Capstone Project 1 : Project Proposal

**What is the problem you want to solve?**

PASSNYC and its partners provide outreach services that improve the chances of students taking the SHSAT and receiving placements in these specialized high schools. The current process of identifying schools is effective, but PASSNYC could have an even greater impact with a more informed, granular approach to quantifying the potential for outreach at a given school. Proxies that have been good indicators of these types of schools include data on English Language Learners, Students with Disabilities, Students on Free/Reduced Lunch, and Students with Temporary Housing.

Part of this challenge is to assess the needs of students by using publicly available data to quantify the challenges they face in taking the SHSAT. The best solutions will enable PASSNYC to identify the schools where minority and underserved students stand to gain the most from services like after school programs, test preparation, mentoring, or resources for parents.

**Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn’t have done otherwise?**

PASSNYC is a not-for-profit organization that facilitates a collective impact that is dedicated to broadening educational opportunities for New York City's talented and underserved students. New York City is home to some of the most impressive educational institutions in the world, yet in recent years, the City’s specialized high schools - institutions with historically transformative impact on student outcomes - have seen a shift toward more homogeneous student body demographics.

PASSNYC uses public data to identify students within New York City’s under-performing school districts and, through consulting and collaboration with partners, aims to increase the diversity of students taking the Specialized High School Admissions Test (SHSAT). By focusing efforts in under-performing areas that are historically underrepresented in SHSAT registration, we will help pave the path to specialized high schools for a more diverse group of students.

PassNYC is conducting a Datascience for Good challenge in collaboration with Kaggle

**What data are you using? How will you acquire the data?**

PASSNYC has shared 2016 School Explorer data which has all the demographic and other relevant data abt the NYC schools , contained in 1272 Rows and x 161columns

In addition to that , it has also shared District 5 ‘s SHSAT Registrations and Testers details in 140\*7 R \* C

In addition to the above I am also using publicly available data from City of New York’spublicly available data at NYC data , which has SHSAT offer details for the alst three academic years

**Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

Data Wrangling:  
Import the file to the database as a pandas DataFrame  
Clean the data using different manipulating techniques to tidy and rearrange the data  
Change column names as applicable.  
  
  
Data Visualization:  
Use Seaborn /Matplotlibto visualize the data with different kinds of plot  
Plot ENI for Black and Hispanic population  
Plot Black and Hispanic population distribution across schools  
Plot the other feature variables to see if there is an trend

Model:

I can use either a regression or a classification algorithm here. Need to do further analysis on this

Linear Regression  
Logistic Regression   
Decision Tree  
Random Forest

**What are your deliverables? Typically, this includes code, a paper, or a slide deck.**

The deliverables for this project would be a complete presentation of the project with all the documentations, visualizations and python codes on notebook. All the model created to predict the schools that qualify for PASSNYC’s training as well schools that qualify for PASNYC’s awareness creation sessions focus. The final submission will include the introduction, the steps and methods involved in this project from the beginning to the end of the project.