Interpretation Document

Outlier Removal (IQR Method)-

1.**Salary**:

- The current boxplot indicates a more symmetric distribution of salaries, that previously existing right-skewness has been corrected by removing outliers.

- The median salary is approximately ₹50,000, indicating that half the individuals earn below ₹50,000 and half earn above.

2.**Purchase Amount**:

- The median purchase amount is around ₹20,000, which represents the central tendency of consumer spending in the dataset.

- The interquartile range (IQR) spans approximately from ₹15,000 to ₹25,000, suggesting that 50% of the customers made purchases within this range.

3.**Satisfaction Score:**

- The median satisfaction score is approximately 7, indicating that customers generally report above-average satisfaction.

- The interquartile range (IQR) spans roughly from 5 to 9, meaning 50% of the customers have satisfaction scores within this range.

Plot Distributions:

1. **Age**:

- The age distribution is multimodal, with noticeable peaks around 18–20 years, 40–45 years, and 55–60 years, indicating clusters of individuals in these age groups.

- The highest frequency occurs at around 18–20 years, suggesting a significant presence of younger individuals, possibly entry-level employees or recent graduates.

- The distribution is fairly spread out, covering a wide range from approximately 18 to 60 years. showing diverse age representation in the data.

2. **Salary:**

- Most salaries are clustered between ₹40,000 and ₹60,000, indicating this is the common earning range in the dataset.

- The peak mode occurs around ₹50,000, suggesting that this is the most frequent salary level.

- There is a gradual decline in frequency beyond ₹60,000, with fewer employees earning higher salaries (above ₹80,000), indicating fewer high earner.

3. Experience Years:

- The distribution of Experience Years is fairly uniform, indicating a diverse workforce with varying levels of experience from 0 to 40 years.

- There are noticeable peaks around 10, 20, and 30 years, suggesting larger employee groups at these career stages.

- The experience distribution shows no strong skew, reflecting a balanced mix of early-career, mid-career, and senior professionals.

4. Satisfaction Score:

- The most frequent satisfaction scores fall between 6 and 8, suggesting that a majority of individuals report moderate to high satisfaction.

- The peak mode is around 7, indicating that many people rated their satisfaction around this value.

- There are fewer respondents with scores below 4 or above 10, suggesting that extreme dissatisfaction or over-satisfaction is rare in the dataset**.**

5. Purchase Amount:

- The majority of purchase amounts fall between ₹15,000 and ₹25,000, indicating this is the typical spending range for most customers.

- The peak frequency occurs around ₹20,000, which serves as the modal purchase amount.

- The spread is moderate, with values ranging roughly from ₹8,000 to ₹34,000, showing a decent diversity in purchasing behavior.

- Very high or very low purchase amounts are relatively rare, as indicated by the thin tails on both sides

6. Tenure:

- The distribution of tenure is fairly uniform, with frequencies spread evenly across all tenure values from 1 to 9 years.

- The employee base is well-distributed across different tenure lengths, indicating consistent hiring or retention practices over time.

Boxplots for Outlier Visualization:

1.**Salary**:

- The current boxplot indicates a more symmetric distribution of salaries, suggesting that previously existing right-skewness has been corrected by removing outliers.

- The median salary is approximately ₹50,000, indicating that half the individuals earn below ₹50,000 and half earn above.

2.**Purchase Amount**:

- The median purchase amount is around ₹20,000, which represents the central tendency of consumer spending in the dataset.

- The interquartile range (IQR) spans approximately from ₹15,000 to ₹25,000, suggesting that 50% of the customers made purchases within this range.

3.**Satisfaction Score:**

- The median satisfaction score is approximately 7, indicating that customers generally report above-average satisfaction.

- The interquartile range (IQR) spans roughly from 5 to 9, meaning 50% of the customers have satisfaction scores within this range.

Count plot for Columns:

1. **Gender:**

- The distribution of gender in the dataset is nearly balanced, with a slightly higher count of males compared to females.

- This indicates that the dataset has a relatively even representation of both genders, which is beneficial for analyses that require gender diversity.

2. **Department:**

- The Sales department has the highest number of employees, slightly ahead of the others.

- The Finance, Marketing, HR, and IT departments have fairly balanced employee counts, with only minor differences.

3. **City:**

- Delhi has the highest number of employees among the listed cities.

- The other cities—Kolkata, Chennai, Mumbai, and Bangalore—have fairly similar employee counts, though slightly fewer than Delhi.

4. **Attrition:**

- The number of employees who left the organization Yes is nearly equal to those who did not No.

- This suggests a high attrition rate, with approximately half of the workforce experiencing turnover.