Standardized Test Analysis – Success by Tutoring to STEM Majors?

General Assembly DSI092021 - Project One

Problem Statement:

The SAT and ACT are benchmarks by which colleges base their entrance decisions. As the college-entry process can be highly competitive, many students seek tutoring for these tests.

Positing we were starting a new tutoring service for the SAT and we wanted to make our service specialized: can we position ourselves as experts for tutoring potential STEM students? Looking at whether or not students with intended college major in a STEM degree on the SAT, the SAT reading section, the SAT math section; could we make reasonable conclusions on which topics to tutor to for those STEM students?

Background:

Which Majors are STEM Majors?

The ACT defines a STEM major as one of the following:

* Agriculture, Agriculture Operations, and Related Sciences

* Biological and Biomedical Sciences

* Computer and Information Sciences and Support Services

* Engineering

* Engineering Technologies/Technicians

* Health Professions and related Clinical Sciences

* Mathematics and Statistics

* Natural resources and Conservation

* Physical Sciences

Background:

Have studies look at this before?

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.880.350&rep=rep1&type=pdf

"Patterns of SAT Scores, Choice of STEM Major, and Gender" by Davison, Jew, and Davenport

From that study: "In this conceptualization, verbal abilities are negatively associated with a STEM choice because high verbal abilities increase the attractiveness of non-STEM alternatives. High verbal abilities raise the student's expected success in non-STEM areas and thereby make the non-STEM alternatives more attractive."

"Second, we hypothesized that the difference, SATQ − SATV, would be associated with choice of a STEM major. Essentially, this means that students for whom the difference is positive will tend more toward STEM majors and that students for whom the difference is negative will tend more toward non-STEM majors." I wondered if that that hypothesis hold true with our SAT data but did not have time to explore this question specifically.

The Data:

What are we working with?

As we were trying to explore whether or not we could make a specialized SAT class for STEM students, we want to look at data relevant both all students and specifically STEM students.

After importing, cleaning, and synthesizing the data we worked with three datasets:

- sat_act_2019
  - which merged state's ACT and SAT data from the year 2019
- sat_2019_int_major
  - which was more or less unchanged from the given data, just cleaned
- int_major_stem
  - which looked solely at 2019's SAT scores for STEM majors (as defined by the ACT)

The 'Data Dictionary' in our .ipynb file fully reviews these in detail.

Exploratory Data Analysis:

What can we see at a brief glance?

- The following states performed the best on the SAT in 2019: Minnesota, Wisconsin, South Dakota, North Dakota, and Nebraska.

- The following states performed the worst on the SAT in 2019: Idaho, Delaware, District of Columbia, Oklahoma, and West Virginia.

- The following states performed the best on the SAT EBRW section in 2019: Minnesota, Wisconsin, South Dakota, Nebraska, and Mississippi.

- The following states performed the worst on the SAT EBRW section in 2019: Rhode Island, Delaware, District of Columbia, Oklahoma, and West Virginia.

Exploratory Data Analysis:

What can we see at a brief glance?

- The following states performed the best on the SAT Math section in 2019: Wisconsin, Minnesota, North Dakota, South Dakota, and Nebraska.

- The following states performed the worst on the SAT Math section in 2019: Delaware, Florida, District of Columbia, Oklahoma and West Virginia.

- The following states had the highest rate of participation on the SAT in 2019: Rhode Island, Illinois, Michigan, Colorado, and Connecticut.

- The following states had the lowest rate of participation on the SAT in 2019: Mississippi, Iowa, Wisconsin, Wyoming, and North Dakota.

Exploratory Data Analysis:

What can we see at a brief glance?

- The intended college majors for which students scored the highest on the SAT were:

- Mathematics and Statistics (**STEM** major)
- Physical Sciences (**STEM** major)
- Social Sciences
- Computer and Information Sciences and Support Services (**STEM** major)
- Multi/Interdisciplinary Studies

- The intended college majors for which students scored the lowest on the SAT were:

- Family and Consumer Sciences/Human Sciences
- Precision Production
- Personal and Culinary Services, General
- Construction Trades, General
- Mechanic and Repair Technologies/Technicians

Exploratory Data Analysis:

What can we see at a brief glance?

- The intended college majors for which students scored the highest on the SAT Math section were:

- Mathematics and Statistics (**STEM** major)
- Physical Sciences (**STEM** major)
- Computer and Information Sciences and Support Services (**STEM** major)
- Engineering (**STEM** major)
- Multi/Interdisciplinary Studies

- The intended college majors for which students scored the lowest on the SAT Math section were:

- Family and Consumer Sciences/Human Sciences
- Precision Production
- Construction Trades, General
- Mechanic and Repair Technologies/Technicians
- Personal and Culinary Services, General

Exploratory Data Analysis:

What can we see at a brief glance?

- The intended college majors for which students scored the highest on the SAT EBRW section were:

- Mathematics and Statistics (**STEM** major)
- Physical Sciences (**STEM** major)
- Social Sciences
- English Language and Literature/Letters
- Liberal Arts and Sciences, General Studies and Humanities

- The intended college majors for which students scored the lowest on the SAT EBRW section were:

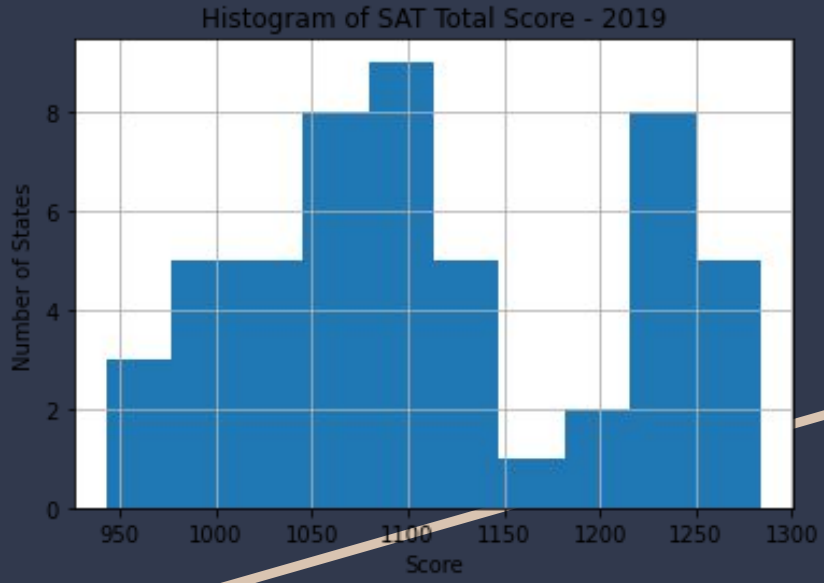
- Parks, Recreation and Leisure Studies
- Personal and Culinary Services, General
- Precision Production
- Construction Trades, General
- Mechanic and Repair Technologies/Technicians

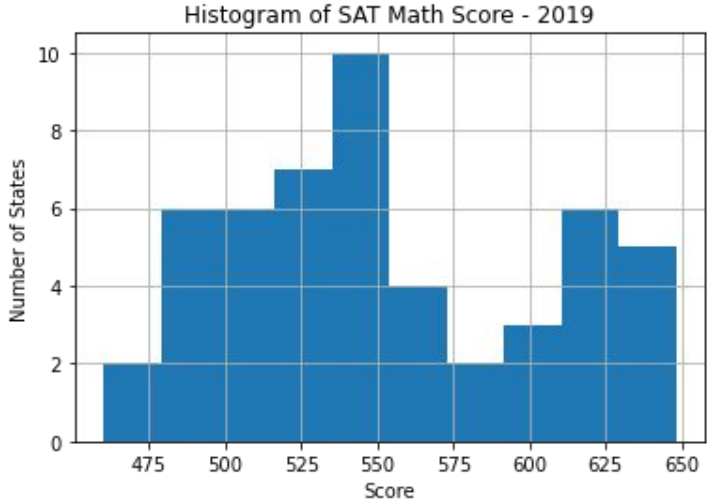
Findings Thus Far:

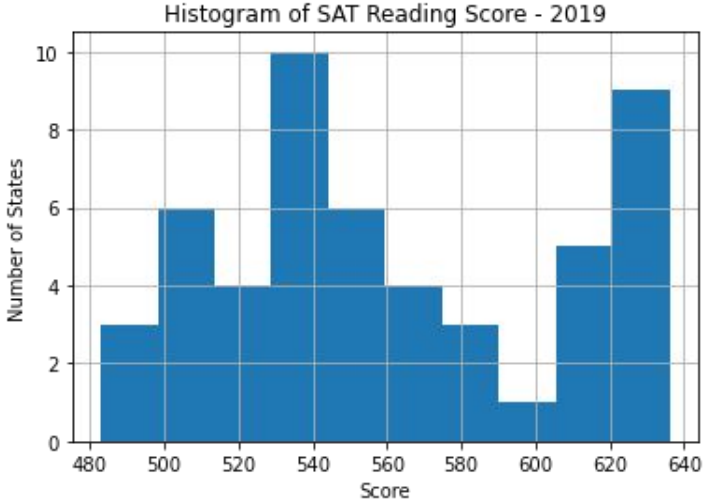
Of all students, regardless of intended major, the mean SAT score was 1058.8 with a standard deviation of 82.45. The range of overall SAT scores was 326. For the Reading section, the average score was 535.5 with a standard deviation of 41.36. The range of Reading scores was 139. For the Math section, the average score was 523.0 with a standard deviation of 43.39. The range of Math scores was 190.

Of those STEM students, the mean SAT score was 1116.56 with a standard deviation of 83.53. The range of overall SAT scores was 265. For the Reading section, the average score was 555.0 with a standard deviation of 35.61. The range of Reading scores was 101. For the Math section, the average score was 561.0 with a standard deviation of 49.89. The range of Math scores was 165.

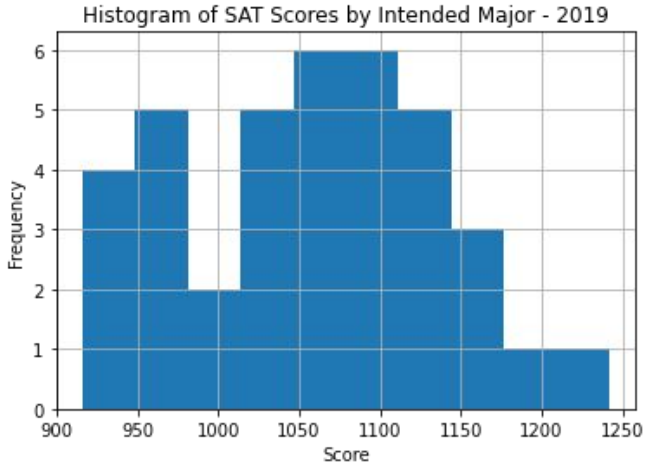
Data Visualizations:

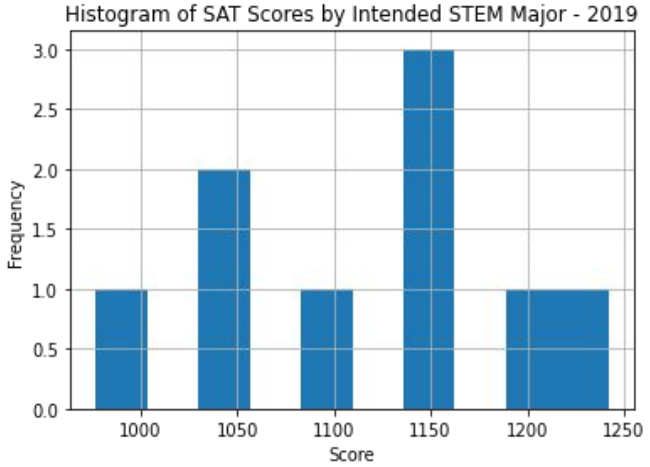




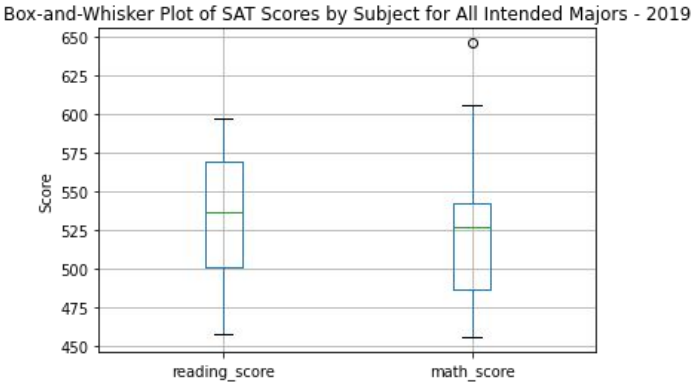


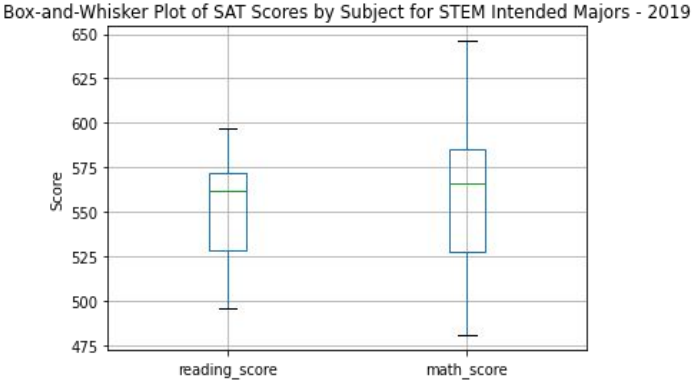
Data Visualizations:



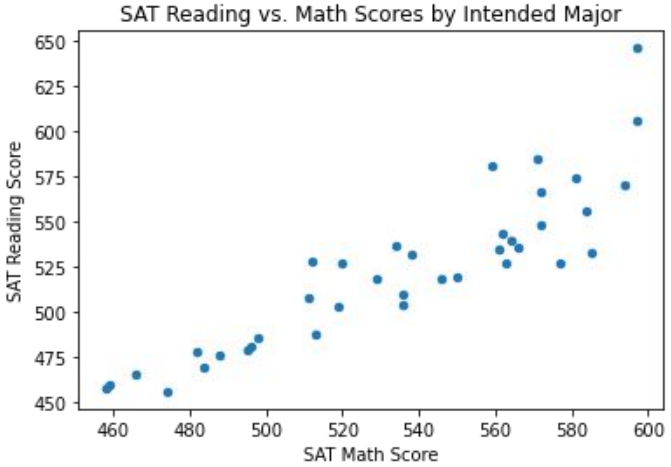


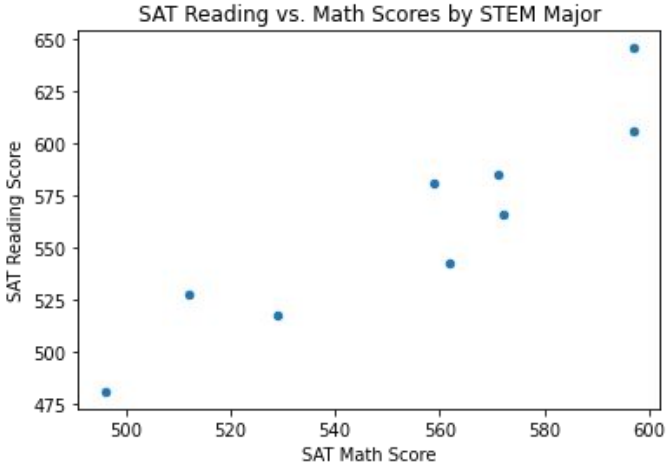
Data Visualizations:





Data Visualizations:





In Conclusion:

Although we attempted to look at states which scored the worst in the SAT and by SAT section, there was no way to account for how many STEM students were coming from each state. In a general sense, as those students from the District of Columbia and Delaware perform the worst on average, it may be a wise business decision to start our tutoring business in those markets. Again, however, as there's no way to account for what percent of STEM students are coming from what state, so this insight is not as valuable as it could be potentially.

Based on the data synthesized, however, we were able to conclude that STEM majors typically perform worse on the Reading section of the SAT than on the Math section, although they still typically outperform their non-STEM peers. This could be attributable to a variety of reason, including personal interests, as explored in the study by Davison, Jew, and Davenport, under Outside Research. Overall, however, STEM students have the most area for improvement in the Reading section.

It would benefit our potential tutoring business to design a module for STEM students and the reading section to help close the gap between their SAT reading and SAT math scores.

The Sources:

External Sources:

https://collegereadiness.collegeboard.org/sat/register/policies-requirements/questionnaire

https://www.act.org/content/act/en/research/reports/act-publications/condition-of-stem-2013/stem-majors-and-occupations/stem-majors-and-occupations.html

https://www.kff.org/other/state-indicator/distribution-by-age/?dataView=1&currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.880.350&rep=rep1&type=pdf

Coding:

https://www.shanelynn.ie/pandas-drop-delete-dataframe-rows-columns/

https://stackoverflow.com/questions/17071871/how-do-i-select-rows-from-a-dataframe-based-on-column-values

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html

https://www.geeksforgeeks.org/saving-a-pandas-dataframe-as-a-csv/

https://www.kite.com/python/answers/how-to-find-the-correlation-between-two-pandas-dataframe-columns-in-pythonv

https://www.statology.org/seaborn-title/#:~:text=To%20add%20a%20title%20to%20a%20single%20seaborn%20plot%2C%20you,set()%20function.&text=To%20add%20an%20overall%20title,suptitle()%20function.

https://www.kite.com/python/answers/how-to-plot-a-histogram-in-matplotlib-in-python
https://matplotlib.org/2.0.2/examples/statistics/histogram_demo_features.html

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.boxplot.html

https://pandas.pydata.org/pandas-docs/version/0.25.0/reference/api/pandas.DataFrame.plot.scatter.html

https://seaborn.pydata.org/generated/seaborn.pairplot.html

https://stackoverflow.com/questions/36813396/how-to-show-the-title-for-the-diagram-of-seaborn-pairplot-or-pridgrid