Day 2 of #21days Challenge

Basic Maths

Types of Data

1. **Numerical data**
    a. quantitative measurement
    ex→ weights of people
    b. Discrete data - integer based
    ex → total purchase that the customer made in a year
    c. Continuous data - infinite number of possible values
    Ex → amount of time it takes for the user to register in a website

2. **Categorical data**
    Qualitative Data - no mathematical insights
    Ex → categories of items
    Id's can be used to represent these categorical values but these categorical values have no meaning.

3. **Ordinal**
    Combo of numerical and categorical data
    Ex → movie review on scale of 1-5 stars
    Here, movie review is categorical data and rating that we provide within 1-5 gives it a mathematical meaning.

Mean

- Average number of values.
- Total sum / Number of counts

Ex → number of chocolates in each box in chocolate shop

Median

- It is used to find the middle most data.
- Determines the point from where 50% of data is more and 50% of data is less.
- Ex → 1,2,3,4,5
- Median here is 3
- Ex → 5, 6, 1, 2, 1, 2, 3
- Data must be sorted and then median must be took
- 1, 1, 2, 2, 3, 5, 6
- Median is 2

Mode

- Repetitive value
- Used with categorical, ordinal, and discrete data which doesn't have much central tendency.

Ex → preferred flavour of ice cream

Standard Deviation

- Standard Deviation is a measure of how spread out numbers are.
- measures the dispersion of a dataset relative to its mean.
- calculated as the square root of variance by determining each data point's deviation relative to the mean.

Variance

- statistical measurement of the spread between numbers in a data set
- measures how far each number in the set is from the mean and thus from every other number in the set.

Population vs. Sample

Using Sample of data in place of entire data set(population) then sample variance must be used

Population

entire group that you want to draw conclusions about.

Sample

- specific group that you will collect data from.
- size of the sample is always less than the total size of the population
- A sample of the population is used in research, as it is easier and cost-effective to process a smaller subset of the population rather than the entire group.

Probability Density Functions

- statistical expression that defines a probability distribution (the likelihood of an outcome) for a discrete random variable (e.g., a stock or ETF) as opposed to a continuous random variable
- probability density plots are used to understand the overall distribution of data.
- used to understand data distribution for a continuous variable and to know the likelihood (or probability) of obtaining a range of values that the continuous variable can assume.

Normal distribution

Gives the probability of a data point falling within some given range of a given value.

Covariance

statistical tool that is used to determine the relationship between the movement of two asset prices

Bayes Theorem

- probability of A given B, is the probability of A times the probability of B given A over the probability of B.
- the probability of something that depends on B depends very much on the base probability of B and A.

Ex → Vaccine testing