Citibike Dataset Analysis

Aleksandar Cvetković

Intro notes

First thing, I want to point out that I am well aware that this way of presenting the results might be considered naïve (and even unacceptable) by the Data Science professionals. Please bear in mind that up so far I have been mainly focused on the scientific research in mathematics and academia-related work, where things are done differently.

I believe that much better way to present the obtained results is to connect Google BigQuery with Python, and compile results into one big Jupyter notebook report. Given the opportunity, I would certainly learn to do that from more experienced colleagues (or on my own, eventually) in a matter of days. But now I wanted to focus and use my time on the analysis and results, and not on the form of presentation.

Intro notes

To answer the questions from the Task 1, I coded and executed SQL queries in Google Big Query and analysed the results. I tried to take each question a bit further and give some additional insights. By doing so, I was simultaneously solving Task 1 and tackling the Task 2 a bit. The insights I obtained this way shaped the way I performed the analysis in the Task 2.

The charts presented in the first section were done in Tableau. The results presented in the second section were obtained by using Google BigQuery alongside with BigQuery Geo Viz.

Task 1

Using SQL to answer some basic questions about the New York Citibike dataset

Question 1: What is the unique number of stations in the dataset?

SQL query:

```
SELECT count(*) as station_count
FROM `bigquery-public-data.new_york_citibike.citibike_stations`
```

Result and analysis:

The query returns the result of 935. So, this is the number of unique Citibike stations. Inspecting the ‘citibike_stations’ table more closely (i.e. ‘is_installed’ and ‘last_reported’ columns), we see that the data is most recent, and so all of the 935 stations are currently installed and active.

Question 2: How many Citibikes are in the dataset?

SQL query:

```
SELECT COUNT(DISTINCT bikeid) as bikes_count
FROM `bigquery-public-data.new_york_citibike.citibike_trips`
```

Result and analysis:

The query returns the result of 17 682, which is the number of bikes that appear in the dataset. As we will note later in our analysis, not all of these bikes are currently in use. We will revisit this question later, to see how many bikes are being used actively.

Question 3: What is the most popular start station and end station?

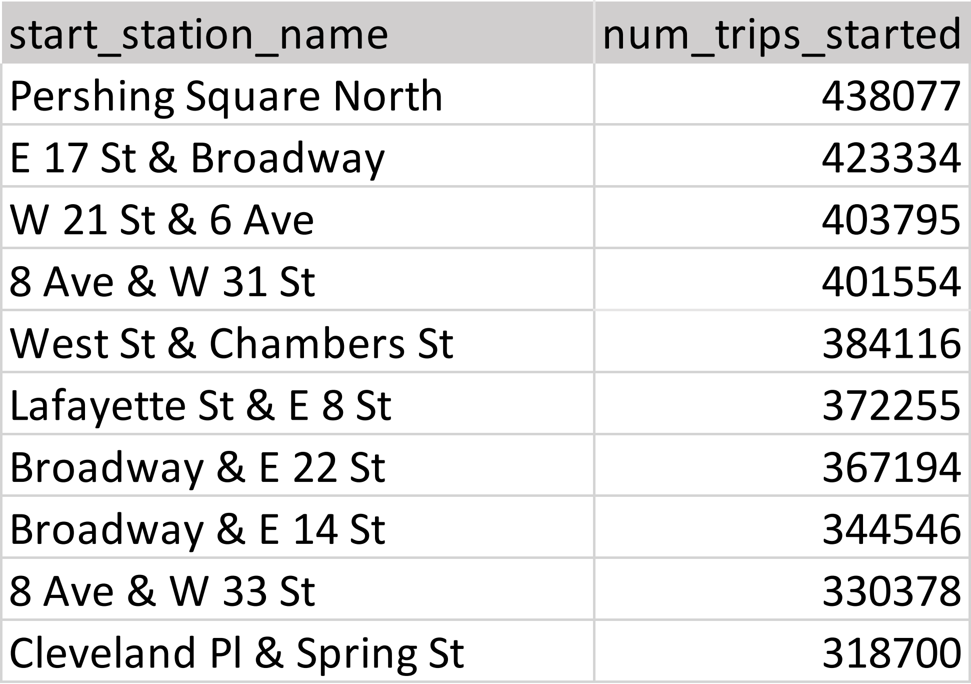
SQL query (for starting stations):

```
SELECT
   start_station_name,
   COUNT(start_station_id) AS num_trips_started

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

GROUP BY start_station_name

ORDER BY num_trips_started DESC

LIMIT 10
```

SQL query (for end stations):

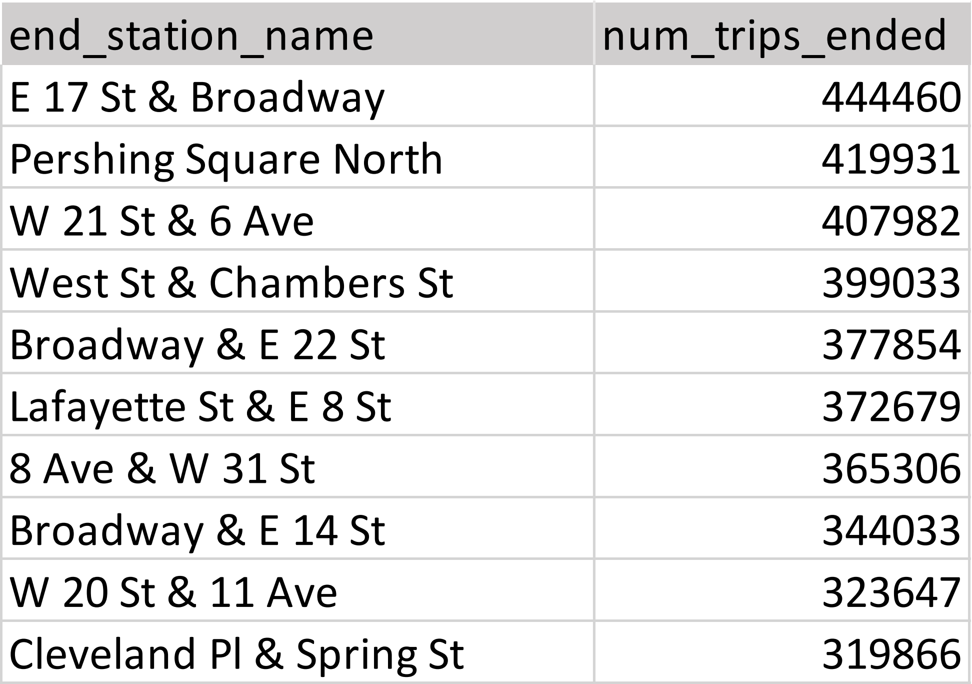
```
SELECT
   end_station_name,
   COUNT(end_station_id) AS num_trips_ended

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

GROUP BY end_station_name

ORDER BY num_trips_ended DESC

LIMIT 10
```

Results and analysis:



start_station_name	num_trips_started
Pershing Square North	438077
E 17 St & Broadway	423334
W 21 St & 6 Ave	403795
8 Ave & W 31 St	401554
West St & Chambers St	384116
Lafayette St & E 8 St	372255
Broadway & E 22 St	367194
Broadway & E 14 St	344546
8 Ave & W 33 St	330378
Cleveland Pl & Spring St	318700

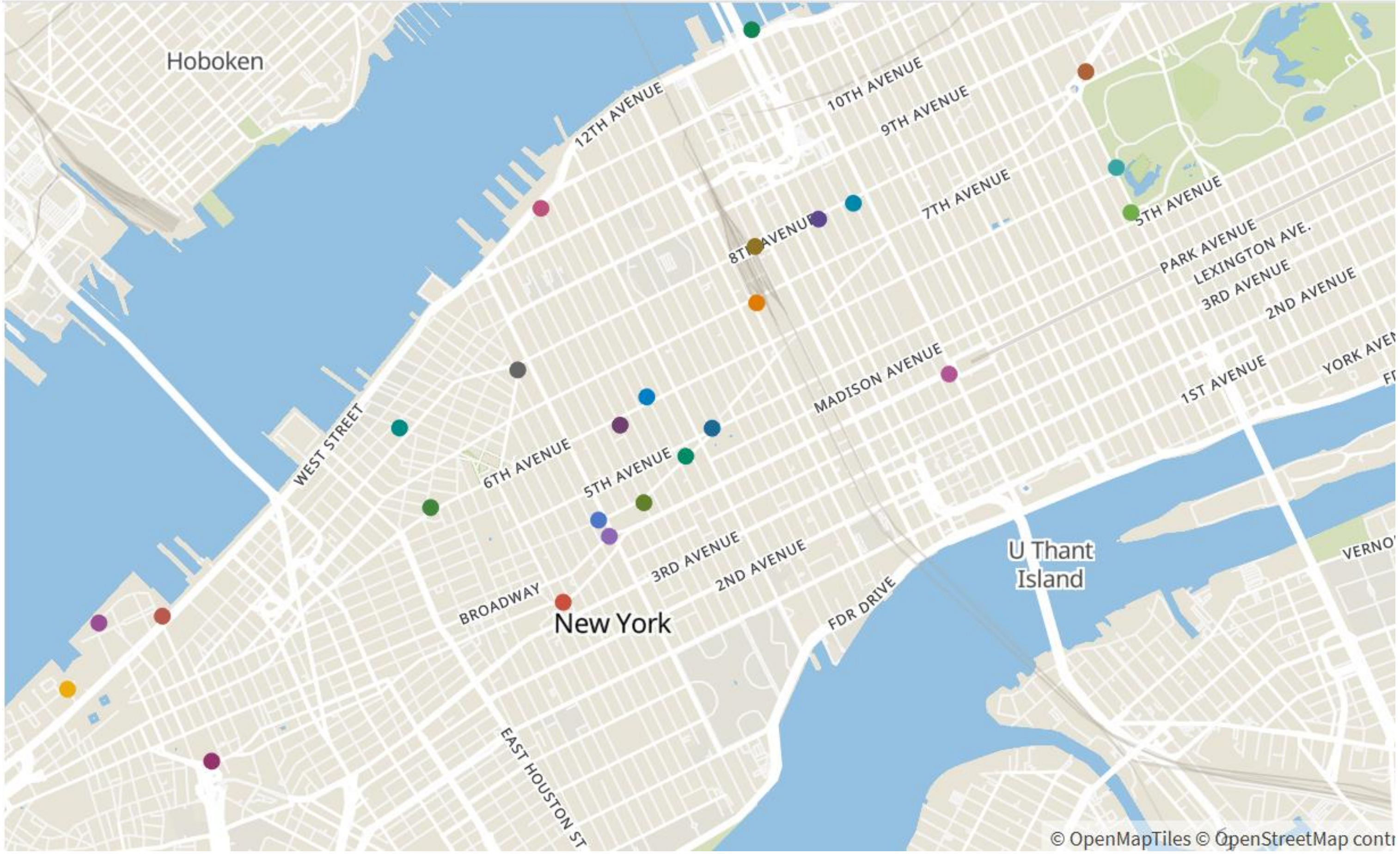
Top 10 most popular starting stations



end_station_name	num_trips_ended
E 17 St & Broadway	444460
Pershing Square North	419931
W 21 St & 6 Ave	407982
West St & Chambers St	399033
Broadway & E 22 St	377854
Lafayette St & E 8 St	372679
8 Ave & W 31 St	365306
Broadway & E 14 St	344033
W 20 St & 11 Ave	323647
Cleveland Pl & Spring St	319866

Top 10 most popular end stations

Here we list the most popular start/end stations, along with the number of times a station was used as starting/ending point of a journey. Pershing Square North and E 17 St/Broadway are the most popular stations.



Top 25 most popular starting stations

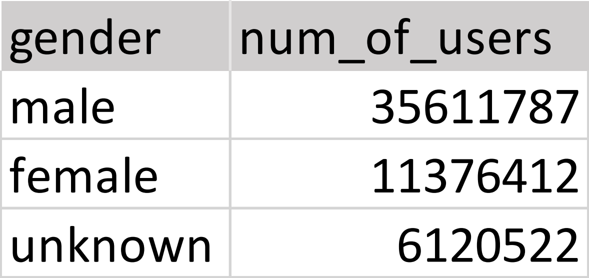
The map of the most popular starting stations (which is almost completely identical to the corresponding end stations map) can give us a clue about most popular NYC areas: they're settled around 3rd, 5th, 6th and 8th avenues.

Question 4: What is the gender distribution of the Citibike users?

SQL query:

```
SELECT
   gender,
   COUNT(gender) AS num_of_users

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE bikeid IS NOT NULL

GROUP BY gender
```

Results and analysis:

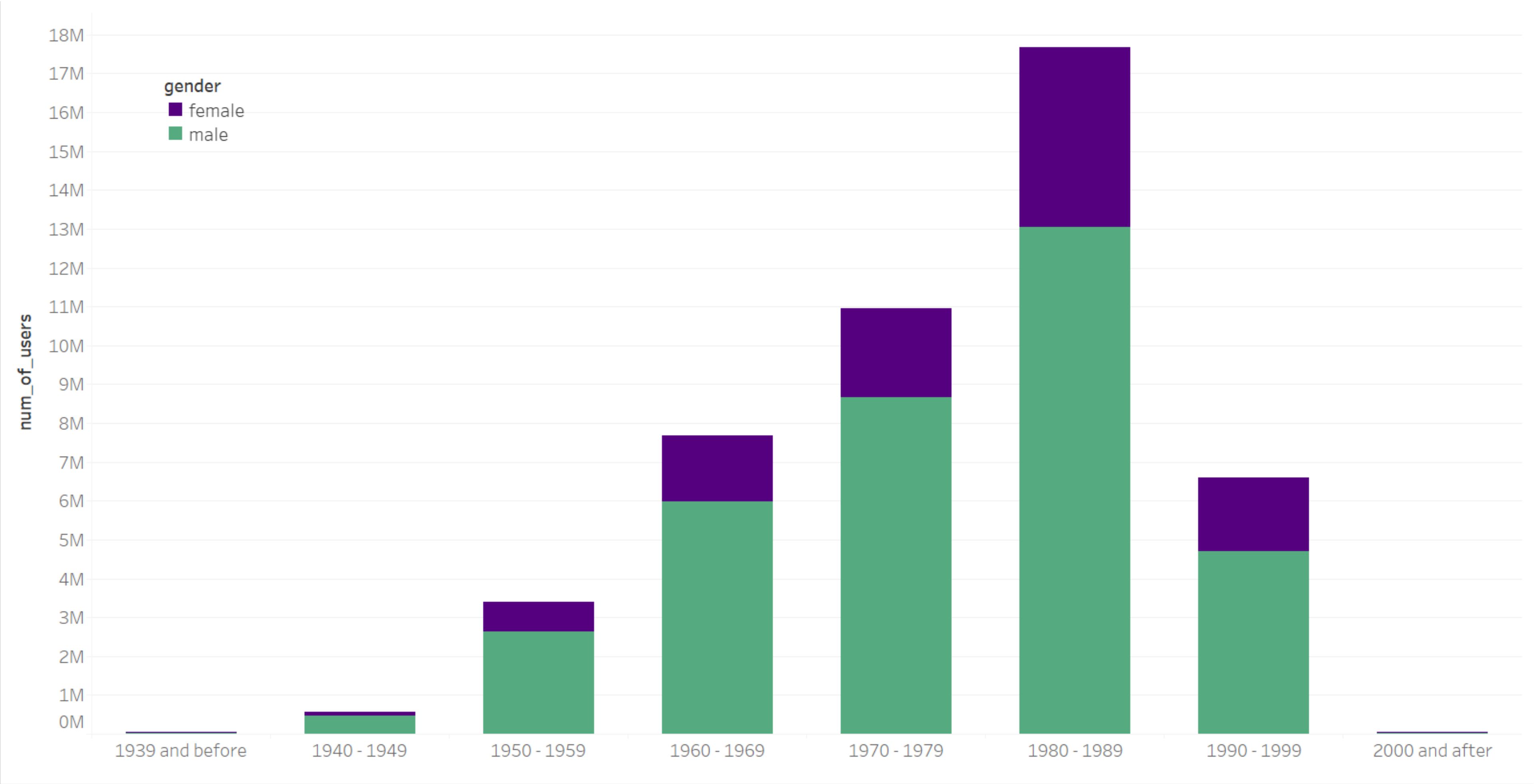


gender	num_of_users
male	35611787
female	11376412
unknown	6120522

User distribution across gender

From the table we see that male users are predominant: men are using Citibike three times more than women. Can the same thing be said about the cycling habits for people of New York in general? Does this three-to-one ratio apply also for people having their own bike?

It’s interesting to see how users are distributed across both gender and age.
We’ll explore this in the next slide.



Citibike users distributed across year of birth and gender

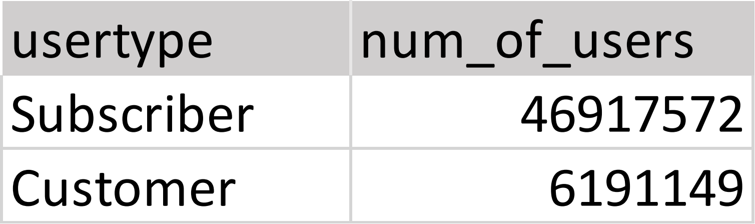
The largest body of Citbike users come from people born during the ’80s. Again, can the same thing be said about NYC bike users in general? In addition, we see that the 1:3 (or even 1:4) distribution of male-to-female users is consistent through all the generations.

Question 5: What is the usertype distribution of the Citibike users?

SQL query:

```
SELECT
   usertype,
   COUNT(usertype) AS num_of_users

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE bikeid IS NOT NULL

GROUP BY usertype
```

Results and analysis:



usertype	num_of_users
Subscriber	46917572
Customer	6191149

User distribution across type of users

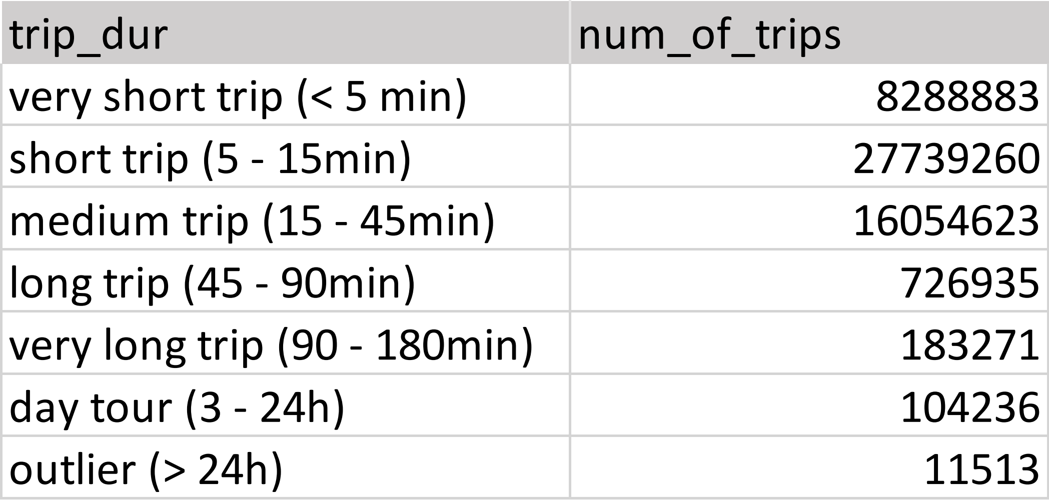
There are significantly more subscribers to Citibike service than the ‘pay-per-ride’ users. We might assume that most of the subscribers are actually NYC residents, while tourists will usually fall into the other category. This can also explain the huge difference in numbers: New York locals who subscribe use the service more often, since they don’t need to pay for every single ride. For many of them, Citibike is a part of their daily life, and they use it to commute around less popular areas, unlike tourists.

Question 6: What is the trip duration distribution of the Citibike trips?

SQL query:

```
SELECT
    CASE
        WHEN tripduration > 0 AND tripduration <= 300 THEN 'very short trip’
        WHEN tripduration > 300 AND tripduration <= 900 THEN 'short trip’
        WHEN tripduration > 900 AND tripduration <= 2700 THEN 'medium trip’
        WHEN tripduration > 2700 AND tripduration <= 5400 THEN 'long trip’
        WHEN tripduration > 5400 AND tripduration <= 10800 THEN 'very long trip’
        WHEN tripduration > 10800 AND tripduration <= 86400 THEN 'day tour’
        WHEN tripduration > 86400 THEN 'outlier’
    END AS trip_dur,
    COUNT(bikeid) AS num_of_trips

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE tripduration IS NOT NULL AND bikeid IS NOT NULL

GROUP BY trip_dur
```

Results and analysis:

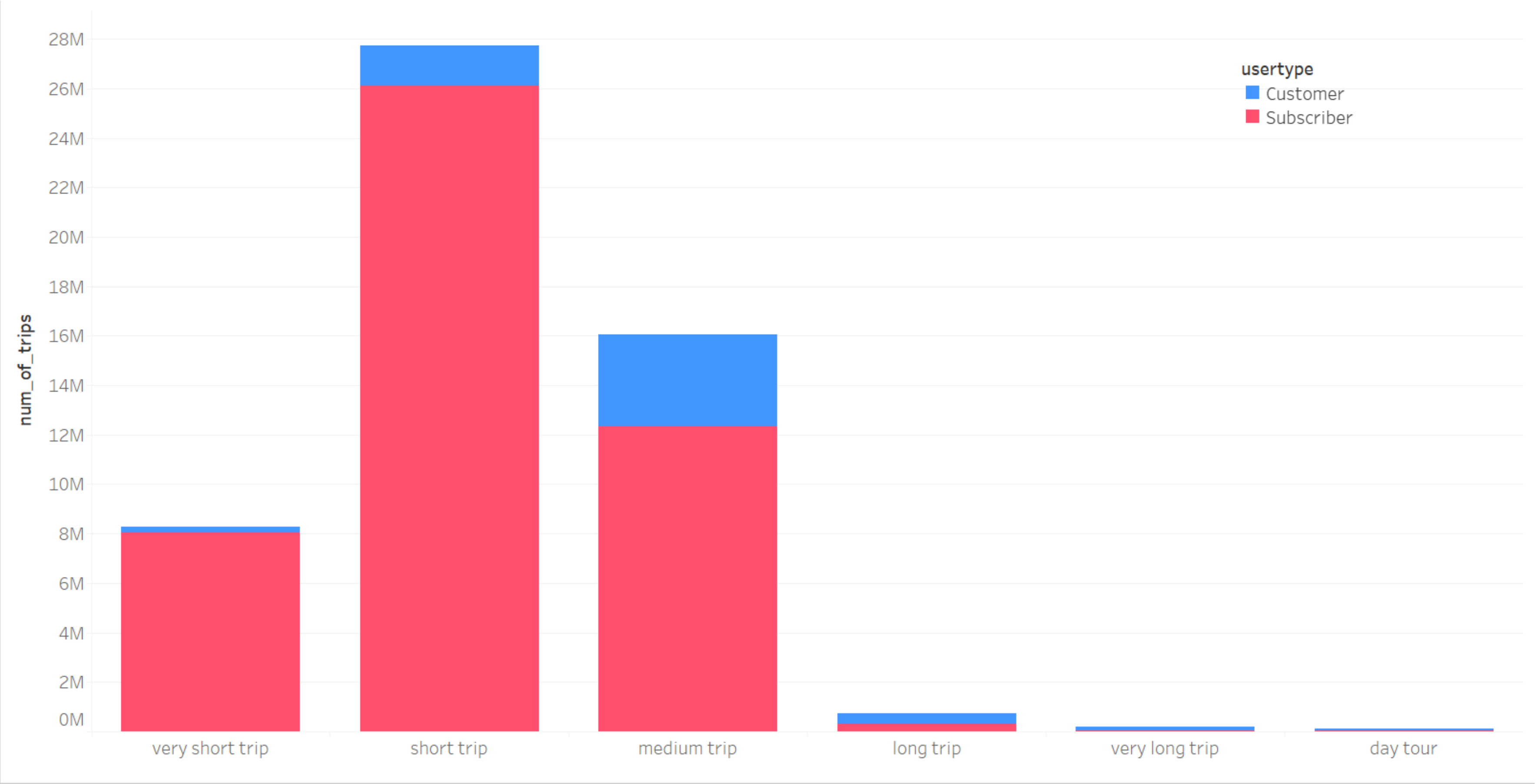


trip_dur	num_of_trips
very short trip (< 5 min)	8288883
short trip (5 - 15min)	27739260
medium trip (15 - 45min)	16054623
long trip (45 - 90min)	726935
very long trip (90 - 180min)	183271
day tour (3 - 24h)	104236
outlier (> 24h)	11513

Trip duration distribution

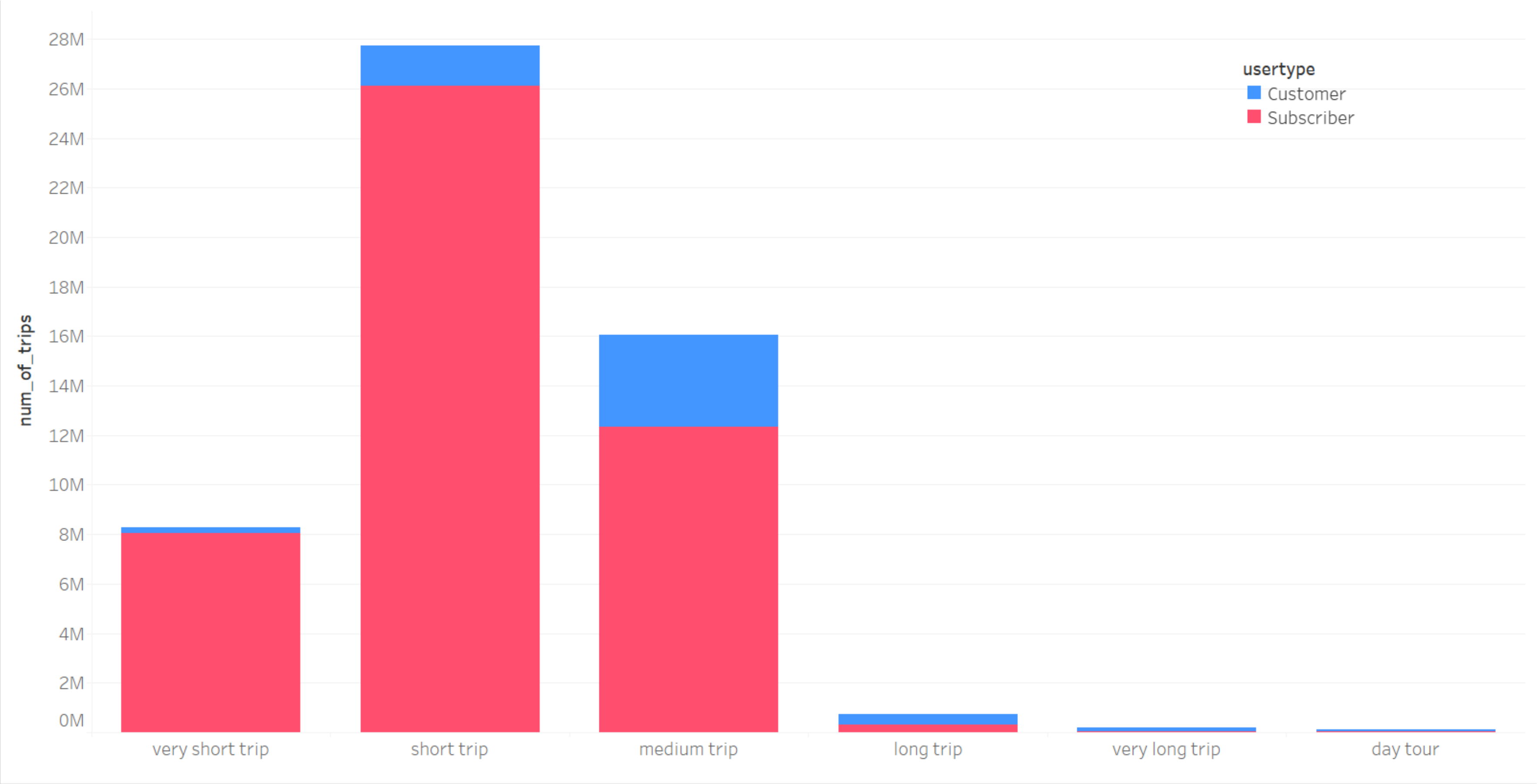
Medium and short trips are by far the most popular ones. Quite short trips are not taken as much, but still much more frequently than trips that last an hour or more.

It will be interesting to see how trips are distributed across both duration and user type. We’ll explore this in the next slide.



Citibike trips distributed across trip duration and user type

Very short trips are almost always undertaken by subscribers. The situation is also similar with the short trips. If we assume that subscribers are NYC locals, this makes more sense: Citibike is a part of their daily routine, and they use it to move around the neighbourhood. The difference between subscribers and customers is smaller for medium trips, but subscribers are still predominant, as they might be using Citibike to commute from home to work, for example.



Citibike trips distributed across trip duration and user type

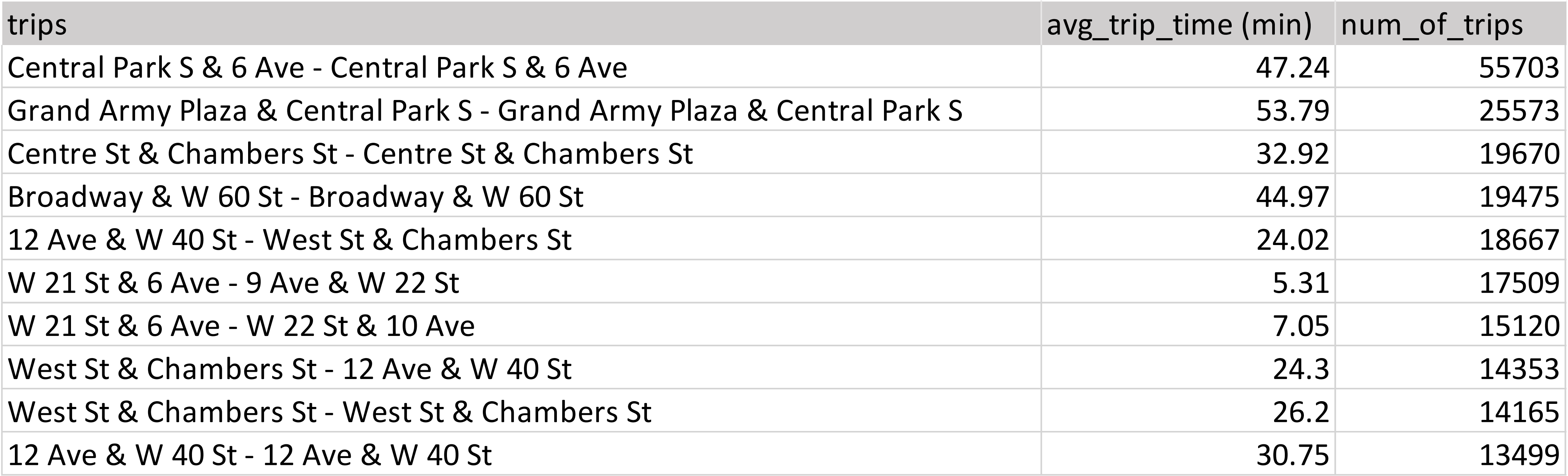
But the story is completely different for longer trips! In this case ‘pay-per-ride’ customers overtake subscribers. If we assume that many customers are also tourists, then this also makes sense: they will want to rent a bike for a bike-tour around the city, probably commuting around the most popular areas.

Question 7: What is the most popular Citibike trip?

SQL query:

```
SELECT
   CONCAT(start_station_name,' - ',end_station_name) AS trips,
   ROUND(AVG(tripduration)/60, 2) AS avg_trip_time,
   COUNT(bikeid) AS num_of_trips

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE bikeid IS NOT NULL

GROUP BY trips

ORDER BY num_of_trips DESC

LIMIT 10
```

Results and analysis:

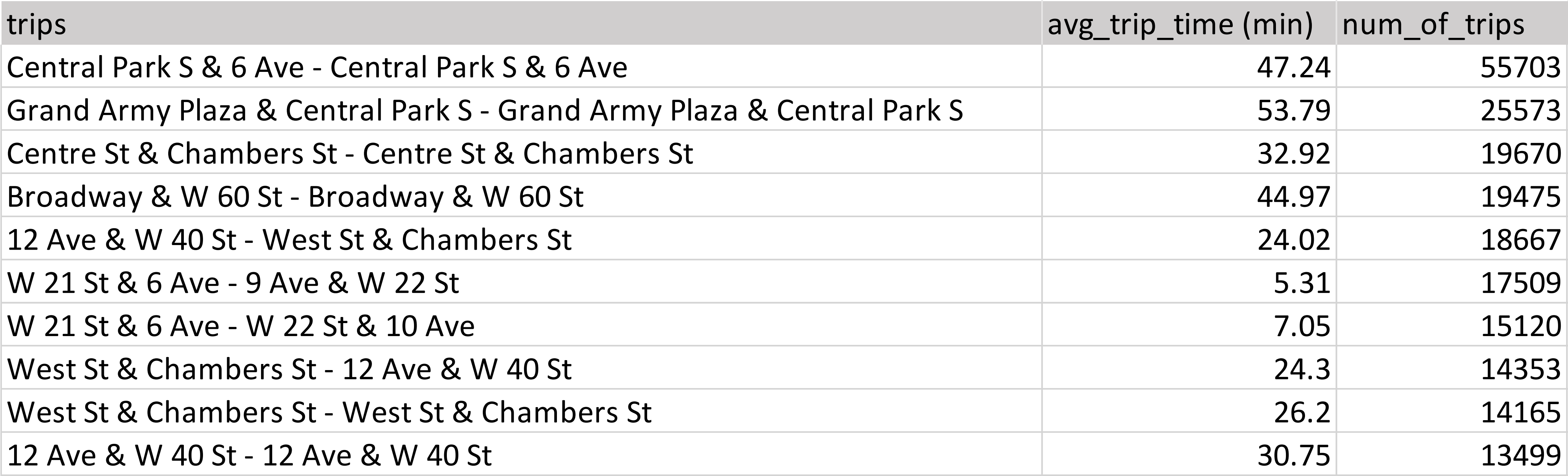


trips	avg_trip_time (min)	num_of_trips
Central Park S & 6 Ave - Central Park S & 6 Ave	47.24	55703
Grand Army Plaza & Central Park S - Grand Army Plaza & Central Park S	53.79	25573
Centre St & Chambers St - Centre St & Chambers St	32.92	19670
Broadway & W 60 St - Broadway & W 60 St	44.97	19475
12 Ave & W 40 St - West St & Chambers St	24.02	18667
W 21 St & 6 Ave - 9 Ave & W 22 St	5.31	17509
W 21 St & 6 Ave - W 22 St & 10 Ave	7.05	15120
West St & Chambers St - 12 Ave & W 40 St	24.3	14353
West St & Chambers St - West St & Chambers St	26.2	14165
12 Ave & W 40 St - 12 Ave & W 40 St	30.75	13499

Top 10 most popular trips

Let’s observe top two routes: they are considerably more popular than other routes, they have the same starting and end station, they are both medium/long trips, and they both refer to Central Park South. All this can reveal a favourite habit of Citibike users: taking a nice long tour through the Central Park.

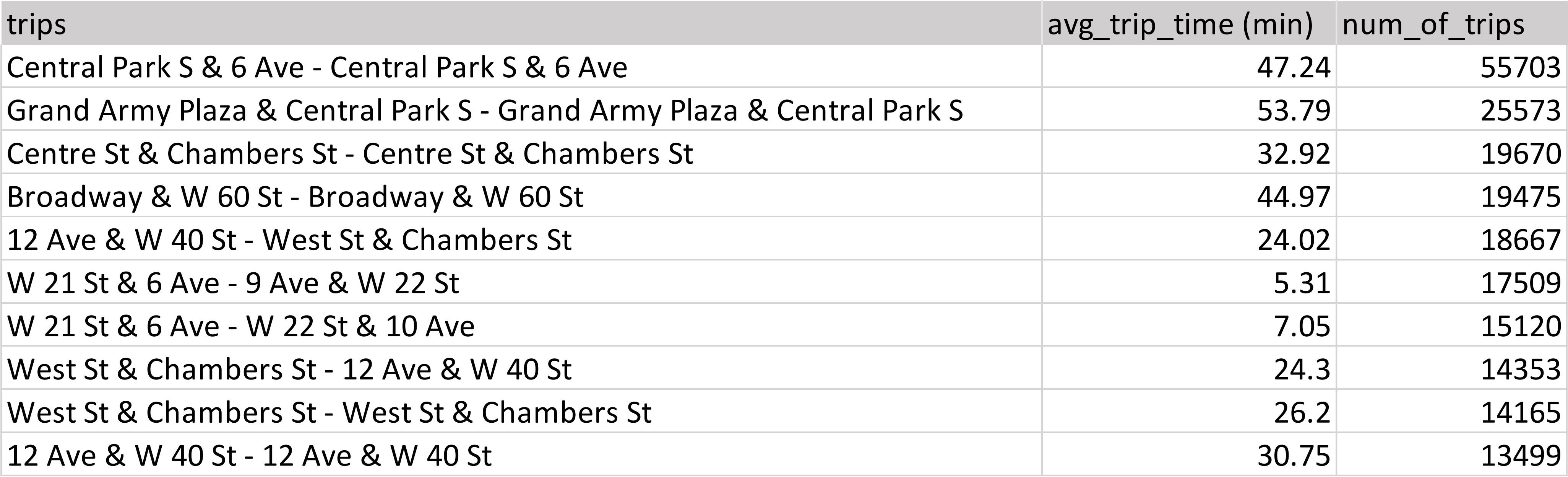
This can also explain why there’s almost an equal number of subscribers (locals) and costumers (tourists) taking longer rides: whether you are a local or not, you’d want to cycle around Central Park.



trips	avg_trip_time (min)	num_of_trips
Central Park S & 6 Ave - Central Park S & 6 Ave	47.24	55703
Grand Army Plaza & Central Park S - Grand Army Plaza & Central Park S	53.79	25573
Centre St & Chambers St - Centre St & Chambers St	32.92	19670
Broadway & W 60 St - Broadway & W 60 St	44.97	19475
12 Ave & W 40 St - West St & Chambers St	24.02	18667
W 21 St & 6 Ave - 9 Ave & W 22 St	5.31	17509
W 21 St & 6 Ave - W 22 St & 10 Ave	7.05	15120
West St & Chambers St - 12 Ave & W 40 St	24.3	14353
West St & Chambers St - West St & Chambers St	26.2	14165
12 Ave & W 40 St - 12 Ave & W 40 St	30.75	13499

Top 10 most popular trips

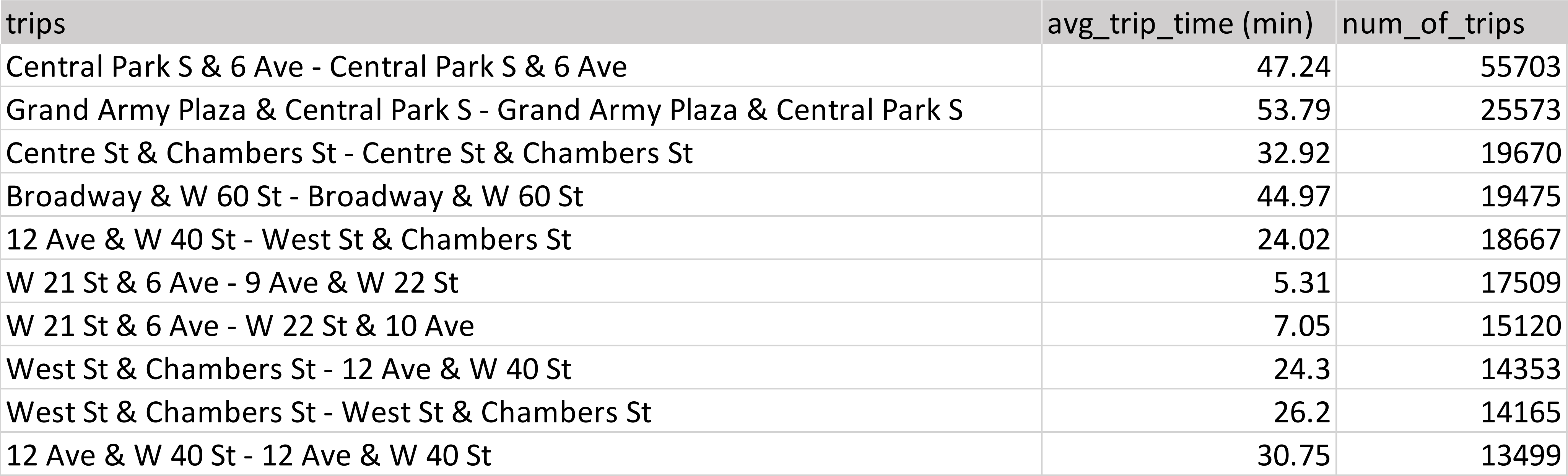
The third entry is also interesting to note: 30 minute trips starting and ending at the same station, located in southern New York. We can assume that biking around south NY (and enjoying all the sights there) is also quite popular.



trips	avg_trip_time (min)	num_of_trips
Central Park S & 6 Ave - Central Park S & 6 Ave	47.24	55703
Grand Army Plaza & Central Park S - Grand Army Plaza & Central Park S	53.79	25573
Centre St & Chambers St - Centre St & Chambers St	32.92	19670
Broadway & W 60 St - Broadway & W 60 St	44.97	19475
12 Ave & W 40 St - West St & Chambers St	24.02	18667
W 21 St & 6 Ave - 9 Ave & W 22 St	5.31	17509
W 21 St & 6 Ave - W 22 St & 10 Ave	7.05	15120
West St & Chambers St - 12 Ave & W 40 St	24.3	14353
West St & Chambers St - West St & Chambers St	26.2	14165
12 Ave & W 40 St - 12 Ave & W 40 St	30.75	13499

Top 10 most popular trips

Observing this table we can see that there is another way to interpret the customer data: it can represent not only NYC visitors, but locals who don’t own a bike, but rent a Citibike for occasional recreation.



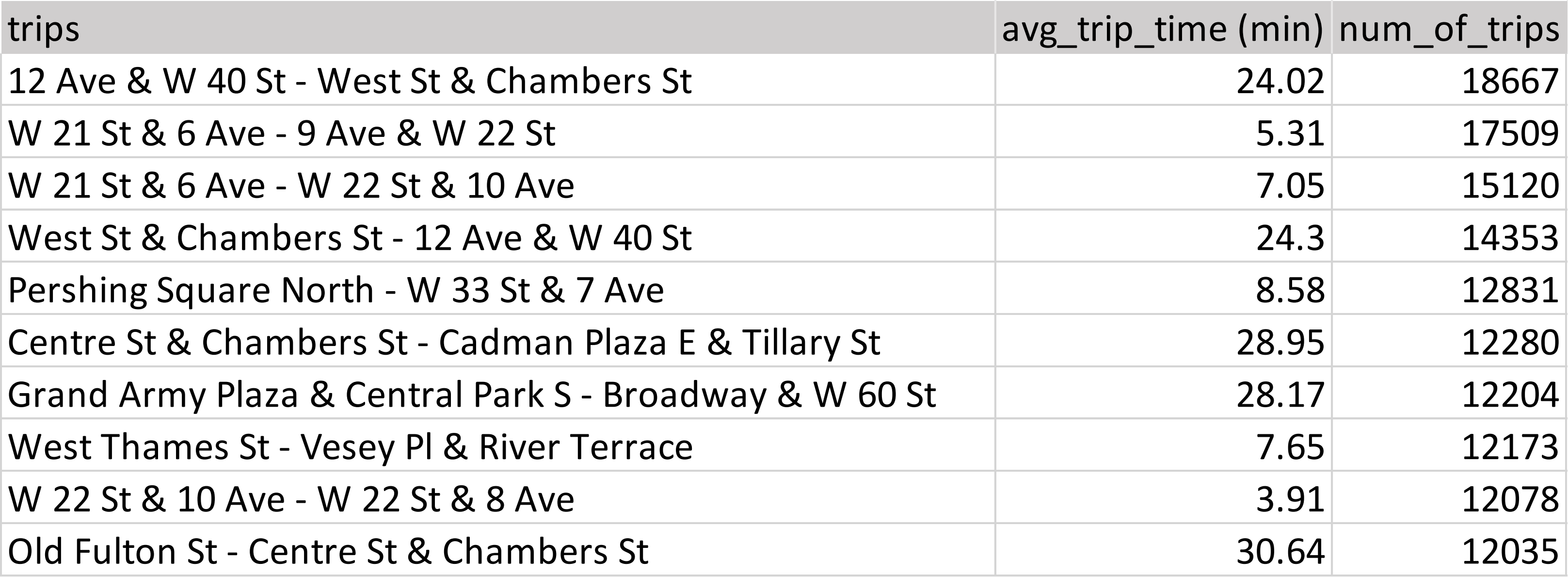
trips	avg_trip_time (min)	num_of_trips
Central Park S & 6 Ave - Central Park S & 6 Ave	47.24	55703
Grand Army Plaza & Central Park S - Grand Army Plaza & Central Park S	53.79	25573
Centre St & Chambers St - Centre St & Chambers St	32.92	19670
Broadway & W 60 St - Broadway & W 60 St	44.97	19475
12 Ave & W 40 St - West St & Chambers St	24.02	18667
W 21 St & 6 Ave - 9 Ave & W 22 St	5.31	17509
W 21 St & 6 Ave - W 22 St & 10 Ave	7.05	15120
West St & Chambers St - 12 Ave & W 40 St	24.3	14353
West St & Chambers St - West St & Chambers St	26.2	14165
12 Ave & W 40 St - 12 Ave & W 40 St	30.75	13499

Top 10 most popular trips

Another thing to note is that most of the listed routs are loops, having the same start and end stations. Let’s explore what’s the situation with routes having different start/end stations.

SQL query (for non-loop trips):

```
SELECT
   CONCAT(start_station_name,' - ',end_station_name) AS trips,
   ROUND(AVG(tripduration)/60, 2) AS avg_trip_time,
   COUNT(bikeid) AS num_of_trips

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE start_station_name != end_station_name
   AND bikeid IS NOT NULL

GROUP BY trips

ORDER BY num_of_trips DESC

LIMIT 10
```

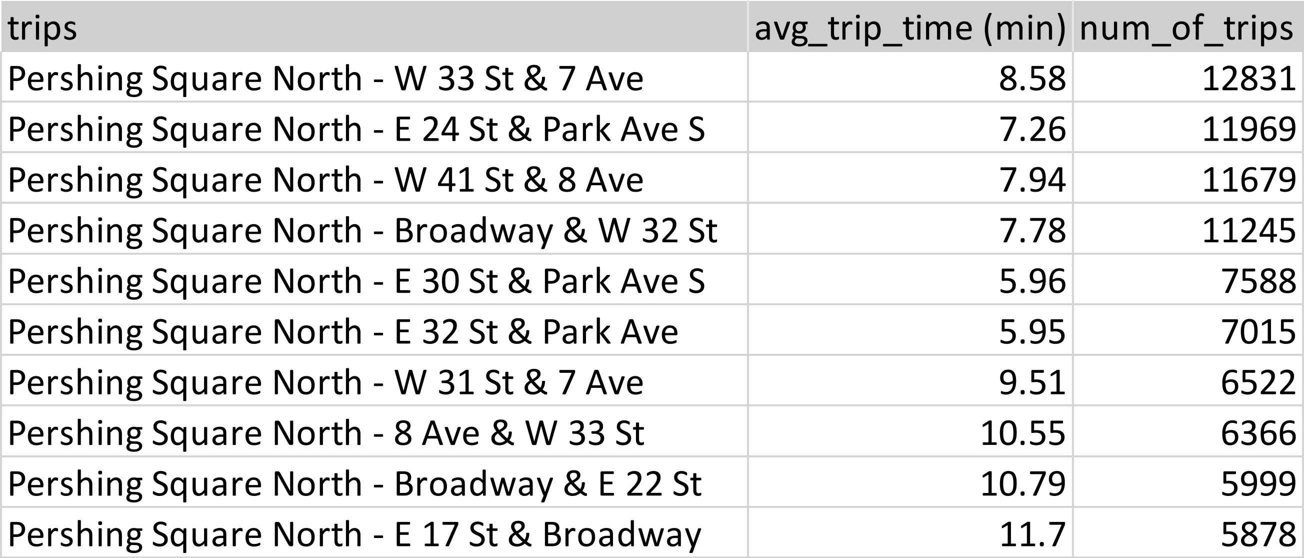


trips	avg_trip_time (min)	num_of_trips
12 Ave & W 40 St - West St & Chambers St	24.02	18667
W 21 St & 6 Ave - 9 Ave & W 22 St	5.31	17509
W 21 St & 6 Ave - W 22 St & 10 Ave	7.05	15120
West St & Chambers St - 12 Ave & W 40 St	24.3	14353
Pershing Square North - W 33 St & 7 Ave	8.58	12831
Centre St & Chambers St - Cadman Plaza E & Tillary St	28.95	12280
Grand Army Plaza & Central Park S - Broadway & W 60 St	28.17	12204
West Thames St - Vesey Pl & River Terrace	7.65	12173
W 22 St & 10 Ave - W 22 St & 8 Ave	3.91	12078
Old Fulton St - Centre St & Chambers St	30.64	12035

Top 10 most popular non-loop trips

The situation is not much different for non-loop trips: West Side and Central Park are predominant. The noticeable difference is that shorter trips are more often in this case.

What about the Pershing Square North? We saw that it’s the most popular starting station, but it didn’t appear anywhere on our top routes results. Let’s dig in by listing the most popular routes having Pershing Square North as a start station.



trips	avg_trip_time (min)	num_of_trips
Pershing Square North - W 33 St & 7 Ave	8.58	12831
Pershing Square North - E 24 St & Park Ave S	7.26	11969
Pershing Square North - W 41 St & 8 Ave	7.94	11679
Pershing Square North - Broadway & W 32 St	7.78	11245
Pershing Square North - E 30 St & Park Ave S	5.96	7588
Pershing Square North - E 32 St & Park Ave	5.95	7015
Pershing Square North - W 31 St & 7 Ave	9.51	6522
Pershing Square North - 8 Ave & W 33 St	10.55	6366
Pershing Square North - Broadway & E 22 St	10.79	5999
Pershing Square North - E 17 St & Broadway	11.7	5878

Top 10 most popular trips starting from PSN

All of the listed trips are short and thus probably mostly used by locals. While tourists tend to stick to popular routes, locals tend to take more diverse routes, according to their daily needs. And it seems that the daily life of locals is more centered on the East Side.

Question 8: Was there new bike introduced or removed at any point in time? What makes you think it did or didn’t? What about the citibike stations?

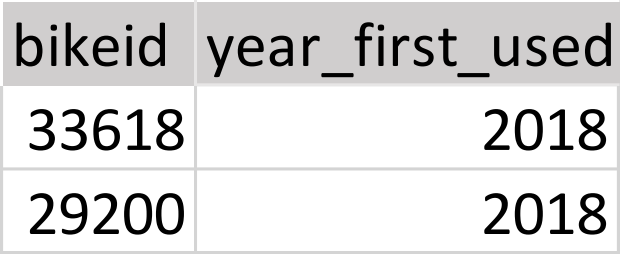
SQL query (for added bikes):

```
SELECT
   bikeid,
   EXTRACT(YEAR FROM MIN(starttime)) AS year_first_used,

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE bikeid IS NOT NULL

GROUP BY bikeid

ORDER BY year_first_used DESC

LIMIT 2
```

SQL query (for removed bikes):

```
SELECT
   bikeid,
   EXTRACT(YEAR FROM MAX(starttime)) AS year_last_used,
   COUNT(bikeid) as times_used

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE bikeid IS NOT NULL

GROUP BY bikeid

ORDER BY year_last_used

LIMIT 2
```

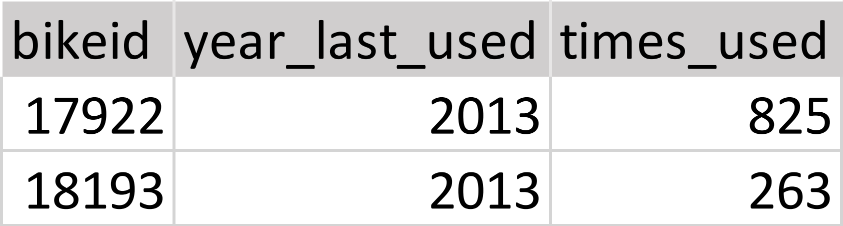
Results and analysis:



bikeid	year_first_used
33618	2018
29200	2018

Some recently introduced bikes

Citibike started its operation in 2013, and from the queried table we see that some new bikes are introduced in 2018.



bikeid	year_last_used	times_used
17922	2013	825
18193	2013	263

Bikes not in use anymore

Table above shows that there are also bikes removed since 2013, and used just few hundred times.

Since some of the bikes are not in use anymore, it is interesting to revisit Question 2 from this point, and see how many bikes were being used in 2018 (since this is the last time the database in BigQuerry was updated).

SQL query:

```
  SELECT COUNT(DISTINCT bikeid) as bikes_count
  FROM `bigquery-public-data.new_york_citibike.citibike_trips`
  WHERE EXTRACT(YEAR FROM starttime) = 2018
```

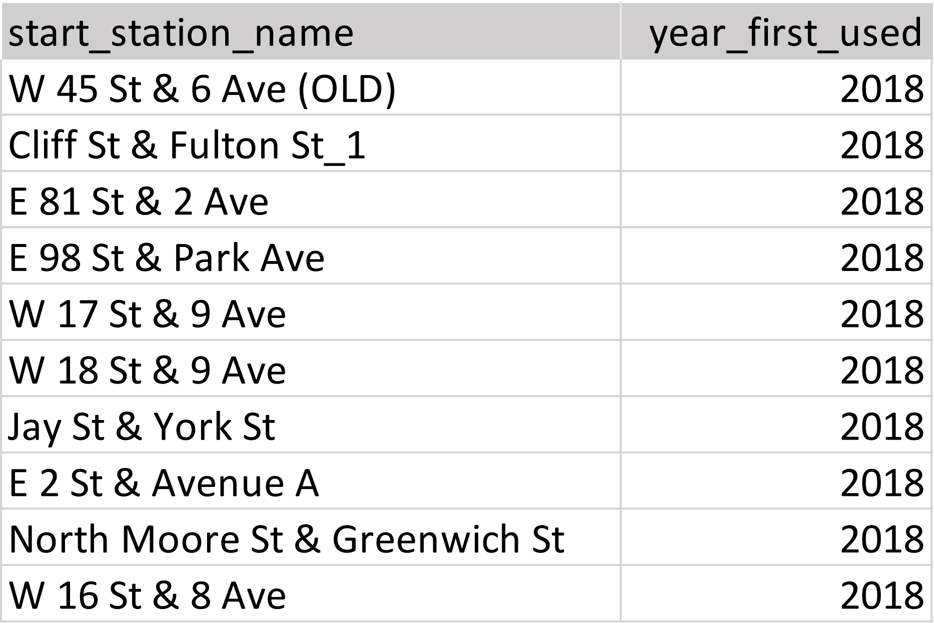
The query returns bike count of 12 947.
(And the same query for year 2013 returns a count of 6 503)

SQL query (for added stations):

```
SELECT
   start_station_name,
   EXTRACT(YEAR FROM MIN(starttime)) AS year_first_used

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE start_station_name != ‘’

GROUP BY start_station_name
ORDER BY year_first_used DESC

LIMIT 10
```

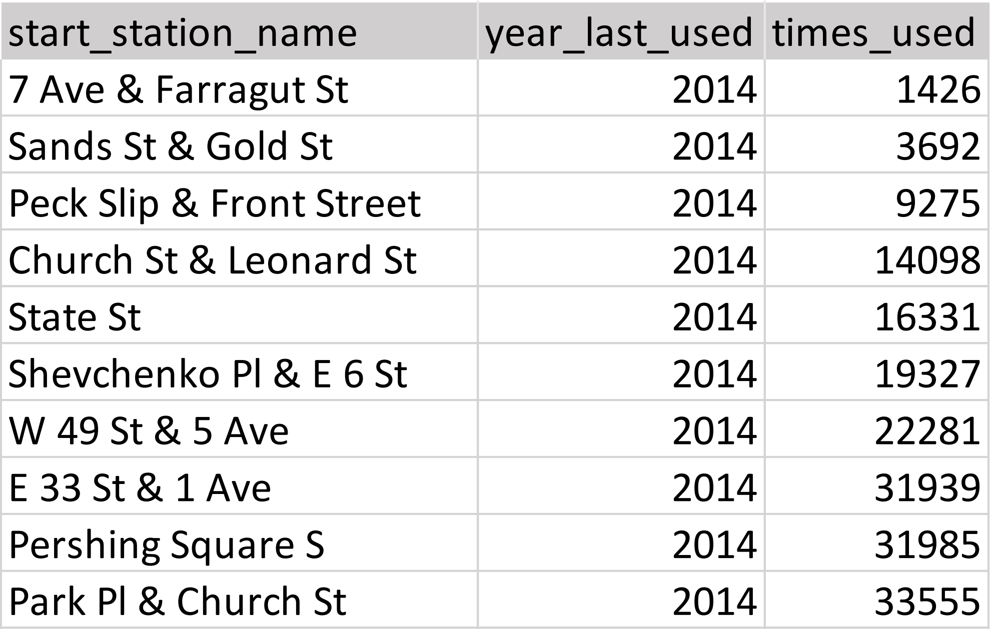
SQL query (for removed stations):

```
SELECT
   start_station_name,
   EXTRACT(YEAR FROM MAX(starttime)) AS year_last_used,
   COUNT(*) AS times_used,

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

WHERE start_station_name != ‘’

GROUP BY start_station_name

ORDER BY year_last_used, times_used

LIMIT 10
```



start_station_name	year_first_used
W 45 St & 6 Ave (OLD)	2018
Cliff St & Fulton St_1	2018
E 81 St & 2 Ave	2018
E 98 St & Park Ave	2018
W 17 St & 9 Ave	2018
W 18 St & 9 Ave	2018
Jay St & York St	2018
E 2 St & Avenue A	2018
North Moore St & Greenwich St	2018
W 16 St & 8 Ave	2018

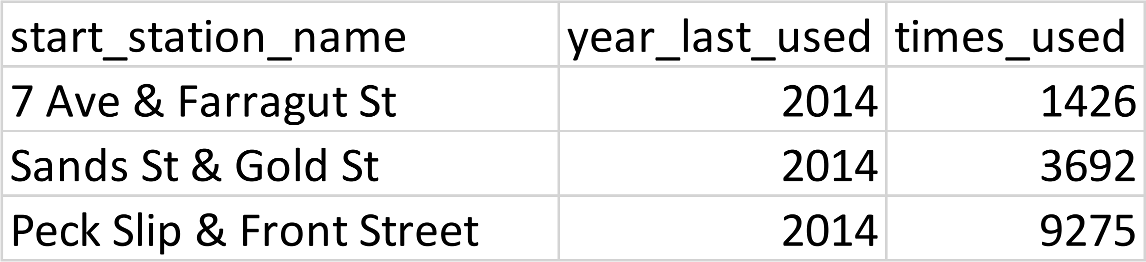
Some recently introduced stations



start_station_name	year_last_used	times_used
7 Ave & Farragut St	2014	1426
Sands St & Gold St	2014	3692
Peck Slip & Front Street	2014	9275
Church St & Leonard St	2014	14098
State St	2014	16331
Shevchenko Pl & E 6 St	2014	19327
W 49 St & 5 Ave	2014	22281
E 33 St & 1 Ave	2014	31939
Pershing Square S	2014	31985
Park Pl & Church St	2014	33555

Stations not in use anymore

Following the same reasoning as with the added/removed bikes, we can conclude that there were changes in stations infrastructures, some of the being added, while others were removed.



start_station_name	year_last_used	times_used
7 Ave & Farragut St	2014	1426
Sands St & Gold St	2014	3692
Peck Slip & Front Street	2014	9275

It might be interesting to think about why the stations listed above were used so little and then ultimately removed. Closer inspection reveals that all three stations are located near East River, first two in Brooklyn the third one in Manhattan. They are also not to far from each other. Maybe biking along East River is not that popular? Maybe people from New York don’t like to use bikes to cross the bridges between Brooklyn and Manhattan?

Task 2

Popular areas in NYC

We already gave some insights about popular NYC areas by going through the previous task. We saw that

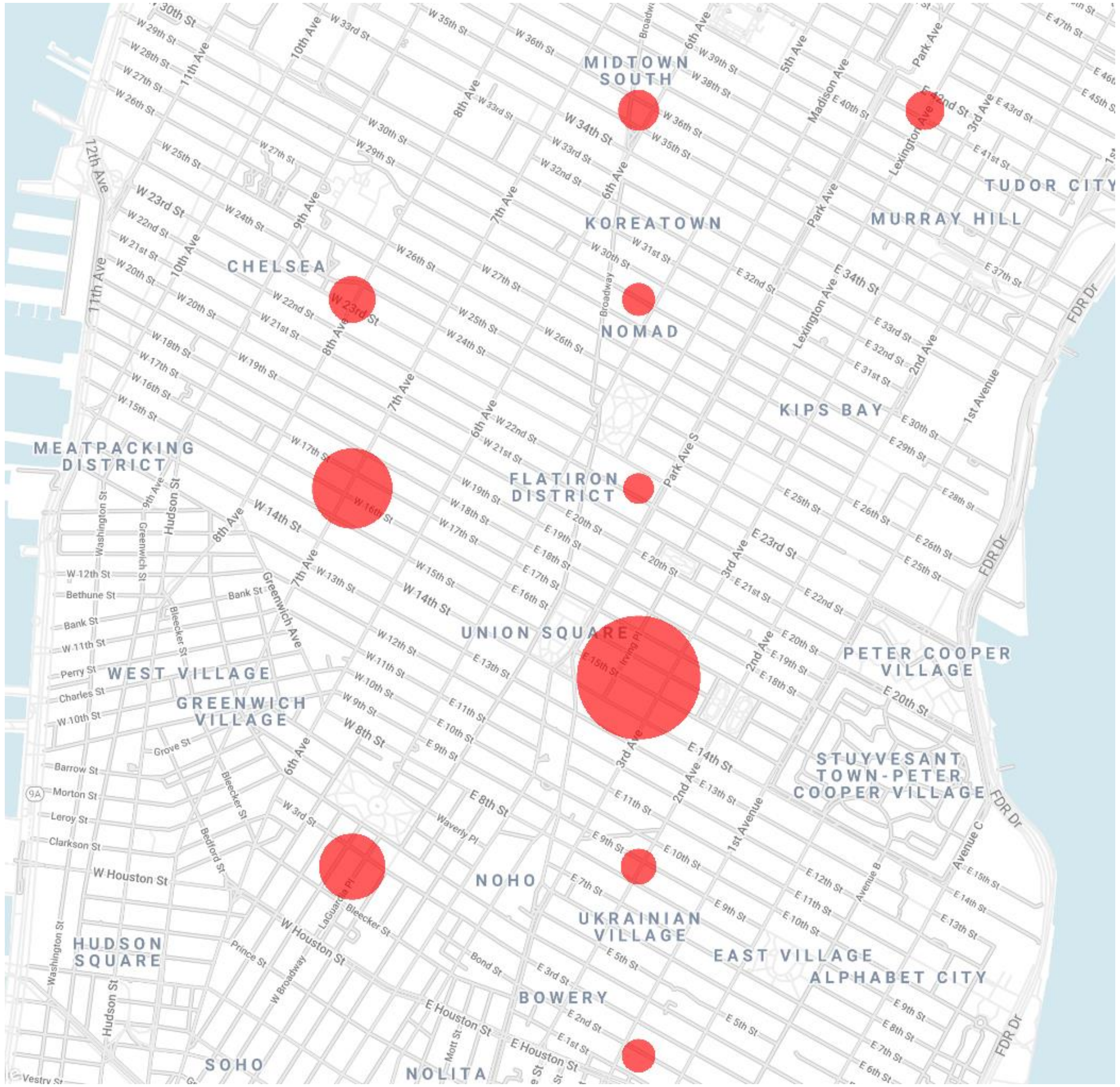
- There is a noticeable difference between subscriber and customer user types. We can make assumptions about the behaviour of NYC locals based on subscriber data, and assumptions about NYC visitors (and locals’ leisure habits) based on the customer data.

- Making tours around Central Park and Lower Manhattan is quite popular among visitors and locals as well.

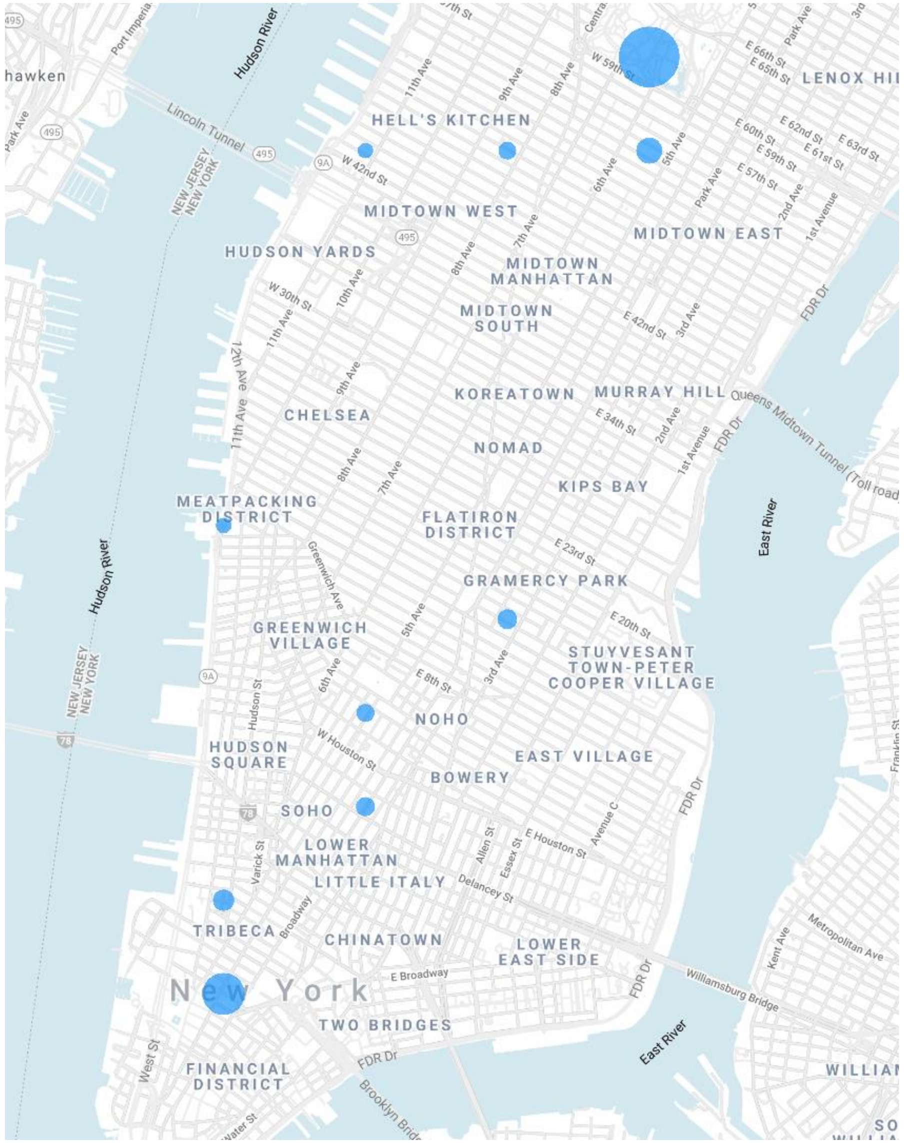
- Manhattan Midtown might be the main point of interst.

NYC top areas

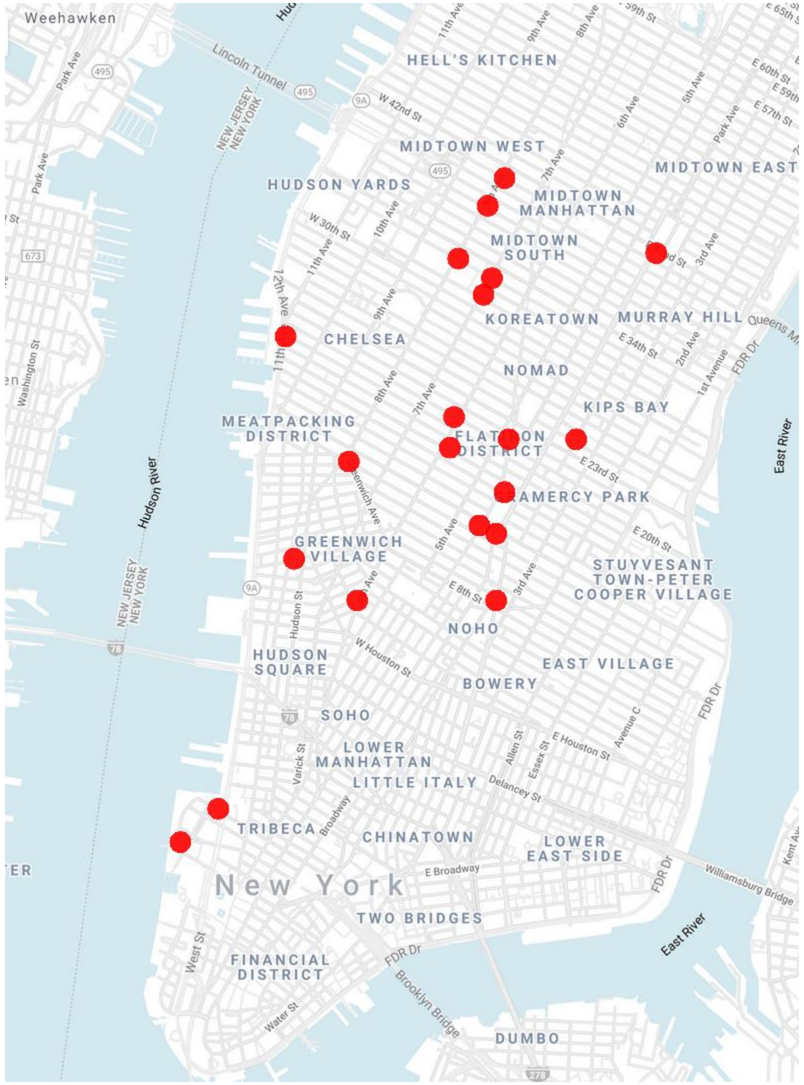
Now we will further explore which areas are most popular in New York, given certain conditions. First, we will see which areas are most popular among subscribers and customers in general, and then we’ll play a variation on a theme, by changing settings.



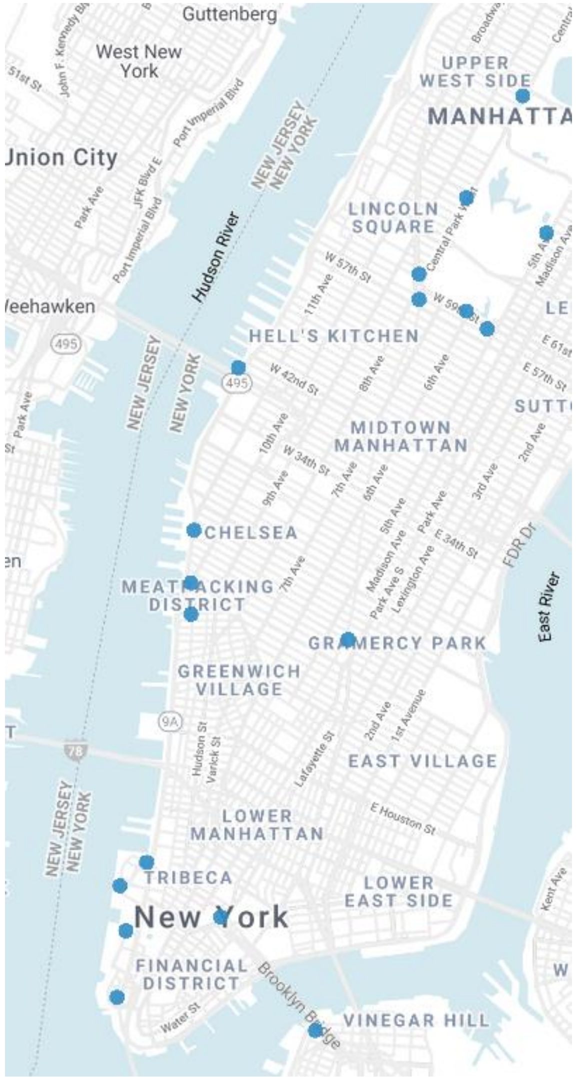
Most popular NY areas,
based on end stations and subscriber data



Most popular NY areas,
based on end stations and customer data



Most popular NY areas, based on end stations and subscriber data



Most popular NY areas, based on end stations and customer data

Location charts on the previous slide were made by counting how many trips were finished in a given area. Each circle represent a center of an area. For the first slide, bigger circle means that more trips were finished there, i. e. we can assume that the area is more popular. By comparing them, we can conclude the following:

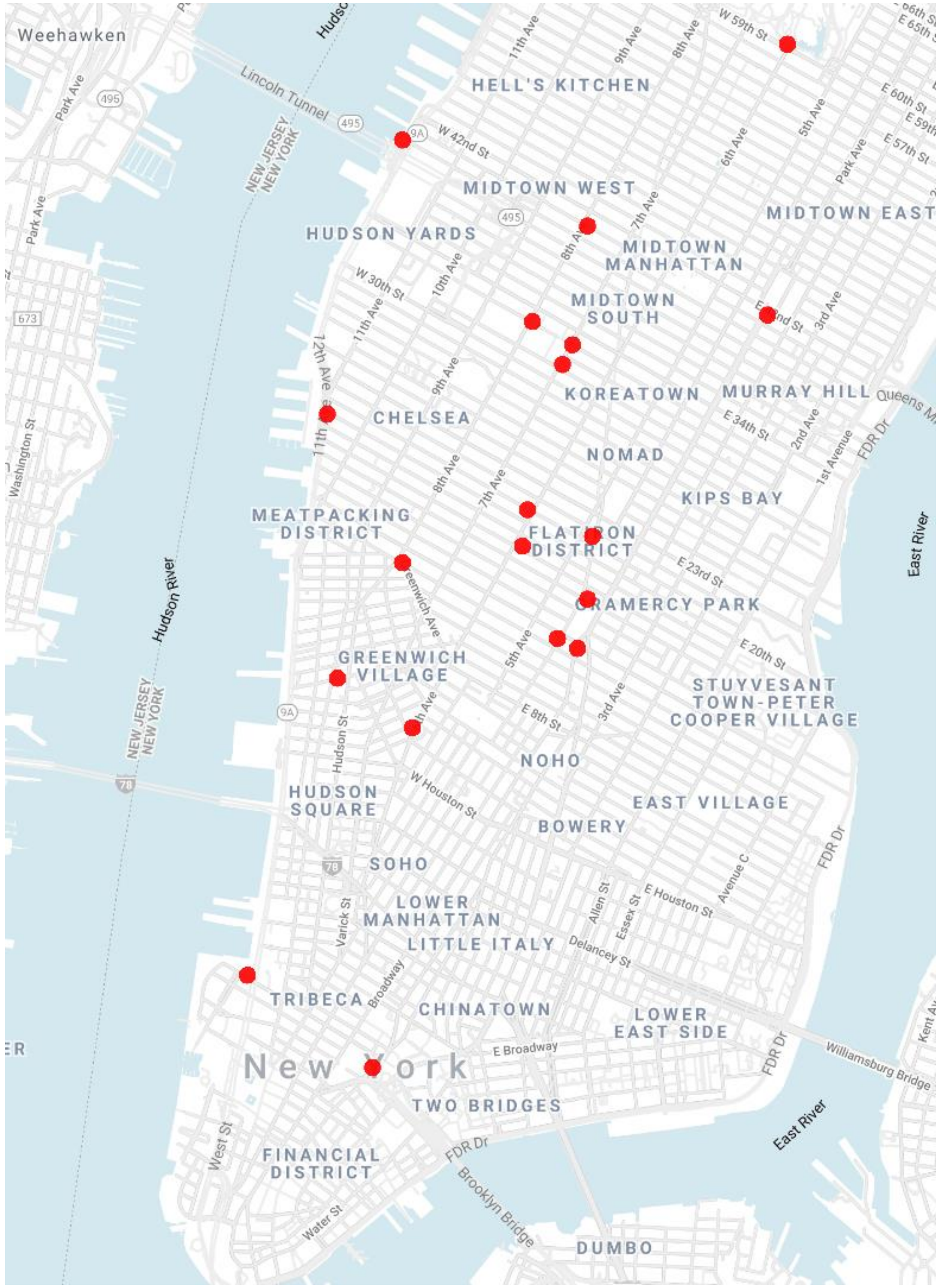
- Locals tend to spend more time in Midtown and Lower East Side, with area around the Union Square being the most active.

- Whole span form E Huston Street to E 42nd St, along the Broadway, is most popular among locals, with the area around Union Square being the biggest point of interest.

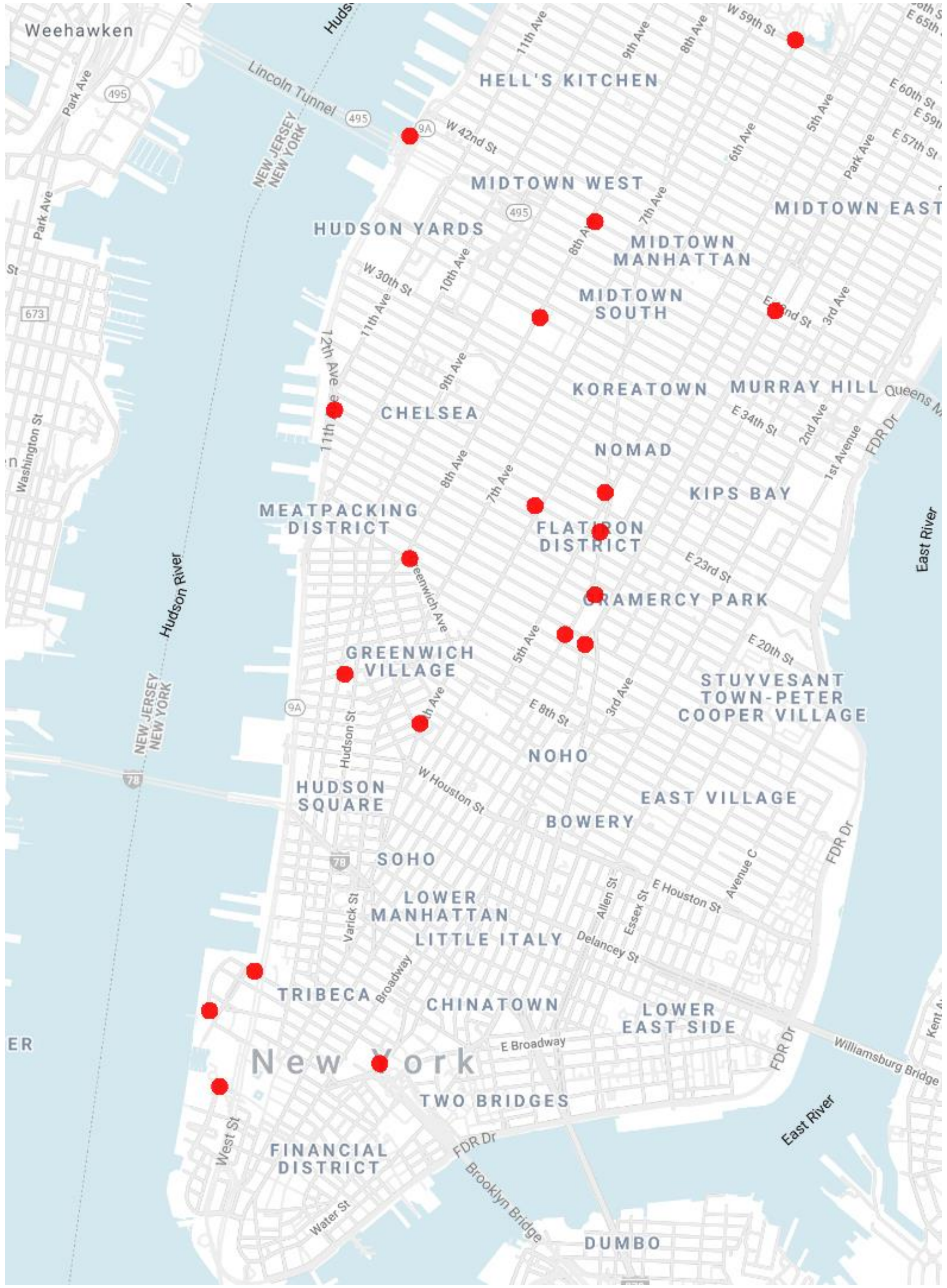
- Visitors, and locals who want to recreate, prefer touring around Central Park, and alongside Hudson river.

Autumn in New York

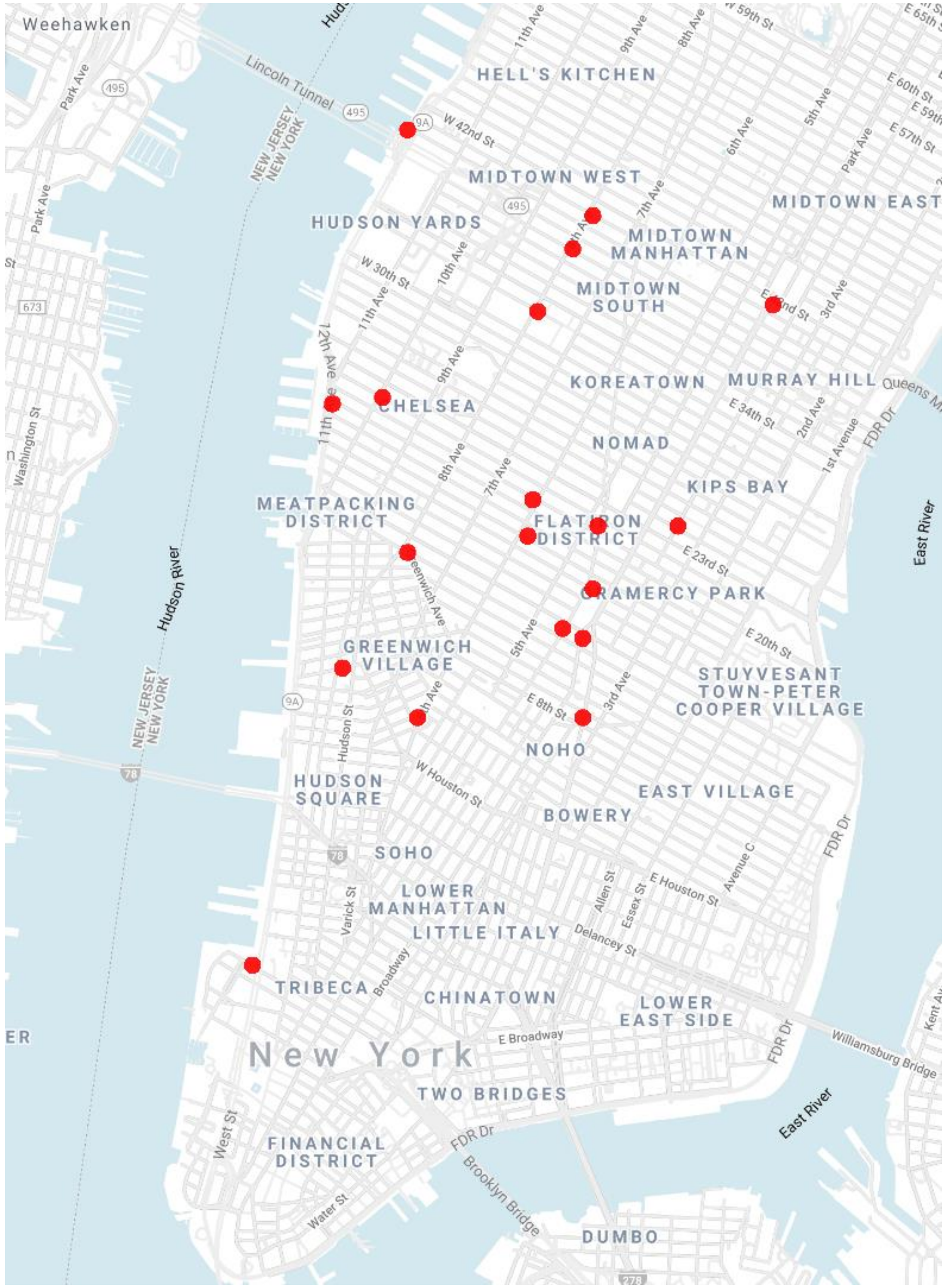
Let’s now explore how trends change over seasons. First we’ll focus on NY locals.



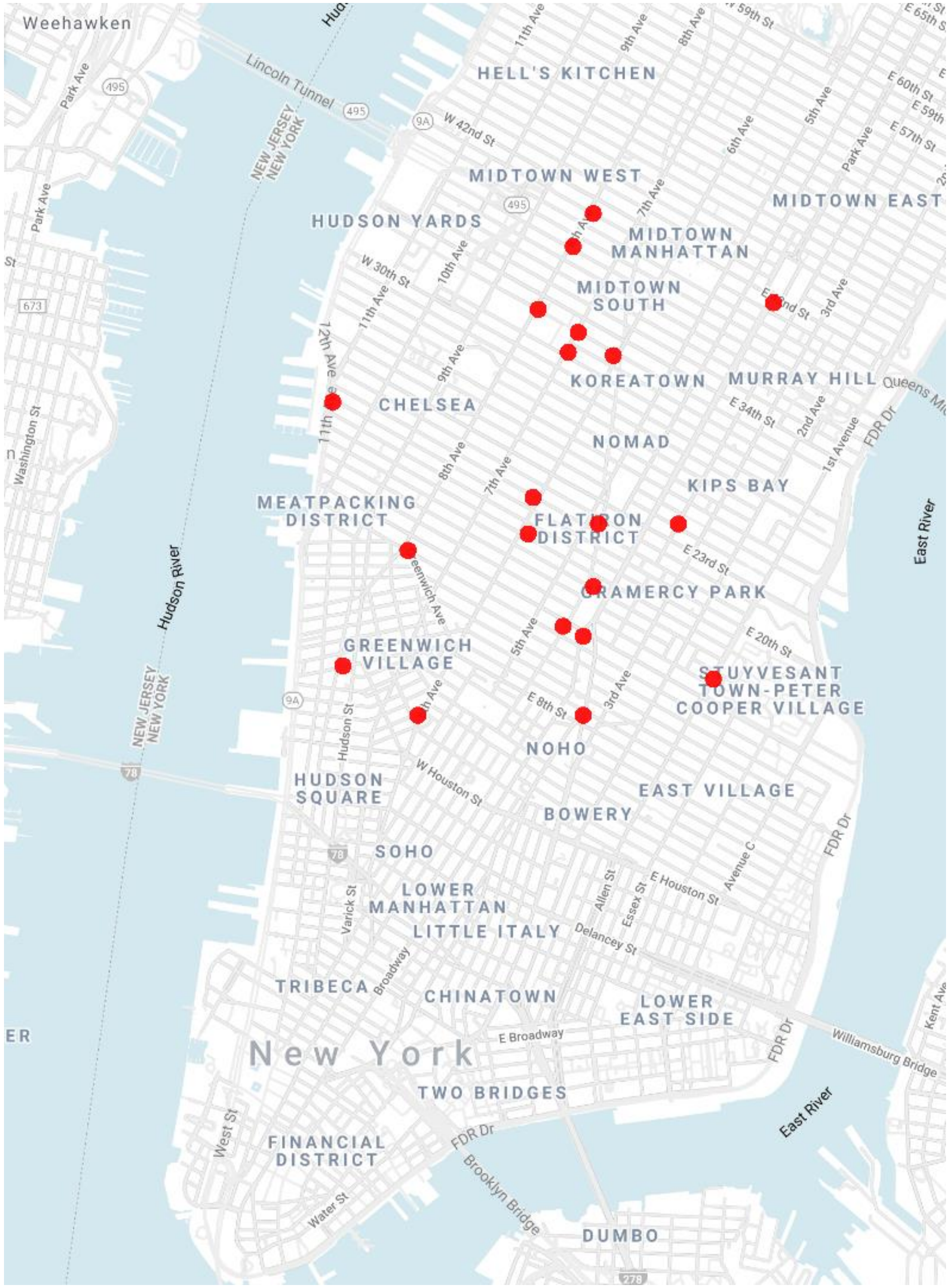
Spring



Summer



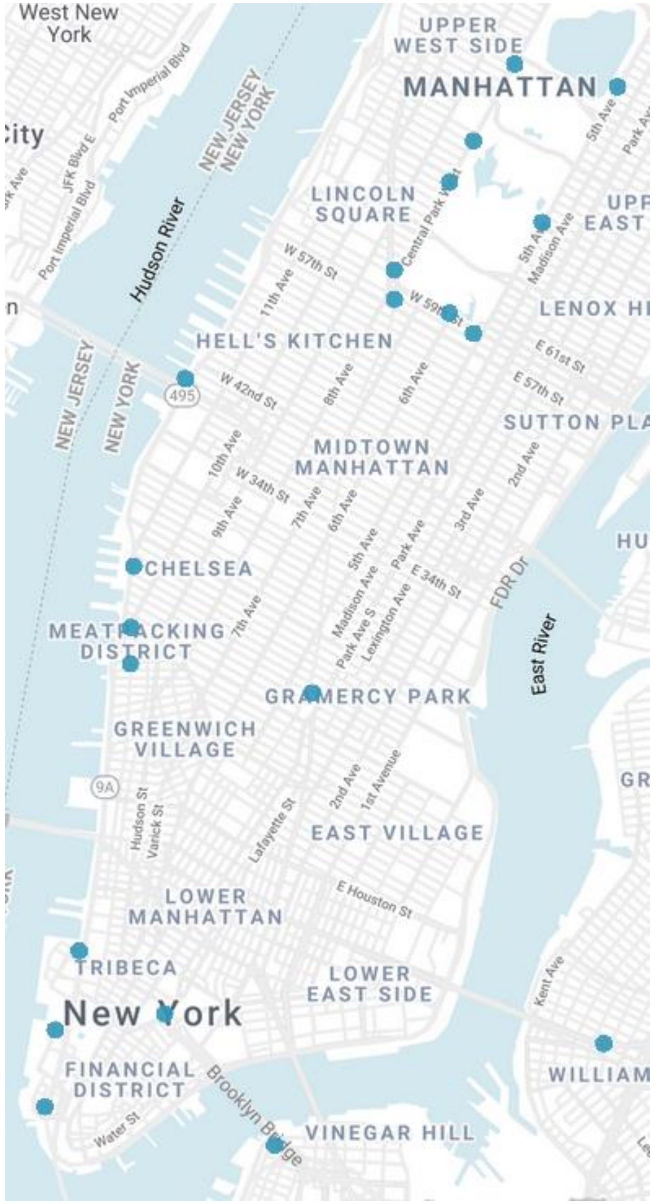
Autumn



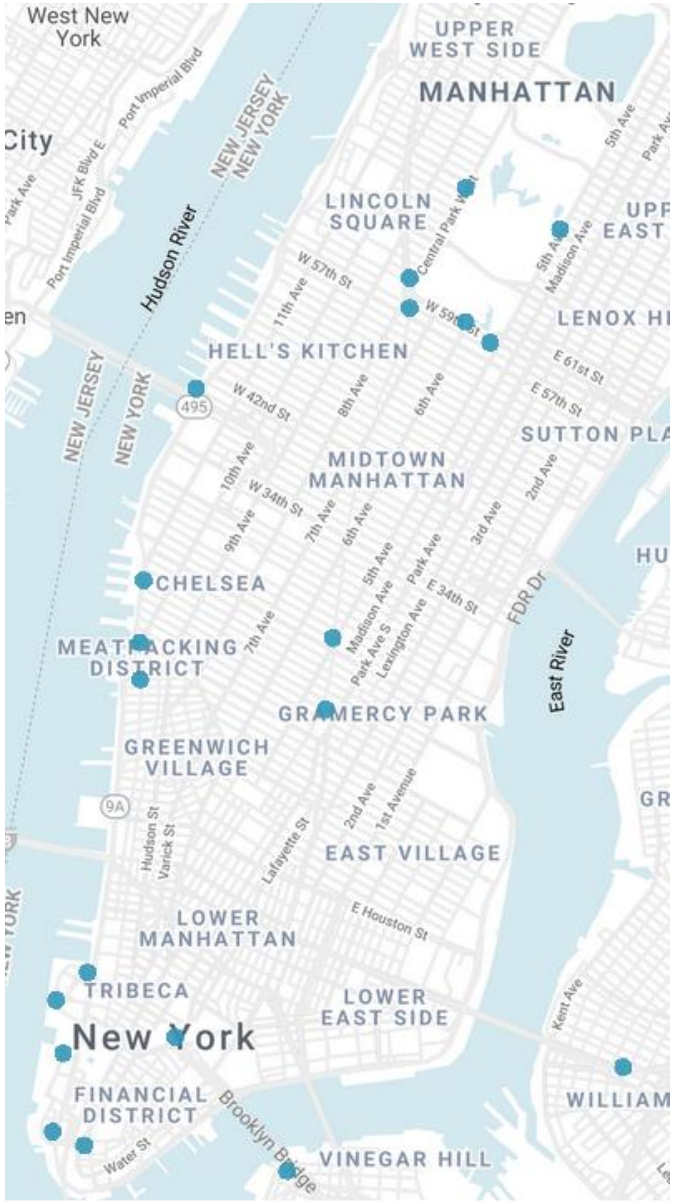
Winter

As weather gets colder, we notice that New Yorkers will get away from the river coast and converge to the Midtown.

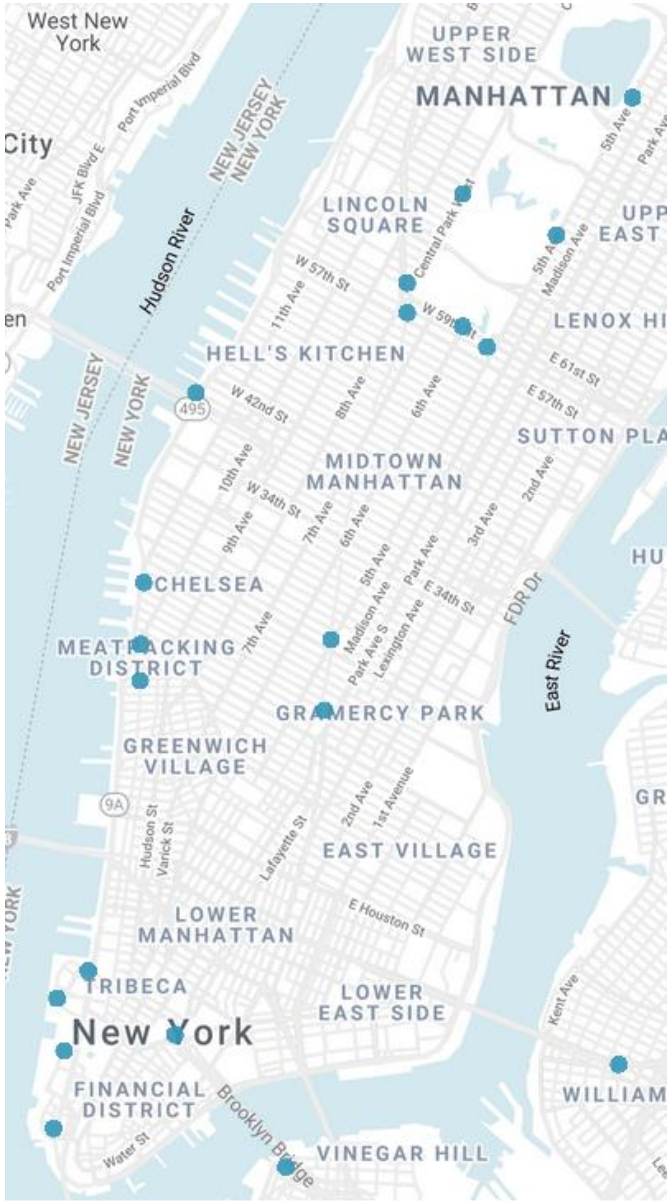
Let’s now explore customer data.



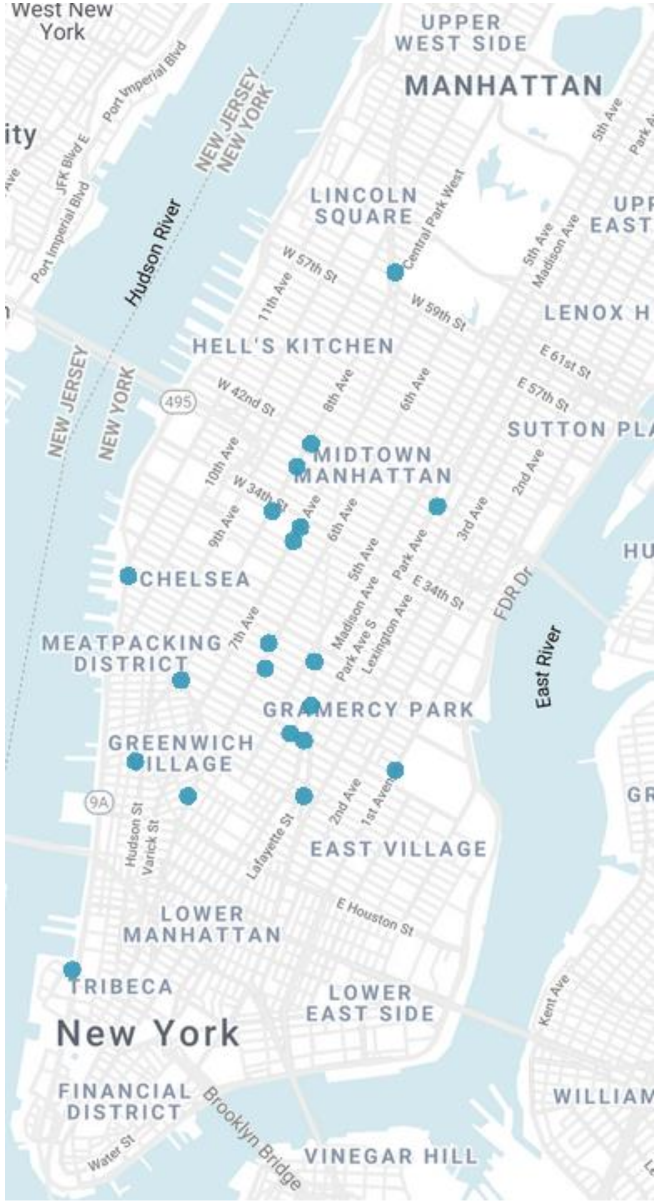
Spring



Summer



Autumn



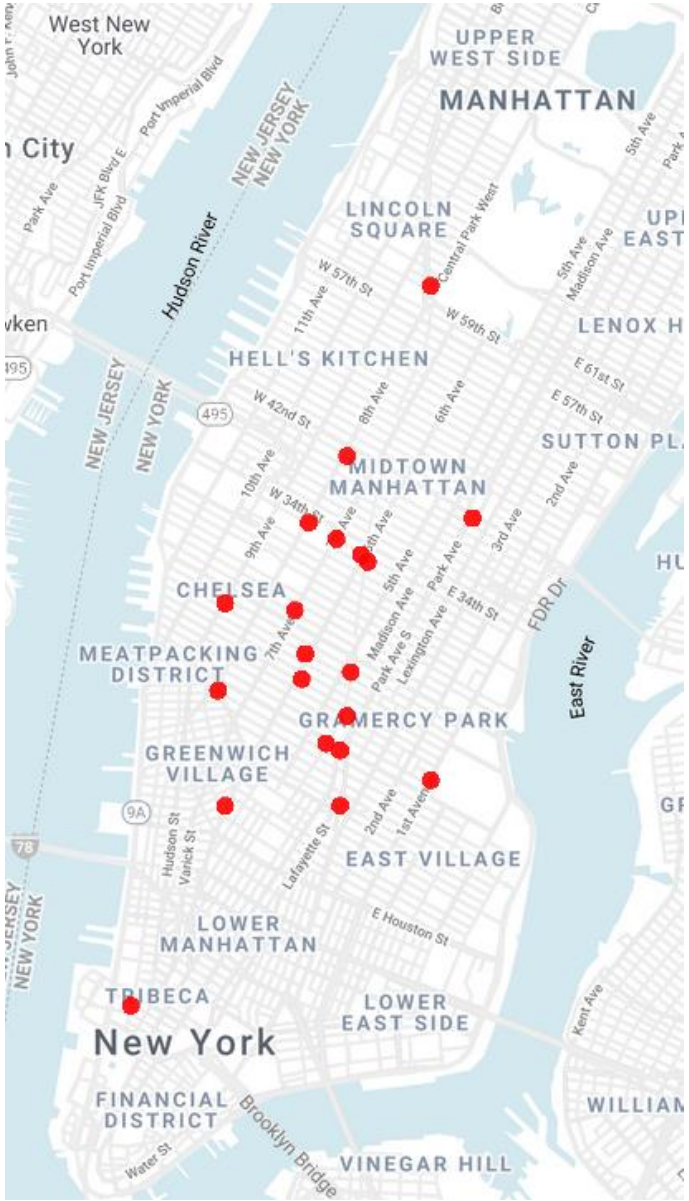
Winter

The similar trend applies to customers: they will prefer to group around the Midtown as temperatures start falling.

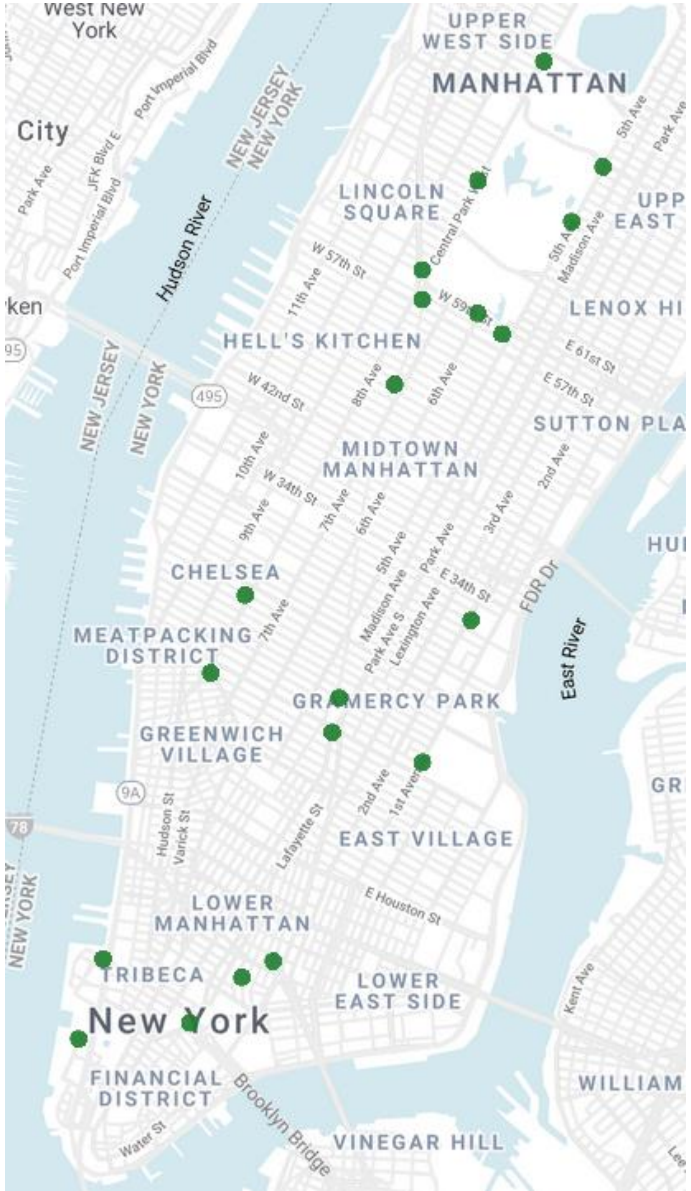
During the warmer days people like to relax along the Hudson river and Central Park. Interestingly, Central Park is popular in autumn as it is in spring.

Celebration time

It’s fun to see which areas are most lively during the biggest American holidays: Christmas and Independence Day.

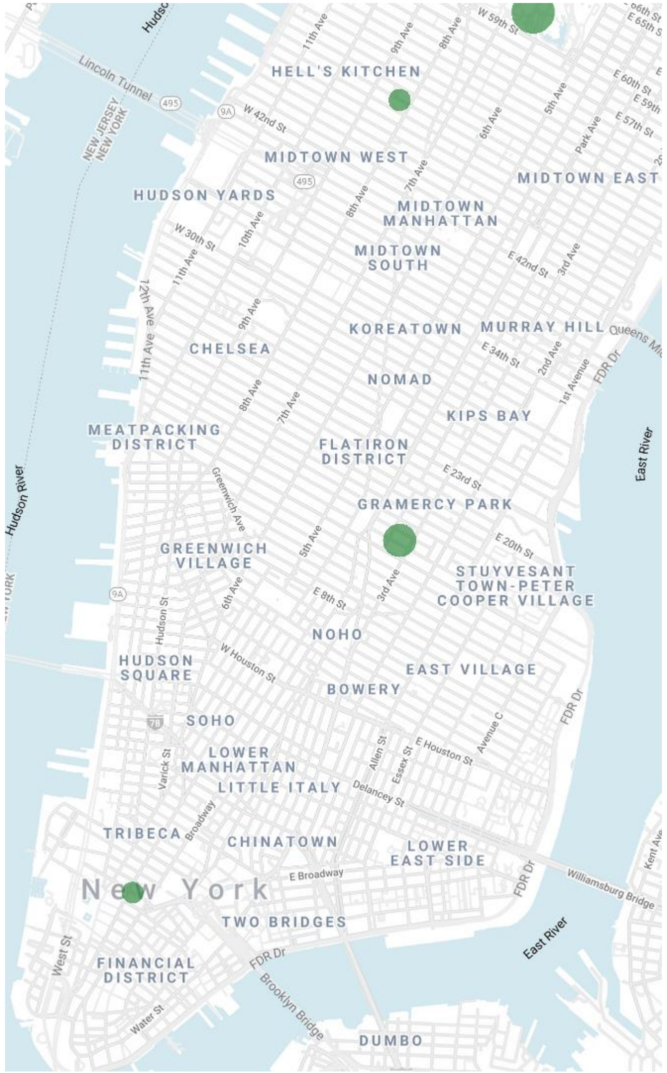


The day before Christmas

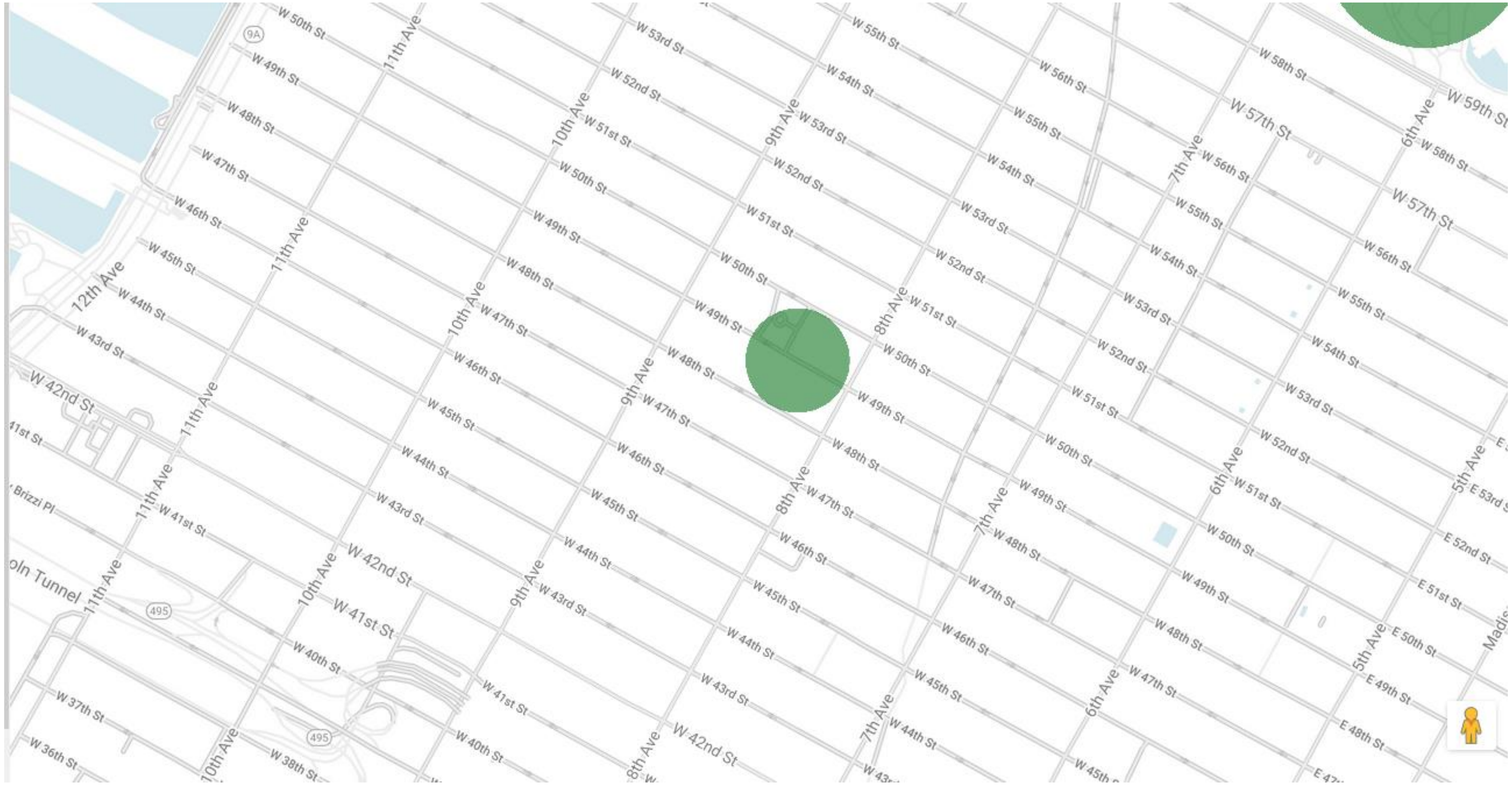


Xmas time!

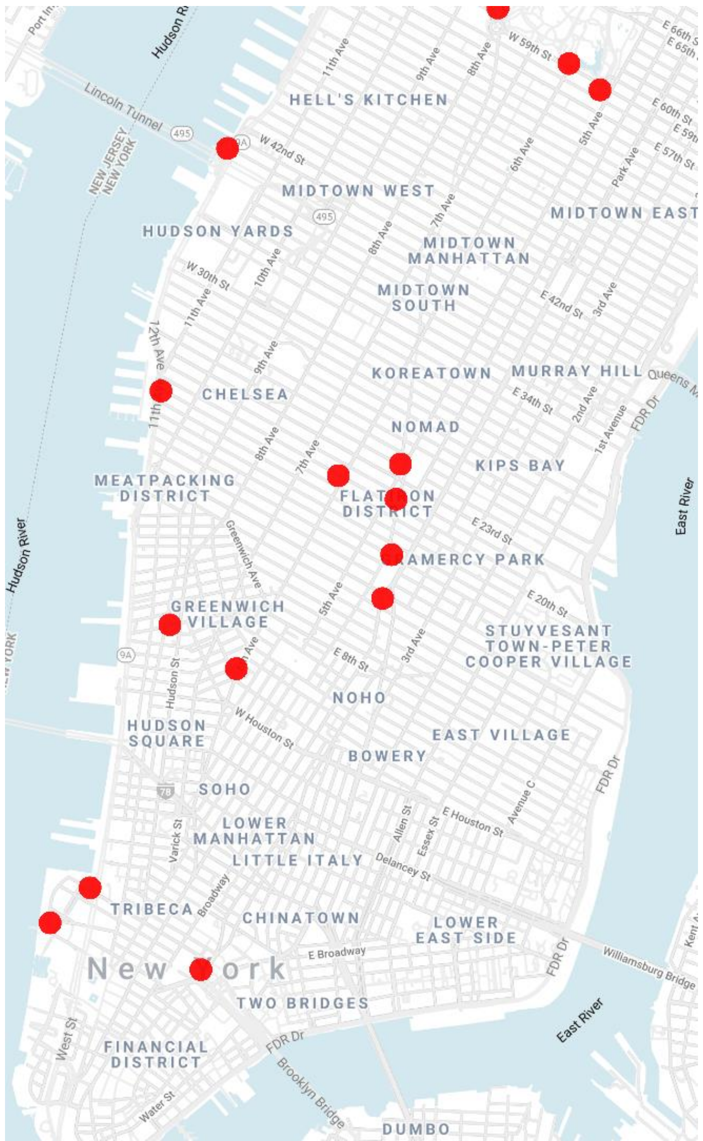
Even though it’s winter, everyone gets away from the Midtown to enjoy Christmas in Central Park and goes even close to the river! Let’s see the greater top four most frequented areas on Christmas:



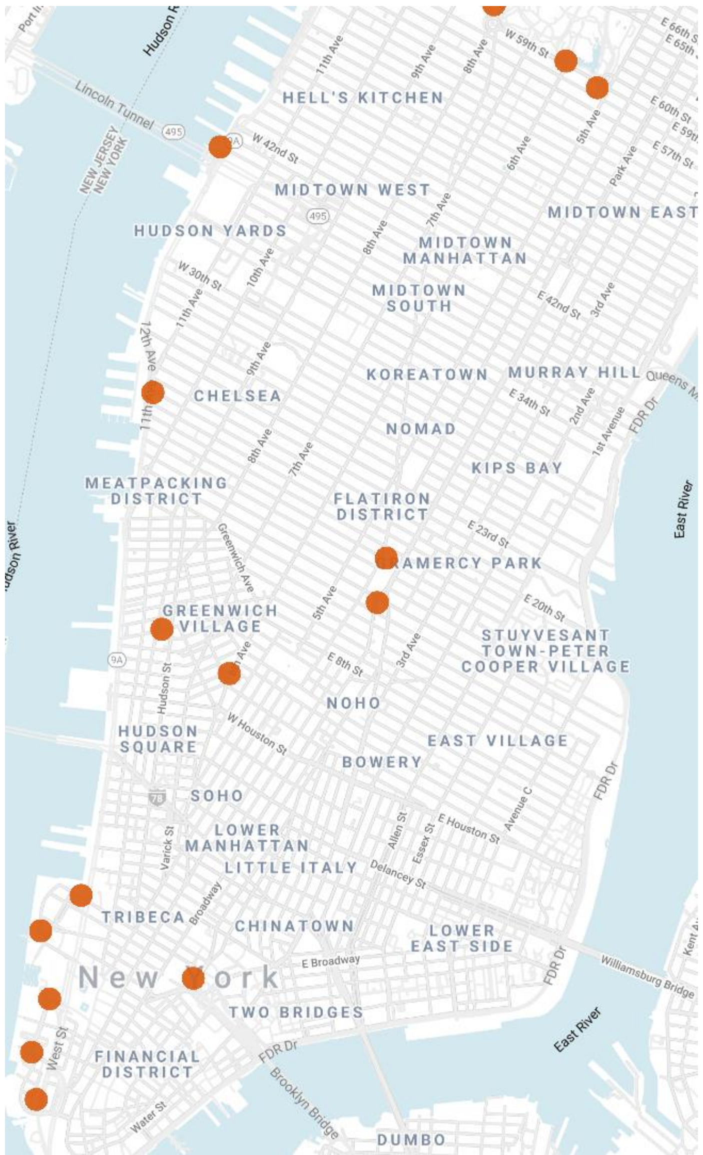
Three of the results are pretty clear from the previous picture: Central Park, the ever-active Union Square and southern part of the island. But what about the fourth point, the one near Hell’s Kitchen? Let’s zoom on it.



It centers on the W 49th Street, exactly the street of the Rockefeller Center where each Christmas a giant Christmas tree is displayed, which everyone comes to see.



3rd July

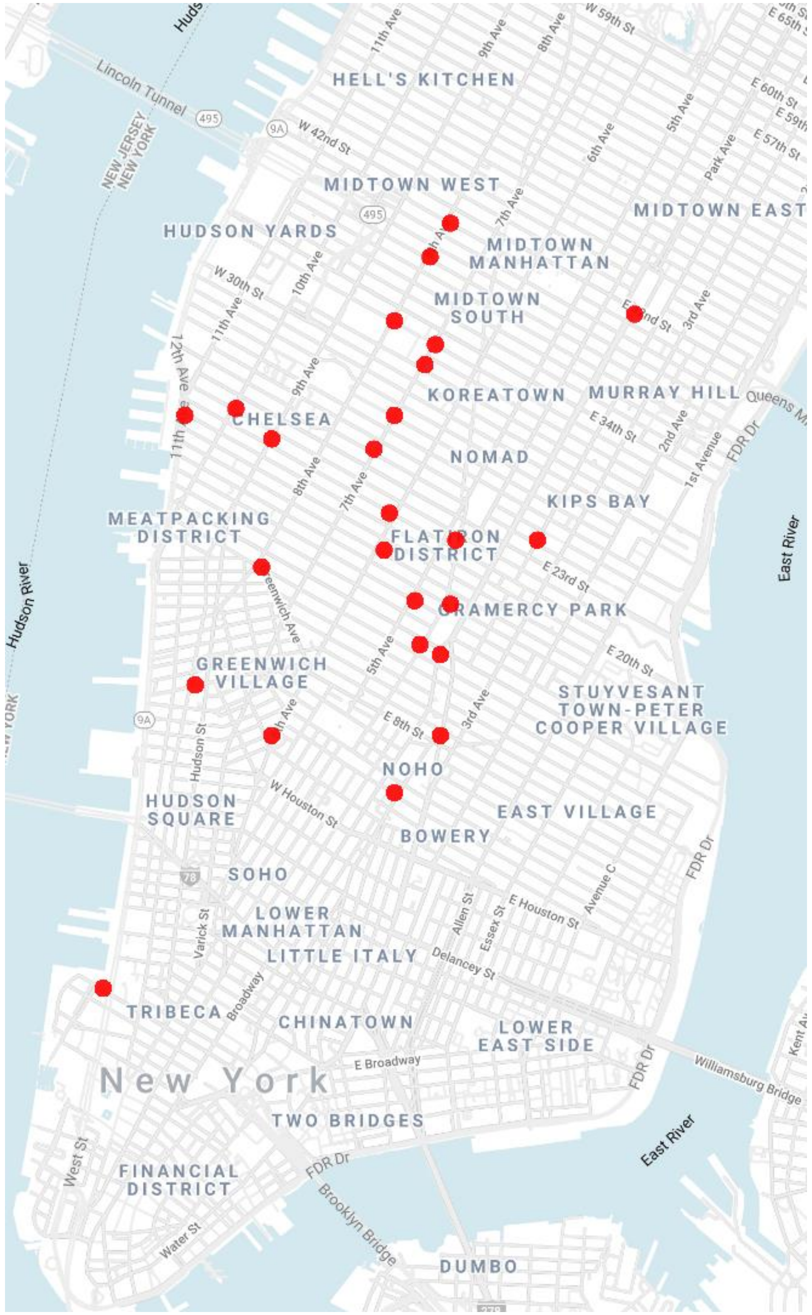


4th July

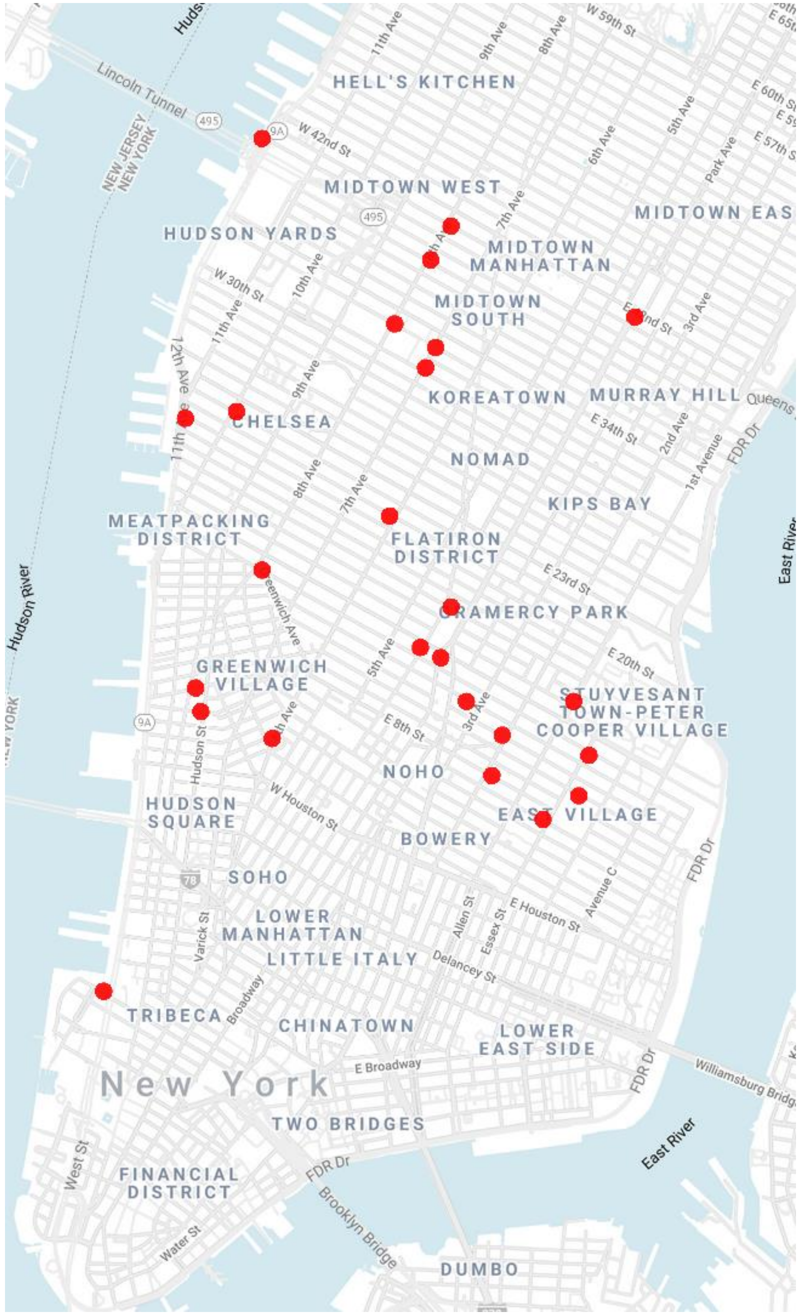
Everyone gets away from Midtown to enjoy fireworks along the coast during the 4th July!

NY ‘round the clock

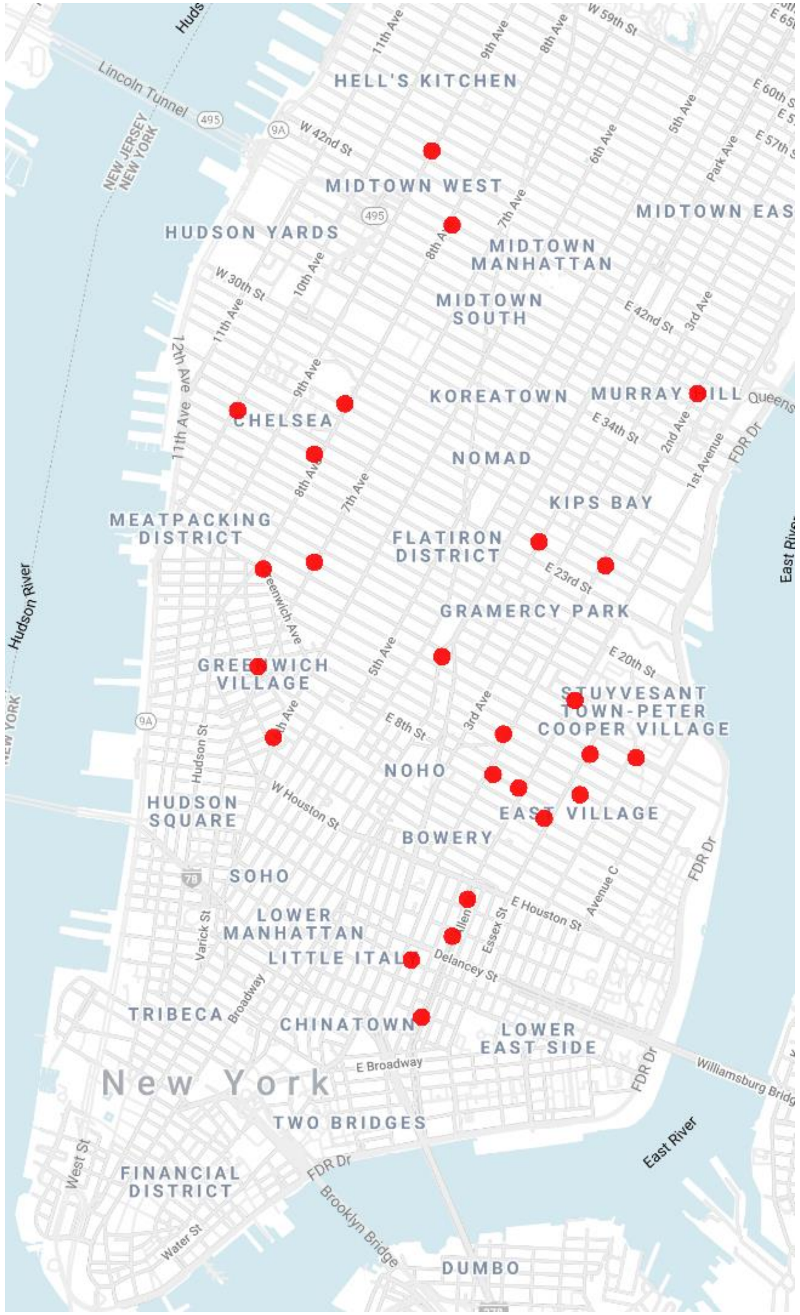
Let’s now examine city’s daily rhythm.



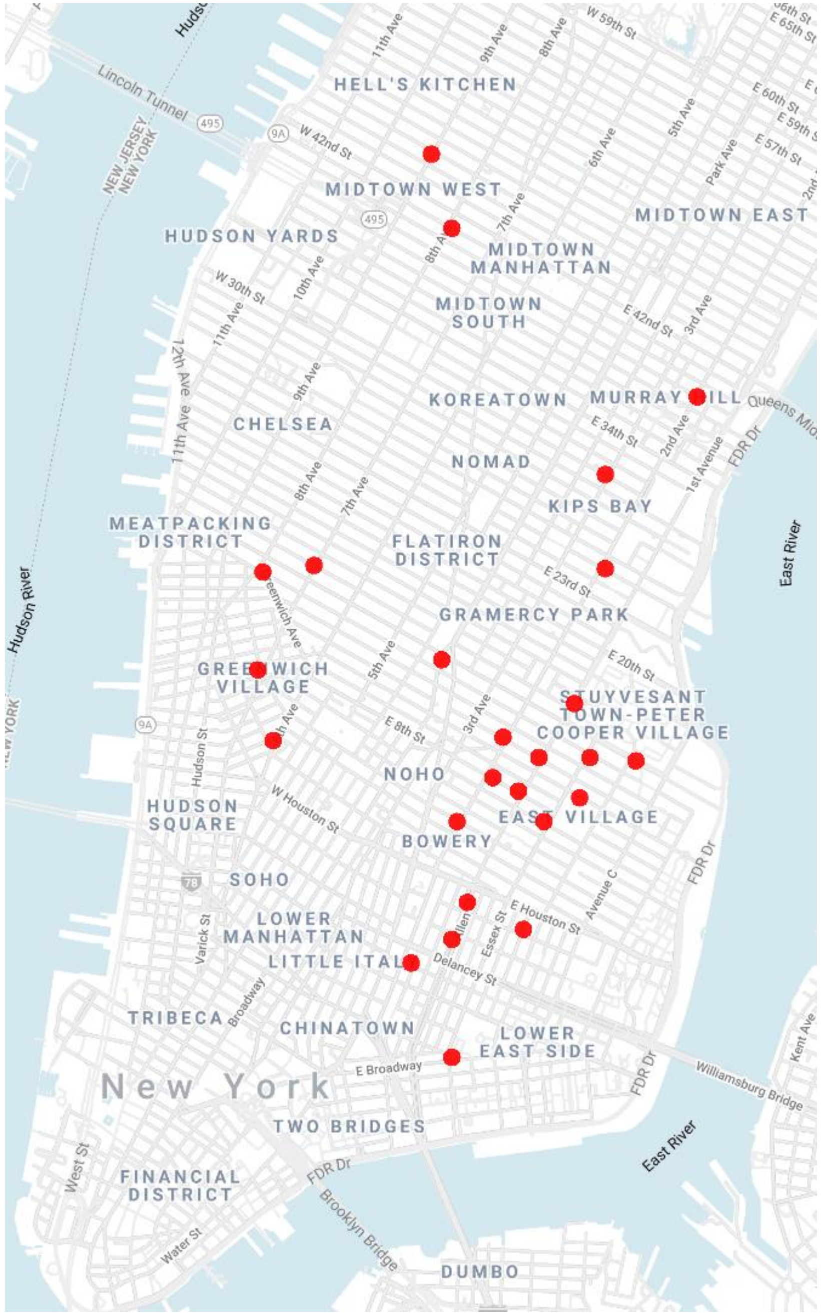
Work hours
(9 – 17h)



After work hours
(17 – 21h)



Before midnight
(21 – 23h)



After midnight
(00 – 04h)

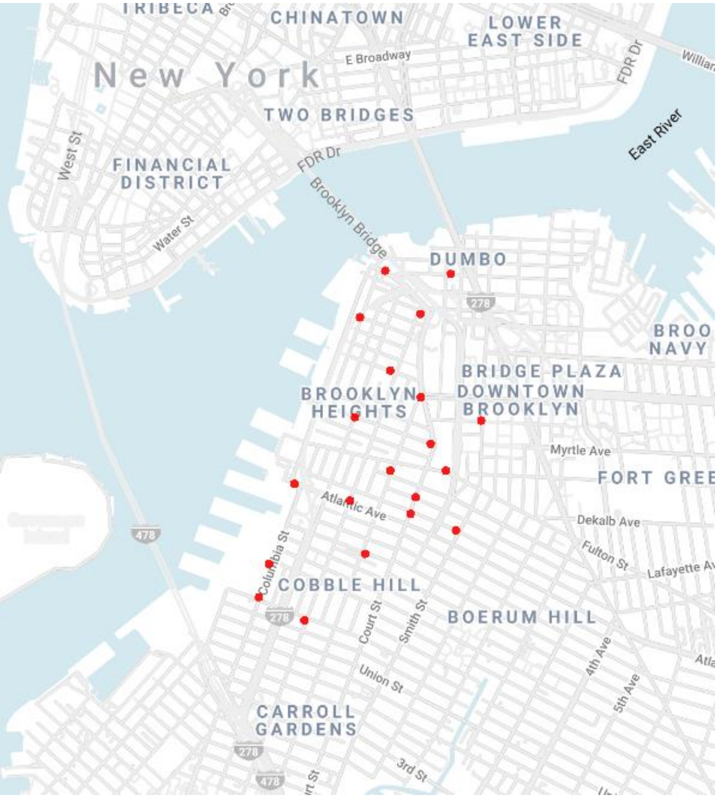
One trend is clearly apparent: as the day goes by, people are converging from the West Side to the East Side. We cam make an assumption that many people in New York live on the East Side and travel to the West Side for work and other daily activities.

Another assumption we can make is that East Side (around the Union Square) is bursting with social activities and rich nightlife.

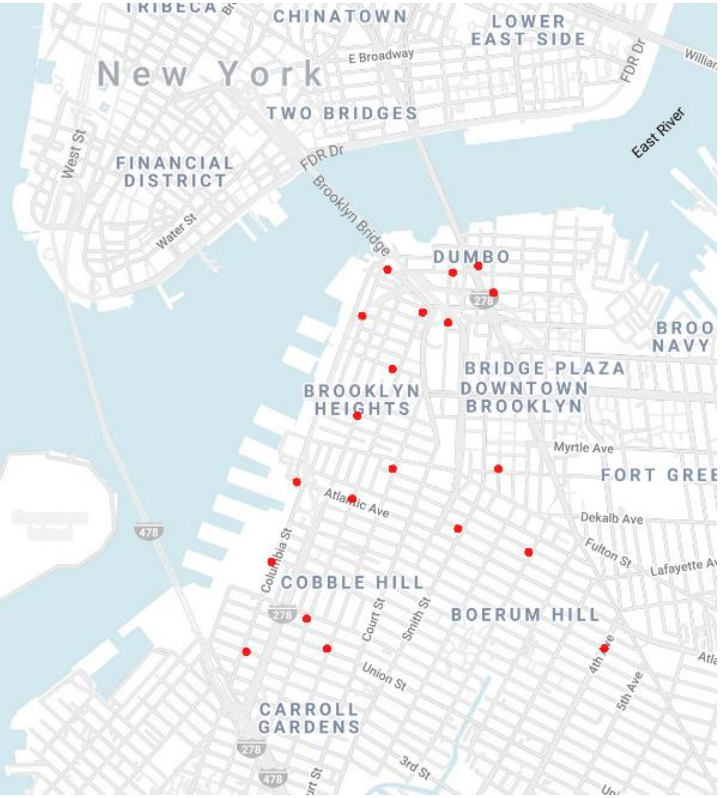
NYC ≠ Manhattan

What about the other parts of New York? We focused only on Manhattan, since it's the by far the most active core of the city. But why don't we use this data to say something about residents of other parts?

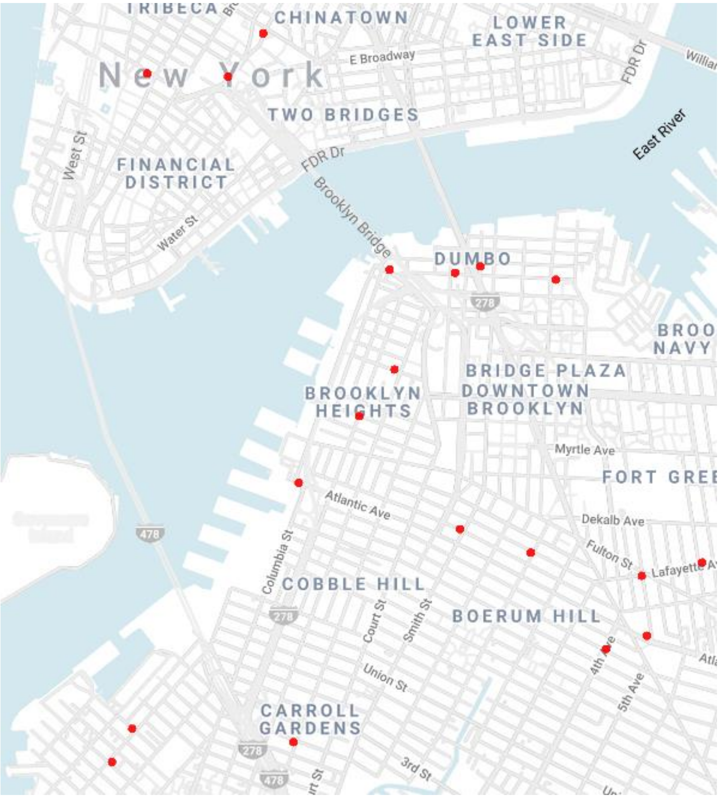
We can observe Brooklyn, for example. Let's track the routes starting from Brooklyn Heights to see where people staying there like to commute.



5 minute rides from Brooklyn Heights



10 minute rides from Brooklyn Heights



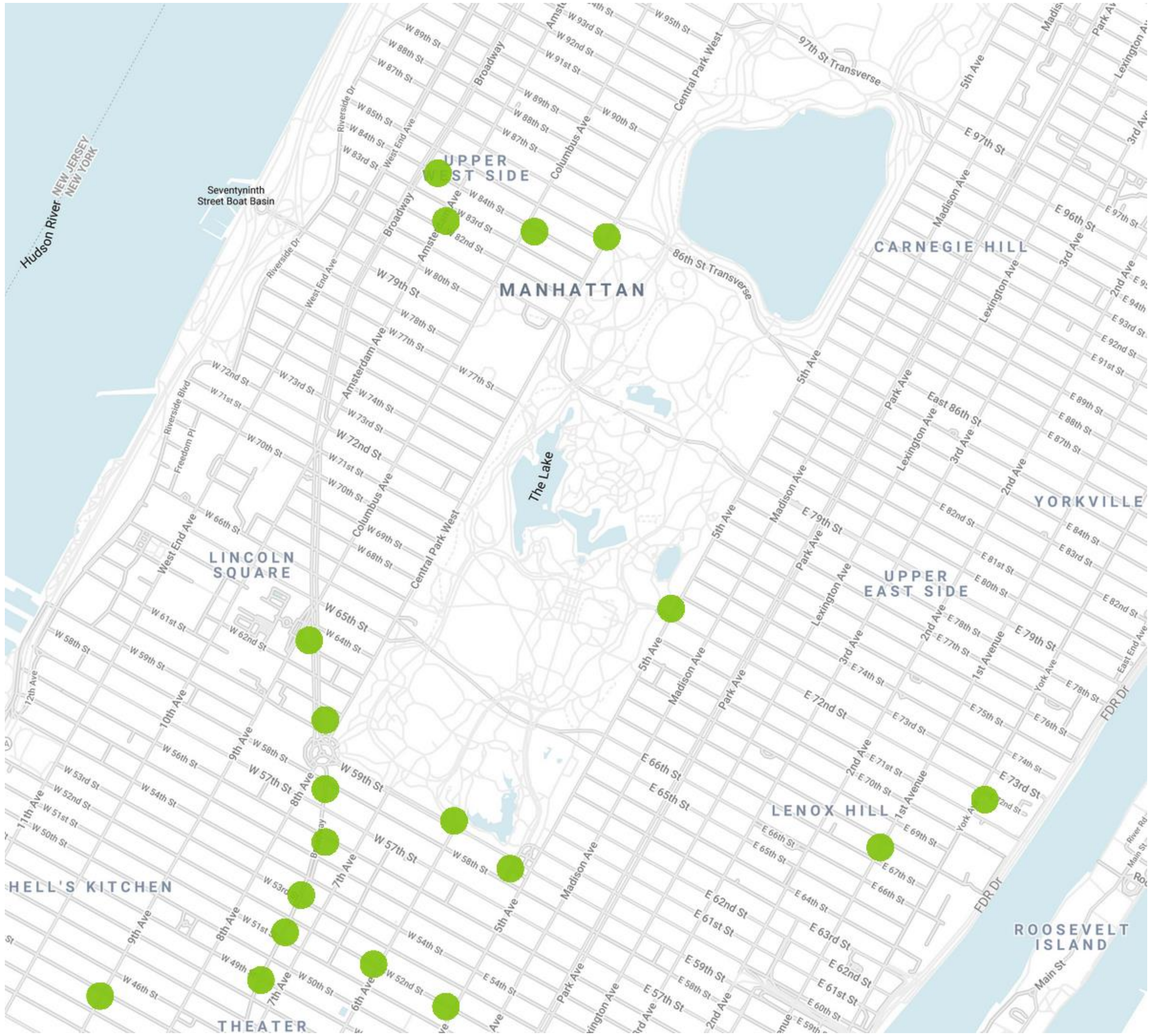
> 15 min. rides from Brooklyn Heights

People from Brooklyn Heights, usually visit Dumbo, area around Metropolitan Ave and East River coast, Chinatown and southern part of the Manhattan island.

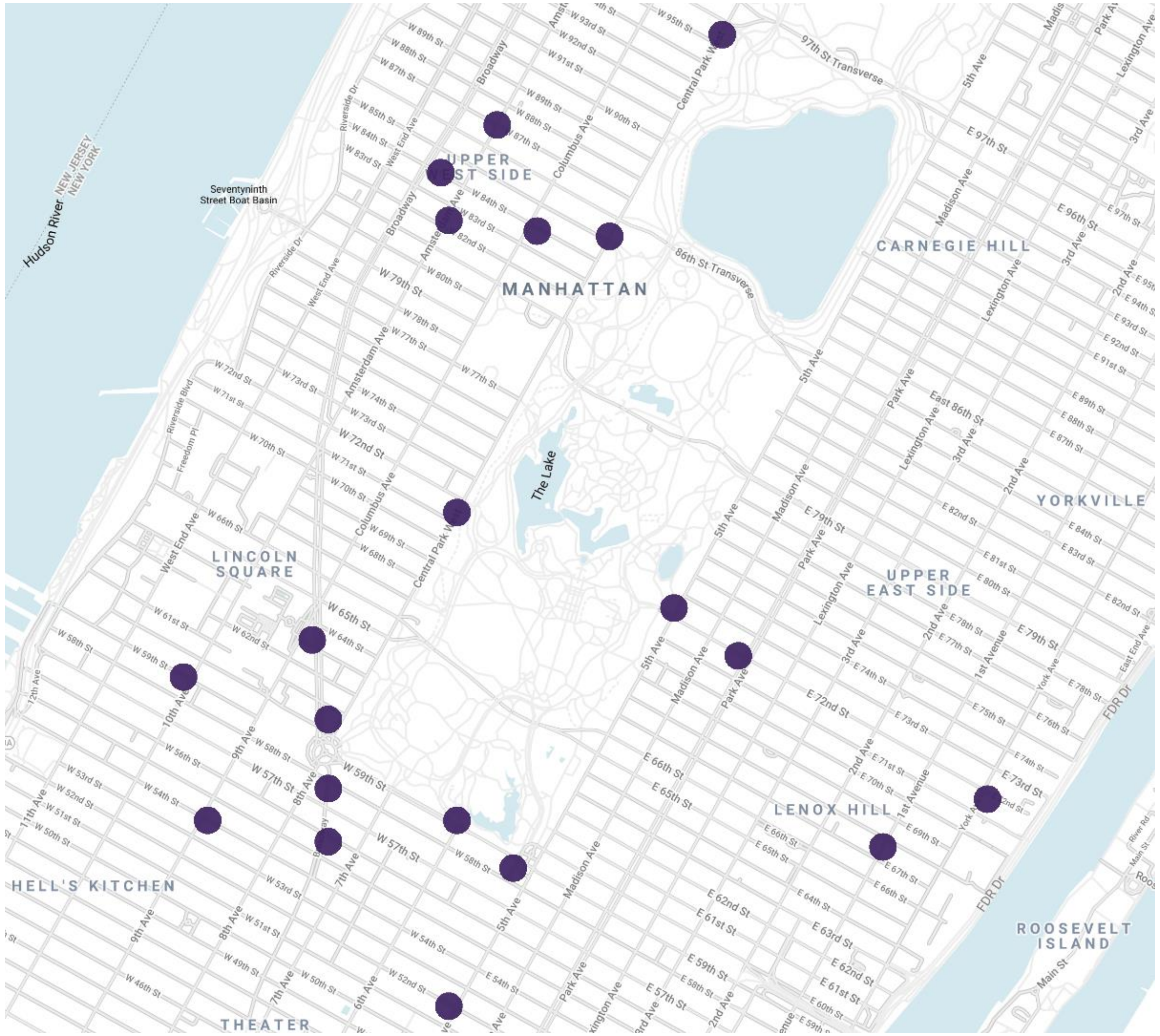
Knowing this information can be quite useful: we can consider those areas as ‘extensions of Brooklyn Heights’. Having this in mind, we can (for example) place there stores, brands and bank ATMs which are already popular among people living in Brooklyn Heights. By opening a store there, we will have some customers from Brooklyn Heights to start with, while attracting new customers from the new area.

Upper West Side Story

We saw that there is a substantial difference between male and female users. Let’s check if their habits also vary that much on some sample. We’ll do this by checking favourite routes for people living in Upper West Side. We focus only on Citibike subscribers and trips that took more than 15 minutes.



Male users



Female users

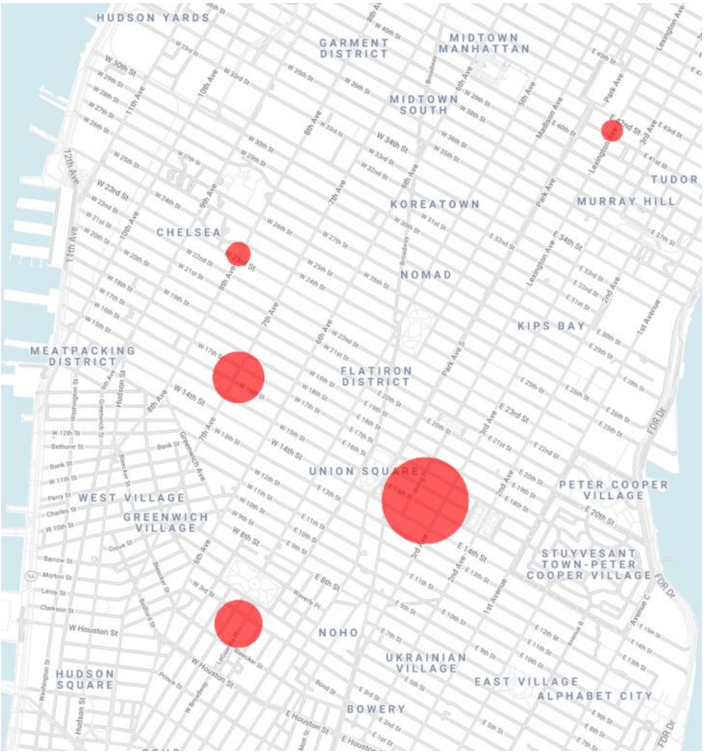
The picture is not significantly different, but we can notice something interesting: men tend to commute to Broadway more.

This can mean that Broadway is more popular among men living in Upper West Side than their female counterparts.

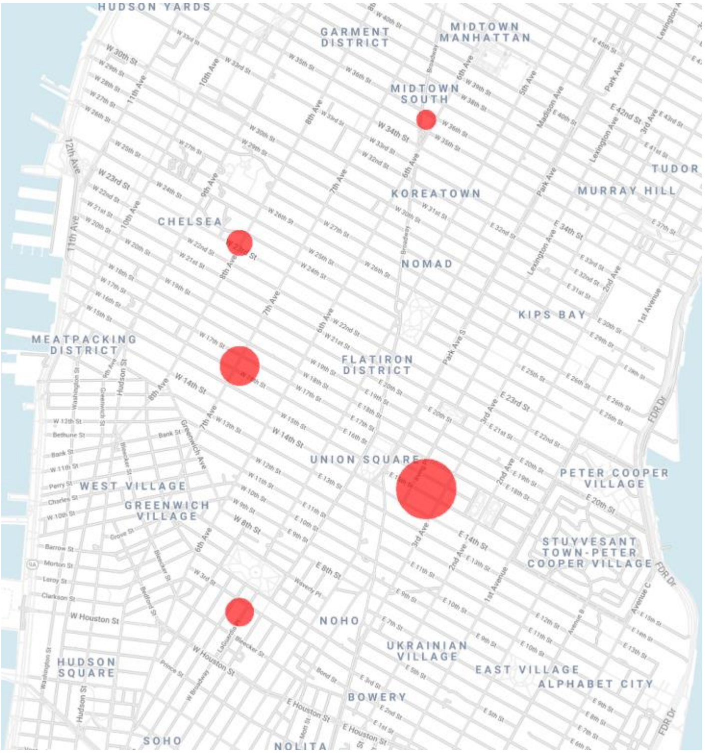
Another conclusion we can derive is that, perhaps, there are more men working in Broadway then women.

Getting famous

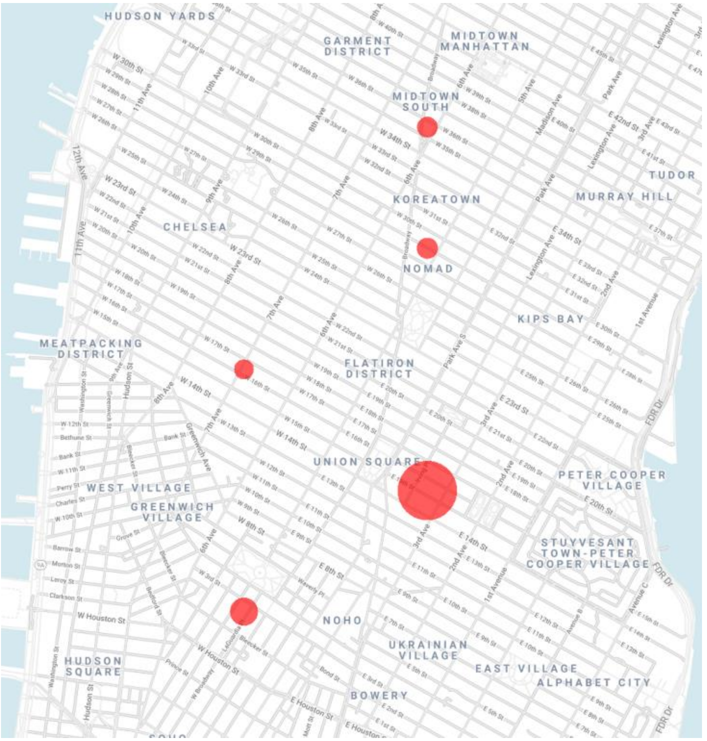
We already determined which areas are most popular. But how did their popularity changed over time? Which areas are became more famous recently? We will see how the top five areas changed from 2013 to 2015.



2013



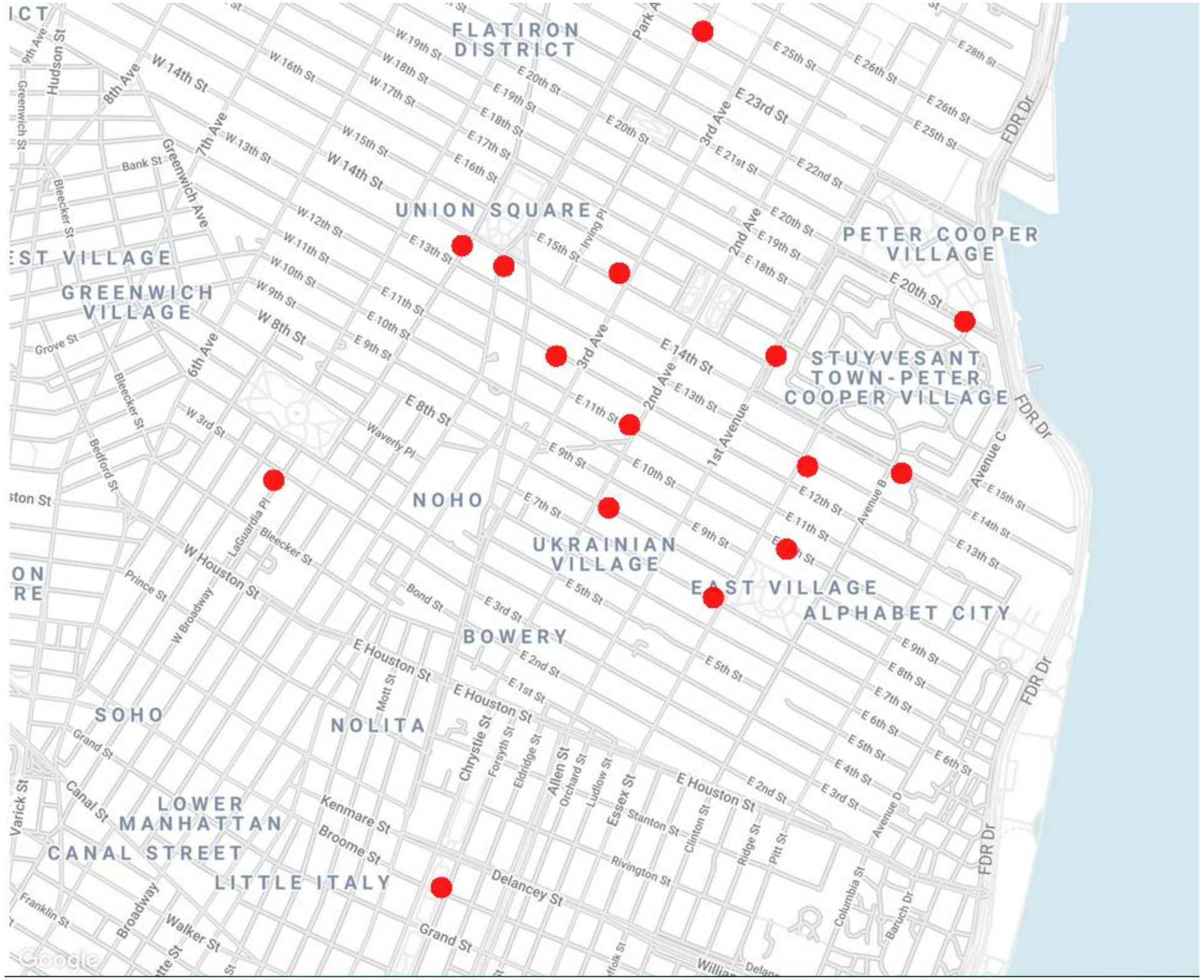
2015



2018

We conclude that Midtown South and Koreatown have been becoming increasingly popular in the recent years, with the area around the Union Square keeping its status as the most active.

To finish the analysis off, we will check where do the young people like to go out in the evening the most.



Data from people born after 1994

Area between Union Square, Ukrainian, East and Cooper Village seems to be the most visited by young crowd that goes out after 8pm, until the morning hours.

Outro notes

Here I managed to outline some of the ways we can utilise the given dataset and draw insights about the dynamics of life and habits in NYC. One of the possible ways to further explore the dataset is to combine the presented approaches in different ways; we can, for example, explore where people from New Jersey like to go out, or women from Upper East Side like to recreate.

There are, of course, many other different ways how this dataset could be used, especially if we manage to come across some additional data such is subscriber’s profession, whether they use some other means of transport, and so on.