Student Score Analysis Based on Parental Marital Status and Ethnic Group

Introduction

The objective of this analysis is to explore how students' academic performance (Math, Reading, and Writing scores) is influenced by their **parents' marital status** and **ethnic group**. The analysis involves data aggregation, visualization using a heatmap, and a pie chart representation of student distribution across ethnic groups.

---

1. Aggregating Student Scores Based on Parental Marital Status

Objective

We aim to analyze the average performance of students in **Math, Reading, and Writing** based on their parents' marital status.

Steps Performed

- The dataset was grouped by the `"ParentMaritalStatus"` column.
- The mean scores for `"MathScore"`, `"ReadingScore"`, and `"WritingScore"` were calculated for each marital status category.
- The resulting DataFrame (`gb1`) contains aggregated mean values.

Observations

From the computed mean values:

- Students from **widowed** families scored the highest in Math.
- Students from **divorced** families had slightly higher writing scores.
- **Reading scores** were fairly consistent across different marital statuses.

---

2. Visualizing Student Performance with a Heatmap

Objective

To represent the average scores visually across different parental marital statuses.

Steps Performed

- A heatmap was generated using **Seaborn's `heatmap()` function**.
- The values were annotated on the heatmap for clarity.
- The color scheme was adjusted (`cmap="coolwarm"`) to improve contrast.
- Text annotations were formatted for visibility.

Insights from the Heatmap

- **Reading scores** showed the most variation across groups.
- **Math scores** had the least difference between different marital statuses.
- Overall, no drastic performance gaps were observed, but subtle variations exist.

---

3. Analyzing Ethnic Group Distribution Using a Pie Chart

Objective

To visualize the distribution of students across different ethnic groups.

Steps Performed

- The `"EthnicGroup"` column was extracted and cleaned to remove `NaN` values.
- The dataset was divided into sub-groups based on ethnicity (`grpA`, `grpB`, etc.).
- The number of students in each group was counted.
- A **pie chart** was created to display the percentage distribution of each ethnic group.

Fixes & Enhancements

- The initial error (`ValueError: incomplete format`) was due to incorrect formatting in the `autopct` parameter. It was fixed by using `"%1.2f%%"`.
- The data passed to `plt.pie()` was changed from categorical values to numeric counts.

Insights from the Pie Chart

- Some ethnic groups have significantly larger student populations than others.
- The diversity in student representation is evident, which can be further analyzed for academic performance trends.

Conclusion

This analysis provides valuable insights into student performance and demographic distribution:

- **Parental marital status has some influence on scores**, but the variation is not drastic.
- **Reading scores show the highest difference across different marital statuses.**
- **The student population is distributed across multiple ethnic groups**, with some groups being more represented than others.

Future analysis could further explore:

- The impact of **parental education level** on student performance.
- Correlation between **socioeconomic factors** and academic success.

This study highlights how **demographic factors** might play a role in student performance, opening pathways for more in-depth research and interventions in education.