



STUDENT ANALYSIS AND PREDICTION

BY: ISMAIL MOHAMED



CONTENT

- 01 INTRODUCTION
- 02 NOTEBOOK SUMMARY
- 03 DATA PREPROCESSING
- 04 EXPLORATORY DATA ANALYSIS (EDA)
- 05 PERFORMANCE ANALYSIS
- 06 BEHAVIOR ANALYSIS
- 07 HYPOTHESIS TESTING

INTRODUCTION

- This presentation provides an in-depth analysis of student demographics, performance, and behavior based on the provided Jupyter notebook.
- The analysis includes data preprocessing, exploratory data analysis (EDA), and hypothesis testing.
- Key insights and visualizations will be presented to understand the dataset better.



NOTEBOOK SUMMARY

- **Demographics:**

- **Gender distribution:** 51% Male, 49% Female.
- **Age distribution:** Males are more than females, except at ages 20 & 23.
- **Highest frequency of ages:** 21 → 23 → 22 → 24 → 19 → 18 → 20.
- **Females outnumber males in all departments.**
- **Extracurricular Activities:** About 70% of students do not participate.
- **Internet access:** About 90% of students have internet at home.
- **Family income levels:** Low (39.7%), Medium (39.5%), High (20.9%).
- **Parental Education Levels:** PhD (52.3%), Master's degree (lowest).

NOTEBOOK SUMMARY

- **Behavior Analysis:**

- **Study hours per week:** Mathematics students study the least (17 hours/week).
- **Stress levels:** Same across all departments (5.5).
- **Sleep hours per night:** Same across all departments (6.5 hours/night).
- **Grades and behavior correlation:** Study hours (17.5), Stress level (5.5), Sleep hours (6.5).
- **Family income level, parental education, internet access, and extracurricular activities** do not significantly impact study hours, stress levels, or sleep hours.

NOTEBOOK SUMMARY

- Hypothesis Testing:

- Males have a higher probability of achieving a higher total score than females.
- Internet access and extracurricular activities have a positive impact on total score and student behavior.

DATA PREPROCESSING

Key Steps:

- Loading Libraries: Importing necessary libraries such as NumPy, Pandas, Seaborn, and Matplotlib.
- Loading Data: Reading the dataset from a CSV file.
- Data Cleaning: Dropping unnecessary columns, handling missing values, and filling them with median values.
- Handling Duplicates: Checking for and removing duplicate rows.
- Data Description: Providing summary statistics and information about the dataset.



(EDA)

- **Key Visualizations:**

- **Countplot of Gender Distribution:** Visualizing the distribution of genders in the dataset.
- **Pie Chart of Gender State:** Showing the percentage distribution of genders.
- **Distribution of Study Hours per Week:** Analyzing the study habits of students across different departments.
- **Correlation Analysis:** Exploring the relationship between different variables such as study hours, stress levels, and sleep hours.

PERFORMANCE ANALYSIS

- **Key Insights:**

- Grade Distribution: Males have a higher count in A grade, while females outnumber males in all other grades.
- Attendance vs. Scores: High attendance does not necessarily correlate with high scores in other areas.
- Impact of Extracurricular Activities: Minimal impact on overall performance and scores.

BEHAVIOR ANALYSIS

- **Key Insights:**

- Study Hours: Mathematics students study the least (17 hours/week).
- Stress Levels: Consistent across all departments (5.5).
- Sleep Hours: Consistent across all departments (6.5 hours/night).
- Correlation with Grades: Study hours, stress levels, and sleep hours show minimal correlation with grades.

HYPOTHESIS TESTING

- **Key Hypotheses:**

- Males vs. Females: Males have a higher probability of achieving a higher total score than females.
- Internet Access and Extracurricular Activities: Both have a positive impact on total score and student behavior.

CONCLUSION

Summary:

- The analysis provides valuable insights into student demographics, performance, and behavior.
- Key findings include the impact of gender on grades, the minimal impact of extracurricular activities on performance, and the consistent stress and sleep levels across departments.
- Further analysis could explore additional variables and their impact on student outcomes.





**THANK
YOU**