Exploratory Data Analysis (EDA)PS

Table Of Contents

Part 01

Ask Meaningful Questions

Before starting let's solve :

1. What is the main purpose of the dataset? (prediction, description, classification, etc.)
2. What business or research questions can it answer?

Part 02

Explore Data Structure

Understand the dataset's composition:

Variables & Data Types: categorical, numerical, datetime, text

Dimensions: number of rows and columns

Missing Values: how many, where, and why

Unique Values: check for duplicates or identifiers

Tools: df.info(), df.describe(), df.head()

Part 03

Identify Trends, Patterns & Anomalies

Look for relationships and distributions:

Univariate Analysis:

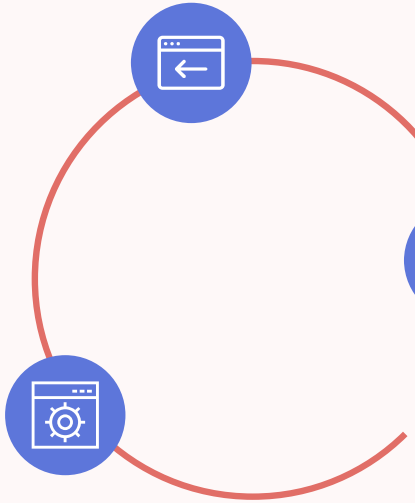
Histograms, Boxplots,

Frequency Counts

Multivariate Analysis:

Clustering, PCA,

Heatmaps



Bivariate Analysis:

Scatterplots,

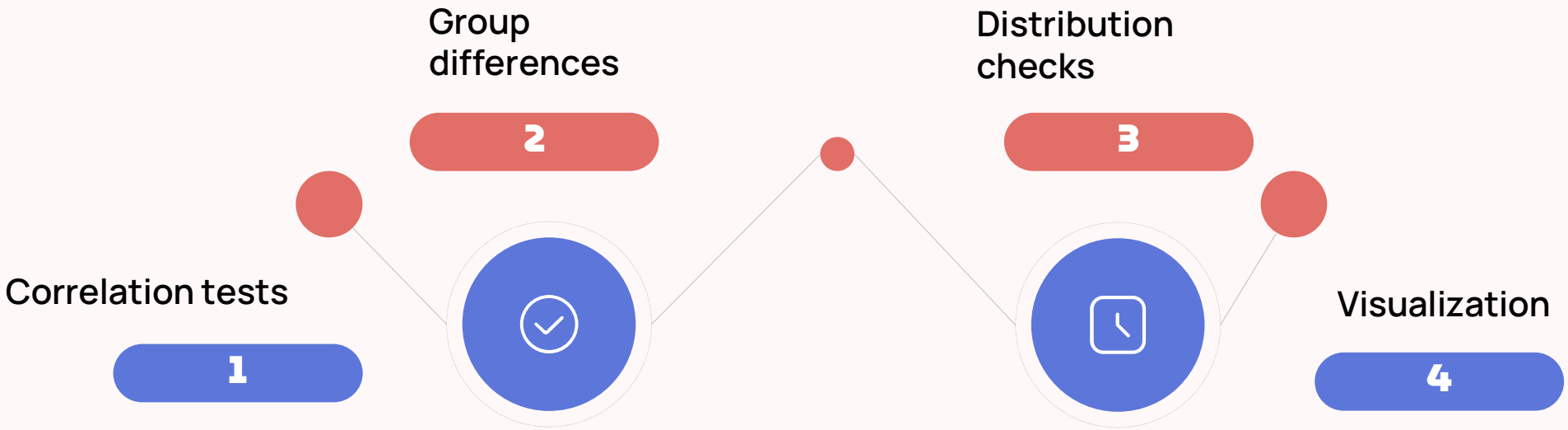
Correlation Matrices,

Group Comparisons

Part 04

Test Hypotheses & Validate Assumptions

Use statistics and visualization to confirm or reject ideas:



Detect Potential Data Issues

1. Missing or inconsistent values
2. Incorrect data types (e.g., numbers stored as strings)
3. Outliers that may distort models
4. Class imbalance in categorical variables
5. Data leakage (variables that reveal the target)

THANK YOU

BY : JAQRITI CHHABRA

Date:2026.02.20