

Data Analysis Report: Students' Performance in Exams

Name: RAHUL YADAV SURESH

Student Number: 23072902

GitHub Repository: https://github.com/rs0206/Students-Performance-in-Exams_assignment

Introduction

The following report discusses relationships in the "Students' Performance in Exams" dataset sourced from Kaggle. This dataset includes information about the profiles of the students as well as their academic performance in different areas. The goal of this study is to find any patterns or connections that could exist between these characteristics, also plot three graph for the Data.

Overview of Dataset

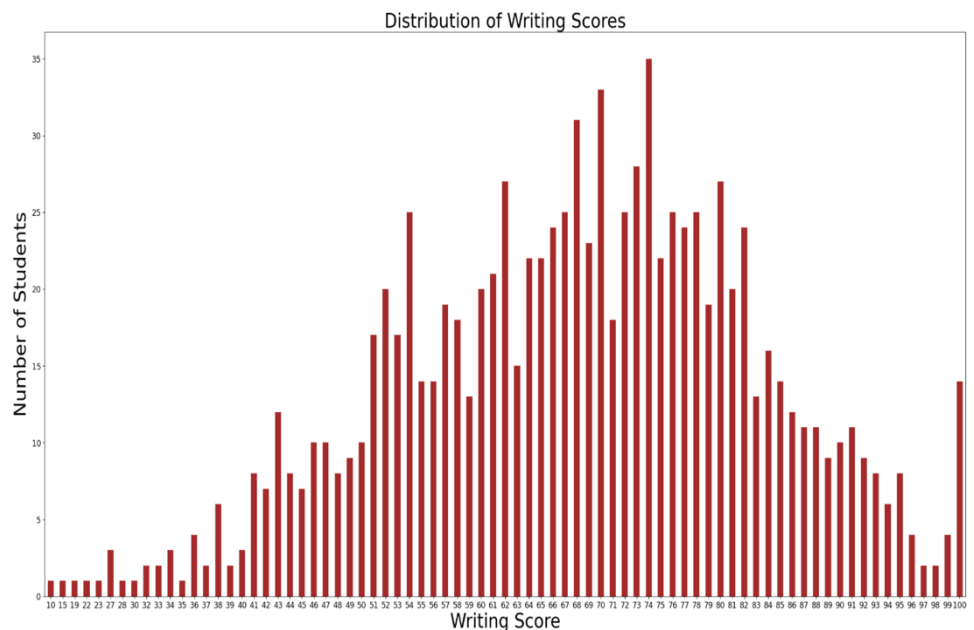
This dataset contains the following columns:

- **Gender:** The student's gender.
- **Race/Ethnicity:** The student's race or ethnicity.
- **Parental Level of Education:** The most degree of education the student's parents have attained.
- **Lunch:** The kind of meal consumed (reduced or free).
- **Test Preparation Course:** Shows if the student has finished a course designed to help them prepare for the test.
- **Math Score:** The student's math score.
- **Reading Score:** The student's reading score.
- **Writing Score:** The student's writing score.

Data Visualizations

1. Bar Chart: Distribution of Writing Scores Among Students

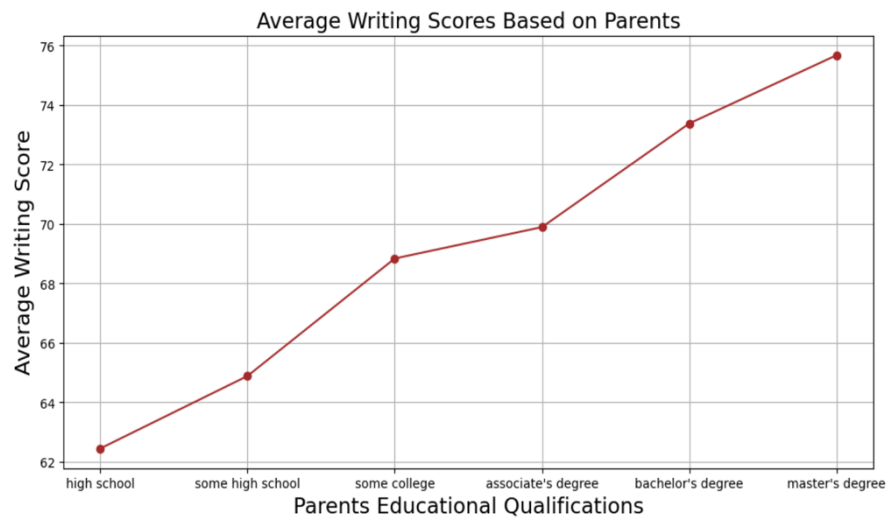
This bar chart shows the distribution of scores from a DataFrame, which is being produced by the `bar(data)` function. Here, I've chosen writing score as the topic of the bar graph. `Value_counts()` is used to determine the frequency of each individual writing score, and the results are then categorised. I have set the size of the x and y labels to 30 and the size of the data in the x and y to 13.



I have also applied the colour brown to the graph to indicate display. A figure that is 30 by 15 inches in size is made. Brown bars are used to plot the bar chart, with the number of students on the y-axis and writing scores on the x-axis. To make it easier to read, titles and axis labels are added with specific font sizes. to correctly understand the data, the plot is shown.

2. Line Graph: Average Writing Scores by Parental Educational Attainment

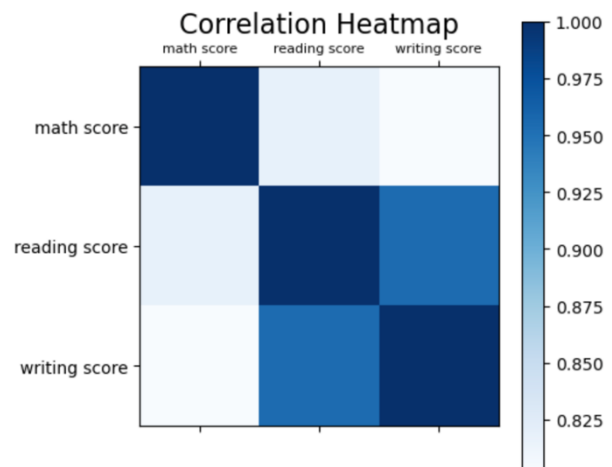
This graph displays the average writing score based on parents' educational background. I have labelled the y-axis as the average writing score and the x-axis as the parents' educational background. Both axes are marked with values up to 17, and I have included grid lines to improve the



appearance of the graph. The figure measures 12 by 6 inches. The graph uses `sort_values()` to sort the values, and the data is re-entered from a DataFrame. It ranks the results after calculating the mean writing score for each educational level using `groupby()`.

3. Heatmap: Correlation Between Scores

To see how math, reading, and writing scores relate to one another, the `heatmap(data)` method creates a heatmap from a DataFrame. First, it uses the `corr()` method to calculate the correlation matrix for these three score columns. A 12-by-12-inch figure is made for the heatmap. the heatmap is shown in blue color. There are various kinds of colormaps here; I've chosen blue, which goes in the `cmap` section. The x and y ticks are labelled with the names of the related scores, and a colour bar is added to show the strength of the connections. To visually represent the strength of the relationships, a colour bar is placed next to the heatmap.



Conclusion

This assignment taught me how to create bar graphs, line graphs, and heat maps in the coding section. The report details how students performed on exams, which are shown in various graph formats, and it also includes information on parent education levels. The reports also include colour charts that visually represent the information contained in the code and what it actually represents.

Reference

Kaggle: Students' Performance in Exams Data's