

Final Technical Report

By Priyanka Balineni

Abstract:

There are many passionate and talented kids who are willing to achieve great things in life but have very limited resources because of their financial background. So, such students deserve to have an opportunity. A non-profit organization like PASSNYC is dedicated to focus on these students to broaden their educational opportunities in New York city as it is the city of most impressive educational institutions in the world. PASSNYC uses public data to identify such students within New York city. It advises students to take Specialized High School Admission Tests (SHSAT), which helps them to decide underperformed students and can help them.

Objectives:

The goal of this project is to analyze which schools in New York has such students and also which part of test are student facing difficulty. This helps PASSNYC to concentrate on the needs of students by using the data and prepare them for the challenges facing in taking SHSAT. The dataset is collected from Kaggle and below is the link to the dataset: <https://www.kaggle.com/passnyc/data-science-for-good>

Functional requirements:

- Number of schools by city
- To find the racial distribution among schools in New York cities.
- To find the schools by low income estimate
- Check any correlation between the city and income estimate
- Find out correlation between city and economic index
- Correlation between student's attendance and their grades
- Correlation between grades and their family, collaborative teaching, environment
- Finding what are key factors of student achievement rating
- Deep Analysis of student grade to find out their weaknesses.

System architecture and description:

The architecture of the project includes six modules that help to achieve the objective of the project:

- Data collection: The data is collected from Kaggle website. PASSNYC uses public data to identify such students in New York city.
- Data cleaning: Checked for null values, performed manipulations such as changing the type of attributes. Data cleaning is performed in python and tableau.
- Data source: Upload the cleaned dataset in goggle API which can be read using JavaScript to perform the analysis further.
- Code: The project will be displayed through a webpage which is coded in HTML, styled by CSS and plots are presented by JavaScript using a package d3.js using Notepad++ editor.

Developed platforms:

The project is developed using tools:

- HTML
- CSS
- JavaScript(d3.js)
- Python (pandas, plotly, matplotlib, seaborn)
- Tableau dashboards.

Proposed Visualizations:

Visualizations can the explain the analysis, so it's important to choose according to each analysis. As this is about New York city distribution Geo maps would explain better, carto is used to produce the geo maps as it has better features and explains better the distribution clearly. A 3D scatter plot is done using a package in python called plotly which clearly describes the distribution of Economic index among schools. To analyze the data further, Bar charts, Bubble charts, Box plots, Pie charts are used for comparing and are created using JavaScript.

Experimental Analysis and Conclusions:

By analysis, the data revealed that Brooklyn has highest number of schools followed by Bronx, Queens, New York and Staten Island. The next step of analysis is finding the racist distribution among schools and resulted the following:

- Hispanics and Blacks are majority of school population followed by Asians and Whites.
- Correlation exist between the school's economic index and racial distribution and resulted that more black students have the higher economic index rate.
- Large Asian/ white students have more school income estimate.

Further analysis is done to check the student's attendance rate and found that absence rate is more in upper part of New York. Students with low attendance has less grade and also the family economic status is affecting student's absence rate. Found that low economic index has more absence rate.

Further analysis has done by checking the variation of grades and found that Asian students has better math grades than other students for all the grades. While ELA scores are better for white people and Asian people which is varying by grades. Students in upper part of New York has comparatively less scores than other students followed by Central part.

The findings conclude that Asians has good performance in math, and black/Hispanic has better ELA grades. Students in upper and central part of New York schools need more support.

References:

- <http://www.passnyc.org/shsat/>
- <https://www.kaggle.com/passnyc>