**EDA:**

**This part was done by :Omar Garawani**

Exploratory Data Analysis Summary:

Based on the visualizations, here are some key findings related to customer churn:

* **Gender:** Churn rate is similar for both male and female customers.
* **Senior Citizens:** Senior citizens have a higher churn rate compared to non-senior citizens, particularly in the initial months of their service.
* **Partners and Dependents:** Customers without partners and dependents have higher churn rates.
* **Tenure:** Churn rate is highest in the first year of service and decreases as tenure increases.
* **Monthly Charges:** Customers with longer tenure tend to have higher monthly charges. There are instances where customers with short tenure have high monthly charges, potentially due to subscribing to premium service bundles early on.
* **Total Charges:** Total charges generally increase with tenure for both churned and non-churned customers, but at a faster rate for non-churned customers.
* **Monthly/Total Charges Ratio:** Churned customers tend to have a higher Monthly/Total Charges Ratio, especially at lower tenures.
* **Additional Services:** Customers with Online Security and Tech Support tend to have lower churn rates. Interestingly, churned customers who do not have additional services can have similar monthly charges to non-churned customers who *do* have additional services, suggesting that perceived value for money is important.
* **Internet Service:** Customers with Fiber Optic internet service have the highest churn rate. Customers without internet service have a very low churn rate.
* **Phone Service:** Customers with phone service have a slightly higher churn rate than those without phone service.
* **Payment Method:** Customers using Electronic check as their payment method have a higher churn rate compared to other payment methods. For churned customers, those using Electronic check tend to have lower total charges at any given tenure, suggesting they might be churning earlier.
* **Contract Type:** Customers with month-to-month contracts have a significantly higher churn rate compared to customers with one-year or two-year contracts.

These findings suggest that customer churn is influenced by factors such as demographics (seniority, partnership, dependents), length of service, the type of services subscribed to, contract type, and payment method. Further analysis and modeling can be done to identify the most significant predictors of churn and develop targeted retention strategies.

**Feature building:**

**This was done by :**  **Fares Elkhayat**

* **Feature Engineering Report – Telco Customer Churn Dataset**
* **1. Data Cleaning / Preprocessing**

| * **Step** | * **Column(s)** | * **Reason** | * **Result** |
| --- | --- | --- | --- |
| * **Drop unnecessary columns** | * **customerID** | * **Unique identifier, not useful for the model** | * **Column removed** |
| * **Handle missing values** | * **TotalCharges** | * **Some values are non-numeric** | * **Converted to numeric (pd.to\_numeric) and filled missing values with 0** |
| * **Remove duplicate rows** | * **All columns** | * **Ensure unique data for the model** | * **All duplicate rows removed** |

* **2. Binary Features Encoding**

| * **Column** | * **Original Values** | * **Transformation** | * **Reason** |
| --- | --- | --- | --- |
| * **gender** | * **Male/Female** | * **Male → 1, Female → 0** | * **Convert text to numeric for model use** |
| * **Partner** | * **Yes/No** | * **Yes → 1, No → 0** | * **Same reason** |
| * **Dependents** | * **Yes/No** | * **Yes → 1, No → 0** | * **Same reason** |
| * **PhoneService** | * **Yes/No** | * **Yes → 1, No → 0** | * **Same reason** |
| * **PaperlessBilling** | * **Yes/No** | * **Yes → 1, No → 0** | * **Same reason** |
| * **Churn** | * **Yes/No** | * **Yes → 1, No → 0** | * **Convert target to numeric** |

* **Note: Binary features don’t require One-Hot Encoding; 0/1 is sufficient.**
* **3. Multi-class Categorical Features Encoding**
* **Step 1: Replace special values**
* **Some columns contain values like 'No internet service' or 'No phone service'.**
* **Replaced these with 'No' in the following columns:  
  MultipleLines, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies**
* **Step 2: Label Encoding**

| * **Column** | * **Type** | * **Transformation** |
| --- | --- | --- |
| * **MultipleLines** | * **Yes/No/No phone service** | * **LabelEncoder → 0,1,2** |
| * **InternetService** | * **DSL/Fiber optic/No** | * **LabelEncoder → 0,1,2** |
| * **OnlineSecurity** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **OnlineBackup** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **DeviceProtection** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **TechSupport** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **StreamingTV** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **StreamingMovies** | * **Yes/No** | * **LabelEncoder → 0,1** |
| * **Contract** | * **Month-to-month/One year/Two year** | * **LabelEncoder → 0,1,2** |
| * **PaymentMethod** | * **Electronic check/Mailed check/Bank transfer/Credit card** | * **LabelEncoder → 0,1,2,3** |

* **4. Numeric Features Handling**

| * **Column** | * **Transformation / Action** | * **Reason** |
| --- | --- | --- |
| * **TotalCharges** | * **Convert to numeric, fill missing values with 0** | * **Ready for model input** |
| * **MonthlyCharges, tenure** | * **Analyze skewness** | * **Identify columns that may need transformation/scaling** |

* **5. Summary of Feature Engineering**
* **Data is now ready for model training:**
* **Binary columns → 0/1**
* **Multi-class categorical columns → numeric values 0,1,2…**
* **Numeric columns → clean and processed, missing values handled**
* **Each step, including cleaning, binary mapping, label encoding, and missing value handling, is part of Feature Engineering.**
* **Next steps could include scaling/normalization for numeric features depending on the model requirements.**