**Assignment 2 : Outliers and Percentiles in Data Science**

**Question 1: Identification of Outliers**

* **Dataset**: [5, 7, 9, 10, 12, 13, 15, 18, 19, 21, 23, 50]
* **Task**:
  1. Calculate the first quartile (Q1), third quartile (Q3), and the interquartile range (IQR) for the dataset.
  2. Use the IQR method to determine which data points, if any, are outliers.
  3. Create a box plot using any software or programming language of your choice to visually inspect and confirm the presence of any outliers.
* **Deliverables**: Answers to calculations, identification of outliers, and a snapshot or code of the box plot.

**Question 2: Understanding Outlier Impact**

* **Scenario**: In a dataset of exam scores from 100 students, one student scored 100 while the rest scored between 40 and 60.
* **Task**:
  1. Discuss the potential impact of the outlier score on the mean and median of the dataset.
  2. How would the outlier affect the interpretation of the dataset’s central tendency? Provide your rationale.
* **Deliverables**: A detailed written analysis addressing the effects of the outlier.

**Question 3: Practical Application of Percentiles**

* **Dataset**: [120, 130, 145, 160, 170, 180, 190, 200, 210, 220, 230, 240, 300]
* **Task**:
  1. Calculate the 25th, 50th, and 75th percentiles.
  2. Discuss how these percentiles could be useful in determining the compensation structure in a company where these values represent salaries (in thousands) of employees.
  3. Suggest how outliers might influence decisions if the highest salaries are considered outliers.
* **Deliverables**: Calculation of percentiles, a discussion on their utility in compensation structuring, and an analysis of the influence of outliers.