

Dataset Used

PISA is a survey of students' skills and knowledge as they approach the end of compulsory education. This study was conducted to analyse the data of around 500,000 students, to find their readiness for the life beyond school. Around 500,000 students in 65 economies took part in the PISA 2012 for proficiency in Mathematics, Reading and Science representing about 28 million 15-year-olds globally. Of those economies, 44 took part in an assessment of creative problem solving and 18 in an assessment of financial literacy.

Question 1: Does gender and age has any effect on the performance of students in math, reading and science? Is there any gender gap in participant's countries?

- ST04Q01, "Gender"
- CNT, "Country code 3-character"
- AGE, "Age of student"
- Average Math Score = $PV1MATH + PV2MATH + PV3MATH + PV4MATH + PV5MATH / 5$
- Average Reading Score = $PV1READ + PV2READ + PV3READ + PV4READ + PV5READ / 5$
- Average Science Score = $PV1SCIE + PV2SCIE + PV3SCIE + PV4SCIE + PV5SCIE / 5$

Question 2: Does Immigration status has any influence on student performance?

- Immigration Status = IMMIG
- Country of Birth National Categories Student = COBN-S
- Country of Birth National Categories Mother = COBN-M
- Country of Birth National Categories Father = COBN-F

Question 3: Does math anxiety influence math scores, and is there any correlation between math friendliness and countries?

- Math_motivation = SUBNORM
- Math Interest = INTMAT
- Parents Like Math = ST35Q06
- Math Anxiety = ANXMAT

Main Findings

FEATURES AND RELATIONSHIPS BETWEEN DIFFERENT VARIABLES

- My feature of Interest was Math, Reading and Science Scores and how different factors affect the academic performance of Students. We found out that Male participants were always performing better than the female participants in Math.
- I also looked into the relationship between Math, Reading and Science Scores and found out that all three academic indicators were showing the same results. This means that students take care of all three subject areas.
- I found out that the smaller math anxiety level of participants, the better scores in average math performance.
- Correlation between math performance and math anxiety is (-0.54). If the students can control their anxiety levels they can score better in Math.

- Location has an influence on the overall scores of students. Hong Kong China have shown extremely high scores in math, reading and science. Estonia, is next as it is on the 2nd place for Math and Science and on the 3rd place for Reading.
- Immigration status for the USA and Canada showed interesting features for different countries and there was no pattern that was being followed.
- The Science scores are higher throughout the range. The smallest age 15.17 has lower scores in all three subjects. Age value of 15.78 is the mean where the age column has null data values and we filled then with the average, so that value is not showing a true result. We do not have a big range of age, but we can see that the Science scores are higher throughout the range.
- On average boys outperformed girls by 11 score points in mathematics, but girls significantly outperformed boys in reading by 45 score points. In science there is almost no gap on average across all countries.
- There are some countries where girls perform better in math than boys, these are countries Jordan, Qatar, Thailand, Malaysia, United Arab Emirates, Singapore. There also are several countries in which there is no significant gender gap like Albania, Lithuania and Kazakhstan. Colombia and Chile are among the top countries in terms of gender gap.

INTERESTING FEATURES

- An interesting feature was noticed in Immigration Status, Natives showed high Math Scores, whereas Chile, Canada, Peru and Florida showed a mixed relationship. Canada and Florida have same Math Scores for all three generations.
- I found out that most of the parents were math friendly and lesser number of parents did not like math and just a very little number strongly disliked math.
- Qatar is one of the countries showing a reversal in gender gap, the reason might be that Qatar's female population is more than the male population.

Reasons

I used Jupyter notebook to analyze the relationship between the features of interest. The investigation started with data wrangling and then the cleaned up data was used to generate visualizations to get insight into PISA study. This presentation includes the interesting features and key findings shown with the help of visualizations. I have zoomed in on USA and Canada for one part of this investigation, where we are analysing the relationship of Immigration status and math likeliness on math scores of the participants.

- Influence of Gender on students performance in Math, Reading and Science literacy and the presence of gender gap in different parts of the world.
- Influence of location on student performance
- Influence of Immigration Status and Parent's Math Likelihood on Math Scores for participants from USA and Canada

I wanted to see how different parts of the world show different patterns in the data, that's why I used the Country against math likelihood and gender and investigated the gender gap between different countries. Immigration status is another important aspect that influences academic performance. We are also immigrants and can relate to the differences shown.

Feedback – (Family – Son and Husband)

The analysis visually is very nice but I want the graphs to tell a story those things are lacking. Maybe that is the requirement of the assignment so that may exactly be according to the requirement and nothing wrong with it.

1. What questions do you have about the data

Ans. None, actually the dataset is quite clear

2. What do you notice about visualizations?

Ans. The visualizations are interesting however there is mix of varying level

a. Some graphs can be created with better scales

b. Some graphs can be improved by dropping them from the matrix type multi plots

c. Some interpretation that the user can draw from graphs is missing.

3. What relationships you notice.

Ans. I noticed some obvious positive relations for example the "Math Score " increase with "Math Anxiety" but no explanation is provided for some anomalous data.

4. Is there anything you understand from plots.

Ans. Yes plots tell lots of interesting obvious relations but most of the plots do not tell me anything about the less obvious questions.

References

https://pandas.pydata.org/pandas-docs/version/0.23.1/generated/pandas.read_csv.html

<https://seaborn.pydata.org/generated/seaborn.pointplot.html>

<https://stackoverflow.com/questions/41659188/how-to-adjust-subplot-size-in-seaborn>

<https://seaborn.pydata.org/tutorial/categorical.html>

<https://python-graph-gallery.com/46-add-text-annotation-on-scatterplot/>

https://matplotlib.org/3.3.0/api/_as_gen/matplotlib.axes.Axes.text.html

<https://stackoverflow.com/questions/57417970/how-to-set-custom-colors-on-a-count-plot-in-seaborn>

<https://developers.google.com/chart/interactive/docs/gallery/sankey>

<https://plotly.com/python/animations/>

<https://stackoverflow.com/questions/42404154/increase-tick-label-font-size-in-seaborn>

[https://stackoverflow.com/questions/36220829/fine-control-over-the-font size in seaborn plots for academic papers/36222162](https://stackoverflow.com/questions/36220829/fine-control-over-the-font-size-in-seaborn-plots-for-academic-papers/36222162)