

Week 1 – Introduction to AI & ML Foundations Summary Report.

Project: Student Performance Analysis

1. Project Overview:

The objective of this project was to analyze the relationship between student study habits and academic results. By using a dataset containing "Hours Studied" and "Exam Scores," I performed basic data exploration, cleaning, and visualization to identify trends that could help educators improve student outcomes.

2. Data Cleaning Rationale:

Data cleaning is a critical first step in the AI/ML pipeline to ensure the accuracy of the model. I performed the following:

- **Handling Missing Values:** I checked for null entries using `df.isnull().sum()`. Removing missing values prevents skewing averages and ensures the visualization accurately represents the data.
- **Column Standardization:** I renamed the columns to `hours_studied` and `score`. Using consistent, lowercase naming avoids syntax errors and makes the code more readable.
- **Data Validation:** I verified that the data types were numerical to allow for mathematical operations and plotting.

3. Visualizations & Insights:

I used two primary visualizations to explore the data:

- **Histogram:** Showed the frequency of study hours, revealing that most students in this sample study between 2 and 8 hours.
- **Scatter Plot:** Revealed a strong positive correlation (approximately 0.98) between study time and performance.

4. Actionable Insights for Educators:

Based on the analysis, I have identified the following insights:

- **Direct Impact:** There is a clear link between effort and result; as study hours increase, scores increase significantly.

- **Risk Identification:** Students who study for fewer than 3 hours are at high risk of failing or scoring very low.
- **Recommendation:** Educators should encourage students to aim for a minimum of 5 hours of structured study per week to consistently achieve passing marks above 50%.