STATISTICS AND ANALYTICS REPORT

(DATE:29-11-2025)

INTRODUCTION:-

This report comprises about Statistics and Analytics data

✓ Descriptive Statistics

✓ Probability & Distributions

✓ Correlation Analysis

✓ Regression Analysis (Prediction)

✓ Hypothesis Testing (Decision Making)

DATASET DESCRIPTION:-

For this analysis, four different datasets were used—StudentScores, Employee_Salary, Sales_Marketing, and Health_Data. Each dataset contains a mix of numerical and categorical variables, making them suitable for a variety of statistical techniques such as descriptive statistics, probability, correlation, regression, and hypothesis testing.

TASK 1-DESCRIPTIVE STATISTICS:-

Mean = 70.6

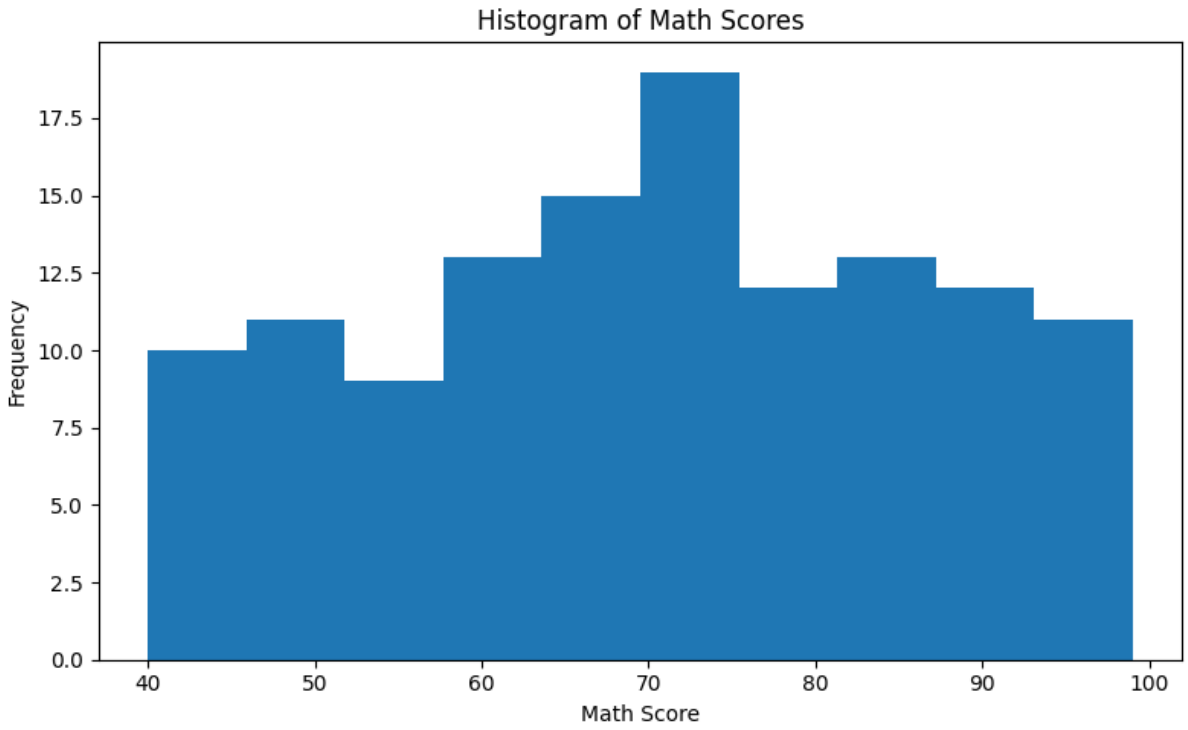
Median = 72

Mode = 72

Range = 59

Standard Deviation = 16.51.

A histogram of Math Scores displays overall distribution.



KEY INSIGHTS:-

Overall Math Performance Is Above Average

The mean Math score is 70.6, which indicates that most students perform better than the passing as per average mark (typically 50 or 60).

The median (72) confirms that half the students score above 72 — a good indicator of consistent performance.*

Score Distribution Is Fairly Balanced

Mean (70.6), Median (72), and Mode (72) are close → No extreme skewness .This suggests a healthy distribution without extreme outliers pulling the mean up or down.

Students Scores Vary Widely

Standard deviation is 16.5 → quite high. It states that:Some students score extremely high- 90+.Some score low -below 50The range of 59 points shows a large performance gap.

Majority of Students Score Between 60 and 90 In the histogram: Most bars fall in the middle range. Very few are in the low 40–50 range, and a handful are in the 90+ range.

Implication: Class performance is overall strong, but improvement for weak performers is possible.

No Indication of Data Issues or Anomalies Distribution is smooth, without spikes or gaps. No clustering at boundaries like exactly 0 or 100. This suggests the dataset is clean and realistic.

Thus, Math score analysis reveals that : Students perform at a generally good level, with most scoring around 70–75, but the presence of high variability suggests the need for targeted interventions for low scorers and opportunities for advanced learners.

TASK 2-PROBABILITY:-

Probabilities

DEPARTMENT	PROBABILITY
Marketing	0.248
HR	0.232
IT	0.208
Finance	0.184
Sales	0.128

Q1.

What is the probability that a randomly selected employee belongs to the Marketing department?

$$P(\text{Marketing}) = 0.248$$

Q2.

If we randomly select two employees (without replacement), what is the probability that both are from HR?

$$P(First\ HR) = 29/125$$
$$P(Second\ HR) = 28/124$$
$$P(Both\ HR) = \frac{29}{125} \times \frac{28}{124} = 0.052$$

Q3.

What is the probability that a randomly chosen employee is NOT from Sales?

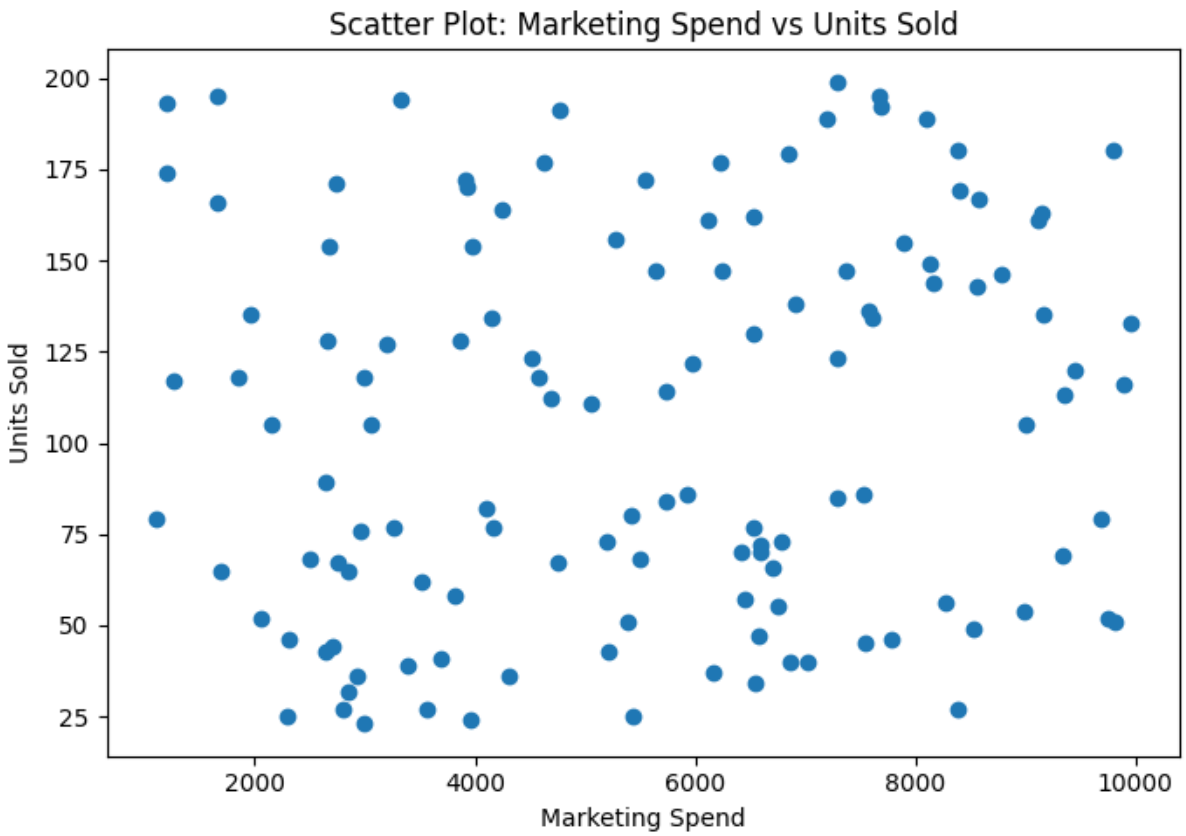
$$P(Not\ Sales) = 1 - P(Sales) = 1 - 0.128 = 0.872$$

INSIGHTS:-

- ***Marketing has the highest probability (0.248).***
- ***Sales has the lowest probability (0.128).***
- ***Real employee distribution does not follow theoretical equal distribution.***
- ***Experimental probabilities help understand actual workforce structure.***

TASK 3-CORRELATION ANALYSIS:-

Correlation between Marketing Spend and Units Sold = 0.1286 (Weak Positive).



INSIGHTS:-

Correlation Result r=0.1286

Weak correlation (because |0.12| < 0.3)

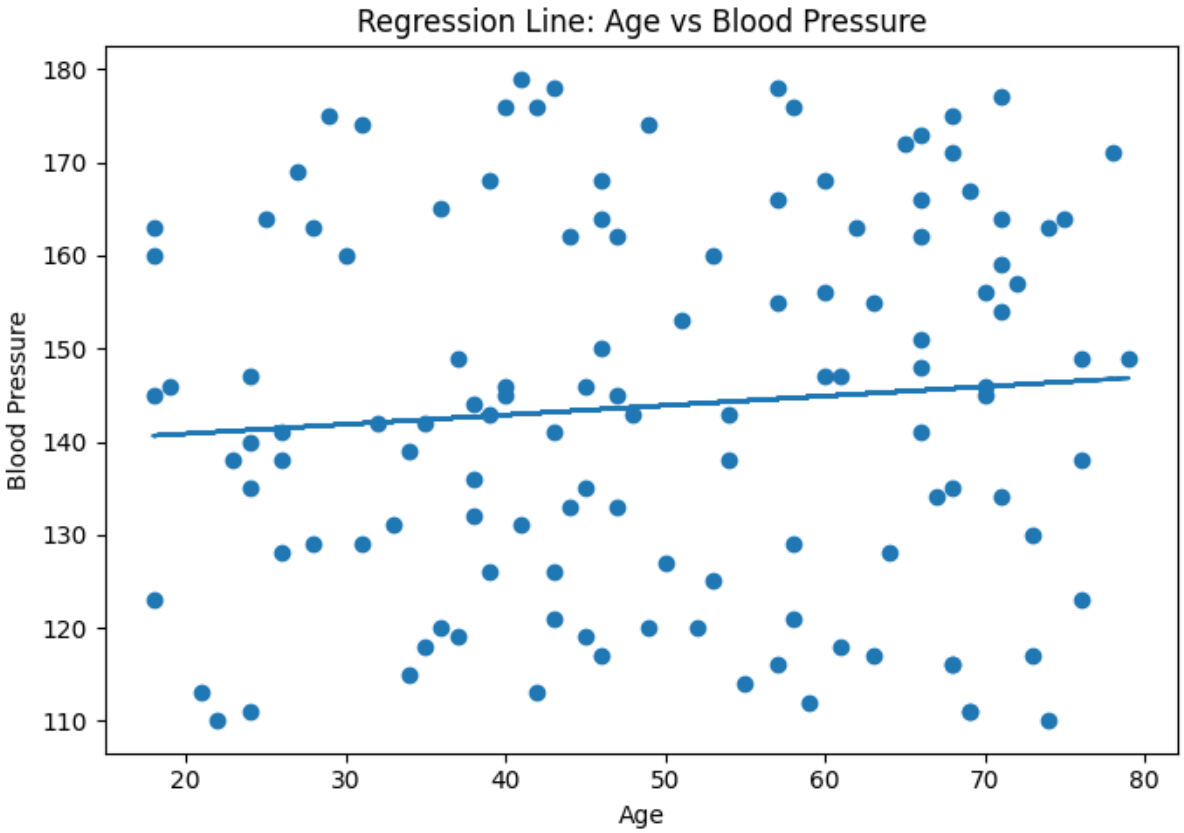
Positive correlation (because r > 0)

There is a weak positive correlation (r = 0.1286) between marketing spend and units sold. While higher marketing spending shows a small tendency to increase sales, the relationship is not strong enough to predict sales accurately.

TASK 4-REGRESSION ANALYSIS:-

Simple Linear Regression Analysis using the dataset Health Data.xlsx shows prediction:

- **Independent Variable (X): Age**
- **Dependent Variable (Y): Blood Pressure**



INSIGHTS:-

Component Result Model -Simple Linear Regression X (Independent Variable)- Age Y (Dependent Variable)- Blood Pressure

BP = 0.1014 × Age + 138.8149 Prediction (Age = 50)- 143.88 mmHg Relationship-Weak Positive Plot Successfully created

TASK 5-HYPOTHESIS TESTING:-

ONE-WAY ANOVA TEST:-

H0: All department salary means are equal.H1: At least one department has a different mean salary.p-value = 0.6807 -Fail to Reject H0.

Conclusion:-

Since p_value=0.6807>0.05 ,failed to reject the Null Hypothesis.

There is no statistically significant difference in the average salary across different departments.

FINAL INSIGHTS:-

Based on the dataset and findings stated above, stating the recommendations under different areas:-

Area	Recommendation
Student Performance	Introduce personalized learning plans & extra support
HR & Staffing	Assess workforce balance and hiring priorities
Sales & Marketing	Optimize marketing spend and adopt data-driven campaigns
Health Analytics	Include additional predictors for better health insights
Salary Policies	Continue current practice but review annually.

LEARNING SUMMARY:-

Week 5 taught me how to:

- Describe data using mean, median, mode, and standard deviation
- Measure probability and make predictions based on uncertainty
- Identify relationships (correlation) and make forecasts (regression)
- Perform hypothesis testing to make evidence-based decisions

These concepts form the core foundation of Data Analytics, enabling me to analyze real datasets, understand patterns, predict outcomes, and validate conclusions using statistical evidence.

**********THE END**********