

Locating A Suitable Place of A Student in NYC

**Capstone Project - The Battle of the Neighborhoods
Applied Data Science Capstone by IBM/Coursera**

**Vergara, Enrico Marco
December 23, 2019**



Content

01

Introduction

- A. Background Information
- B. Business Problem
- C. Target Interest

02

Data Definition

03

Methodology

- A. Exploratory Data Analysis
- B. Modelling

04

Cluster Analysis

05

Results and Discussion

06

Conclusion





INTRODUCTION

Scholastic Assessment Test (SAT)



A standardized examination usually taken by high school juniors and seniors used for college admissions in United States. This test is conducted in order to measure literacy, numeracy and writing skills which are three important components needed for academic success in college. In addition, SAT is said to be an assessment on how well the students analyze and solve problems which are skills learned in school that is vital in college. However, the test is administered under time constraint in order to help produce a range of scores.

With future as the biggest consequence of this assessment specially for incoming college students, it is important for parents to consider which accredited high school would help students achieve the needed skills to be learned as preparation for college. Preferably, parents want to know which public high school's students received the highest overall SAT score and which borough has the highest performing schools. This will serve as a guide for them in order to select the best school for their children in the hope of a bright future.

Business Problem

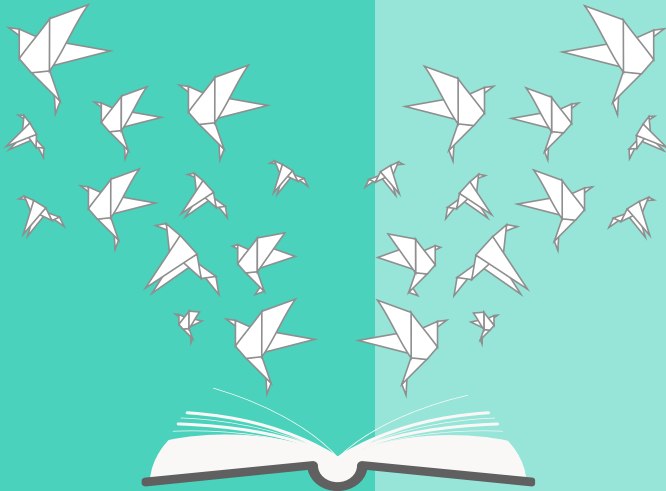
The average SAT scores for NYC public schools dataset from Kaggle provides a rich data compilation of average scores on each SAT test section for the 2014-2015 school year.

This project aims to:

- help parents select which area or borough in NYC is recommended for high performance on SAT
- explore the neighborhoods of that borough to find the most venues in each neighborhood that may affect the life of the students and cluster the neighborhoods using k-mean clustering.

Target Interest

The project is intended for parents who are considering which borough they are to select with the highest satisfactory performance on SAT. In addition, check the public high school in that area they can consider to be the next home of their children and what are the most common venues on that boroughs that will certainly affect the life of the students.



Business Problem

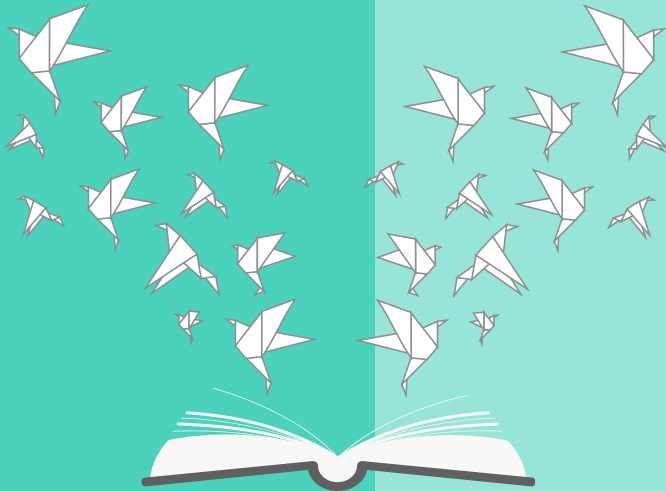
The average SAT scores for NYC public schools dataset from Kaggle provides a rich data compilation of average scores on each SAT test section for the 2014-2015 school year.

This project aims to:

- help parents select which area or borough in NYC is recommended for high performance on SAT
- explore the neighborhoods of that borough to find the most venues in each neighborhood that may affect the life of the students and cluster the neighborhoods using k-mean clustering.

Target Interest

The project is intended for parents who are considering which borough they are to select with the highest satisfactory performance on SAT. In addition, check the public high school in that area they can consider to be the next home of their children and what are the most common venues on that boroughs that will certainly affect the life of the students.

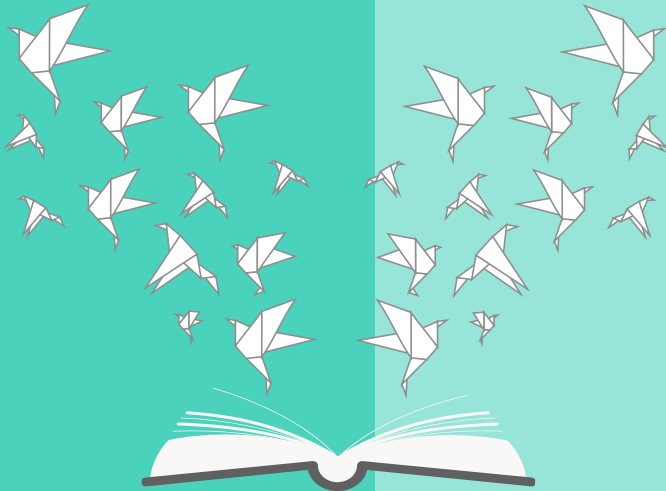




DATA DEFINITION

The dataset consists of a row for every accredited high school in New York City with the following attributes:

- School ID: The unique id of the school
- School Name: The name of the school
- Borough: Common name for NYC borough
- Building Code: The unique identifier of the building
- Street Address: The specific address
- City: The city where school is located
- State: NYC
- Zip Code: The address zipcode
- Latitude: The latitude of the location
- Longitude: The longitude of the location
- Phone Number: The school's phone number
- Start Time: Start of classes
- End Time: End of classes
- Student Enrollment: No. of enrollment for the school year
- Percent White: Percent of white from the total population
- Percent Black: Percent of black from the total population
- Percent Hispanic: Percent of hispanic from the total population
- Percent Asian: Percent of asian from the total population
- Average Score (SAT Math): Average score in Math
- Average Score (SAT Reading): Average score in Reading
- Average Score (SAT Writing): Average score in Writing
- Percent Tested: Percent of the sample tested from the whole population





METHODOLOGY

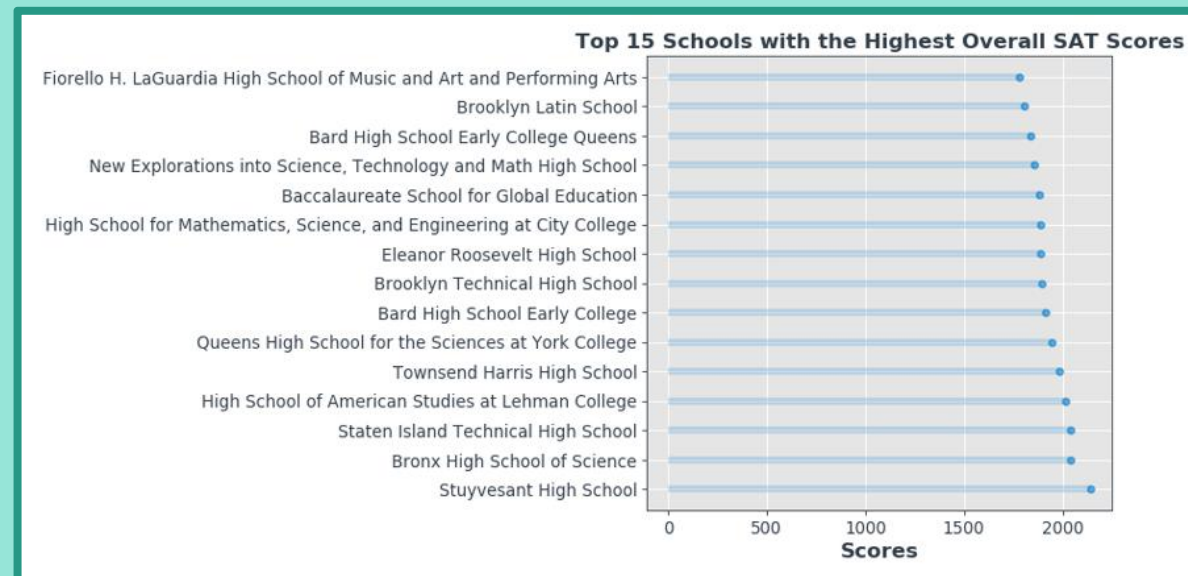
EXPLORATORY DATA ANALYSIS

	Ave_SAT_Score_Math	Ave_SAT_Score_Reading	Ave_SAT_Score_Writing	Ave_SAT_Score_Overall
count	375.000000	375.000000	375.000000	375.000000
mean	432.944000	424.504000	418.458667	1275.906667
std	71.952373	61.881069	64.548599	194.906283
min	317.000000	302.000000	284.000000	924.000000
25%	386.000000	386.000000	382.000000	1157.000000
50%	415.000000	413.000000	403.000000	1226.000000
75%	458.500000	445.000000	437.500000	1330.500000
max	754.000000	697.000000	693.000000	2144.000000

SAT has three sections (Math, Reading, and Writing) worth 800 points each with a maximum total score of 2400.

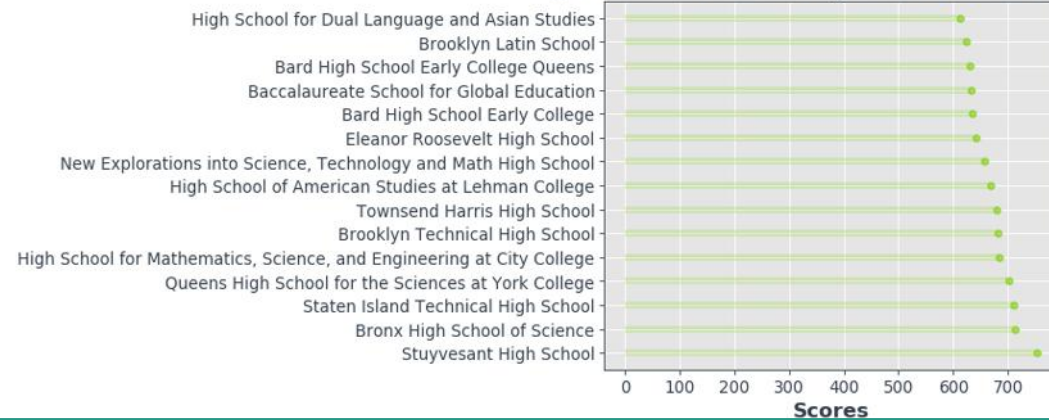
In the dataset used, the mean of public school SAT score is 1275 and the median is 1226, meaning the distribution is skewed slightly to the right with more high-scoring outliers.

Significantly, the highest average score is attained in Math and the lowest average score is Writing. All scores on each section is skewed slightly to the right with more high-scoring outliers.



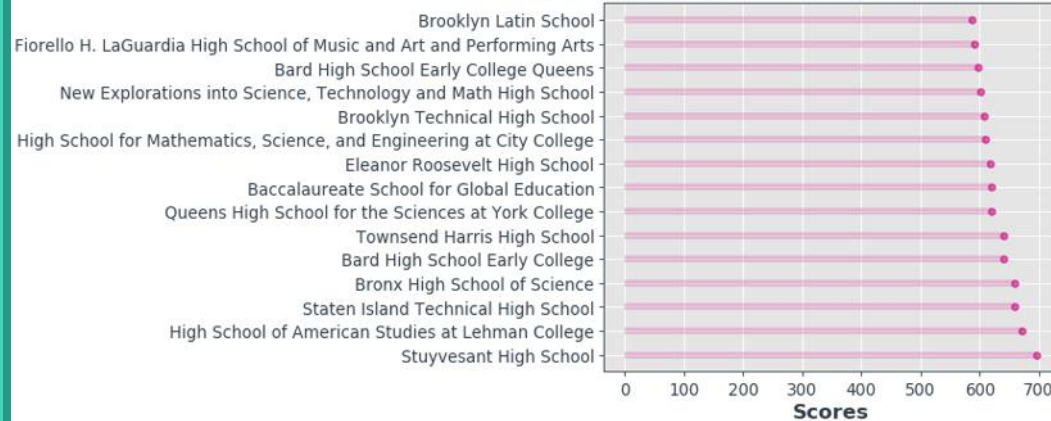
Overall, the top three performing schools are: Stuyvesant High School, Bronx High School of Science and Staten Island Technical High School.

Top 15 Schools with the Highest SAT Scores on Math



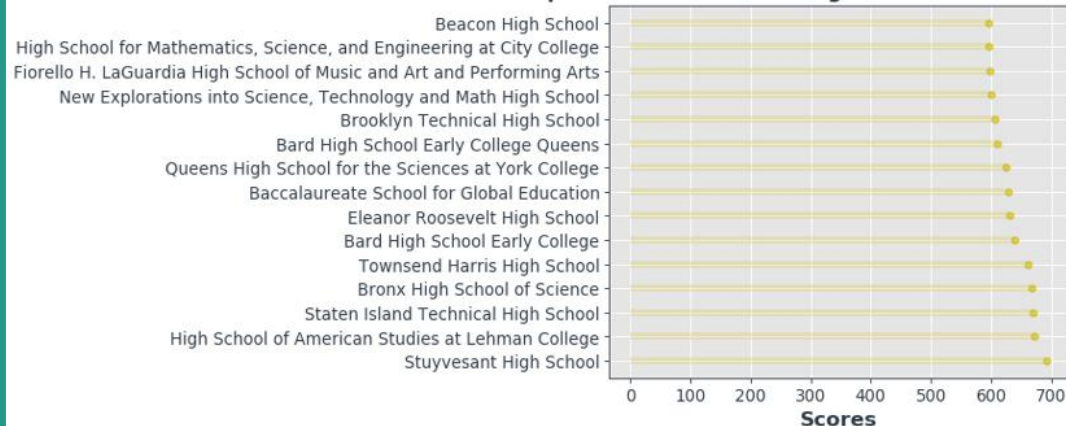
For the Math Section, the top 3 public high school with the highest scores are still: Stuyvesant High School, Bronx High School of Science and Staten Island Technical High School.

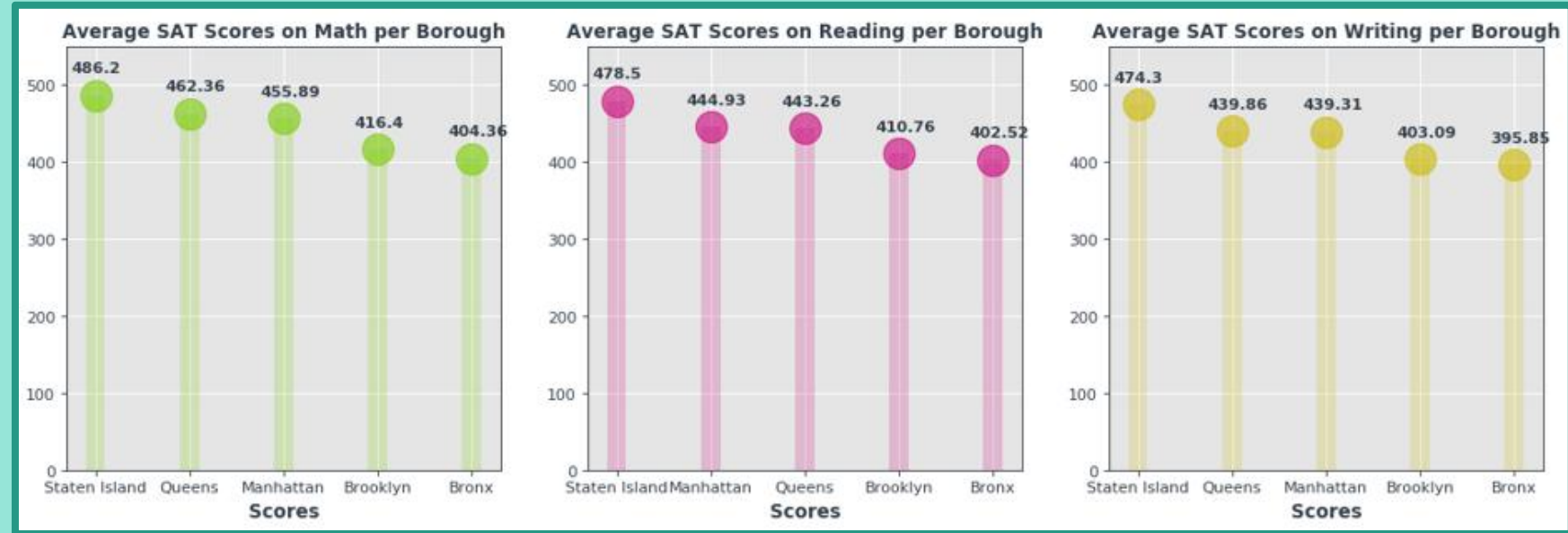
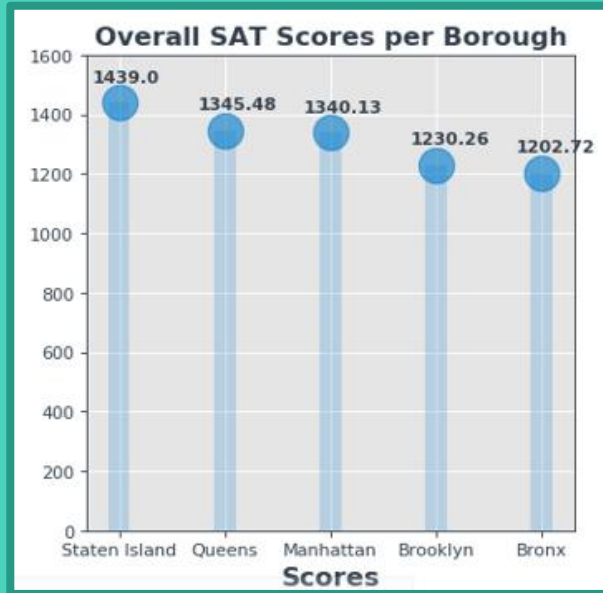
Top 15 Schools with the Highest SAT Scores on Reading



On the other hand, High School of American Studies at Lehman College claimed the top 2 spot for Reading and Writing Section while Stuyvesant High School remain at the top spot. Bronx High School of Science and Staten Island Technical High School claimed the top 3 and 4 spot for the said section.

Top 15 Schools with the Highest SAT Scores on Writing

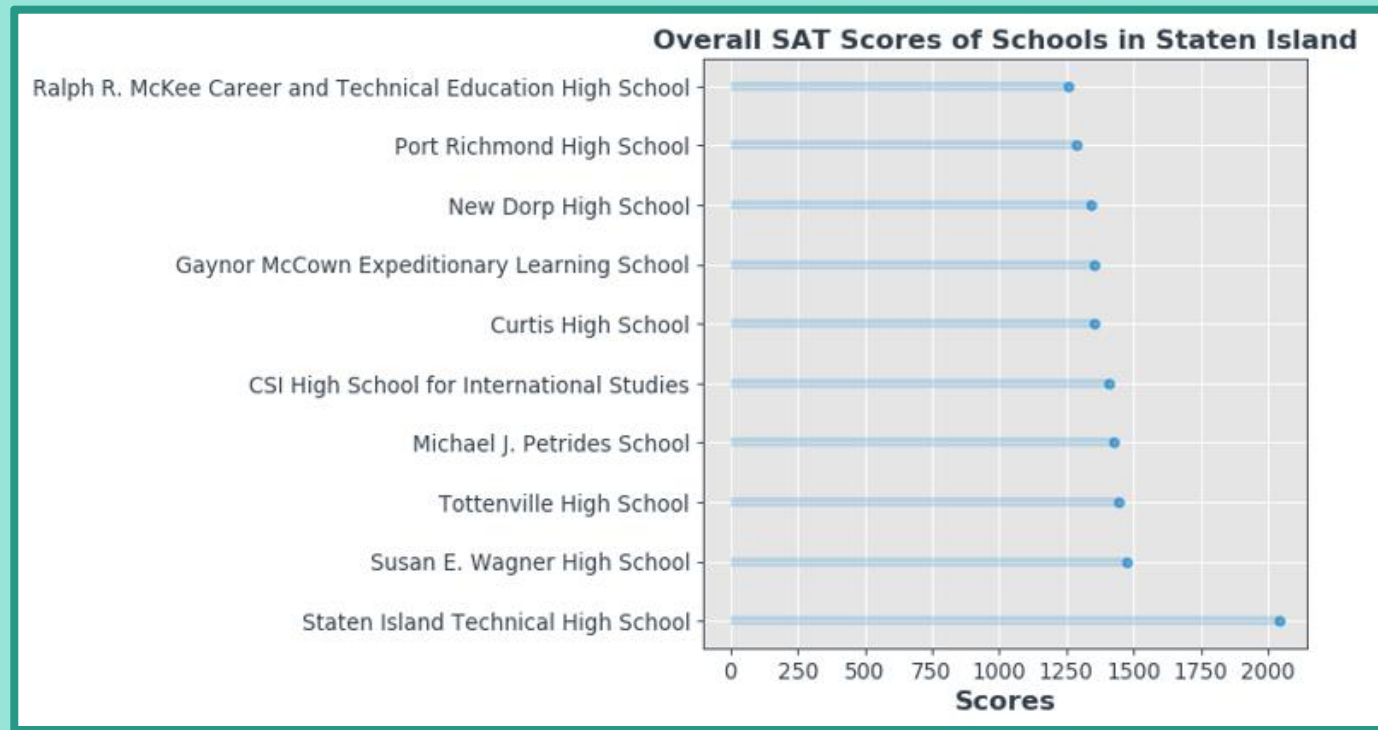




Even though the top school is located at Manhattan, in terms of average score per borough, it seems that public schools at Staten Islands have the consistent high performance. It is followed by Queens and Manhattan. On the other hand, Bronx seems to have the lowest scores in SAT.

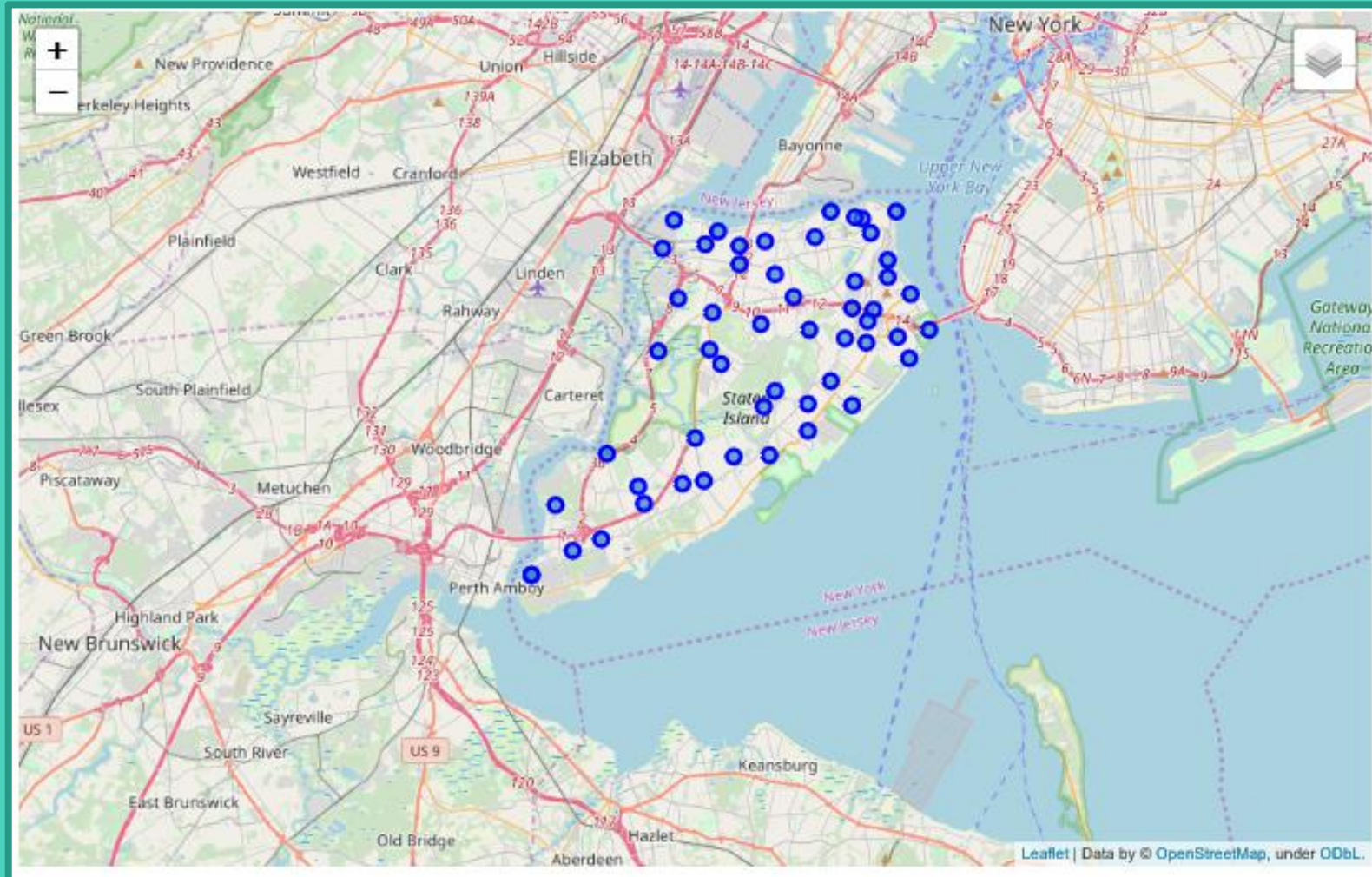
From the figure above, it is very consistent that public schools at Staten Islands have the highest scores across each section of SAT even though the top school is located at Manhattan. On the other hand, Bronx still has the lowest scores across each section of SAT.





The top performing school at Staten Island is Staten Island Technical School which is also part of the Top 3 Performing School in the whole New York City.





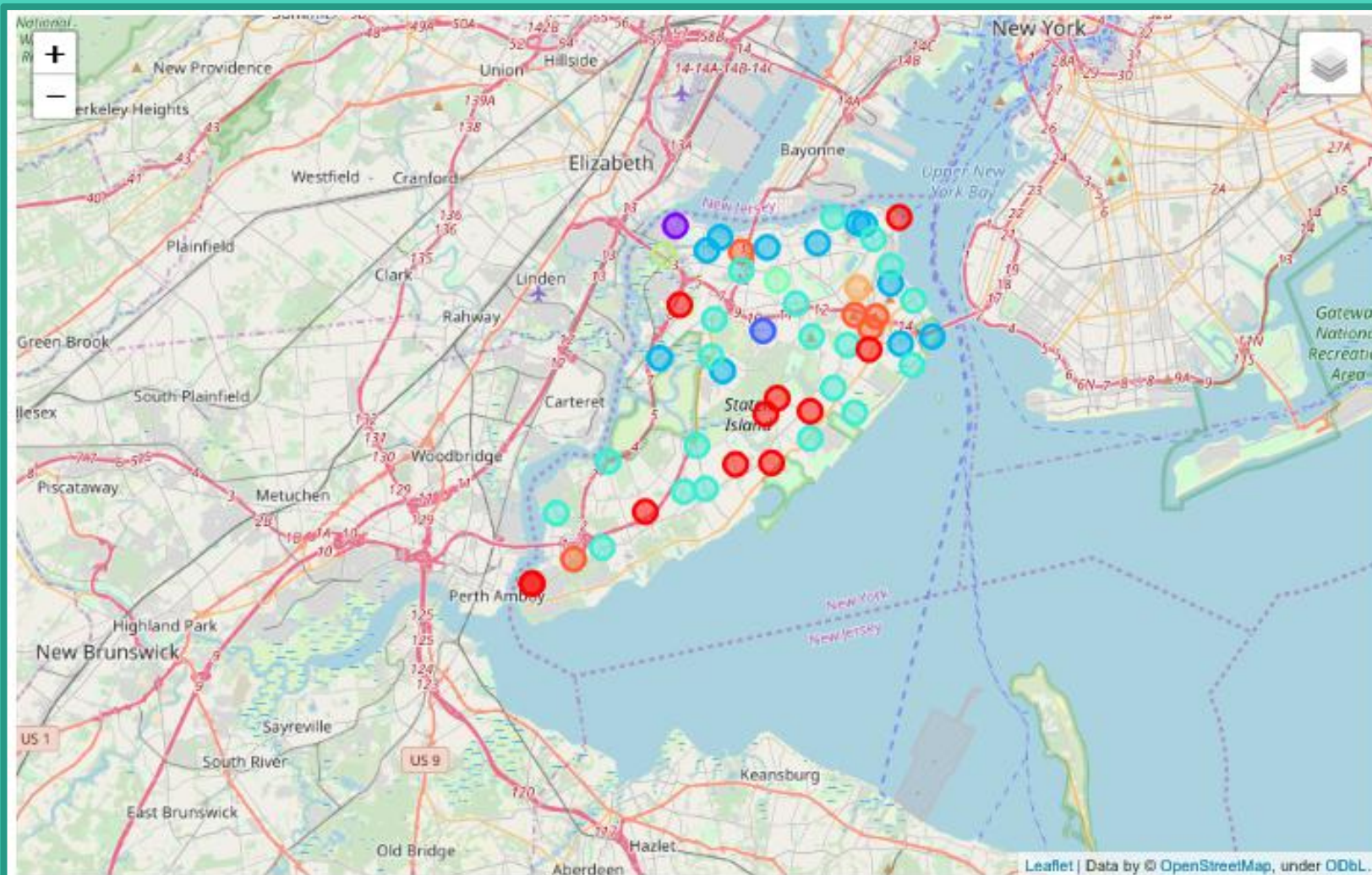
Map of the Neighborhood at Staten Island





METHODOLOGY

MODELLING (Clustering)



After applying K-Means Clustering, 9 clusters of neighborhoods were generated namely:

- Dining and Leisure Place
- Pragmatic District
- Food District
- Travel Friendly Place
- All Around District
- Technohub
- Recreational Place
- Physical Activity District
- Deli Place

Map of the Neighborhood at Staten Island with Clustering





CLUSTER ANALYSIS

The Nine Clusters

Dining and Leisure Place

The first cluster has the second to the highest number of Neighborhood. Most of the common venues are food related places such as food trucks, restaurants and pizza place. In addition, other common venues includes, playgrounds, bar and department store. Therefore, this cluster can be known as "Dining and Leisure Place".

Pragmatic District

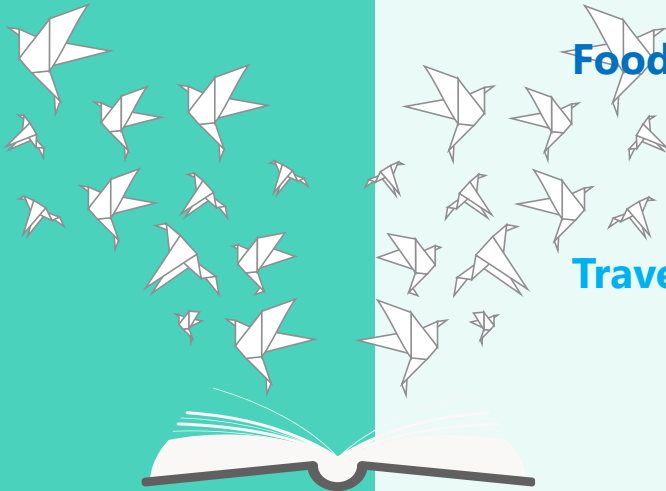
Most common venue for the second cluster is a locksmith place. Thus, it can be named as the "Pragmatic District".

Food District

Different restaurants, cafe and markets are the most common venues in the third cluster. Therefore, it can be known as the "Food District".

Travel Friendly Place

It is very apparent that the most common venue in the fourth cluster are bus stops and convenience stores. Thus, this cluster can be called the "Travel Friendly Place".



The Nine Clusters

All Around District

The fifth cluster has a good mixed of different venues such as food place, stores, fields for recreational activities, banks, parks and more. In addition it is the largest cluster of 19 neighborhoods. Therefore, it can be known as the "All Around District".

Technohub

The sixth cluster is mainly electronic stores and arcades. This cluster seems more advanced than the other cluster, making it a "Technohub".

Recreational Place

The seventh cluster is mainly for those who are seeking peace to reflect on place such as harbor and park. Thus, it can be called the "Recreational Place".

Physical Activity District

The second to the last cluster is mainly for those who are looking for physical activities for it is comprised of dance studios, trails, golf courses and gyms. We call it the "Physical Activity District".

Deli Place

The last cluster is mainly composed of deli/bodega venues. Therefore, it is the "Deli Place".





RESULTS AND DISCUSSION



The aim of this project is to help parents identify which borough in NYC is ideal for their children to be their next home for a satisfactory performance in SAT. From the data that was studied, although the **top performing school for overall and across NYC was Stuyvesant High School located at Manhattan**, when it comes to the performance per borough, it seems that schools at Staten Island are more consistent in achieving satisfactory performance on SAT both overall and across sections (Math, Reading and Writing). In addition **Staten Island Technical High School, which also is included in the Top 3 Schools with high performance in SAT is located in this area.** Thus, we focused more on studying the neighborhood of this borough.

The most ideal place for students is the **All Around District** for it is comprised of good mixed of different venues such as food place, stores, fields for recreational activities, banks, parks and more. Other options would be, the **Technohub** for those who are seeking advancement and technology lovers or **Physical Activity District** for parents who want their children to be physically active or learn dancing and sports.



CONCLUSION



The project provided a means for parents to select which borough in NYC has more consistent satisfactory score in SAT and is most suitable to shelter their children for a brighter future ahead of them since SAT is a tool for college applications. It also helps in finding a comfortable neighborhood for a student depending on the most common venues located in the place which will greatly affect the students' life. In the future, this project can **consider other factors** such as **cost of living** in the areas, **demographics** and **safety** on the location for finding the most suitable area to live for a student.