SUMMER BOOTCAMP 2024

DATA SCIENCE PROJECT
YASHIKA
UBER DATA ANALYSIS

INDEX

LIST OF TABLES

LIST OF FIGURES

PROBLEM STATEMENT

The data contains the details for the Uber rides across various boroughs (subdivisions) of New York City at an hourly level and attributes associated with weather conditions at that time.

BASIC EXPLORATION

1.First-Five Rows

	pickup_dt	borough	pickups	spd	vsb	temp	dewp	slp	pcp01	pcp06	pcp24	sd	hday
0	1/1/2015 1:00	Bronx	152.0	5.0	10.0	30.0	7.0	1023.5	0.0	0.0	0.0	0.0	Y
1	1/1/2015 1:00	Brooklyn	1519.0	5.0	10.0	NaN	7.0	1023.5	0.0	0.0	0.0	0.0	?
2	1/1/2015 1:00	EWR	0.0	5.0	10.0	30.0	7.0	1023.5	0.0	0.0	0.0	0.0	Y
3	1/1/2015 1:00	Manhattan	5258.0	5.0	10.0	30.0	7.0	1023.5	0.0	0.0	0.0	0.0	Y
4	1/1/2015 1:00	Queens	405.0	5.0	10.0	30.0	7.0	1023.5	0.0	0.0	0.0	0.0	Y

2. Last-Five Rows

	pickup_dt	borough	pickups	spd	vsb	temp	dewp	slp	pcp01	pcp06	pcp24	sd	hday
29096	30-06-2015 23:00	EWR	0.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29097	30-06-2015 23:00	Manhattan	3828.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29098	30-06-2015 23:00	Queens	580.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29099	30-06-2015 23:00	Staten Island	0.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N
29100	30-06-2015 23:00	NaN	3.0	7.0	10.0	75.0	65.0	1011.8	0.0	0.0	0.0	0.0	N

3. Shape

```
(29101, 14)
```

4. Info

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29101 entries, 0 to 29100
Data columns (total 19 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pickup_dt  11542 non-null  datetime64[ns]
 1   borough    26058 non-null  object
 2   pickups    29099 non-null  float64
 3   spd        29101 non-null  float64
 4   vsb        29101 non-null  float64
 5   temp       28742 non-null  float64
 6   dewp       29101 non-null  float64
 7   slp        29101 non-null  float64
 8   pcp01      29101 non-null  float64
 9   pcp06      29101 non-null  float64
 10  pcp24      29101 non-null  float64
 11  sd         29101 non-null  float64
 12  hday       29101 non-null  int64
 13  sd_bins    29101 non-null  category
 14  month      11542 non-null  float64
 15  season     11542 non-null  object
 16  hour       11542 non-null  float64
 17  weekday    11542 non-null  float64
 18  spd_bins   29101 non-null  category
dtypes: category(2), datetime64[ns](1), float64(13), int64(1), object(2)
```

Observations: The pickup_dt should be date or int type but it is object in info

5. Statistical Summary

	pickups	spd	vsb	temp	dewp	slp	pcp01	pcp06	pcp24	sd	hour
count	29099.000000	29101.000000	29101.000000	28742.000000	29101.000000	29101.000000	29101.0	29101.0	29101.000000	29101.000000	11542.000000
mean	282.060930	5.961970	9.487576	47.900262	30.823065	1017.810618	0.0	0.0	0.030223	1.861165	11.599203
std	386.669555	3.631521	0.906658	19.800541	21.283444	7.701187	0.0	0.0	0.049091	3.104397	6.908025
min	0.000000	0.000000	7.750000	0.000000	-16.000000	996.900000	0.0	0.0	0.000000	0.000000	0.000000
25%	1.000000	3.000000	9.100000	32.000000	14.000000	1012.500000	0.0	0.0	0.000000	0.000000	6.000000
50%	54.000000	6.000000	10.000000	46.500000	30.000000	1018.200000	0.0	0.0	0.000000	0.000000	12.000000
75%	449.000000	8.000000	10.000000	65.000000	50.000000	1022.900000	0.0	0.0	0.050000	2.958333	18.000000
max	1121.000000	15.500000	10.000000	89.000000	73.000000	1038.500000	0.0	0.0	0.125000	7.395833	23.000000

6. Null Values

	pickup_dt	borough	pickups	spd	vsb	temp	dewp	slp	pcp01	pcp06	pcp24	sd	hday	hour
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	True	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
29096	True	False	False	False	False	False	False	False	False	False	False	False	False	True
29097	True	False	False	False	False	False	False	False	False	False	False	False	False	True
29098	True	False	False	False	False	False	False	False	False	False	False	False	False	True
29099	True	False	False	False	False	False	False	False	False	False	False	False	False	True
29100	True	True	False	False	False	False	False	False	False	False	False	False	False	True

29101 rows × 14 columns

7. Duplicate Values

```
0        False
1        False
2        False
3        False
4        False
         ...
29096    False
29097    False
29098    False
29099    False
29100    False
Length: 29101, dtype: bool
```

8. Outliers and their Authenticity

```
<Axes: >
```

9. Anomalies or wrong entries

```
(pickup_dt     17559
 borough        3043
 pickups           2
 spd               0
 vsb               0
 temp            359
 dewp              0
 slp               0
 pcp01             0
 pcp06             0
 pcp24             0
 sd                0
 hday              0
 hour          17559
 dtype: int64,
 pickup_dt     object
 borough       object
 pickups      float64
 spd          float64
 vsb          float64
 temp         float64
 dewp         float64
 slp          float64
 pcp01        float64
 pcp06        float64
 pcp24        float64
 sd           float64
 hday          object
 hour         float64
 dtype: object)
```

PICKUP ANALYSIS

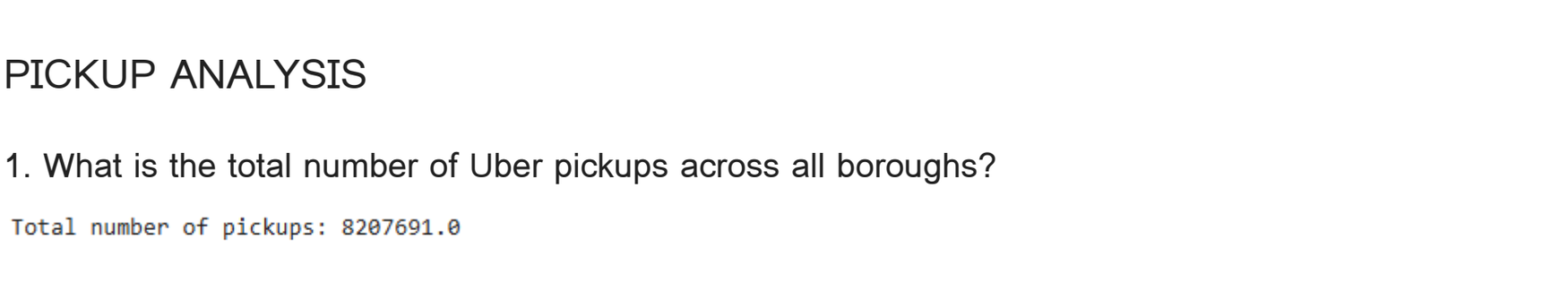
1. What is the total number of Uber pickups across all boroughs?

```
Total number of pickups: 8207691.0
```

2. Which borough has the highest average number of hourly pickups?

```
Manhattan
```

3. How do the number of pickups vary across different hours of the day?



4. Which day of the week has the highest number of pickups?

```
Day of the week with the highest number of pickups: 4.0
Number of pickups on the highest day: 856062.0
```

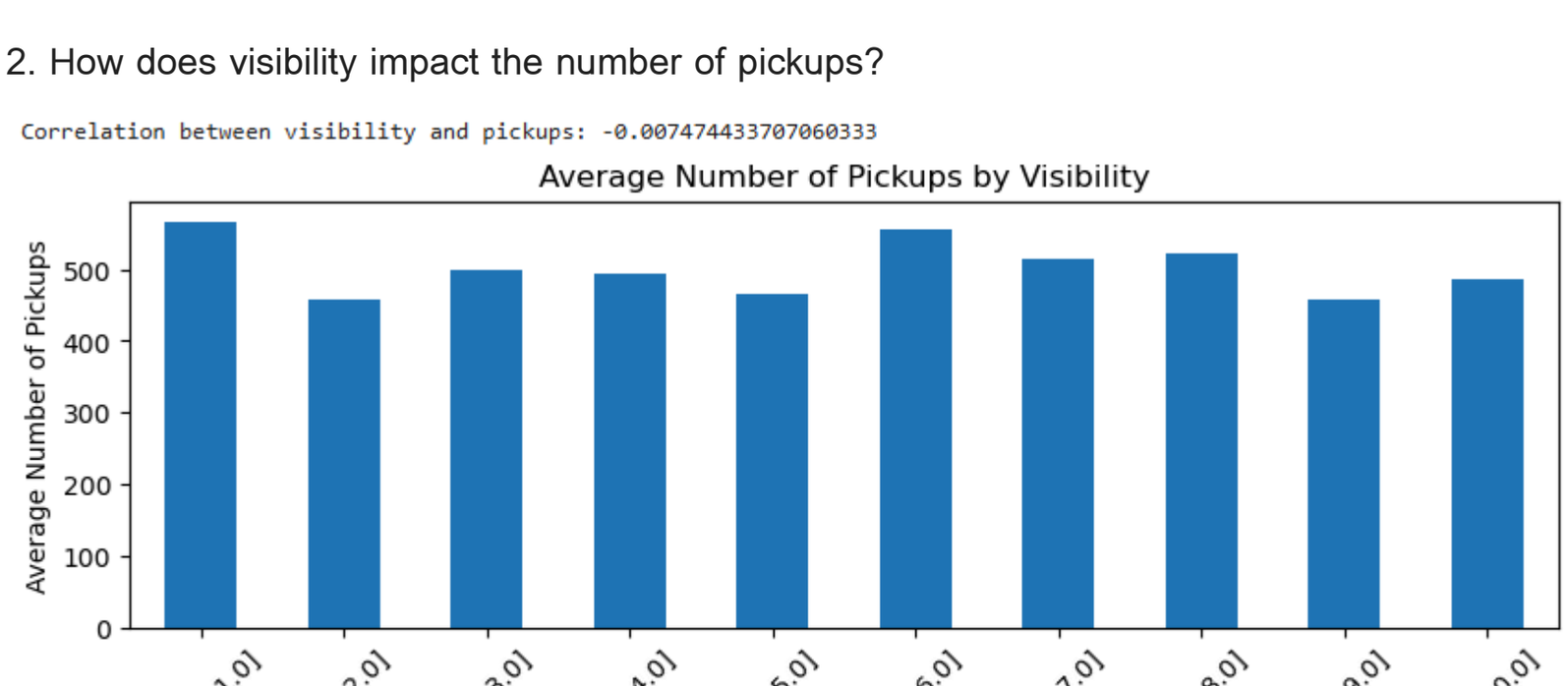
WEATHER IMPACT

1. What is the correlation between temperature and the number of pickups?

```
Correlation between temperature and pickups: 0.06851304320789024
```

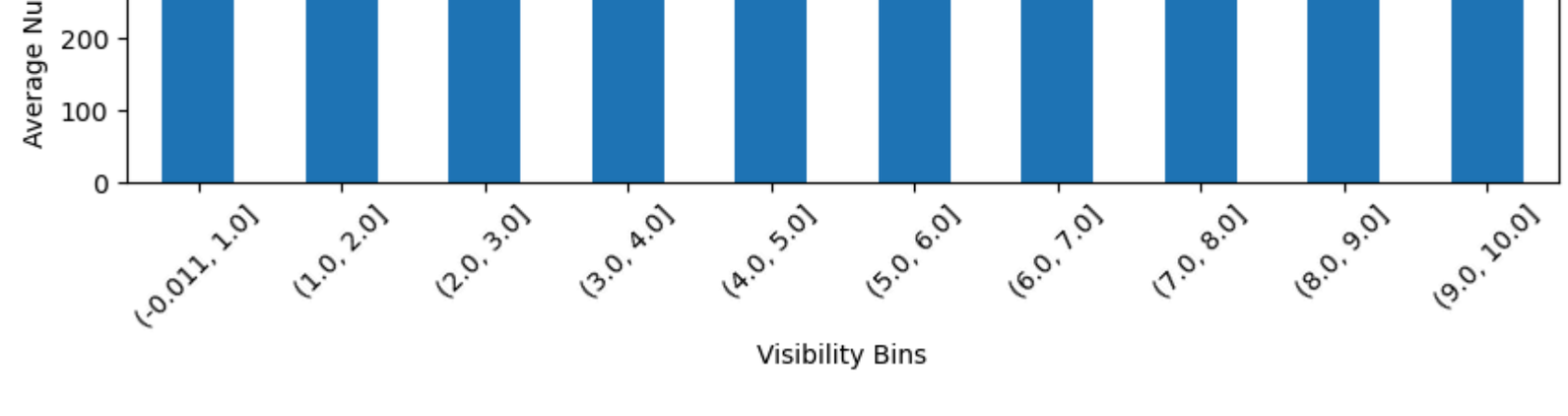
2. How does visibility impact the number of pickups?

```
Correlation between visibility and pickups: -0.007474433707060333
```



3. Is there a relationship between wind speed and the number of pickups?

```
Correlation between wind speed and pickups: 0.011110156997631699
```

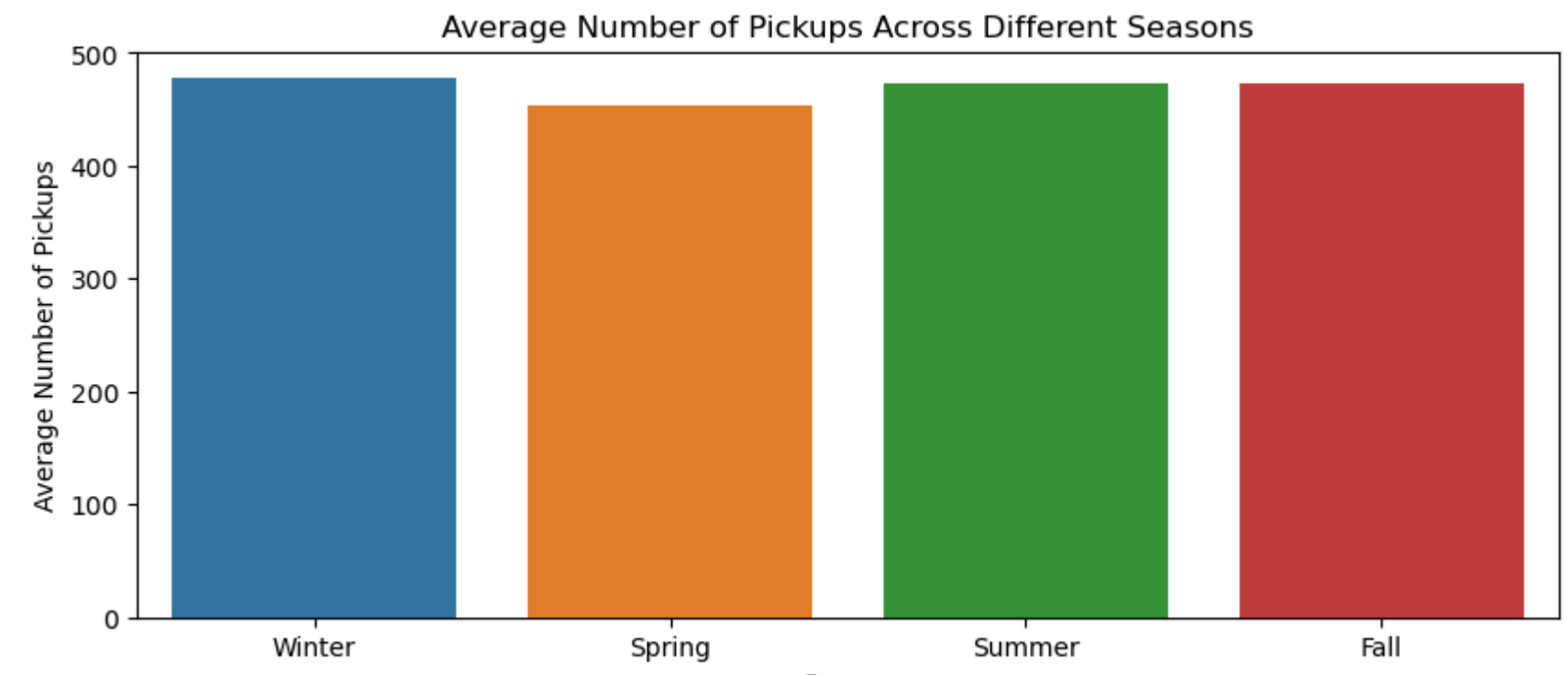


4. How does precipitation (1-hour, 6-hour, 24-hour) affect the number of pickups?

```
Correlation between 1-hour precipitation and pickups: -0.007967284671963636
Correlation between 6-hour precipitation and pickups: 0.0058164541143686495
Correlation between 24-hour precipitation and pickups: -0.005404637226941016
```

SEASONAL TRENDS

1. How do the number of pickups vary across different seasons (winter, spring, summer, fall)?



2. What is the average number of pickups during holidays compared to non-holidays?

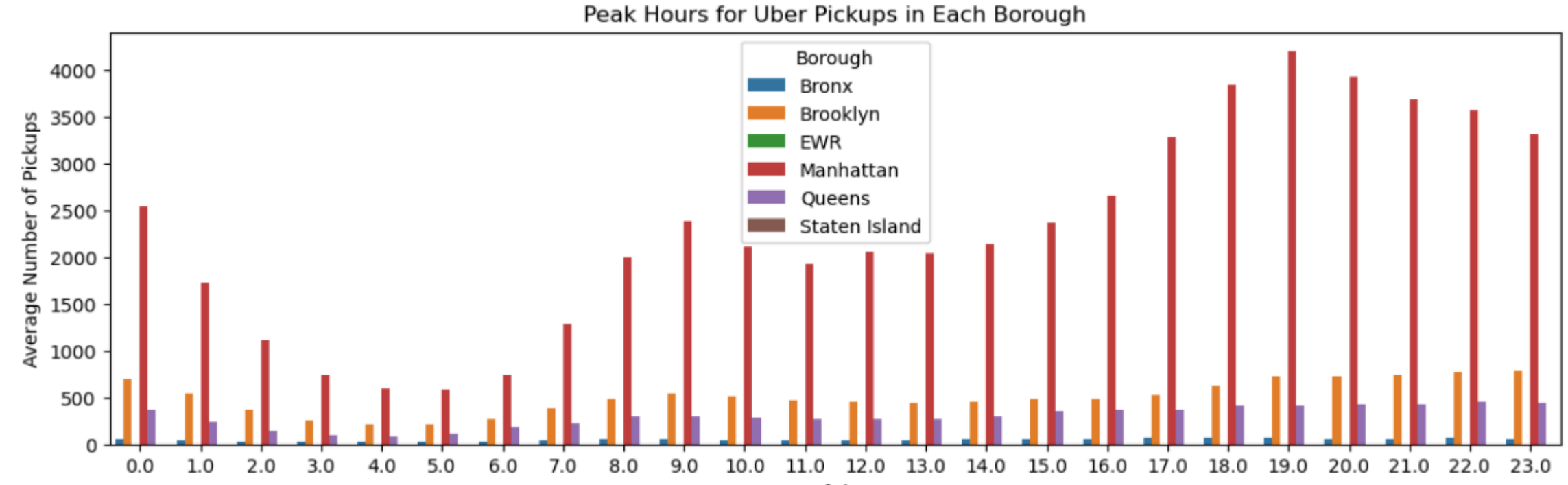
```
hday
0    490.236022
Name: pickups, dtype: float64
```

3. How does snow depth influence the number of pickups?


```
Correlation between snow depth and pickups: -0.008252751568239175
```

HOURLY TRENDS

1. What are the peak hours for Uber pickups in each borough?



2. How do the number of pickups change during rush hours (e.g., 7-9 AM, 5-7 PM)?

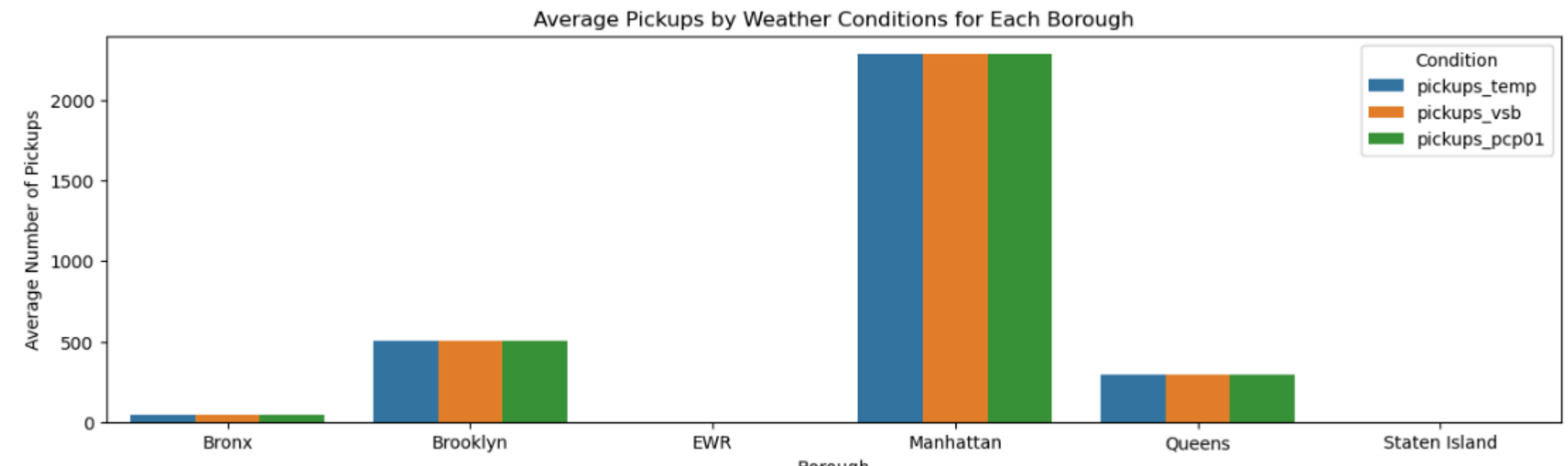
```
Average pickups during morning rush hours: hour
7.0    286.163934
8.0    424.556936
9.0    482.843943
Name: pickups, dtype: float64

Average pickups during evening rush hours: hour
17.0    626.764344
18.0    730.893443
19.0    797.774127
Name: pickups, dtype: float64
```

3. What is the average number of pickups during late-night hours (e.g., 12 AM - 4 AM)?

```
Average pickups during late-night hours:
 hour
0.0    549.811441
1.0    386.981013
2.0    253.736052
3.0    172.779221
4.0    140.923913
Name: pickups, dtype: float64
```

BOROUGH COMPARISON

1. How do pickup trends differ between boroughs during different weather conditions?



2. Which borough shows the highest increase in pickups during holidays?

```
Increase in Pickups During Holidays by Borough: borough
Brooklyn         67.344458
Queens           16.326933
Manhattan         8.288854
Bronx             6.529360
Staten Island     0.300787
EWR               0.018821
Name: pickups, dtype: float64
```

3. How does the number of pickups compare between weekdays and weekends for each borough?

```
Comparison of Pickups Between Weekdays and Weekends by Borough:
                  Weekday      Weekend
borough
Bronx            43.788925    52.797917
Brooklyn        444.393745   665.404167
EWR               0.028869     0.012500
Manhattan      2334.920609  2154.093750
Queens          285.127608   318.941667
Staten Island     1.380112     1.716667
```

WEATHER EXTREMES

1. How do extreme weather conditions (e.g., very high or very low temperatures, heavy rainfall, snowstorms) affect the number of pickups?

```
Impact of Extreme Weather Conditions on Pickups by Borough:
borough
Bronx              0.587522
Brooklyn          -3.296325
EWR                0.007290
Manhattan        179.240176
Queens            21.194939
Staten Island     -0.162415
Name: pickups, dtype: float64
```

2. What is the impact of visibility less than 1 mile on the number of pickups?

```
Impact of Visibility Less Than 1 Mile on Pickups by Borough:
borough
Bronx             -4.772583
Brooklyn         -59.038468
EWR                0.062637
Manhattan        680.338159
Queens            11.860018
Staten Island     -0.169306
Name: pickups, dtype: float64
```

DATA CORRELATIONS

1. Is there a correlation between sea level pressure and the number of pickups?

```
Correlation between sea level pressure and pickups: -0.0030808606955312957
```

2. How do different weather variables (temperature, dew point, wind speed, visibility) collectively impact the number of pickups

```
Correlation Matrix:
          pickups       spd       vsb      temp      dewp       slp     pcp01
pickups  1.000000  0.013278 -0.006679  0.059496  0.037832 -0.003081 -0.007967
spd      0.013278  1.000000  0.104570 -0.238851 -0.264187  0.007485  0.009288
vsb     -0.006679  0.104570  1.000000  0.042633 -0.241430  0.005404 -0.532323
temp     0.059496 -0.238851  0.042633  1.000000  0.881690 -0.184618 -0.078406
dewp     0.037832 -0.264187 -0.241430  0.881690  1.000000 -0.007252  0.084268
slp     -0.003081  0.007485  0.005404 -0.184618 -0.007252  1.000000 -0.002379
pcp01   -0.007967  0.009288 -0.532323 -0.078406  0.084268 -0.002379  1.000000
pcp06    0.005816 -0.032245 -0.172031 -0.042832  0.031678 -0.002919  0.227623
pcp24   -0.005405 -0.084814 -0.066955  0.045938  0.061509 -0.004155  0.063034
sd      -0.013425  0.012192 -0.095789 -0.508410 -0.436254 -0.006058  0.081803

            pcp06     pcp24        sd
pickups  0.005816 -0.005405 -0.013425
spd     -0.032245 -0.084814  0.012192
vsb     -0.172031 -0.066955 -0.095789
temp    -0.042832  0.045938 -0.508410
dewp     0.031678  0.061509 -0.436254
slp     -0.002919 -0.004155 -0.006058
pcp01    0.227623  0.063034  0.081803
pcp06    1.000000  0.205595  0.123141
pcp24    0.205595  1.000000  0.168233
sd       0.123141  0.168233  1.000000
```

3. .What is the relationship between holiday status and weather conditions on the number of pickups?

```
Average pickups by holiday status: hday
0    273.811279
1    283.902362
Name: pickups, dtype: float64
```

GROWTH INSIGHTS

1. Which weather conditions are most favorable for Uber pickups, and how can this information be used to optimize driver availability?

Solution:

1. Mild Temperature: Moderate temperatures, typically in the range where people find it comfortable to walk short distances to reach a pickup point without being overly hot or cold.

2. Clear or Mild Precipitation: Conditions where there is no precipitation or light rain are generally favorable. Heavy rain, snow, or storms can deter riders from requesting trips due to discomfort or safety concerns.

3. Good Visibility: Clear visibility conditions improve safety and ease of travel, which can lead to increased demand for Uber pickups.

4.Non-Extreme Weather: Avoiding extreme weather events such as heavy snowfall, blizzards, hurricanes, or severe thunderstorms is crucial. These conditions not only reduce demand but also pose safety risks for both drivers and passengers

2. Based on the data, what recommendations can be made to Uber to increase pickups during low-demand periods?

Solution:

1. Offer discounts or special promotions during low-demand times.

2. Implement lower prices during off-peak periods to attract more customers.

3. Provide bonuses or higher earnings to encourage driver availability during low-demand periods.

4. Run campaigns to promote rides during off-peak times.

5. Adjust driver availability and promotions based on favorable weather conditions.