Advanced Analytics Assignment 01

248518 Gaurav B Raskar

Q.1.

Data analysis is the process of evaluating, cleaning, and interpreting data in order to extract relevant insights and perform decision-making. It involve applying statistical, mathematical, and visualization tools to identify patterns.

Q.2.

Data visualization is the process of showing data in charts, graphs, or images to help people understand recognize trends.

* Bar chart - used to analysis of dimensions with respect to measure.
* Pie chart - Divides circle in the percentage of measure according to each dimension object.
* Bubble chart – Size of circle represents Measure by Dimension.
* Tree Map – Similar to Bubble chart, Uses Area of Rectangle. Cannot be sorted.
* Line Chart – used for Time-Series data, to represent Continuous or Discrete Data.
* Histogram – Measures are distributed into bins with specified size.

Q.3.

Mean = Sum Observations / Total Observations

Q.4.

Advantage: Mean gives value that represent all data.

Disadvantage: Cannot help to find comparative Consistency.

Q.5.

Advantage: Gives middle value which divides data in equal parts

Disadvantage: Nominal data cannot have Median.

Q.6.

Advantage: Gives value with max frequency.

Disadvantage: Can have multiple values if equal Frequency.

Q.7.

Continuous variable can take any value within a range.

Discrete variable can only take specific, separate values.

E.g. Continuous: height, time, distance, temperature, weight

E.g. Discrete: number of students, number of books, shoe size, number of cars, number of phone calls.

Q.8.

A Sample is a subset drawn from the Population, whereas the Population consists of every member of a group.

Population: Total Blood in Body. Sample: 10ml blood drawn with syringe.

Population: Total Voters. Sample: Data collected for Exit Poll.

Population: Total No. of Students. Sample: 10 Selected for Interview.

Population: Total No. of Images for Classification. Sample: Few used to Train Model.

Population: Total Soil in Land. Sample: few Grams selected for testing.

Q.9.

Descriptive Analytics: Analyzes past data to understand what happened.

Diagnostic Analytics: Explores data to understand why something happened.

Predictive Analytics: Uses historical data to predict future outcomes.

Prescriptive Analytics: Recommends actions to optimize outcomes.

Q.10.

1. 89,6,90,34,65,234,8,3000,7,567,6,2,45,20

Mean - 298.071

Median - 39.5

Mode – 6

2. 5677, 60, 22, 34,6500,23, 869,67,900,1,2,6,1,70,1

Mean - 948.867

Median - 34

Mode – 1

Q.11.

1) Decide on the objective.

2) Set measurement priorities.

3) Data Collection.

4) Data Cleaning

5) Analysis of Data

6) Interpreting the reults.

Q.12.

-Nominal Scale => Numeric or Character data used to represent individuals or categories.

E.g. Gender, Roll. No. etc.

Can only calculate Mode, Not Mean/Median

-Ordinal Scale=> Used to represent relations between individuals but, cannot tell difference/ratio in between.

E.g. Ranks, Grades, etc.

Can calculate Median only.

-Interval Scale=> Assigned numbers are such that the difference in numbers is valid but not ratios.

E.g. Ratings, Temperature, etc.

Can calculate Mean, Median, Mode.

-Ratio Scale=> Scale which contains Absolute Zero.

E.g. Marks, Price, Profit, etc.

Can perform all operations.

Q.13.

Skewness and kurtosis are measured to represent the distribution of data.

Skewness helps tells if sample space biased in either side of median or not.

Left Skewed data -> -ve Skewness

Right Skewed data-> +ve Skewness

Symmetric Distribution-> 0 Skewness.

Kurtosis tells the uniformness of spread of data across the median on distributions.

LeptoKurtic-> +ve Kurtosis

Mesosurtic-> Normal Distribution

Platykurtic-> -ve Kurtosis

Q.14.

**34,56,12,34,32,89,70,65,45,678,90,890,760,55**

|  |  |
| --- | --- |
| Mean Deviation | 243.4897959 |
| Standard Deviation | 311.4891916 |
| Kurtosis | 0.910770753 |
| Skewness | 1.621173806 |