In [1]:

```
import pandas as pd
import numpy as np
```

In [7]:

```
np.random.seed(42)

# Defining the size
n=5000

# CustomerID
customer_ids = np.arange(1, n + 1)

# Age: Random ages between 18 and 90 with some missing values
age = np.random.randint(18, 90, n).astype(float)
age[np.random.choice(n, 100, replace=False)] = np.nan

# Gender: Random selection with some missing values
gender = np.random.choice(['Male', 'Female'], n)
gender[np.random.choice(n, 50, replace=False)] = np.nan

# ContractType: Random selection from three options
contract_type = np.random.choice(['Month-to-month', 'One year', 'Two year'], r

# MonthlyCharges: Random charges between $20 and $120 with some outliers
monthly_charges = np.random.uniform(20, 120, n).astype(float)
monthly_charges[np.random.choice(n, 20, replace=False)] = np.nan
monthly_charges[np.random.choice(n, 10, replace=False)] = np.random.uniform(20

# Tenure: Random number of months between 0 and 72
tenure = np.random.randint(0, 72, n)

# TotalCharges: MonthlyCharges multiplied by tenure with some inconsistencies
total_charges = (monthly_charges * tenure).astype(float)
total_charges[np.random.choice(n, 30, replace=False)] = np.nan
inconsistent_indices = np.random.choice(n, 10, replace=False)
total_charges[inconsistent_indices] = total_charges[inconsistent_indices] * np

# TechSupport: Random selection with some missing values
tech_support = np.random.choice(['Yes', 'No'], n)
tech_support[np.random.choice(n, 60, replace=False)] = np.nan

# InternetService: Random selection with some missing values
internet_service = np.random.choice(['DSL', 'Fiber optic', 'No'], n)
internet_service[np.random.choice(n, 40, replace=False)] = np.nan

# PaperlessBilling: Random selection with some missing values
paperless_billing = np.random.choice(['Yes', 'No'], n)
paperless_billing[np.random.choice(n, 30, replace=False)] = np.nan

# PaymentMethod: Random selection from four options with some missing values
payment_method = np.random.choice(['Electronic check', 'Mailed check', 'Bank t
payment_method[np.random.choice(n, 30, replace=False)] = np.nan

# Churn: Random selection with some missing values
churn = np.random.choice(['Yes', 'No'], n)
churn[np.random.choice(n, 50, replace=False)] = np.nan

# Creating the DataFrame
data = pd.DataFrame({
    'CustomerID': customer_ids,
    'Age': age,
    'Gender': gender,
    'ContractType': contract_type,
    'MonthlyCharges': monthly_charges,
    'TotalCharges': total_charges,
    'TechSupport': tech_support,
    'InternetService': internet_service,
    'Tenure': tenure,
    'PaperlessBilling': paperless_billing,
    'PaymentMethod': payment_method,
    'Churn': churn
})

# Introducing some duplicates
duplicates = data.sample(20, replace=True)
data = pd.concat([data, duplicates], ignore_index=True)

# Shuffling the dataset
data = data.sample(frac=1).reset_index(drop=True)

data.head(10)  # Display the first 10 rows to preview

data.to_csv('Customer.csv')
```