
# Where to grab my coffee?

## A Site Survey of Coffee Shops for NYC Subway Riders


Yongkang Liu

January 2020

# Outlines

- Problems
- Data
- Methodology
- Results
- Discussion
- Conclusion

# Problems

- Q1. Any relationship between the subway ridership and the nearby service venue density?
  - Some intuition: More passengers (commuters and travelers) mean more business opportunity, therefore, it means more venues nearby.
- Q2. Is there any missed subway station with potentials for new business venues?
  - Identify the subway neighborhood with a high ridership but few venues
- Q3. Are current players doing a good job?
  - Compared with the peers, what is the performance of Starbucks?

# Data

- NYC subway turnstile data
  - Source: http://web.mta.info/developers/turnstile.html (Format: txt/csv)
  - Feature: passenger counts (enters and exits) at each turnstile of MTA subway stations. The counts are collected every 4 hours from each turnstile unit in a 24x7 schedule.
  - Usage: The passenger density in different subway stations at different time.

| | C/A | UNIT | SCP | STATION | LINENAME | DIVISION | DATE | TIME | DESC | ENTRIES | EXITS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 10/26/2019 | 00:00:00 | REGULAR | 7247322 | 2455491 |
| 1 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 10/26/2019 | 04:00:00 | REGULAR | 7247336 | 2455499 |
| 2 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 10/26/2019 | 08:00:00 | REGULAR | 7247351 | 2455532 |
| 3 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 10/26/2019 | 12:00:00 | REGULAR | 7247463 | 2455623 |
| 4 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 10/26/2019 | 16:00:00 | REGULAR | 7247755 | 2455679 |

# Data (Cont'd)

- NYC subway station data
  - Source: http://web.mta.info/developers/data/nyct/subway/Stations.csv
  - Features: station name, service lines/division, and location (latitude and longitude).
  - Usage: The location information (latitude and longitude) of each subway station in the Manhattan island.

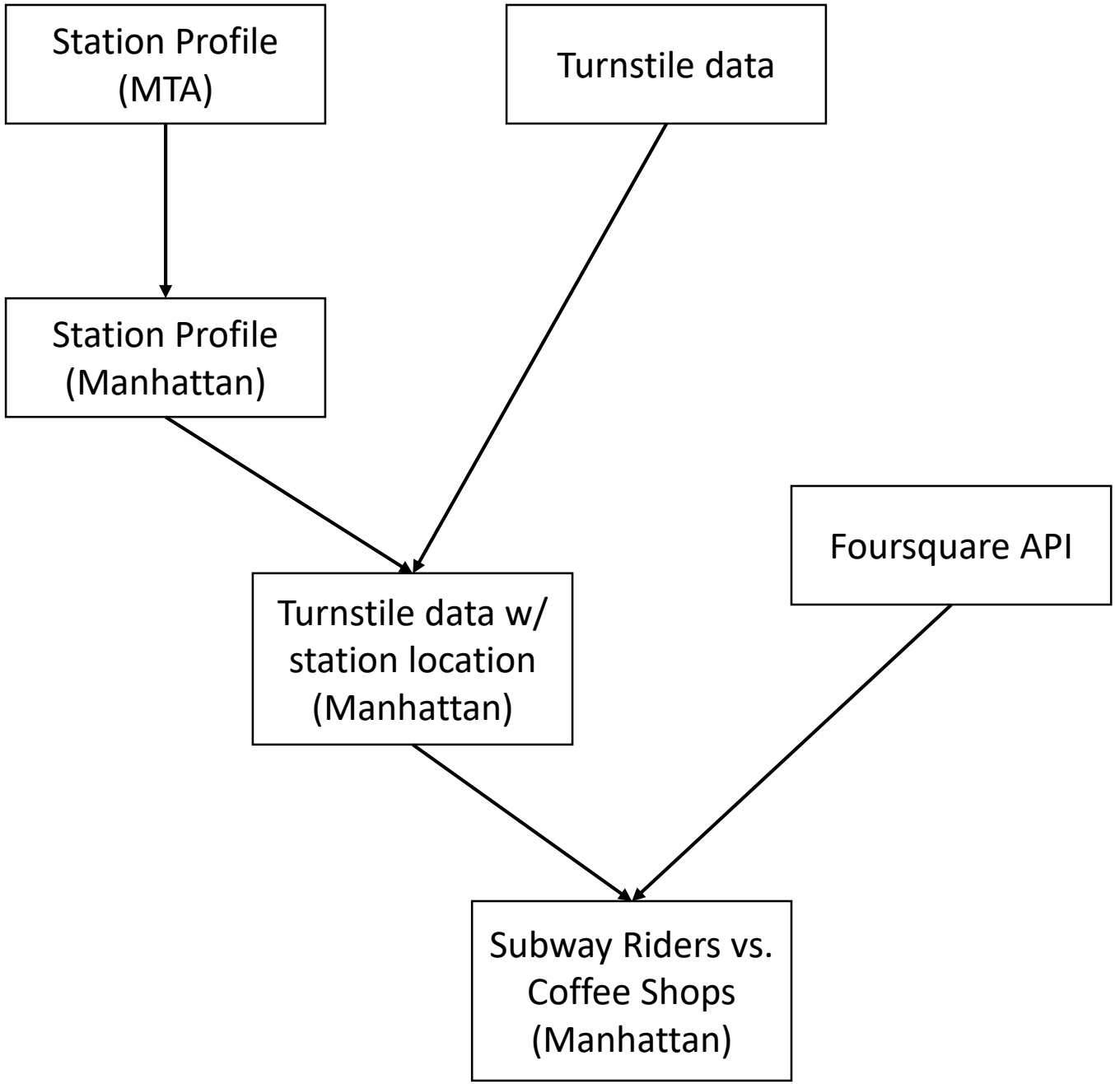
| | Station ID | Complex ID | GTFS Stop ID | Division | Line | Stop Name | Borough | Daytime Routes | Structure | GTFS Latitude | GTFS Longitude | North Direction Label | South Direction Label |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | R01 | BMT | Astoria | Astoria - Ditmars Blvd | Q | N W | Elevated | 40.775036 | -73.912034 | NaN | Manhattan |
| **1** | 2 | 2 | R03 | BMT | Astoria | Astoria Blvd | Q | N W | Elevated | 40.770258 | -73.917843 | Ditmars Blvd | Manhattan |
| **2** | 3 | 3 | R04 | BMT | Astoria | 30 Av | Q | N W | Elevated | 40.766779 | -73.921479 | Astoria - Ditmars Blvd | Manhattan |
| **3** | 4 | 4 | R05 | BMT | Astoria | Broadway | Q | N W | Elevated | 40.761820 | -73.925508 | Astoria - Ditmars Blvd | Manhattan |
| **4** | 5 | 5 | R06 | BMT | Astoria | 36 Av | Q | N W | Elevated | 40.756804 | -73.929575 | Astoria - Ditmars Blvd | Manhattan |

# Data (Cont’d)

- Foursquare location data
  - Source: https://api.foursquare.com/v2/venues/search
  - Format: Foursquare API queries
  - Feature: The venues around the searched locationalysis.

# Methodology

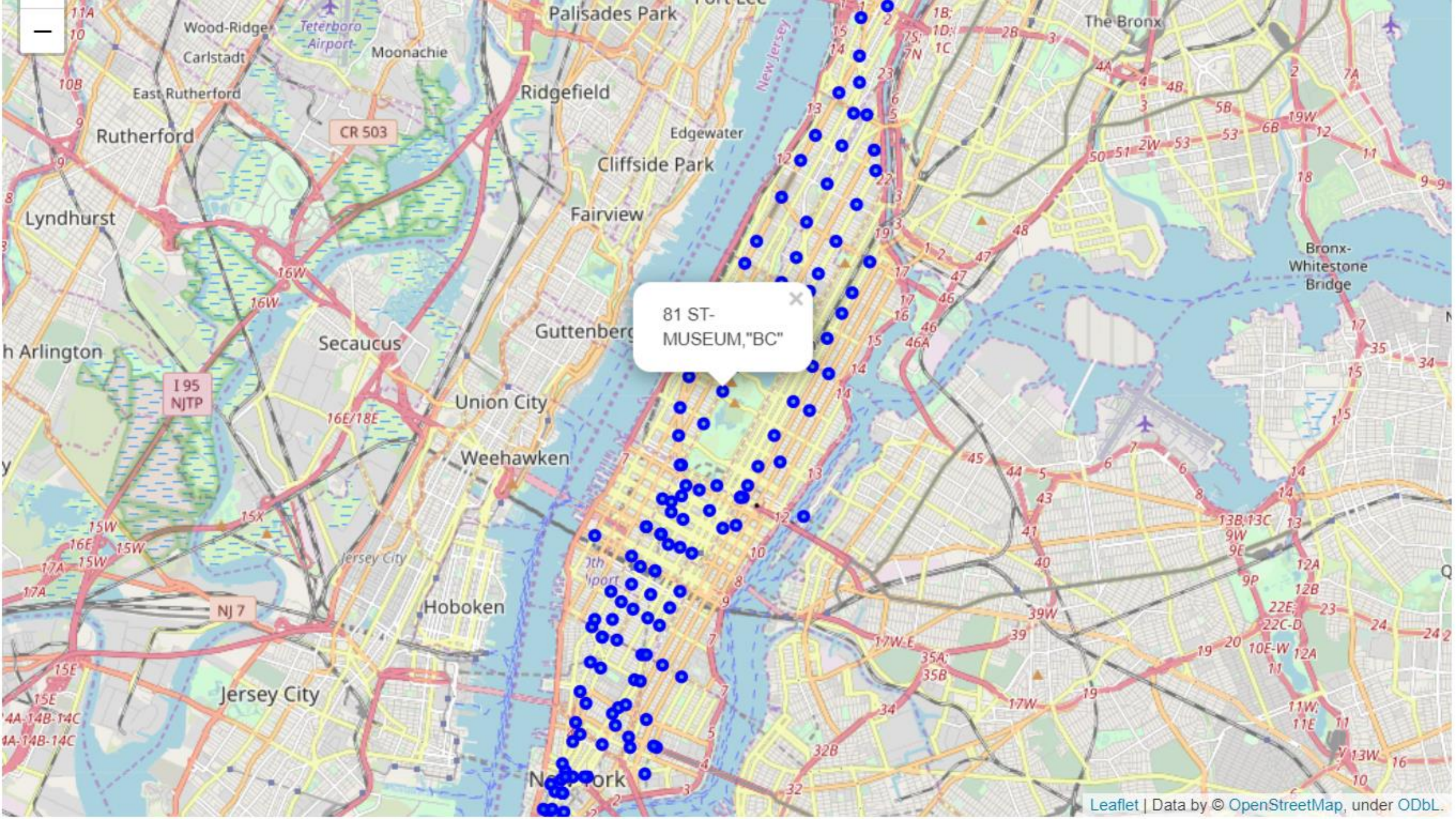
- Integrating data from different sources
- Study the statistical and geolocation features of the data
- Visualize and calculate the correlation between features



| | STATION_IDX | STATION | LINENAME | LOCATION | PASSENGERS | ENTRIES | EXITS | B_Counts | L_Counts | S_Counts | N_Counts | LATITUDE | LONGITUDE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | t0000000 | 59 ST | NQR456W | (40.762526, -73.967967) | 21683.8 | 12998.2 | 8685.6 | 6148.6 | 5485.4 | 7660.4 | 2389.4 | 40.762526 | -73.967967 |
| **1** | t0001001 | 5 AV/59 ST | NQRW | (40.764811, -73.973347) | 28250.4 | 16195.6 | 12054.8 | 7665.2 | 7832.8 | 11732.4 | 1018.8 | 40.764811 | -73.973347 |
| **2** | t0001002 | 57 ST-7 AV | NQRW | (40.764664, -73.980658) | 62275.4 | 36188.6 | 26086.8 | 16510.4 | 16266.8 | 22055.6 | 7442.6 | 40.764664 | -73.980658 |
| **3** | t0001003 | 49 ST | NQRW | (40.759901, -73.984139) | 43546.8 | 24051.8 | 19495.0 | 10989.6 | 10251.8 | 15957.6 | 6347.8 | 40.759901 | -73.984139 |
| **4** | t0002004 | TIMES SQ-42 ST | ACENQRS1237W | (40.755983, -73.986229) | 40079.4 | 16989.6 | 23089.8 | 12394.2 | 9851.2 | 14409.4 | 3424.6 | 40.755983 | -73.986229 |

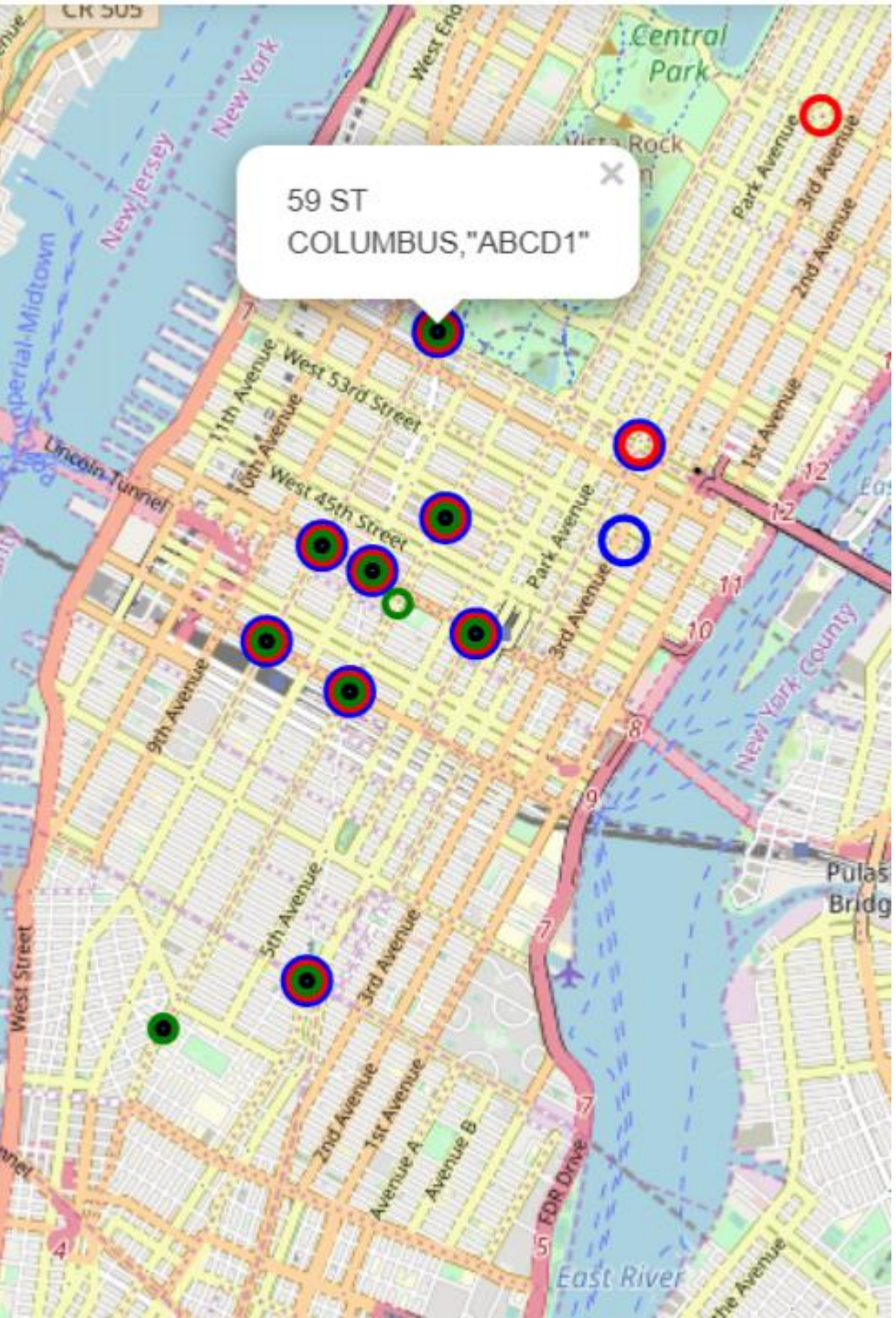
# Results

- Locations of subway stations and their ridership
- Coffee shops around subway stations


81 ST-
MUSEUM,"BC"
Palisades Park
Wood-Ridge
Carlstadt
Teterboro Airport
Moonachie
East Rutherford
Rutherford
CR 503
Ridgefield
Edgewater
Cliffside Park
Fairview
Lyndhurst
Secaucus
Guttenberg
Union City
Weehawken
Hoboken
Jersey City
I 95
NJTP
NJ 7
The Bronx
Bronx-
Whitestone
Bridge
New Jersey
Leaflet | Data by © OpenStreetMap, under ODbL.


# Top 10s in a Week

- 10 busiest stations in a weekday (RED)
- 10 busiest stations in a weekend (BLUE)




# Top 10s in a Day

- Top 10 busiest stations in a weekday in different periods
  - Breakfast time (05:00-11:00) in BLUE
  - Lunch time (11:00-16:00) in RED
  - Supper time (16:00-22:00) in GREEN
  - Night time (22:00-05:00+1Day) in BLACK

# Foursquare Venue Data

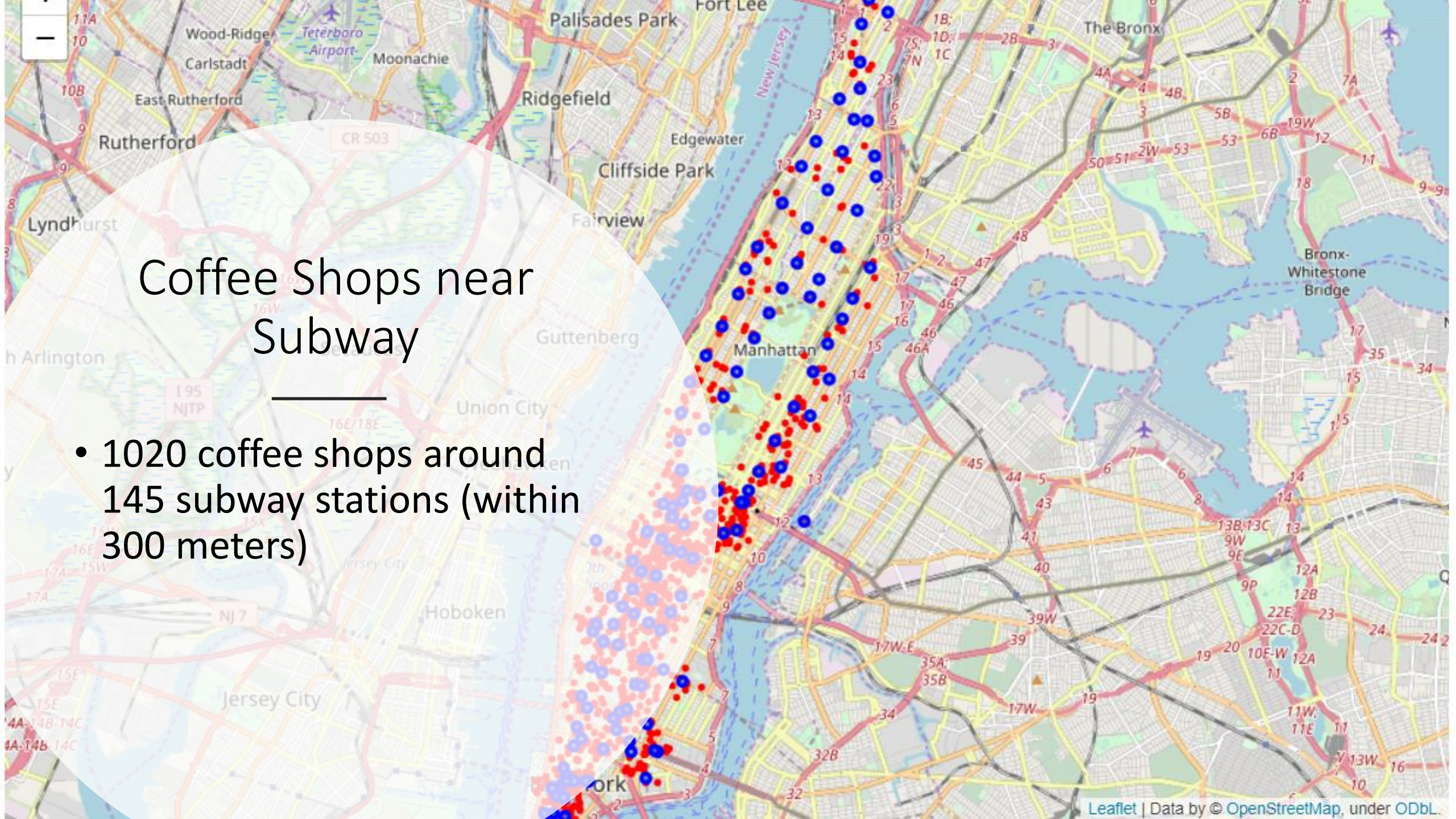
- 300M: the coffee shop number within 300 meters from a station, returned by venue search of “coffee”
- Starbucks: the Starbucks shop number in the “300M”
- StarbucksAll: the Starbucks shop number within 300 meters from station, returned by venue search of “Starbucks”

| | STATION_IDX | STATION | LINENAME | LATITUDE | LONGITUDE | Starbucks | StarbucksAll | 100M | 200M | 300M |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | t0000000 | 59 ST | NQR456W | 40.762526 | -73.967967 | 4 | 6 | 1 | 9 | 20 |
| 1 | t0001001 | 5 AV/59 ST | NQRW | 40.764811 | -73.973347 | 1 | 2 | 1 | 3 | 7 |
| 2 | t0001002 | 57 ST-7 AV | NQRW | 40.764664 | -73.980658 | 9 | 12 | 6 | 18 | 31 |
| 3 | t0001003 | 49 ST | NQRW | 40.759901 | -73.984139 | 7 | 11 | 8 | 21 | 34 |
| 4 | t0002004 | TIMES SQ-42 ST | ACENQRS1237W | 40.755983 | -73.986229 | 6 | 10 | 7 | 19 | 50 |

- Foursquare API has a limit of 50 venues per query

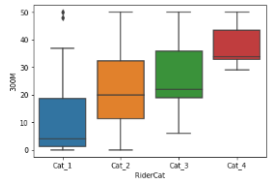
# Coffee Shops near Subway

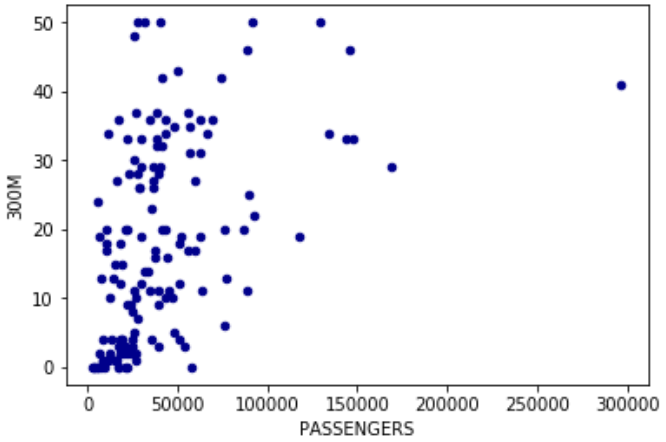
- 1020 coffee shops around 145 subway stations (within 300 meters)

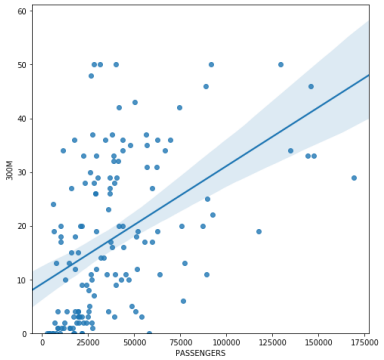


# More riders, more shops?

- The answer is partially Yes
- A relatively strong positive relationship between the per-station passengers and the nearby coffee shop number
- Outliners
- Different clusters of stations



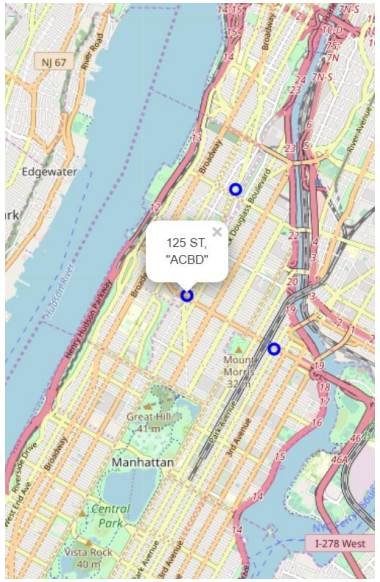



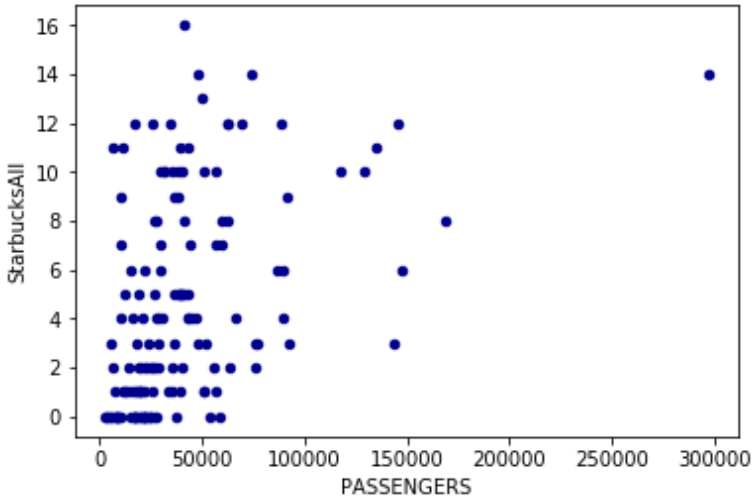

| | PASSENGERS | 300M |
|---|---|---|
| **PASSENGERS** | 1.000000 | 0.461635 |
| **300M** | 0.461635 | 1.000000 |

# Location Factors

- Hub stations (Top Left)
- Stations with low ridership and few shops (Top Right)
- Small stations with affluent coffee options (Bottom Left)
- Stations of potentials (good ridership figure, but fewer coffee shops)







# How Starbucks Play?

- 100+ stores that cover almost 90% NYC subway ridership
- The average accessibility from stations

# Conclusions

- Subway and Coffee fuel the busy New Yorkers
- Transportation ridership is among the factors to determine new service businesses
- Data unveils the temporal and spatial distribution of varying customer needs
- Data science is a powerful tool to answer many questions in future marketing

- Thank you!