**1. What is Mean, Median, Mode?**

* **Mean**: The average of a set of numbers, calculated by dividing the sum of the values by the number of values.
* **Median**: The middle value in a set of numbers, which separates the higher half from the lower half.
* **Mode**: The value that appears most frequently in a data set.

**2. What is Dictionary?**

A dictionary in Python is an unordered collection of key-value pairs. Each key is unique and is used to store and retrieve corresponding values. Example:

my\_dict = {"name": "Alice", "age": 25}

**3. Difference between Bagging and Boosting**

* **Bagging (Bootstrap Aggregating)**:
  + Involves training multiple models independently using random subsets of the data and averaging their predictions.
  + Reduces variance and helps in avoiding overfitting.
  + Example: Random Forest.
* **Boosting**:
  + Involves training models sequentially, each new model correcting errors made by previous ones.
  + Focuses on reducing bias and improving accuracy.
  + Example: AdaBoost, Gradient Boosting.

**4. What is Logistic Regression?**

Logistic Regression is a statistical method for predicting binary outcomes from data. It estimates the probability of a binary response based on one or more predictor variables.

**5. What is Linear Regression?**

Linear Regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables.

**6. What is Supervised and Unsupervised Learning?**

* **Supervised Learning**:
  + The model is trained on labeled data.
  + Examples include classification and regression.
* **Unsupervised Learning**:
  + The model is trained on unlabeled data.
  + Examples include clustering and association.

**7. What is Descriptive, Inferential, Predictive Analytics?**

* **Descriptive Analytics**:
  + Analyzes historical data to understand past behaviors and trends.
* **Inferential Analytics**:
  + Uses sample data to make generalizations about a population.
* **Predictive Analytics**:
  + Uses statistical models and machine learning techniques to predict future outcomes based on historical data.

**8. What is Binomial Theorem?**

The Binomial Theorem describes the algebraic expansion of powers of a binomial. It states that (x + y)^n can be expanded into a sum involving terms of the form a \* x^b \* y^c.

**9. Advantages and Disadvantages of Decision Tree**

* **Advantages**:
  + Easy to understand and interpret.
  + Requires little data preprocessing.
  + Can handle both numerical and categorical data.
* **Disadvantages**:
  + Prone to overfitting.
  + Can be unstable with small variations in data.
  + Biased towards features with more levels.

**10. Advantages and Disadvantages of KNN**

* **Advantages**:
  + Simple to understand and implement.
  + No training phase required.
  + Flexible to feature/distance choices.
* **Disadvantages**:
  + Computationally intensive with large datasets.
  + Sensitive to irrelevant features and scale of the data.
  + Requires a good choice of k.

**11. What is P-value?**

The p-value is a measure of the evidence against a null hypothesis. A lower p-value indicates stronger evidence against the null hypothesis.

**12. Difference between z-test and t-test**

* **z-test**:
  + Used when the population variance is known or the sample size is large (n > 30).
  + Assumes normal distribution.
* **t-test**:
  + Used when the population variance is unknown and the sample size is small (n < 30).
  + More robust for small sample sizes.

**13. What is the F-statistic?**

The F-statistic is used in ANOVA to determine if the means of different groups are significantly different. It compares the variance between group means to the variance within the groups.

**14. What is the P-Statistic?**

The term "P-statistic" is not commonly used. However, in hypothesis testing, the p-value is often referred to as a statistic that helps determine the significance of results.

**15. What is ANOVA?**

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more groups to determine if at least one group mean is different from the others.

**16. What is a Decorator?**

A decorator in Python is a design pattern that allows you to add new functionality to an existing object without modifying its structure. Decorators are often used for logging, authentication, and authorization.

**17. What are the 4 V's in Big Data?**

* **Volume**: The amount of data.
* **Velocity**: The speed at which data is generated and processed.
* **Variety**: The different types of data.
* **Veracity**: The uncertainty of data.

**18. What is Right Join?**

A right join (or right outer join) returns all records from the right table, and the matched records from the left table. The result is NULL from the left side if there is no match.

**19. Difference between SQL and NoSQL**

* **SQL**:
  + Relational databases.
  + Structured query language (SQL).
  + Fixed schema.
  + Suitable for complex queries.
* **NoSQL**:
  + Non-relational databases.
  + No fixed schema.
  + Suitable for unstructured and semi-structured data.
  + Scales horizontally.

**20. Difference between Modules and Packages**

* **Module**:
  + A single file (or files) that are imported under one import and used.
  + Example: math, os.
* **Package**:
  + A collection of modules in directories that give a package hierarchy.
  + Example: numpy, pandas.

**21. What Data Analysis Techniques are Present?**

* **Exploratory Data Analysis (EDA)**:
  + Summarizing main characteristics of data.
  + Visualization techniques.
* **Descriptive Statistics**:
  + Measures of central tendency (mean, median, mode).
  + Measures of dispersion (variance, standard deviation).
* **Inferential Statistics**:
  + Hypothesis testing.
  + Confidence intervals.
* **Predictive Analytics**:
  + Regression analysis.
  + Classification techniques.
* **Data Mining**:
  + Clustering.
  + Association rule mining.