Analysis of:

New York City Department of Education

Pre-Kindergarten School Quality Report 2017-2018

August 2020

Introduction

Numerous studies have shown that children who attend pre-kindergarten (Pre-K) programs are better prepared for kindergarten. The New York City Department of Education ("NYCDOE") performs and publishes School Quality Reports for Pre-K programs. According to the NYCDOE, it evaluates how well schools are organized to support student learning and teaching. It was developed to assist NYCDOE schools in raising student achievement by looking behind a school's performance statistics to ensure that the school is engaged in effective methods of accelerating student learning. The NYCDOE uses the Classroom Assessment Scoring System (CLASS) and the Early Childhood Evaluation Rating Scale (ECERS) scales to assess its Pre-K programs.

Most of the data in the NYDOE data is obtained from surveys of parents and teachers. I was interested in knowing whether the parents' participation in the survey was correlated to the school's enrollment and whether I could use a linear model to forecast the parents' participation rate. I calculated the ratio of parent surveys received to the enrollment of the program and found that the average and median were around 51%. I also clustered the schools based on the ECER data and created a map of the school locations.

Finding a nearby quality pre-kindergarten program is a priority for many parents and this information should be useful in determining what schools parents and teachers view as the best, the similarities among the ECER data and the location of the schools.

The NYCDOE is continually trying to improve their performance and may find the results of this analysis interesting.

Data

The 2017-2018 School Quality Report - Pre-K Programs was published on July 21, 2020 and includes data on over 1,600 Pre-K programs. The data includes 45 columns. Most of the data are based on survey results of parents and teachers that are categorized into the following metrics:

- School information such as the school name, location and enrollment
- Results of site visits
- CLASS Survey Results and
- ECER Survey Results

The parent survey asks parents to rate the schools on communications, school decision making and questions such as:

- I feel good about the way that my child's teacher helped my child adjust to pre-K.
- My child's teacher gives me helpful ideas about how I can support my child's learning.
- My child's teacher lets me know that I can make a difference in my child's learning.
- My child's teacher gives me opportunities to share what I know about my child.
- Someone at my child's pre-K or 3-K program has helped me consider which schools or programs would be best for my child for next school year.

The teacher's survey looks at whether teachers participate in opportunities to develop, grow, and contribute to the continuous improvement of the school community.

The data does not include the responses to individual questions.

The NYCDOE provides an API for this data, which was used for this project.

Methodology

Being that I believe parents want the best education for their children, I focused on schools that had both a parent and teacher survey rating greater than ninety. This reduced the number of schools in the data to one hundred.

I used Sci-kit learn linear model for forecasting the parents' scores, the standard scaler and K-means for clustering the schools. Finally, I used geopy geocoders and Nominatim for the mapping of the schools.

Results

There was strong correlation (.86) between the parents survey scores and enrollment data. The linear model to determine whether the enrollment could predict the parents scores produced the following measure of accuracy:

Mean absolute error: 4.28
Residual sum of squares (MSE): 30.54
R2-score: 0.58

The above statistics means that the was an average difference of four between the predicted scores and the observed scores. As indicated by the Residual sum of squares value, the model could not explain 30% of the data. The R2 score is similar in that it is an a percentage of parents' survey responses that are explained by the enrollment.

My understanding is that it is not unusual for the R2 for data that describes human behavior to be approximately .50 so I am willing to accept this model.

Discussion

My initial though on using this data set was to identify the availability of Pre-K programs to distinguish full day from half day and those programs with early drop off and late pick up options so that parents can determine which school programs would work best with their schedules. After exploring the data, I discovered that almost every NYC Pre-K programs offer these options so I decided to look at the survey data as the parents input is such a critical component of these ratings. My thinking is that if I was living in NYC, I would want my child to attend the best Pre-K program. I decided to use the survey results from the parents and

teachers to determine the best schools because this data set looks beyond the academic performance of the students and weighs heavily on the results.

There was one school that I considered to be an outlier and was removed from the data set. This school received 98 parent responses and has a total enrollment of 108 students. The rest of the sample had a mean of 51 and a median of 57 parent responses.

I clustered the data into 3 clusters using the Kmeans algorithm. Many of the scores provided in the surveys are on a scale from one to four. I graphed the results based on the parents and teachers survey results and the enrollment data. The resulting clusters had clear boundaries between parent result scores above .6 and between .4 and .6. The third cluster had the most variation among the parents and teacher survey results and the enrollment.

The mapping of the top schools revealed significant geographic gaps. For example, only three schools in Manhattan were in the top schools and all of these are located in downtown Manhattan. Eastern Brooklyn and Queens were also lacking in highly rated schools.

Conclusion

With an average participation rate of 51%, the NYCDOE might want to consider how it communicates the importance of the parent survey and the methods that it collects the data to improve the surveys.

In this project, I analyzed the NYDOE Pre-K data for the 2017-2018 school year and analyzed the schools that received above a 90 on surveys from teachers and parents. I explored the relationship between these score results and the number of students enrolled in the program and found that while there is a strong correlation among the data, a linear model for predicting the parents scores could not account for 30% of the data.

I believe there are economic and social factors that influence the parent response rate, which may account for large geographic areas not having highly rated schools as is evident on the map of the top schools. This is a topic that could be investigated further.

Parents should explore the data provided by the NYCDOE before selecting a public Pre-K program and determine if it is suitable for their children.