

## Key Attributes of the Dataset

- **Rows (Records):** Each row represents a unique customer.
- **Columns (Features):** The dataset has 21 columns, covering customer demographics, account details, service usage, and the target label, **Churn**.

### 1. Customer Demographic Information

- **customerID:** A unique identifier for each customer. While useful for tracking individual customers, it's irrelevant for predictive modeling, so we usually drop it.
- **gender:** The customer's gender (Male/Female).
- **SeniorCitizen:** Indicates whether the customer is a senior citizen (1 for Yes, 0 for No). This feature may correlate with churn as older customers might have different usage patterns or loyalty tendencies.
- **Partner:** Whether the customer has a partner (Yes/No).
- **Dependents:** Whether the customer has dependents (Yes/No). This information can provide insights into customers' financial situations or service usage preferences.

### 2. Customer Account Information

- **tenure:** The number of months the customer has been with the company. This is often one of the most important predictors, as longer-tenure customers are generally less likely to churn.
- **PhoneService:** Whether the customer has a phone service (Yes/No).
- **MultipleLines:** If the customer has multiple lines (No phone service/No/Yes).
- **InternetService:** The type of internet service the customer subscribes to (DSL, Fiber optic, or No internet).
- **Contract:** The type of contract the customer has chosen (Month-to-month, One year, or Two year). Customers on longer contracts may have lower churn rates.
- **PaperlessBilling:** Whether the customer is billed electronically (Yes/No).
- **PaymentMethod:** The payment method used by the customer (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)). Payment methods like automatic payments can be indicative of a lower churn rate.
- **MonthlyCharges:** The amount the customer is charged per month. High charges can sometimes be associated with higher churn rates, as customers may feel their services are too expensive.
- **TotalCharges:** The total amount charged to the customer over the entire tenure. This feature often needs preprocessing, as some values may be non-numeric or missing.

### 3. Customer Services Information

- **OnlineSecurity:** Whether the customer has online security add-on (No internet service/No/Yes).
- **OnlineBackup:** Whether the customer has online backup add-on (No internet service/No/Yes).
- **DeviceProtection:** Whether the customer has device protection add-on (No internet service/No/Yes).
- **TechSupport:** Whether the customer has tech support add-on (No internet service/No/Yes).
- **StreamingTV:** Whether the customer has a TV streaming service (No internet service/No/Yes).
- **StreamingMovies:** Whether the customer has a movie streaming service (No internet service/No/Yes).

These service-related features indicate the types of additional services customers subscribe to, which can provide insights into customer satisfaction and the likelihood of retention.

### 4. Target Variable

- **Churn:** This is the target variable for prediction, indicating whether the customer has left the company (Yes/No). In a predictive model, this variable would be converted to binary, where 1 represents churn and 0 represents no churn.