

# IBM Data Science Capstone

Keshav Narasimhan

# Introduction

- New York City: very diverse & populous
  - 8.4 million ppl
- Contains hundreds of high schools
- Good education important for students
- Good metric of education = SAT
- SAT: standardized test used for college admissions
- Higher score likely means better school



# Business Problem

- Newcomers to NYC
- Don't know area
- Applies to such individuals who have children
- Need good schools for kids
- Trying to answer:
  - Borough with highest average SAT score?
  - borough(s) with most number of schools better than national average on SAT?
  - Recurring venues around good schools?



# Data

- Need data that:
  - Contains info of NYC high schools
  - Average SAT scores
  - School locations
  - Name of schools
  - Surrounding venues
- Database on Kaggle
  - <https://www.kaggle.com/nycopendata/high-schools>
- Foursquare API
  - Retrieve surrounding venues for each school
  - Explore recurring venues

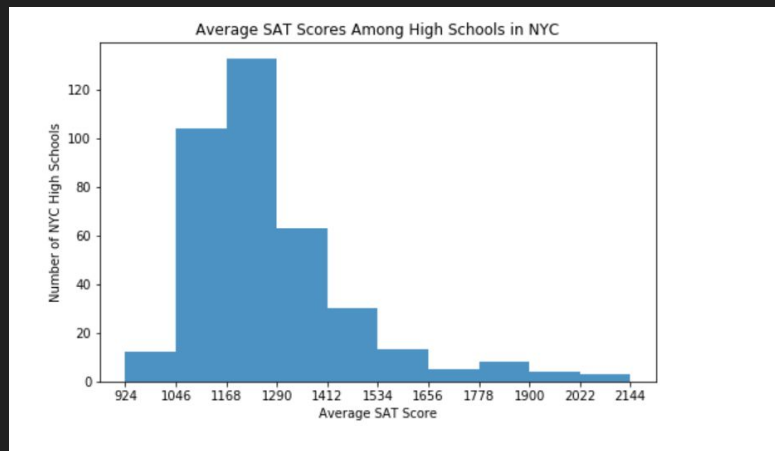
# kaggle



# Methodology

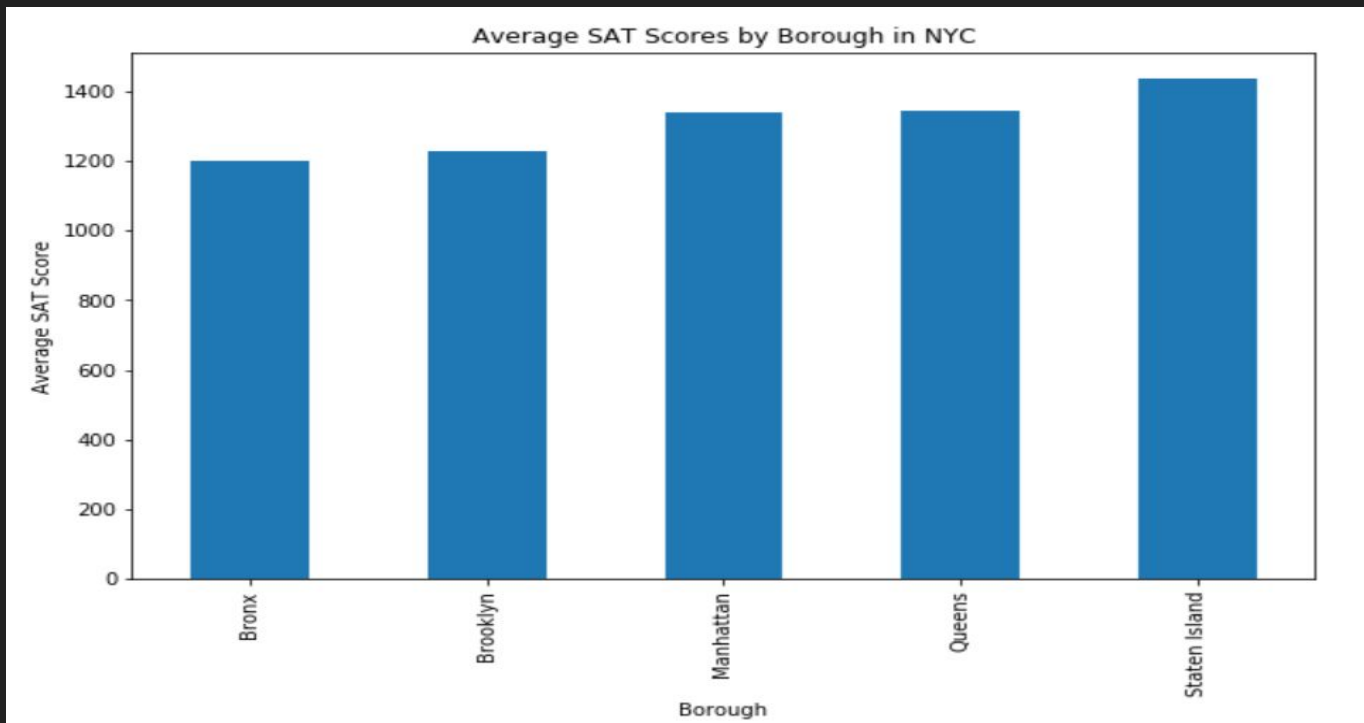
- Download data into pandas dataframe from Kaggle source
- View average SAT score in NYC high schools

	School ID	School Name	Borough	Building Code	Street Address	City	State	Zip Code	Latitude	Longitude	...	End Time	Student Enrollment	Percent White	Percent Black	Percent Hispani	
	0	02M260	Clinton School Writers and Artists	Manhattan	M933	425 West 33rd Street	Manhattan	NY	10001	40.75321	-73.99786	...	NaN	NaN	NaN	NaN	
	1	06M211	Inwood Early College for Health and Informatio...	Manhattan	M052	650 Academy Street	Manhattan	NY	10002	40.86605	-73.92486	...	3:00 PM	87.0	3.4%	21.8%	67.8%
	2	01M539	New Explorations into Science, Technology and ...	Manhattan	M022	111 Columbia Street	Manhattan	NY	10002	40.71873	-73.97943	...	4:00 PM	1735.0	28.6%	13.3%	18.0%
	3	02M294	Essex Street Academy	Manhattan	M445	350 Grand Street	Manhattan	NY	10002	40.71687	-73.98953	...	2:45 PM	358.0	11.7%	38.5%	41.3%
	4	02M308	Lower Manhattan Arts Academy	Manhattan	M445	350 Grand Street	Manhattan	NY	10002	40.71687	-73.98953	...	3:00 PM	383.0	3.1%	28.2%	56.9%
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	430	27Q302	Queens High School for Information, Research, ...	Queens	Q465	8-21 Bay 25th Street	Far Rockaway	NY	11691	40.60199	-73.76283	...	4:10 PM	381.0	2.1%	49.1%	43.6%



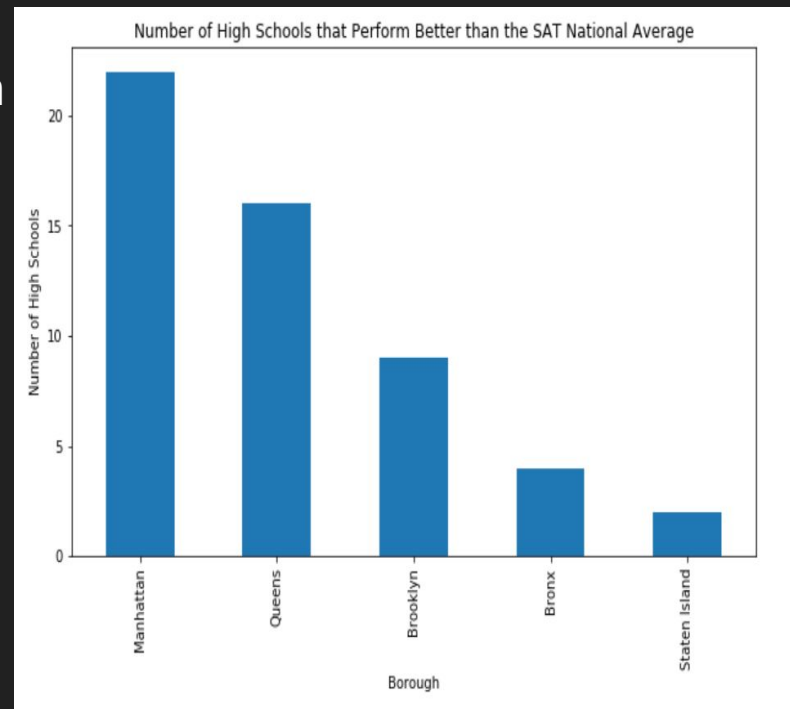
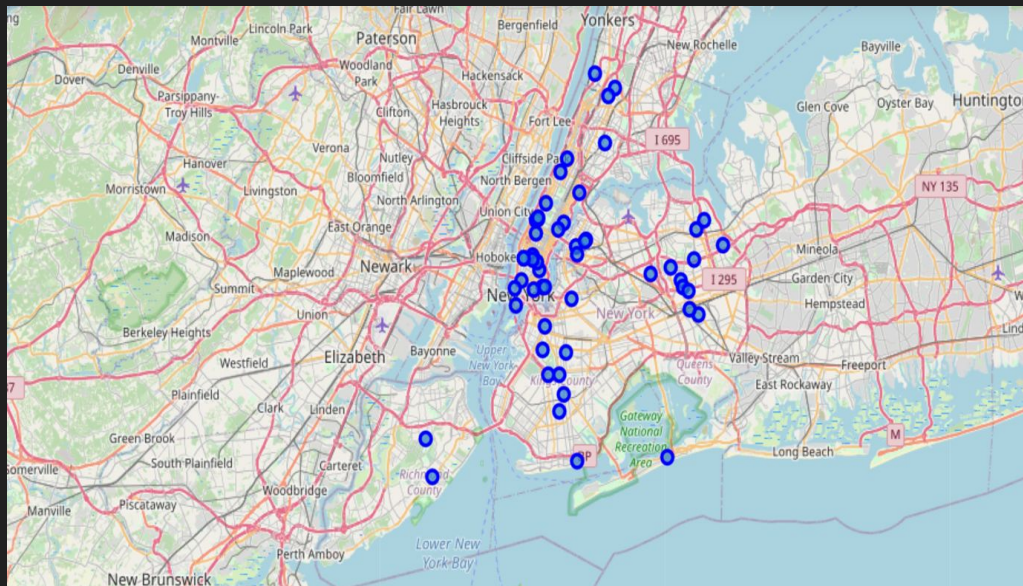
## Methodology cont'd

- View borough with highest average SAT score



# Methodology cont'd

- Focus only on schools with average SAT's higher than national average (1450)
- View their locations & number per borough





# Methodology cont'd

- From analysis: most high performing schools come from Manhattan & Queens
- Made new dataframe to hold only schools from these two boroughs

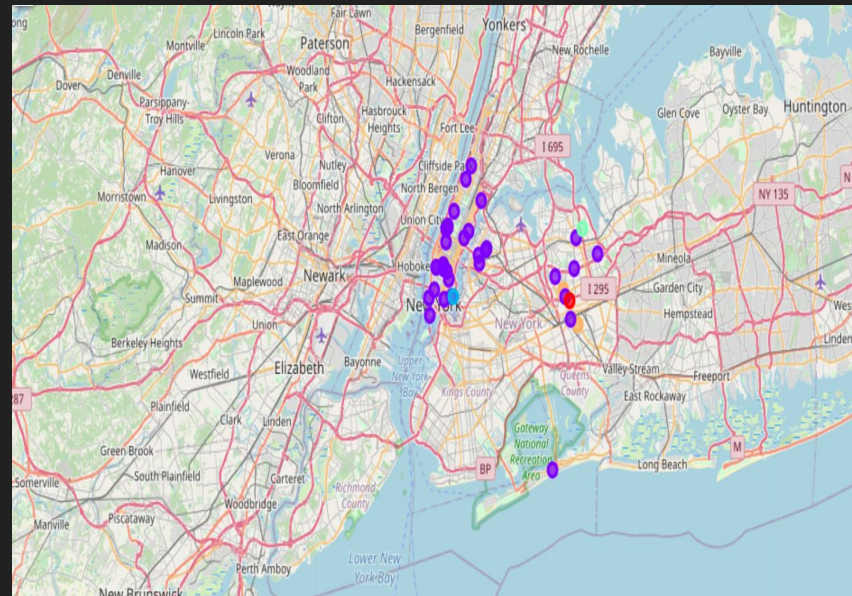
	School ID	School Name	Borough	Building Code	Street Address	City	State	Zip Code	Latitude	Longitude	...	Student Enrollment	Percent White	Percent Black	Pe His
0	01M539	New Explorations into Science, Technology and ...	Manhattan	M022	111 Columbia Street	Manhattan	NY	10002	40.71873	-73.97943	...	1735.0	28.6%	13.3%	1
1	02M545	High School for Dual Language and Asian Studies	Manhattan	M445	350 Grand Street	Manhattan	NY	10002	40.71687	-73.98953	...	416.0	1.7%	3.1%	
2	01M696	Bard High School Early College	Manhattan	M097	525 East Houston Street	Manhattan	NY	10002	40.71896	-73.97607	...	545.0	45.3%	17.2%	1
3	02M407	Institute for Collaborative Education	Manhattan	M475	345 East 15th Street	Manhattan	NY	10003	40.73249	-73.98305	...	482.0	56.5%	14.1%	1
4	02M418	Millennium High School	Manhattan	M824	75 Broad Street	Manhattan	NY	10004	40.70492	-74.01151	...	659.0	32.8%	7.6%	1



## Methodology cont'd

- Generated nearby venues using Foursquare API & clustered the schools using K Means

	School Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Academy of American Studies	Hotel	Coffee Shop	Bar	Pizza Place	Deli / Bodega	Café	Donut Shop	Cocktail Bar	Italian Restaurant	Mexican Restaurant
1	Baccalaureate School for Global Education	Mexican Restaurant	Bar	Gym	Discount Store	Athletics & Sports	Italian Restaurant	Seafood Restaurant	Bakery	Restaurant	Café
2	Bard High School Early College	Track	Park	Tennis Court	Baseball Field	Nightclub	Yoga Studio	Electronics Store	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant
3	Bard High School Early College Queens	Café	Coffee Shop	Bakery	Halal Restaurant	Deli / Bodega	Donut Shop	Juice Bar	Tennis Court	Cafeteria	Cocktail Bar
4	Baruch College Campus High School	Coffee Shop	Hotel	Yoga Studio	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Café	Hotel Bar	Japanese Restaurant	Grocery Store	American Restaurant
	Bayside High School	Farmers Market	Gym / Pool	Empanada Restaurant	Bar	Donut Shop	Café	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	English Restaurant



# Results

- Borough with highest average SAT score: [Staten Island](#)
- borough(s) with most number of schools with average SAT's above national average: [Manhattan & Queens](#)
- Recurring venues surrounding such schools: [coffee shops & restaurants](#) (Indian, Chinese, American, Italian, etc.)

# Discussion

- Analysis from one database primarily based on SAT scores of NYC high schools
- Recommendations:
  - Parents should choose either Manhattan or Queens: most number of high performing schools
  - Staten Island has highest average, but results skewed due to few schools
  - Need to incorporate other metrics and use more data in further analysis to improve

# Conclusion

- Used only one database of NYC high schools: improve using more data
- Used only one metric of high school performance (SAT): improve by using & incorporate other metrics
- Solutions presented may be skewed & not completely accurate
- Further data and more metrics needed to improve accuracy and reliability of solutions