</talentlabs>

Data Analysis
Assignment Chapter 2

Instructions

1. You can take help from the lecture notes to revise the concepts that we have covered
2. Choose the best suitable answer and submit the word document
3. Please submit the assignment through TalentLabs Learning System.

Question 1. (5 points):
There are different methods of summarizing data. In this exercise you have to identify the methods which best fit the following scenarios:

#	Scenario	Best Method
E.g.	Average Salary of software engineers in the market	Average
1	Highest and lowest marks of students in the subject of Math	Max - Min
2	Outlier product sold by a company	Interquartile Range
3	Most trending topic on Internet	Mode
4	Average exam score with different weight for different subjects	Weighted Mean
5	Number of products sold by product category	Group by

Question 2:

Suppose that a marketing firm conducts a survey of 100 households to determine the average number of Air Conditioners each household owns. Every household in the sample has at least one Air Conditioner and no household has more than four. Find the expected number of Air Conditioners per household. (Hint: Weighted Mean)

Number of refrigerators per Household	Number of Households
1	47
2	18
3	23
4	12

The expected number of Air Conditioners per household is 2.

Question 3:

Consider a data set having the number of toys owned by different kids in a society. Find the median number of toys found in the society

Number of toys owned by each kid = [1, 2, 7, 6, 4, 3, 3, 8, 7, 6]

The median number of toys found in the society is 5.

Question 4 (1 point):
Measures of Spread is a type of data summary, and it helps you understand the spread of your data set. In the given multiple choices you need to identify the methods which do not belong to Measures of Spread.

☐ A – Interquartile Range
☐ B – Weighted Mean
☐ C – Data Skewness
☐ D – Range
☐ E – Median

Question 5:
Consider a store that sells different categories of products

Skin care	Frozen Foods	Imported Cookies	Chocolates	Clothes	Electronics	Stationery

Now they have a dataset of all the sales (Number of sales of each product every day, with categories information) that they made **in one year**. If they want to get some insights from that data:

a) What are the metrics that can be generated from the dataset of products sold using different data aggregation methodologies? List 2 of them. (2 points)
b) And explain the significance of insights generated using the data aggregation methods in part(a)? (1 point)

a)

1. Sum of sales per product category
2. Maximum and minimum sales per product category

b)

1. **Sum of sales per product category**: Helps to identify the total contribution of each category to overall sales, providing insights into which categories are the most profitable.
2. **Maximum and minimum sales per product category**: Helps in understanding the range and variability in sales within each category, which can assist in inventory management and marketing strategies.

Question 6 (2 points):
There are several ways of Data Summary, identifying which of the following falls under the methods of Data Aggregation.

☐ A – Groupby
☐ B – Mode
☐ C – Unique Values
☐ D – Count
☐ E – Quartiles

Question 7 (2 points):
Data Skewness is the difference of some values from the majority of other values in a dataset. It brings the asymmetry. As we know that there are two types of Data Skew; can you formulate an example of Positive and Negative Skew in Data ?

Skewness	**Example**
Positive Skew	Income distribution in a city where most people earn low to moderate income, but a few people earn significantly higher, creating a long tail on the right.
Negative Skew	The age of retirement in a company where most employees retire around the same age, but a few retire much earlier, creating a long tail on the left.

Question 8 (2 points):
Let's take an example of a data set recorded at different branches of a bank where they recorded the time taken by an ATM for doing one transaction. The Maximum and minimum of the IQR of data set is as follows:
Maximum of IQR = 12 min
Minimum of IQR = 3 min
Now identify the Outlier ATM's involved in the experiment
Time take by ATM (in minutes) = [7,5,3,9,6,22,10,11,4,2]

2 and 22 minutes