Datascience Scenario based AnsS1:

1.Calculate Q1 and Q3-

Q1(25th Percentile)-25% of data below this value

Q3(75th Percentile)—75%of data below this value

Formula:

Q1=0.25*(n+1)Q1=0.25\times(n+1)Q1=0.25*(n+1)Q3=0.75*(n+1)Q3=o.75\times(n+1)Q3=0.75*(n+1)

Q1=4th value-40

Q3=11th value-75

Find IQR:

Interquartile Range(IQR) measures the spread of the middle 50% of data

Formula:

IQR = Q3-Q1 ,75_40=35IQR=Q3-Q1=75-40=35

Detect Outliers:

Lower Bound=Q1-1.5xIQR\text{LowerBound}=Q1-1.5\timesIQRLower Bound=Q1-1.5xIQR

Upper Bound=Q3+1.5xIQR\text{Upper Bound}=Q3+1.5\timesIQRUpper Bound=Q3+1.5xIQR

Lower Bound=40-1.5*35=12.540-1.5*35=-12.5\time35=-12.540-1.5*35-no lower outliers

Upper Bound=75=1.5x35=127.575+1.5|times 35=127.575+1.5x35=127.5-no Upper outliers

Conclusion:No outliers in this dataset.

2.Scenario:Scores are[45,50,55,60,62,63,65,90,95}

Mean=sum of scores/totalnumber of score

45+50+55+60+62+63+65+90+9510=65.5\frac{45+50+55+60+60+62+63+65+90+95}{10}=65.51045+50+55+60+60+63+65+90+95=65.5

Median=Middle value=(60+62)/2=61(60+62)/2=61(60+62)/2=61

Mode:60(occurs twice)

Median is a better measure because high outliers90 and 95,skew themean.Median is not affected by extreme values,better student performance

3.Grocery store manager tracks how many customers visit store daily for a month

Create frequency distribution table for this data

4.Real estate model has three variables

House size

Numberof Rooms

Number of bathrooms

Logic:

CalculateVIF

VIF>10 indicates multicollinearity

Ans:high VIF –variables are correlated,impacting model auuracy

5Scenario: Check Medicine works

Make guess Hypothesis

H0:The medicine doesNot lower blood pressure

H1:The medicine lowers blood pressure

T-test

Find P-value

If P-value<0.05,it means medicine works

Ans:if P-value is small,the medicine is effectivw

6.Find unusual spike in sales

calculaturIQR

Identify outliers using formula

Outliers (Data<Q1-1.5xIQR)or(Data>Q3+1.5xIQR)outliers=(Data<Q1-1.5\timesIQR)\text{or}(Data>Q3+1.5\timesIQR)outliers=(Data<Q1-1.5xIQR)or(Data>Q3+1.5xIQR)

7.Understanding Customer Satisfaction

Find the Mode to see most commen rating

Calculate Mean and Median for further insights

If most rati gs are 4 or 5,high satisfaction