SUMMARY AND RECOMMENDATION

📌 Project Title: Telco Customer Churn Analysis

🧠 Objective

The primary aim of this project is to perform **Exploratory Data Analysis (EDA)** on a customer churn dataset sourced from a telecommunications company. The key goals include:

- Identifying **factors** that significantly influence customer churn — the act of customers discontinuing their subscription to the service.
- Understanding the **demographics, behavioral trends, and service usage** patterns of customers who tend to churn.
- Generating **actionable business insights** through data visualizations that can inform customer retention strategies and service improvements.

By uncovering the hidden patterns in the dataset, this project provides a data-driven foundation for making **targeted business decisions** and improving customer satisfaction.

📥 Data Overview

- **Dataset Used**: `Telco_Customer_Churn.csv`
- **Source**: Public dataset from Kaggle
- **Total Records**: Approx. 7,000+ customer entries.
- **Target Column**: `Churn` — a binary categorical variable with two values:
    - `"Yes"`: Customer **left** the service.
    - `"No"`: Customer **remained** subscribed.

This dataset includes a **wide range of features**, such as:

- **Customer Demographics**: Gender, SeniorCitizen, Partner, Dependents.
- **Account Information**: Tenure, Contract type, Monthly and Total Charges.
- **Service Usage**: Internet Service, Streaming TV, Tech Support, etc.
- **Payment Details**: Payment method, Paperless billing, etc.

🧹 Data Cleaning & Preprocessing

✅ 1. Handling Missing/Blank Values:

- The column `TotalCharges` contained **blank entries** represented as `" "`.
- These entries were first **replaced with `0`**, and the column was **converted to `float`** type for accurate numeric operations.
- This ensured that computations like mean, sum, and plotting would not result in errors.

✅ 2. Data Type Verification:

- The `df.info()` function was used to inspect the dataset's structure.
- Necessary conversions (like string to float) were performed to ensure **type consistency** across features.

✅ 3. Checking for Missing & Duplicate Values:

- After cleaning:
  - **No missing values** were found in the dataset.
  - **No duplicate rows** were identified (`df.duplicated().sum()` returned `0`).

✅ 4. Feature Transformation:

- The `SeniorCitizen` column initially had **binary numeric values** (0 and 1).
- A custom function was applied to make it **more human-readable**, mapping:
  - `0 → "No"`
  - `1 → "Yes"`

This enhances interpretability during visualization and improves communication of insights.

📊 Exploratory Data Analysis (EDA)

📍 1. Overall Churn Count and Percentage

- **Count Plot**: Displayed how many customers churned (`Yes`) vs. stayed (`No`).
- **Pie Chart**: Visualized the proportion of churned customers in the entire dataset.

Insight:

- A significant percentage of customers had churned, signaling a potential **customer retention problem** for the telecom provider.

📍 2. Churn by Gender

- A grouped count plot was used (`hue="Churn"`) to compare male vs female churn behavior.

Insight:

- Churn was **almost evenly split between genders**, indicating that **gender is not a major driver** of churn.

📍 3. Churn by Senior Citizen Status

- Bar plots and **stacked bar charts with percentage labels** were created using `pd.crosstab()`.

Insight:

- **Senior citizens** had a **higher churn rate** than non-seniors.
- Possible reasons include:
  - Less digital engagement
  - Higher price sensitivity
  - Lack of comfort with new technology or services

🔍 Planned/Additional EDA Sections (Based on Notebook Headings)

Although code may not have been fully shown, markdown suggests further in-depth analysis is conducted or planned:

📍 4. Churn by Contract Type

- **Hypothesis**: Long-term contracts lead to **lower churn**.
- Customers with:
  - **Month-to-month plans** tend to churn more due to **flexibility and low commitment**.
  - **1-year or 2-year plans** are likely to stay longer due to **contractual lock-in**.

Expected Visualization: Grouped bar plots comparing churn across contract types.

📍 5. Churn vs Service Usage

- Investigates churn behavior with respect to usage of services such as:
  - `InternetService`
  - `StreamingTV`
  - `StreamingMovies`
  - `OnlineBackup`

Expected Observations:

- Customers who use more **value-added services** might be **less likely to churn**.
- Conversely, customers with **"No internet service"** often show **higher churn rates**.

📍 6. Churn by Payment Method

- Analyzed payment methods:
  - `Electronic check`, `Mailed check`, `Bank transfer`, `Credit card`.

Key Trend:

- Customers using **Electronic checks** tend to **churn more**.
- Those using **automatic recurring payments** (Credit card, Bank transfer) show **greater loyalty**.

Possible Reason: Auto-pay customers experience **less friction**, improving retention.

🛠️ Tools and Libraries Used

Library	Purpose
Pandas	Data manipulation, filtering, aggregation
NumPy	Numerical operations (e.g., type conversion)
Matplotlib	Basic plotting (e.g., bar, pie charts)
Seaborn	Advanced plots (e.g., countplot, hue-based)
Custom Functions	For transformation (e.g., encoding binary values)

🧾 Key Takeaways

- **Customer churn** is a significant concern that must be addressed for long-term business stability.
- **Senior citizens** and those on **month-to-month contracts** represent **high-risk churn segments**.
- **Service engagement** (e.g., subscribing to backup, streaming, or tech support services) is often **associated with reduced churn**.
- **Payment methods** matter: customers using **electronic checks** are more likely to churn.
- **Data cleaning and transformation** steps are critical in ensuring accurate insights.
- Well-designed **visualizations** help translate raw data into **meaningful business actions**.