FELIX RENÉ FRITZ

BIKE SHARING IN NEW YORK CITY: HOW THE CITI BIKE SYSTEM SERVES POINTS OF INTEREST

UNIVERSITY OF TWENTE.

BACHELOR OF SCIENCE THESIS f.r.fritz@student.utwente.nl

Advanced Technology Database Group 2017

BACHELOR ASSIGNMENT COMMITTEE

Chairman	Dr. ir. Maurice van Keulen
Supervisor	Dr. Doina Bucur
External member	Dr. P.J. Dickinson

ABSTRACT

With an increasing popularity, bike sharing systems are getting deployed in more and more cities around the world, offering alternative transport methods in cities with busy traffic and increase public health. Since 2013, Citi Bike operates in New York City and publishes usage data of the system online, offering insights into the system. This system data was mined and the top 50 rides between two stations for three defined plateaus were aggregated. Based on these rides, this thesis investigates how the system is serving common points of interest (POI). Every station was classified with up to three POI types describing it and its surroundings. An interactive application (www.nycmap.bike) was built as a tool for and to assist with the research, with a map visualizing the top rides as a weighted graph, while offering insights into trip durations and times, user type distribution and user specific data such as gender and age. A strong relation was found between areas of housing and nearby transportation hubs, serving each other and being part of last mile rides from and to work. An overall increase in female ridership was noted. Popular rides among tourists include rides around Central Park, over the Brooklyn Bridge and alongside West Side Highway, and are found to be either in areas with less motorized traffic or dedicated bicycle lanes, similar to rides with a high share of female ridership.

ACKNOWLEDGEMENTS

I would like to thank Dr. Doina Bucur for supervising this assignment and for her assistance with with obstacles which came up during the research.

CONTENTS

1	INTRODUCTION 1.1 Overview	2 2 3 4
2	RELATED WORK 2.1 Trip Aggregation	6 6 6
3	RESEARCH QUESTIONS	8
4	METHOD 4.1 Overview	9 9 10 11 12
5	 RESULTS 5.1 Central Park	17 19 21 21 23 25 25 27 29
6	CONCLUSION	31
7	DISCUSSION	32
A	APPENDIX	37

ABBREVIATIONS & TERMINOLOGY

Point of Interest

A user with either a 3 or 7 day pass to the Citi Bike System A user with an annual subscription to the Citi Bike System Bike Sharing System POI Customer Subscriber BSS

1 INTRODUCTION

1.1 OVERVIEW

In the last couple of years, bike-sharing has taken a major role in the transportation network of big cities as an alternative way of getting around the city. It consists of a network docking stations (Figure 1) from which bikes can be borrowed and returned to, where it does not matter from which station a bike is borrowed or where it is returned.



Figure 1: A Citi Bike station with bikes locked into place. [1]

People interested in using the system can sign up for a membership, often offering unlimited bike rides for a yearly fee, where there is usually a limit on the duration of borrowing a bike, to prevent users from just keeping the bike instead of returning it to a station. Furthermore, access passes spanning just a couple of days are offered as well, targeting tourists who want to explore the city by bike.

Originated in 1965 in Amsterdam, the first bike sharing system (BSS) was establishing by a political group called *Provos*, which placed *witte fiets* around the city, white painted bikes which anyone could take to get around. The goal was to decrease pollution and congestion in Amsterdam. However since there were no return stations for the bikes and the systems regulations were rather simple, it collapsed after a short time [2].

The first station-based system was introduced in Denmark in 1991 with four stations, and in 1995 *Bycyklen* was established in Copenhagen as a first system on a bigger scale [2]. This was also the first system to emphasize on urban usage, providing bikes with more solid tires designed for the everyday use in the city. While solving problems of the earlier systems, *By-cyklen* had issues with bikes being stolen, which was a result of the system only requiring a

coin as deposit. This was resolved in a new generation of BSS which was established at the Portsmouth University. It used magnetic cards to authenticate a user and gave only those who could authenticate themselves access to a bike [3]. The most recent generation of such systems combine the positive aspects of their predecessors, offering sturdy and easy to operate bikes to their registered members.

Considering the positive effects on the environment, public health and overall traffic, bikesharing offers a way of traveling around the city which relieves pressure from the congestion and results in less impact on the wear of roads. This makes the incorporation of a bike-sharing system attractive for many big cities facing those issues.

Looking at New York City and especially Manhattan, which doubles its population to 3.9 Million during work days due to commuters [4], the city's Department of City Planning in 2008 determined several points of action as it was investigating into transport alternatives for the city, and in order to decrease traffic related issues such as road wear, collisions and congestion. This plan involved the expansion of bike lanes throughout the city and installation of bike racks [5]. One year later in 2009, the Department published its 'Bike-Share Feasibility Report', where it investigates in the opportunities of a bike-share system within New York City. It was found that for New York City, the station density would be a major factor of the potential success of such a system [6]. In 2011, the city then decided to partner with 'Alta Bicycle Share' to build and run a BSS [7].

Citi Bike, New York's BSS, received great popularity after its launch in 2013. With initially 330 stations and 5,000 bikes, Citi Bike has expanded to 470 stations and 7,000 bikes in spring 2016 [8][9]. It plans to add another 5,000 bikes and 280 stations during the course of 2017 [10].

Citi Bike announced plans to further expand their system throughout the years in the boroughs of Manhattan, Brooklyn and Queens by increasing station density and tapping into neighborhoods where there are no stations yet. Next to that, a main goal is to create a reliable transport network by decreasing the distance to a station in the areas where Citi Bike already is [10].

1.2 THE PROBLEM

Several issues can arise during planning and operation of a BSS. Estimating the demand at a station can be one of them, since the it is dependent on many factors like nearby points of interest (POI), the local demographic and the type of area. There can be great demand during rush hours at major stations nearby apartments or transportation hubs, such that there are only a small number of bikes available outside of rush hours, until customers bring back the bikes on their way home and thereby re-balancing the system. But weather and spatial variables can also have impact on the decision to use the BSS, such that the bikes might not be brought back in this particular example. Stations nearby offices could potentially only experience heavy usage within a short period during the day, given that it mostly holds bikes for workers who went to work using the system, in which case there must be sufficient slots available in the morning. An asynchronous flow is thus almost guaranteed, which needs to be counteracted through rebalancing, a rather complex task. Citi Bike does this using bike transporters and encourages people via their *Bike Angels* program to help relocating bikes to earn points in return, which can be spent on different ways, for example sharing a ride with a friend for free or extending the membership. After a certain number of points, it is also possible to exchange them for money

[11]. However, re-balancing remains a tedious process, involving the investigation in optimal routes and times for bike transporters to re-balance the system.

1.3 AIMS & APPROACH

Over the last years, several studies researched bike usage and the flow within the network, using different approaches, such as population density, bike lane availability or retail destinations [12]. In addition to simply increasing the number of stations and their density, expansion could be further optimized by identifying top routes and analyzing how common points of interest are served. By identifying a station as a symbolic representation of a nearby point of interest (POI), this thesis tries to accomplish the analysis of the usage of shared bikes in relation to the POI. The results for a certain type of area can be mapped onto similar areas, making it possible to hypothesize whether an area is expected to receive a higher or lower number of commuters from bike sharing in comparison to similar areas. This can improve planning for future station expansion, e.g. by mapping the results of this research onto similar areas in the city. This thesis investigates the top routes, the correlation between different points of interest and user and trip specific data such as gender and trip time, which can help understand this flow and how it can be counteracted in the case of re-balancing. The results can improve the understanding of the numerical requirements of a certain area or district towards Citi Bike for future expansions. Business owners in areas of Citi Bike stations could also make use of the result, e.g. by investigating in the times when a nearby station is most busy or from where people arrive there, and then adjusting local advertisement according to it.

The goal of this thesis is identifying the top POI-to-POI routes in the recent years which people prefer to cover using the Citi Bike BSS. This is done using a directed weighted graph created from the aggregated data mined from the Citi Bike trip data set. Furthermore it is investigated whether there exists a correlation between the type of POI (such as office, tourist or apartment area) and the likelihood that it is served by the Citi Bike system, in comparison to other POIs. This graph will give a geographical overview of the role of the BSS in the overall city transportation. Additionally, it is investigated whether there exists a correlation between the users gender, the date and time of the day, and the likelihood of using the BSS for traveling to a specific type of POI. This can further help understanding user-specific requirements towards the system, and can help analyzing usage patterns of different types of users. Finally, it is hypothesized about the underlying causes of the results.

Citi Bike publishes anonymous usage data of their system on its website. Next to trip duration and user gender and age, it contains start and end station of every trip since the start of the system, offering great opportunity for an analysis of several points.

The Citi Bike data is analyzed using methods of graph theory, such as creating a directed graph from the aggregated data with according to their popularity weighted edges, representing a route. To visualize the result the graph is plotted onto a map of Manhattan containing the stations and edges between them. Moreover, the correlation between the type of POI and the popularity of it within the BSS system and its causes are interpreted. Important input variables are the POIs at first, which need to be identified and circumscribed. Stations of the system need to be related to such POI, from where on the trips to and from the stations can be analyzed in correspondence to the POI.

The related work is presented first, after which the research questions are formulated. The

method used for aggregation of the trips and mapping of the stations is described. The results for each questions are described in sections corresponding to neighborhoods of New York City, after which the results are then discussed and concluded.

2 | RELATED WORK

2.1 TRIP AGGREGATION

Investigating in related work regarding the aggregation of *top trips* within the system, not many papers come up. Gordon-Koven et al. (2014) [13] for example discusses top stations, meaning stations which had the most trip starts and ends. The result can help classify the surroundings of a station, but it does not tell about the relationship between two stations, a crucial aspect of investigating the flow of a station based system.

This research focuses its investigations on top trips *between* stations in different plateaus defined based on the number of stations in the system. These top trips represent the base of then carried out research into user data, built environment and trip time data.

2.2 USER GENDER

In Garrard et al. (2008) [14], it was investigated whether females would be more likely to cycle on routes which provide a separation from regular motor traffic. It was concluded that improved cycling infrastructure like cycling paths and a maximum separation from motor traffic is an important factor for the female representation in the cycling demographic. In addition, this agrees with the differences in risk aversion of the male and female gender.

This result is further supported by Kaufman, S. (2014) [15], who found that the number of traffic lanes and heavy traffic influences the choice of females taking a Citi Bike. Furthermore, it was found that the female user share is higher in less densely populated areas such as Fort Greene in Brooklyn. Kaufman suggests that from this, the safety of a location can be derived.

Regarding user gender variables, this research investigates in the change over time in gender distribution on several trips. It is then, based on the information available, hypothesized what the cause for a shift in the distribution is.

2.3 BUILT ENVIRONMENT VARIABLES

Regarding the impact of surroundings of a station on the usage of the station and system, Faghih-Imani et al. (2016) [16] looked at built environment variables and found that for example subway stations have a positive impact on a stations departure and arrival rates for subscribers. This is also the case for parks on weekends, and restaurants in the area increases rates as well for both user types. Furthermore, it was found that the job density of an area around a station has an impact on arrival and departure rates, in particular higher arrivals in the morning and higher departures in the evening. It confirms the use of the system for daily commutes to and from work.

Similar, Rixey (2013) [17] investigates in the effects of demographic and built environment

variables on the bike sharing ridership levels for different cities in the US. Built environmental variables considered were parks and colleges. Furthermore it was looked at transportation network variables, namely bus stops, bicycle paths and the number and distance of other bike sharing stations around. While it was found that colleges and bus stops do have a correlation with ridership, it is not the case for parks.

Lindsey et al. (2015) [18] investigates in the effects of businesses and jobs on station activity of a BSS in Minnesota by incorporating multiple built environment variables such as parks, water bodies and food related businesses as a economic activity variable. Additionally, nearby trails and other stations are used as transportation variables. It was found that stations nearby parks and water bodies serve recreational purpose, different to rides in downtown which users take in order to commute.

While the previously done research regarding built environment variables only looks at transportation variables (subway and bus stations, bicycle paths) and sometimes into variables related specifically to the local area such as restaurants and parks, this research classifies every station according to the land use around it and also accounts for details in the direct neighborhood of a station such as apartment and housing areas, transportation hubs, offices and commercial districts, tourist sights and recreational facilities such as museums, galleries and churches, for which it is also tried to differentiate between recreational POIs for locals and tourists.

3 | RESEARCH QUESTIONS

• **R**: How does the Citi Bike system serve common Points of Interest (POI) in New York City?

R1: What are the topmost POI-to-POI routes which the systems users cover by bike?

R2: Does a correlation between the type of POI and the likelihood that the system serves this type of POI exist? What could be the causes for the result?

R3: Does a correlation between user type and gender, date and time of day and the likelihood of the systems usage exist? What could be the causes for the result?

4 | method

4.1 OVERVIEW

On their website, Citi Bike provides the trip data from July 2013 up until March 2017 for download in the form of csv files [19] for every month. Each trip has the following entries available, except for customer trips, which do not have age or gender data recorded.

- Trip Duration
- Start Time
- Stop Time
- Start Station ID
- Start Station Name
- Start Station Latitude
- Start Station Longitude
- End Station ID

- End Station Name
- End Station Latitude
- End Station Longitude
- Bike ID
- User Type
- Birth Year
- Gender

From the Citi Bike system data, it was decided to built an analysis tool to support the research, increasing the insights into the system and possibly allow for optimizations to it. Except for the *bike id*, all of the data is relevant for this assignment, and considering the number of all trips combined is close to 40 million, optimizations to the structure of the data needed to be made, in order to build a responsive interactive map with the available resources.

The idea is to provide an application where it is possible to select a station and view the outbound or inbound trips for a defined date interval and other data relevant to the trips of a station, such as the distribution of gender, age, type of user and time of the trip.

In order to give a good visualization of the data and allow for investigation into the trip data, some key requirements for the application were determined, which had impact on the structure of the data. The basic requirements for the application were as follows:

- Clicking a station should display the trips from and to this station as a weighted graph visualizing the popularity of these trips.
- When a station is clicked, the user and trip data specific to the stations trips should be displayed.
- It should be possible to select a date range for the displayed data, allowing for investigations of different times of the year.

For the performance of the application, a different set of requirements was determined in order to work with the available tools:

• Keeping server and database load low: Only query for data which is needed at a point, and do not deliver unnecessary data.

- "Minify" the data set to save space and speed up delivery and processing, while keeping as much of the datas internal relationships as possible.
- Allowing for different *depths* of querying, i.e. only trip counts, then user data in a second step.

4.2 CLASSIFYING STATIONS

In total, five different types of points of interest were defined. Every station was then assigned up to three of them, depending on the area they are located in and the kind of POIs around them. The types defined are listed in Table 1.

Type Number	Type Name	Description	Map Icons
1	Transport	Stations nearby transportation hubs such as train stations or piers.	[20]
2	Commercial & Work	Stations surrounded by offices, shops or other commercial buildings.	[20]
3	Apartments & Housing	Stations nearby apartments and non-commercial areas.	[20]
4	Tourist Sights & Attractions	Stations nearby tourist sights and attractions.	[0] _[21]
5	Recreation	Stations nearby parks and museums.	• [20] or 111[20]
6	Service Stations	Stations used for service and maintenance.	× _[20]

The classifications were done by visual inspection of between 1 and 3 blocks around a station, with the help of Google Maps and Google Earth in some cases, the NYC Zoning and Land Use Map (ZoLa) [22] and the official NYC tourist guide [23]. ZoLa provides a detailed land use map of the city and can also be used to display offices and recreational areas, while Google Maps provides an overall visualization of the type of businesses or buildings around, and the tourist guide gives an overview over the most common sights. An example is shown in Figures 2 and 3.



Figure 2: ZoLa map for area around Congress Street & Clinton Street. [22]



Figure 3: Station 346 with types 3 (Apartments & Housing), 5 (Recreation) and 2 (Commercial & Work). [24]

Because New York City being a very densely populated area and visual inspection using maps can only only give a general overview of the surroundings of a station and is often subjective to the inspector, the types assigned to a station should be seen as equally weighted. One could argue that in some cases weighing them is rather simple even without having local knowledge of the area, but it was quickly noticed during the process that it is not simple to determine in more dense areas and would thus introduce inconsistencies across the classifications.

4.3 CHOOSING A DATABASE

Looking at the different options of databases, it mainly comes down to relational and nonrelational databases, both having advantages and disadvantages for different applications. Classic relational databases are normally referred to as *SQL* databases, while in the recent years, a trend in *NoSQL* databases can be noticed. While there is no general agreement of the definition of *NoSQL* (standing for "Not only SQL" or "Not relational") [25], there are several key differences between the two systems. NoSQL databases store data in a non-relational way and they can be scaled up horizontally, while relational databases are designed to be operated on one machine [26], meaning a decrease in performance of relational databases when the amount of data increases [27].



Figure 4: Comparison of relational and document database. [28]

Relational databases in general split up the data to different tables, as seen in Figure 4. For example, there might be a table for users, and another table for blog-posts of a user. Both of these tables first need to be defined in their structure (the schema), before data can be put in. This means introducing limitations to the rows and columns and the kind of data to be stored. Modifying this structure later on is more complex, since it needs to be made sure the structure is obeyed across tables.

Combining a user and a blog-post from above example requires a *join*, where the user-key of the user table is matched with the same key in the blog-posts table, after which the data is delivered. Looking at the given trip data set, it could be modeled as a table for the stations, and another table for trips.

NoSQL databases can be classified into different categories, such as Key-Value Store databases, Graph databases and Document Store databases. The latter stores all of the data in form of documents, which are schema less, meaning two documents do not have to obey the same schema and can be different in their structure. Likewise, updating them simply means adding or modifying an entry in a document. While schema-less was not a deciding factor, since the trip data does in general obey a structure, another factor is the high performance of such a NoSQL document database, as they put great emphasis on throughput of the data requested.

Since the trip data could be modeled in both the relational and non-relational way, other factors needed to be investigated in. Especially for this application, speed was the most deciding factor for choosing a database, considering the size of the data and the resources available while still being able to deliver quick responses from the database when querying. While *write* performance does not matter here - since the data is given and once it is stored, no new data needs to be written - the *read* performance is a crucial factor. Comparing different databases, it became noticeable that NoSQL databases in general deliver better performance than relational databases. In Yishan Li et. al. [29], a performance comparison between several databases was done and *MongoDB* and *CouchDB* came out to be the two fastest overall in read, write and delete operations. The differences between them are less important for this application, but MongoDB provides a more expressive query language while for CouchDB, often map/reduce functions need to be written, which are in general viewed as more complex.

Because of the speed advantage and the simpler options for querying the data, MongoDB was chosen for this application.

4.4 AGGREGATING & STRUCTURING THE DATA

After choosing MongoDB for storing the data, a document structure needed to be developed, allowing for quick access of the data for those different queries. For the database structure, the points defined in section 4.1 meant that the key to deliver the data quickly lies in indexing the stations, since every query is based on a specific station and MongoDB scans an index very fast in comparison to a collection. Therefore it was determined to use the station ID as an index.

For the weighted graph, it was decided to be more efficient to calculate the necessary values beforehand and thereby making it possible to deliver an almost instant weighted representation. These values are simply the number of trips, here called *trip counts*, and were easily determined by first aggregating all the trips leaving a station (and in a second step arriving at a station), and then counting a trip specific entry which is guaranteed to be non empty, for example the trip duration.

For the user data, it was important to not loose the relation to the specific trip and also not to each other, meaning that they should be stored with respect to the trip they originated from.

In order to aggregate this data, every file from the Citi Bike data set was imported into its own collection, resulting in a document for every trip, split over collections for every month of the given data. Having the data in MongoDB, it is possible to perform aggregations on it using MongoDBs *aggregation framework*.

```
1
2
3
4
5
6
7
8
9
10
```

{

```
"_id": ObjectID("59429f1e55e1832543c113c8"),
"dur": 594,
                                // Trip Duration in seconds
"sta": "2017-03-01 15:05:00",
                                // Start Time
"sto": "2017-03-01 15:10:00",
                                // Stop Time
"sid": 3441,
                                // Start Station ID
"eid": 360,
                                // End Station ID
"u": 1,
                                // User Type (1 = Subscriber, 2 = Customer)
"b": 1981,
                                // User Birth Year
"g": 2
                                // User Gender (1 = Male, 2 = Female)
```

```
10
11 }
```

Listing 1: Example document for a single trip of trip data after shortening field names

One of the constraints MongoDB put on the aggregation is its document size limit of 16 megabytes. Since the desired structure was to have a single document per station, which holds all of the trips for all months available from the trip data set, it means that all of the trips

combined of a station were not allowed to exceed this limit. This was a problem at first for some stations being very popular and having many trips, resulting in the need of shrinking the data set while not loosing any information.

A first optimization was to replace all mentions of *Subscriber* and *Customer* in the original files, which strings are encoded in UTF-8 with numbers, where 1 would correspond to Subscriber and 2 to Customer, thus saving 9 and 7 bytes per entry respectively.

Another major impact on the size of a document is the length of the field names. As seen in listing 1 the field names were shortened as much as possible while keeping them distinguishable. Less expressive field names with the benefit of shrinking the document size were more appropriate, especially since there are not too many field names to keep track of. Due to the way MongoDB stores field names, namely that even identical field names across arrays or documents are stored as an own entity, shortening them results in a linear decrease in document size. Simply by replacing *BirthYear* by *b*, *Gender* by *g* and doing the same for the other field names shrinked the document sizes to a level where all of the trips for a station could fit in a single document for that station, except for the start and stop times.

While attempting to store *all* trip related information from one station in a single document, it was noticed that the length of the strings of both the start and stop times was a major contributor to the data quickly exceeding the allowed document size. After trying many different formats, it was decided to convert them to BSON date objects and to have four collections in total, two for each the trip and time data, thus splitting up time data from the user data:

- 1. TripsOut: Holds all the trip counts and user data for outbound trips.
- 2. TripsIn: Holds all the trip counts and user data for inbound trips.
- 3. TimesOut: Holds all the date objects for outbound trips.
- 4. TimesIn: Holds all the date objects for inbound trips.

Aggregating Outbound and Inbound Trips

In Listing 2, the aggregation query which was used to aggregate all the outbound trips for a specific month from a station along with its user data is shown. It consists of a group stage at first, where the *_id* field determines which fields define a document and what gets aggregated together, in this case the definition lies in the relation between the start station (*sid*) and end station (*eid*). Everything after this definition is always in relation to these two stations. The count array is used to aggregate the number of trips, and details array holds all of the user data for the trips.

The second group stage determines more specifically how the document will look in the output. For the *_id* field, it the start station id is taken from the definition of the first group stage. Then, an array is created and the aggregated data - again, all specific to that start-end-station relation from the *_*id field - is inserted into that array. At last, an out stage defines the collection of the database where the result of the query should be saved to.

```
db.tripcollection201703.aggregate([
 1
 2
       {
 3
          "$group": {
                               // First Group Stage
 4
             "_id": {
                               // Defining the document id by
 5
               "sid": "$sid", // Start Station ID and
                "eid": "$eid" // End Station ID
 6
 7
            },
 8
             "count": {
                              // Array for the Durations
 9
                "$push": {
10
                   "dur": "$dur"
11
               }
12
            },
13
             "details": {
                              // Array for User Data
14
                "$push": {
15
                  "u": "$u", // User Type
16
                  "b": "$b", // Birth Year
17
                  "q": "$q"
                             // Gender
18
               }
19
            }
20
          }
21
      },
22
       {
          "$group": {
23
                                    // Second Group Stage
                                  // Assigning Start ID to be the Document ID for indexing
24
            "_id" "$_id.sid",
25
             "t1703": {
                                    // Array for all Trips of March 2017
26
                "$push": {
27
                   "eid" "$_id.eid",
                                              // End Station ID
28
                   "cnt": { "$size": "$count" }, // Trip count
29
                   "tps": "$details"
                                              // User Data array
30
               }
31
            }
32
         }
33
      }
34
   ],{
35
      "$out": "t1703" // Out Stage defining the output collection
36
     }
37
   )
```

Listing 2: MongoDB aggregation framework query for aggregating trips from a station.

For the aggregation of the inbound trips to a station, the same query can be used while simply changing "\$_id.sid" in the second group stage to "\$_id.eid", and "eid": "\$_id.eid" with "sid": "\$_id.sid" for consistency reasons.

The result of described query can be seen in Listing 3.

```
1
    {
 2
       "_id": 3441,
                                   // Start Station ID
 3
       "t1703": [
                                  // Array for all Trips of March 2017
 4
          {
                                   // from Station ID 3441
             "eid": 360,
 5
                                   // End Station ID
 6
             "cnt": 2,
                                   // Count of Trips
 7
             "tps": [
                                  // Array for Trips to Station ID 360
 8
                {
 9
                    "u": 1,
                                // Type of User
10
                   "b": 1981, // User Birth Year
11
                    "g": 2
                                 // User Gender
12
                },
13
                {
                   "u": 1,
14
15
                   "b": 1981,
                   "g": 2
16
17
                }
18
             ]
19
          }
20
    }
```

Listing 3: Example entry of the aggregated data structure used in the database.

Aggregating Duration and Start/Stop Times

While looking into alternative ways of storing the trip dates and times, MongoDBs BSON *Date Object*, being a 64 bit integer representation of a date in milliseconds since the UNIX epoch [30], turned out as a good way of not only storing dates in a smaller format to save space and speed up processing, but also allowing for more functionality of the application in combination with the library *moment.js*, which can return the day for a given date and thus enabling filtering the time data by day of the week. Converting all of the dates turned out to be a resource and time intensive task taking several hours to convert the close to 80 million date strings in the database to BSON date objects.

For the creation of the date objects, a simple query was used to generate date objects from the given strings. Unfortunately, the Citi Bike data set stores these dates in a very unconventional way from a programming perspective, and has furthermore differences in such across the monthly files, requiring several modifications to meet the input requirements of a javascript date object, after which the collections *TimesIn* and *TimesOut* were created on a similar way like the two trip collections. The query for this can be found in Listing 4.

```
db.tripcollection201703.aggregate([
 1
 2
       {
 3
           "$group": {
                                       // Grouping by
             "_id": { "sid": "$sid", // Start Station ID and
"eid": "$eid" // End Station ID
 4
 5
 6
             },
 7
              "details": {
 8
                 "$push": {
 9
                     "d": "$dur", // Trip Duration
10
                    "b": "$sta", // Start Time
                    "e": "$sto" // Stop Time
11
12
                 }
13
             }
14
          }
15
       },
16
       {
17
           "$group": {
              "_id": "$_id.sid",
18
              "t1703": {
19
20
                 "$push": {
                    "eid": "$_id.eid",
21
                    "tps": "$details"
22
23
                 }
24
              }
25
           }
       }, {"$out": "d1703"}
26
27
    ]
```

Listing 4: Query for aggregating times and duration.

5 | RESULTS

In order to determine the *top* routes within the system, it first was decided to define three plateaus which can be found in Table 2. They were defined based on the number of stations available in the system, plotted in Figure 5, because keeping the number of stations constant and thus eliminating fluctuations in start and destination possibilities will lead to a more meaningful result when determining the top most trips. For this, a graph was generated visualizing the number of stations in the city over the time Citi Bike operates in the city.

Station Count

The number of stations over the whole timespan.



Figure 5: Number of stations over the years.

Plateau Number	Number of Stations	Time Span	Total Rides	Rides in the Top 50
1	≈ 332	07/04/2013 - 07/29/2015	17.584.004	295.006
2	≈ 510	11/16/2015 - 05/25/2016	6.007.078	91.062
3	pprox 690	09/30/2016 - 02/27/2017	13.500.380	85.538

Table 2: Defined plateaus.

For each of these plateaus, the 50 top routes were aggregated. The result can be found in Tables 4, 5 and 6 in the appendices.

Weighted maps were created from the aggregations, which can be seen in Figure 7.

An example of a map with weighted edges can be found in Figure 6. The icons represent a station, corresponding to the POI types defined in Table 1. The stroke weight of the path between two stations is a result of linearly mapping the trip count from that route into a range between 2 and 15 pixel based on the minimum and maximum trip count of the trip set which is currently viewed (e.g. plateau 3), which turned out to be appropriate to still see the smallest weights while not overlapping too much space by the greatest weights. The number next to the station represents the station ID, corresponding to the 'Start Station ID' and 'End Station ID'

in Tables 4, 5 and 6. For crowded areas with many routes where detailed investigations were done, additionally trip numbers were defined, like *3.1* in this example, where 3 corresponds to the plateau and the second number simply helps with numbering them throughout the map. All map images were created with the Google Maps API and as such belong to Google Maps.



Figure 6: Example of a map with weighted edges, station numbers and trip number [31].



Figure 7: Weighted graph of top 50 routes of plateau 1, 2 and 3 from left to right respectively. [32]

5.1 CENTRAL PARK



R1

In all three plateaus, the path with the greatest trip count and thus the most popular ride by count is from station 2006 to itself, which can be seen right at the bottom end of Central Park in Figure 8. It represents the most popular station-to-station ride whole Citi Bike system. Further trips in plateau 1 from the top 50 in the area are from the bottom right end of Central Park to Columbus Circle, and again to stations itself.

In plateau 2 and 3, Citi Bike expanded more into the north, having a significant impact on the top rides which now reach to the Metropolitan Museum of Art and Pilgrim Hill on the east side of the park. On the west side, both stations 3165 and 3168 are nearby subway stations. Plateau 3 features even more destinations alongside the park, all of them serving nearby sights. Furthermore, rides in the upper west side and upper east side are now also among the top 50, which are discussed in section 5.1.1. The ride between station 2006 and 3374 with 3.3 mile distance is furthermore the longest in the top 50 for all plateaus. **R2**

The mentioned routes are listed on the Citi Bike website under the popular rides section, indicating the popularity especially among tourists which want to discover Central Park by bike [35]. In plateau 1, the stations from Figure 8 were the most north stations in the whole system; additional stations were only added later in plateau 2 and 3, explaining the difference in the top rides in Figures 9 and 10 with additional stations in the north of central park. While in plateau 1 most riders seem to have taken a trip through the park and back to station 2006 and stations near Columbus Circle, in plateau 2 and 3 more riders seem to use the system to get to the museums, galleries and recreational spots on the east and west side of the park, such as the Pilgrim Hill, the Metropolitan Museum of Art, the Guggenheim and the Jewish Museum. Before the station expansion during the end of 2015, it was not possible to visit the museums using Citi Bike due to the lack of stations and the 45 minutes limit. Users were bound to have a trip around Central Park and then return to a station they originated from, while now trips in the uptown direction are among the most popular.

For Central Park, an area with little to no motorized traffic, it can be concluded that it is very likely that people make use of the Citi Bike system to reach nearby sights.

R3

Station 2006 is seen as a representative for all trips reaching stations on the borders of Central Park, as it acts as main hub for them. The route from station 2006 to itself has the significantly high customer to subscriber ratio, with 80% of trips on this route coming from customers and 20% from subscribers in plateau 1, 83% and 17% for plateau 2 and 81% and 19% for plateau 3. Over the whole time span and including all destinations, 66% of all the outbound trips from this station are from customers, and 34% from subscribers. Looking at inbound trips of station 2006, it is nearly the same with 65% customers and 35% subscribers. These results strengthen Central Parks stations and rides as very popular among tourists.

Regarding the gender distribution on this route, 65% of the users gender is not available, corresponding to the customer share, as these do not get collected for this type of user. For the available 35% of subscribers for which the gender is available, 26% is male and 9% is female, which is however unlikely to be representative for the whole group, as it is expected that the female share would be higher if the genders of customers would be known, as suggested by additional results discussed in this thesis.

From the time data specific to this trip plotted in Figure 11, it can be noted how there is an almost equal amount of popularity to start a trip from station 2006 between 11 am and 4 pm, and an even higher popularity in the evening rush hour around 5 - 6 pm, representing a station which is also busy during the middle of the day. It is likely that the evening rush hour is a result of subscribers taking a bike home, unlike tourists which are hardly bound to times and thus make use of the system almost over the whole day.



Figure 11: Distribution of trip start times for all outbound trips of station 2006.

Since most hotels in Manhattan are in midtown and thus below station 2006, it is reasonable to assume that most tourists start their tour at the south end of Central Park where station 2006 is, an area which is very busy and thus can be seen as a sight itself. From this and looking at the destinations of the routes and previous results, it can be further hypothesized that it is even more likely for tourists to use the system when two stations are connecting two sights, which is the case in all three plateaus.

From this, it can be seen how tourist are very likely to discover Central Park by bike and the additional station availability over the years lead to an increase in tourists also using Citi Bike to reach sights around the park. It seems very likely that this usage pattern will further increase in the coming years as Citi Bike expands more.

5.1.1 Upper East Side, Yorkville & Upper West Side

R1

In plateau 3 on Figure 10, other popular trips around the area include a ride on East 85th street from an apartment dominated area to a nearby subway station (3150 to 3147). A little below that two more similar rides were found, one of which starts in a business area (3142) and ends at St. Catherine's Park (3141), the other starting at the Park and ending at the station near Lexington Av/63 St subway station (3155). On the upper west side next to Central Park, a similar short ride between station 3167 and 3171 is among the top 50, showing a similar pattern to the previous ones.

R2

All of these rides follow a similar pattern in their types of POI: They all are connecting a housing area and a transportation hub, and a park in one case. Since both the upper west and upper east side feature much fewer sights by comparison, locals are the dominating users on them and they use Citi Bike for their last mile rides between apartments and subway stations, most likely to get to work.

R3

All of the rides feature a subscriber share greater than 99%, a male user share between 68.5% and 75% and a female user share between 23.4% and 29.2%. In addition they all show a typical morning and afternoon rush hour, except for the ride between stations 3142 and 3141 which has a small uptick around midday during weekdays. From this data, it can be deducted that the users on these rides represent the subscribers to the system who are local to the stations, suggesting a high likelihood of the systems usage for last mile rides between an apartment and a transportation hub among locals.

5.2 MIDTOWN MANHATTAN

5.2.1 Grand Central Station



Figure 12: Popular rides in midtown in plateau 1, 2 and 3 from left to right respectively. [36]

R1

Looking at midtown Manhattan and especially the area around Grand Central Station, simi-

larities between the plateaus can be noticed as well. In plateau 1, the popular rides start at Grand Central Station (519) and end near the Port Authority Bus Terminal on 42nd street (477), which is close to Times Square, and at Pennsylvania Station (492). There is thus two common destinations with a start at Grand Central Station. This is continued in plateau 2 and 3, here additionally with also the direction back from station 477 to 519 being in the top 50, and further station 498 near the Empire State Building and station 472 in a general business area being destinations.

R2

A clear pattern can be noticed regarding the type of POIs in this area. All trips originate at Grand Central Station and have either another transportation hub or areas of business as destinations, indicating that these rides serve for the purpose of switching the transport method for workers who arrive by train in the city or directly lead to a work place. However, Grand Central Station and the Empire State building are just two among the high density of very common sights in the area. It is thus further investigated if there is a significant number of tourists using these rides.

R3

By looking at Figure 13, the times when trips are started from station 519, showing a typical morning and afternoon rush hour in all three plateaus, and the user type distribution being all between 95% and 98% of subscriber share on these routes strengthen the rides classification as dominated by workers which need to switch transport or need to reach a business related POI in the area. As the popularity of the rides and the number of destinations from station 519 increased over the years, it is reasonable to assume that Citi Bike does increase in popularity among this group of people as a method of last mile transportation.

Since there is no significant customer share on any of those trips, even though many sights are around the area, it can be assumed that the extremely busy traffic in midtown has negative impacts on the ridership of tourists, and that this group does not use the system to reach destinations like it was noted in section 5.1. Similar the share of female users is very low, in between 1% and 4% on most rides, in comparison to the male share with 86 - 96%, except for one ride. In both plateau 2 and 3, the ride between station 519 and 498 features a 12% female ridership, thus significantly higher than the surrounding trips. While this trip was not in the top 50 in plateau 1, it had a female share of 4.6% during the time span of plateau 1, comparable to the surrounding trips in Figure 12, and has since risen to 12%. It is however difficult to determine an underlying cause for this, since it could be due to local changes like improvements of cycle paths, or changes in the business areas nearby.





Figure 13: Start time and user type distribution of the station "Pershing Square N" for all outbound trips of the whole time span.

5.2.2 Chelsea



Figure 14: Popular rides in Chelsea in plateau 1, 2 and 3 from left to right respectively. [37]

R1

Going further towards downtown Manhattan, rides between business areas and apartments become more noticeable. All three plateaus have a ride between station 435 at W 21 St & 6 Ave and station 509 at 9 Ave & W 22 St among the most popular ones (1.3, 2.2, 3.1). In addition, in every plateau a ride on the west side crossing through Chelsea can be seen (1.8, 2.7, 3.4).

R2 & R3

Stations in Chelsea feature a mixture of POI types, ranging from transportation hubs (station 521 in plateau 1) to offices and business areas further downtown and some housing areas as well. The Chelsea neighborhood overall is filled with many galleries, restaurants and is known for its cultural aspects [38].

Station 435, which is present in all three plateaus and is surrounded by almost exclusively businesses and offices, faces greater popularity in the afternoon in comparison to the morning, indicating usage of the station mainly for leaving the business districts around 6th avenue after the end of a work day, which is visualized in Figure 15.



Figure 15: Distribution of start times of all trips from station 435.

Furthermore, the subscriber share of these rides (1.1 - 1.4, 2.1 - 2.4 and 3.1) in all plateaus is well above 95%, showing that they are almost exclusively used by locals. A significant difference can be noticed in the gender of users on these rides, being between 21.2% and 27.8% female, depending on the plateau, while the earlier discussed popular stations around Grand Central Terminal are almost exclusively used by men.

From this, it can be hypothesized that women are more likely to use the system for rides be-

tween businesses and housing areas in comparison to rides serving for the purpose of switching methods of transport. Another reason could be that the area is considered to be safer for cyclists among female users, which is however unlikely due to the density and location of it.

Rides 1.8, 2.7 and 3.4 are the same rides and are as such popular in all plateaus. Additionally, ride 1.8 and 3.4 are among the top 50 with both directions. This ride between station 514 and 426 follows the West Side Highway, being next to the Hudson River. Along the highway, multiple sights and attractions for tourists can be found, such as sightseeing cruises at station 514, parks and sports fields. This ride connects midtown to Battery Park City in lower Manhattan, which is discussed in more detail in section 5.3.2. Both end and start of this ride are sight dominated areas and with the dedicated bicycle path next to the highway, it offers a connecting ride between these two areas with a waterfront, parks and sights next to it. With the customer share being 45%, 59% and 52%, it is clearly shown that it is a popular ride among tourists. From the subscriber share, between 18% and 26% is female, which is a greater share than most rides in Manhattan. Since this ride follows a bicycle path separated from the busy highway next to it, it can be concluded that this path does have positive impact on tourist and female ridership, making the ride more safe. It is also the second longest ride among the top 50 by distance with roughly 3 miles.

Ride 1.5, being a ride from a business area to Pennsylvania Station, and ride 1.6, being a ride of similar length and type in the opposite direction, can offer more insights by comparing them, as seen in Table 3.

Rides	1.5	1.6
Description	Business District to Train Station	Train Station to Business District
Total Trips	8230	4826
Female Users	15%	9.22%
Male Users	71.4%	83.7%
Subscriber Share	86.9%	81.8%
Customer Share	13.1%	18.2%

Table 3: Comparison of ride 1.5 and 1.6 from Figure 14.

The higher share of female users on rides 1.5 could indicate that female users are a little more likely to use a Citi Bike to get from work to a transport hub in comparison to taking a bike in the morning when arriving in the city to get to work, strengthening the hypothesis made earlier. The cause could be for example the exhaustion some of those trips could have as a result, which is likely not preferred when arriving at work, but less of a problem when leaving work. Investigating further into these two rides by looking at the whole data set, it can be noticed that the male share being 71.4% and 83.7% does change, however this can be explained with the customer share where the gender is unknown, namely 18% for ride 1.6 in comparison to 13% for 1.5. While this shows that more users in this area general prefer to use the bike in the direction towards Pennsylvania Station, it can be deducted with the help from start time data from Figure 16 that at least some share of the female users prefer to only use the bike in the afternoon, due to the mentioned reasons, and not during the morning rush hour, which is much stronger in comparison to the afternoon.



Figure 16: Distribution of start times of all trips from station 521 at Pennsylvania train station.

It plateau 2 and 3, more very similar rides can be seen, for example rides 2.6 and 3.2, being the same. Comparing the shares of the users genders, an increase from 25.2% female users in plateau 2 to 40.9% in plateau 3 can be noticed. Moreover, the share of male users dropped from 72.4% to 57.1% in the same time span.

Another study by Kaufman, S. suggests that there is a strong relationship between the safety of riding a bike next to the car traffic and women taking a bike, meaning that women are more likely to choose a Citi Bike in areas with fewer traffic lanes [15]. Road works and construction will thus have an impact on the female user share, however this leaves the decline in the male user share unexplained. It is therefore more likely to be either a temporary shift in the users behavior due to changes in the local area, like events, openings of related businesses and offices. The fact that there were no rides in April, might also have something to do with this change.

5.3 LOWER MANHATTAN

5.3.1 Greenwich Village & East Village



Figure 17: Popular rides in Greenwich & East Village in plateau 1, 2 and 3 from left to right respectively.
[39]

R1

Going further downtown, a different pattern in the popular rides can be seen in Greenwich Village & East Village(Figure 17), for example rides connecting two or more recreational points

of interest. In plateau 1, the rides are connecting several parks and station 151 in SoHo, a district housing many restaurants, art galleries and shops. Similar trends are found in plateau 2 and 3, especially rides 1.2, 1.3, 2.2, 2.3 and 3.2. Important to notice is that for rides 1.2, 1.3, 3.2 and also 3.3, both directions are in the popular rides of these plateaus.

R2

Significant in all three plateaus is the number of recreational POIs, which in addition, are connected to stations of the same POI or nearby housing and business areas. As seen in section 5.1, rides between two or more recreational POIs seem to be common in areas where there is a great density of such. The fact that in plateau 3, rides 3.3 and 3.2 are popular in both directions does suggest a relation between areas of work and business and recreational areas. As for nearby subway stations, the closest one is Astor Place Station, being right next to station 293 and slightly above station 3263. From this, it can be hypothesized that locals arrive by subway and use Citi Bike for the last mile to go to Tompkins Square Park and the area around it, and the same ride is also used for the way back. To further investigate if these rides serve the purpose of reaching a recreational POI, the start times of all inbound trips to station 445 were examined and are plotted in Figure 18.



Figure 18: Distribution of start times of all trips having station 445 as destination.

The plot shows a nice example of increasing popularity towards the end of the day, suggesting that station 445 becomes more popular as more people reach the end of their work day, indicating that station 445 is indeed popular as an 'after work' destination.

R3

The user gender share for these rides are very similar across plateaus, being roughly between 20% and 25% female, and between 70% and 75% male. The user type share shows that these rides are mostly used by locals, with the subscriber share being above 90% for all of them. The fact that the rides have a very similar popularity in both directions supports the hypothesis that locals use these to go from station 293 to Tompkins Square Park and back. It seems thus very likely that recreational POIs are a popular last mile destination for locals using the Citi Bike system.

5.3.2 Financial District, Tribeca & Two Bridges



Figure 19: Popular rides around the Financial District, Tribeca & Two Bridges in plateau 1, 2 and 3 from left to right respectively. [40]

R1

Popular rides around the south end of Manhattan include rides around Battery Park City (rides 1.4 - 1.6, 2.2 - 2.4 and 3.2 - 3.3) and a ride from station 387 to itself in all plateaus. They are visualized in Figure 19. Furthermore, plateau 2 and 3 both have a ride between stations 502 and 307 among their popular ones, both being around the neighborhood Two Bridges. Station 387 is also start and destination for rides to and from Brooklyn in plateau 1 (rides 1.1 and 1.2). **R2 & R3**

Investigating the area of station 387, popular tourist sights can be found, like the World Trade Center Memorial and the One World Trade Center, Wall Street, Trinity Church, Brooklyn Bridge and Battery Park City, which is also frequented often by ride 1.3 in plateau 1. In addition being right at City Hall Park, which is a common spot for workers in the area of the financial district to spend their breaks [41], the area around it is also among the most densely populated in the city during the day [42]. Rides 1.1 and 1.2 being across the Brooklyn Bridge clearly underline that the system is a likely choice for trips across the Brooklyn Bridge, which does offer great views of the city and certainly has a recreational aspect. Battery Park City features a mixture of POIs: Many office buildings, tourist sights and shops are dominating the area.

Exploring user type shares and trip durations in Figure 20, it can be seen that station 387 is both used by tourists and regular subscribers equally. Tourists will most likely use the station to get to Brooklyn via the Brooklyn Bridge and to the mentioned sights nearby. Citi Bike subscribers working in the area are likely to use the system to grab lunch at a nearby shop and return to City Hall Park for their break, which can be seen from the relatively great percentage of trips started around midday in Figure 20. In addition, the ride durations show next to the typical peak around 4 minutes (which are most likely the subscribers trips), also a great share of trips in the 10 - 25 minute range.





Comparing rides 2.1 and 3.1 for their respective user gender shares, it can be noted that while there is a 62.3% male and 36.1% female share for the first plateau, with 98.4% subscribers on this ride, there is a shift in plateau 3 to 54.9% male and 44.7% female users, while the subscriber share remains high at 99.6%, again underlining a general trend of an increasing popularity of the system among women.

As for popular tourist rides, in addition to in section 5.1 discussed station 2006, rides 1.1 and 1.2 also feature a very high customer share, both being around 90%. These rides, connecting City Hall Park with Cadman Plaza Park via the Brooklyn Bridge are popular among tourists, since the Bridge is itself a very popular sight and taking a bike instead of walking reduces the time of this three mile round trip from roughly an hour of walking to 20 minutes of cycling (assuming no stop is taken).



Figure 21: Distribution of ride durations for station 232.

In contrary to most rides in the city featuring a great share of subscribers in comparison to customers, trip durations are on average much longer for customers. Figure 21 emphasizes this, showing ride durations for a customer dominated station which are still very significant in the 15 - 25 minute range, while most subscriber dominated stations have a peak around 4 - 7 minutes and fall of quickly after that.

As for rides in Battery Park City, steady popularity can be seen across the plateaus. While in plateau 1 and 2, the customer share on rides 1.4 - 1.6 and 2.3 - 2.4 was smaller than 10%, rides 3.2 and 3.3 have a respective 24.8% and 38.7% customer share, indicating that rides around Battery Park City become more popular among tourists. Since rides 1.4 - 1.6 and 2.3 - 2.4 almost exclusively have subscribers using them - in both directions - they could serve the same purpose as station 387, namely to get lunch and spent the break in the park, and then return to the office. But in addition, station 327 has a strong rush hour in the morning and evening, suggesting a typical last mile type of station for people around the area using the Citi Bike system to get to work. However since the closest subway station is near station 387, it is most likely that those people are either living in Battery Park City or arrive by boat at the WFC ferry terminal near station 327. Rides 2.2 and 3.2 are more customer dominated, with 20% and 25% customer share respectively. Both are the same ride, starting at the World Financial Center featuring many shops and ending at Battery Park at the south end of Manhattan.

5.4 BROOKLYN & QUEENS



Figure 22: Popular rides in Brooklyn & Queens in plateau 1, 2 and 3. [43][44]

R1

While plateau 1 had only around 80 stations in Brooklyn, during the second and third expansion Citi Bike installed many more in both Brooklyn and Queens, which can be seen in the respective top rides for each plateau as well (Figure 22), as popular rides tend to be further south and east as Citi Bike expands in those directions. The first plateaus only top trip is from station 398 to itself. Plateau 2 and 3 have due to the expansions more top trips, both in Long Island City and throughout Brooklyn.

R2 & R3

For plateau 1, station 398 serves as the base for rides through Brooklyn Bridge Park and and the Brooklyn Heights Promenade, offering great views of Manhattans skyline, thus being an interesting ride for tourists, also being featured on Citi Bikes website under the popular rides section [45].

For plateau 2, ride 2.1 includes both directions between station 3119 at Long Island City train station to station 3124, being surrounded by mostly industrial and manufacturing facilities is another example of a last mile ride from a transportation hub to work and back. Having 27.3% female and 71.2% male user gender share with a 98.9% subscriber share, ride 2.1 underlines the previous hypothesis of a subscriber dominated last mile ride. It serves the purpose of getting to work and back as an alternative to a cab or walking, as there is no subway station covering this small route and younger people prefer to use Citi Bike to a bus, with the advantage of not being bound to the buses times and often reaching the destination quicker with the bike.

Looking closer at ride 2.2 and 2.3, it turns out that they feature almost the same user properties as ride 2.1, and in addition also go in both directions and have one of the stations nearby a transportation hub, namely Graham Avenue station nearby station 3086 and Bedford Avenue station at station 3093.

Rides between station 460 and 3093, and 2002 and 3093 are of the same type as 2.2 and 2.3, where station 3093 is at Bedford Avenue subway station and 2002 and 460 are surrounded by houses and apartments. This pattern continues with ride 3.3, notably having almost a 40% female user share, this time with only one direction being in the top 50 with 1,442 trips taken, while the opposite direction is less popular with 325 trips in the same time span. Station 258 has a strong rush hour in the morning, while station 324 is stronger in the evening but less significant overall, further underlining this asymmetric relationship between these two stations. Looking at a topographic map of the area, it can be seen that station 258 elevation is roughly 82ft, while station 3093 has an elevation of only 26ft, explaining the preferred downhill direction of the route.

It can be concluded that all discussed popular rides in Brooklyn and Queens in plateau 2 and 3 are of the described last mile type and follow the same pattern. They represent a *local* usage pattern of the Citi Bike System, being more typical for people living in the area and working outside (e.g. in Manhattan). The cause is most likely the absence of tourist dominated stations and rides, since there are much fewer sights around the area in comparison to Manhattan. However this is supported by rides in the Upper East Side in Manhattan (see section 5.1.1), and these type of rides seem to be the most common in areas where there is fewer businesses and sights. Thus the Citi Bike system seems to be a popular choice for shorter trips in order to reach the nearest train or subway station, or go the other way around.

6 | CONCLUSION

The Citi Bike system data was mined and the most popular rides of the Citi Bike system in New York City on three defined plateaus were aggregated using MongoDB and its aggregation framework. From it, an application (www.nycmap.bike) was built to help visualizing these rides as a weighted graph and in addition, offering insights into user and trip data for clickable stations and time ranges. The popular rides within the plateaus were discussed and from them derived how the system is serving different types of points of interest and whether a correlation between these POIs, user data, trip time and the popularity of the system exists.

It was found that a significant number of the rides serve for the purpose of reaching a business related destination starting at a transportation hub, so called *last mile* rides. Especially in Brooklyn & Queens, this type of ride is very popular. If available, workers use the system during their work day to get to a nearby park and spent the break there. While most rides are dominated by male ridership, over the years an overall increase in female ridership within the system was noticed, indicating that Citi Bikes efforts to increase it do have positive impact, or that additionally the safety for bicyclers increased over time. Popular rides among tourists are in and around Central Park, often having nearby sights such as museums as destination. A ride over the Brooklyn Bridge and a ride from midtown Manhattan along the west side highway to Battery Park City was also found to be very popular. It was found that all tourist dominated rides have lower car traffic next to the assumed routes in common, or there are designated bicycle paths available, which is also the case for rides with an increased in female ridership.

From the results it can be said that expansion of the BSS in neighborhoods with low expected tourism should put emphasis on connecting housing areas with transportation hubs. Tourists are more likely to use the system for recreational purposes, and the system should thus provide stations enabling such rides in areas of little to no motorized traffic. Increasing the local share of female ridership needs to be done through increasing the safety of riding a bicycle on the routes that locals usually cover, often with busy traffic. This could be done by further separation of the bicycle path and the motorized traffic, e.g. by moving parking spots between them, where it is then clearly harder for a car to interfere with cyclists.

The results can further be helpful for business owners nearby Citi Bike stations. They could easily investigate, from where cyclists arrive at a station, the user data and rush hours of a station, and could then adjust their street advertising according to it.

For Citi Bike users themselves, extrapolating the results can give an estimation of how busy a station is during a certain time, and it could help with planning a ride.

Overall, a significant number of the results are expected and their cause can be identified rather easily by looking at built environment variables. However, some of the changes in user gender are harder to investigate in, either due to missing knowledge of the local area but also due to a too small sample size. This also represents a limitation to this method, since New York City is one of the most densely populated areas in the US, it is expected that changes in rides with only a few hundred trips taken over several weeks can be hard to explain, which was definitely the case for rides here.

Significant to see but often expected is the difference in usage patterns for subscribers and customers, ultimately representing tourists. Their rides can easily be identified but there are only very limited stations and routes which reach a tourist share as high as 40%, most are between 1% and 5%. It is very likely that tourists do not see riding a bike as a safe method of transportation, especially in Manhattan, which was also the case in personal experiences. New York City is known for its extremely busy traffic and being used to average car traffic from different countries, tourists are most likely unwilling to ride a bike next to the traffic. It should be noted however that so far, there has been a single deadly Citi Bike related accident [46], while the data set contains more than 39 Million rides. Rides dominated by tourists clearly serve the purpose of discovering the city on routes with as little car traffic as possible and have a recreational use. It suggests that tourists do use other means of transport to get to destinations in Manhattan, and use the BSS only as a free time activity. It should be possible to have an increased number of tourists also taking rides next to more motorized traffic, which could be done by increasing safety for cyclists, which will also increase the female share.

Other limitations were found in the classification of the stations. While 5 different types were defined and every station was assigned up to 3 of them, it was found that while trying to visualize this *POI to POI* relation from a station to all of its destinations by dividing up the trip-count from a station by the number of POI types of the destination stations, the distribution hardly changes. This means that the categories defined are most likely too general and thus stations were assigned too similar types too often, or the types should be weighted differently. A good example are stations nearby subway stations, which in Manhattan often do not seem to have a huge effect on ridership, while they are a driving force behind the ridership of stations in Brook-lyn and Queens. For further research into the POI-to-POI relationships, stations would need to be classified further, either with more types or weights, a task hard to accomplish without detailed knowledge of the area. If stations could be classified more detailed, it should be possible to define a more detailed relation between types of POI, which could as such be helpful during system expansion planning.

In Gordon-Koven et al. (2014) [13], stations such as Pershing Square at Grand Central Station were seen as one station, while they are actually two stations in the data set (IDs 517 and 519). This was not accounted for in this research, as it discusses the top rides between two in the data set defined stations, and the aggregation did not take this example into account. Therefore, both stations are seen as own entities and thus only one of them shows up in the

top aggregated trips; however this should be negligible since one station can be seen as as a good representation of a neighboring station.

While it was found that the user gender share on the top rides in Manhattan is predominantly male and for rides in Brooklyn it can reach close to a 40% female share, again suggesting that many routes in Manhattan are seen as less safe, an overall increase in female users over the years was found, which could be the effect of increasingly safer cycle paths but also Citi Bike actively trying to promote the system among females with their *#WomenWhoBike* events, where free day passes to the system are given out to women [47]. This agrees with related work such as Kaufman, S. (2014) [15] and also shows results of Citi Bike efforts to increase popularity of the system among women.

BIBLIOGRAPHY

- [1] Alan Levine. The city with citi bikes. Flickr, Jun 2013. https://www.flickr.com/photos/ cogdog/8912565722/in/photolist-ezzehf-eDjtjM-eDqzTA-eDjsQt-eDqBMb-eDk6tX/, visited 2017-07-10.
- [2] Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, 12(4):3, 2009.
- [3] C Black and S Potter. Bikeabout in portsmouth: Ein automatisiertes fahrradverleihsystem mit smart-card (german) [bikeabout in portsmouth: An automatic bike-lending-system with smart-card]. 1999.
- [4] Amy O'Leary. Everybody inhale, how many people can manhattan hold? *New York Times*, 1, 2012.
- [5] New York City Department of City Planning. Sustainable streets strategic plan for the new york city department of transportation 2008 and beyond. Welcome to NYC.gov / City of New York, Apr 2008. http://www.nyc.gov/html/dot/downloads/pdf/stratplan_ compplan.pdf, visited 2017-07-03.
- [6] Amada M Burden, Richard Barth, et al. Bike-share opportunities in new york city. *New York: Department of City Planning*, 2009.
- [7] Christine Haughney. New york chooses company to run bike-share program. *The New York Times*, 2011.
- [8] Sarah M Kaufman, Lily Gordon-Koven, Nolan Levenson, and Mitchell L Moss. Citi bike: The first two years. 2015.
- [9] Inc. Motivate International. About citi bike: Company, history, motivate. *Citi Bike NYC*. https://www.citibikenyc.com/about, visited 2017-07-03.
- [10] New York City Department of Transportation. Citi bike expansion: Updated final plan. City of New York. http://a841-tfpweb.nyc.gov/bikeshare/files/2016/07/ Manhattan-CB-8-Updated-Final-Plan-Presentation.pdf, visited 2017-07-03.
- [11] Motivate International Inc. Citi bike | bike angels. *Citi Bike | Bike Angels*. http: //bikeangels.citibikenyc.com/.
- [12] Divya Singhvi, Somya Singhvi, Peter I Frazier, Shane G Henderson, Eoin O'Mahony, David B Shmoys, and Dawn B Woodard. Predicting bike usage for new york city's bike sharing system. 2015.
- [13] Lily Gordon-Koven and Nolan Levenson. Citi bike takes new york. *Rudin Center for Transportation Management and Policy*, 2014.
- [14] Jan Garrard, Geoffrey Rose, and Sing Kai Lo. Promoting transportation cycling for women: the role of bicycle infrastructure. *Preventive medicine*, 46(1):55–59, 2008.
- [15] Sarah Kaufman. Citi bike and gender. NYU Wagner, 2014. https://wagner.nyu.edu/ rudincenter/2014/05/citi-bike-and-gender, visited 2017-07-20.

- [16] Ahmadreza Faghih-Imani and Naveen Eluru. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of new york citibike system. *Journal of Transport Geography*, 54:218–227, 2016.
- [17] R Rixey. Station-level forecasting of bikesharing ridership: Station network effects in three us systems. *Transportation Research Record: Journal of the Transportation Research Board*, (2387):46–55, 2013.
- [18] Xize Wang, Greg Lindsey, Jessica E Schoner, and Andrew Harrison. Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, 142(1):04015001, 2015.
- [19] Motivate International Inc. Citi bike system data. *Citi Bike NYC*. https://www.citibikenyc.com/system-data, visited 2017-07-01.
- [20] Freepik. Office, Apartments, Tree, Repairing & Subway Icon designed by Freepik from *Flaticon*. Freepik. https://www.flaticon.com/, visited 2017-07-01.
- [21] Madebyoliver. Photo Camera designed by Madebyoliver. https://www.flaticon.com/ authors/madebyoliver, visited 2017-07-01.
- [22] The City of New York. Nyc dcp zoning and land use (zola). NYC.gov. http://maps.nyc. gov/doitt/nycitymap/template?applicationName=ZOLA, visited 2017-07-15.
- [23] Inc NYC & Company. The official new york city guide / nycgo.com. NYCgo.com. https://www.nycgo.com/assets/files/pdf/nyc-official-visitors-map-2017.pdf, visited 2017-07-02.
- [24] Google Maps. Google maps. Google Maps, 2017. https://www.google.de/maps/@40. 7367219, -74.0058767, 17.62z, visited 2017-07-12.
- [25] Rick Cattell. Scalable sql and nosql data stores. Acm Sigmod Record, 39(4):12–27, 2011.
- [26] Vatika Sharma and Meenu Dave. Sql and nosql databases. International Journal of Advanced Research in Computer Science and Software Engineering, 2(8), 2012.
- [27] Ameya Nayak, Anil Poriya, and Dikshay Poojary. Type of nosql databases and its comparison with relational databases. *International Journal of Applied Information Systems*, 5(4):16–19, 2013.
- [28] Couchbase Inc. Comparing document-oriented and relational data. Comparing documentoriented and relational data. https://developer.couchbase.com/documentation/ server/3.x/developer/dev-guide-3.0/compare-docs-vs-relational.html, visited 2017-07-01.
- [29] Yishan Li and Sathiamoorthy Manoharan. A performance comparison of sql and nosql databases. In *Communications, computers and signal processing (PACRIM), 2013 IEEE pacific rim conference on*, pages 15–19. IEEE, 2013.
- [30] MongoDB Inc. Bson types. BSON Types MongoDB Manual 3.4. https://docs. mongodb.com/manual/reference/bson-types/#document-bson-type-date, visited 2017-07-01.
- [31] Google Maps. Google maps upper west side. *Google Maps*, 2017. https://www.google. de/maps/@40.7768592, -73.9512002, 17.61z, visited 2017-07-12.

- [32] Google Maps. Google maps manhattan. Google Maps, 2017. https://www.google.de/ maps/@40.7552877, -73.9941502, 12.57z, visited 2017-07-09.
- [33] Google Maps. Google maps central park. Google Maps, 2017. https://www.google.de/ maps/@40.7664135, -73.9765463, 17.12z, visited 2017-07-15.
- [34] Google Maps. Google maps central park. *Google Maps*, 2017. https://www.google.de/ maps/@40.7802723, -73.9681985, 14.42z, visited 2017-07-10.
- [35] Motivate International Inc. Central park full loop. *Citi Bike NYC*. https://www.citibikenyc.com/rides/central-park-full-loop, visited 2017-07-06.
- [36] Google Maps. Google maps central station. Google Maps, 2017. https://www.google. de/maps/@40.7517035, -73.9864906, 16.19z, visited 2017-07-14.
- [37] Google Maps. Google maps chelsea. Google Maps, 2017. https://www.google.de/ maps/@40.7457235, -73.9990371, 16.1z, visited 2017-07-15.
- [38] Heather Liang. Everything you need to know about chelsea, nyc high line, chelsea piers, chelsea market. NYCgo.com, May 2016. https://www.nycgo.com/articles/ must-see-chelsea-slideshow, visited 2017-07-22.
- [39] Google Maps. Google maps greenwich and east village. *Google Maps*, 2017. https: //www.google.de/maps/@40.7305573, -73.9903861, 15.54z, visited 2017-07-16.
- [40] Google Maps. Google maps lower manhattan. Google Maps, 2017. https://www.google. de/maps/@40.7058441, -73.9948061, 14.94z, visited 2017-07-11.
- [41] The City of New York. City hall park. City Hall Park : NYC Parks. https://www. nycgovparks.org/parks/city-hall-park, visited 2017-07-07.
- [42] Joe Lertola. Day vs. night population maps, 2007. http://www.joelertola.com/grfx/ population/d_n.html, visited 2017-07-02.
- [43] Google Maps. Google maps brooklyn bridge park. Google Maps, 2017. https://www.google.de/maps/@40.6934078, -73.9994354, 16.53z, visited 2017-07-12.
- [44] Google Maps. Google maps brooklyn and queens. Google Maps, 2017. https://www.google.de/maps/@40.7294569, -73.9484712, 14.24z, visited 2017-07-15.
- [45] Motivate International Inc. Brooklyn bridge park. Citi Bike NYC. https://www. citibikenyc.com/rides/brooklyn-bridge-park, visited 2017-07-22.
- [46] Cyclist killed by bus in new york's first citi bike fatality, Jun 2017. https://www.nytimes. com/2017/06/12/nyregion/citi-bike-death-manhattan.html, visited 2017-08-07.
- [47] Motivate International Inc. Women who bike. https://www.citibikenyc.com/ women-who-bike-2017, visited 2017-07-06.

A APPENDIX

Start Station Name	Start Station ID	End Station Name	End Station ID	Trip Count
Central Park S & 6 Ave	2006	Central Park S & 6 Ave	2006	29771
Grand Army Plaza & Central Park S	281	Grand Army Plaza & Central Park S	281	12437
Broadway & W 60 St	499	Broadway & W 60 St	499	12334
E 43 St & Vanderbilt Ave	318	W 41 St & 8 Ave	477	7555
Centre St & Chambers St	387	Centre St & Chambers St	387	7455
Grand Army Plaza & Central Park S	281	Broadway & W 60 St	499	7118
West St & Chambers St	426	West St & Chambers St	426	6836
E 7 St & Avenue A	432	Lafayette St & E 8 St	293	6540
Broadway & W 58 St	457	Broadway & W 58 St	457	6408
West Thames St	363	Vesey PI & River Terrace	327	6291
W 21 St & 6 Ave	435	9 Ave & W 22 St	509	6283
12 Ave & W 40 St	514	West St & Chambers St	426	6210
E 42 St & Vanderbilt Ave	519	W 33 St & 7 Ave	492	6084
Washington Square E	294	University PI & E 14 St	382	6059
W 17 St & 8 Ave	116	8 Ave & W 31 St	521	5976
W 21 St & 6 Ave	435	W 22 St & 10 Ave	462	5975
Central Park S & 6 Ave	2006	Broadway & W 60 St	499	5533
12 Ave & W 40 St	514	12 Ave & W 40 St	514	5497
Broadway & W 60 St	499	Grand Army Plaza & Central Park S	281	5470
Grand Army Plaza & Central Park S	281	Central Park S & 6 Ave	2006	5395
Vesev PI & River Terrace	327	West Thames St	363	5290
F 10 St & Avenue A	445	Lafavette St & F 8 St	293	5284
Greenwich St & N Moore St	329	Greenwich St & Warren St	147	5255
E 43 St & Vanderbilt Ave	318	W 33 St & 7 Ave	492	5124
Vesev PL & Biver Terrace	327	Greenwich St & N Moore St	329	5089
E 42 St & Vanderbilt Ave	519		477	5086
Lafavette St & F 8 St	293	F 10 St & Avenue A	445	5003
West St & Chambers St	426	12 Ave & W 40 St	514	4836
Cleveland Pl & Spring St	151	Lafavette St & E 8 St	293	4678
E 32 St & Park Ave	472	2 Ave & E 31 St	528	4668
Lafavette St & E 8 St	203		432	4575
Greenwich St & Warren St	147	Centre St & Chambers St	387	4563
8 Ave & W 31 St	521	11 Ave & W 27 St	458	4545
Central Park S & 6 Ave	2006	Grand Army Plaza & Central Park S	281	4539
W 22 St & 10 Ave	462	$W 22 $ St & 8 Δv_{P}	453	4529
Greenwich St & N Moore St	329	Vesev PL& Biver Terrace	327	4340
Vesev PI & Biver Terrace	327	Vesey PL& Biver Terrace	327	4302
11 Ave & W 27 St	458	8 Δvo & W 31 St	521	4147
W 22 St & 10 Ave	462	W 21 St & 6 Ave	435	4143
Lafavette St & F 8 St	202	F 17 St & Broadway	497	4140
Old Eulton St	217	Centre St & Chambers St	387	4047
W 14 St & The High Line	217	West St & Chambers St	426	4047
W 21 St & 6 Avo	125		420	4011
Contro St & Chambora St	433	Codmon Plaza E & Tillony St	400	2000
	507		202	3999
Grand Army Plaza & Control Park S	0∠1 091	Broadway & W 59 St	403	2055
Liniversity DI & E 14 St	201	Graanwigh Ave 8 9 Ave	407	3900
$\frac{14 \text{ Sl}}{14 \text{ Sl}}$	JOZ 457	Greenwich Ave a 6 Ave	204	2029
Divauway & W 26 St	437	Granu Army Flaza & Gentral Park S	201	3930
Divadway & vv 24 St	444	E 1/ St & Broadway	497	3894
Aliantic Ave & Furman St	398	Aliantic Ave & Furman St	398	3860

Table 4: Top Routes for Plateau 1.

Table 5: Top Routes for Plateau 2.

Start Station Name	Start Station ID	End Station Name	End Station ID	Trip Count
Central Park S & 6 Ave	2006	Central Park S & 6 Ave	2006	5966
Central Park S & 6 Ave	2006	5 Ave & E 78 St	3143	2703
W 21 St & 6 Ave	435	9 Ave & W 22 St	509	2485
E 42 St & Vanderbilt Ave	519	W 33 St & 7 Ave	492	2346
E 42 St & Vanderbilt Ave	519	W 41 St & 8 Ave	477	2221
W 21 St & 6 Ave	435	W 22 St & 10 Ave	462	2174
E 85 St & York Ave	3150	E 85 St & 3 Ave	3147	2129
Centre St & Chambers St	387	Centre St & Chambers St	387	2088
E 42 St & Vanderbilt Ave	519	E 24 St & Park Ave S	491	2078
N 6 St & Bedford Ave	3093	Wythe Ave & Metropolitan Ave	2002	2065
12 Ave & W 40 St	514	West St & Chambers St	426	1995
Vernon Blvd & 50 Ave	3119	46 Ave & 5 St	3124	1986
1 Ave & E 78 St	3140	E 77 St & 3 Ave	3154	1937
Central Park S & 6 Ave	2006	Central Park West & W 85 St	3168	1869
E 7 St & Avenue A	432	Lafayette St & E 8 St	293	1857
E 42 St & Vanderbilt Ave	519	Broadway & W 32 St	498	1851
Wythe Ave & Metropolitan Ave	2002	N 6 St & Bedford Ave	3093	1830
1 Ave & E 68 St	3141	Lexington Ave & E 63 St	3155	1772
Grand Army Plaza & Central Park S	281	Grand Army Plaza & Central Park S	281	1757
5 Ave & E 78 St	3143	Central Park West & W 85 St	3168	1749
W 17 St & 8 Ave	116	W 20 St & 11 Ave	459	1701
1 Ave & E 62 St	3142	1 Ave & E 68 St	3141	1679
Central Park West & W 85 St	3168	Central Park S & 6 Ave	2006	1663
South End Ave & Liberty St	3002	Bus Slip & State St	427	1648
E 30 St & Park Ave S	546	E 42 St & Vanderbilt Ave	519	1627
Greenwich St & N Moore St	329	Vesey PI & River Terrace	327	1627
N Henry St & Richardson St	3097	Graham Ave & Conselyea St	3086	1616
W 41 St & 8 Ave	477	E 42 St & Vanderbilt Ave	519	1581
West Thames St	363	Vesey PI & River Terrace	327	1568
E 7 St & Avenue A	432	Shevchenko PI & E 7 St	300	1556
Central Park S & 6 Ave	2006	5 Ave & E 73 St	3137	1556
Lafayette St & Jersey St	250	Lafayette St & E 8 St	293	1552
Vesey PI & River Terrace	327	Greenwich St & N Moore St	329	1551
Central Park West & W 85 St	3168	Central Park West & W 85 St	3168	1541
E 77 St & 3 Ave	3154	1 Ave & E 78 St	3140	1539
46 Ave & 5 St	3124	Vernon Blvd & 50 Ave	3119	1529
W 22 St & 10 Ave	462	W 21 St & 6 Ave	435	1527
11 Ave & W 41 St	478	8 Ave & W 33 St	490	1519
Lexington Ave & E 63 St	3155	1 Ave & E 68 St	3141	1517
W 22 St & 10 Ave	462	W 22 St & 8 Ave	453	1514
Central Park S & 6 Ave	2006	Central Park West & W 72 St	3165	1492
E 47 St & Park Ave	359	W 41 St & 8 Ave	477	1484
E 24 St & Park Ave S	491	E 42 St & Vanderbilt Ave	519	1484
W 26 St & 8 Ave	494	11 Ave & W 27 St	458	1459
Graham Ave & Conselyea St	3086	N Henry St & Richardson St	3097	1454
W 26 St & 10 Ave	388	W 26 St & 8 Ave	494	1452
E 32 St & Park Ave	472	E 33 St & 1 Ave	527	1448
Lafayette St & E 8 St	293	E 10 St & Avenue A	445	1441
Henry St & Grand St	502	Canal St & Rutgers St	307	1440
W 21 St & 6 Ave	435	W 22 St & 8 Ave	453	1439

Table 6: Top Routes for Plateau 3.

Start Station Name	Start Station ID	End Station Name	End Station ID	Trip Count
Central Park S & 6 Ave	2006	Central Park S & 6 Ave	2006	2982
Central Park S & 6 Ave	2006	5 Ave & E 88 St	3282	2979
E 7 St & Avenue A	432	Cooper Square & E 7 St	3263	2691
W 21 St & 6 Ave	435	9 Ave & W 22 St	509	2535
E 42 St & Vanderbilt Ave	519	W 33 St & 7 Ave	492	2133
E 42 St & Vanderbilt Ave	519	E 24 St & Park Ave S	491	2095
E 42 St & Vanderbilt Ave	519	Broadway & W 32 St	498	2083
E 85 St & York Ave	3150	E 85 St & 3 Ave	3147	2049
N 6 St & Bedford Ave	3093	Wythe Ave & Metropolitan Ave	2002	2034
Greenwich Ave & Charles St	383	Greenwich Ave & Charles St	383	2023
Wythe Ave & Metropolitan Ave	2002	N 6 St & Bedford Ave	3093	1882
Cooper Square & E 7 St	3263	E 6 St & Avenue B	317	1815
12 Ave & W 40 St	514	West St & Chambers St	426	1781
W 26 St & 8 Ave	494	W 27 St & 10 Ave	3258	1741
1 Ave & E 62 St	3142	1 Ave & E 68 St	3141	1738
Central Park S & 6 Ave	2006	5 Ave & E 78 St	3143	1729
E 6 St & Avenue B	317	Cooper Square & E 7 St	3263	1724
Central Park S & 6 Ave	2006	Central Park North & Adam Clayton Powell Blvd	3374	1724
Centre St & Chambers St	387	Centre St & Chambers St	387	1721
E 42 St & Vanderbilt Ave	519	W 41 St & 8 Ave	477	1716
1 Ave & E 68 St	3141	Lexinaton Ave & E 63 St	3155	1683
E 33 St & 1 Ave	527	W 33 St & 7 Ave	492	1670
S 4 St & Wythe Ave	460	N 6 St & Bedford Ave	3093	1660
Bichardson St & N Henry St	3430	Graham Ave & Conselvea St	3086	1651
F 30 St & Park Ave S	546	F 42 St & Vanderbilt Ave	519	1643
Central Park S & 6 Ave	2006	5 Ave & F 73 St	3137	1572
N 6 St & Bedford Ave	3093	S 4 St & Wythe Ave	460	1552
11 Ave & W 41 St	478	8 Ave & W 33 St	490	1546
Cooper Square & F 7 St	3263	F 7 St & Avenue A	432	1545
6 Ave & W 33 St	505	E 42 St & Vanderbilt Ave	519	1535
Bedford Ave & Nassau Ave	3107	N 8 St & Driggs Ave	3090	1526
Vernon Blvd & 50 Ave	3119	46 Ave & 5 St	3124	1491
W 27 St & 10 Ave	3258	W 26 St & 8 Ave	494	1485
F 24 St & Park Ave S	491	F 42 St & Vanderbilt Ave	519	1484
DeKalb Ave & Vanderbilt Ave	258	DeKalb Ave & Hudson Ave	324	1442
Henry St & Grand St	502	Canal St & Butgers St	307	1435
5 Ave & F 88 St	3282	Central Park North & Adam Clayton Powell Blvd	3374	1433
Amsterdam Ave & W 73 St	3167	Amsterdam Ave & W 82 St	3171	1425
9 Ave & W 22 St	509	W 21 St & 6 Ave	435	1420
F 32 St & Park Ave	472	E 33 St & 1 Ave	527	1412
South End Ave & Liberty St	3002	Bus Slip & State St	427	1404
West St & Chambers St	426	12 Ave & W 40 St	514	1403
West St & Chambers St	426	Little West St & 1 Pl	2008	1400
Broadway & E 14 St	285	Washington PL & Broadway	335	1399
Broadway & E 22 St	402	F 42 St & Vanderhilt Ave	519	1377
McGuinness Blvd & Eagle St	3118	Vernon Blvd & 50 Ave	3119	1376
Lexington Ave & F 63 St	3155	1 Ave & E 68 St	3141	1362
E 42 St & Vanderbilt Ave	519	Broadway & E 22 St	402	1349
Central Park North & Adam Clayton Powell Plud	3374	Central Park S & 6 Ave	2006	1344
W 26 St & 8 Ave	1014		458	1330
W 20 St & D AVE	734	TI AVE & W 21 JL	-100	1009