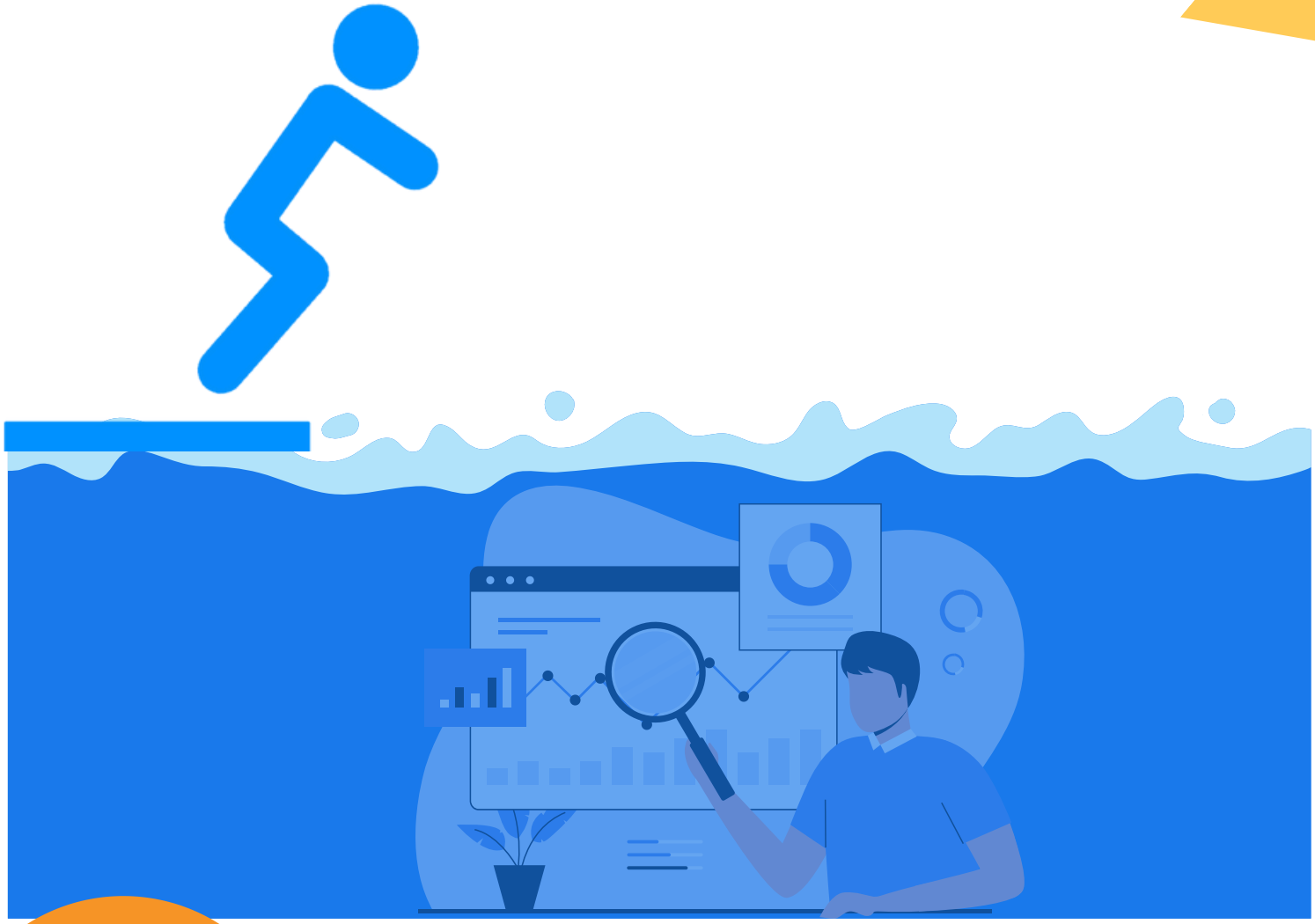
Data Analyst Perspective

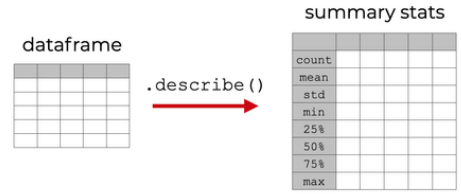
STATISTICS

WHAT ,WHY ,HOW ?????

Let’s dive in Statistics



STATISTICS AROUND US



Virat Kohli ODI Average: 58.34

Just 5% may purchase electric cars this year, most want EVs under INR 10 lakh: Survey

Car sales in India crossed 4 million units last year while 72,321 electric four-wheelers were registered, according to the survey conducted by online consumer pulse aggregator platform, LocalCircles.

WHAT
IS STATISTICS

- Statistics is the science of collecting, analyzing, interpreting, presenting, and organizing data.
- Statistics is a vital branch of mathematics that significantly impacts businesses and the world at large.
- By providing a framework for making data-driven decisions, statistics enables businesses and organizations to operate more efficiently, innovate, and adapt to changing circumstances.

- **Descriptive Statistics**: Summarizes and describes the main features of a dataset. Common measures include mean, median, mode, variance, and standard deviation.
- **Inferential Statistics**: Makes inferences and predictions about a population based on a sample of data. Techniques include hypothesis testing, confidence intervals, and regression analysis.
- **Data Collection**: Methods of gathering data, which can include surveys, experiments, observational studies, and simulations.
- **Statistical Models**: Mathematical models that represent data relationships and patterns. These can range from simple linear regression to more complex models like logistic regression and time series analysis.

WHY

STATISTICS IMPORTANT FOR A DATA PROFESSIONAL

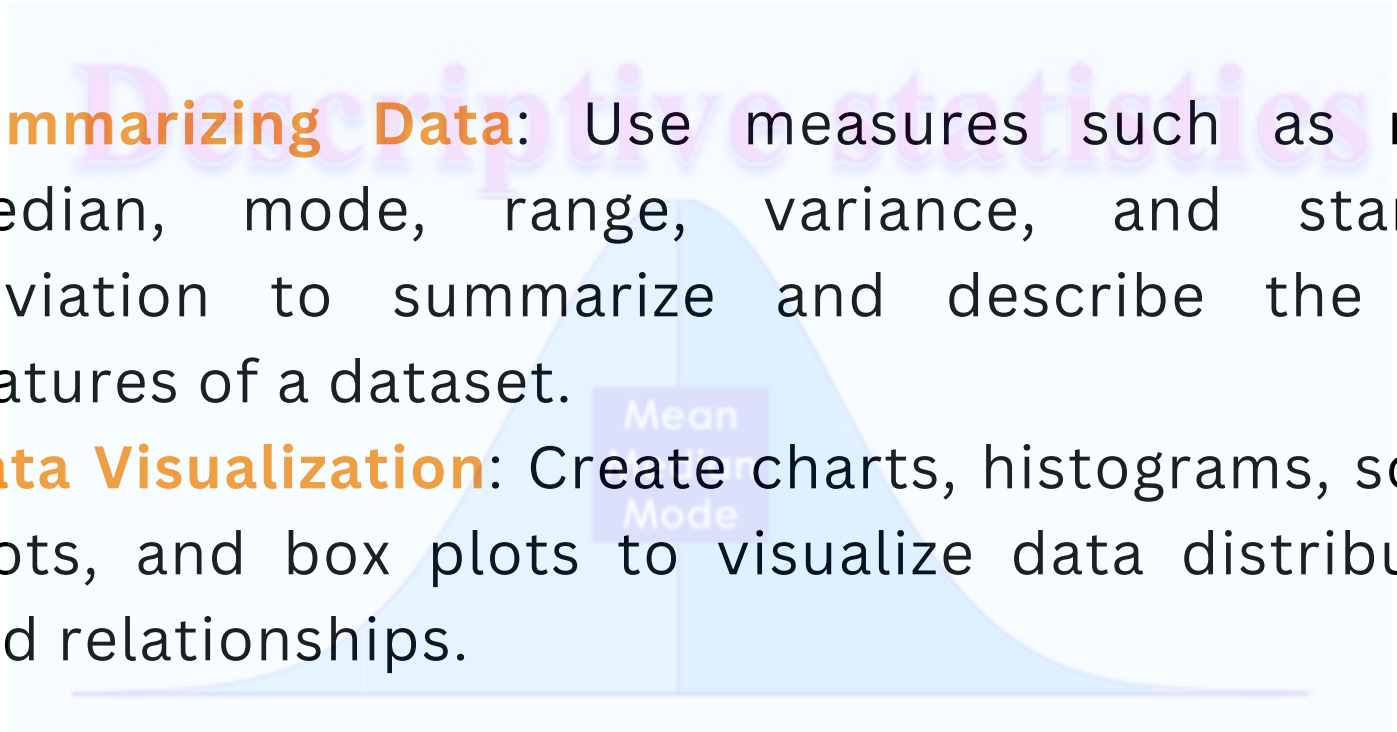
- To **analyze and interpret data accurately, make informed decisions, and provide valuable insights** that drive business strategy and operational improvements.
- Statistics is an essential component of data analytics, as it provides the necessary tools to **turn data into knowledge, and knowledge into action.**

HOW

STATISTICS CAN BE UTILZED BY DATA PROFESSIONAL

Let’s explore the **methods and strategies** through which we can utilize Statistics in the Data Domain.

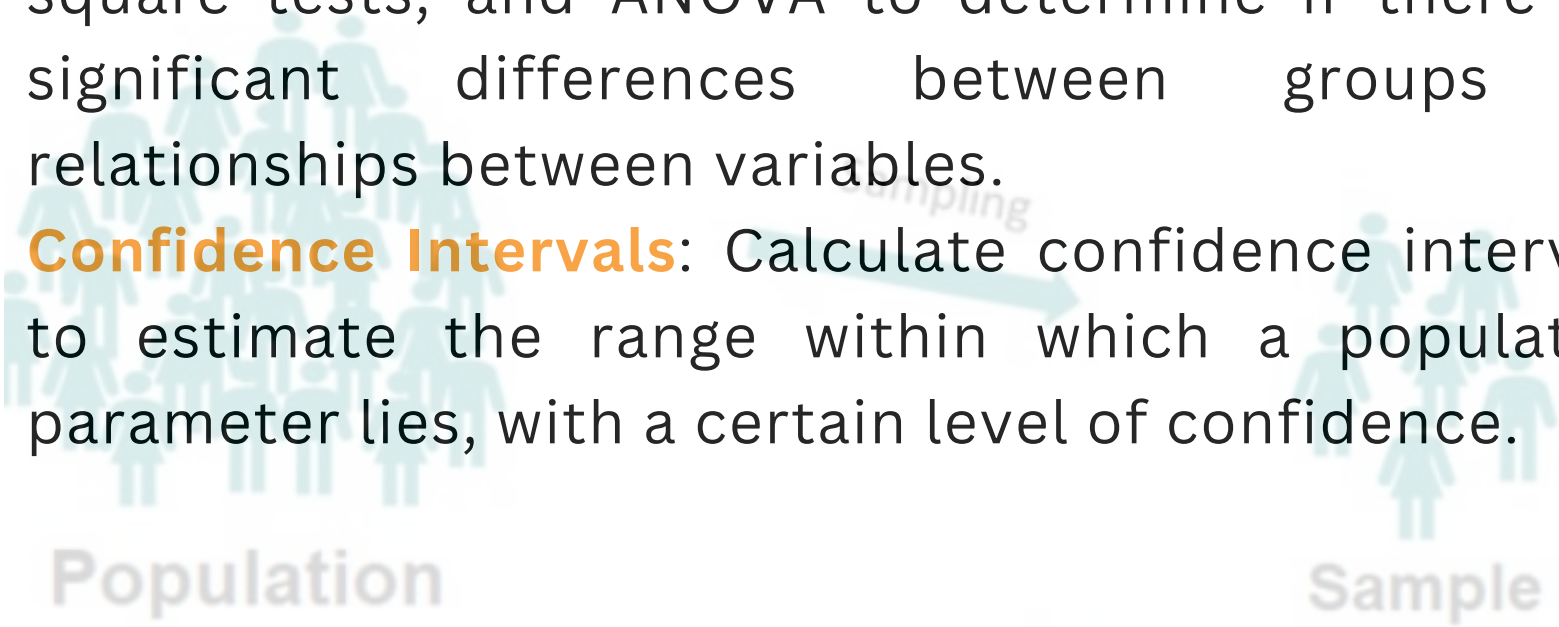
DESCRIPTIVE ANALYSIS



- **Summarizing Data**: Use measures such as mean, median, mode, range, variance, and standard deviation to summarize and describe the main features of a dataset.
- **Data Visualization**: Create charts, histograms, scatter plots, and box plots to visualize data distributions and relationships.

INFERENTIAL STATISTICS

- **Hypothesis Testing**: Perform tests like t-tests, chi-square tests, and ANOVA to determine if there are significant differences between groups or relationships between variables.
- **Confidence Intervals**: Calculate confidence intervals to estimate the range within which a population parameter lies, with a certain level of confidence.



PREDICTIVE MODELING

- **Regression Analysis**: Use linear and logistic regression to model relationships between variables and make predictions about future outcomes.
- **Time Series Analysis**: Analyze data collected over time to identify trends, seasonal patterns, and make forecasts.

DATA CLEANING AND PREPARATION

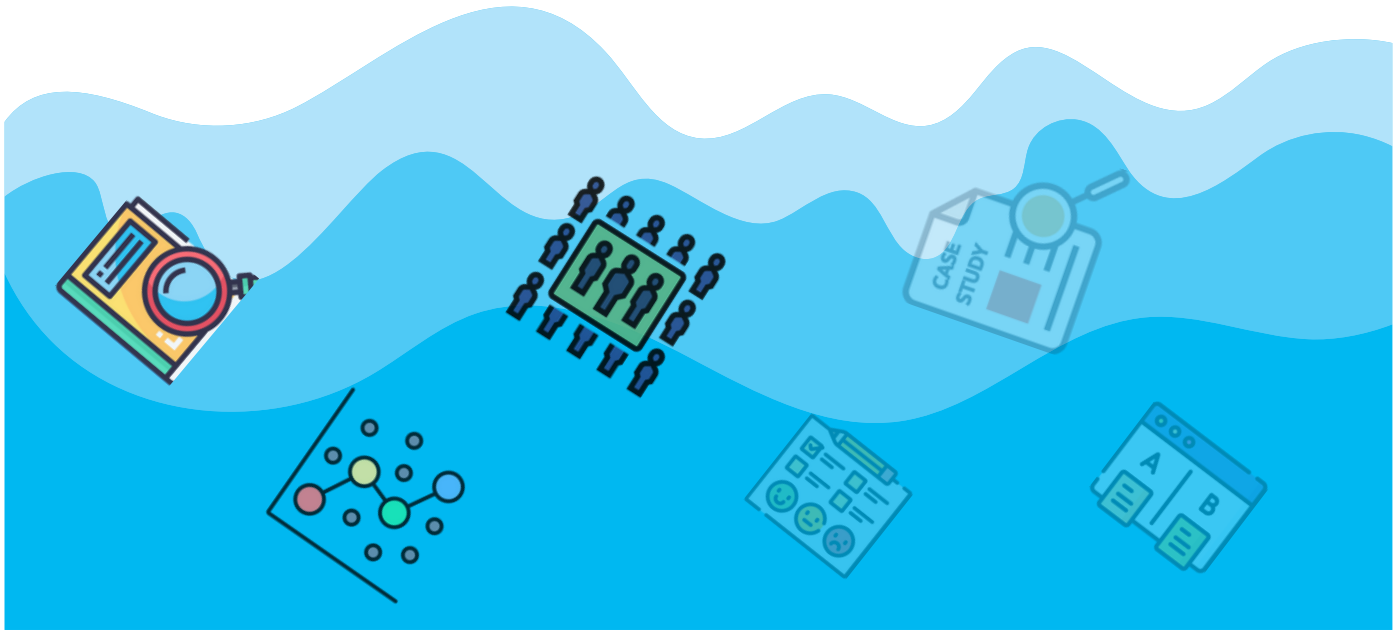
- **Outlier Detection**: Identify and handle outliers using statistical methods like z-scores and interquartile ranges (IQR).
- **Missing Data Handling**: Impute or handle missing data appropriately using techniques such as mean imputation or multiple imputation.

COMMUNICATION OF RESULTS

- **Data Storytelling**: Use statistical findings to tell compelling stories with data, supported by visualizations and clear interpretations.
- **Reporting**: Prepare comprehensive reports that summarize statistical analyses and highlight key insights and recommendations.

STAY TUNED FOR MORE!

WE'LL DEEP DIVE MORE INTO THE OCEAN OF STATSITCS



THANK YOU

Share your thoughts and feedback !!