Bradley Reardon

Time Series Forecasting and Analysis

HW9

```
    customerID  gender  SeniorCitizen  ...  MonthlyCharges  TotalCharges  Churn
0   7590-VHVEG  Female              0  ...           29.85         29.85     No
1   5575-GNVDE    Male              0  ...           56.95        1889.5     No
2   3668-QPYBK    Male              0  ...           53.85        108.15    Yes
3   7795-CFOCW    Male              0  ...           42.30       1840.75     No
4   9237-HQITU  Female              0  ...           70.70        151.65    Yes

[5 rows x 21 columns]
       SeniorCitizen        tenure  MonthlyCharges
count    7043.000000   7043.000000     7043.000000
mean        0.162147     32.371149       64.761692
std         0.368612     24.559481       30.090047
min         0.000000      0.000000       18.250000
25%         0.000000      9.000000       35.500000
50%         0.000000     29.000000       70.350000
75%         0.000000     55.000000       89.850000
max         1.000000     72.000000      118.750000
```

```
customerID             object
gender                 object
SeniorCitizen           int64
Partner                object
Dependents             object
tenure                  int64
PhoneService           object
MultipleLines          object
InternetService        object
OnlineSecurity         object
OnlineBackup           object
DeviceProtection       object
TechSupport            object
StreamingTV            object
StreamingMovies        object
Contract               object
PaperlessBilling       object
PaymentMethod          object
MonthlyCharges        float64
TotalCharges          float64
Churn                   int64
dtype: object
customerID              0
gender                  0
SeniorCitizen           0
Partner                 0
Dependents              0
tenure                  0
PhoneService            0
MultipleLines           0
InternetService         0
OnlineSecurity          0
OnlineBackup            0
DeviceProtection        0
TechSupport             0
StreamingTV             0
StreamingMovies         0
Contract                0
PaperlessBilling        0
PaymentMethod           0
MonthlyCharges          0
TotalCharges           11
Churn                   0
dtype: int64
```
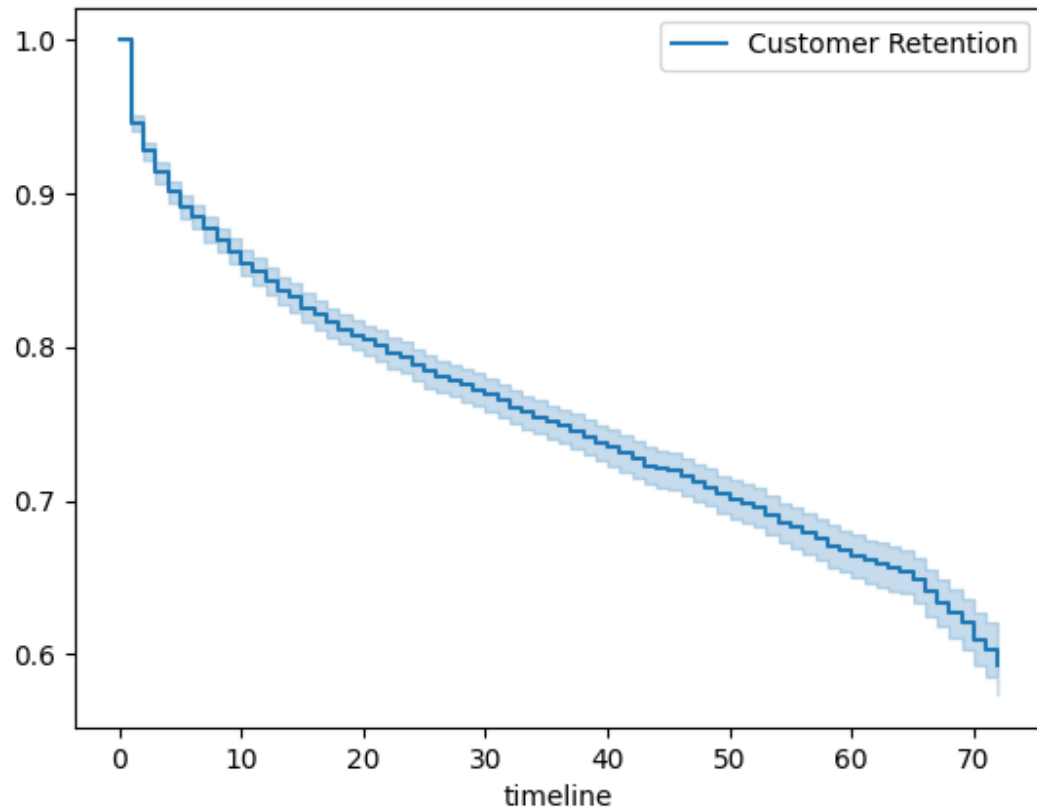
```python
from lifelines import KaplanMeierFitter
import pandas as pd
import matplotlib.pyplot as plt

# q 1
df = pd.read_csv(r'C:\Users\brear\OneDrive\Desktop\Grad School\Time-Series-
Analysis-and-Moldeing\Datasets\WA_Fn-UseC_-Telco-Customer-Churn.csv')
print(df.head())
print(df.describe())

# q 2
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'],errors='coerce')

# q 3
df['Churn'] = df['Churn'].apply(lambda x: 1 if x == 'Yes' else 0)

# q 4
print(df.dtypes)
features = df.columns
for feature in features:
    if df[feature].dtype == 'object':
        df[feature].fillna(value=df[f'{feature}'].mode(), inplace=True)
    elif df[feature].dtype == 'int64' or df[feature].dtype == 'float64' or
df[feature].dtype == 'numeric':
        df[feature].fillna(value=df[f'{feature}'].mode(), inplace=True)
print(df.isna().sum())
```

```python
# q 5 - 8
durations = df['tenure']
event_observed = df['Churn']
ax = plt.subplot(111)
kmf = KaplanMeierFitter()
kmf.fit(durations, event_observed,label='Customer Retention')
kmf.plot_survival_function(ax=ax)
plt.show()
```