Data science homework 3 - Mayank Sharma - ms14662

```
import pandas as pd
import matplotlib.pyplot as plt

url = "https://data.cityofnewyork.us/api/views/6fi9-q3ta/rows.csv?accessType=DOWNLOAD"
df = pd.read_csv(url)

df['hour_beginning'] = pd.to_datetime(df['hour_beginning'])
df['Day_of_Week'] = df['hour_beginning'].dt.day_name()
df['Year'] = df['hour_beginning'].dt.year
df['Hour'] = df['hour_beginning'].dt.hour

weekdays = df[df['Day_of_Week'].isin(['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday'])]
weekday_counts = weekdays.groupby('Day_of_Week')['Pedestrians'].mean()
weekday_counts = weekday_counts.reindex(['Monday', 'Tuesday', 'Wednesday', 'Thursday',
'Friday'])
plt.plot(weekday_counts.index, weekday_counts.values, marker='o', linestyle='-', color='b')
plt.show()

df_2019 = df[(df['Year'] == 2019) & (df['location'] == 'Brooklyn Bridge')]
df_encoded = pd.get_dummies(df_2019, columns=['weather_summary'], drop_first=True)
correlation_matrix = df_encoded.corr()
plt.imshow(correlation_matrix, cmap='coolwarm', interpolation='nearest')
plt.colorbar()
plt.show()

def categorize_time(hour):
    if 5 <= hour < 12:
        return 'Morning'
    elif 12 <= hour < 17:
        return 'Afternoon'
    elif 17 <= hour < 21:
        return 'Evening'
    else:
        return 'Night'

df['Time_of_Day'] = df['Hour'].apply(categorize_time)
time_of_day_counts = df.groupby('Time_of_Day')['Pedestrians'].sum()
time_of_day_counts = time_of_day_counts.reindex(['Morning', 'Afternoon', 'Evening', 'Night'])
plt.bar(time_of_day_counts.index, time_of_day_counts.values, color='skyblue')
plt.show()
```