Report on Data Cleaning and Exploratory Data Analysis (EDA) of Students Grading Dataset

1. Introduction

This report presents an in-depth analysis of the Students Grading Dataset, focusing on data cleaning, handling outliers, and exploratory data analysis (EDA). The objective is to understand the structure of the data, identify any anomalies, and uncover patterns that might influence student performance.

---

2. Data Cleaning Process

Before conducting the analysis, the dataset underwent preprocessing to ensure data integrity. The following steps were taken:

- **Missing Values:** No significant missing values were detected in the dataset, ensuring that all student records were complete.
- **Duplicate Records:** The dataset was checked for duplicate entries, and no redundant records were found.
- **Data Type Consistency:** All numerical fields (such as Age, Attendance, and Grades) were confirmed to be correctly formatted for statistical analysis.

---

3. Handling Outliers

- The dataset was analyzed for extreme outliers using boxplots.
- No extreme values were found outside the whiskers, indicating that the dataset does not contain significant anomalies.
- The overall spread suggests a relatively even distribution of scores and attendance rates.

---

4. Univariate Analysis

4.1 Age Distribution

- The age distribution is relatively uniform, with students predominantly between 18-24 years old.
- The dataset does not exhibit extreme variations, making it a good representation of a typical university-aged population.

4.2 Attendance Percentage

- The attendance rate peaks between **70-80%**, indicating that most students attend a substantial portion of their classes.
- The distribution suggests that while a few students have lower attendance rates, the majority maintain a consistent presence in class.

4.3 Grade Distribution

- The grades are normally distributed, with a central tendency around the mid-range values.

- No extreme skewness is observed, meaning most students perform within a reasonable range.
- The distribution does not exhibit excessive failing grades, suggesting a well-balanced grading system.

---

5. Bivariate Analysis

5.1 Correlation Between Attendance and Grades

- A positive correlation was observed between **attendance percentage and grades**, implying that students with higher attendance tend to score better.
- However, the relationship is not perfectly linear, suggesting that other factors might also influence student performance.

5.2 Gender-Based Performance Analysis

- The dataset was analyzed for any gender-based performance disparities.
- No significant difference in grades was found between male and female students, indicating that academic performance is not heavily influenced by gender.

5.3 Impact of Age on Performance

- There is no significant correlation between **age and academic performance**, suggesting that students across different age groups perform similarly.

---

6. Key Insights and Conclusion

- The dataset is well-structured with no missing or duplicate values, making it suitable for further analysis.
- Attendance plays a crucial role in student performance, with higher attendance generally leading to better grades.
- Grades follow a normal distribution with no extreme outliers, ensuring a balanced assessment system.
- No significant disparities were observed based on gender or age.

These findings provide a foundation for more advanced predictive modeling, such as predicting student performance based on various attributes. Further analysis can explore multivariate interactions to uncover deeper insights.

---

7. Recommendations

- **Encouraging Regular Attendance:** Since attendance is positively correlated with performance, implementing attendance-based incentives could improve overall student outcomes.

- **Further Investigation of Other Influencing Factors:** Additional factors such as study habits, extracurricular activities, and socio-economic background should be considered in future studies.

This concludes the exploratory analysis of the Students Grading Dataset. Future research can build upon these insights for enhanced academic planning and student support strategies.