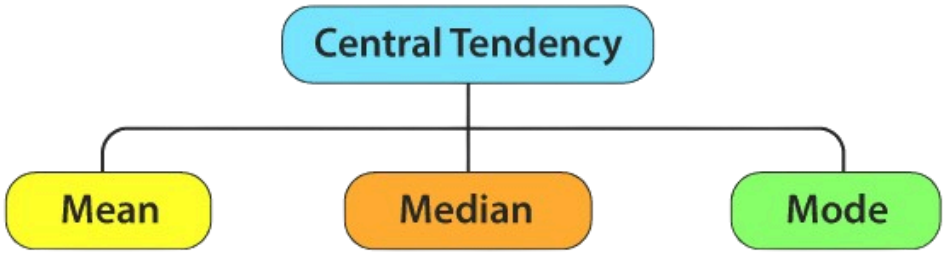
Data Analyst Perspective

STATISTICS

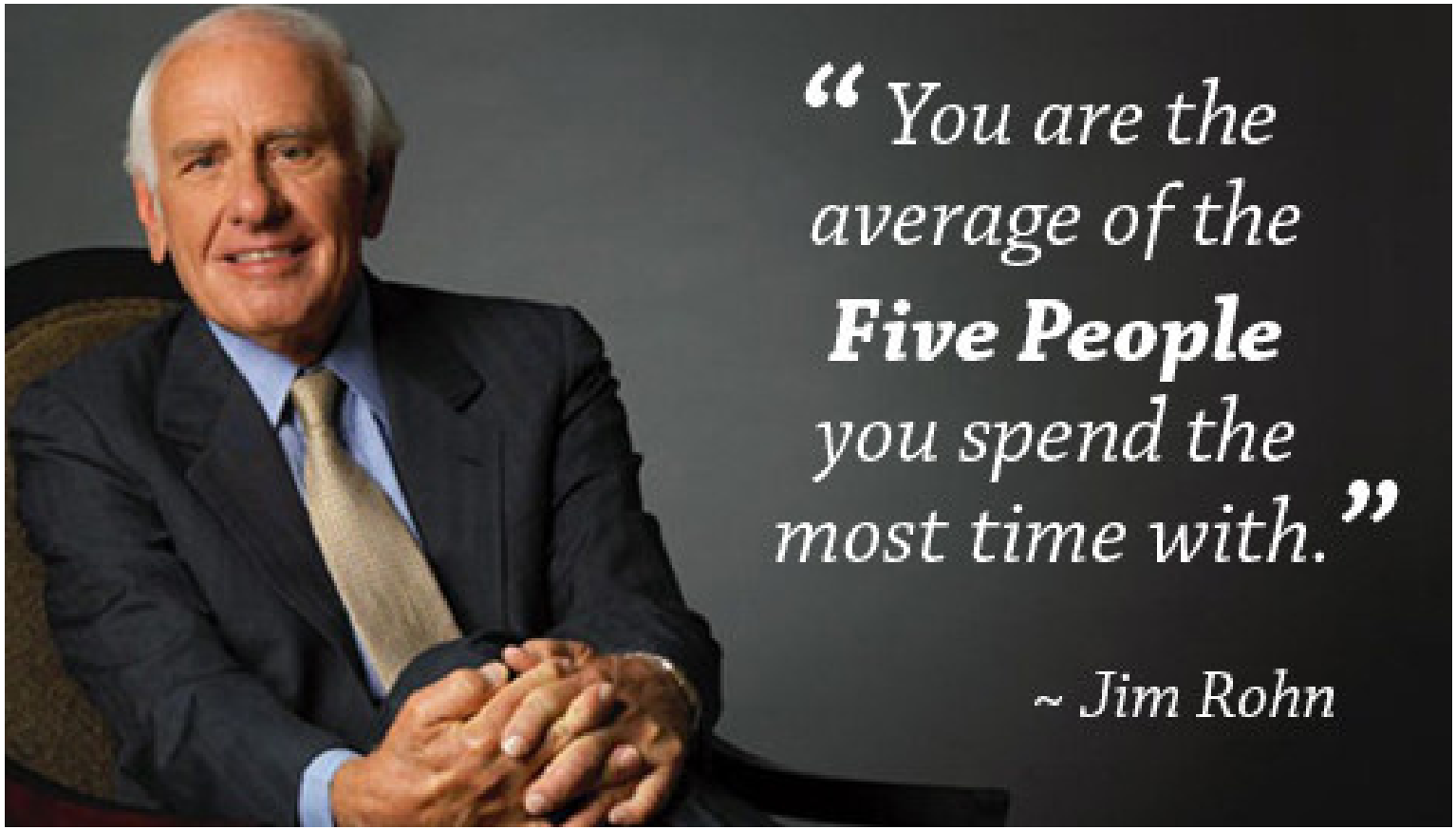


Agenda

- WHAT IS CENTRAL TENDENCY
- KEY COMPONENTS OF CENTRAL TENDENCY
- GET TO KNOW EACH COMPONENTS
- APPLICATIONS OF EACH COMPONENTS

WHAT
IS CENTRAL TENDENCY

- Central tendency is a statistical measure that identifies a **single value as representative of an entire distribution**.
- It aims to provide an accurate description of the entire data set with a single value that is the center point of the data.
- **Number crunching** perfectly encapsulates the process of calculating measures of central tendency. It involves **performing mathematical operations on a dataset to arrive at a single value** that represents the dataset's typical value.


“ You are the
average of the
Five People
you spend the
most time with.”
~ Jim Rohn

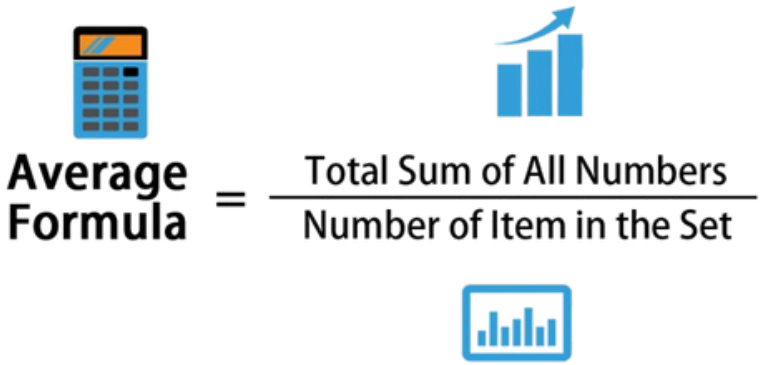
KEY COMPONENT

OF CENTRAL TENDENCY

- Mean
- Median
- Mode

MEAN

- Often referred to as the **average,** summarizes a set of data points by identifying their central value.
- It is calculated by adding up all the data points and then dividing by the number of data points.



CHARACTERISTICS
OF MEAN

- **Sensitive to Outliers**: The mean can be greatly affected by extreme values **(outliers)** in the data set.
- Consider the Mean Salary of following group and think, is mean appropriate here ???


Data Scientist


Senior Analyst

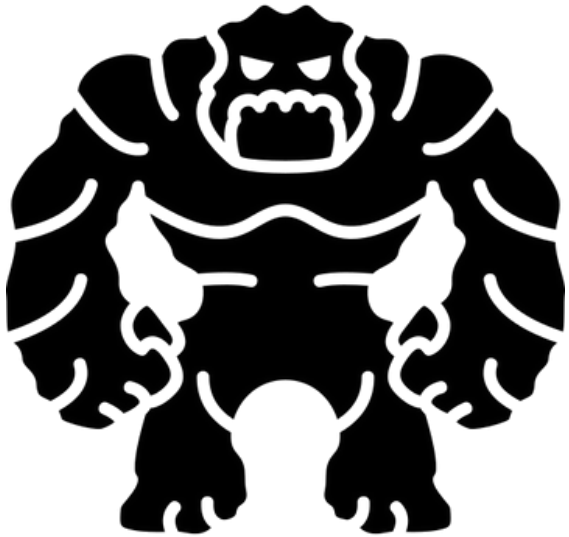

Data Engineer


Junior Analyst


CEO Meta Platforms

APPLICATION
OF MEAN

- **Statistics and Data Analysis:** The mean is widely used in various fields such as economics, engineering, and social sciences to analyze data and draw conclusions.
- **Performance Metrics:** It helps in calculating average scores, average incomes, average sales, etc.
- **Comparative Studies:** Used to compare different data sets or groups to understand overall trends.



Outliers

Mean

Median

MEDIAN

- Median is a measure of central that represents the middle value of a data set when it is **ordered from smallest to largest.**
- The median is particularly useful in understanding the distribution of data, especially **when the data set contains outliers or is skewed.**
- if *n* is even,

$$\text{median} = \frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$$

if *n* is odd,

$$\text{median} = \left(\frac{n+1}{2}\right)^{th}$$

CHARACTERISTICS OF MEDIAN

- **Resistant to Outliers:** Not affected by extreme values (outliers) in the data set. This makes it a **reliable measure** of central tendency for skewed distributions.
- **Useful for Ordinal Data:** The median can be used with ordinal data, where the values have a meaningful order but the distances between them are not necessarily equal.

APPLICATION
OF MEDIAN

- **Income and Wealth Distributions:** The median is often used to report typical incomes or wealth levels, as it is not skewed by extremely high or low values.
- **Real Estate Prices:** Median home prices provide a better indication of the typical property value in an area compared to the mean, which can be influenced by very high or low prices.

MODE

- Measure of central tendency that identifies the **most frequently occurring value(s)** in a data set.
- Mode provides a straightforward way to identify the **most common value in a data set**, offering insights into the most typical or popular items, responses, or categories.
- Unlike the mean and median, the **mode can have more than one value or none at all**, depending on the data set.

CHARACTERISTICS OF MODE

- **Only measure** of central tendency that can be used with **nominal data** (data that represents categories).
- **Resistant to Outliers:** Not affected by extreme values (outliers) in the data set.
- **Uniqueness :**
  1. UNIMODAL: A DATA SET WITH ONE MODE.
  2. BIMODAL: A DATA SET WITH TWO MODES.
  3. MULTIMODAL: A DATA SET WITH MORE THAN TWO MODES.
  4. NO MODE: A DATA SET WHERE NO VALUE REPEATS.

APPLICATION
OF MODE

- **Market Research:** Identifying the most popular product or service.
- **Education:** Determining the most common grade or score on an exam.
- **Public Health:** Identifying the most frequently occurring health condition or symptom.
- **Retail:** Finding the most sold item or the most preferred size or color.

THANK YOU

Share your thoughts and feedback !!