



PRESIDENCY SCHOOL

Keynotes

Grade: X
Month: MARCH

Subject: Data Science

Topic: USE OF STATISTICS IN DATA SCIENCE

Use of Statistics in Data Science

Keynotes

1. Subsets in Data Science

- **Subset:** A smaller section of a larger dataset used for analysis.
- Methods of subsetting:
 - **Row-based subsetting:** Selecting specific rows.
 - **Column-based subsetting:** Selecting specific columns.
 - **Data-based subsetting:** Filtering data based on a condition.

2. Two-Way Frequency Tables

- A method to organize and analyze categorical data in two dimensions.

3. Statistical Measures

- **Mean:** The average value of a dataset.
- **Median:** The middle value when data is arranged in order
- **Mean Absolute Deviation:** Measures variability by calculating the average distance between each data point and the mean.
- **Standard Deviation:** A measure of the spread of data points around the mean.

4. Applications in Data Science

- Helps in **data cleaning**, **predictive modeling**, and **trend analysis**.
-