ASSIGNMENT – 1

Name – Hiyansh Sharma Roll no – 2401730263 Course – BTech CSE AI ML Section – A

Q1. The prices of five bikes in a shop are: ₹12,000, ₹15,000, ₹18,000, ₹20,000, ₹50,000

The solutions below are calculated using NumPy.

- **Mean Price:** The mean price is the average of all the prices. **np.mean([12000, 15000, 18000, 20000, 50000]) Mean = ₹23,000**
- **Median Price:** The median is the middle value of the sorted data. **np.median([12000, 15000, 18000, 20000, 50000]) Median = ₹18,000**
- **Advertising Strategy:** The shop owner should advertise the median price of ₹18,000. This is because the mean is heavily influenced by the ₹50,000 outlier, making the average price seem much higher than what most bikes cost. The median offers a more accurate representation of a typical bike's price**.**

Q2. Describe the relationship between a histogram's shape and the values of its mean, median, and mode in a left-skewed distribution.

In a left-skewed distribution, the tail of the data is longer on the left side of the histogram. This causes the relationship between the measures of central tendency to be:

Mean < Median < Mode

The mode is at the peak of the distribution, the median is to the left of the mode, and the mean is pulled further to the left by the lower values in the tail.

Q3. The following data represents the exam scores (out of 100) for 10 students: 85, 92, 78, 90, 85, 67, 88, 95, 85, 72

The solutions below are calculated using NumPy .

- **Sorted Scores: np.sort([85, 92, 78, 90, 85, 67, 88, 95, 85, 72]) Scores (sorted) = [67, 72, 78, 85, 85, 85, 88, 90, 92, 95]**
- **Mean: np.mean(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72]) Mean = 83.7**
- **Median: np.median(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72]) Median = 85.0**
- **Mode: The mode is the value that appears most frequently, which is 85.**
- **Distribution Shape: The distribution is left-skewed because the mean () is less than the median ().**
- **Five-Number Summary:**

ASSIGNMENT – 1

Name – Hiyansh Sharma Roll no – 2401730263 Course – BTech CSE AI ML Section – A

- **Minimum: np.min(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72]) = 67**
- **Q1 (25th percentile): np.percentile(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72], 25) = 79.75**
- **Median: np.median(...) = 85.0**
- **Q3 (75th percentile): np.percentile(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72], 75) = 89.5**
- **Maximum: np.max(([85, 92, 78, 90, 85, 67, 88, 95, 85, 72]) = 95**

Q4. Differentiate between a population parameter and a sample statistic with an example of each.

- A population parameter is a numerical measure that describes a characteristic of an entire population. It is a fixed, but often unknown, value. For example, the average height of all adults in a country is a population parameter.
- A sample statistic is a numerical measure that describes a characteristic of a sample, which is a subset of the population. It is a variable value calculated from the sample data. For example, the average height of a randomly selected group of 1,000 adults from that same country is a sample statistic**.**

Q5. A small company has 5 employees with the following annual salaries: $45,000, $50,000, $55,000, $60,000, $250,000

The solutions below are calculated using NumPy.

- **Mean Salary: np.mean([45000, 50000, 55000, 60000, 250000]) Mean = $92,000**
- **Median Salary: np.median([45000, 50000, 55000, 60000, 250000]) Median = $55,000**
- **Advertising Strategy:** The company would likely advertise the mean salary ($92,000) to attract new hires because it is much higher. A potential employee, however, would prefer to see the median salary ($55,000) as it is a more realistic measure of a typical salary in the company.