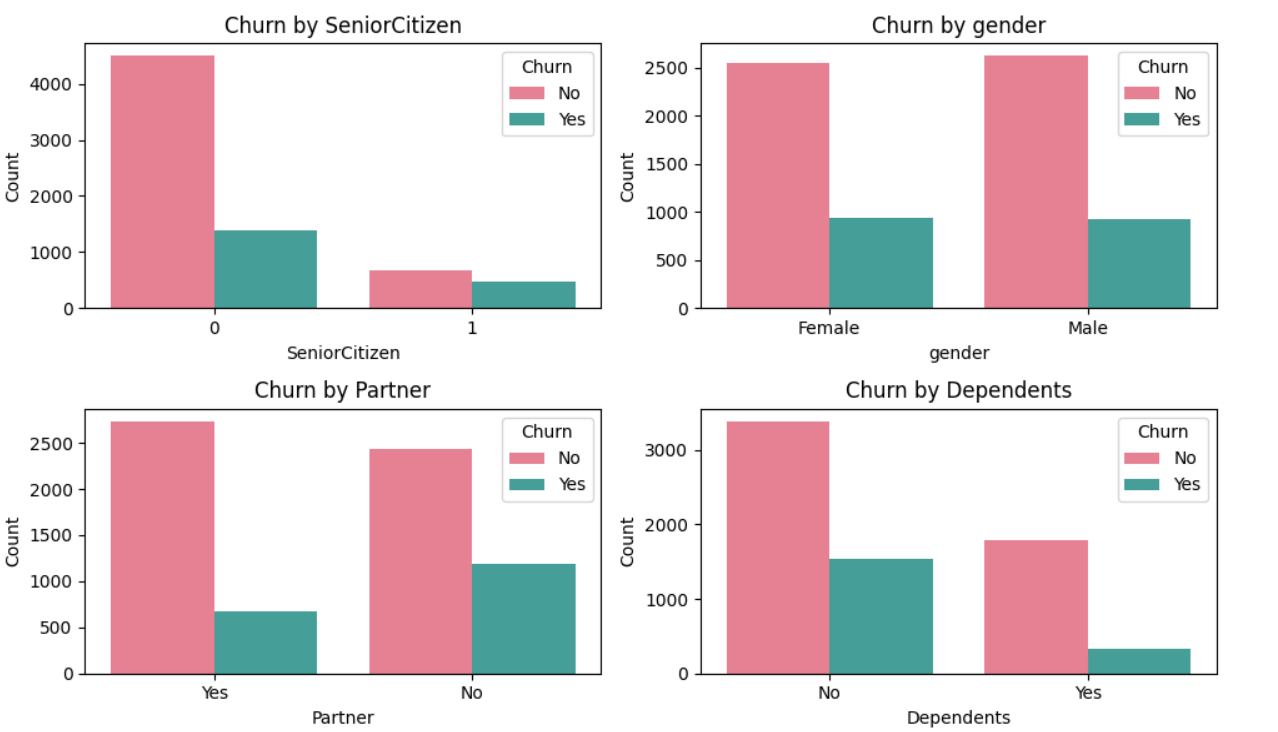
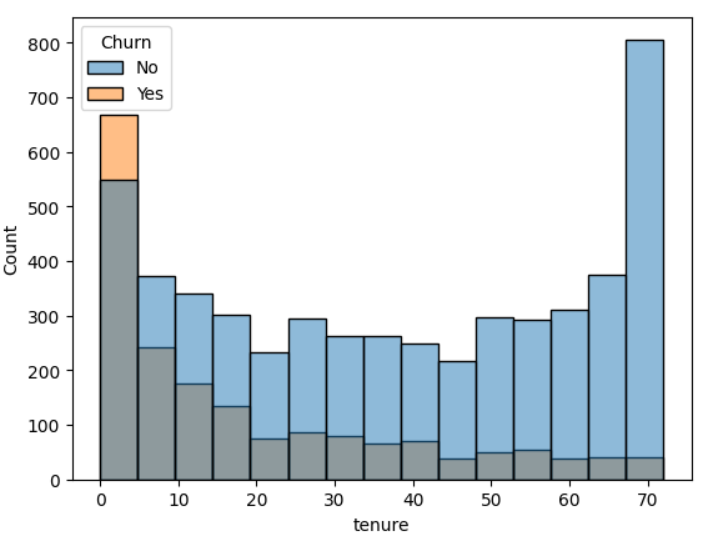
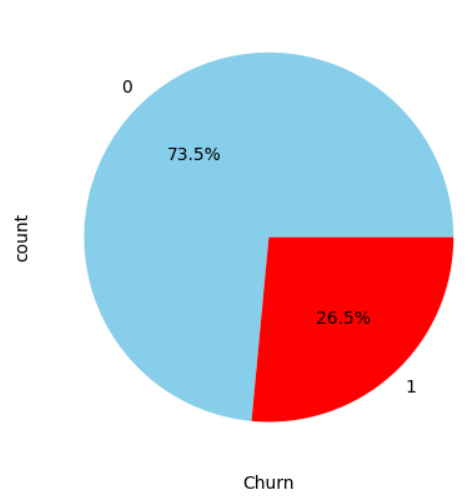


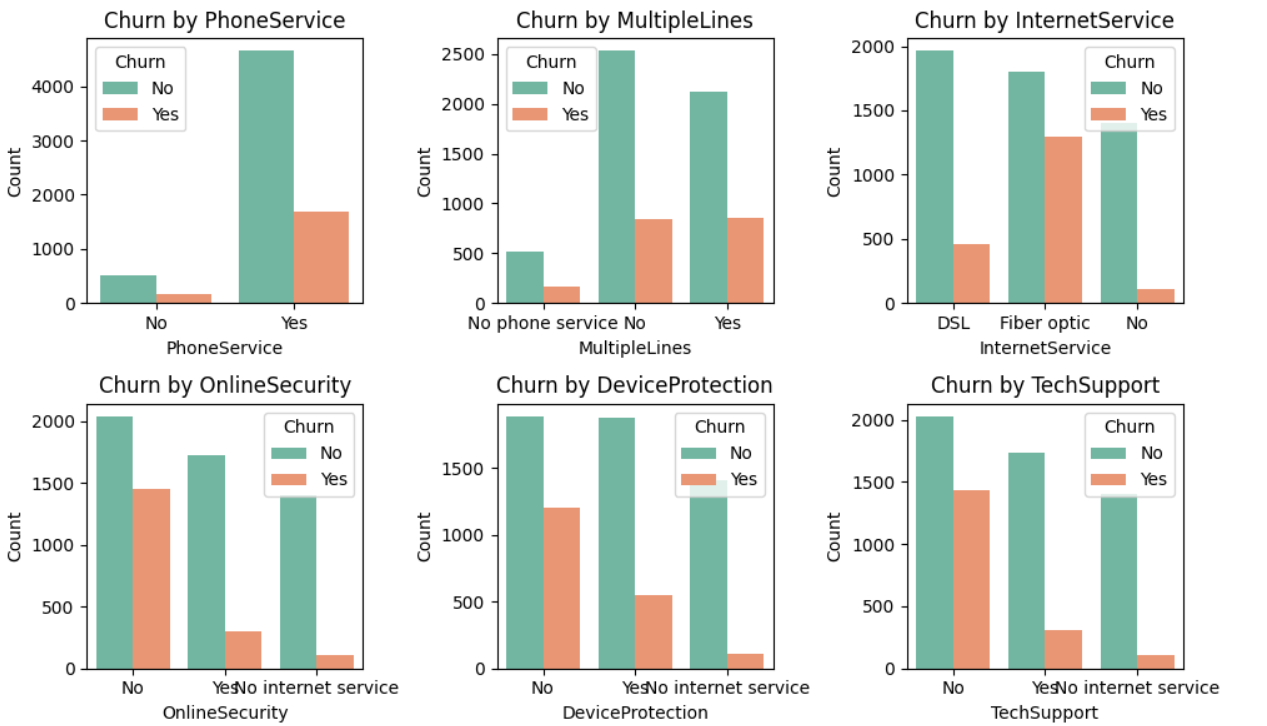
Checking for Null Values

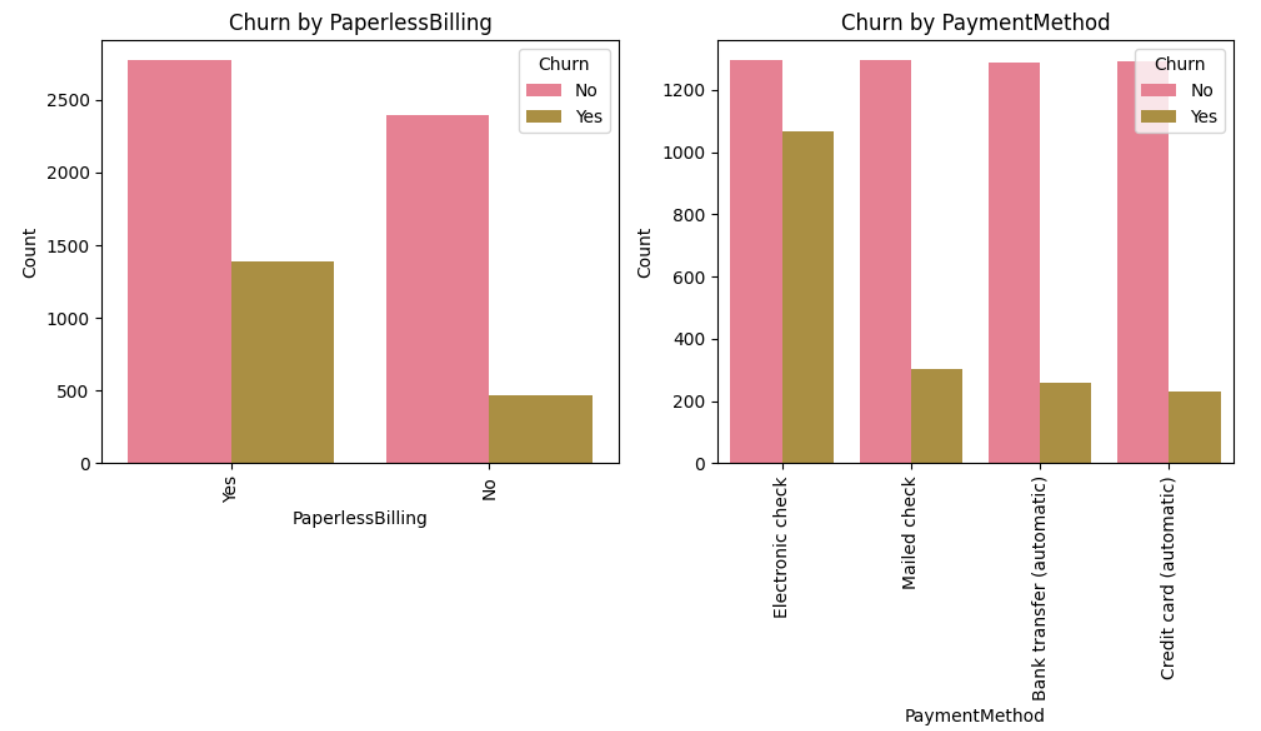


Checking for duplicated values and dropping them

1. EDA
   1. Data Visualization

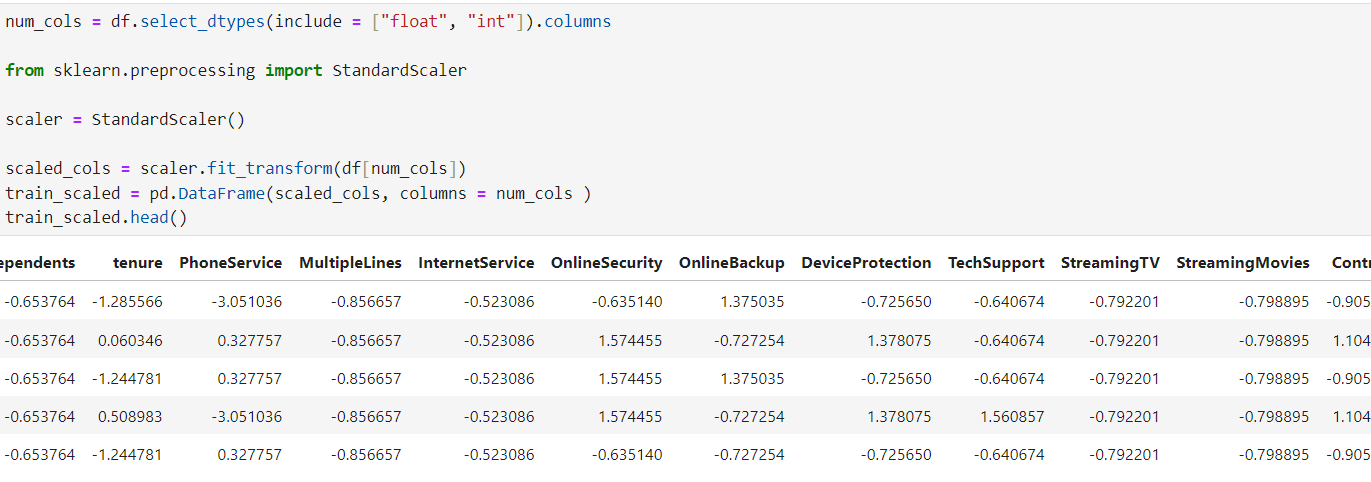




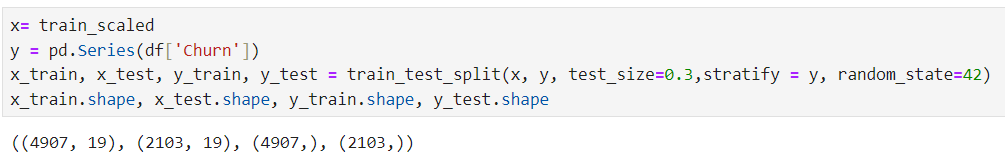


* 1. Data Preprocessing and Feature Engneering

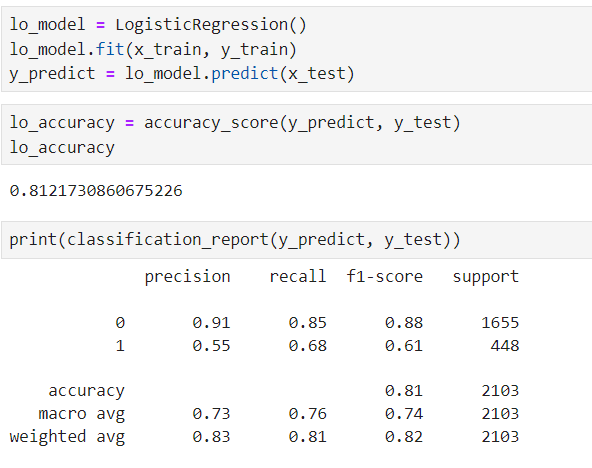
1. Dropping CustomerID column which doesnot has importance in model building and Analysis
2. Dealing with Total Charges Column
3. Mapping Category Columns with integers like 0,1,2 etc
4. Detecting Outliers
5. Converting Datatypes of Category to int for model Building
6. Scaling the Data with Standard Scaler

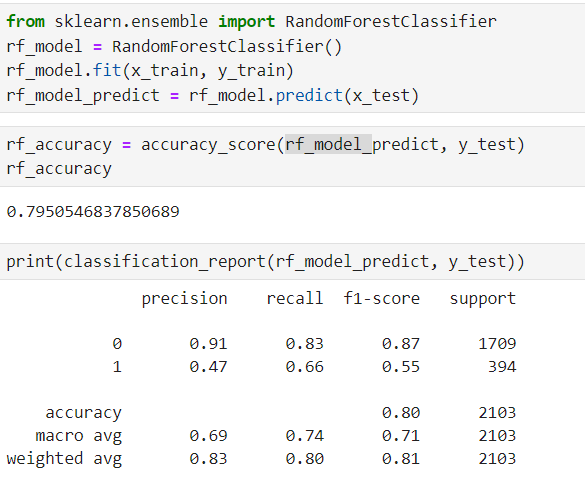


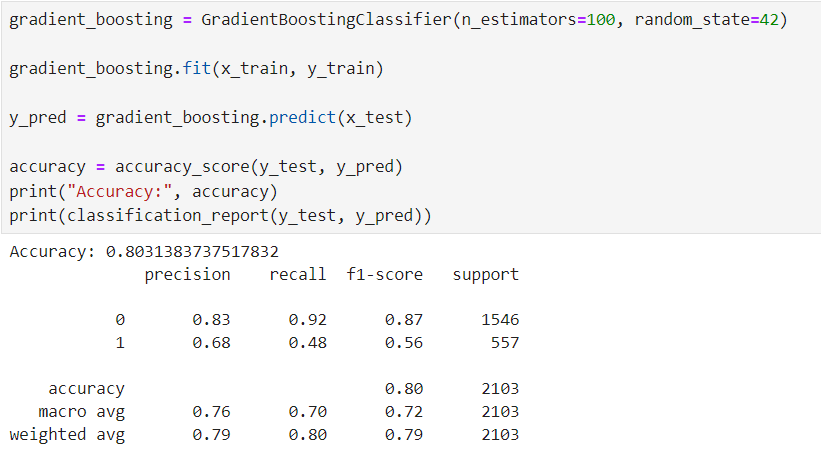
* 1. Model Building
     1. Splitting the Data



* + 1. Building Classification Algorithms like Logistic Regression, Random Forest and Gradient Boosting
    2. Testing models using metrics like accuracy, classification report







Challenges Faced During the Project:

Incomplete Data: Missing values in critical fields

Feature Selection: Determining which variables (e.g., service usage patterns, customer demographics, billing information) are most predictive of churn can be challenging.

Model Complexity: Choosing the right model (e.g., logistic regression, decision trees, random forests, gradient boosting, neural networks) involves balancing complexity, interpretability, and performance

Insights from Project:

Gender: Analyzing churn rates across different genders can reveal if there's a significant difference in churn behavior between male and female customers.

SeniorCitizen: Senior citizens might have different service needs and preferences. Higher churn rates among senior citizens could indicate a need for more tailored services or support for this demographic.

Partner and Dependents: Customers with partners and dependents might prefer more stable services. Higher churn rates among single customers or those without dependents might suggest the need for different engagement strategies.

Tenure: Customers with shorter tenure might be more likely to churn, indicating the importance of engagement and satisfaction efforts early in the customer lifecycle.

PhoneService and MultipleLines: Analyzing these can reveal if customers with more comprehensive phone services are more or less likely to churn.

InternetService: The type of internet service (DSL, Fiber, etc.) can significantly impact churn. For instance, customers using slower internet services might be more likely to switch providers.

OnlineSecurity, DeviceProtection, TechSupport: Customers utilizing these additional services might show different churn patterns. Lower churn rates among these customers might highlight the value of bundled services.

StreamingTV and StreamingMovies: These services might be crucial for customer satisfaction. Churn analysis can show if customers who use streaming services are more loyal.

Contract: The type of contract (monthly, yearly, etc.) often influences churn. Customers on longer contracts might be less likely to churn due to contractual obligations.

PaperlessBilling: Preferences for paperless billing might correlate with certain demographics or tech-savviness, impacting churn rates.

PaymentMethod: Different payment methods (credit card, electronic check, etc.) can show varying churn patterns. For example, customers using electronic checks might have higher churn rates due to perceived convenience or lack of flexibility.

MonthlyCharges and TotalCharges: Higher monthly charges might correlate with higher churn if customers feel they are not getting value for their money. Similarly, the total amount spent can influence churn, especially if the perceived value diminishes over time.

Churn: This column is the key outcome variable. Analyzing its relationship with the other variables helps identify the main drivers of customer churn.