DATA 608: Knowledge and Visual Analytics

Final Project

Michael ODonnell

**Project Overview**

**Title**: NYS Educational Success vs Teachers’ Salary by School District

**Aim**: explore New York State educational data to determine if a relationship exists between School Districts’ student success and Teachers’ Salary in New York State.

**Hypothesis**: there will not be a direct correlation between student success and teachers’ salary. However, there will be a visible increase in student success above a specific teachers’ salary (example, student’s test scores discernably increase for school districts with median teachers’ salaries >= $90,000).

**Link to public visualization**: <https://nys-student-success.herokuapp.com/>

**Data Sources:**

1. New York State Education Department datasets:
   1. <https://data.nysed.gov/downloads.php>
   2. Source: New York State Education Department, (governmental data)
2. New York State teacher salaries
   1. <https://www.seethroughny.net/teacher_pay>
   2. Source: Empire Center (non-profit)

**Data Parameters:**

1. Location: New York State public school districts only
2. Time Frame: all statistics and numbers are based on the 2018-2019 school year
3. Outlier handling: removed all New York State public school districts that did not graduate at least one student in 2019
4. Data points: each data point (in scatter plot) is one New York State school district
5. Scatter Plot x-axis: The x-axis of the scatter plot is the Median Teachers’ Salary by School District (independent variable). This remains constant regardless of user interaction.
6. Scatter Plot y-axis: The y-axis of the scatter plot is the measure of student success (dependent variable), which the user chooses in the above dropdown.

**Data Analysis Process:**

1. Download first dataset from New York State Education Department datasets
   1. Go to <https://data.nysed.gov/downloads.php>
   2. Download “2018-2019 3-8 Assessment database”
   3. Filter in excel for only district-level data, not more granular than all-students at the district-level
   4. Load database into python: <https://github.com/odonnell31/NYS_School_District_Success/blob/master/data_analysis/data_analysis_v1.py>
2. Download second dataset from New York State Education Department datasets
   1. Go to <https://data.nysed.gov/downloads.php>
   2. Download “2018-2019” Graduation Rate Database”
   3. Filter in excel for only all-students, district-level data
   4. Load the database into python: <https://github.com/odonnell31/NYS_School_District_Success/blob/master/data_analysis/data_analysis_v1.py>
3. Download Teachers Pay by School District information from See Through NY
   1. Go to <https://www.seethroughny.net/teacher_pay>
   2. Copy the full dataset into excel
   3. Load data into python: <https://github.com/odonnell31/NYS_School_District_Success/blob/master/data_analysis/data_analysis_v1.py>
4. Manipulate the first dataset loaded as a pandas dataframe (titled “nys\_3through8\_data”) to be more horizontal than vertical (unstack)
   1. Code in line 27-30 of: <https://github.com/odonnell31/NYS_School_District_Success/blob/master/data_analysis/data_analysis_v1.py>
5. Merge the three pandas dataframes together into one final dataframe, and export to excel with xlsxwriter: line 35-50 of <https://github.com/odonnell31/NYS_Educational_Success/blob/master/data_analysis_v1.py>
6. Remove any school district that did not graduate a student in 2019
7. Create Heroku account: [www.heroku.com](http://www.heroku.com)
8. Create dash app in app.py file: <https://github.com/odonnell31/NYS_School_District_Success/blob/master/app.py>

**Write-up on Visualization:**

My motivation for this visualization was a personal past experience. I left high school teaching due to low salaries. I wanted to teach in a school district that was not paying teachers a competitive salary, and for that reason there was not much competition for the job. Since that experience, I was curious if teachers’ salaries effected students’ success. Thus, I built this project to compare those variables and used New York State as a case study.

Hence, this project is relevant to two social movements: One, push government to increase teachers’ salaries. Two, push government to provide equal educational opportunities to all students.

From this project, I concluded that as students get older (high school vs elementary school), their success is impacted by teachers’ salaries. But, this visualization is exploratory so the user can come to their own conclusions.

To create this project, I build a Dash app with python and deployed it on Heroku. Personally, this was a huge victory. For over a year I have wanted to deploy a public, interactive visualization and this project gave me the focus and opportunity! Furthermore, I think dash was the appropriate vehicle for this visualization because it allowed me to cleanly display a large amount of information that can be easily ingested by the user.