1. What is Data Analysis?

* Data Analysis is a process of collecting, inspecting, cleaning, transforming and modelling data with the goal of discovering useful information for making informing conclusion and supporting decision making.

1. Explain is Data Analysis Process?
2. Asking Questions
3. Data Wrangling
4. Exploratory Data Analysis
5. Drawing Conclusions
6. Communicating results
7. What is Data Wrangling?

* Data wrangling is the process of transforming and mapping data from ‘raw’ data form into a proper format with the intent of making it appropriate and valuable for analysis.

1. What is Data Gathering?

* Data Gathering is a process of collecting data from difference sources such as CSV files, API’s, webscraping, and databases.

1. What is Data Accessing?

* In data accessing, the data is to be understood more deeply. Before implementing methods to clean it, we will definitely need to have a better idea about what the data is about.

1. Types of unclean data.
2. Dirty data: Dirty data is also known as low quality data which has quality issues or content issues. Ex. Duplicated data, Missing data, inaccurate data.
3. Messy data: Messy data is also known as untidy data. It has structural issues.
4. Data Quality Dimensions.
5. Completeness: Is data missing ?
6. Validity: Is data valid/invalid? Ex. Negative heights
7. Accuracy: Data is valid but inaccurate. Ex. Weight of 10 years boys is 1 kg.
8. Consistency: But Valid and accurate but written differently. Ex. New York / NY
9. Order of severity:

* Completeness -> Validity -> Accuracy -> Consistency

1. EDA from Notes.
2. Explain the difference between Descriptive, Predictive, and Prescriptive analysis. / Types of Analysis.
3. Descriptive Analysis: Descriptive analysis helps us understand what has happened in the past by summarizing data in a simple way. It focuses on providing a clear picture of historical data. It uses techniques like data aggregation, charts, graphs, and statistics (like averages, percentages, and trends).
4. Predictive Analysis: Predictive analysis focuses on what could happen in the future by using historical data to predict outcomes. It uses statistical models and machine learning techniques to make forecasts. It identifies patterns in past data and uses these patterns to predict future trends or behaviors.
5. Prescriptive Analysis: Prescriptive analysis helps decide what actions should be taken to achieve a desired outcome. It provides recommendations by analyzing different possible future scenarios. It uses advanced algorithms, simulations, and optimization techniques to suggest the best course of action.
6. What are the different types of sampling techniques used by data analysts?

* Sampling is a method used by data analysts to select a subset of data from a larger population. This helps in analyzing and making conclusions without having to deal with the entire population, which may be too large or time-consuming to study.

1. Random Sampling: In random sampling, the data is randomly taken, every member of the population has an equal chance of being selected. You randomly choose individuals from the entire population.
2. Stratified Sampling: This technique involves dividing the population into different groups or "strata" based on specific characteristics (like age, gender, income), and then taking a random sample from each group. After splitting the population into groups, you randomly pick a certain number of people from each group, ensuring that every group is represented in your sample.
3. Systematic Sampling: Systematic sampling involves selecting individuals at regular intervals from a list. First, you choose a starting point and then select every nth person on the list.
4. Cluster Sampling: In cluster sampling, the population is divided into groups (clusters), and a few clusters are randomly selected. Then, all individuals within those chosen clusters are studied. You first divide the population into clusters (such as cities, schools, or neighborhoods), then randomly choose certain clusters, and survey all the people in those selected clusters.
5. Describe univariate, bivariate, and multivariate analysis.
6. Univariate Analysis: Univariate analysis focuses on analyzing one variable at a time. The goal is to understand the characteristics of that single variable. It looks at things like the distribution, central tendency (mean, median, mode), and spread (variance, range) of the data.
7. Bivariate Analysis: Bivariate analysis looks at the relationship between two variables. The goal is to find out how one variable affects or is related to the other. It helps answer questions like whether one variable increases as the other increases or whether they are independent of each other.
8. Multivariate Analysis: Multivariate analysis examines three or more variables at the same time. It looks for relationships and interactions between these variables. It’s useful when you want to understand complex relationships where many factors may influence the outcome.
9. What are the ethical considerations of data analysis?
10. Privacy: Respecting the privacy of individuals whose data is being analyzed. Personal information, like names, addresses, health records, or financial details, should be kept confidential. Analysts must ensure that sensitive data is not exposed to unauthorized people or used inappropriately.
11. Informed Consent: Getting permission from individuals before using their data. People should know how their data will be collected, used, and shared. They should have the choice to allow or deny the use of their data.
12. Accuracy: Ensuring that the data used in analysis is correct and free from errors. Inaccurate data can lead to wrong conclusions and decisions, which can harm individuals or organizations.
13. Bias: Avoiding unfair treatment or biased results in the analysis. Data analysis can sometimes reflect biases (like gender or racial bias) that exist in the data or in the methods used. Analysts should work to eliminate these biases to ensure fair and objective results.
14. Transparency: Being clear about how the data is collected, analyzed, and used. People and organizations should know how conclusions are reached from data. This builds trust and allows others to verify or challenge the findings if needed.
15. Responsibility in Decision Making: Ensuring that data analysis is used responsibly and ethically in making decisions. Decisions based on data can have real-world consequences. Analysts should be mindful of the impact their work may have on individuals, businesses, or society.
16. Avoiding Misleading Results: Presenting data and conclusions honestly without manipulating or misrepresenting the results. Misleading results can create false perceptions and lead to incorrect actions or policies.
17. Data Ownership: Understanding who owns the data and how it can be used. Just because someone has access to data doesn’t mean they own it or have the right to use it however they want.
18. How Data analysis is similar to Business Intelligence?

* Data analysis involves data gathering, inspecting, cleaning, transforming and finding relevant information, So, that it can be used for the decision-making process.
* Business Intelligence(BI) also makes data analysis to find insights as per the business requirements. It generally uses statistical and Data visualization tools popularly known as BI tools to present the data in user-friendly views like reports, dashboards, charts and graphs.

1. What is Time Series analysis?

* Time Series analysis is a statistical technique used to analyze and interpret data points collected at specific time intervals. Time series data is the data points recorded sequentially over time. The data points can be numerical, categorical, or both.
* The objective of time series analysis is to understand the underlying patterns, trends and behaviours in the data as well as to make forecasts about future values.