Liliane Effoudou

Data 205- Capstone Project

CRN: 34669

**Property Values and Educational Performance: A Case Study in Montgomery County**

**Introduction:** This project aims to analyze whether there is a measurable relationship between local property values and student academic performance in Montgomery County. By mapping and comparing school-level academic outcomes with corresponding property value data, the project seeks to uncover spatial trends without considering demographic variables.

**Datasets Used:**

* [List of Montgomery County Public Schools](https://data.montgomeryschoolsmd.org/Schools/FY2013-MSDE-School-Participation-Indicators-All-Sc/ajmx-3i9d/data_preview)
  + Taken from [Data Montgomery](https://data.montgomerycountymd.gov), this dataset contains the list of all the public schools located in Montgomery County. It also includes the performance ratings of all those schools such as their individual attendance, graduation, and drop-out rates. In addition to that, it organizes the schools by their respective neighborhood clusters.
* [Median Home Value in Montgomery County](https://api.census.gov/data/2022/acs/acs5)
  + This is an API taken from the [U.S. Census Bureau](https://www.census.gov/), mainly the American Consumer Survey, which analyzes the median home value for properties in Montgomery County. It offers an insight into how much property in different neighborhoods vary in terms of pricing.
* [Locations of Schools in Montgomery County](https://data.montgomerycountymd.gov/Education/Public-Schools/772q-4wm8/data_preview)
  + As for this dataset, that was once again taken from [Data Montgomery](https://data.montgomerycountymd.gov), it contains the locations of all the schools in the county, along with all their coordinates. It enabled me to use latitude and longitude to illustrate certain areas with a heatmap.

**Goals:**

* **Evaluating the Correlation Between Property Values and Academic Success**
  + Determining if higher property values are associated with better student performance (e.g., graduation rates, college admissions rate) using only publicly available data.
* **Identifying Spatial Patterns in Education Outcomes**
  + Using geographic data like school clusters to visualize whether regions with higher property values show consistently higher academic results.
* **Compare School-Level Academic Metrics Across Property Value Ranges**
  + Analyzing differences in performance between schools located in high-value vs. low-value property areas.

**Tools Used:**

* Python
  + All the analysis such as data cleaning, visualization, and modeling were performed with Python.
    - Packages used:
      * Pandas: used to read the datasets.
      * Seaborn and Matplotlib for visualization, analysis, and modeling.
      * Folium for heatmap display.
      * Squarify to visualize the clusters by order of value.

**Data Cleaning and Processing:**

* **List of School Dataset:** 
  + Cleaning:
    - This dataset included all the public schools in Montgomery County, along with elementary, middle, charter, and alternative schools. Those needed to be eliminated from the dataset, since only high schools were necessary for the analysis. Alternative and charter schools, although technically high schools, had to be omitted to prevent them from skewing the analysis, given that those schools also contained a lot of missing values.
  + Processing:
    - Schools were already organized by their respective clusters so that part was taken care of. After cleaning up the unwanted values, analysis could start. This dataset was merged with the median home value one to generate models and visuals, and the last process was combining this dataset with the one containing the locations of the schools to generate a heatmap.
* **Median Home Values API:**
  + Cleaning:
    - Removing suffixes such as ‘CDP’ and ‘Town’ to rename the cities to properly organize and group them into clusters.
    - There were some cities that had to be omitted since they were added by the Census Bureau because of their similar names to other cities in the County (Potomac Heights and Potomac Park which are in Washington DC were included as being part of Potomac, Maryland), and their proximity to the county itself (Glen Burnie, Baltimore City).
  + Processing
    - The high schools in Montgomery County are divided into clusters. The API from the census bureau did not organize the cities that way, so clusters had to be created and included all the cities in their respective clusters. Then the datasets needed to be combined to continue the analysis.
* **All Datasets:** 
  + The Median Home Value and List of Schools dataset had to be merged so that they could both be processed together. The Location dataset was only used to draw the coordinates of the schools.

**Descriptive Statistics:**

* **Median Home Value Statistics:**
  + The Median Home Value dataset consists of two columns divided into cities and their respective property values. The data is only numerical, so the descriptive statistics are straight-forward.

|  |  |
| --- | --- |
| **Count** | 35 |
| **Mean** | 820805.7 |
| **Standard Deviation** | 389852.0 |
| **Minimum** | 393700.0 |
| **First Quartile** | 535550.0 |
| **Second Quartile** | 626200.0 |
| **Third Quartile** | 1032400.0 |
| **Maximum** | 2000001.0 |

* **List of School Statistics:** 
  + The original dataset contains 200 rows and 11 columns, but most of them had to be cleaned out for the analysis. It ended up with 25 rows representing the high schools and nine relevant columns.



**Description of Final Data Product**

There are several factors to consider. Montgomery high school students tend to perform well overall academically, with some schools performing better than others. As for property value, there are some school clusters that are more expensive to live in than others. Academically, the analysis’s findings were close. For the overall academic performance, we have:

* Mean attendance rate: 94%
* Mean graduation rate: 88%
* Mean college acceptance rate: 93%
* Mean dropout rate: 7%

And when we compare some of those statistics by clusters, along with their respective median home values we get:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **% Attendance rate** | **% Graduation rate** | **% College acceptance rate** | **Mean of home value** |
| **Bethesda-CC** | 95 | 94 | 95 | $ 1,445,743.00 |
| **Clarksburg** | 95 | 89 | 93 | $ 604,000.00 |
| **Col. Zadok Magruder** | 95 | 89 | 91 | $ 558,000.00 |
| **Damascus** | 94 | 92 | 95 | $ 474,800.00 |
| **Downcounty Consortium** | 95 | 81 | 91 | $ 556,100.00 |
| **Gaithersburg** | 92 | 77 | 86 | $ 687,267.00 |
| **Northeast Consortium** | 94 | 86 | 94 | $ 525,667.00 |
| **Northwest** | 95 | 91 | 95 | $ 393,700.00 |
| **Poolesville** | 95 | 95 | 93 | $ 606,300.00 |
| **Rockville** | 95 | 86 | 93 | $ 559,800.00 |
| **Sherwood** | 95 | 93 | 95 | $ 615,700.00 |
| **Walt Whitman** | 95 | 95 | 95 | $ 921,767.00 |
| **Walter Johnson** | 95 | 95 | 94 | $ 909,300.00 |
| **Winston Churchill** | 95 | 95 | 95 | $ 907,450.00 |

As we can see, attendance rates are not dependent on the students’ neighborhoods’ values, neither are college acceptance rates. We can, however, see graduation rates being affected, if even by a little. To better understand and see it here is a visual:

A graph of a number of people

AI-generated content may be incorrect.

The most expensive clusters also boast the most successful students. However, there are other clusters that rival the highest four. The former’s consistency can lead us to deduce that there could be a correlation or at least a positive relationship between high graduation rates and expensive neighborhoods.

To determine whether a correlation exists between median home value and student’s success, a linear regression model was generated:

A graph with a red line

AI-generated content may be incorrect.

The results of this model found that the correlation between graduation rate and median home value is at 0.56, indicating that a positive relationship exists between the two, albeit a very weak one. To better test the strength of this correlation, an R-squared test was also performed and these were the results:

**R-squared**: 0.32

**Slope**:1.1

**Intercept**: 78.1

**P-value**: 4.316003746385565e-06 (less than zero)

**Standard Error**: 2.2

**P-value interpretation:** The null hypothesis of our analysis states that median home values do not affect graduation rate, while the alternative analysis tells us the opposite. Since the analysis found the p-value at lower than 0%, the null hypothesis is rejected in favor of the alternate one. There is an association between the two, if a weak one.

**R-squared interpretation:** the r-squared that measures the relationship between median home values and graduation rate falls at just a little over 31%. This figure means that overall median home value is affecting the graduation rate of students by as little as 31%. The other 69% has to do with different factors in the dataset.

**Other factors to consider**: the relationship between graduation rates, attendance rates and college acceptance rates. When graduation and attendance rates are compared, we have this analysis:

Correlation: 0.76

Slope: 5.4

Intercept: -423.5

R-squared: 0.58

P-value: 8.446200132738759e-06

As for graduation rates vis-vis college acceptance rate we get this:

Correlation: 0.69

Slope: 1.4

Intercept: -46.1

R-squared: 0.47

P-value: 0.00013296313206659566

Both of those comparisons give us a better explanation for the successful graduation rates in the county. As students’ attendance increases, graduation rates rise, as graduation rates rise, so do college acceptance rates.

In conclusion, median home value in Montgomery County do not account for its students’ success, but diligence and attendance do.

**References and Acknowledgements**

**References:**

* U.S. Census Bureau. (2022). *American Community Survey*. <https://api.census.gov/data/2022/acs/acs5>
* Data Montgomery. (2023). *Public Schools*. <https://data.montgomerycountymd.gov/Education/Public-Schools/772q-4wm8/data_preview>
* Data Montgomery. (2013). *School Participation Indicators*. [FY2013 MSDE School Participation Indicators - All Schools | MCPS - Open Data Portal](https://data.montgomeryschoolsmd.org/Schools/FY2013-MSDE-School-Participation-Indicators-All-Sc/ajmx-3i9d/data_preview)

**Acknowledgements:**  
I would like to thank Professor Lori Perine for her guidance, feedback, and support throughout this project. I would also like to thank all the professors who have helped me through my Data Science Certificate:

* Professor Amy Thantou
* Professor Rebin Muhammed
* Professor Abdirisak Mohammed

I also acknowledge the use of certain programs, including:

* **Python** for statistical analysis and data cleaning
* **Seaborn, Folium, Matplotlib** for data visualization
* **Excel** for initial data exploration