Data Analysis Report

This report presents a comprehensive analysis of a dataset obtained from Kaggle, utilizing Python libraries such as Pandas and Seaborn.

A by Ahmed Masoud

Data Source: Kaggle

Kaggle is a platform for data science competitions and a valuable resource for finding publicly available datasets.

Dataset Selection

The dataset chosen for this analysis aligns with the project goals and research objectives.

Data Description

The data focuses on student performance in schools and the various factors that influence it..

```
# Look at the Data
print(data.shape)
print(data.info())
print(data.describe())
```

```
# Convert data to numeric
data["Parental_Involvement"] = data["Parental_Involvement"].replace({"Low" : 1,"Medium" : 2,"High" : 3})
data["Access_to_Resources"] = data["Access_to_Resources"].replace({"Low" : 1,"Medium" : 2,"High" : 3})
data["Extracurricular_Activities"] = data["Extracurricular_Activities"].replace({"Yes" : 2,"No" : 1})
```

```
# Save the data
data.to_csv("NEWData.csv")
```

```
# We have null values at Parental_Education_Level & Distance_from_Home & Teacher_Quality
print(data["Parental_Education_Level"].mean())      # approx. (1.7) ==> College  ==> 2
data["Parental_Education_Level"].fillna(2,inplace=True)
```

```
# Group Columns:
# Family Columns: Parental_Involvement - Motivation_Level - Family_Income - Parental_Education_Level
# School Columns: Distance_from_Home - Physical_Activity - School_Type - Teacher_Quality - Internet_Access
# Student Columns: Hours_Studied - Attendance - Extracurricular_Activities - Sleep_Hours - Previous_Scores
```

Data Cleaning: Pandas in Python

Pandas, a powerful Python library, enables efficient data manipulation and cleaning.

1 Handling Missing Values

Missing values were addressed using appropriate methods, such as imputation or removal.

2 Data Transformation

Data was transformed to ensure consistency and prepare it for analysis.

3 Data Validation

The cleaned data was validated to ensure accuracy and completeness.

Exploratory Data Analysis

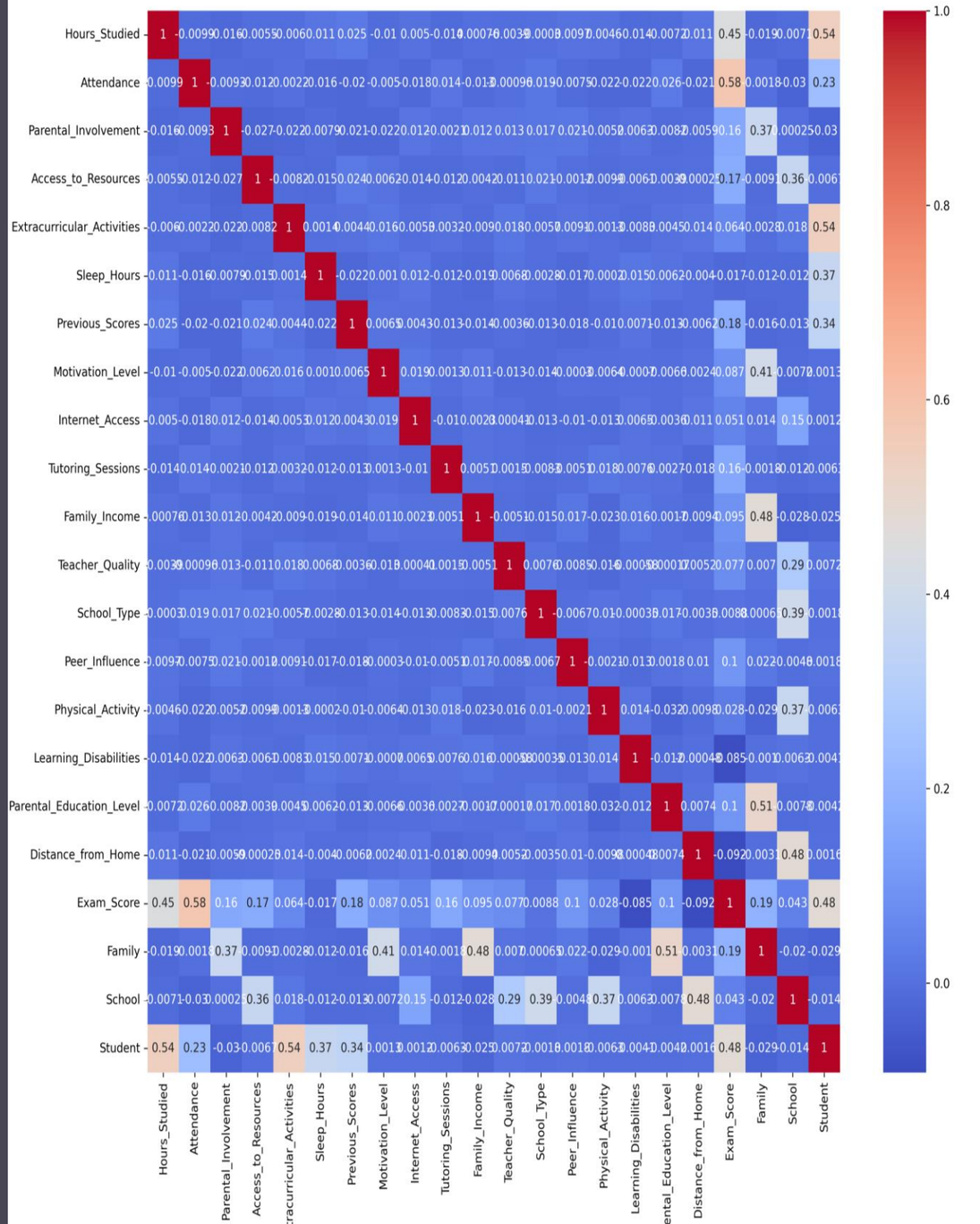
Exploratory Data Analysis (EDA) helps uncover patterns and insights within the data.

1 Descriptive Statistics

Summary statistics were calculated to understand the basic characteristics of the data.

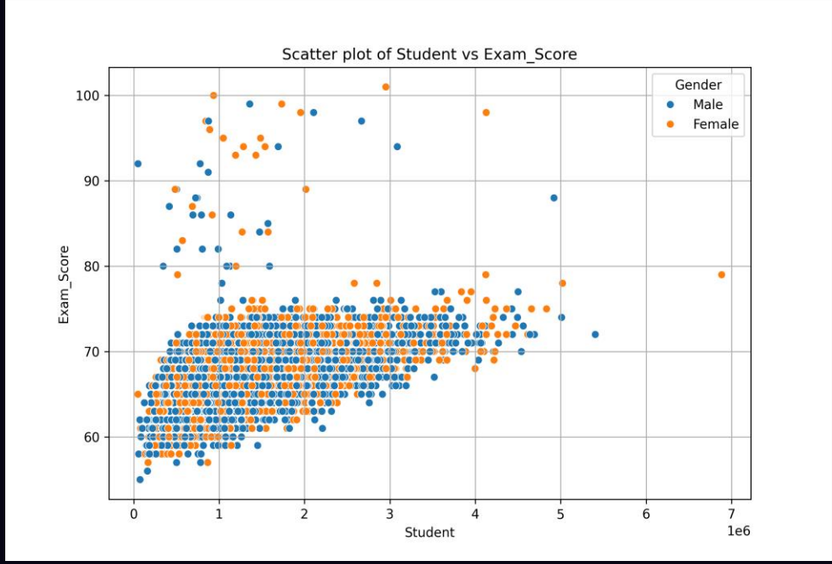
2 Correlation Analysis

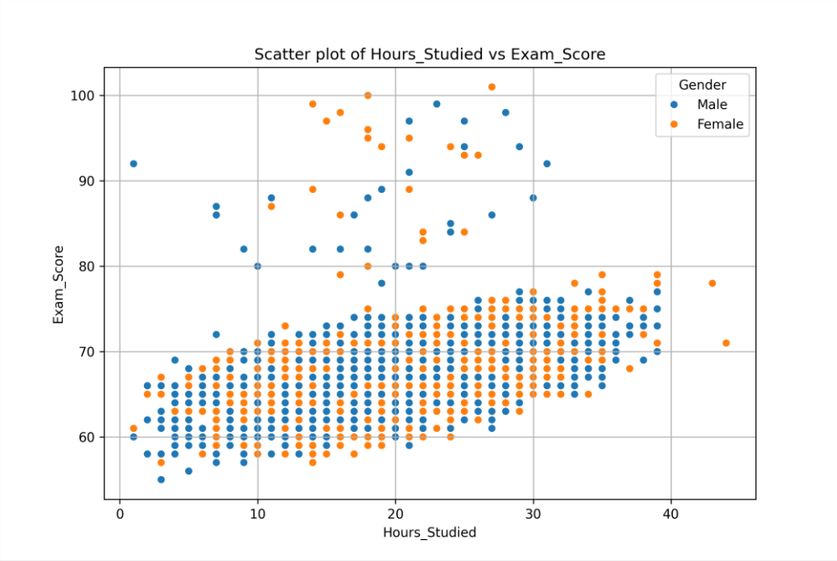
Correlation coefficients were calculated to identify relationships between variables.

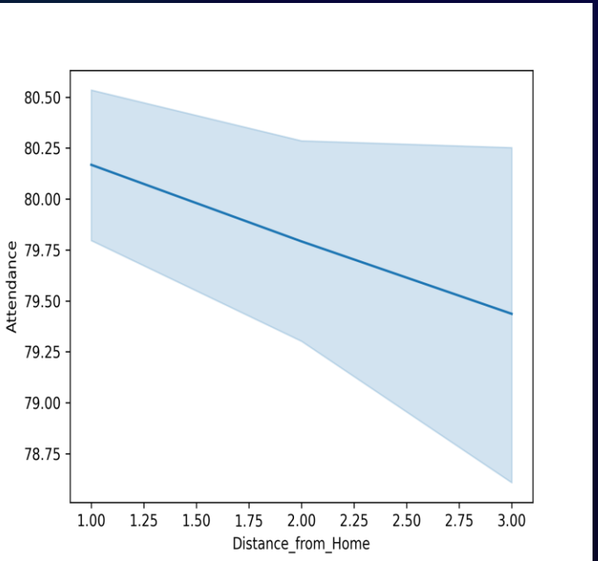


Visualizations: Seaborn and Matplotlib

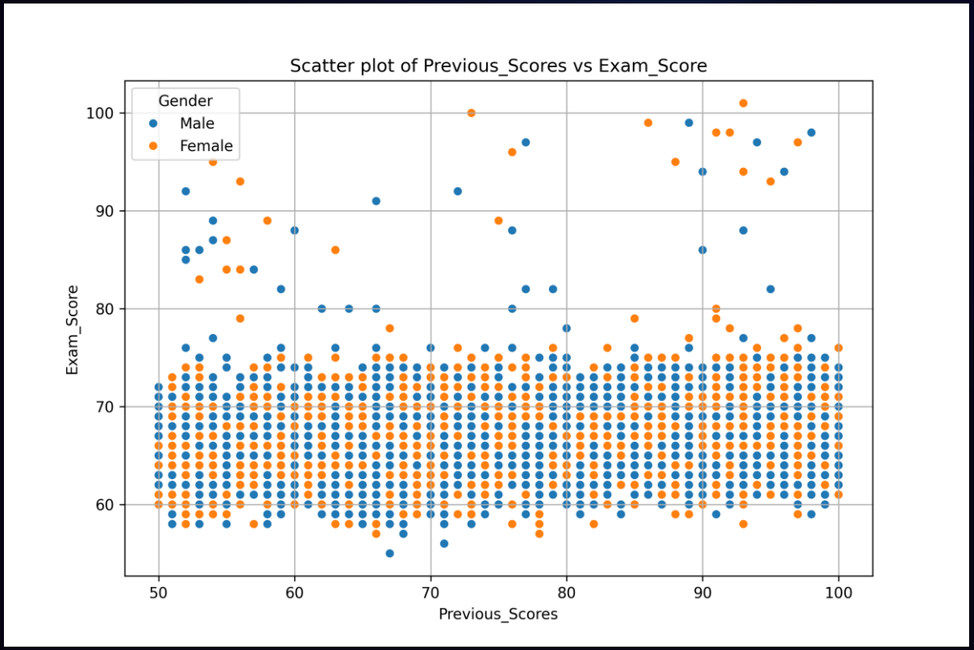
Seaborn and Matplotlib, popular Python libraries, facilitated the creation of informative visualizations.

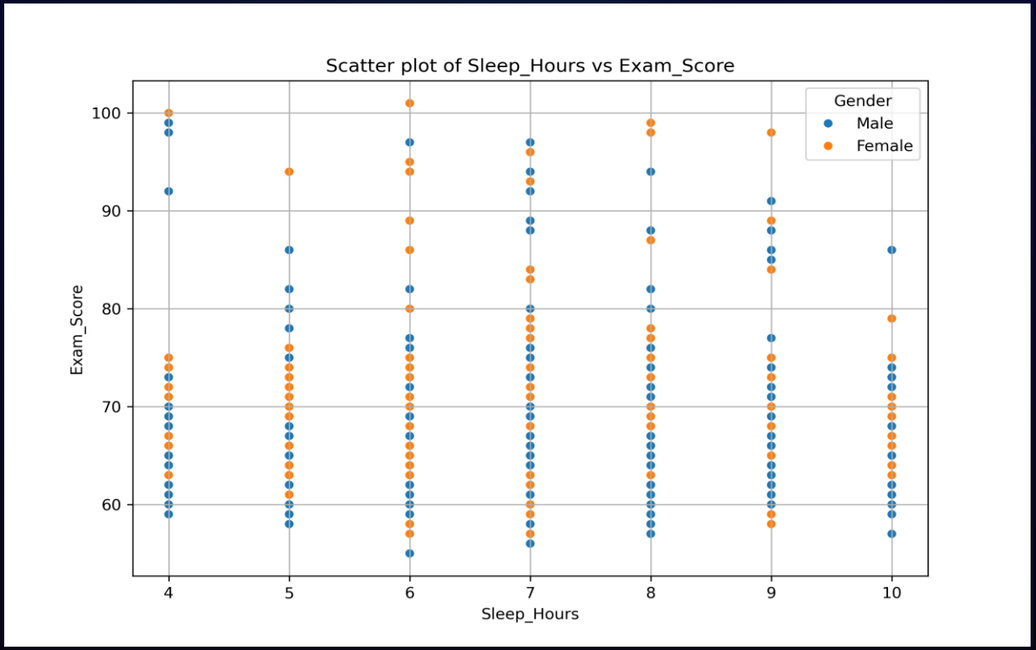






Identifying Relationships in the Data





Key Findings and Insights

The analysis revealed several key findings and insights about the data:

1. School Proximity:

- Attendance is influenced by the distance to the nearest school.

2. Exam Score Influences:

- School Factors:
  - Access to resources significantly impacts student progress.
  - School type and teacher quality show minimal effect on progress.
- Family Factors:
  - Family involvement has a slight influence on progress.
  - Notably, students from low-income families can achieve high grades.
- Student Factors:

The most significant influence on progress comes from:

- Hours studied
- Attendance rates
- Previous exam scores

Recommendations and Next Steps

Based on the findings, several recommendations and next steps are proposed.

1. Encourage Family Engagement:

- Families should actively support and monitor their children's academic progress to foster a positive learning environment.

2. Transportation Solutions:

- Schools should consider implementing bus services for students who live at greater distances to ensure consistent attendance and reduce travel barriers.

Conclusion

For dataset and code check:

https://github.com/AhmedMasoud135/Student-Performance-Analysis

For more inquires you can contact me at:

ahmedmasoud2098@gmail.com

www.linkedin.com/in/ahmed-masoud12