Basics of Statistics and ML
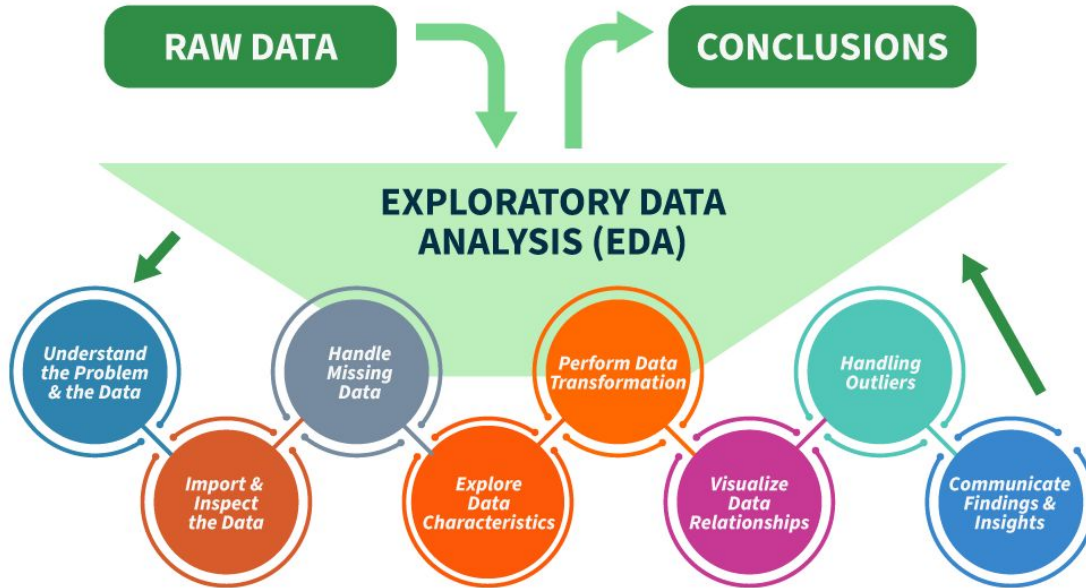
Oct '24

Data-Driven Decision Making
- Businesses can adapt data informed strategies
- Businesses can identify trends and patterns

Predictive Analytics
- Prediction of future outcomes and optimization
- Risk assessment and mitigation

Performance Measurement
- Tracking KPIs
- Significance of drivers, and evaluation of their explainability

Automation and Innovation
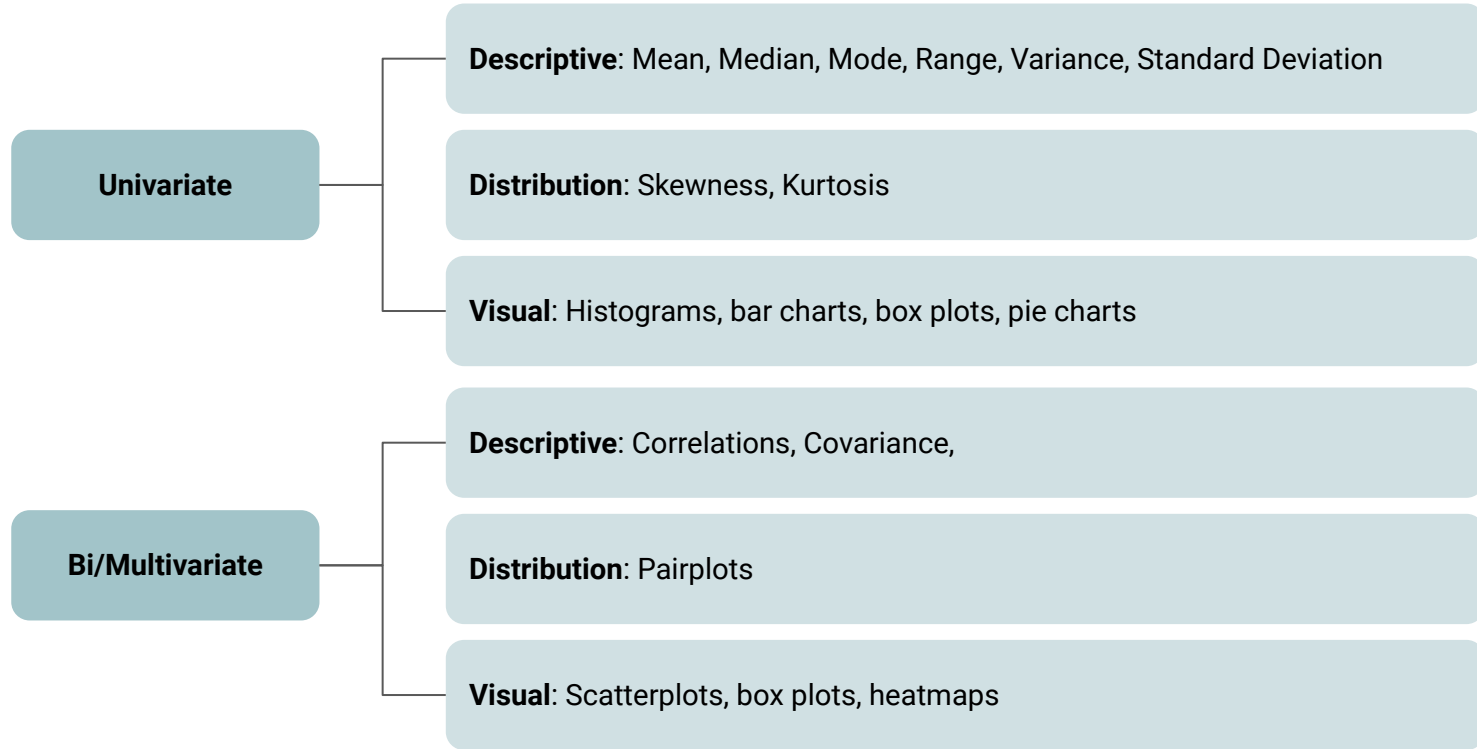- Automating Routine tasks
- Identifying new opportunities

Introduction

Statistics and machine learning enable data-driven decision-making, enhance customer insights, optimize operations, and foster innovation, making them essential tools for achieving competitive advantage in the workplace

The Art of EDA…

Steps for Performing Exploratory Data Analysis

RAW DATA → CONCLUSIONS

EXPLORATORY DATA ANALYSIS (EDA)

- Understand the Problem & the Data
- Import & Inspect the Data
- Handle Missing Data
- Explore Data Characteristics
- Perform Data Transformation
- Visualize Data Relationships
- Handling Outliers
- Communicate Findings & Insights

Univariate and Multivariate Analysis…

Univariate

Descriptive: Mean, Median, Mode, Range, Variance, Standard Deviation

Distribution: Skewness, Kurtosis

Visual: Histograms, bar charts, box plots, pie charts

Bi/Multivariate

Descriptive: Correlations, Covariance,

Distribution: Pairplots

Visual: Scatterplots, box plots, heatmaps

Identifying and Treating Outliers…

Ways to identify outliers

- **Visually**: box plots, scatter plots, or histograms
- **Statistically**: Z-score or IQR

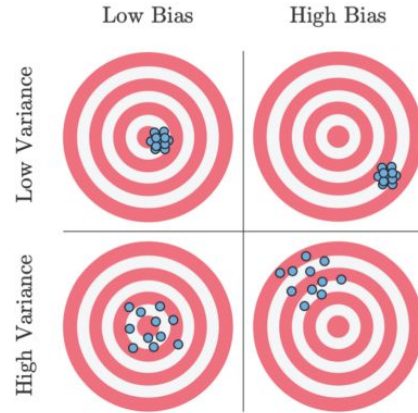

Detecting Outliers with z-Scores
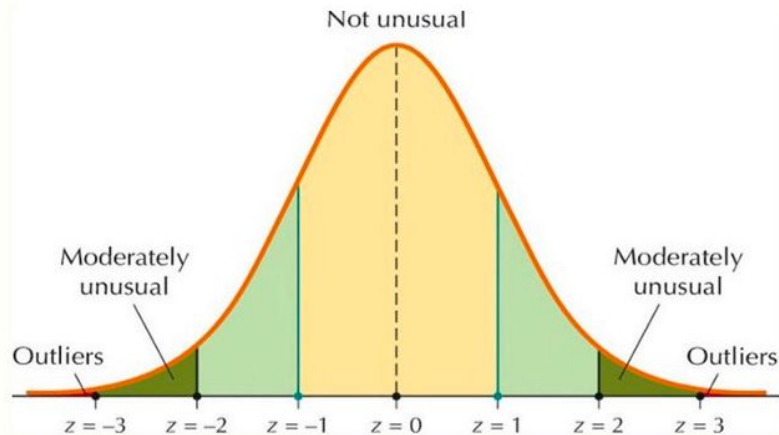
Ways to treat outliers

- **Remove outliers**
- **Apply transformations to reduce the impact**
- **Replace using mean/median**
- **Use more robust models like Random Forest**

Key Concepts of Statistics…

Bias and Variance

Confusion Matrix

Questions

- **What are the null hypothesis in our MaternAI case?**
- **What are some other concepts like p-values, R2, adj-R2?**

Machine Learning Types and Common Algorithms…