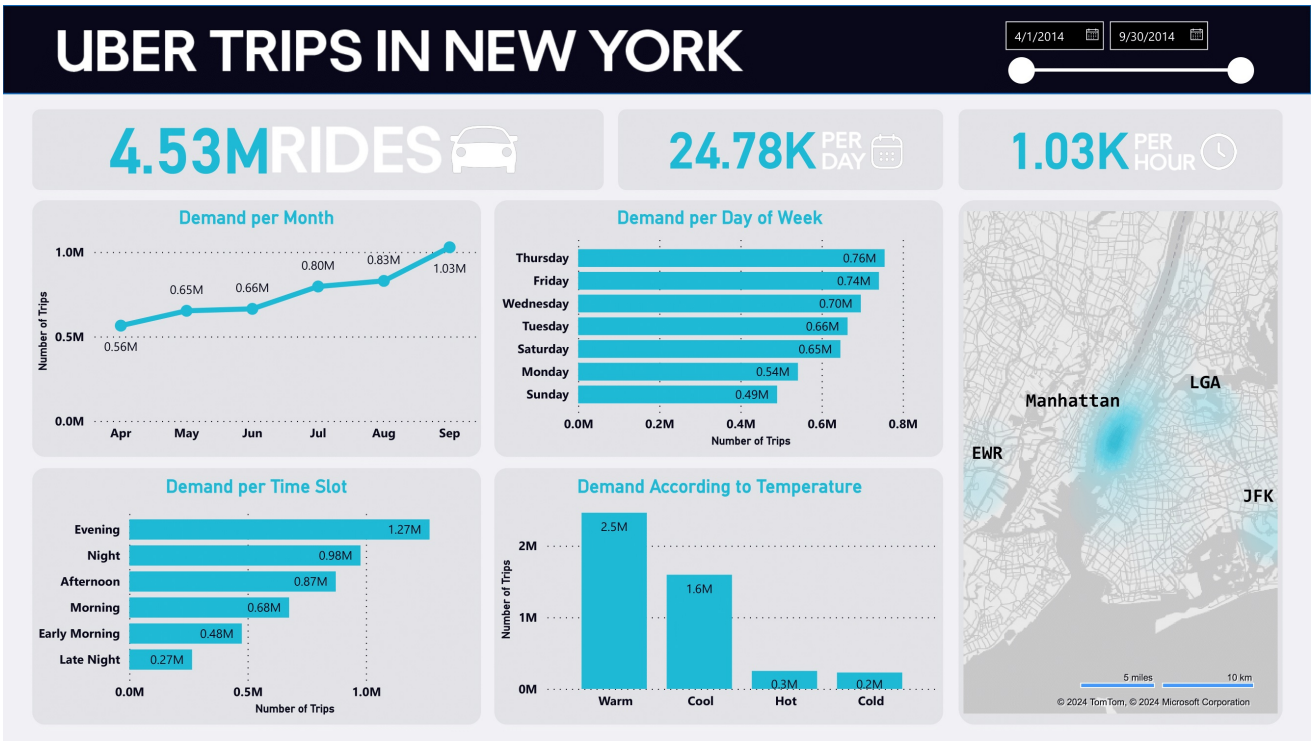
In [1]:
```
%load_ext sql
```



This project aims to explore and analyze Uber pickup data in New York during the period from April to September 2014. The primary objectives of this project are to gain insights into Uber pickup patterns, explore correlations between pickups and time, location, and weather conditions, and create visualizations to better understand the trends in the data.

Disclaimer: It's important to consider that this analysis is limited by the time frame of the data and that the patterns of trips may vary seasonally due to factors like weather conditions, holidays, or special events. Further analysis with data over a longer time frame could reveal more about seasonal trends.

How many trips were provided from April to September?

In [2]:
```
%%sql

SELECT COUNT(*) AS total_number_of_trips
FROM uber_pickups;
```

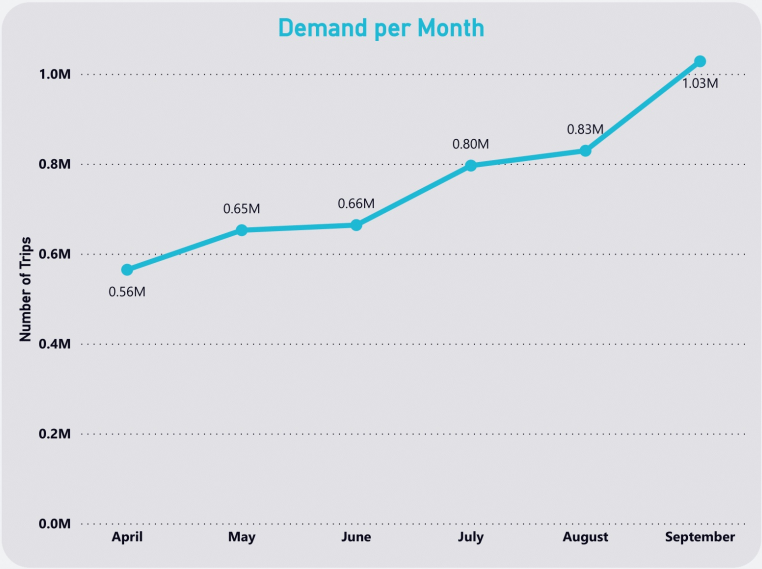
```
 * mysql://root:***@localhost/test
1 rows affected.
```

Out[2]:

total_number_of_trips
4534327

How does the demand vary across different months?

How does the demand vary across different months?



In [3]:
```
%%sql

SELECT MONTHNAME(date) AS month,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups
GROUP BY month
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
6 rows affected.
```

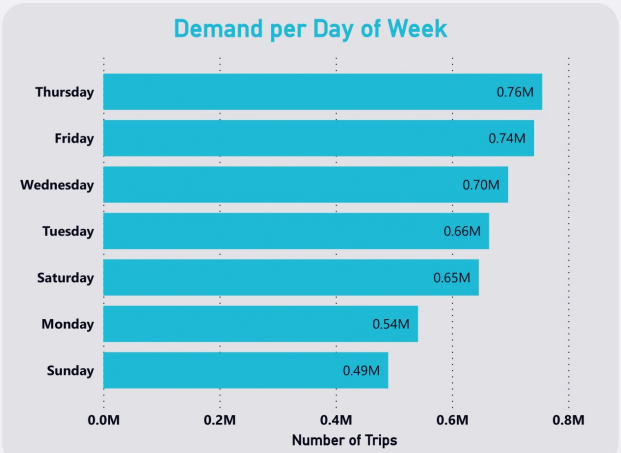
Out[3]:

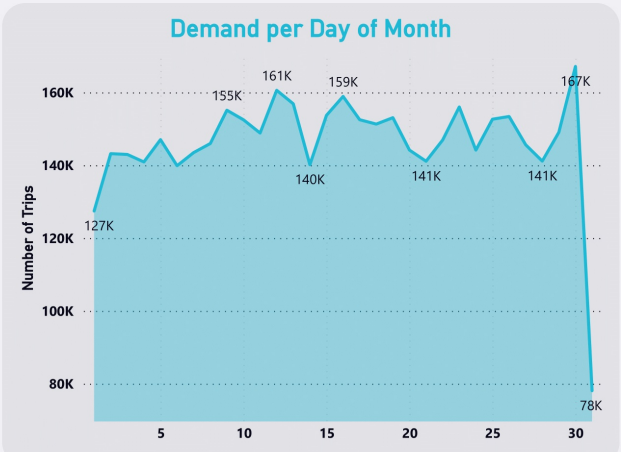
month	number_of_trips	percentage
September	1028136	22.67
August	829275	18.29
July	796121	17.56
June	663844	14.64
May	652435	14.39
April	564516	12.45

There is a clear pattern of increasing demand throughout the analyzed months.
The begining of the fall (September) witnessed the highest number of trips, followed by the summer months. This can be attributed to various factors, including better weather, tourists visiting the city, and people going out more frequently for leisure activities. Uber can use this seasonal information to allocate drivers more effectively, especially during peak months, and potentially adjust pricing strategies to meet increased demand. During the lower-demand months like April and September, Uber may want to focus on marketing campaigns or promotions to stimulate ridership.

Are there any specific days with higher demand?

Are there any specific days with higher demand?





Number of Trips in Top 10 Days

9/13/2014	9/5/2014	9/19/2014	9/6/2014	9/18/2014	9/12/2014	9/27/2014	9/20/2014	9/4/2014	9/25/2014
43205	42319	41017	40520	40274	39540	39468	38864	38360	38203

In [4]:
```
%%sql

SELECT DAYNAME(date) AS day,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups
GROUP BY day
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
7 rows affected.
```

Out[4]:

day	number_of_trips	percentage
Thursday	755145	16.65
Friday	741139	16.35
Wednesday	696488	15.36
Tuesday	663789	14.64
Saturday	646114	14.25
Monday	541472	11.94
Sunday	490180	10.81

Thursday and Friday stand out as the peak days for Uber rides. This also highlights that Uber experiences the highest demand on weekdays while the demand drops significantly on weekends, with Sunday having the lowest number of trips, which could be attributed to factors such as people commuting to work/school on weekdays moreso than leisure trips on weekends.
As for weekdays, Wednesday and Tuesday have fewer trips compared to Thursday and Friday. This might reflect a midweek peak in demand, possibly related to work or midweek events. Monday consistently has the lowest number of Uber trips. This can be explained by the "Monday blues" phenomenon, where people may be less likely to go out or travel for leisure immediately after the weekend.

In [5]:
```
%%sql

SELECT DAYOFMONTH(date) AS day,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups
GROUP BY day
ORDER BY number_of_trips DESC
LIMIT 10;
```

```
 * mysql://root:***@localhost/test
10 rows affected.
```

Out[5]:

day	number_of_trips	percentage
30	167160	3.69
12	160606	3.54
16	158921	3.50
13	156892	3.46
23	156032	3.44
9	155135	3.42
15	153726	3.39
26	153405	3.38
19	153088	3.38
25	152667	3.37

The demand for Uber rides appears to be relatively evenly distributed throughout the month, with the top days all having a similar percentage of total trips. This indicates a consistent need for ride-sharing services across various days of the month.

In [6]:
```
%%sql

SELECT date, DAYNAME(date) AS day,
       COUNT(*) AS number_of_trips
FROM uber_pickups
GROUP BY date
ORDER BY number_of_trips DESC
LIMIT 10;
```

```
 * mysql://root:***@localhost/test
10 rows affected.
```

Out[6]:

date	day	number_of_trips
2014-09-13	Saturday	43205
2014-09-05	Friday	42319
2014-09-19	Friday	41017
2014-09-06	Saturday	40520
2014-09-18	Thursday	40274
2014-09-12	Friday	39540
2014-09-27	Saturday	39468
2014-09-20	Saturday	38864
2014-09-04	Thursday	38360
2014-09-25	Thursday	38203

All of the top 10 days fall in September, which can be explained by the fact that September is a transitional month. Schools and universities usually start their academic years, people return from vacations, and routines resume after the summer break. This period of change can lead to increased transportation needs, especially in a bustling city like New York.
Moreover, these high-demand days are spread across Saturdays, Fridays, and Thursdays, indicating that weekends and the days leading up to the weekend consistently experience higher demand.

In [7]:
```
%%sql

SELECT COUNT(*) / COUNT(DISTINCT date) AS average_trips_per_day
FROM uber_pickups;
```

```
 * mysql://root:***@localhost/test
1 rows affected.
```

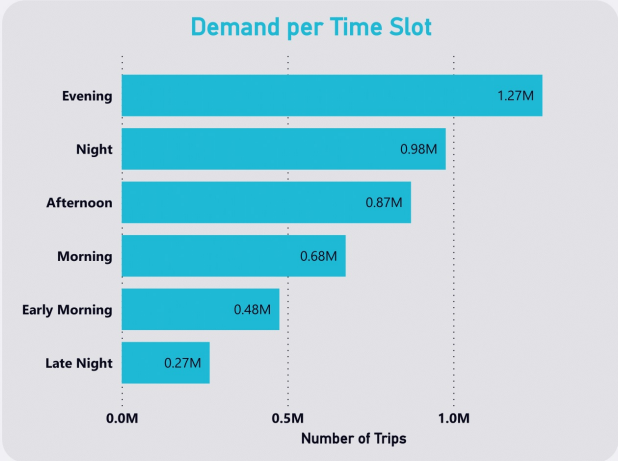
Out[7]:

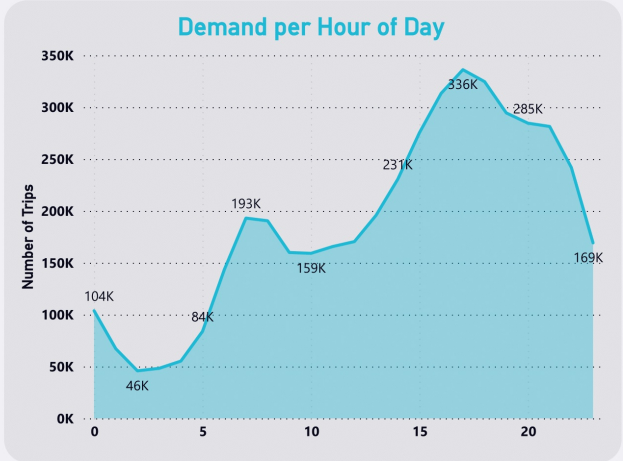
average_trips_per_day
24777.7432

According to 2014 Taxicab Fact Book, New York Yellow Taxis provided an average of 485,000 trip per day in 2014, which is approximately 19 times more trips per day than Uber. The high demand for Yellow Taxis suggests that there might be specific factors contributing to their popularity, such as availability, pricing, or customer preference. Further analysis is required to identify these factors.

What are the peak hours for Uber pickups?

What are the peak hours for Uber pickups?





In [8]:
```
%%sql

SELECT hour,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups
GROUP BY hour
ORDER BY number_of_trips DESC
LIMIT 5;
```

```
 * mysql://root:***@localhost/test
5 rows affected.
```

Out[8]:

hour	number_of_trips	percentage
17	336190	7.41
18	324679	7.16
16	313400	6.91
19	294513	6.50
20	284604	6.28

Peak demand occurs during late afternoon and early evening hours.

In [9]:
```
%%sql

WITH hour_slots AS (
    SELECT CASE
            WHEN hour BETWEEN 0 AND 3 THEN 'Late Night'
            WHEN hour BETWEEN 4 AND 7 THEN 'Early Morning'
            WHEN hour BETWEEN 8 AND 11 THEN 'Morning'
            WHEN hour BETWEEN 12 AND 15 THEN 'Afternoon'
            WHEN hour BETWEEN 16 AND 19 THEN 'Evening'
            WHEN hour BETWEEN 20 AND 23 THEN 'Night'
           END AS time_slot
    FROM uber_pickups
)

SELECT time_slot,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM hour_slots
GROUP BY time_slot
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
6 rows affected.
```

Out[9]:

time_slot	number_of_trips	percentage
Evening	1268782	27.98
Night	977112	21.55
Afternoon	872420	19.24
Morning	675322	14.89
Early Morning	475476	10.49
Late Night	265215	5.85

Uber experiences peak demand during the evening hours (4 PM to 7 PM), and night hours (8 PM to 11 PM) also see significant demand. This suggests that people often use Uber services for evening outings, dinner, and late-night activities.
Early morning (4 AM to 7 AM) and late night (12 AM to 3 AM) periods, understandably, exhibit lower demand.

In [10]:
```
%%sql

WITH peak_hours AS (
    SELECT CASE
            WHEN hour BETWEEN 7 AND 10 THEN 'Morning Rush'
            WHEN hour BETWEEN 16 AND 19 THEN 'Evening Rush'
            ELSE 'Off-peak'
           END AS time_slot
    FROM uber_pickups
)

SELECT time_slot,
       COUNT(*) AS number_of_trips,
       ROUND(COUNT(*) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM peak_hours
GROUP BY time_slot
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
3 rows affected.
```

Out[10]:

time_slot	number_of_trips	percentage
Off-peak	2562832	56.52
Evening Rush	1268782	27.98
Morning Rush	702713	15.50

Off-peak hours have the highest demand, indicating that a significant portion of Uber's business occurs outside of the traditional rush hours.
Evening rush hours, although representing only 4 hours throughout the day, witness substantial demand, which suggests that people rely on Uber for their evening commutes or social outings. Morning rush hours, on the other hand, exhibit lower demand.

In [11]:
```
%%sql

SELECT COUNT(*) / (COUNT(DISTINCT date) * 24) AS average_trips_per_hour
FROM uber_pickups;
```

```
 * mysql://root:***@localhost/test
1 rows affected.
```

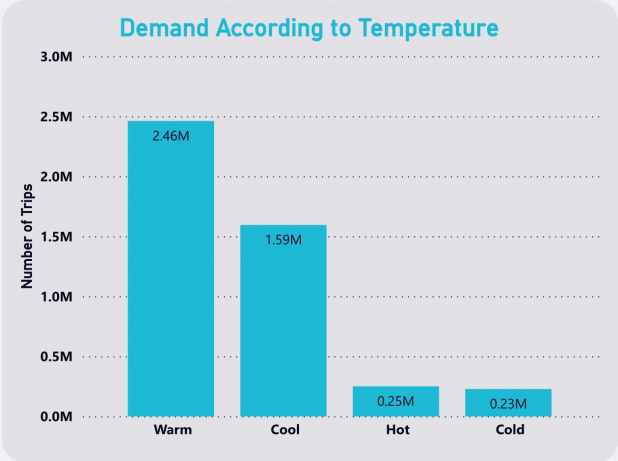
Out[11]:

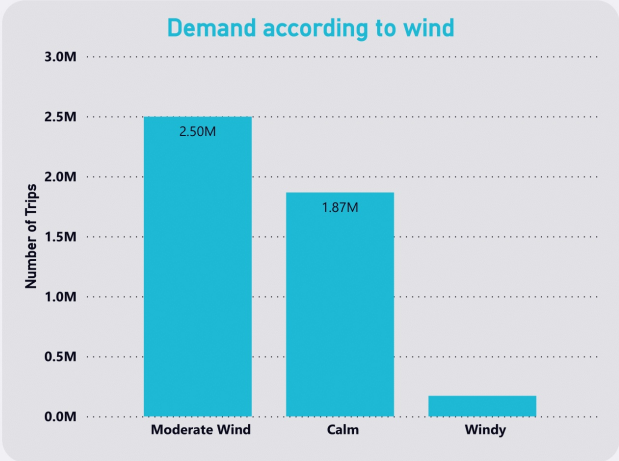
average_trips_per_hour
1032.4060

Is there any correlation between weather conditions and the number of Uber pickups?

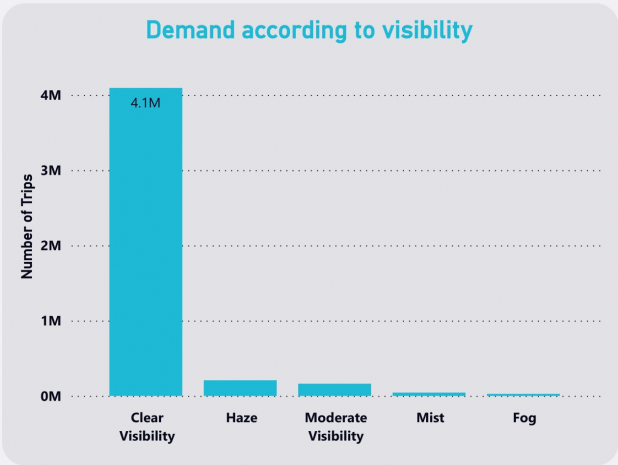
Before analyzing the correlation between weather conditions and the demand for Uber, it's essential to acknowledge that the dataset provided covers a specific time frame, including the summer, end of spring, and beginning of fall. As weather patterns, user behaviors, and transportation demands can vary significantly throughout the seasons, the analysis is limited to the recorded time period. Any insights drawn from this analysis should be interpreted within the context of this specific timeframe.

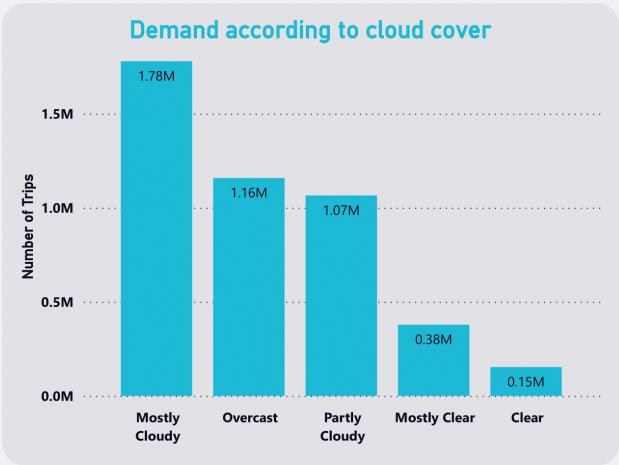
Is there any correlation between weather conditions and the number of Uber pickups?





Is there any correlation between weather conditions and the number of Uber pickups?





In [12]:
```
%%sql

WITH temp_categories AS (
    SELECT CASE
            WHEN temperature > 85 THEN 'Hot'
            WHEN temperature >= 70 AND temperature <= 85 THEN 'Warm'
            WHEN temperature >= 50 AND temperature < 70 THEN 'Cool'
            WHEN temperature < 50 THEN 'Cold'
           END AS temp_category,
           date, hour
    FROM weather
)

SELECT t.temp_category,
       COUNT(u.date) AS number_of_trips,
       ROUND(COUNT(u.date) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups u
JOIN temp_categories t
ON u.date = t.date AND u.hour = t.hour
GROUP BY t.temp_category
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
4 rows affected.
```

Out[12]:

temp_category	number_of_trips	percentage
Warm	2461318	54.28
Cool	1594746	35.17
Hot	250322	5.52
Cold	227941	5.03

More than half of Uber trips were requested during warm temperature. Nevertheless, we can't surely say that people tend to use Uber more during warm weather, because the time frame of this data is skewed in favor of the summer.
But generally, warmer weather encourages people to go out, attend events, and socialize, leading to increased demand for transportation services like Uber. On the other hand, extreme temperatures, whether hot or cold, can deter people from traveling.

In [13]:
```
%%sql

WITH wind_categories AS (
    SELECT CASE
            WHEN wind > 18 THEN 'Windy'
            WHEN wind >= 8 AND wind <= 18 THEN 'Moderate Wind'
            WHEN wind < 8 THEN 'Calm'
           END AS wind_category,
           date, hour
    FROM weather
)

SELECT w.wind_category,
       COUNT(u.date) AS number_of_trips,
       ROUND(COUNT(u.date) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups u
JOIN wind_categories w
ON u.date = w.date AND u.hour = w.hour
GROUP BY w.wind_category
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
3 rows affected.
```

Out[13]:

wind_category	number_of_trips	percentage
Moderate Wind	2497583	55.08
Calm	1865648	41.14
Windy	171096	3.77

Moderate Wind conditions accounted for more than half of the demand. Suggesting that moderate winds do not significantly deter people from going out and commuting.

In [14]:
```
%%sql

WITH vis_categories AS (
    SELECT CASE
            WHEN visibility > 7 THEN 'Clear Visibility'
            WHEN visibility >= 5 AND visibility <= 7 THEN 'Moderate Visibility'
            WHEN visibility >= 2 AND visibility < 5 THEN 'Haze'
            WHEN visibility >= 1 AND visibility < 2 THEN 'Mist'
            WHEN visibility < 1 THEN 'Fog'
           END AS vis_category,
           date, hour
    FROM weather
)

SELECT v.vis_category,
       COUNT(u.date) AS number_of_trips,
       ROUND(COUNT(u.date) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups u
JOIN vis_categories v
ON u.date = v.date AND u.hour = v.hour
GROUP BY v.vis_category
ORDER BY number_of_trips DESC;
```

```
 * mysql://root:***@localhost/test
5 rows affected.
```

Out[14]:

vis_category	number_of_trips	percentage
Clear Visibility	4093272	90.27
Haze	207882	4.58
Moderate Visibility	162532	3.58
Mist	43844	0.97
Fog	26797	0.59

Clear Visibility conditions accounted for a vast majority of trips. This can be explained by how clear weather likely encourages outdoor activities, events attendance, and overall mobility, leading to higher demand for transportation services.

In [15]:
```
%%sql

SELECT w.cloud_cover,
       COUNT(u.date) AS number_of_trips,
       ROUND(COUNT(u.date) * 100 / (SELECT COUNT(*) FROM uber_pickups), 2) AS percentage
FROM uber_pickups u
JOIN weather w
ON u.date = w.date AND u.hour = w.hour
GROUP BY w.cloud_cover
ORDER BY number_of_trips DESC;
```

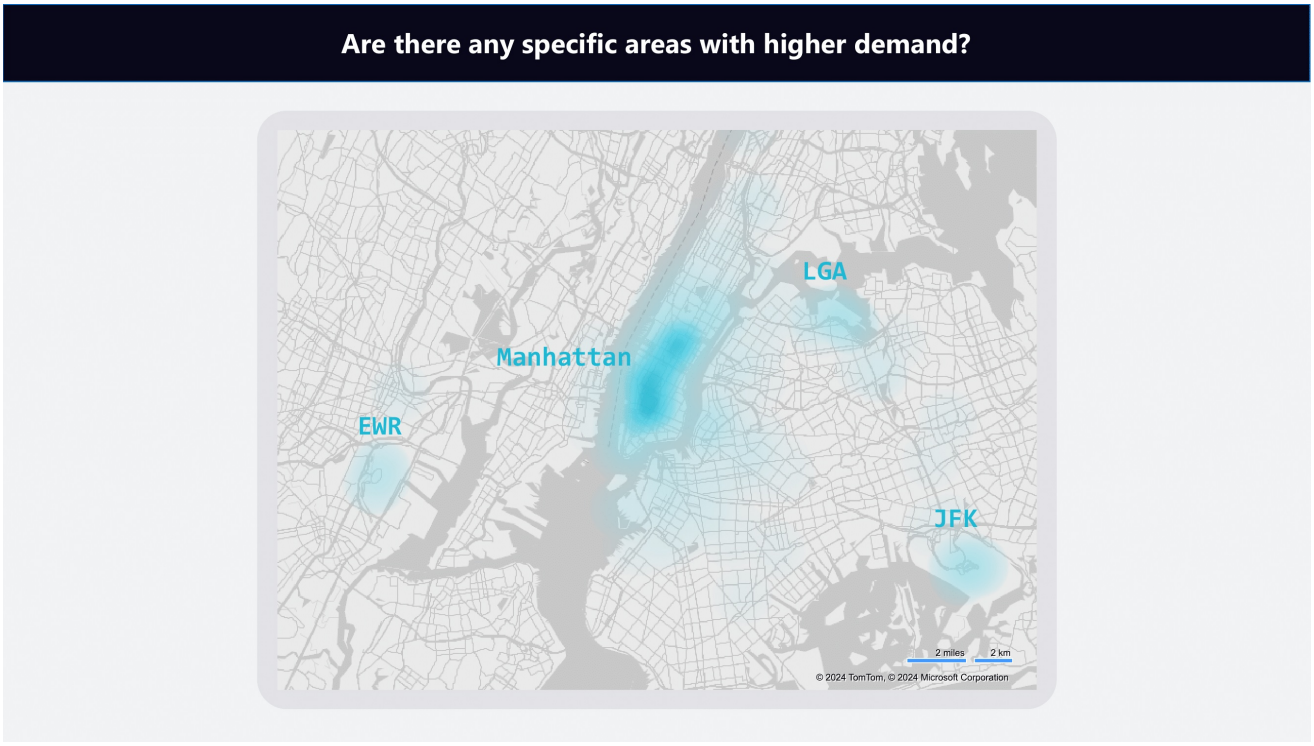
```
 * mysql://root:***@localhost/test
5 rows affected.
```

Out[15]:

cloud_cover	number_of_trips	percentage
Mostly Cloudy	1779150	39.24
Overcast	1158186	25.54
Partly Cloudy	1065507	23.50
Mostly Clear	378324	8.34
Clear	153160	3.38

Surprisingly, cloudy days witnessed far more demand than clear days, which typically offer optimal weather conditions. This might be due to people feeling comfortable walking or using other forms of transportation on such days.

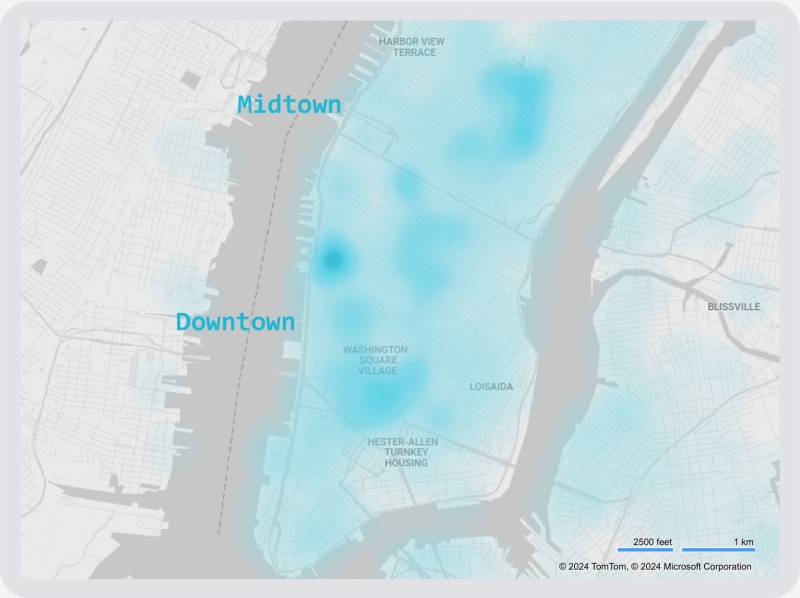
Are there any specific areas with higher demand?



We can see that there few hotspots in the city, namely the areas aound airports and, the heart of the city, Manhattan, and as we move outwards the demand drops significantly.
The three demand clusters at JFK, LaGuardia, and Newark airports are distinct and well-defined, which highlits the fact that airports are natural hubs for ride-sharing demand. Travelers arriving or departing from these airports often require transportation to or from various parts of the city. This demand is consistent and tends to peak during flight arrival and departure times.
We can also see that the demand at JFK and LaGuardia are higher than Newark. This is understandable since most international flights

land in JFK and LGA is the closest to Manhattan, while EWR is located in New Jersey and is the furthest from Manhattan.

Are there any specific areas with higher demand?



Manhattan, expectedly, witnesses the highest demand in the city, especially the dynamic hubs of midtown and downtown districts. It's a bustling business hub, with employees relying on these services for daily commuting, especially during peak hours. Additionally, Manhattan's status as a major tourist destination ensures a consistent need for convenient transportation, with visitors using ride-sharing to explore landmarks, theaters, museums, and restaurants. Moreover, Manhattan's high population density and economic activity create a constant, high-demand environment, prompting ride-sharing companies to strategically focus their services, promotions, and incentives to efficiently serve the transportation needs of both residents and tourists in this vibrant urban landscape.

Uber can take advantage from these insights to allocate more drivers and resources to meet the higher demand at the identified times, conditions, and locations, and potentially adjust pricing strategies to meet increased demand.
During lower demand periods, Uber may want to focus on marketing campaigns or promotions to stimulate ridership and make it an attractive option compared to other modes of transportation.

In [ ]:

In [ ]: