STUDENTS’ EXAM SCORES

|  |  |
| --- | --- |
| BASEMA BDEIR | 42010683 |
| MANAR HOSSAM | 42010603 |
| MENNA SOLIMAN | 42010691 |
| REEM HAITHAM | 42010049 |
| SAMAR HANY | 42010256 |
| MOHAMED ASHRAF | 42010297 |

Problem Definition:

In this data set we had to deal with:

* **Missing values:** Missing values in the 'WklyStudyHours', 'IsFirstChild', 'EthnicGroup', 'ParentEduc', 'TestPrep', and 'ParentMaritalStatus' columns.
* **Data types:** Wrong data type in the 'WklyStudyHours' column.
* **Outliers:** Outliers in the 'NrSiblings' column.
* **Data cleaning:** The code is dropping the 'Unnamed: 0' column.

Method:

Using python, the code handles:

* **Missing values:** The code is handling missing values in several columns ('WklyStudyHours', 'IsFirstChild', 'EthnicGroup', 'ParentEduc', 'TestPrep', and 'ParentMaritalStatus') by replacing them with the mode.
* **Data cleaning:** The code is dropping the 'Unnamed: 0' column, which appears to be an extraneous column with no useful information.
* **Data type conversion:** The code is converting the 'WklyStudyHours' column from a float to a string data type.
* **Outlier:** The code is identifying and handling potential outliers in the 'NrSiblings' column by replacing them with the median value.
* **Data visualization:** Various visualizations, such as pie charts, histograms, and box plots, are created to explore the relationships between different variables in the dataset.

Experiment:

The 'WklyStudyHours' column contains missing values, which are handled by replacing them with the mode. Potential outliers in the 'NrSiblings' column are identified and handled by replacing them with the median value. The 'WklyStudyHours' column is converted from a float to a string data type. The analysis shows that the majority of students are female, and the most common ethnic group is group C. The analysis also reveals that there is a relationship between gender and exam scores, with female students generally performing better than male students. Additionally, the analysis shows that there is a relationship between the lunch type, test preparation, parent's education background, and parent's marital status and exam scores. Students who have free/reduced lunch, completed test preparation, have parents with higher education backgrounds, and who have married parents tend to perform better on exams. The number of siblings and sport practice have a weaker relationship with exam scores.

Add column GPA to calculate OverallScore

References:

1. <https://www.kaggle.com/datasets/desalegngeb/students-exam-scores>

GitHub:

https://github.com/manarhossam/Student-Scores