EDA (EXPLORATORY DATA ANALYSIS) OF MTA TURNSTILE DATA NEW YORK CITY

YALIN YENER

June 13, 2020

Agenda

- Introduction
- Methodology
- Results
- Conclusions
- Future Work
- Appendix

Business Need:	**Optimize** the placement of client's street teams in subway stations who **collect email addresses** for their annual gala which is about women in tech and organised by WomenTechWomenYes (WTWY) International
Solution:	Using **MTA Turnstile** and **New York City Census** data, and analyse to find best placement for street teams. Visualization with **GIS** tools
Objectives:	• Finding most busiest subway **stations**, most crowded **day** and best **time slot** • Finding right person for client's target audience (**woman**, **technology**)

"**Exploratory Data Analysis** is a detective work"

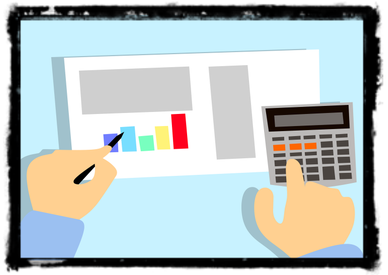
"**EDA** can never be the whole story, but nothing else can serve as the foundation stone--as the first step "

–John W. Tukey, 1970

Business Need

Data Acquire

Data Wrangling



Analyse

Visualization

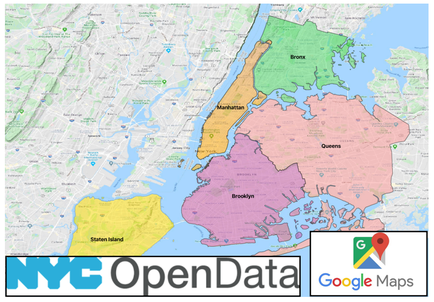
Data Sources

MTA Turnstile Data

- Between 01 January 2019 - 31 May 2019
- 5 months , 22 *.txt file

New York City Census

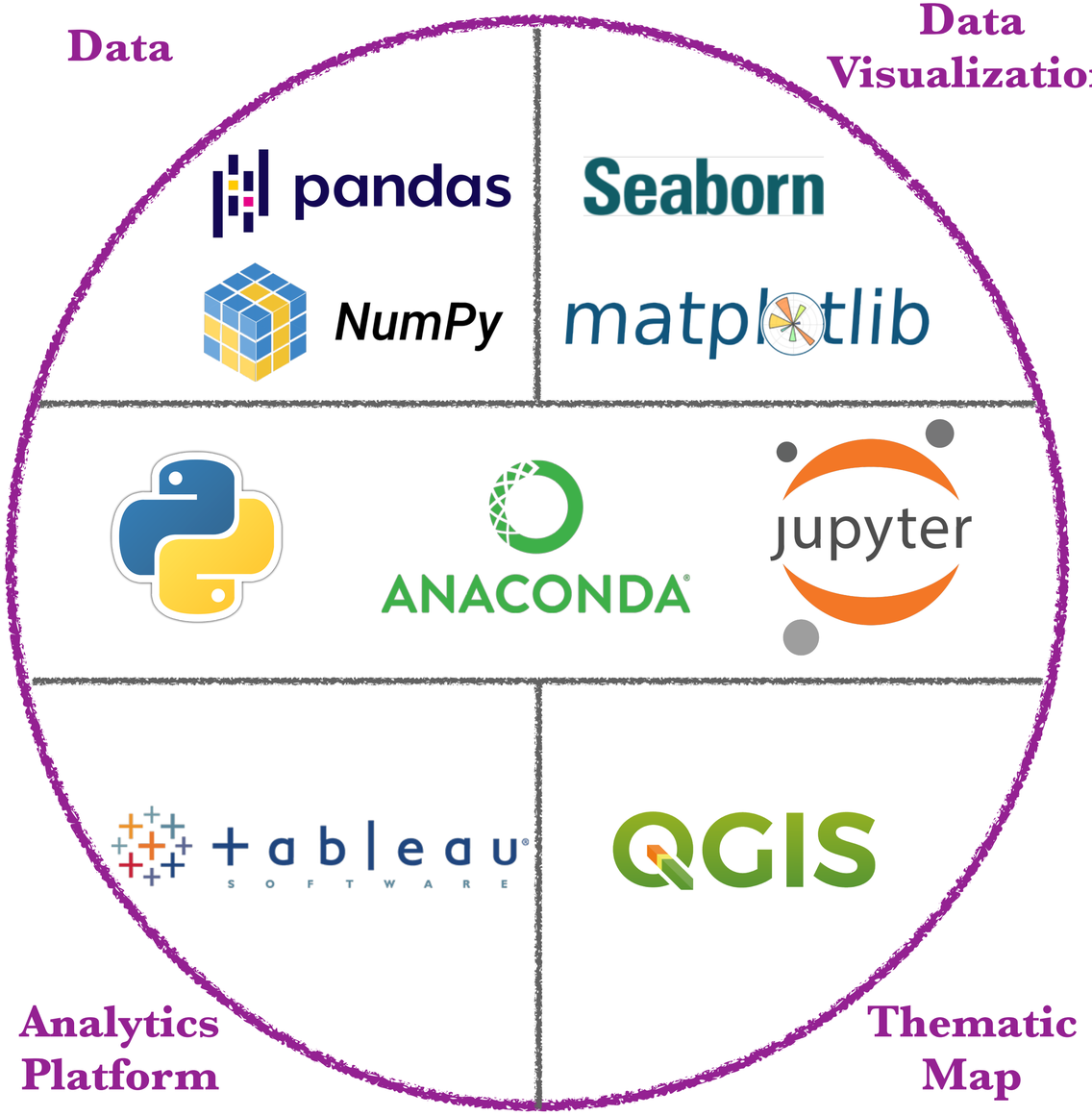
- Population
- Age and Sex
- Education
- Business
- Income



Spatial Data

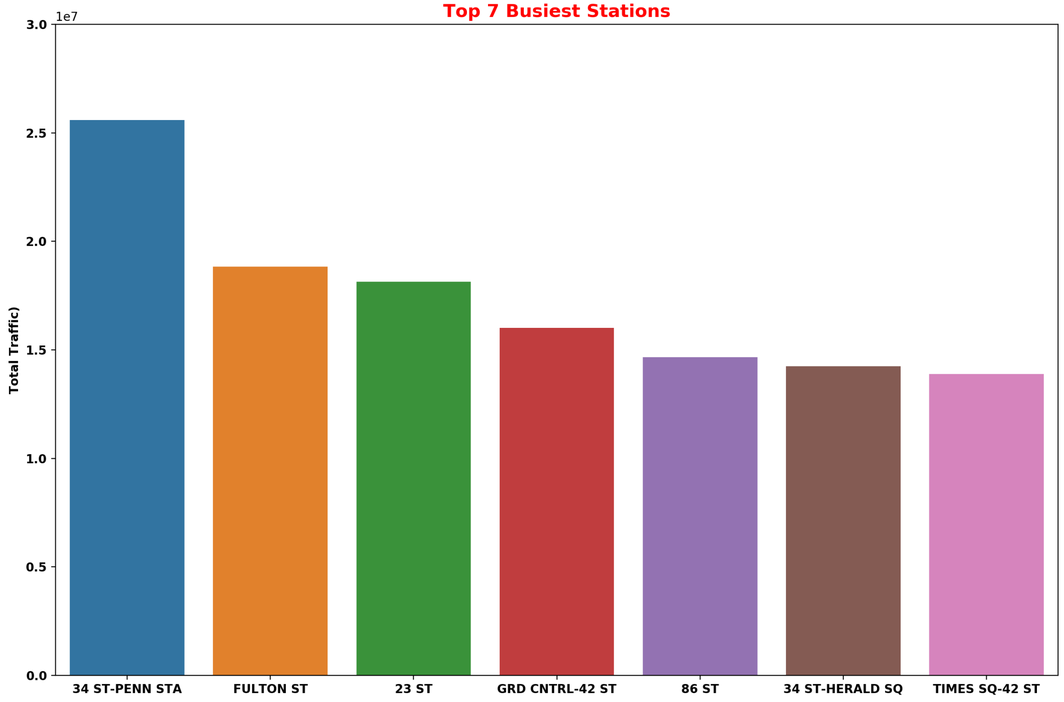
- Borough Border (shp file)
- Stations Point (shp file)
- Base Map (Web Map Service)

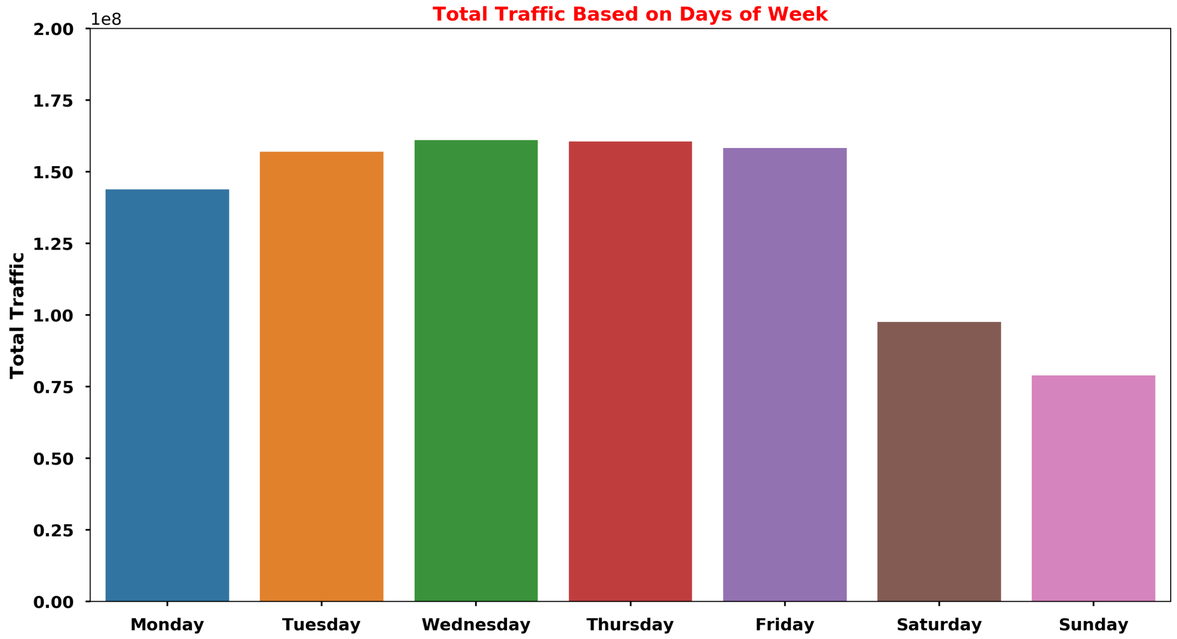
Tools

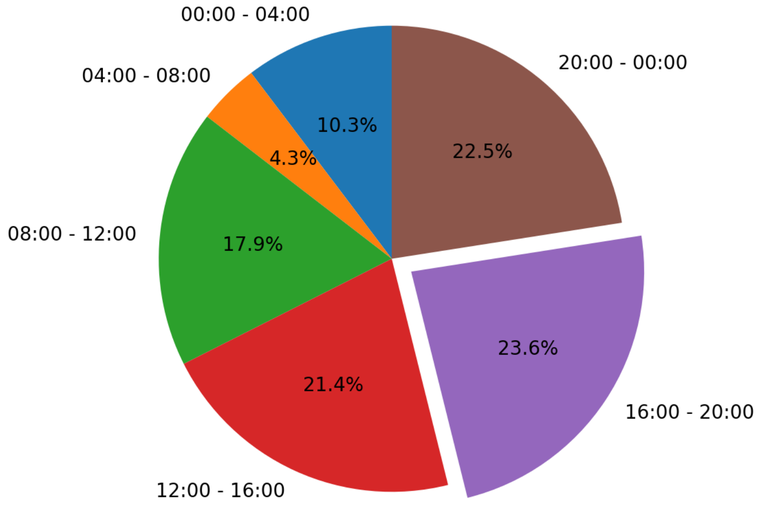


Brief Insight from Data

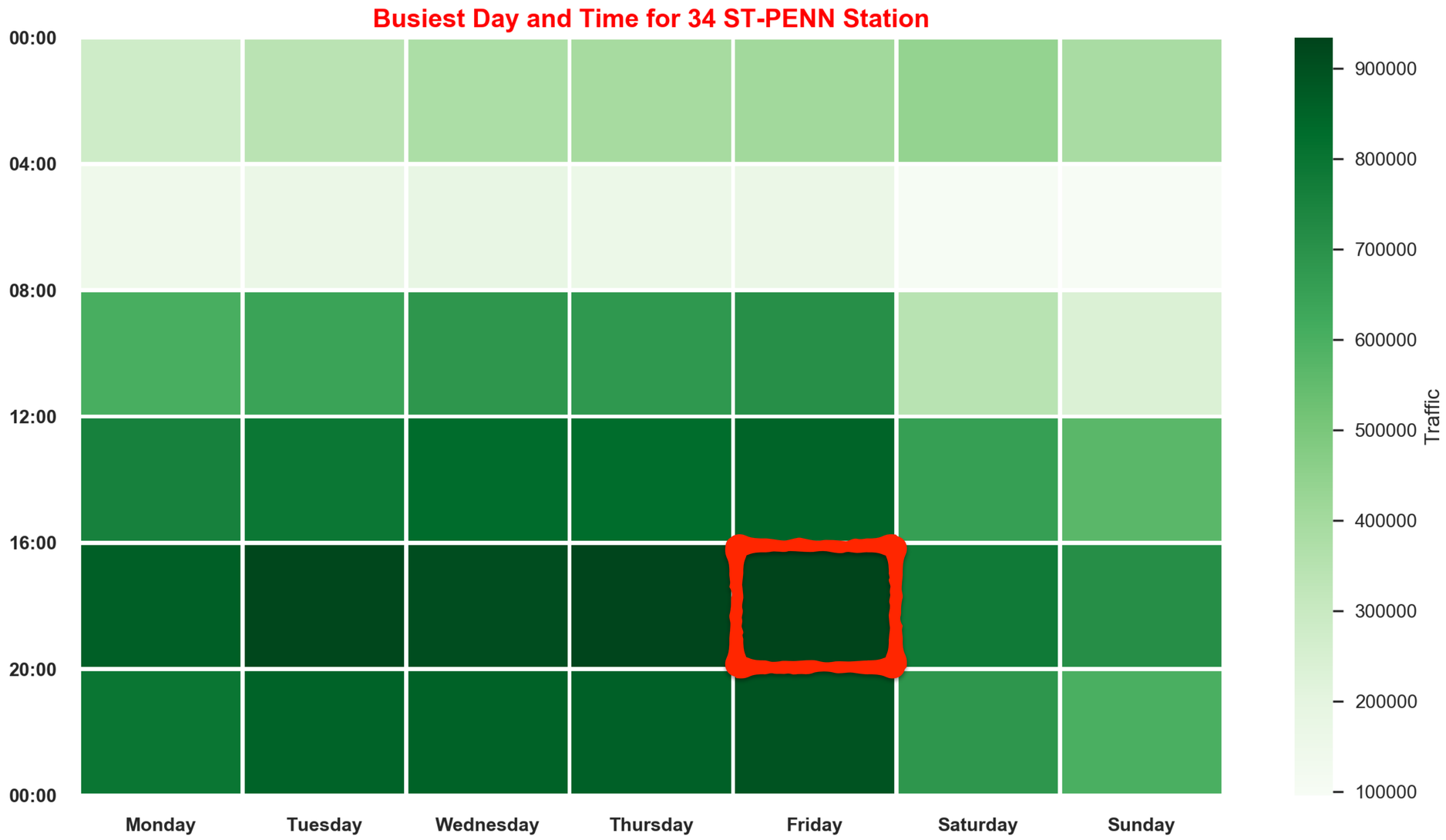
- 379 Stations
- 4932 Turnstile
- 4.48M rows
- **Most Busiest Station:** 34 ST-PENN Station
- **Most Busiest Day:** Wednesday
- **Most Busiest Time Slot:** 16:00 - 20:00



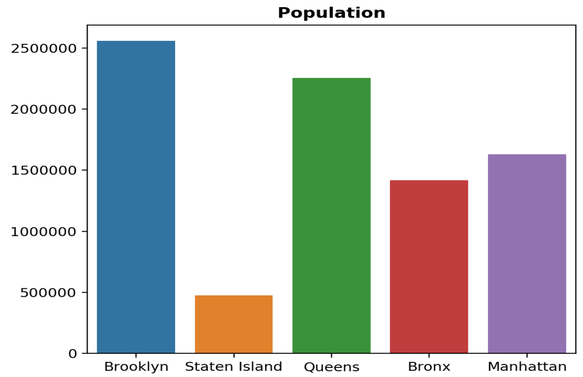


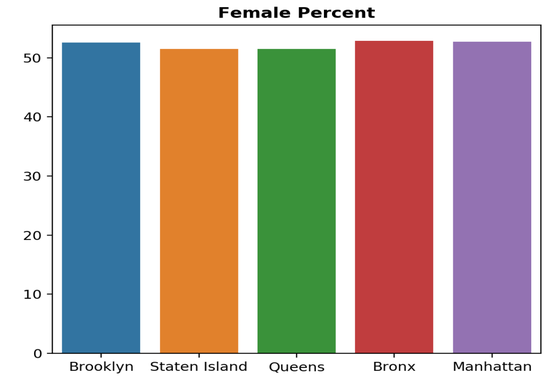


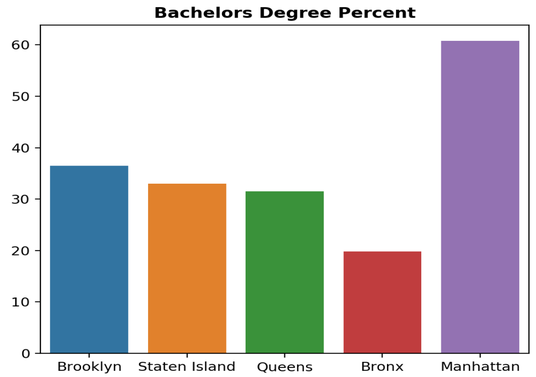
- **Weekday** is more busy
- **Tuesday**, **Thursday** and **Friday** are best days
- **16.00 - 20.00** time slot is best time
- The best option to reach max. people: **Friday, 16.00 - 20.00**

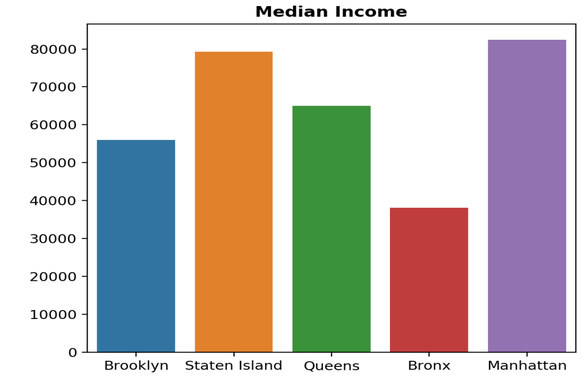


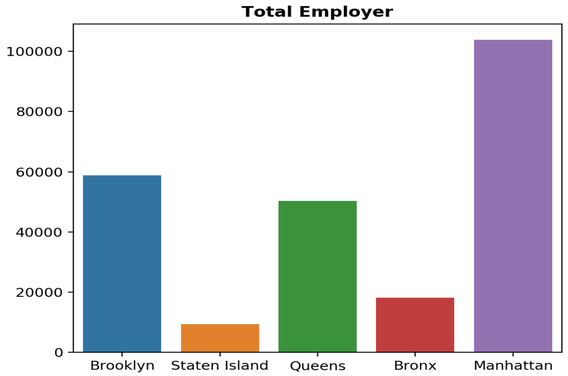
- **Manhattan** is the most suitable borough for income, annual payroll, employer and employment numbers and education
- **Brooklyn**, is the most suitable for Woman Owned Firms

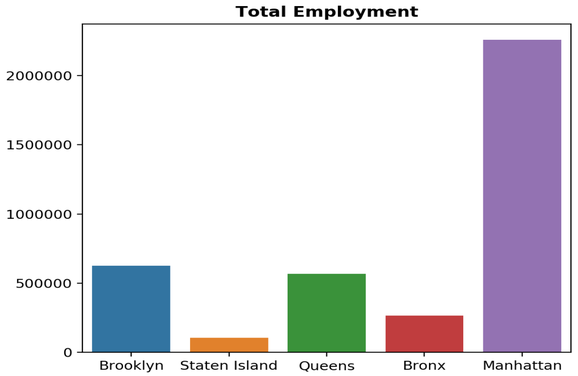


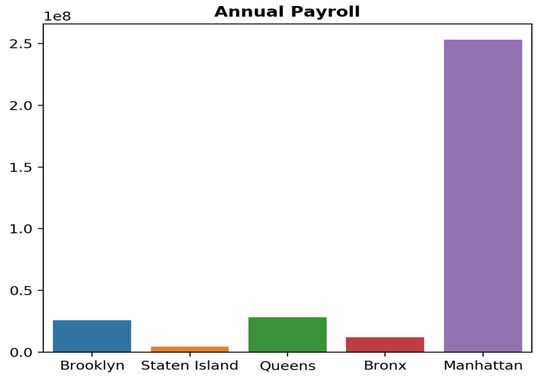


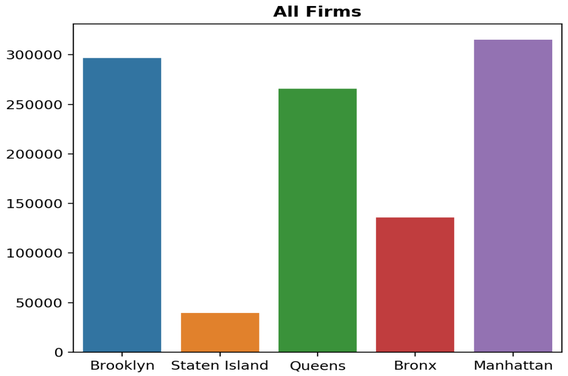


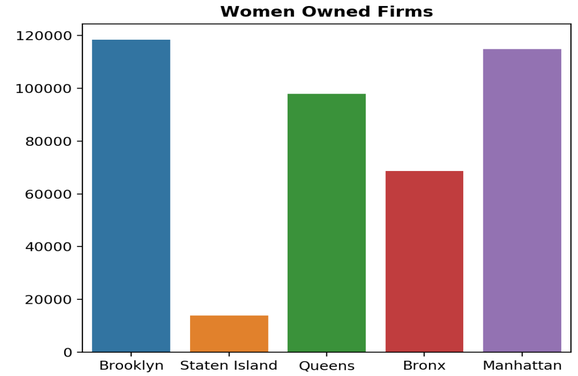








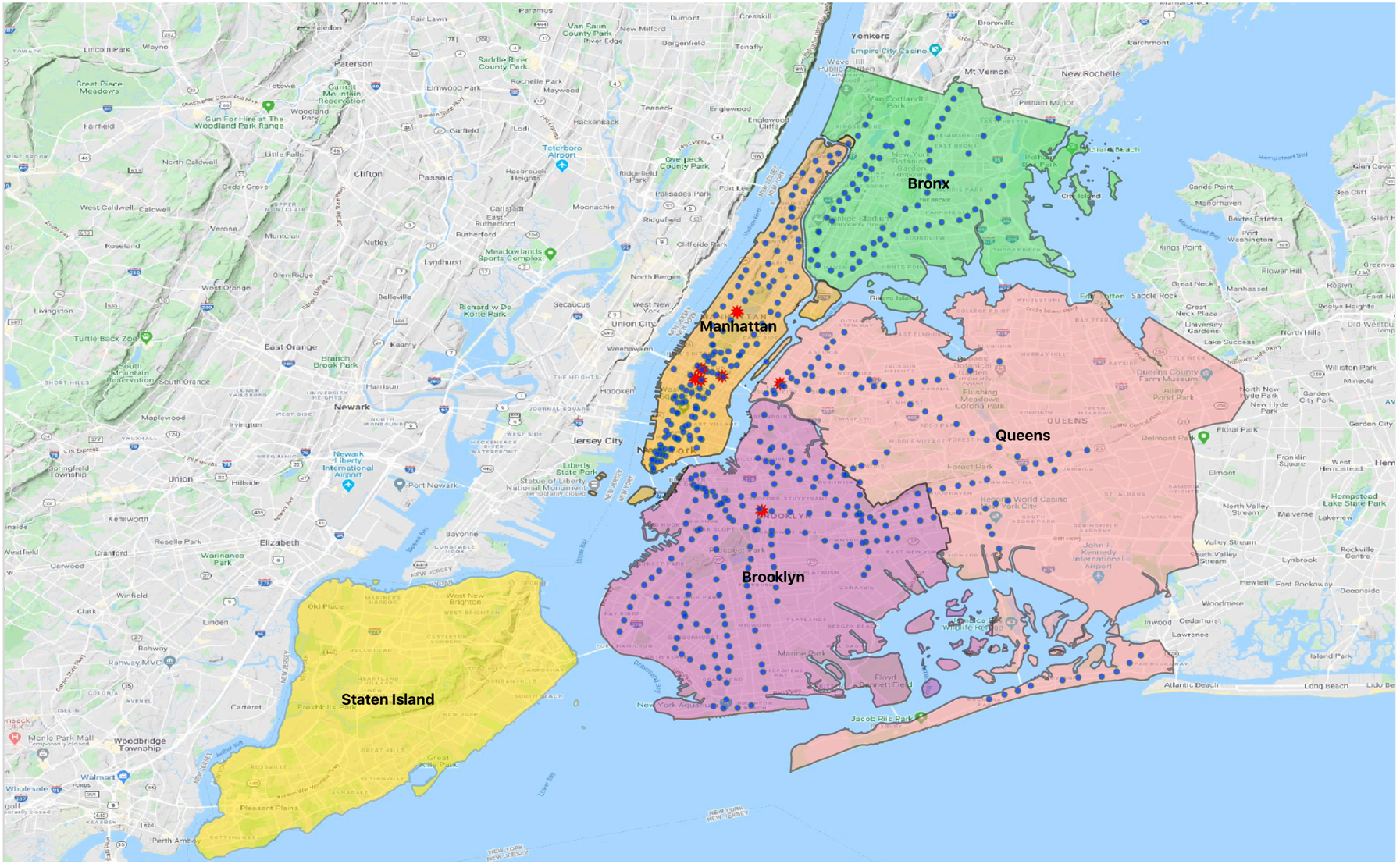




ST @Manhattan
- 34 ST ✓
- GRD CNTR
- 86 ST
- 34 ST
- TIMES SQ 42

ST @Brooklyn
- Fulton ✓

ST @Queens
- 23 ST



This project can be extended further by;

- analysing **census data** based on **zip code.**
- considering the location of **Business** and **Technology Center** throughout the city.
- Using **social media** data (LinkedIn, Twitter, Meetup) and acquire events about women in tech.

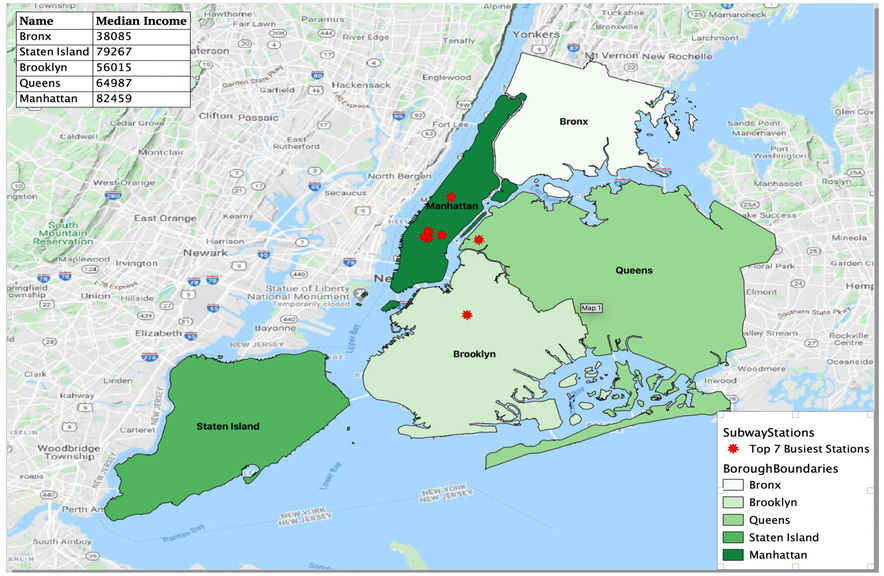
THANK YOU

YALIN YENER
+44 7786 761559
yalinyener@gmail.com
https://github.com/yalinyener
https://medium.com/@yalinyener
https://www.linkedin.com/in/yalinyener

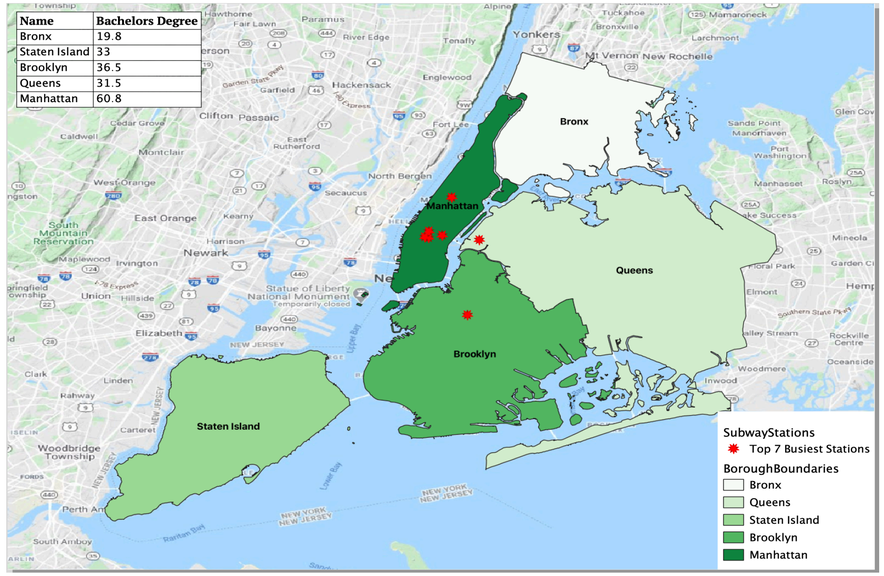
Thematic Maps

- Median Income
- Bachelor Degree
- Total Employment
- Woman Owned Firms

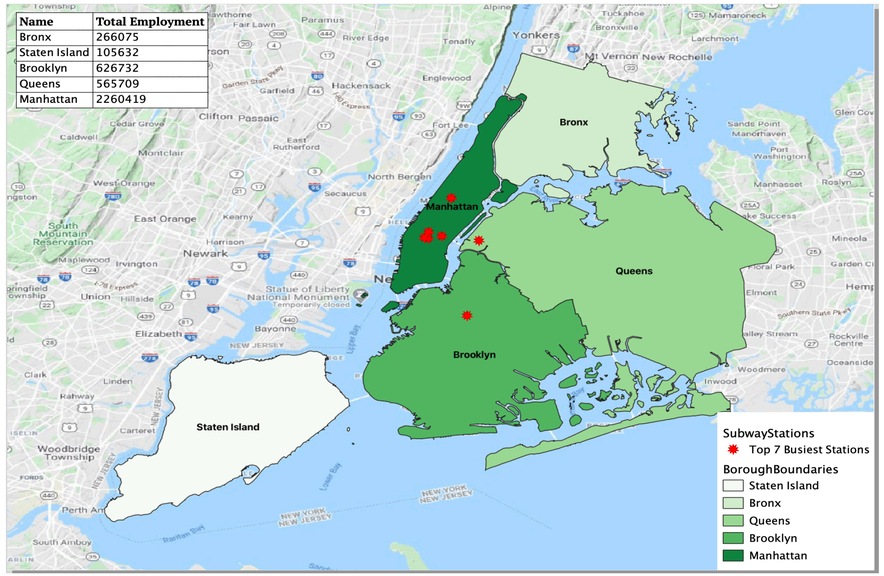
Name	Median Income
Bronx	38085
Staten Island	79267
Brooklyn	56015
Queens	64987
Manhattan	82459



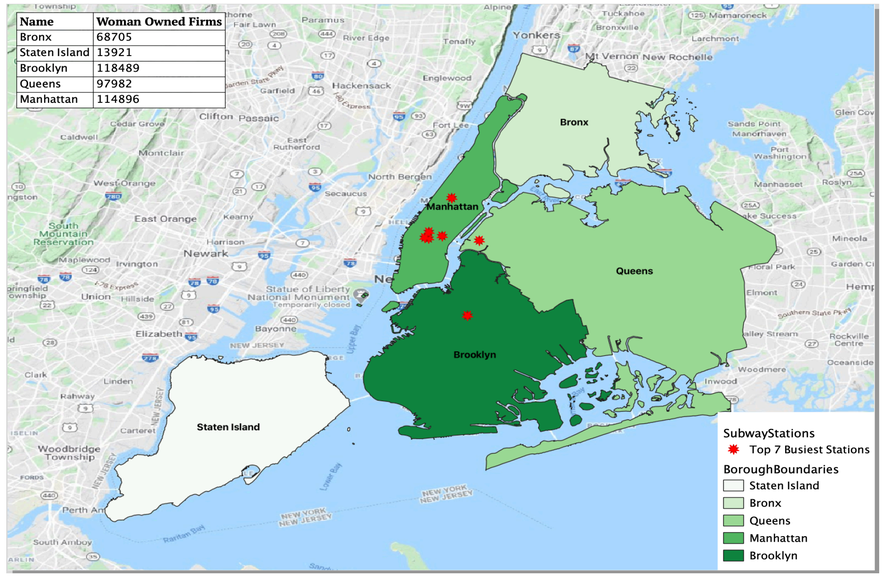
Name	Bachelors Degree
Bronx	19.8
Staten Island	33
Brooklyn	36.5
Queens	31.5
Manhattan	60.8



Name	Total Employment
Bronx	266075
Staten Island	105632
Brooklyn	626732
Queens	565709
Manhattan	2260419



Name	Woman Owned Firms
Bronx	68705
Staten Island	13921
Brooklyn	118489
Queens	97982
Manhattan	114896



	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	03:00:00	REGULAR	6914752	2344809
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	07:00:00	REGULAR	6914764	2344821
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	11:00:00	REGULAR	6914825	2344898
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	15:00:00	REGULAR	6915047	2344971
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	19:00:00	REGULAR	6915367	2345026
5	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/19/2019	23:00:00	REGULAR	6915525	2345049
6	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/20/2019	03:00:00	REGULAR	6915559	2345064
7	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/20/2019	07:00:00	REGULAR	6915565	2345072
8	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/20/2019	11:00:00	REGULAR	6915628	2345125
9	A002	R051	02-00-00	59 ST	NQR456W	BMT	01/20/2019	15:00:00	REGULAR	6915778	2345178

STATION	TURNSTILE	DATE	NAMEOFDAY	ENTRIES	EXITS	TRAFFIC
1 AV	H007-R248-00-00-00	01/01/2019	Tuesday	1457	2039	3496
1 AV	H007-R248-00-00-00	01/02/2019	Wednesday	3456	3647	7103
1 AV	H007-R248-00-00-00	01/03/2019	Thursday	3785	3959	7744
1 AV	H007-R248-00-00-00	01/04/2019	Friday	3806	4077	7883
1 AV	H007-R248-00-00-00	01/05/2019	Saturday	2149	3181	5330
1 AV	H007-R248-00-00-00	01/06/2019	Sunday	1779	2694	4473
1 AV	H007-R248-00-00-00	01/07/2019	Monday	3801	3735	7536
1 AV	H007-R248-00-00-00	01/08/2019	Tuesday	3844	3854	7698
1 AV	H007-R248-00-00-00	01/09/2019	Wednesday	4025	3978	8003
1 AV	H007-R248-00-00-00	01/10/2019	Thursday	4154	3745	7899

STATION	TURNSTILE	DATE	TIME	DATETIME	NAMEOFDAY	ENTRIES	EXITS	DIF_ENTRIES	DIF_EXITS	TRAFFIC
1 AV	H007-R248-00-00-00	01/01/2019	03:00:00	2019-01-01 03:00:00	Tuesday	14276525	16005314	0.0	0.0	0.0
1 AV	H007-R248-00-00-00	01/01/2019	07:00:00	2019-01-01 07:00:00	Tuesday	14276565	16005531	40.0	217.0	257.0
1 AV	H007-R248-00-00-00	01/01/2019	11:00:00	2019-01-01 11:00:00	Tuesday	14276733	16005804	168.0	273.0	441.0
1 AV	H007-R248-00-00-00	01/01/2019	15:00:00	2019-01-01 15:00:00	Tuesday	14277178	16006393	445.0	48.0	493.0
1 AV	H007-R248-00-00-00	01/01/2019	19:00:00	2019-01-01 19:00:00	Tuesday	14277679	16007007	501.0	48.0	549.0
1 AV	H007-R248-00-00-00	01/01/2019	23:00:00	2019-01-01 23:00:00	Tuesday	14277982	16007353	303.0	346.0	649.0
1 AV	H007-R248-00-00-00	01/02/2019	03:00:00	2019-01-02 03:00:00	Wednesday	14278032	16007445	50.0	92.0	142.0
1 AV	H007-R248-00-00-00	01/02/2019	07:00:00	2019-01-02 07:00:00	Wednesday	14278106	16007740	74.0	295.0	369.0
1 AV	H007-R248-00-00-00	01/02/2019	11:00:00	2019-01-02 11:00:00	Wednesday	14279160	16009110	67.0	48.0	115.0
1 AV	H007-R248-00-00-00	01/02/2019	15:00:00	2019-01-02 15:00:00	Wednesday	14279791	16009914	631.0	48.0	679.0