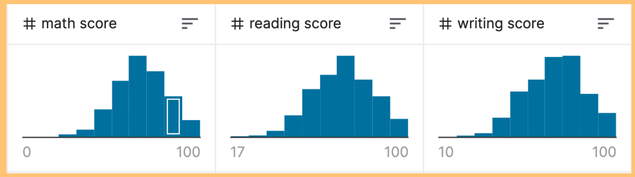
High Score Test Scores

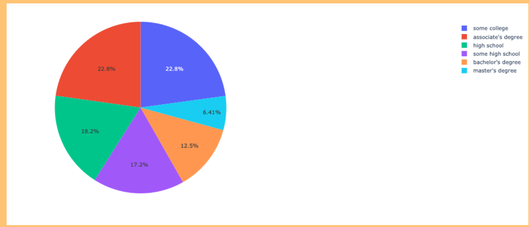
DH 100 Theory and Methods | Channing Lee | 5/30

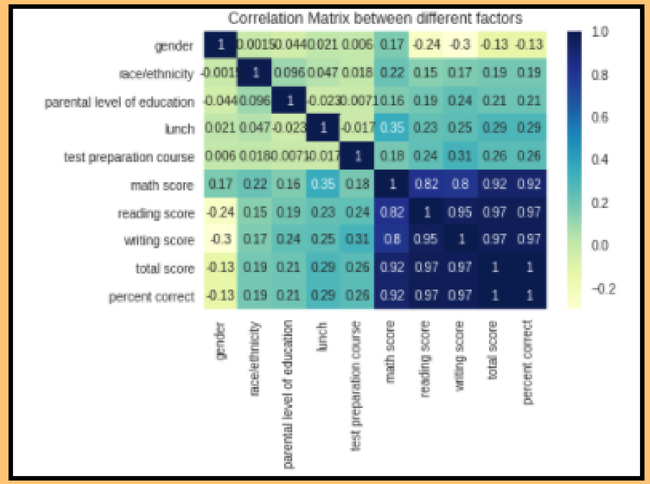
1) Dataset
https://www.kaggle.com/spscientist/students-performance-in-exams
(PDF, TXT, CSV)

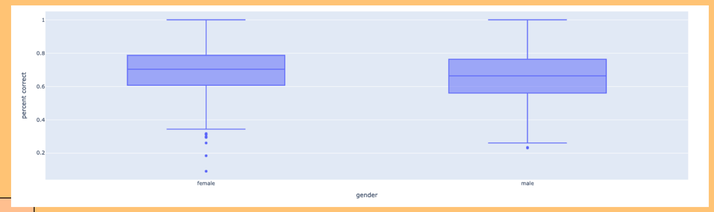
2) How does factors such as gender, race/ethnicity, parental level of eduction, and etc affect student test performance

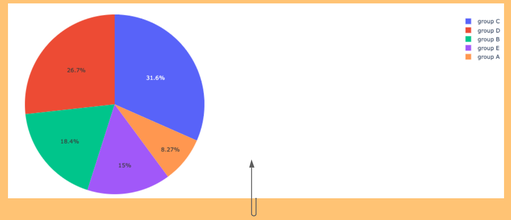
Introduction / "Hook"











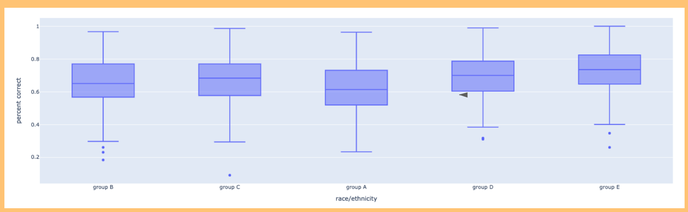
Kaggle Dataset

Google Drive

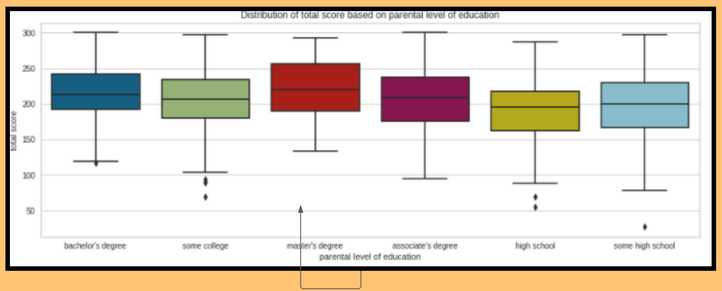
GitHub

Google Colab notebooks

Dataset Imported as CSV



1) High School Test Scores
2) DigHum 100 Theory & Methods in the Digital Humanities
3) Dr. Adam Anderson
4) Channing Lee

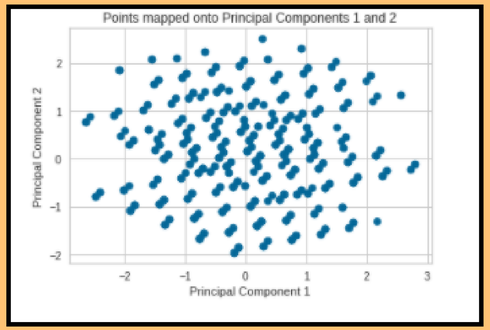


Descriptions:

This Dataset has 8 columns. It includes data on gender, race ethinicity, parent level of eduction, lunch, test preparations, and test score results for math, reading, and writing. I created additional columns such as total score and a percent correct.

I will be using Machine learning methods to figure out which factors affect the test score results the most. I will try methods such as linear regression, random forest, lasso regression, ridge regression, and principal components regression.

.



Discussion of results

Linear Regression:
Math Score: From the analysis we could see that the Parents level of eduction had the least amount of impact on Math scores while whether the students had lunch or not made the most amount of difference.
Reading Score: For reading score, gender had barely any impact at all while lunch and test preparation mattered the most
Writng Score: Gender had a negative impact on the scores and Once again Preparation had the most impact on Writing Performance.

Ridge Regression:
Math Score: Parental level of education had the least amount of impact while lunch had the most amount of impace
Reading Score: Gender had a negative impact on Reading and test preparation mattered the most
Writing Score: Gender also had a negative impact on Reading and Test preparation mattered the most.

Lasso Regression:
Math Score: Parent level of education had the least impact. Lunch had the most impact
Reading Score: Gender had a negative impact, Test prep most impact
Writing Score: Gender had negative impact, test prep most impact

Random Forest
Instad of using Random forest for classificiation, I used it for regression to find predictors of my model. Running Random Forest produced subpar results. Using Root Mean Squared Error as a metric, running Random Forest resulted in a lower Root Mean Squared Error than all the other methods.

PCA:
I was able to find split my dataset into components to reduce the deimisnsion of the data. While plotting the the points on the first and second components of principal components, I found a pattern as seen on the left with the points forming three separate planes. This made it apparent to me that the dataset was fabricated and not real.

Interpreting your results:

1) return to research question & how the results are answered by your methods.
2) It should explain how the visuals can be interpreted, and demonstrate your knowledge of the subject matter & corpus.

I have not ran my Machine learning methods yet but I created some visualizations so I can find trends in the data and hopefully develop some intuition.

Conclusions/Further steps

As a result of my research we are able to see that lunch and test preparation have the largest impact on test scores. To put my results into action I would recommend the school to offer lunch to students that may not have access to it. This would signficantly inhance a student's cognitive ability and allow them to focus in class resulting in high test scores. Another recommendation would be to put more time into standardize test preparation. Though standarize testing is dying out, it is almost impossible to receive admission to a top ranked school without a good ACT or SAT score. By providing students with a class period every week, students will be prepared for the test and have a bright future ahead of them

Work Cited
include the work you are citing:

1) High School Test Scores
2) DH 100
3) Instructor: Dr Anderson
4) Student: Channing Lee.

Seshapanpu, Jakki. “Students Performance in Exams.” *Kaggle*, 9 Nov. 2018, www.kaggle.com/spscientist/students-performance-in-exams.

https://github.com/earthimmortal/DigHum100