

## **Distribution of Private Schools: A Data Analysis using Python**

Sophia Weidner

Bellevue University

DSC 530: Data Exploration and Analysis

Professor Shankar Parajulee

August 12th, 2022

## **Distribution of Private Schools: A Data Analysis using Python**

The outcome of my EDA was not what I expected it to be. When originally looking at my dataset, I felt as if it would suffice for what I wanted to explore. I was more interested in seeing the effect of the pandemic on the amount of teachers or students, but because the data set was from 2019-2020, it was hard to justify any of those numbers being a result of COVID-19. Thus, I changed my hypothesis to see if the type of school had any influence on the amount of students or teachers.

I first started to run into challenges when cleaning my dataset and plotting the histograms. I realized that only having 5 “numbers” which were all either percentages or total amounts does not make for a great analysis. I struggled in finding what was significant to what I wanted to discuss and ended up transposing my data to see if that would help. That helped for plotting the histograms, however, my histograms did not show anything profound and rather just showed the amount of times certain percentages appeared (which was usually just once).

I wish I could have looked at a snapshot of my data and had more data points generated from that. For example, it would have been interesting to see a breakdown of Catholic Schools in each Region, and then the amount of teachers or students in those specific schools. The dataset was almost too generalized to draw any real conclusions from it. The conclusion I was able to draw was one that is pretty much common sense: more students need more teachers and more schools. That isn’t anything new, but I was able to show significant evidence for that when running my correlation and covariance analysis.

Overall, I faced a lot of challenges during this project, but I definitely learned what I could have done better or what data I should have looked for. The concepts we learned

throughout this course have helped me tremendously in analyzing data, which is how I know that the data I had was not the kind that would help me fully explore my hypothesis.