National University of Computer & Emerging Sciences, Lahore
Department of Computer Science (Fall 2024)

CS4048 – Data Science
Quiz#1 CLO#1

Roll No: [redacted]

Name & Section: [redacted]

Time allowed: 20 minutes

Maximum Marks: 20

34

Consider the following dataset and answer the following questions:

Employee ID	Employee Name	Designation	Department	Age	Salary	Employment Status
101	John Smith	Associate (entry-level)	Sales	32	$60,000	Full-Time
102	Jane Doe	Executive (junior-mid level)	Marketing	28	$55,000	Full-Time
103	David Johnson	Manager (top-level in this context)	Human Resources	45	$75,000	Full-Time
104	Lisa Brown	Senior Executive (mid-level)	Finance	38	$65,000	Full-Time
105	Michael Lee	Analyst (mid-senior level)	Sales	29	$58,000	Part-Time
106	Emily Wilson	Senior Analyst (higher seniority)	Marketing	31	$56,000	Full-Time
107	Susan Turner	Senior Executive (mid-level)	Finance	35	$63,000	Part-Time

Question#1: Recognize and name each attribute's data type given in the dataset.

- Employee ID - int - discrete quantitative data
- Employee Name - string - nominal data
- Designation - string - ordinal data
- Department - string - nominal data
- Age - int - discrete (interval data) ratio
- Salary - string - int - discrete (currency $ is concatenated to the number at the time of inputting in table)
- Employement Status - string - nominal data

National University of Computer & Emerging Sciences, Lahore
Department of Computer Science (Fall 2024)

Question#2: **Identify** the outliers in the dataset (any one **valid** attribute) using Z-score analysis. The following threshold should be considered:

$$|Z| > 2$$

[redacted] (Assuming it is a subset of actual dataset)

Age

$$-2 < z < 2$$ [redacted]

$$S.D = \sqrt{\frac{\sum (x - \bar{x})}{n-1}}$$ ✓

Mean = 34

Mean Formula?

$$S.D = \sqrt{\frac{212}{6}}$$ ✓

$$S.D = 5.9442$$ ✓

$$z\text{-score (for 32)} = \frac{x - \mu}{\sigma}$$

$$= \frac{32 - 34}{5.9442}$$

$$= -0.34$$ ✓

z-score (for 28) = −0.34 ✗ −1.0094

z-score (for 45) = 1.85 ✓

z-score (for 38) = 0.67 ✓

z-score (for 29) = −0.84 ✓

z-score (for 31) = −0.50 ✓

z-score (for 35) = 0.17 ✓

Write values upto 4 decimal places like −0.3365

There are no outliers in this dataset for age attribute. ✓