

# Telco Customer Churn Data Analysis and Visualization Assessment

## Assignment

You have been provided with the Telco Customer Churn dataset, which includes detailed information on customer demographics, account details, subscribed services, and churn behaviour. Your task is to leverage your R skills to transform, analyse, and visualise this data, generating actionable insights. Synthesize your findings into a concise report to communicate key patterns, trends, and recommendations.

### Dataset Overview

The Telco Customer Churn dataset provides comprehensive details about customers, including their demographics, account information, service usage, and whether they have churned. Key columns include:

- **customerID**: Unique identifier for each customer.
- **gender**: Customer gender ('Male' or 'Female').
- **SeniorCitizen**: Indicator if the customer is a senior (1 for Yes, 0 for No).
- **Partner**: Whether the customer has a partner ('Yes' or 'No').
- **Dependents**: Whether the customer has dependents ('Yes' or 'No').
- **tenure**: Number of months the customer has stayed with the company.
- **PhoneService**: Indicates if the customer has a phone service ('Yes' or 'No').
- **MultipleLines**: Indicates if the customer has multiple phone lines ('Yes', 'No', or 'No phone service').
- **InternetService**: Type of internet service ('DSL', 'Fiber optic', or 'No').

- **OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies:** Service-specific columns with values ('Yes', 'No', or 'No internet service').
- **Contract:** Customer's contract type ('Month-to-month', 'One year', or 'Two year').
- **PaperlessBilling:** Whether the customer uses paperless billing ('Yes' or 'No').
- **PaymentMethod:** Customer's payment method (e.g., 'Electronic check', 'Mailed check', 'Bank transfer (automatic)', 'Credit card (automatic)').
- **MonthlyCharges:** Amount charged to the customer monthly.
- **TotalCharges:** Total amount charged to the customer.
- **Churn:** Indicates whether the customer has churned ('Yes' or 'No').

## Tasks

### 1. Data Manipulation and Transformation

#### a. Data Import:

- Locate the `telco-customer-churn.csv` dataset at <https://bit.ly/telco-customer-churn> and import it into R.

#### b. Variable Transformation:

- Transform the `Churn` column into a binary format (e.g., 1 for churned, 0 for not churned).
- Recode the `SeniorCitizen` variable into a more descriptive format (e.g., "Yes" for 1 and "No" for 0).
- Create a new variable, such as `AvgChargePerMonth`, calculated by dividing `TotalCharges` by `tenure` (ensuring that cases where `tenure` is 0 are handled appropriately).
- Optionally, develop another metric (e.g., a `ServiceCount` that aggregates the number of additional services to which a customer subscribes).

### 2. Handling Missing and Inconsistent Values

#### a. Identify Issues:

- Scan the dataset for missing or inconsistent values, and document which columns are affected.

#### b. Data Quality Improvement:

- Apply strategies to address any data quality issues (for example, convert data types if necessary, handle missing values, and ensure consistency across columns).

### 3. Analysis and Insights

#### a. Overall Churn Patterns:

- Determine the overall churn rate in the dataset.

#### b. Segmented Analysis:

- Calculate the percentage of churned customers across different segments such as:
  - **Gender:** What proportion of male vs. female customers churn?
  - **Contract Type:** How does the churn rate vary across different contract types?
  - **Internet Service:** What are the churn rates for customers with DSL, Fiber optic, or no internet service?
- For each segment (or combination of segments), compute summary statistics (e.g., counts, averages, medians) for key metrics like `MonthlyCharges` and `tenure`.

#### c. Advanced Aggregation:

- For each gender, determine summary statistics (mean, median, maximum) for monthly charges.
- Identify which customer segment or service bundle is associated with the highest churn rate.

### 4. Data Visualization

#### a. Exploratory Visualizations:

- Create a histogram or density plot to visualize the distribution of customer `tenure`.
- Develop a bar chart that shows the counts of churned and non-churned customers.

#### b. Comparative Visualizations:

- Construct a boxplot to compare `MonthlyCharges` across different `Contract` types.
- Generate a scatter plot displaying the relationship between `tenure` and `MonthlyCharges`, with points colored by churn status. Consider adding a trend line if it enhances interpretation.

#### c. Combined Analysis:

- Filter the dataset to focus on a specific segment (for example, only customers with 'Fiber optic' service) and create additional visualizations (such as a histogram of their `tenure` or a scatter plot of their charges vs. `tenure`).

- 
- For each unique `tenure` value, compute the percentage of customers who churned, and plot these percentages as a line graph.

## **Deliverables**

- **Code:**

Provide your Quarto file with clear, commented code showing your data manipulation, analysis, and visualization steps.

- **Report:**

Write a concise summary that explains:

- Your approach to data cleaning and transformation.
- The key findings from your analysis.
- Insights derived from your visualizations.
- Any recommendations or follow-up questions that your analysis suggests.