DATA ANALYSIS


Motivation
&
Objective

- Taxi is one of the urban public transports in many busy countries. Unlike other public transports, taxi rides provide accessibility, convenience, yet privacy to passengers.
- Millions of taxi trips data are generated on monthly basis, which this data can be useful to gain the insight of the traffic patterns and obtain a clear view of urban city life.

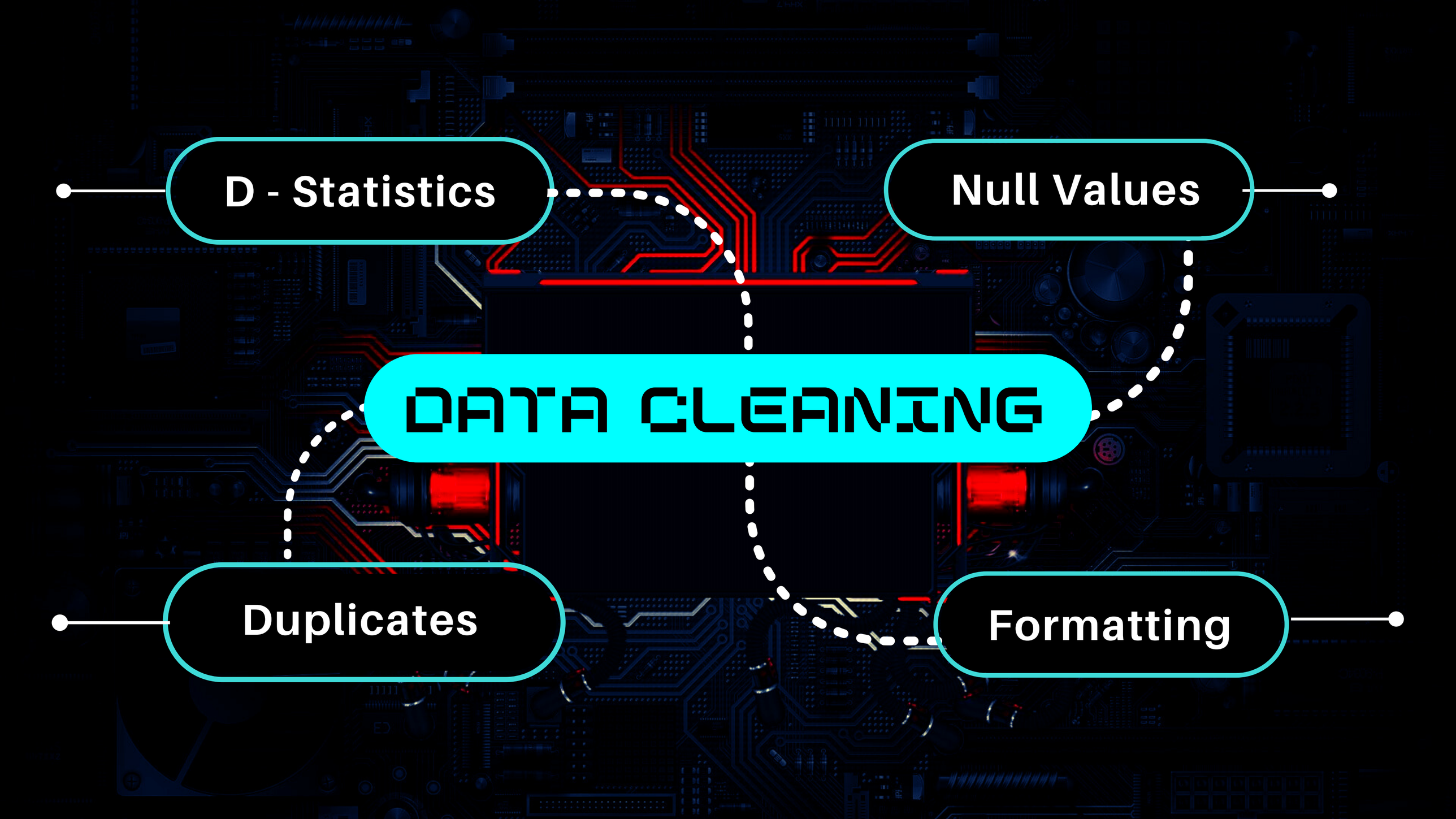
- With automated data collection of taxi movement, a city's operation can be extracted from geospatial data in both spatial and temporal point of view.
- It provides a more accurate depiction of the nature of a region, considering that daily movement and activities found in geospatial data indicate the social-economic properties of urban functions

Dataset Details

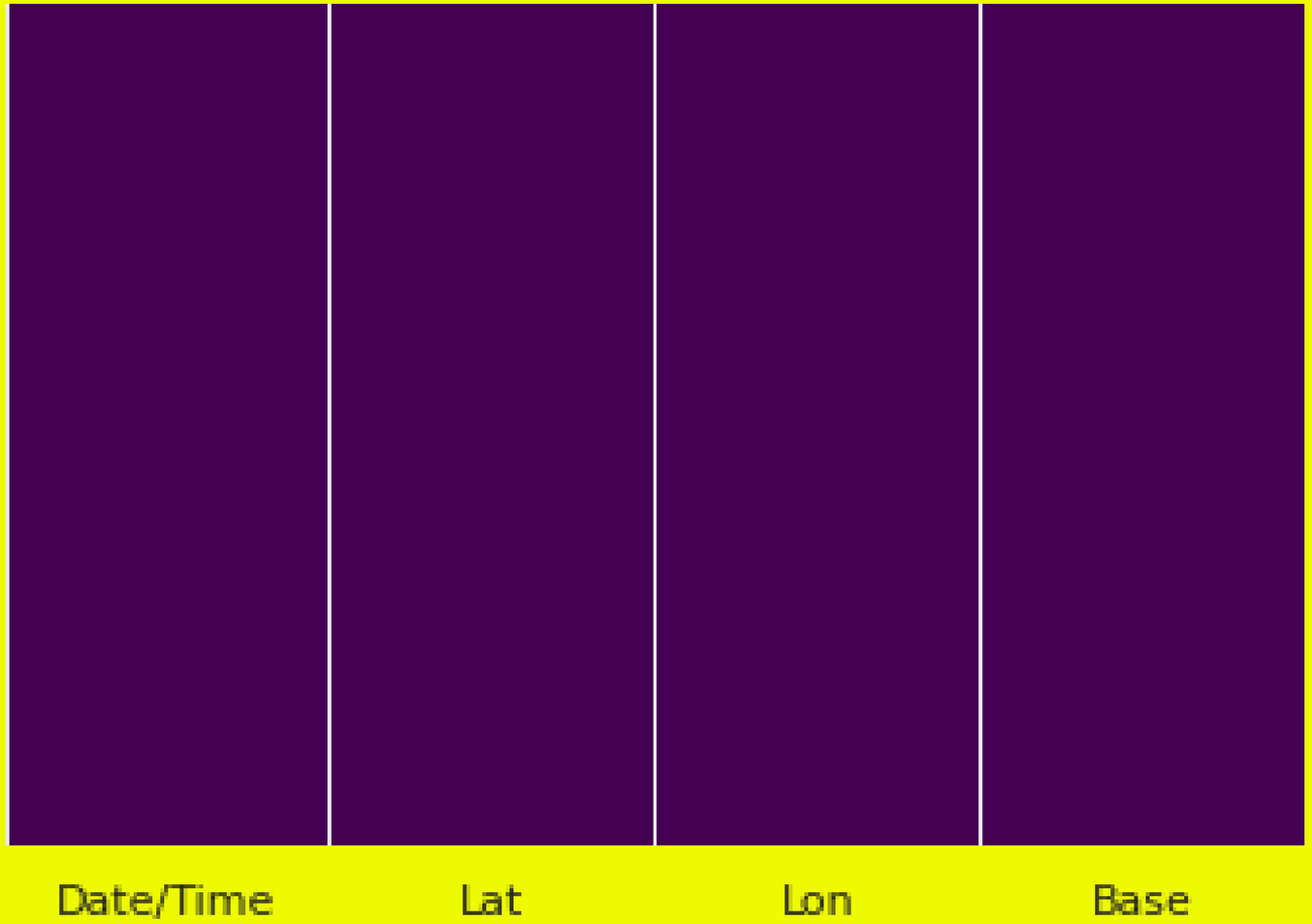
- In this project, data is obtained through Kaggle repository titled “Uber TLC FOI Response”
- This directory contains data on over 4.5 million Uber pickups in New York City from April to September 2014, and 14.3 million more Uber pickups from January to June 2015.

- Trip-level data on 10 other for-hire vehicle (FHV) companies, as well as aggregated data for 329 FHV companies, is also included.
- All the files are as they were received on August 3, Sept. 15 and Sept. 22, 2015

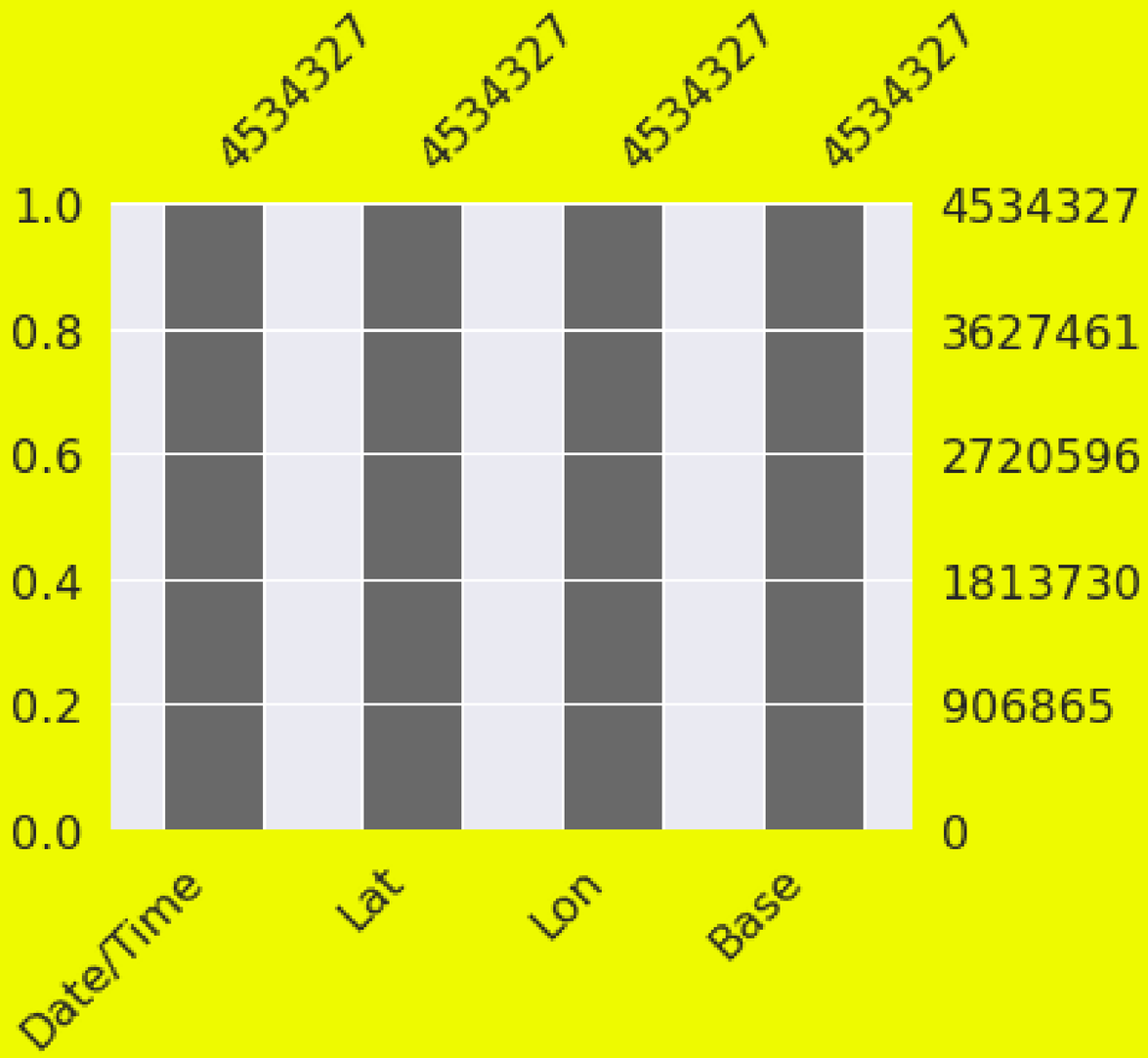
Data Cleaning


D - Statistics
Null Values
DATA CLEANING
Duplicates
Formatting

HEAT MAP

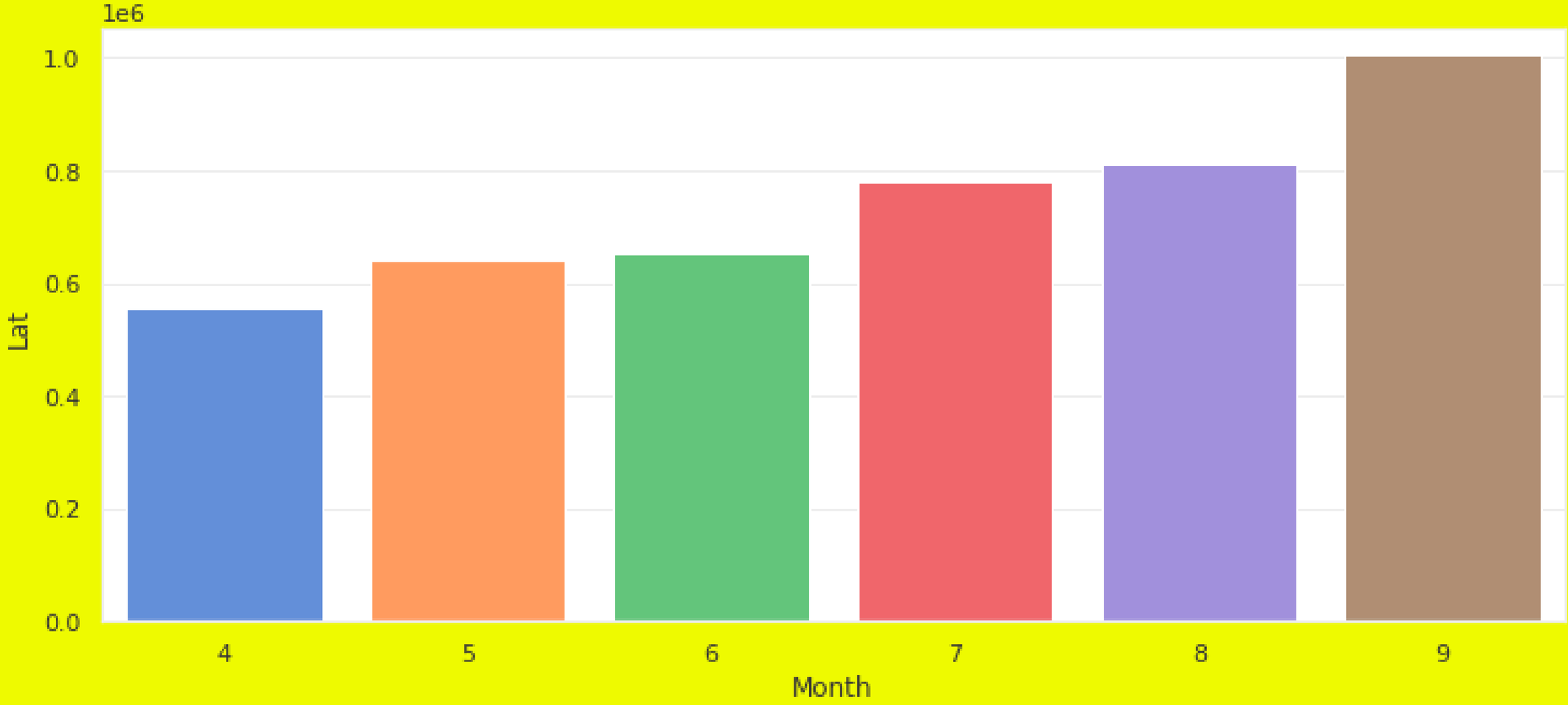


MSNO BAR

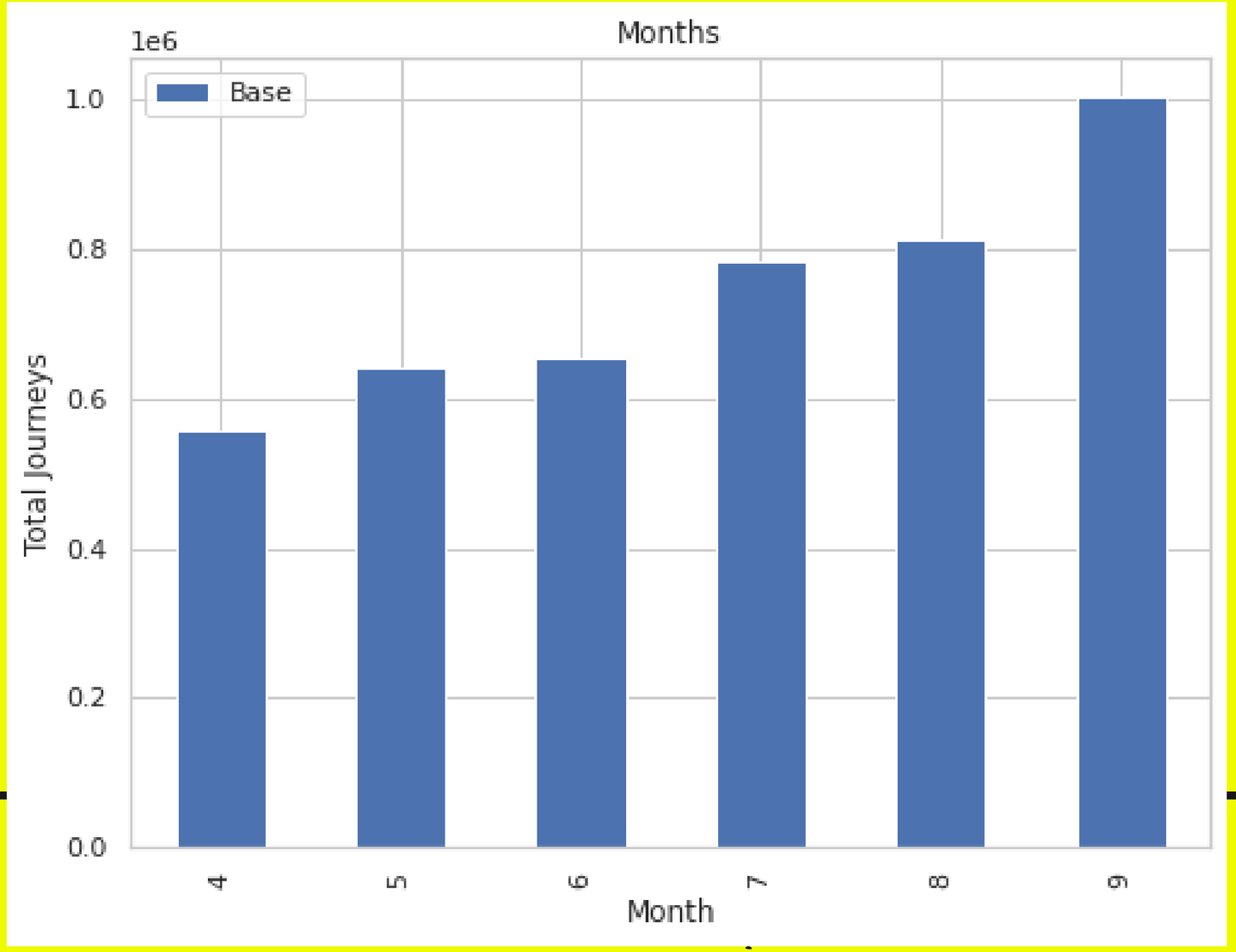

4534327
4534327
4534327
4534327
1.0
0.8
0.6
0.4
0.2
0.0
4534327
3627461
2720596
1813730
906865
0
Date/Time
Lat
Lon
Base

Exploratory Data Analysis

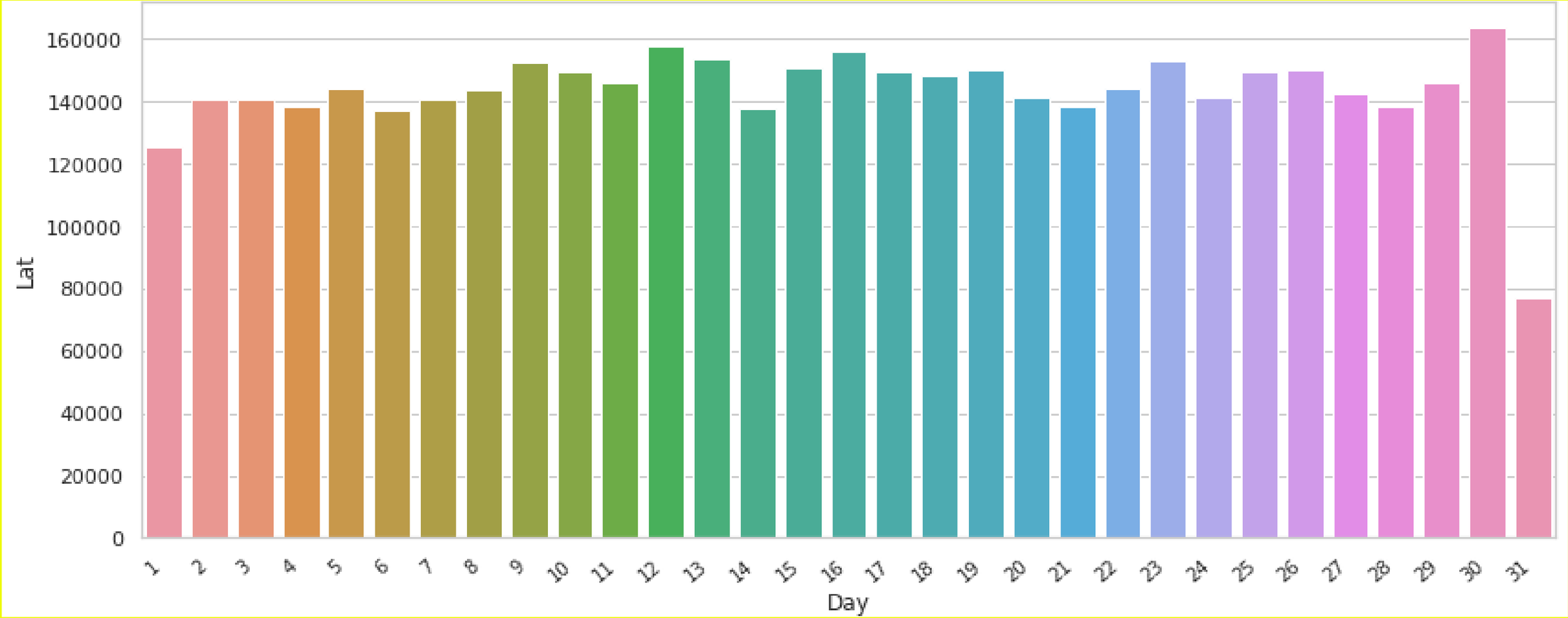
Trip by Month

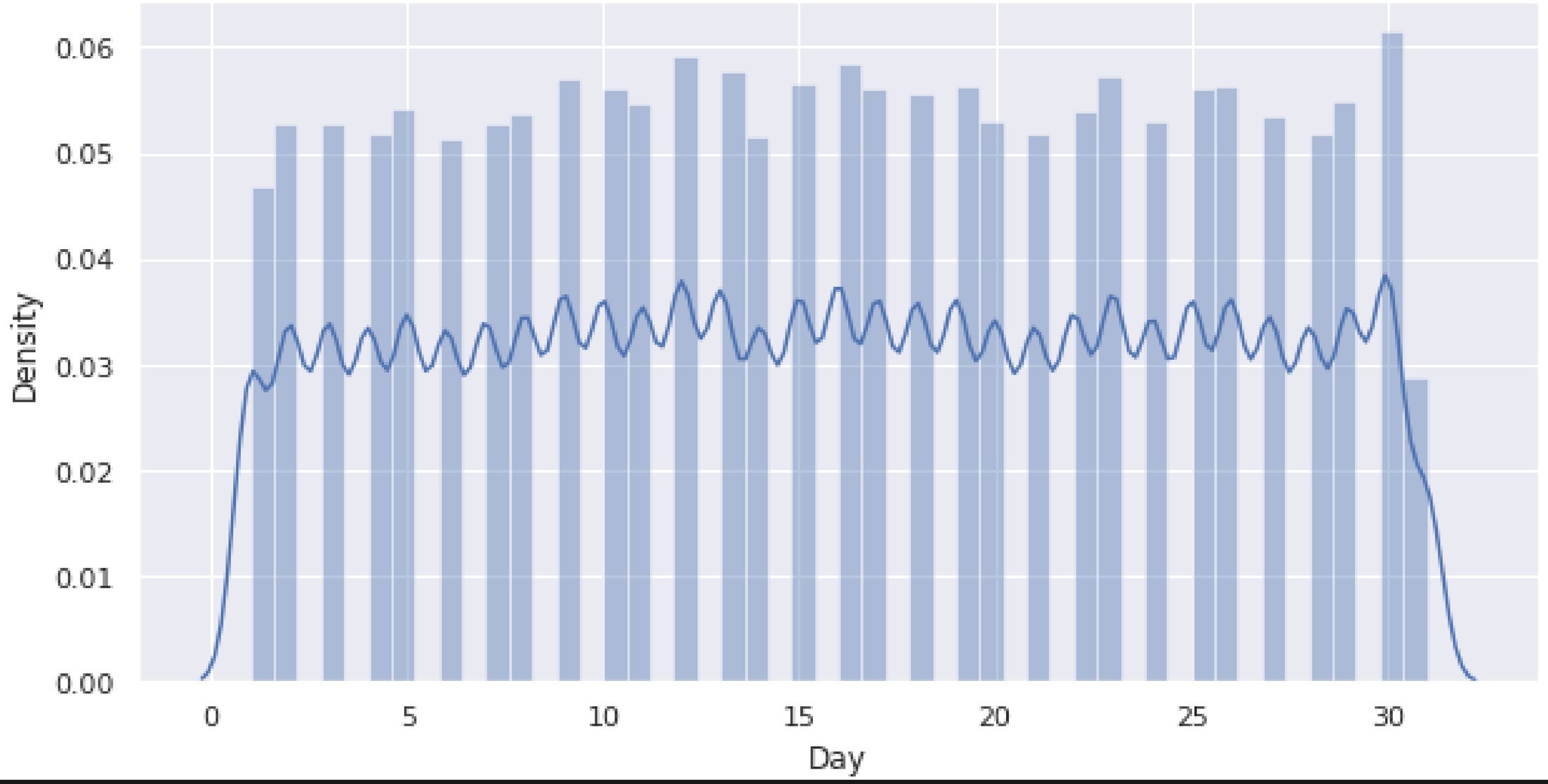


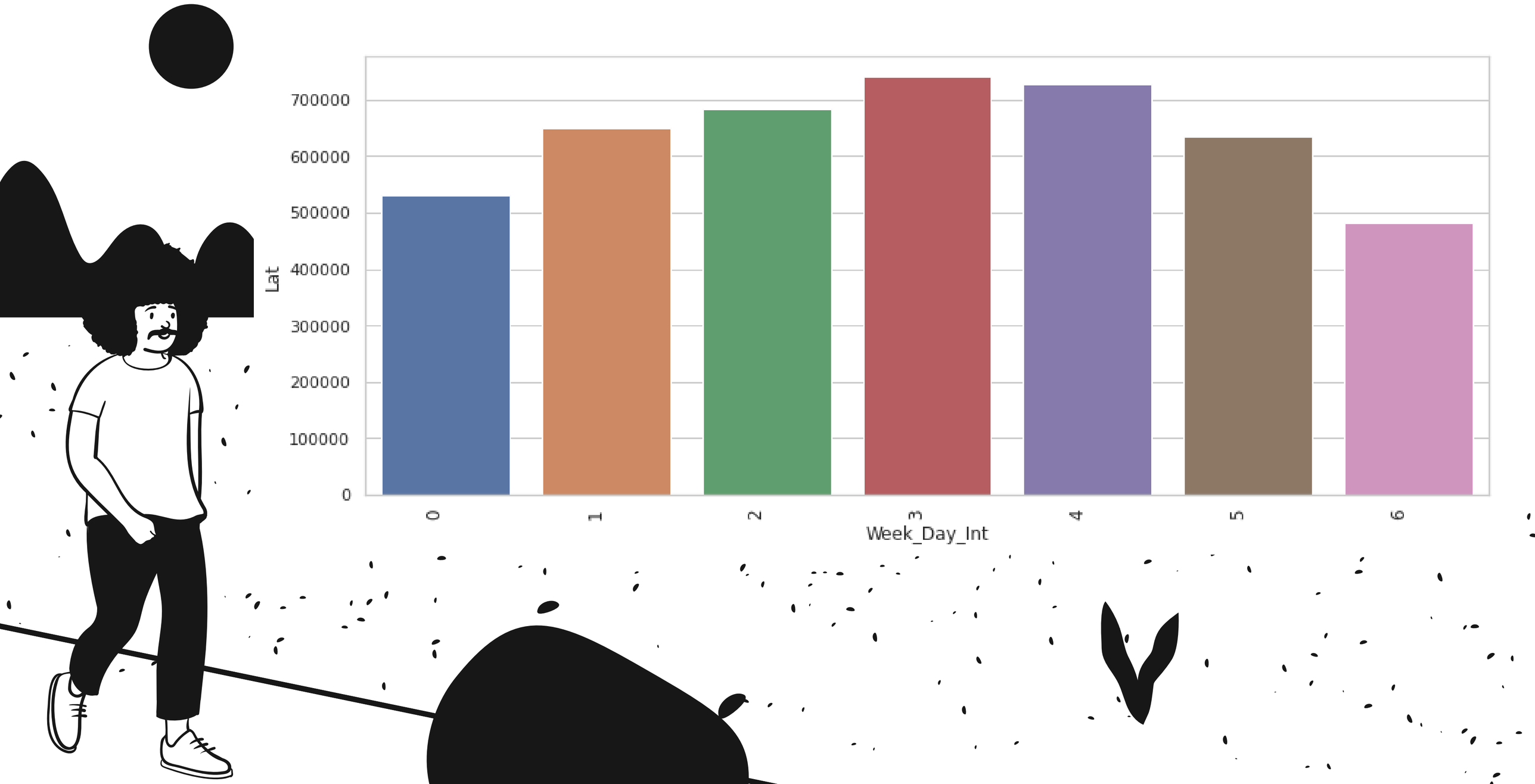
TOTAL JOURNEY

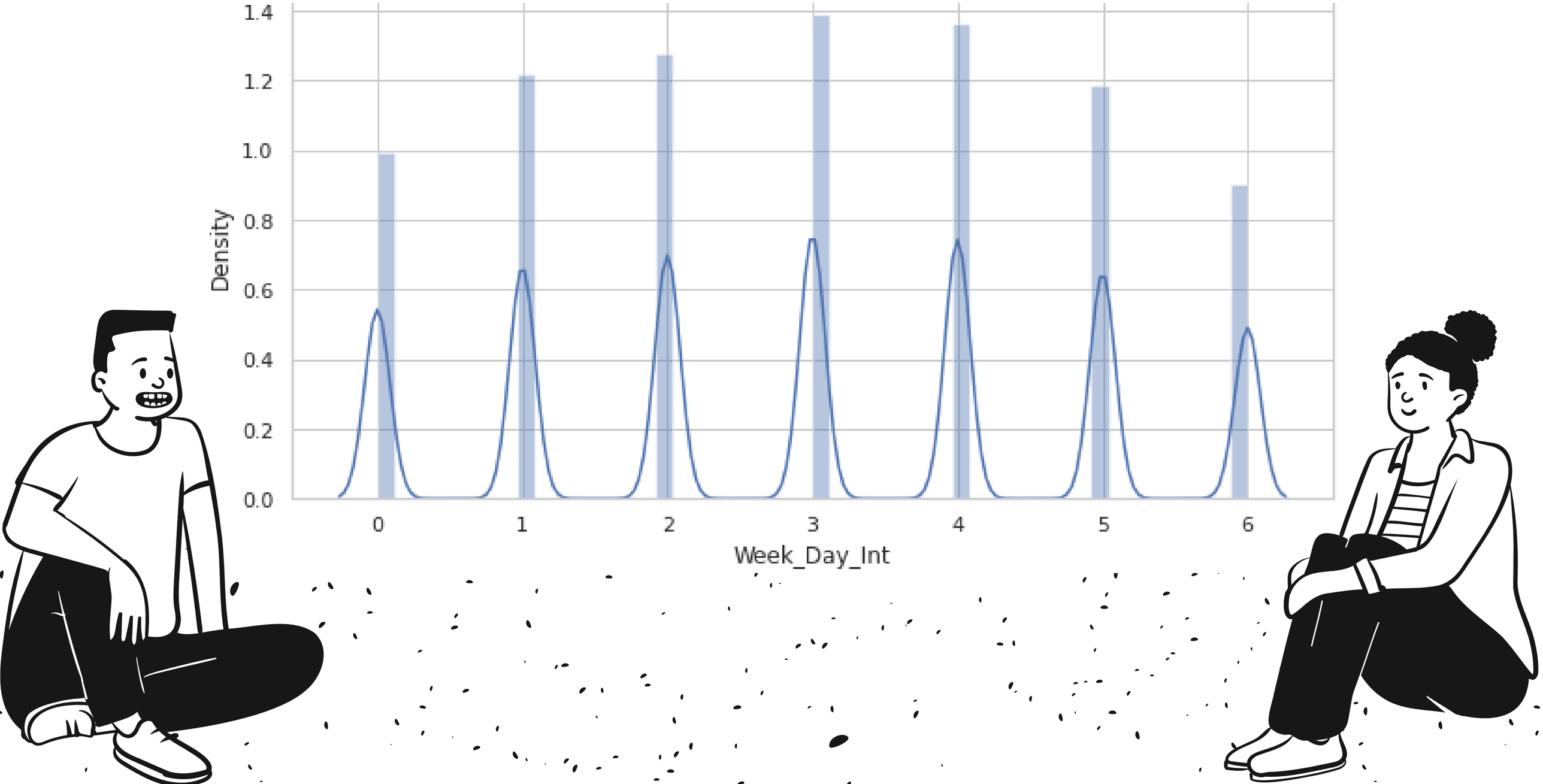


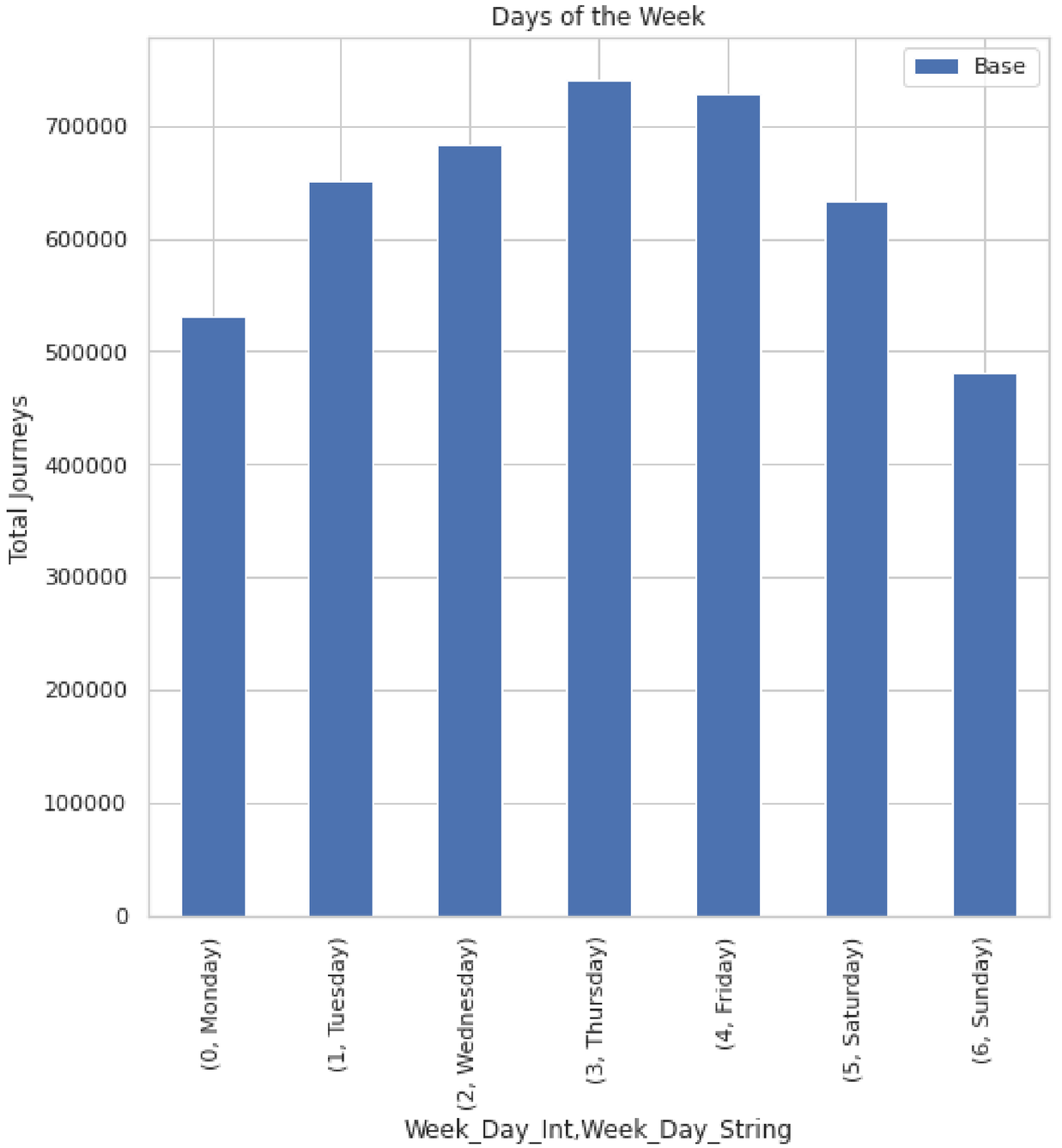
Trip by Week Day

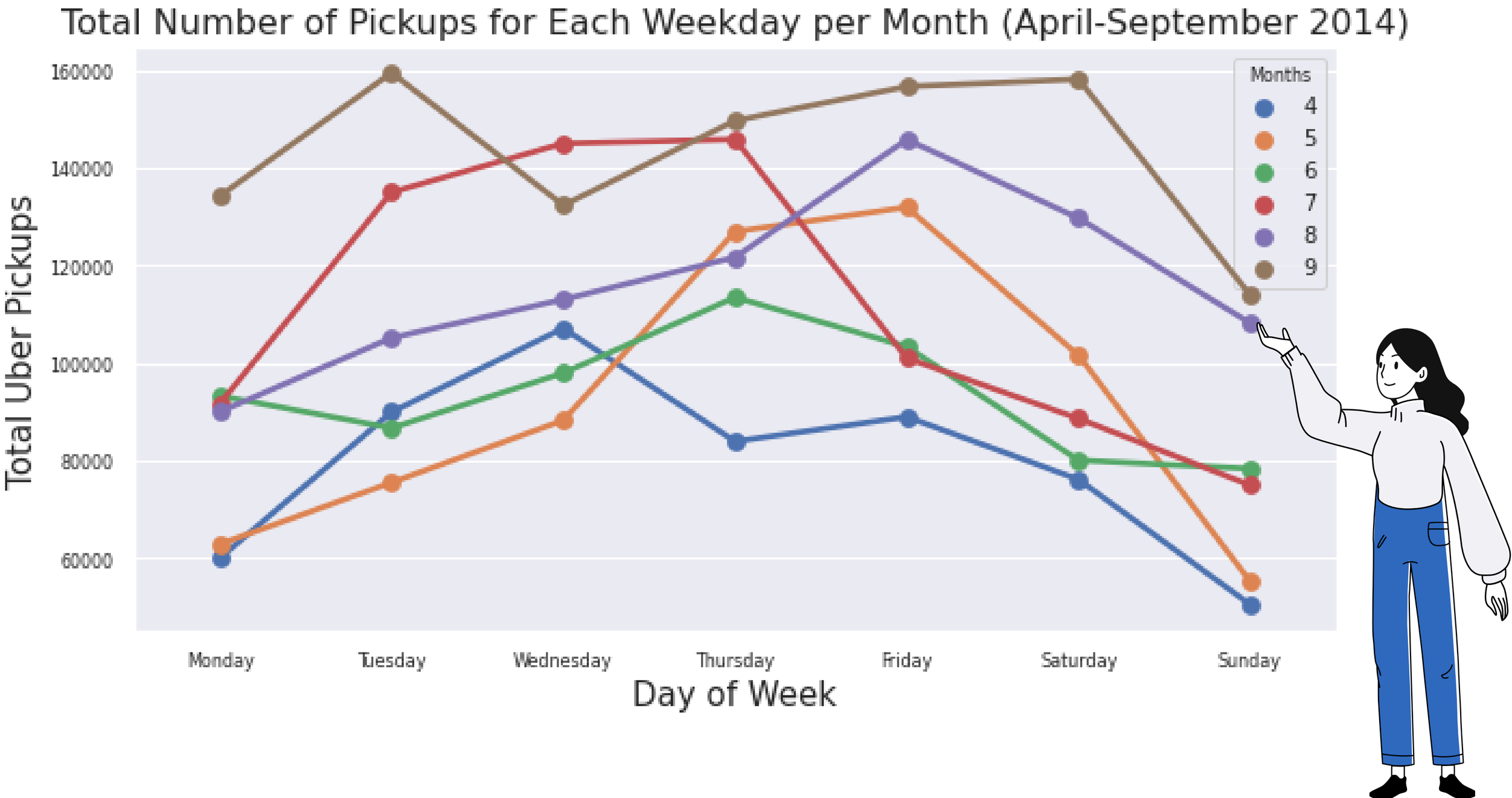



0.06
0.05
0.04
0.03
0.02
0.01
0.00
Density
0
5
10
15
20
25
30
Day

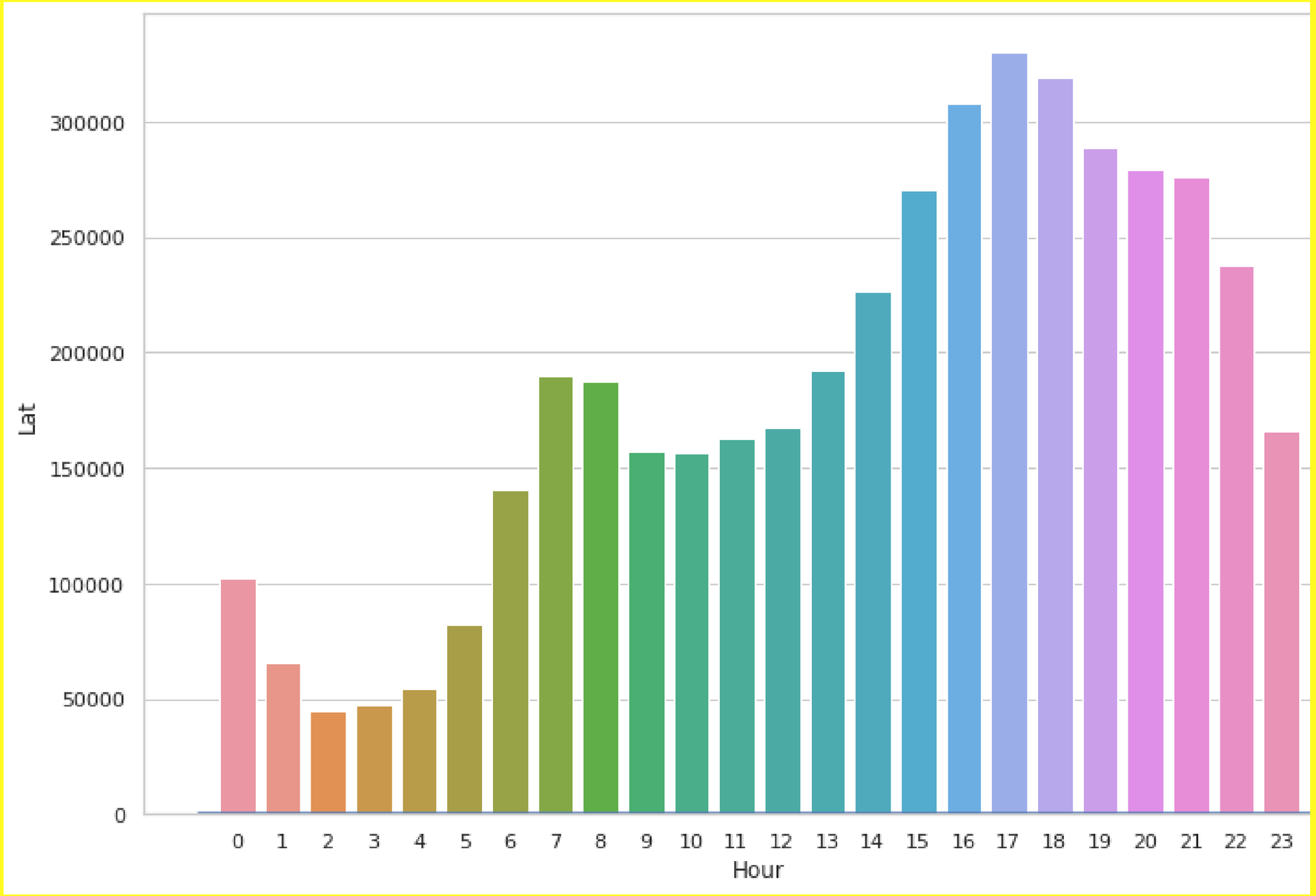

700000
600000
500000
400000
300000
200000
100000
0
Lat
0
1
2
3
4
5
6
Week_Day_Int

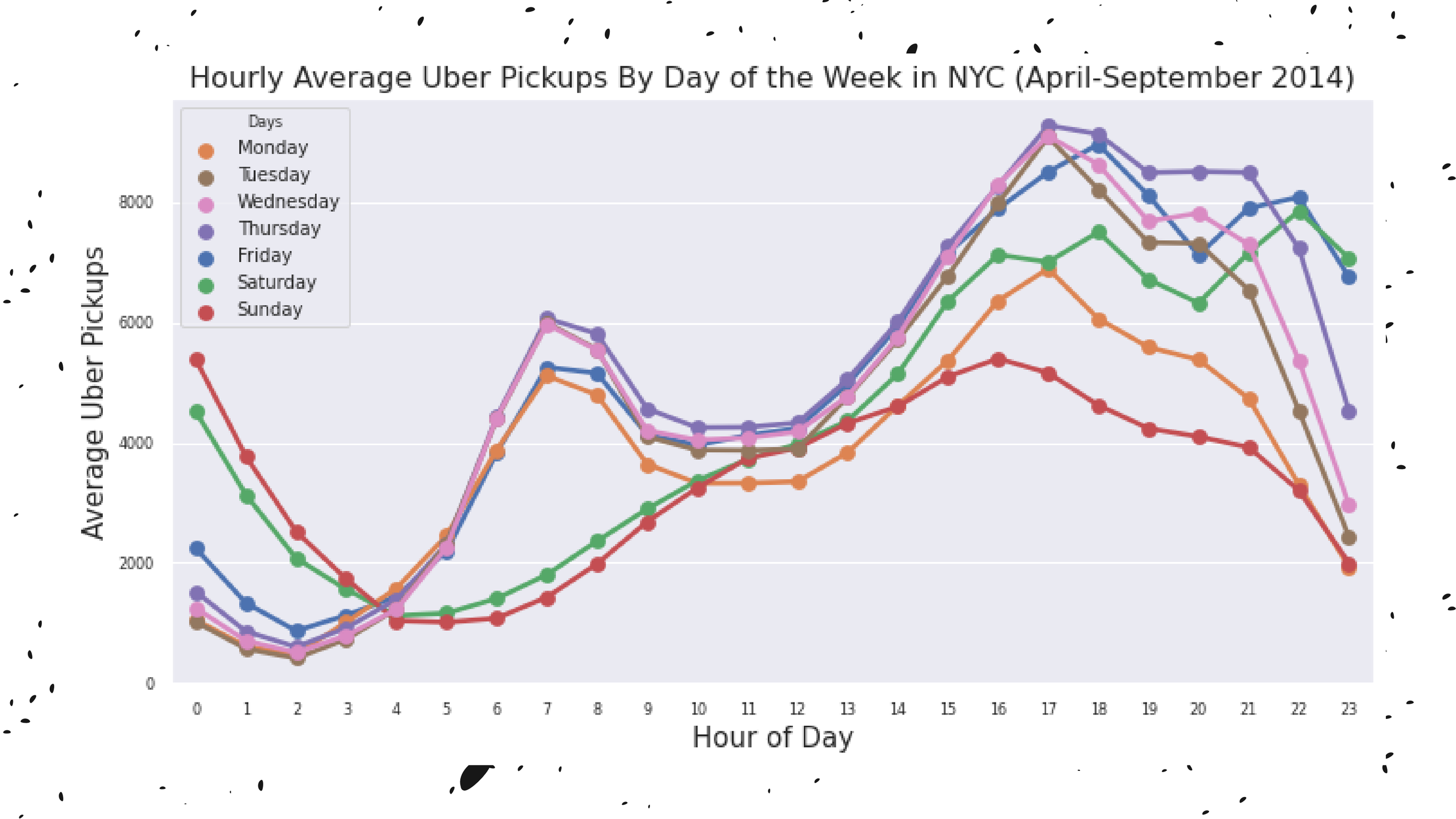

1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Density
0
1
2
3
4
5
6
Week_Day_Int


Days of the Week
Base
700000
600000
500000
400000
300000
200000
100000
0
Total Journeys
(0, Monday)
(1, Tuesday)
(2, Wednesday)
(3, Thursday)
(4, Friday)
(5, Saturday)
(6, Sunday)
Week_Day_Int,Week_Day_String

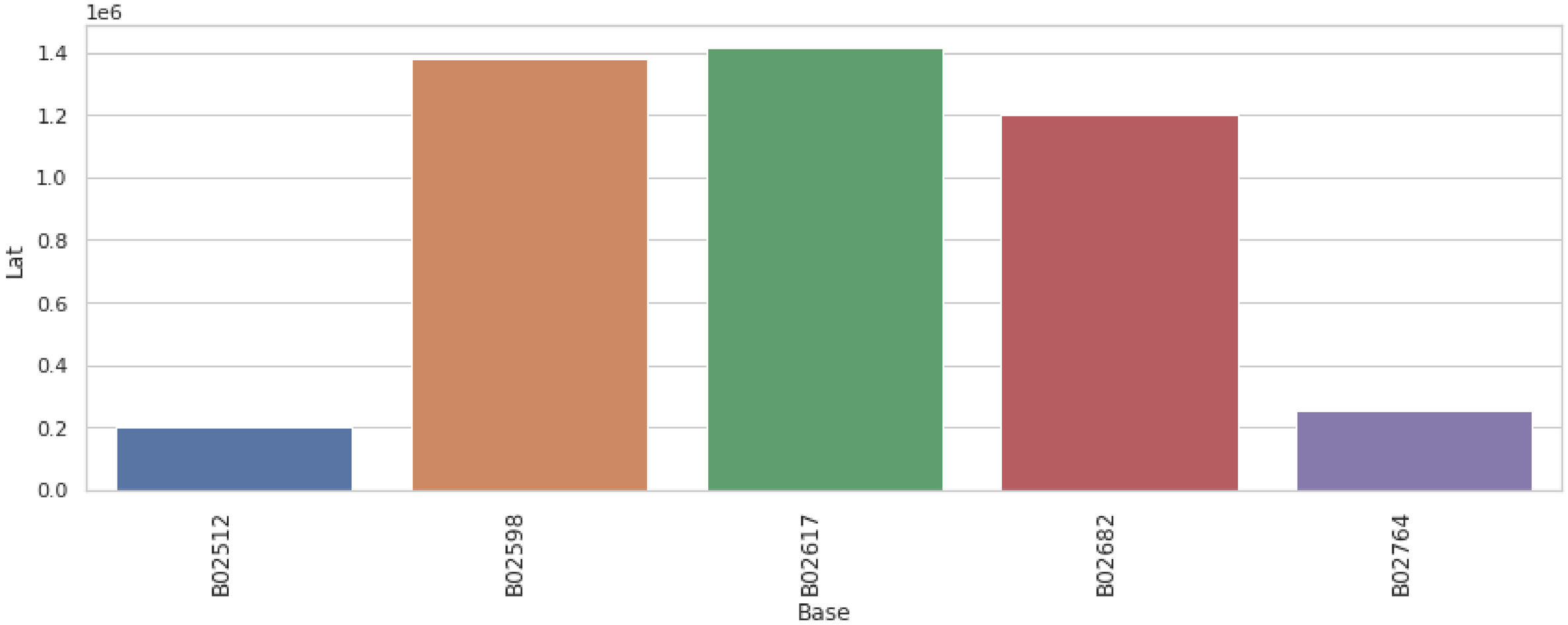

Total Number of Pickups for Each Weekday per Month (April-September 2014)
Total Uber Pickups
160000
140000
120000
100000
80000
60000
Months
4
5
6
7
8
9
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week

Trip by Hour

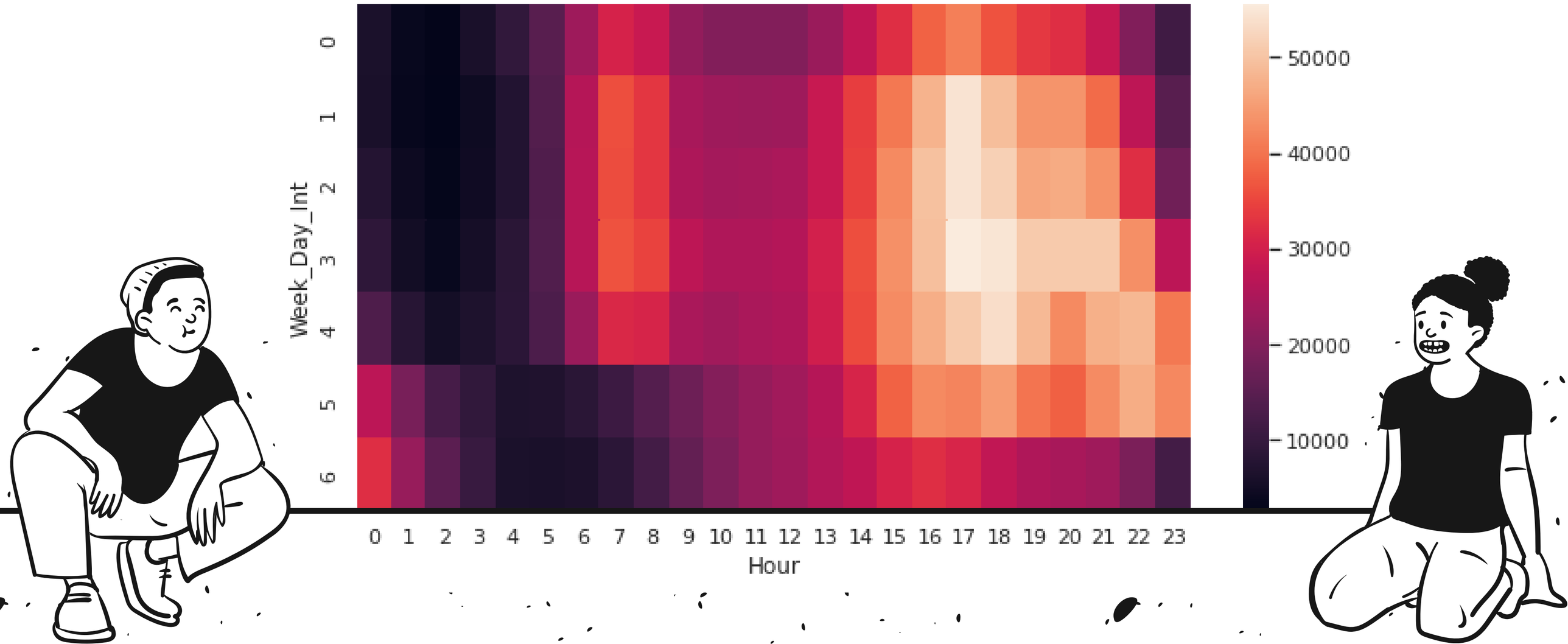

300000
250000
200000
150000
100000
50000
0
Lat
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour


Hourly Average Uber Pickups By Day of the Week in NYC (April-September 2014)
Days
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Average Uber Pickups
0
2000
4000
6000
8000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day

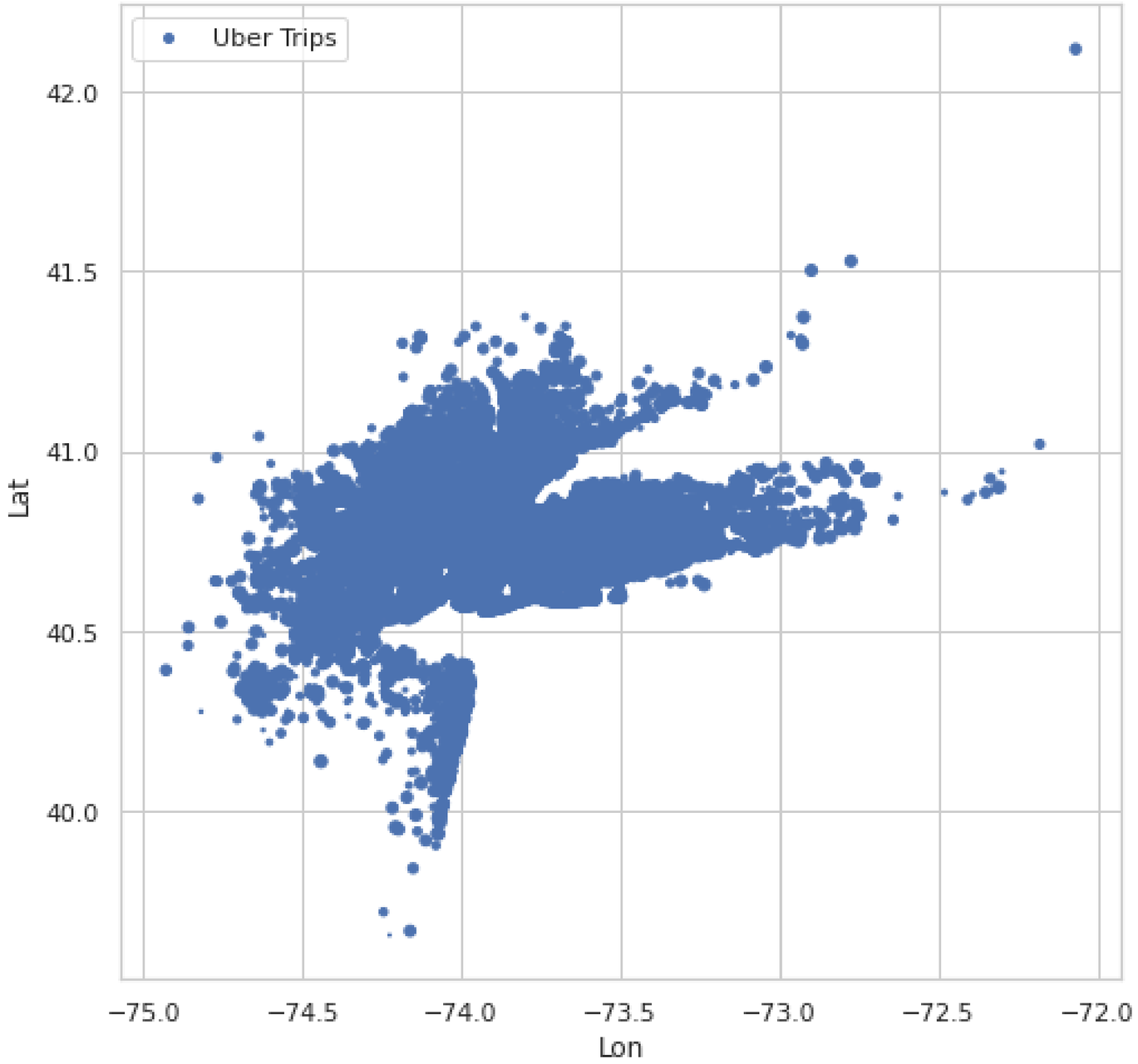
Trip by Base


1e6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Lat
B02512
B02598
B02617
B02682
B02764
Base

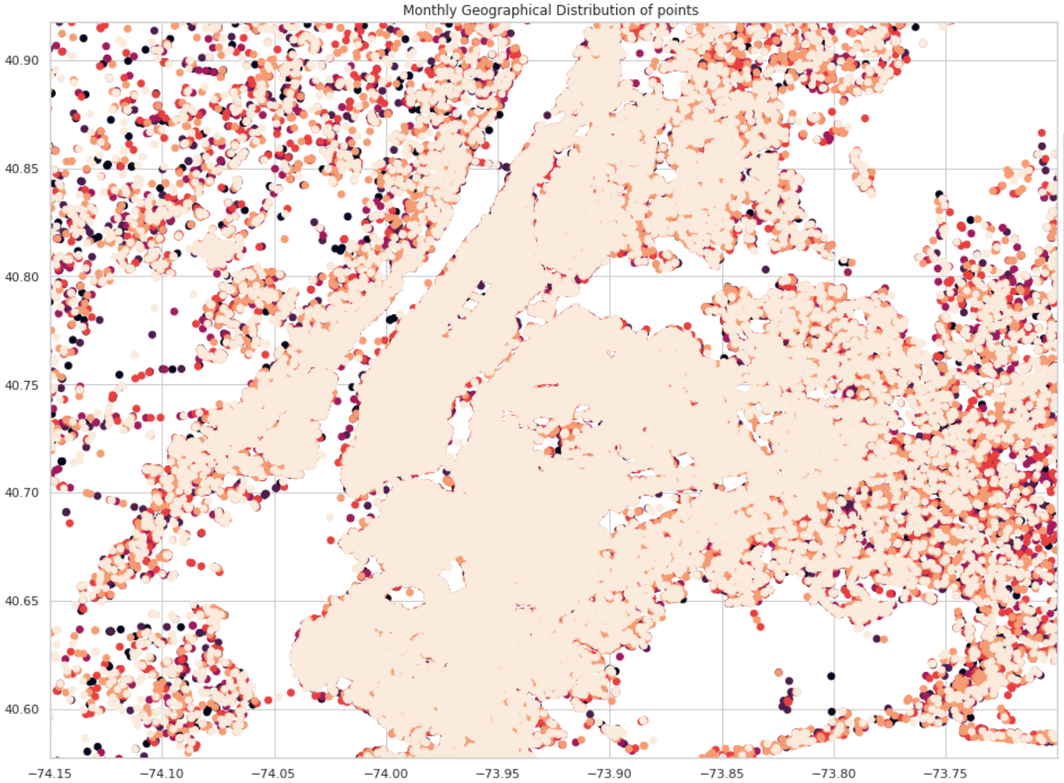
Correlation of Hours & Week Days



Density of Uber Trips

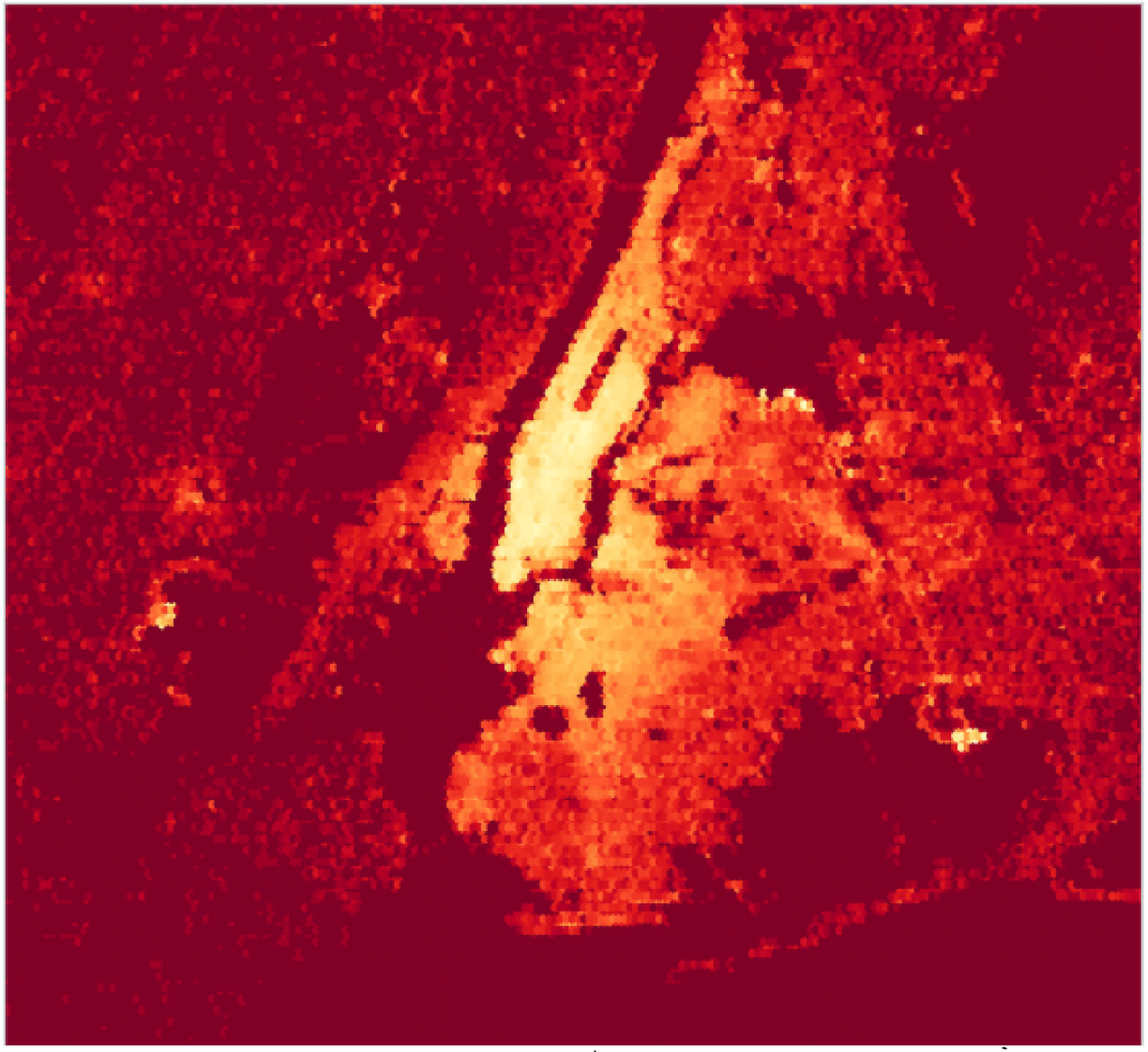

Uber Trips
42.0
41.5
41.0
40.5
40.0
Lat
−75.0
−74.5
−74.0
−73.5
−73.0
−72.5
−72.0
Lon

Geographical Distribution


Monthly Geographical Distribution of points
40.90
40.85
40.80
40.75
40.70
40.65
40.60
−74.15
−74.10
−74.05
−74.00
−73.95
−73.90
−73.85
−73.80
−73.75


Daily Geographical Distribution of points
40.90
40.85
40.80
40.75
40.70
40.65
40.60
−74.15
−74.10
−74.05
−74.00
−73.95
−73.90
−73.85
−73.80
−73.75

HEAT MAP

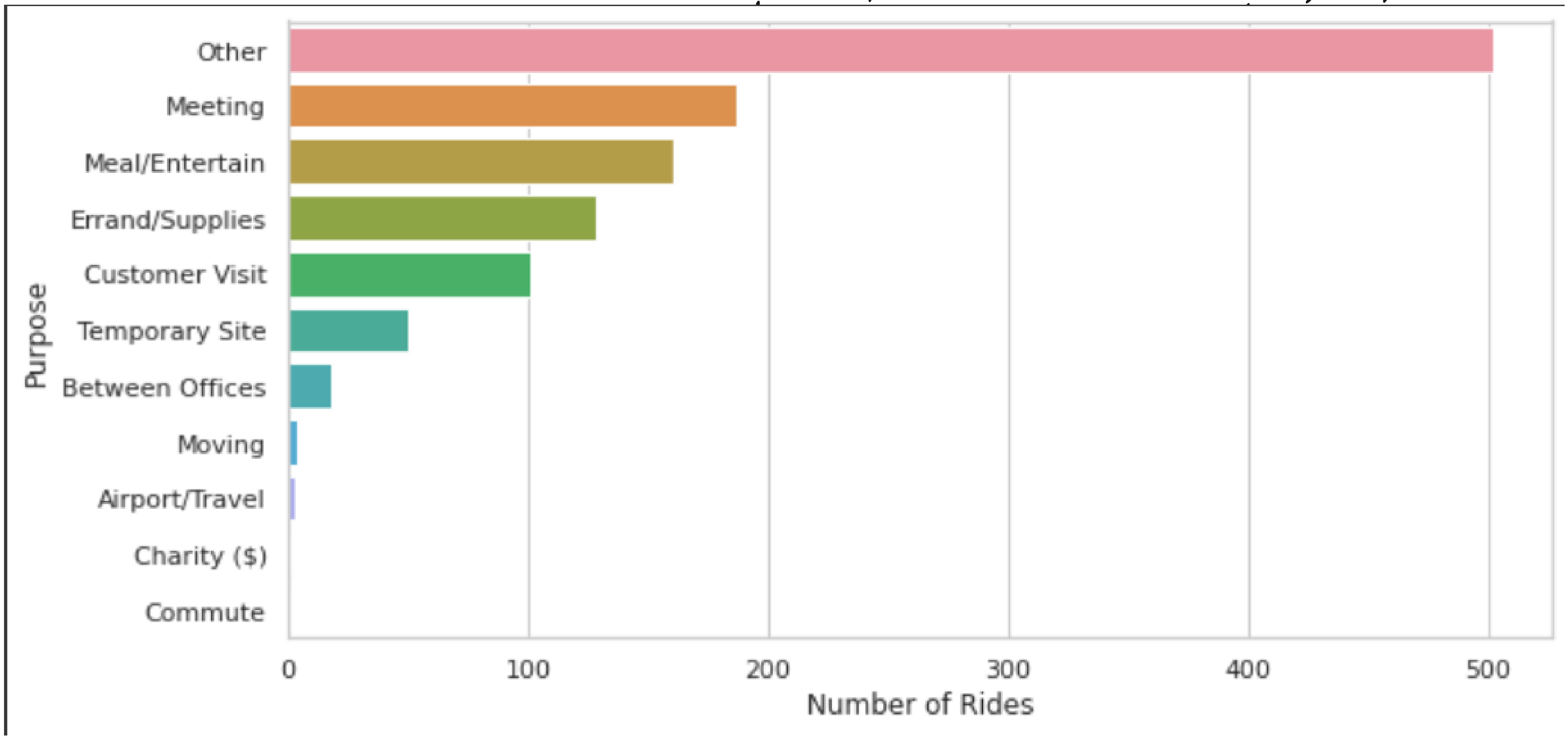


COLAB

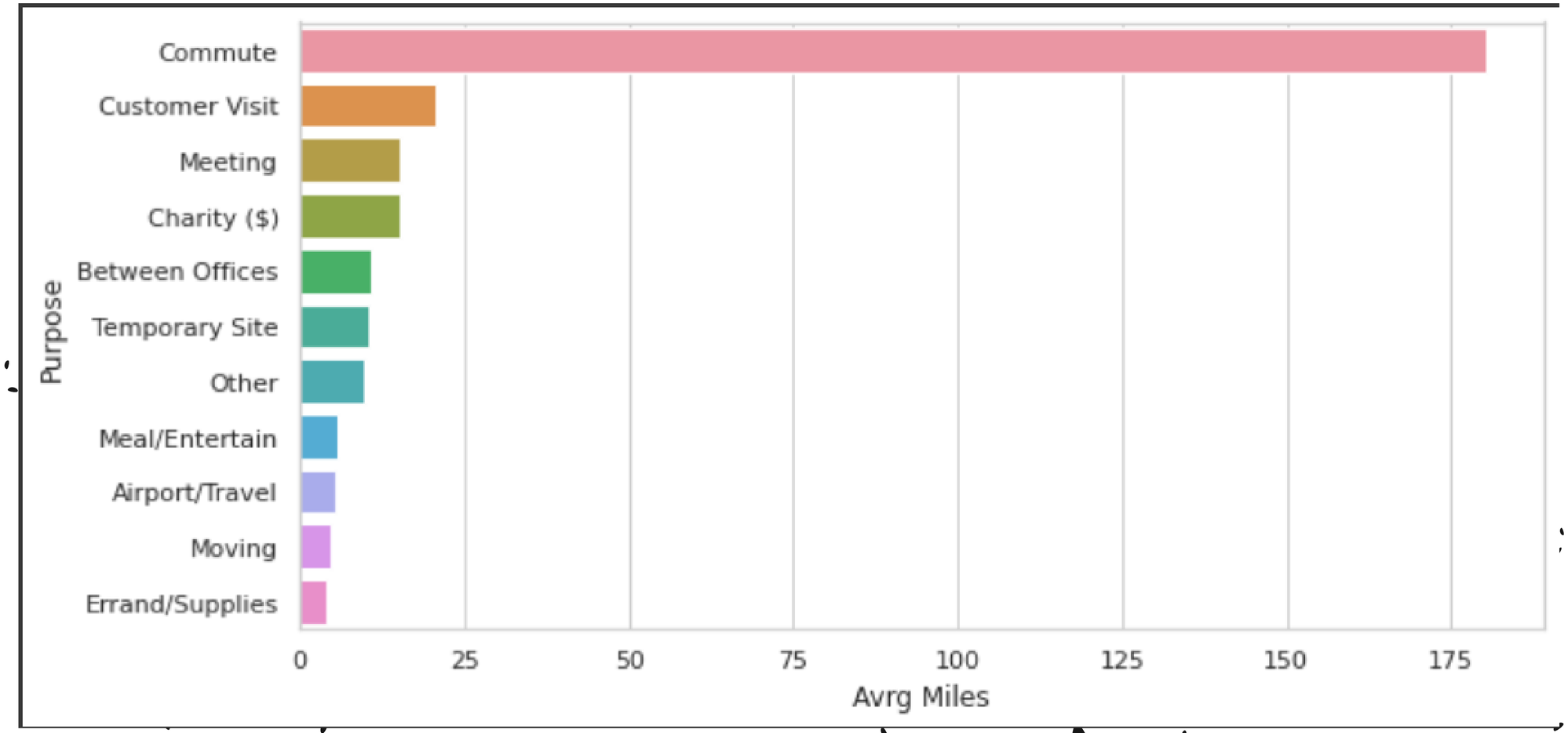
HIGHLY CLUSTERED AREAS

NETWORK ANALYSIS

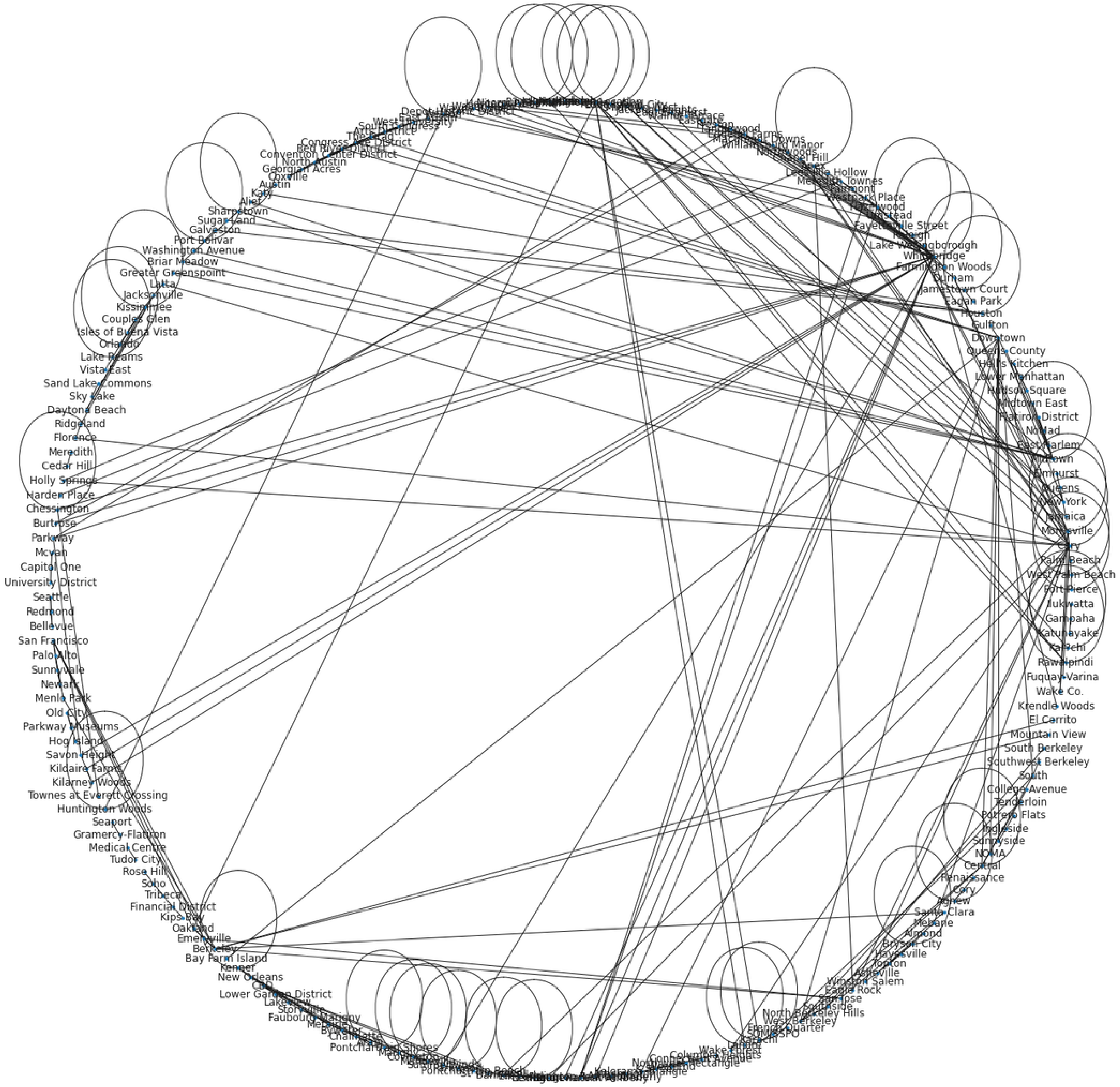
Number of Rides vs Purpose of Travel



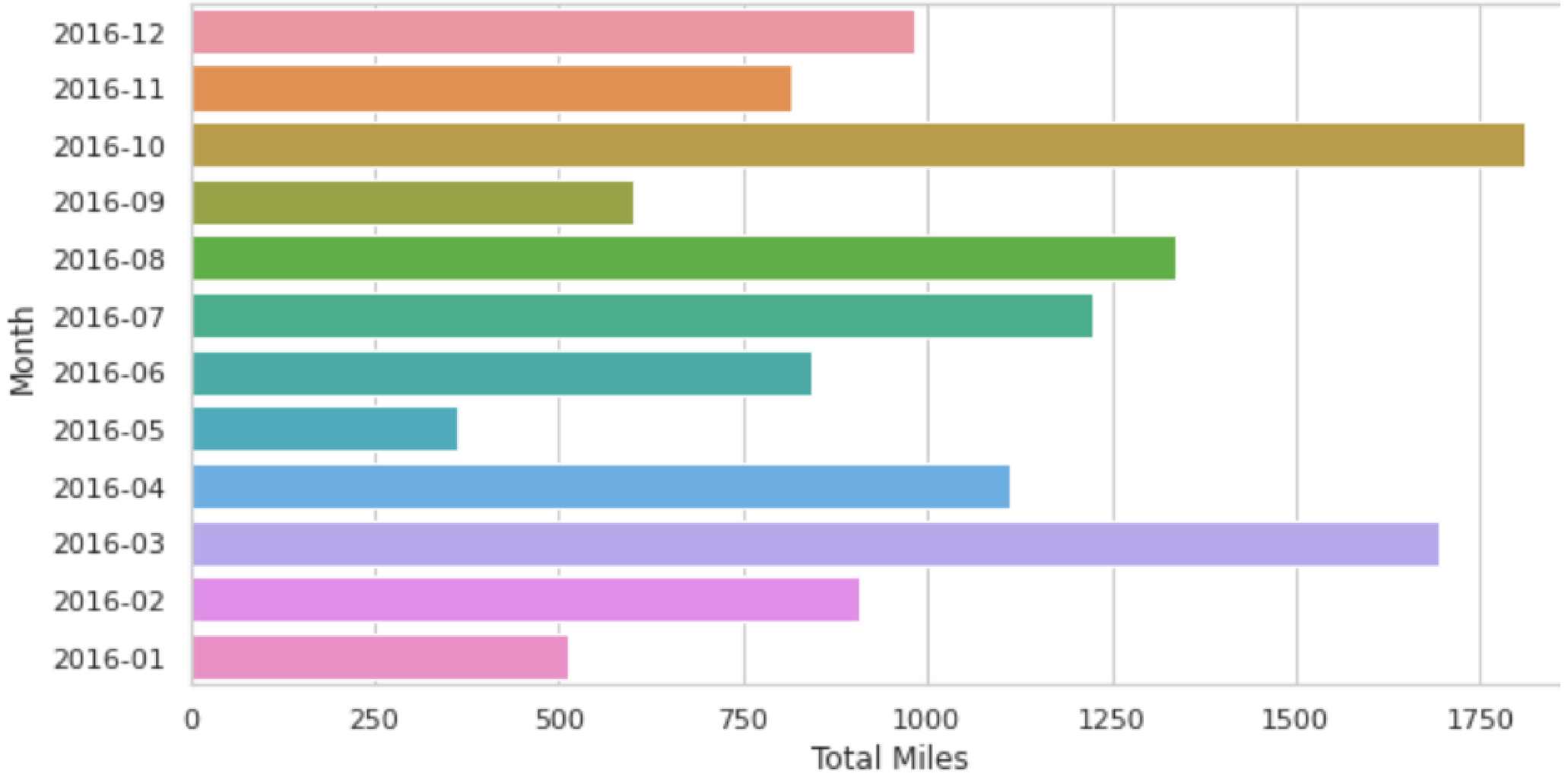
Average Miles travelled for a certain Purpose



RIDE NETWORK



Total miles travelled for each month

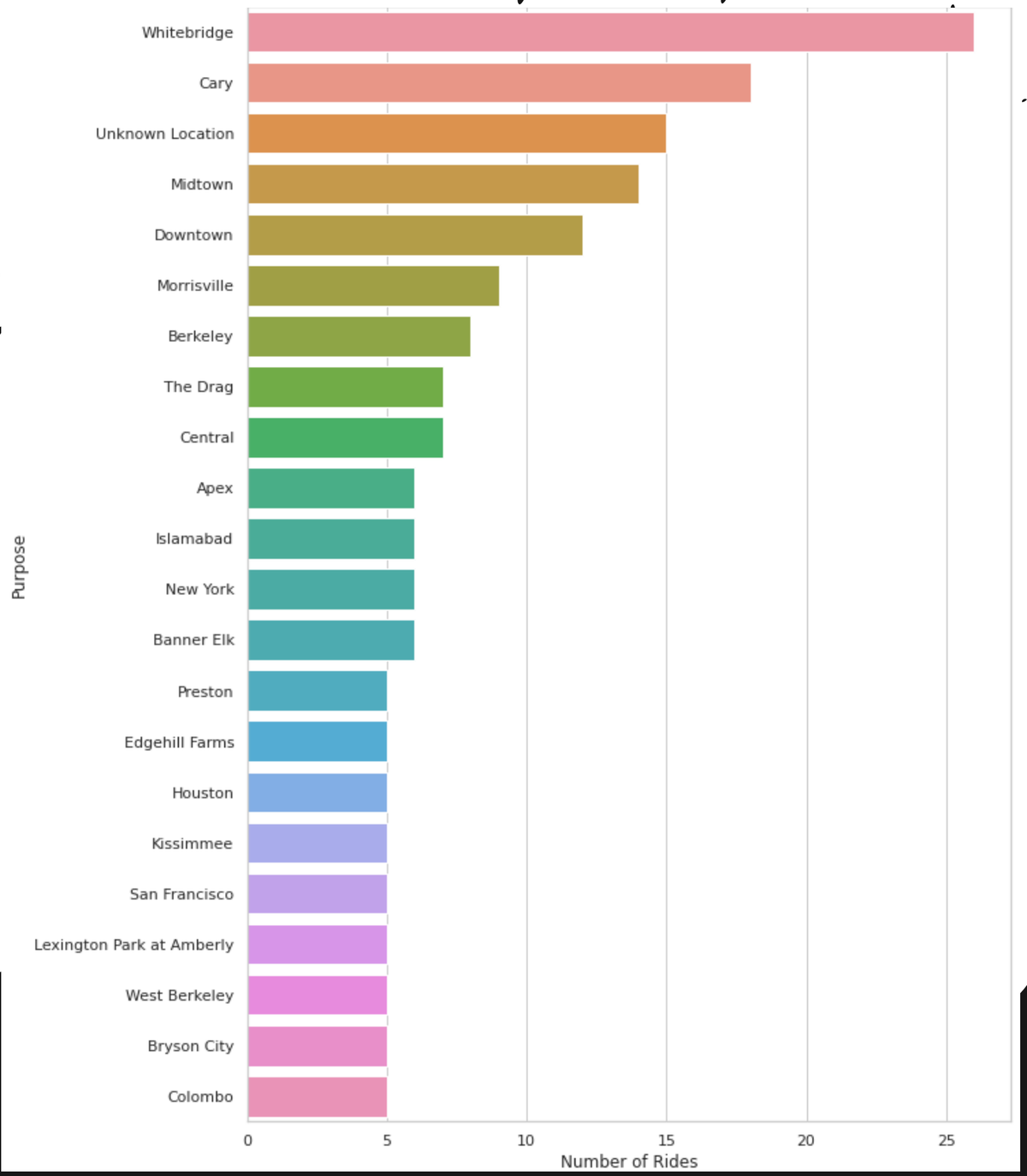


Locations visited atleast 5 times

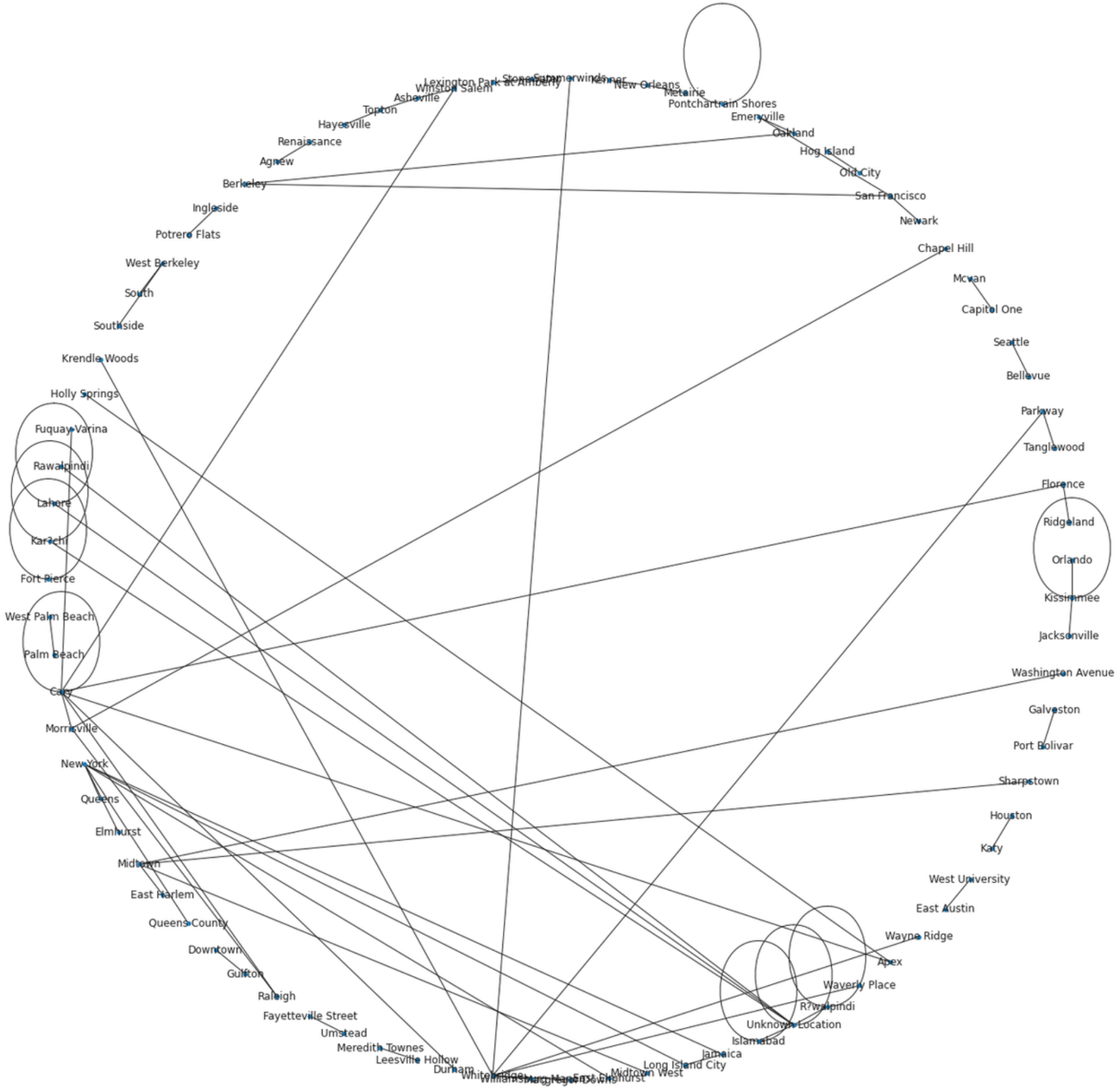
	Location	Degree
25	Whitebridge	26
3	Cary	18
50	Unknown Location	15
9	Midtown	14
18	Downtown	12
...	...	...
89	Vista East	1
86	Isles of Buena Vista	1
81	Greater Greenspoint	1
79	Washington Avenue	1
208	Ilukwatta	1

209 rows × 2 columns

PURPOSE VS NUMBER OF RIDES FOR A PARTICULAR PURPOSE



uber -Ride analysis



DATA ANALYTICS

DATA ANALYTICS

Data analytics focuses on processing and performing statistical analysis of existing datasets. Analysts concentrate on creating methods to capture, process, and organize data to uncover actionable insights for current problems, and establishing the best way to present this data.

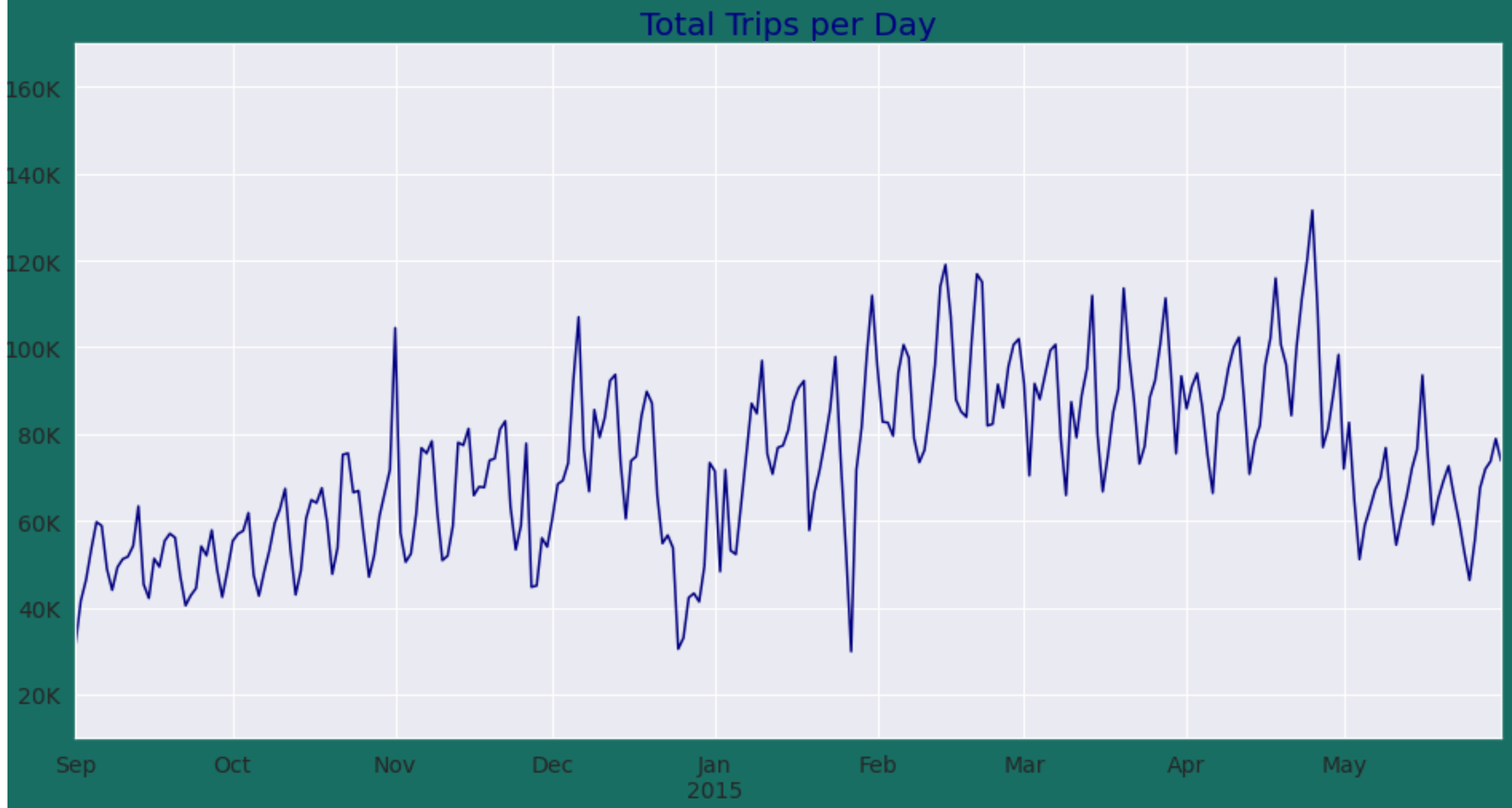
VISUALIZATION

VISUALIZATION

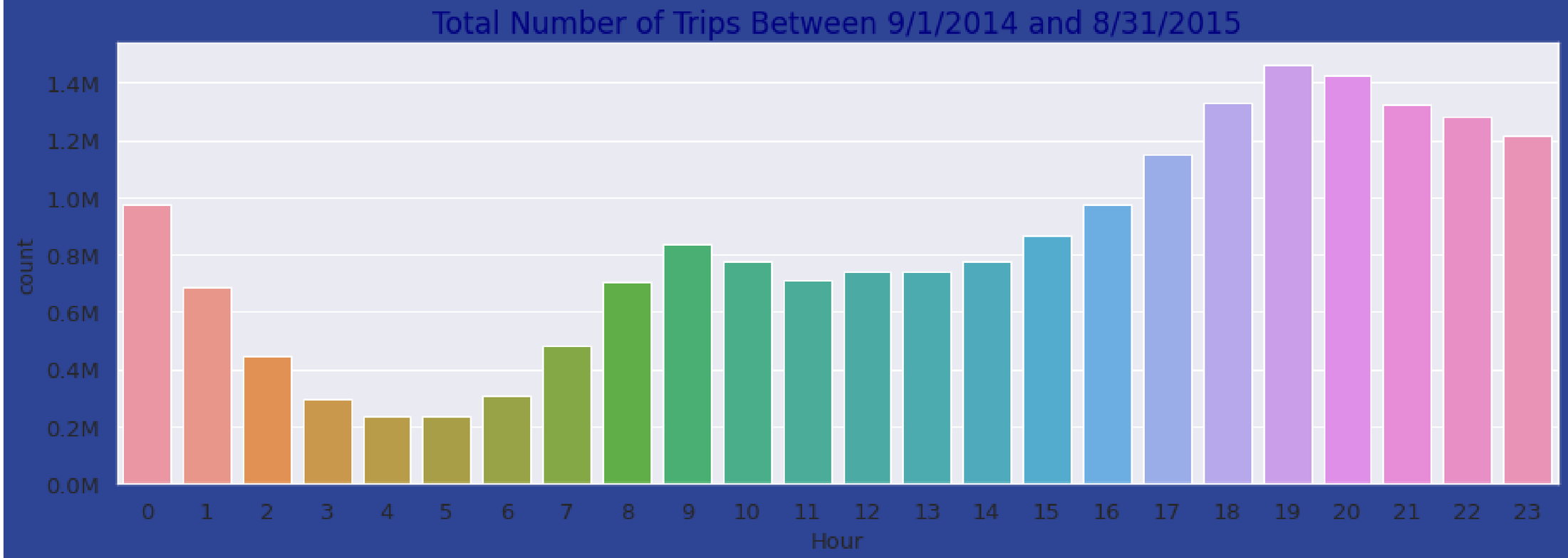
visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Total Trips Per Day

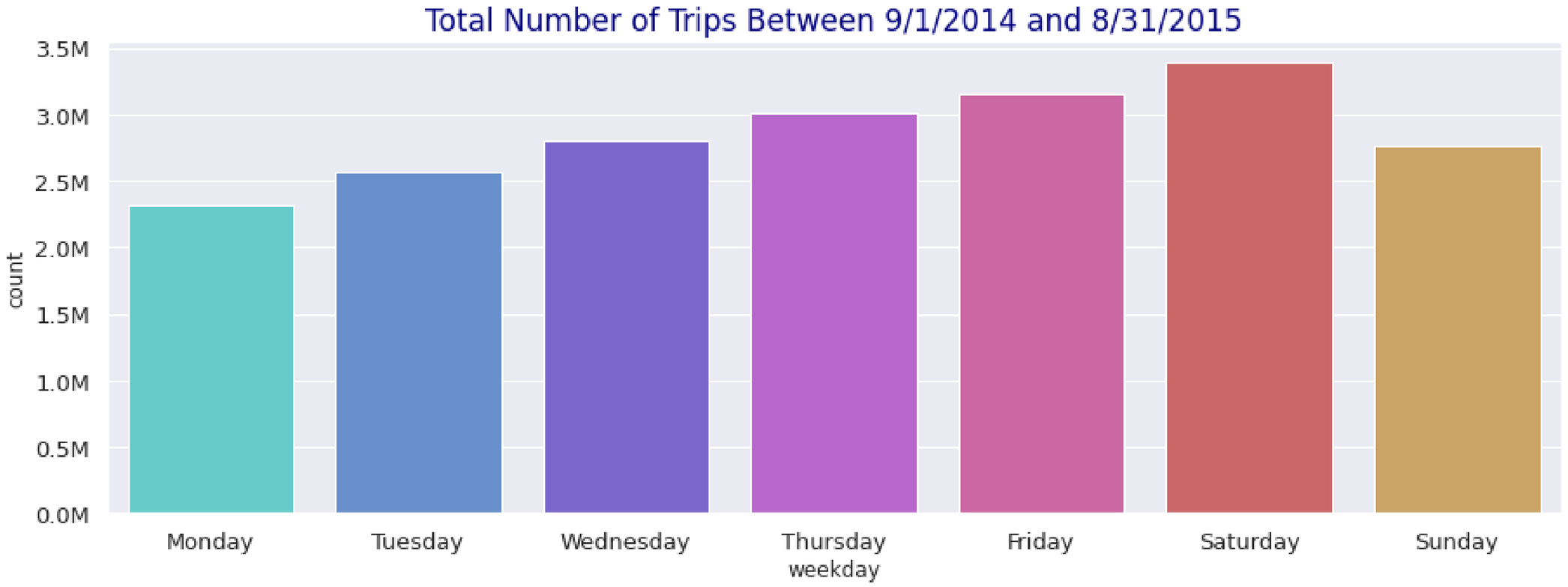
we have to Create a plot with the total number of trips per day, highlighting some changepoints associated with major holidays and other weather and touristic/cultural events.



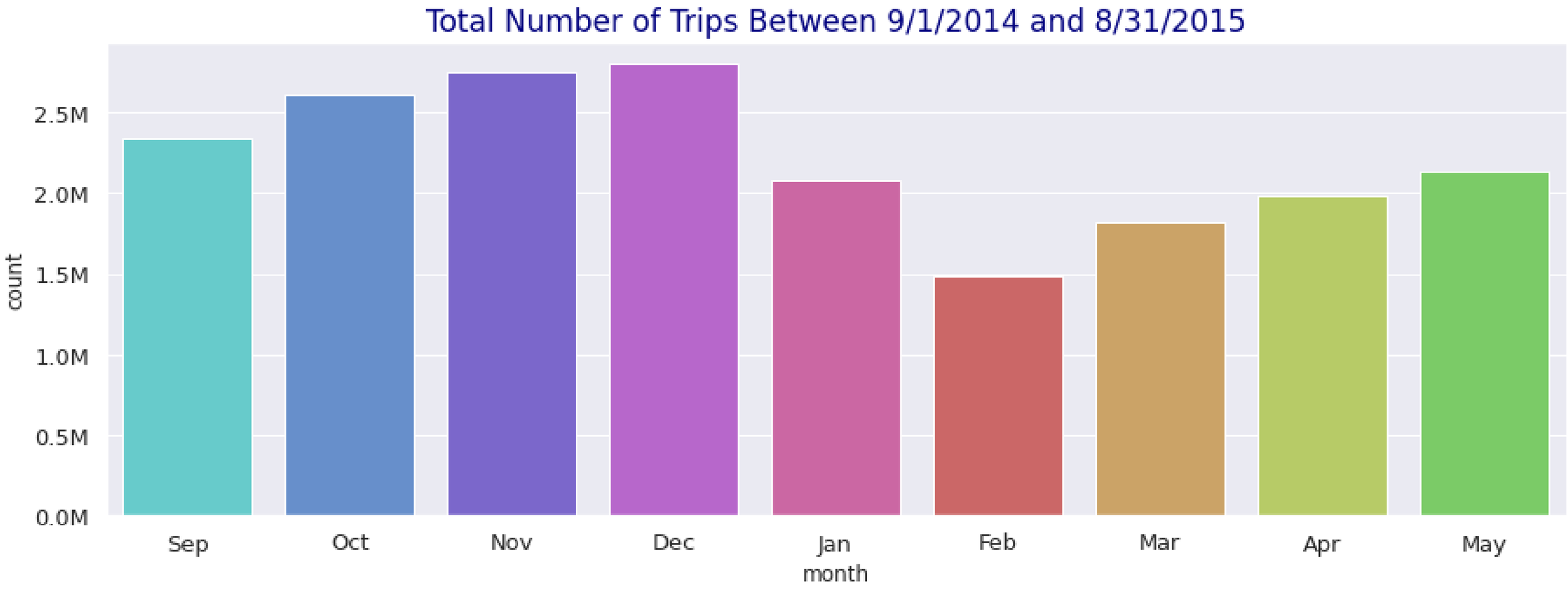
Total Number of Trips in Hours



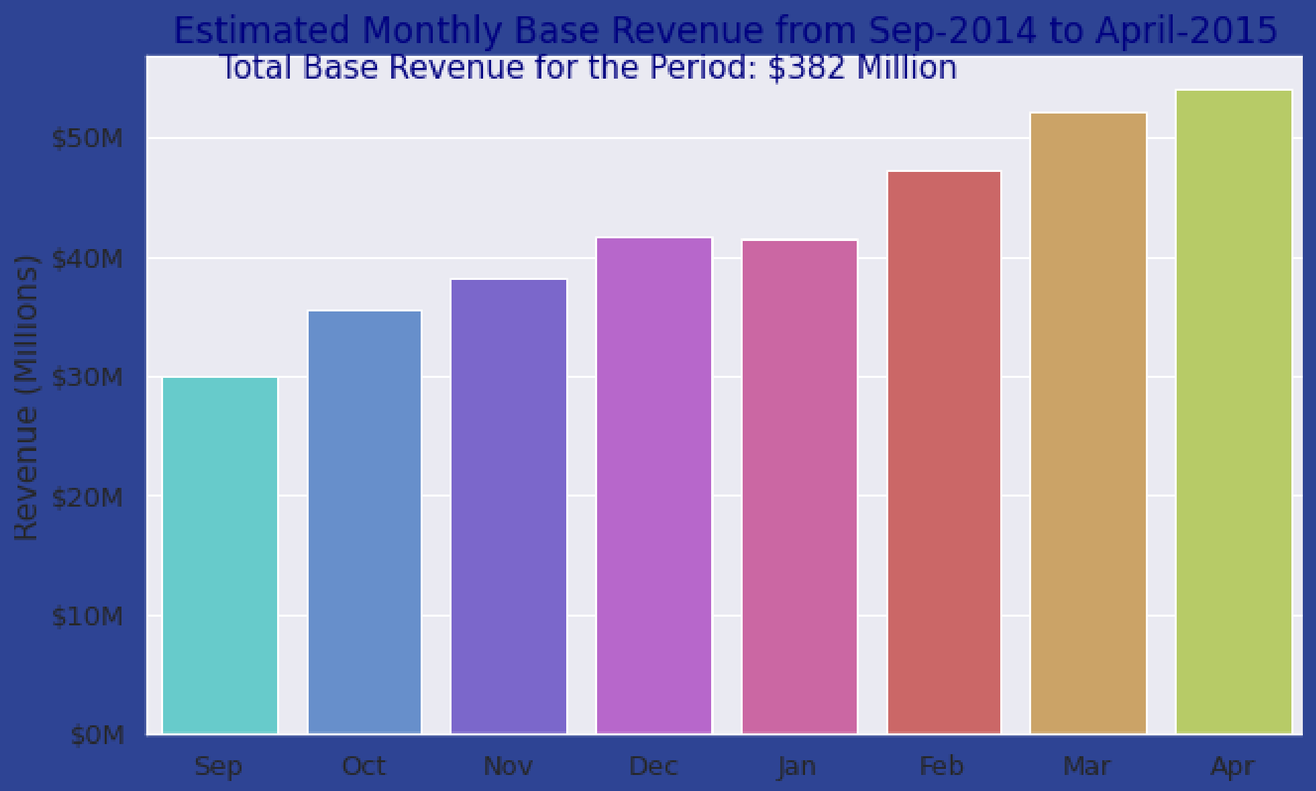
Total Number of Trips in Weekdays

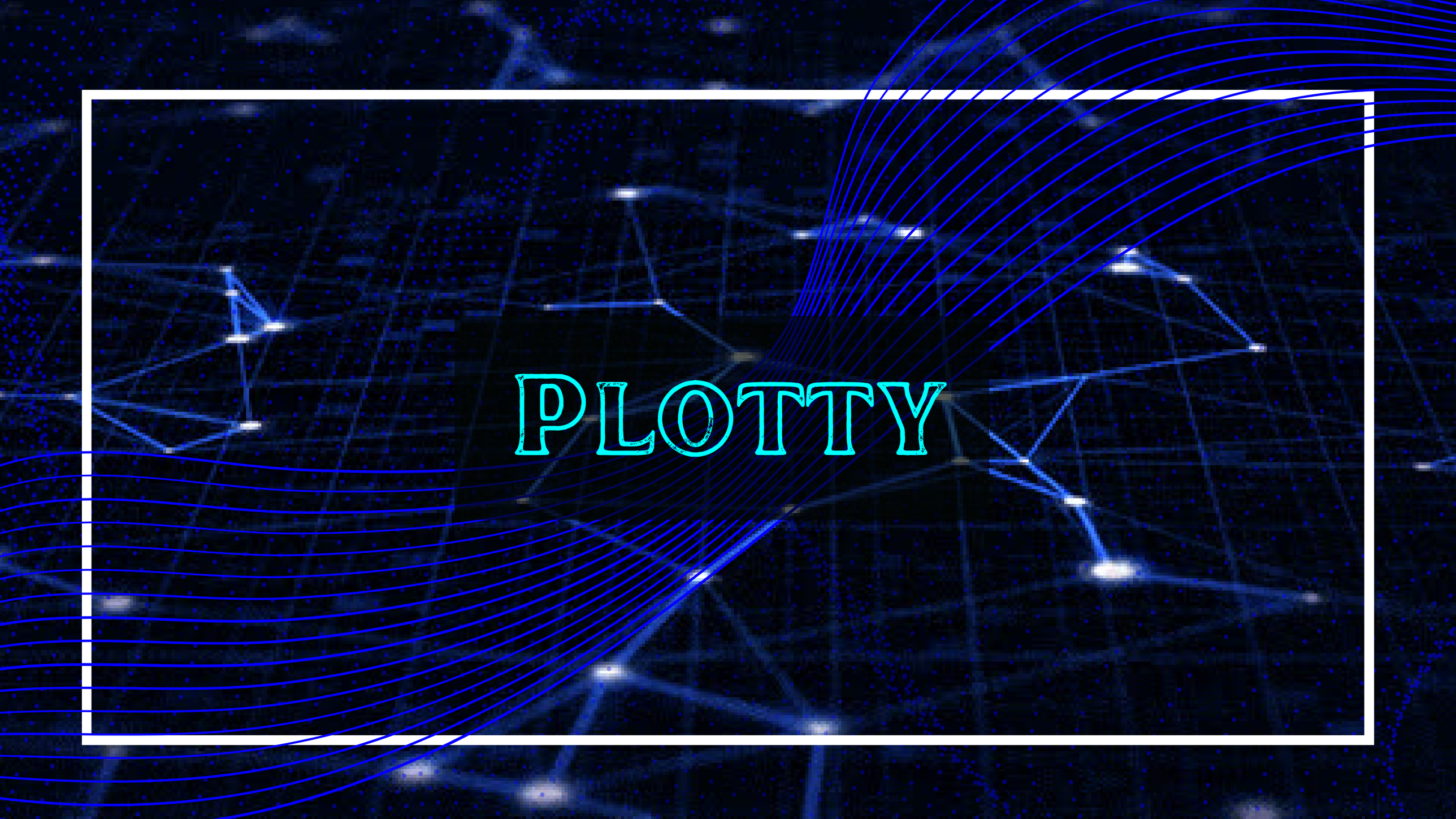


Total Number of Trips in Month

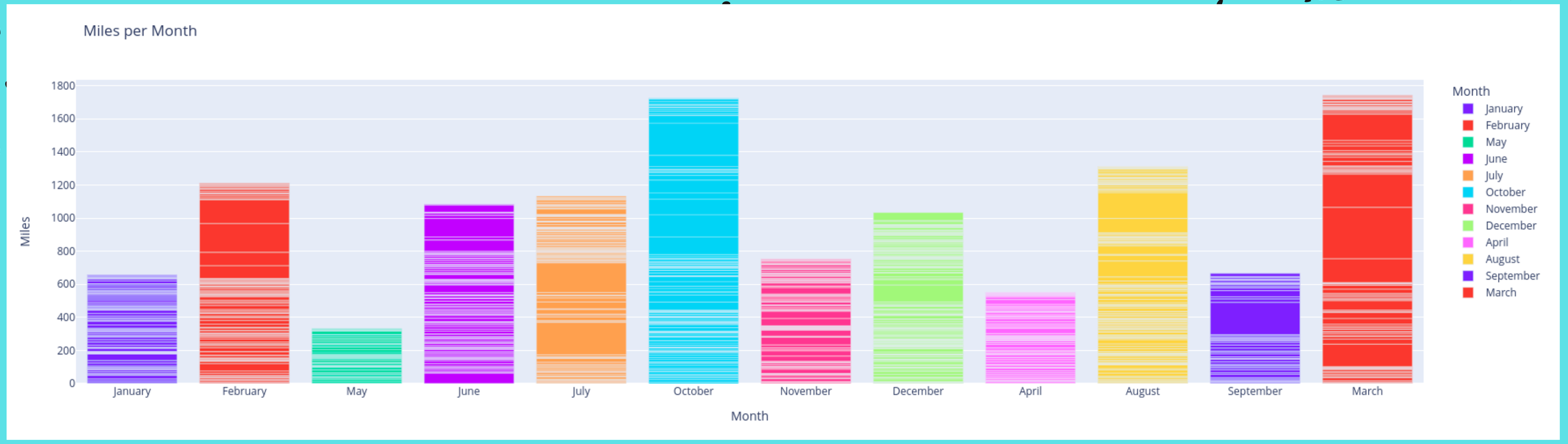


Monthly Base Revenue

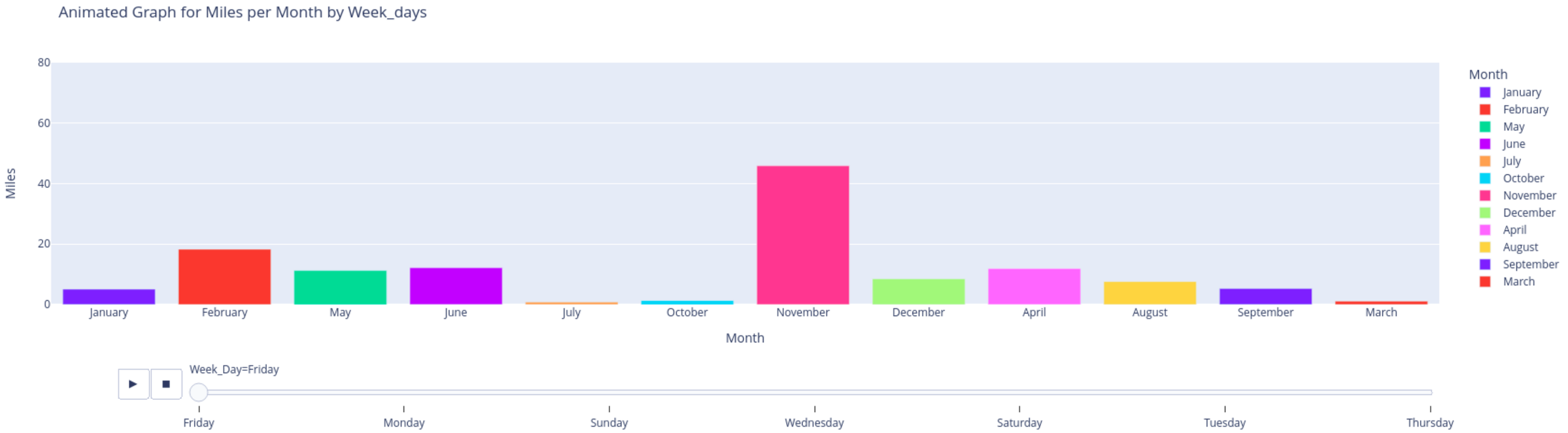



PLOTTY

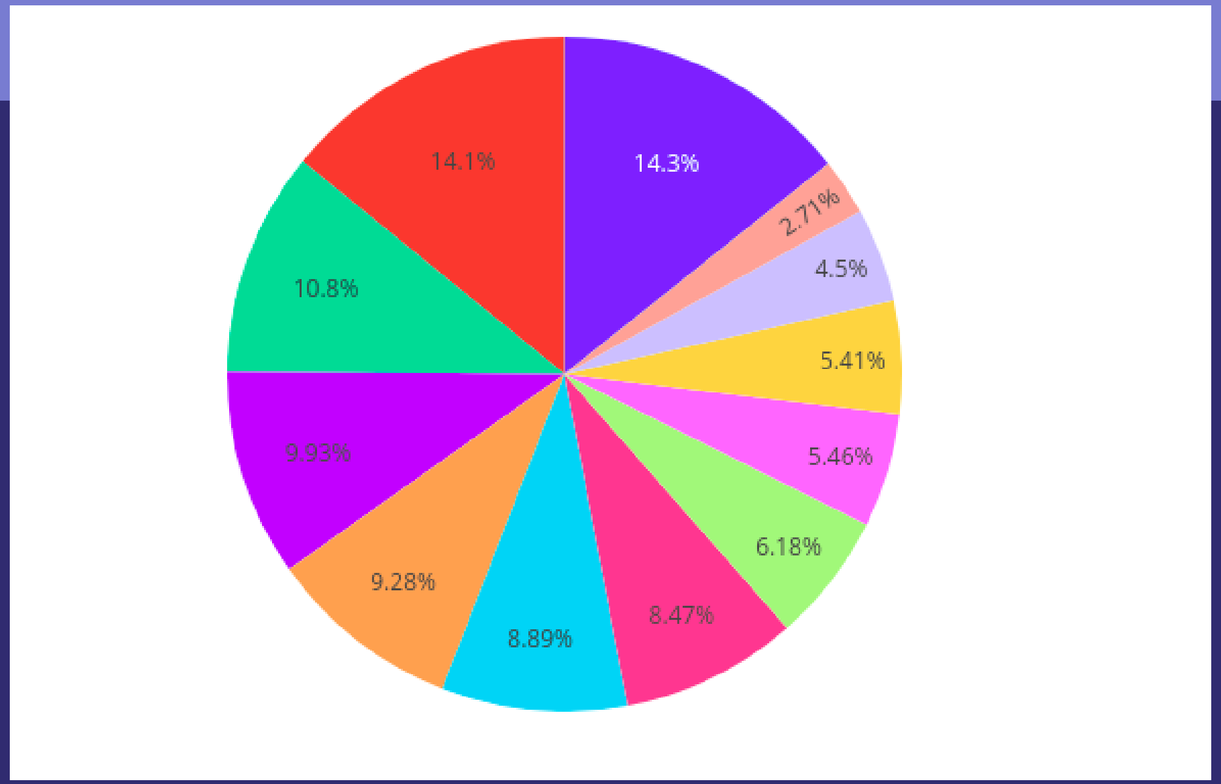
Miles Per Month



Miles Per Month for different days of the week

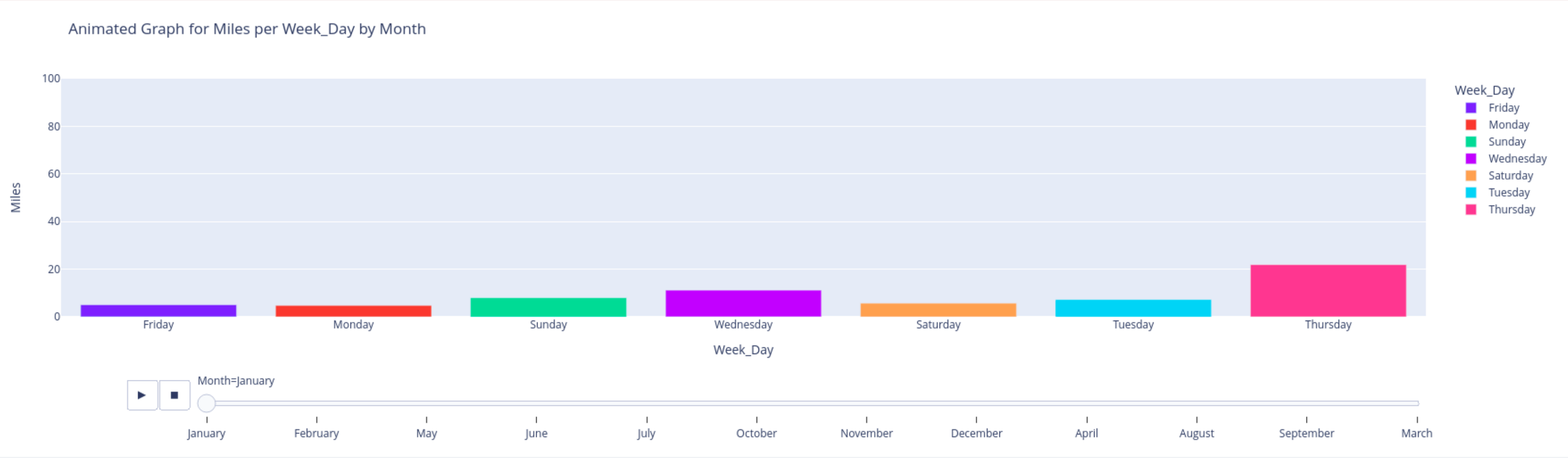


Pi Graph of "Miles" Per "Month"

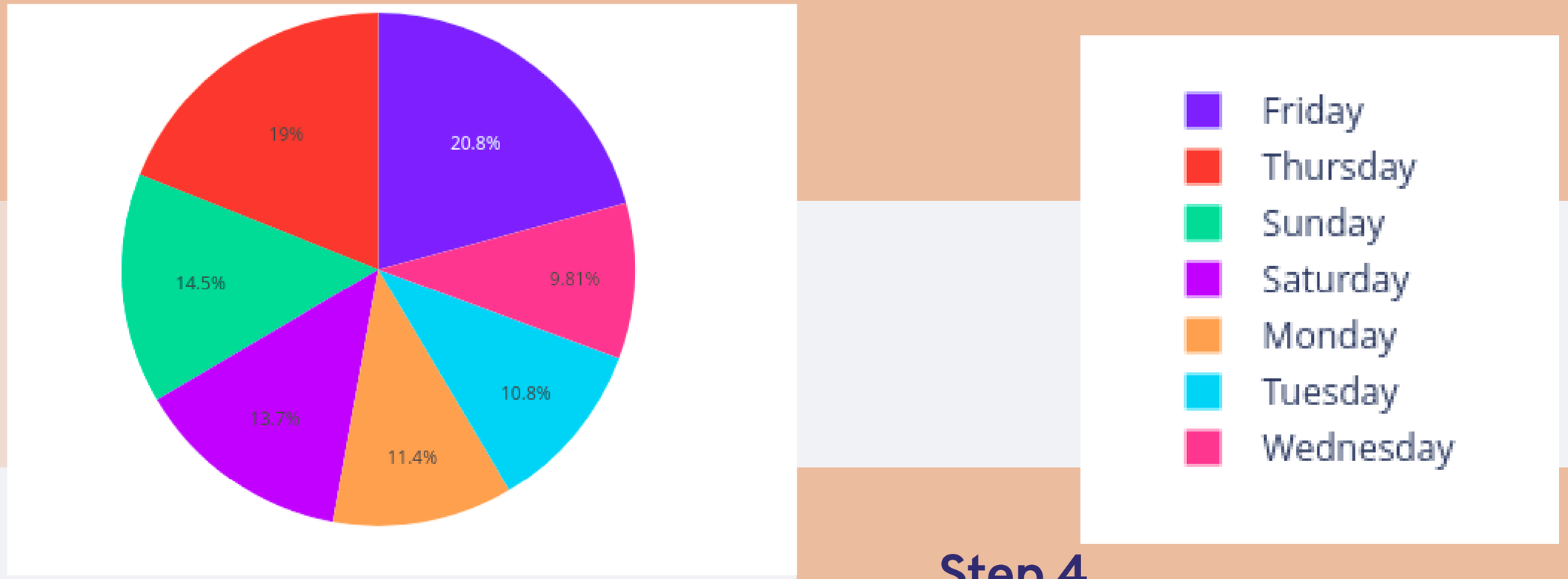


Visualize "Miles" and "Week_Day"

Animated Graph for "Miles" Per "Week_Day" By MONTH

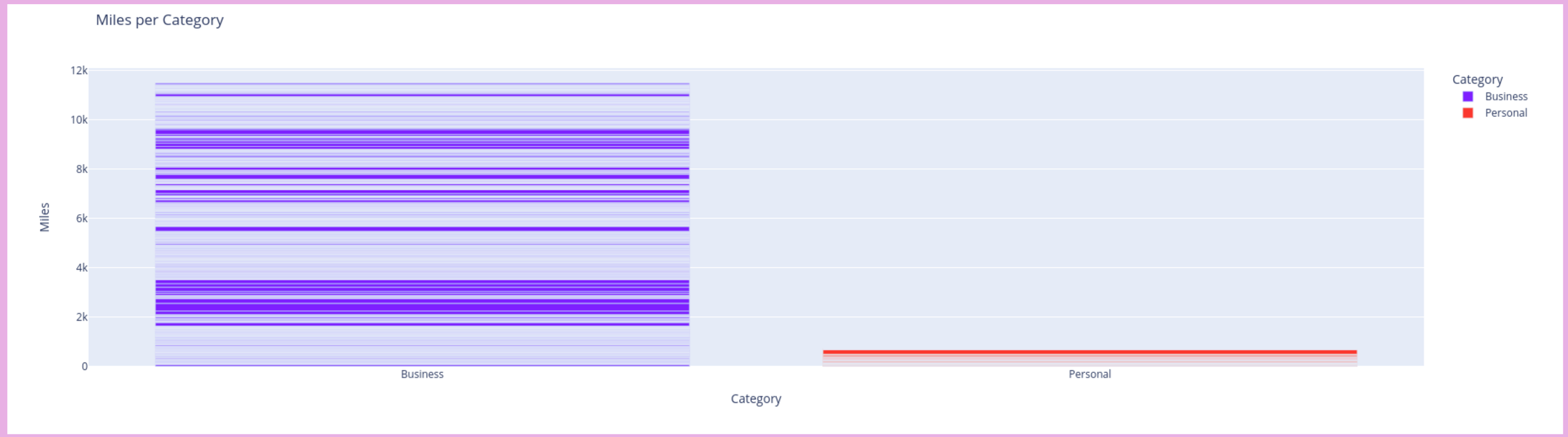


Pi Graph of "Miles" Per "Week_Day"



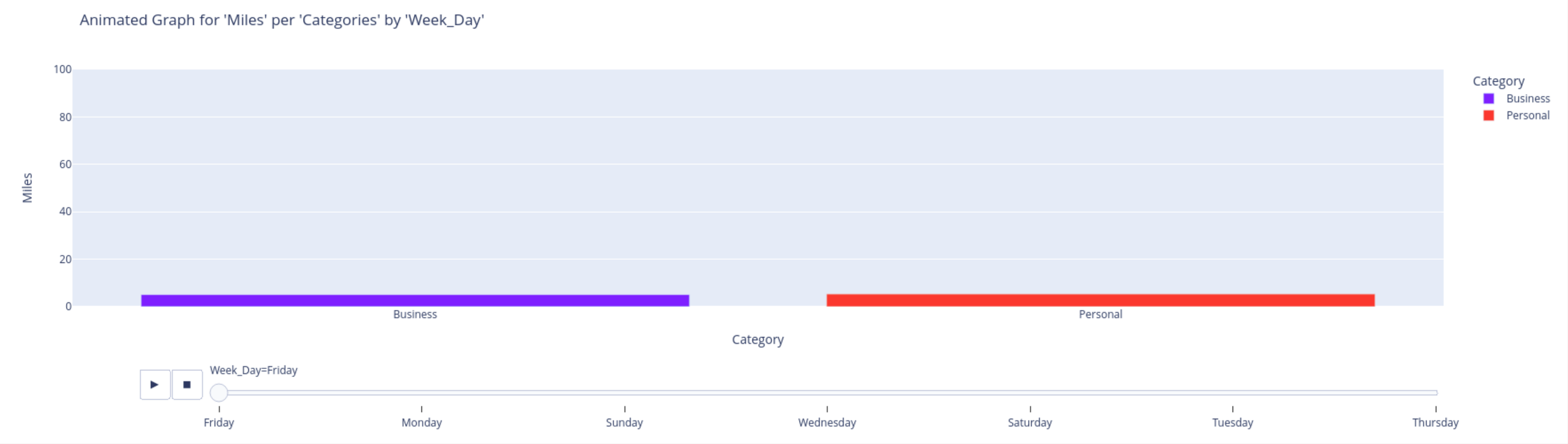
Step 4

Relation Between "Miles" and "Categories"

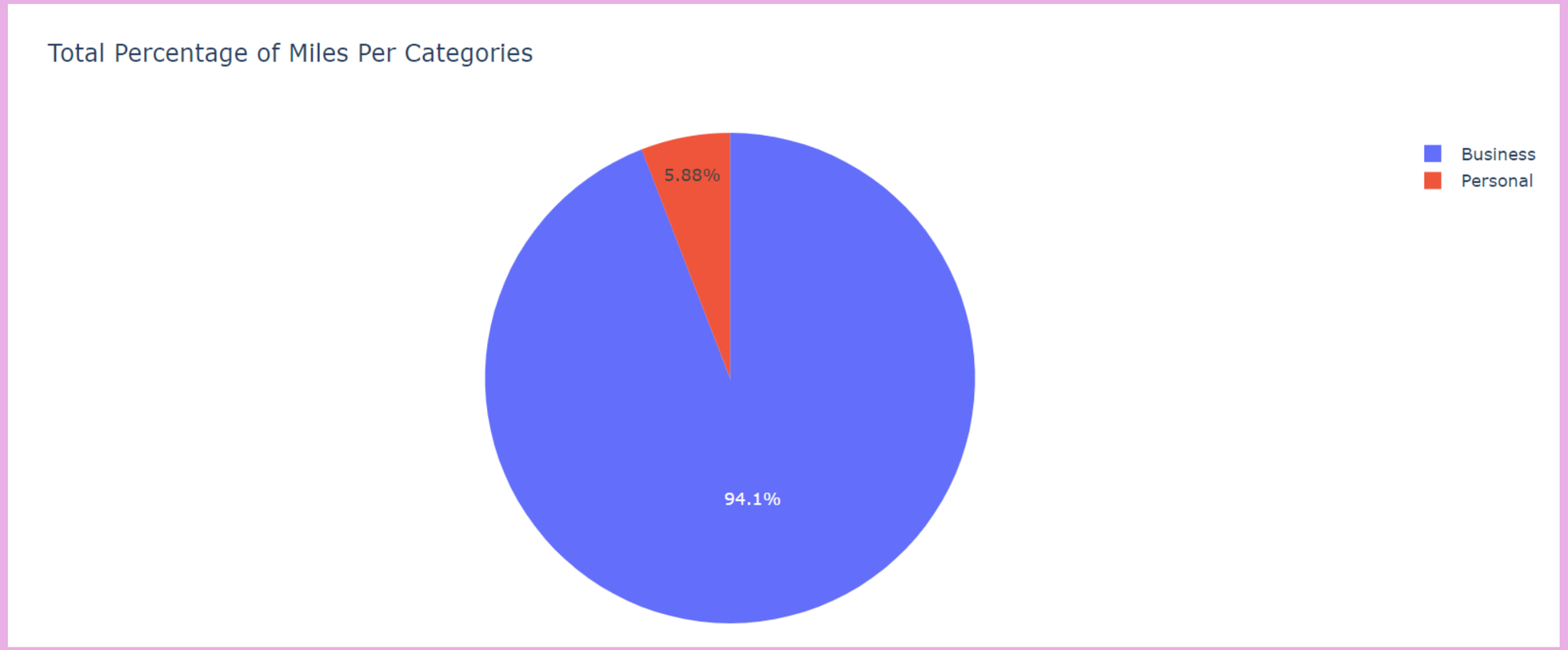


Relation of "Miles" and "Category" per "Month"

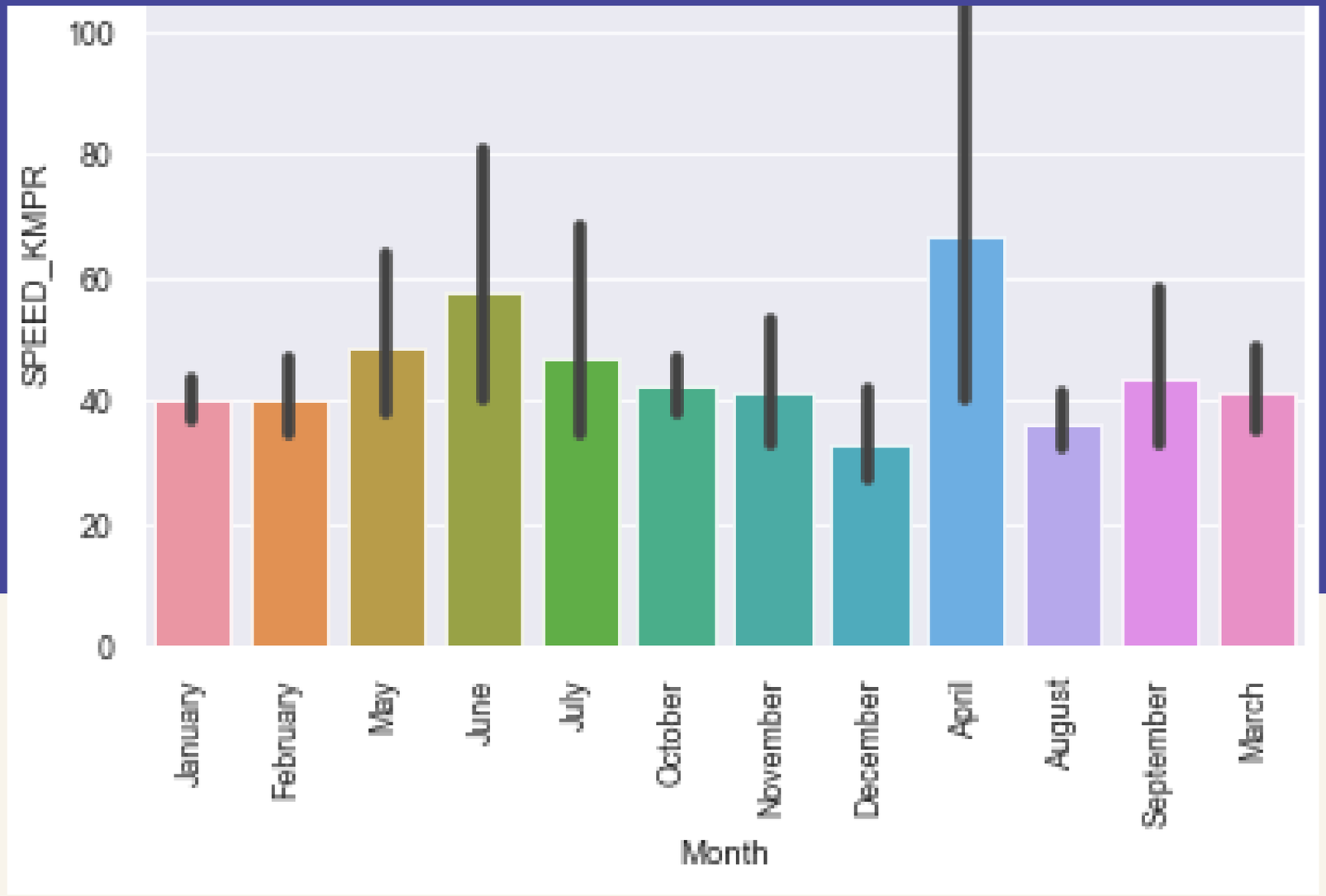
Relation of "Miles" and "Category" per "Week_Day"



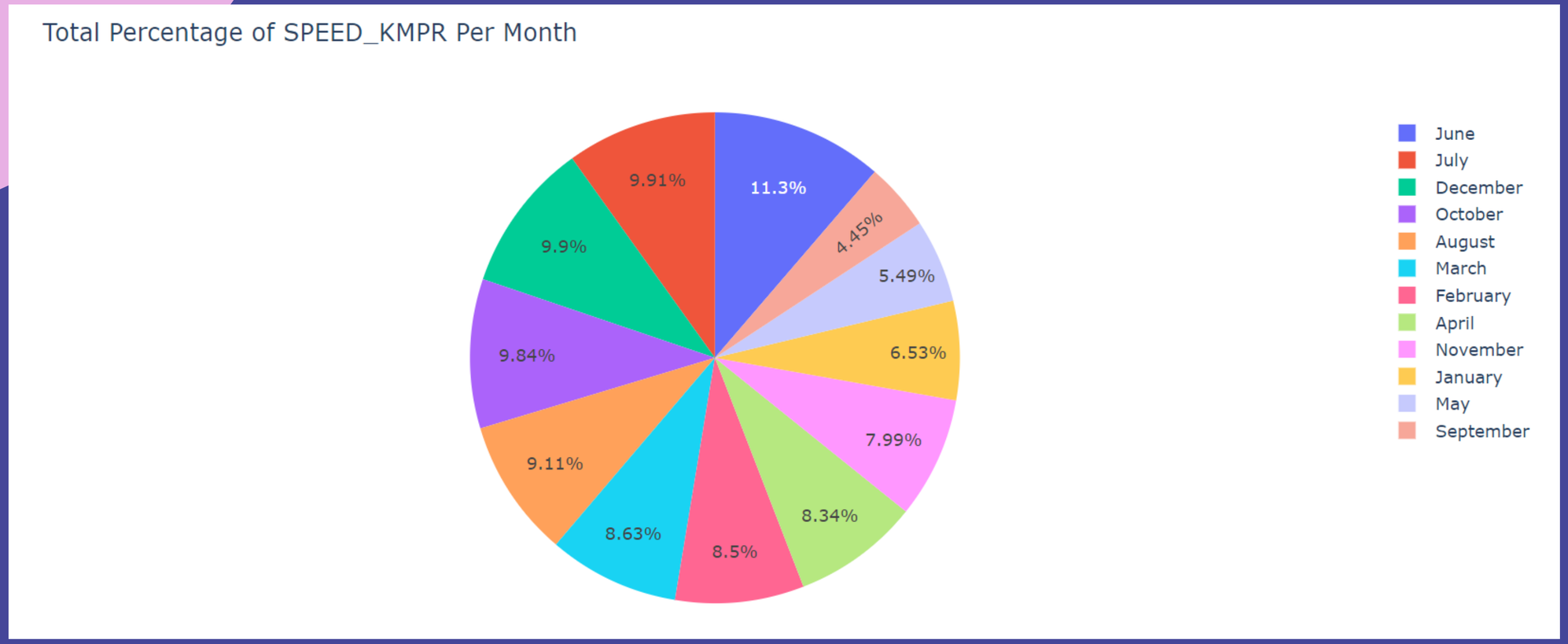
Pie Graph representing the % of Business vs Personal Travel



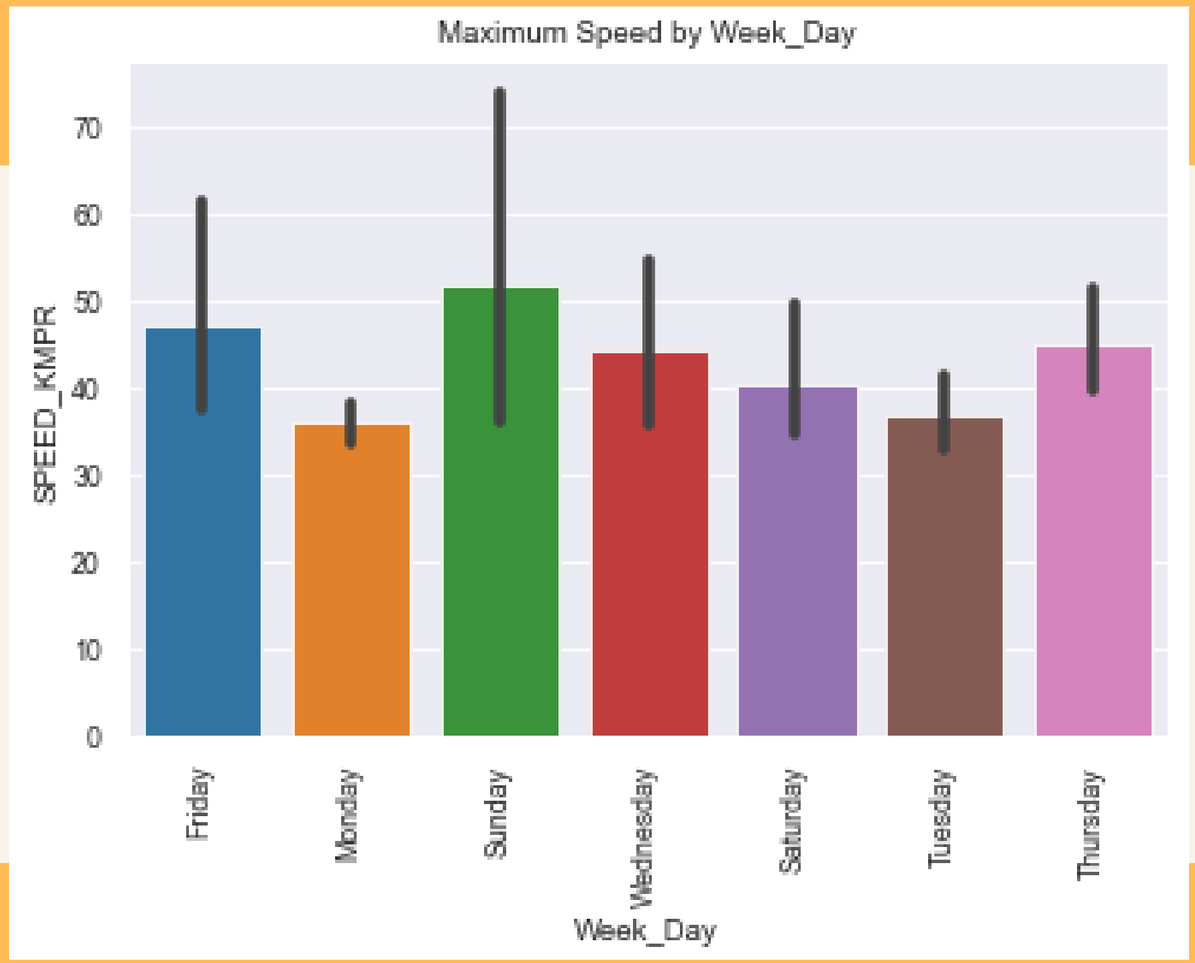
Relation between Speed and Month



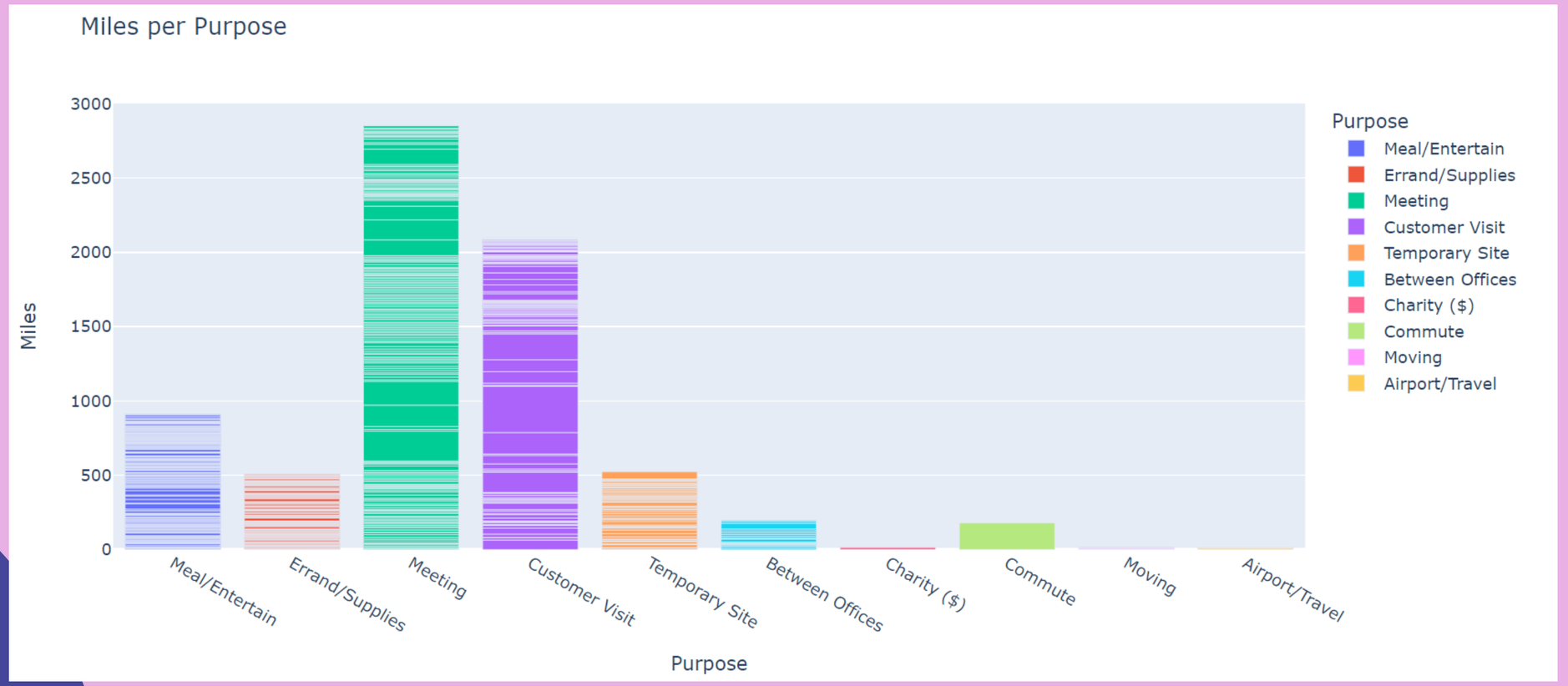
Percentage of Speed in Km/hr for each month



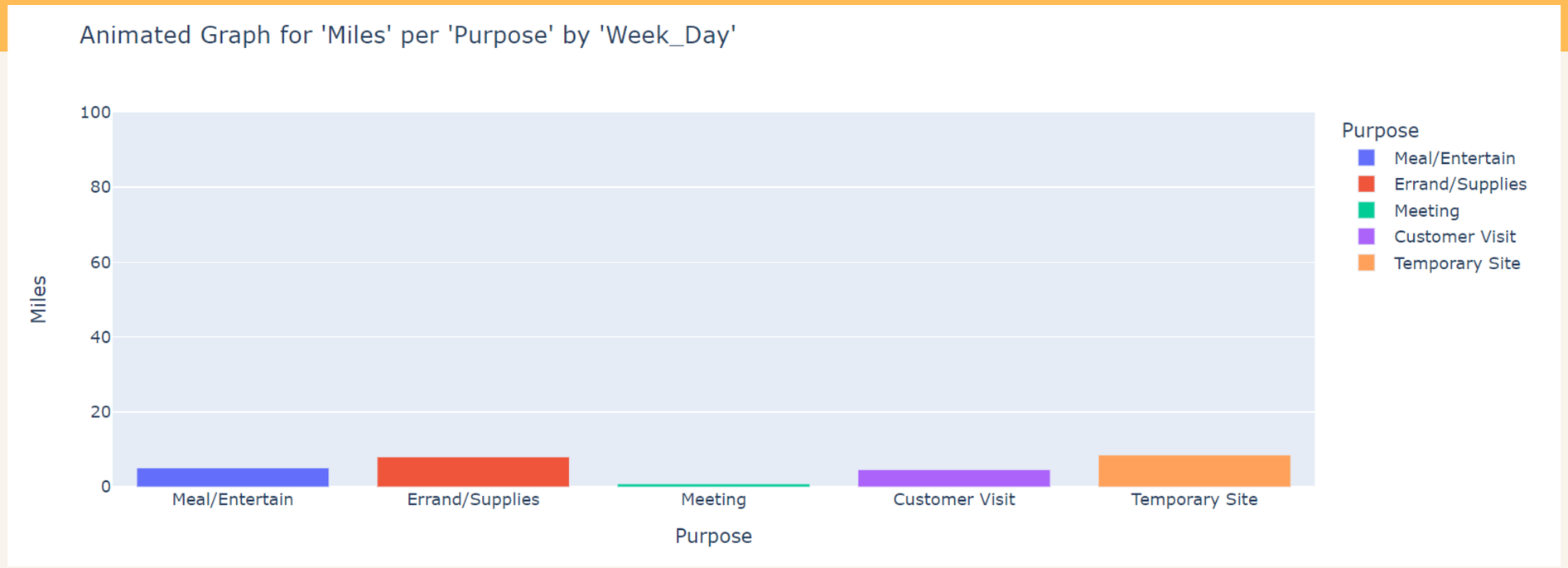
Maximum speed recorded in each Weekday



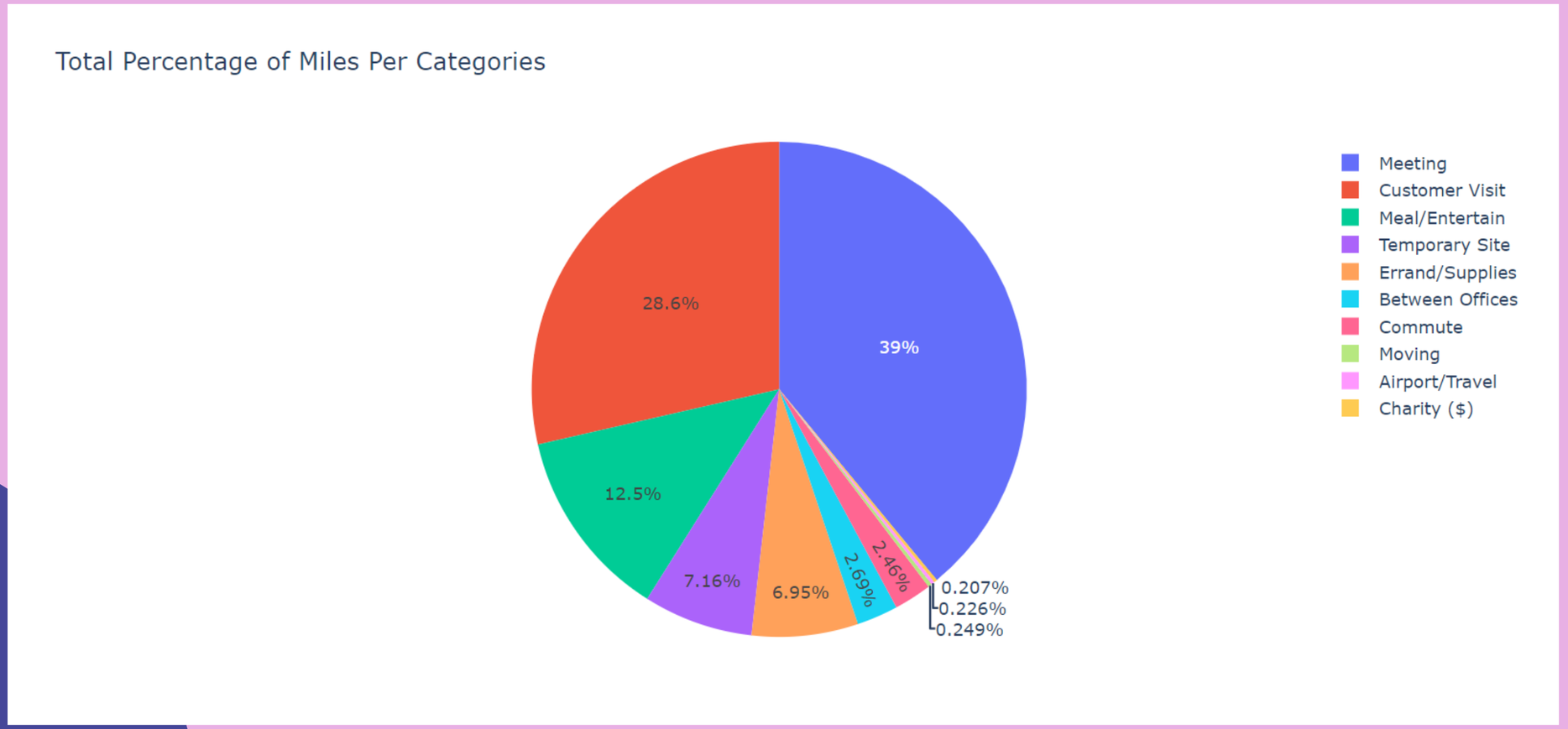
Miles travelled for a certain purpose



Miles per purpose for each weekday



percentage of miles travelled per category



location mapping