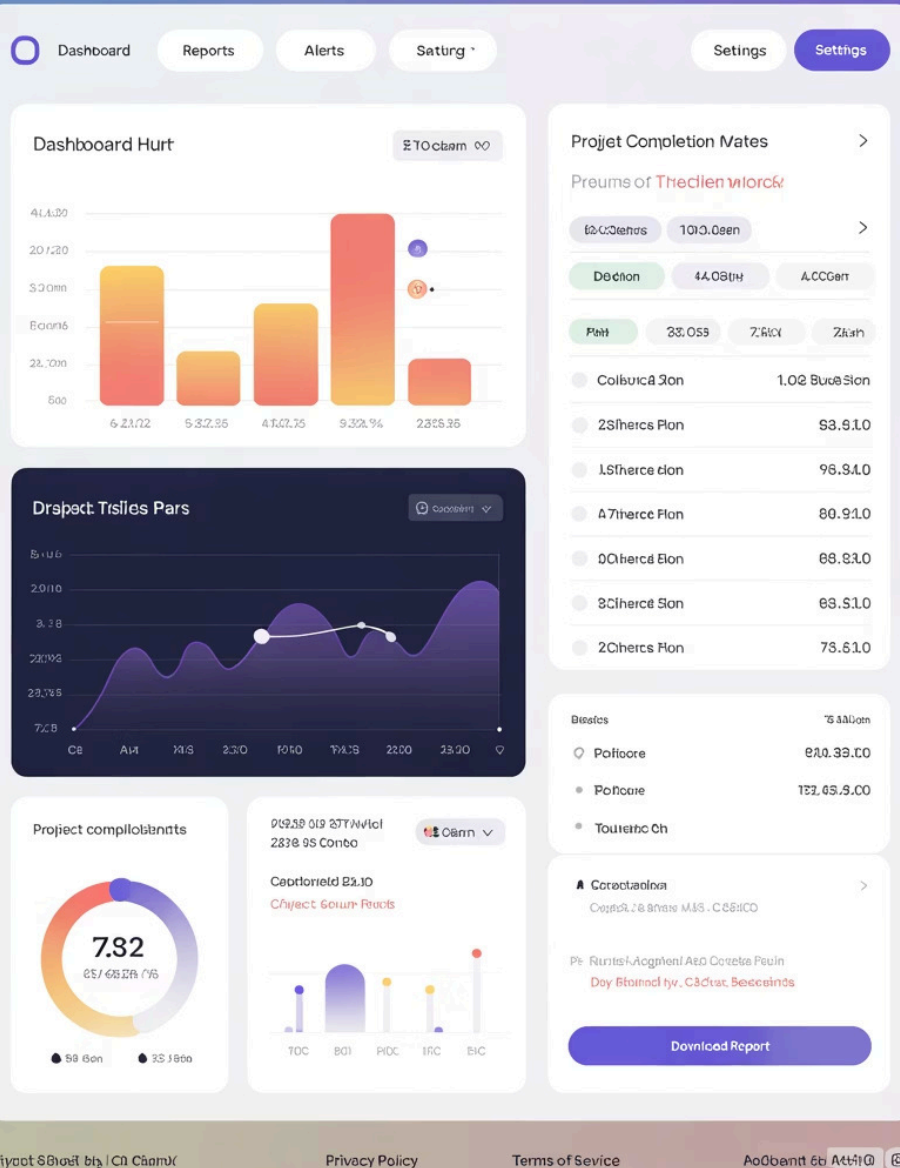# Exploratory Data Analysis & Summary Statistics

Welcome to Week 4! Today we'll discover how to extract meaningful insights from raw data through exploratory techniques and statistical measures.

What is EDA?

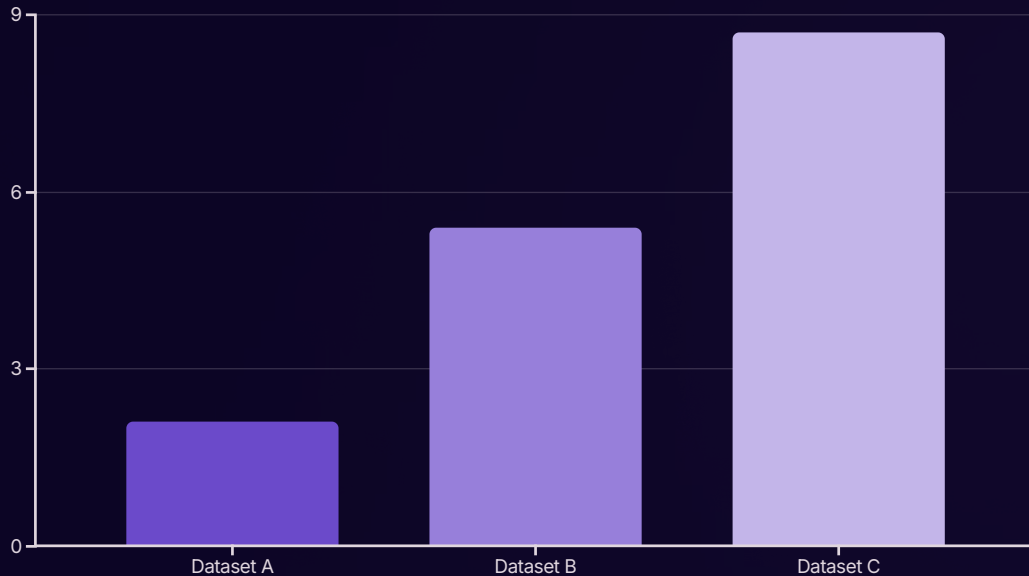Exploratory Data Analysis is the critical first step in understanding your dataset.

It helps identify patterns, anomalies, and relationships between variables.

EDA informs your choices for more advanced analysis and modeling techniques.

Bell Curve with Median and Mode

Measures of Central Tendency

Mean

The arithmetic average of all values. Sensitive to outliers. Calculated by summing all values and dividing by count.

Median

The middle value when data is ordered. Robust against outliers. Represents the 50th percentile of your dataset.

Mode

The most frequently occurring value. Useful for categorical data. A dataset may have multiple modes.

Measures of Dispersion

Key Dispersion Measures

- Variance: Average squared deviation from mean

- Standard Deviation: Square root of variance

- Range: Difference between maximum and minimum values

- IQR: Middle 50% of data values

Outlier Detection

Identify Outliers

Use boxplots, z-scores, or IQR method to detect values that fall significantly outside the expected range.
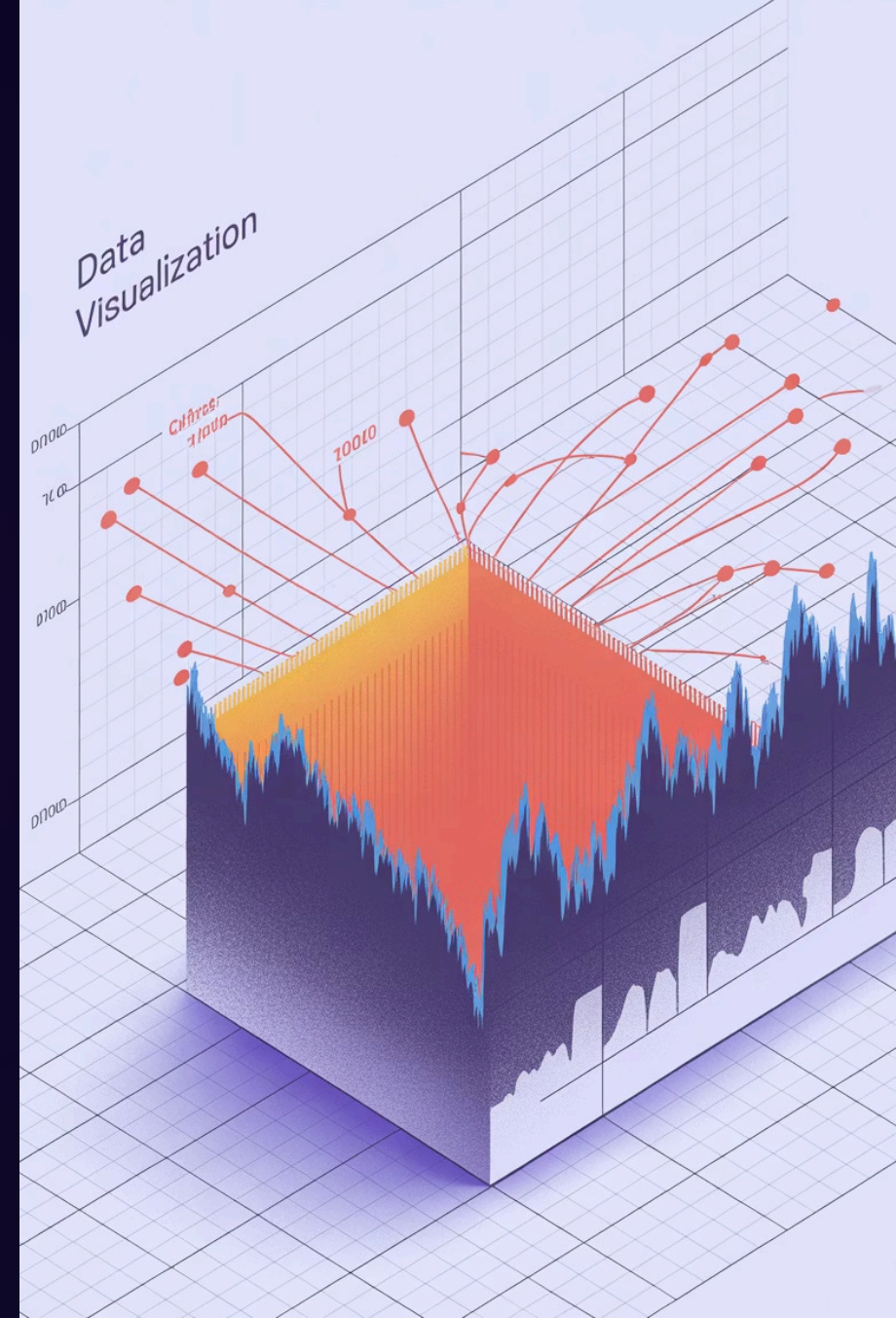
Investigate Causes

Determine if outliers are data errors or genuine but rare observations through domain knowledge.
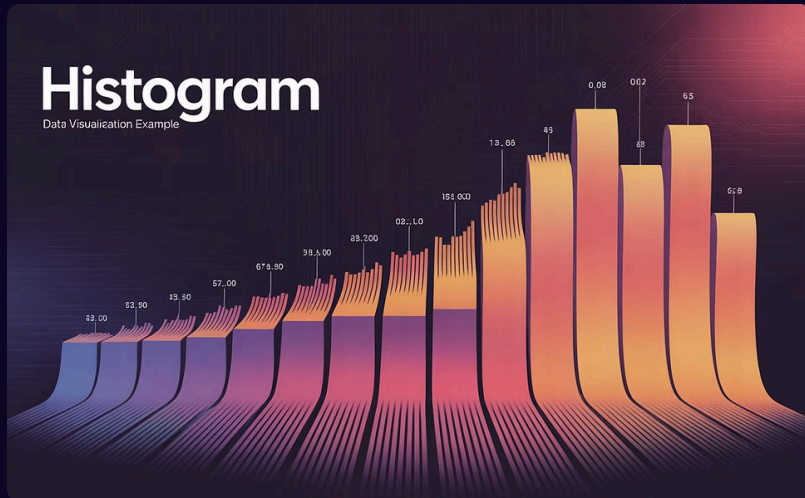
Make Decisions

Choose to remove, transform, or keep outliers based on their nature and your analysis goals.
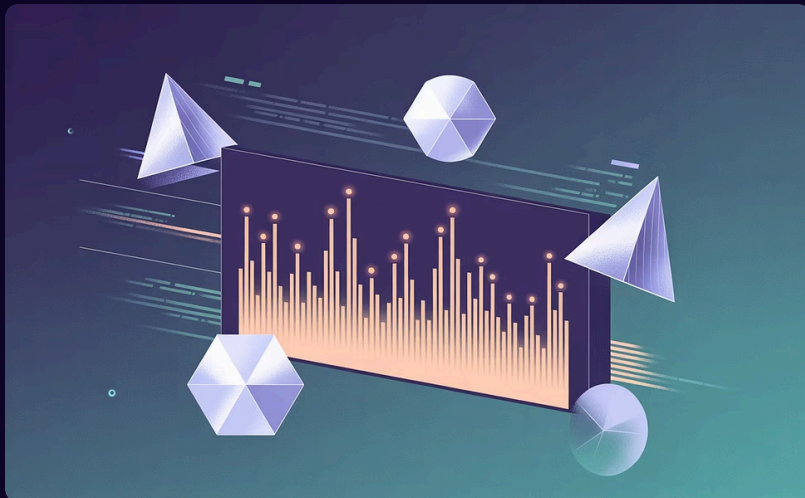

Data Visualization
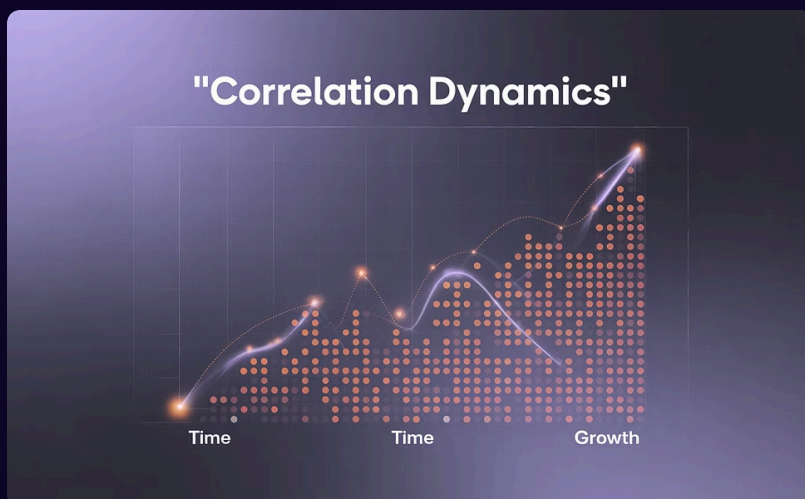
Visualization Techniques

Histograms

Show frequency distribution and overall shape of data. Reveal skewness and modality.

Boxplots

Display median, quartiles, and outliers. Perfect for comparing distributions across groups.

Scatter Plots

Reveal relationships between two variables. Help identify correlations and patterns.

Mini Project Preview

E-commerce Dataset Analysis

You'll apply EDA techniques to a real-world e-commerce dataset containing:

- Customer demographics
- Purchase history
- Product categories
- Sales performance metrics

Expected Outcomes

By completing this mini-project, you'll be able to:

- Calculate key statistical measures
- Create informative visualizations
- Identify sales patterns and trends
- Draw actionable business insights

What You'll Learn Next

1 **Week 5**

Data Cleaning & Preprocessing

2 **Week 6**

Feature Engineering

3 **Week 7**

Introduction to Machine Learning Models

4 **Week 8**

Model Evaluation & Deployment

Get ready to apply your EDA skills in future modules as we build toward complete data science workflows!