* **Software used:**

Anaconda

* Python with pandas, sklearn , seaeborn and other libraries
* Jupyter notebook

**Reason**:

* Python is chosen for its versatility in data manipulation, statistical analysis, and machine learning. It offers powerful libraries like NumPy, Pandas, and Scikit-learn, essential for data preprocessing, modeling, and evaluation.
* Jupyter provides an interactive environment ideal for data exploration, visualization, and iterative development. It supports inline plotting and markdown cells, facilitating clear documentation and analysis.
* Statistical Analysis Techniques and Algorithms:
  + **Descriptive Statistics**:
    - **Technique**: Calculation of mean, median, mode, and standard deviation.
    - **Reason**: Descriptive statistics summarize the main characteristics of data quantitatively, providing insights into the dataset's central tendency and variability.
  + **Inferential Statistics**:
    - **Technique**: Hypothesis testing using t-tests
    - **Reason**: Inferential statistics are used to make inferences or generalizations about a population based on sample data, helping to assess the significance of observed differences.
  + **Statistical Tests**:
    - **Chi-square Statistic**
    - **Reason:** The Chi-square statistic measures the association between categorical variables. A higher Chi-square value indicates a stronger association.
    - **Cramér's V:**
    - **Reason:** Cramér's V is a measure of association between categorical variables, scaled between 0 and 1. It provides a standardized way to interpret the strength of association, with v
  + **Machine Learning Algorithms**:
    - **Algorithm**: Random Forest Classifier
    - **Reason**: Random Forest is chosen for its robust performance in classification tasks, handling non-linear relationships and high-dimensional data effectively. It also provides feature importances, aiding in understanding variable contributions.
    - **Algorithm**: Logistic Regression
    - **Reason**: Logistic Regression is chosen for its effectiveness in binary classification tasks, predicting the probability of a binary outcome based on independent variables. It provides interpretable coefficients, aiding in understanding the impact of predictors on the outcome.