

# DAY 2 STUDENT STUDY HOURS ANALYSIS DATA FINAL REPORT



## Data Cleaning and Analysis Report

### 1. Handling Missing Values

- **Study Hours:**
    - Issue: Some values were missing.
    - Action: Replaced missing values using the **mean (average)** of the column.
    - Insight: This ensures the central tendency is preserved without biasing results too high or low.
  - **Screen Time:**
    - Issue: Some values were missing.
    - Action: Replaced missing values using the **median**.
    - Insight: Median is more robust against outliers (e.g., very high screen times), so it provides a stable replacement.
  - **Test Scores:**
    - Issue: A few records had missing values.
    - Action: Replaced missing values using the **mode**.
    - Insight: Since test scores often have repeated values, mode is a good fit to keep results consistent with common performance levels.
- 

### 2. Data Formatting and Transformation

- Converted all **General columns into Numeric** to allow calculations and pivot table analysis.
  - Ensured **consistent column names** (e.g., "Study\_Hours", "Screen\_Time", "Test\_Scores").
  - Created **new derived columns**:
    - **Age Group** (e.g., 13–14, 15–16, 17+)
    - **Screen Time Range** (Low, Moderate, High)
-

### 3. Pivot Table Analysis

- **Screen Time vs Test Scores:**
    - High screen time students: Average ~75
    - Low screen time students: Average ~78 (highest)
    - Moderate screen time students: Average ~63 (lowest)  
👉 Moderate usage surprisingly underperformed compared to high/low.
  - **Age Group vs Test Scores:**
    - 15–16 years: Highest average (~75.2)
    - 13–14 years: Average ~71.5
    - 17+ years: Lowest (~67.9)  
👉 Performance **peaks in mid-teens** and drops at 17+.
- 

### 4. Dashboard Highlights

- Visual dashboards created in Excel using **Pivot Charts**:
    - Bar chart of Screen Time Range vs Scores
    - Age Group vs Scores trend line
    - Distribution of Study Hours
- 

### 5. Key Insights

1. Data cleaning ensured no missing values remain, improving reliability.
2. Choice of mean, median, and mode preserved **statistical integrity** depending on column type.
3. Study shows **high and low screen time** both correlate with better scores compared to moderate.
4. **15–16 years old students perform best**, suggesting peak focus/learning.
5. Cleaned, numeric-formatted dataset is now **ready for advanced analysis & reporting**.