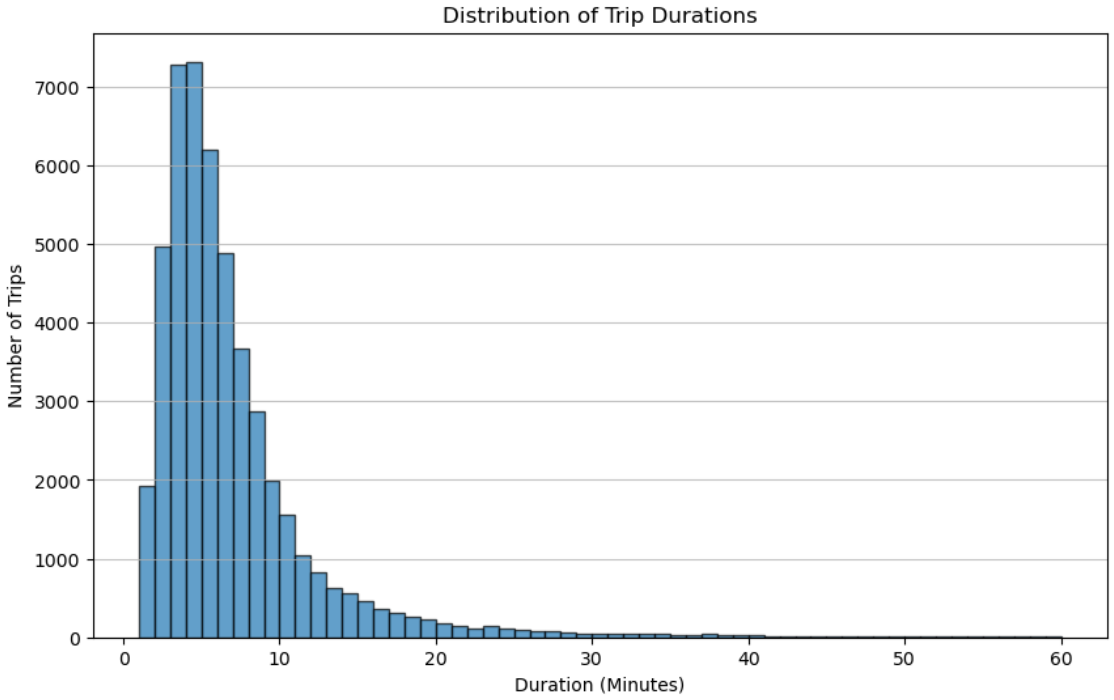
Urban_Mobility_Project

April 7, 2024

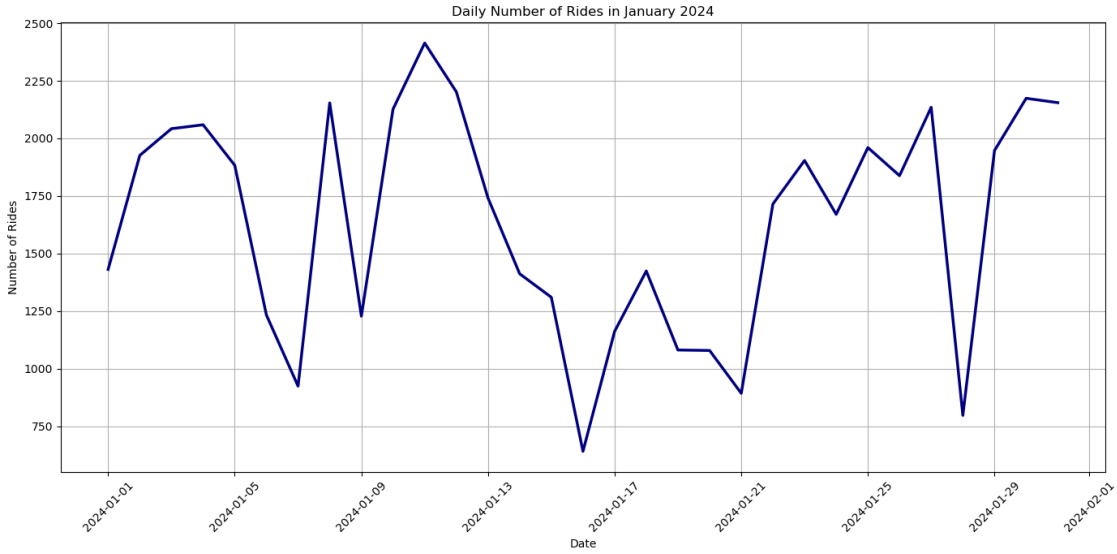
[1]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

[2]:
```
bike_data = pd.read_csv("C:/Users/mattl/OneDrive/Documents/GitHub/DSC 680/
 ↪JC-202401-citibike-tripdata.csv")
```

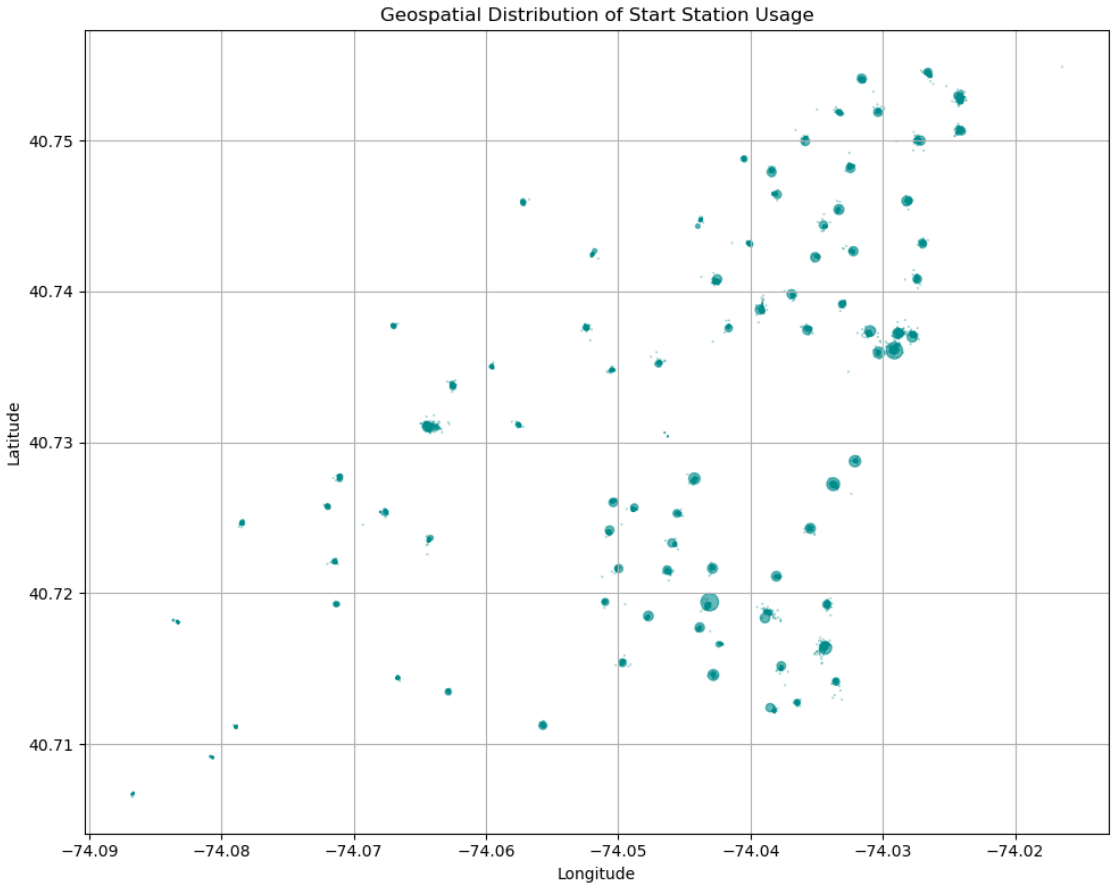
[3]:
```
# Convert 'started_at' and 'ended_at' to datetime
bike_data['started_at'] = pd.to_datetime(bike_data['started_at'])
bike_data['ended_at'] = pd.to_datetime(bike_data['ended_at'])

# Calculate trip durations in minutes
bike_data['trip_duration_min'] = (bike_data['ended_at'] -␣
 ↪bike_data['started_at']).dt.total_seconds() / 60

# Filter out unrealistic trip durations
filtered_data = bike_data[(bike_data['trip_duration_min'] > 1) &␣
 ↪(bike_data['trip_duration_min'] < 1440)]

# Plot histogram of trip durations
plt.figure(figsize=(10, 6))
plt.hist(filtered_data['trip_duration_min'], bins=np.arange(1, 61, 1),␣
 ↪edgecolor='k', alpha=0.7)
plt.title('Distribution of Trip Durations')
plt.xlabel('Duration (Minutes)')
plt.ylabel('Number of Trips')
plt.grid(axis='y', alpha=0.75)
plt.show()
```



[4]:
```
# Aggregate data to get the number of rides per day
daily_rides = bike_data.groupby(bike_data['started_at'].dt.date).size()

# Plot time series of daily rides
plt.figure(figsize=(14, 7))
daily_rides.plot(kind='line', linewidth=2.5, color='navy')
plt.title('Daily Number of Rides in January 2024')
plt.xlabel('Date')
plt.ylabel('Number of Rides')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[5]:
```
# Aggregate data to count trips starting from each station
start_station_usage = bike_data.groupby(['start_station_name', 'start_lat',␣
 ↪'start_lng']).size().reset_index(name='num_trips')

# Plot scatter for geospatial distribution of start station usage
plt.figure(figsize=(10, 8))
plt.scatter(start_station_usage['start_lng'], start_station_usage['start_lat'],
            s=start_station_usage['num_trips']/10, # Scale point size for␣
 ↪better visualization
            alpha=0.6, color='darkcyan')
plt.title('Geospatial Distribution of Start Station Usage')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.grid(True)
plt.tight_layout()
plt.show()
```


Geospatial Distribution of Start Station Usage
Latitude
Longitude
40.75
40.74
40.73
40.72
40.71
−74.09
−74.08
−74.07
−74.06
−74.05
−74.04
−74.03
−74.02

[ ]: