1. **What is data analysis?**

**Data analysis** is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.

A simple example of data analysis can be seen whenever we make a decision in our daily lives by evaluating what has happened in the past or what will happen if we make that decision. Basically, this is the process of analyzing the past or future and making a decision based on that analysis.

**2.What are the tools useful for data analytics?**

Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretation. Here is a complete list of [tools](https://www.guru99.com/big-data-analytics-tools.html) used for data analysis .

* **Excel**

Excel is a popular, basic, and frequently leveraged analytical tool in practically all industries.

### Python

It's a scripting language that is simple to understand, write, as well as maintain. Furthermore, it's a free open-source tool.

### R Programming

Well, R is the industry's premier analytics tool, and it's extensively used for statistics and data modeling.

### Tableau Public

It is a free data visualization application that links to any data source you can think of whether it's a corporate Data Warehouse, Microsoft Excel, or web-based information.

### SAS

SAS is basically a data manipulation programming ecosystem and language that is a market leader in analytics.

**3.What are the various step involved in any analytics project?**

**1. Problem definition** - The first step is to understand the business problem. What is the problem you are trying to solve - what is the business context? Your client can also give you a lot of data and ask you to do something with him.

**2. Data mining**- Once you have defined the problem, the next step is to explore the data and become familiar with it.

**3. Data Preparation** - Now that you have a good understanding of the data, you will need to prepare it for modeling. You will identify and process missing values, detect outliers, transform variables, create binary variables if needed, and so on

**4. Modeling** - Once the data is ready, you can start modeling. This is usually an iterative process where you run a template, evaluate the results, fine tune your approach, run another template, evaluate the results, readjust and so on

**5. Validation** - The final model (or perhaps the two or three best models) should then be submitted to the validation process. During this process, you test the model using entirely new data, that is, data that was not used to build the model.

**4.What are the responsibilities of a Data Analyst?**

A data analyst is responsible for organizing data related to sales numbers, market research, logistics, linguistics, or other behaviors. They utilize technical expertise to ensure data is accurate and high-quality. Data is then analyzed, designed, and presented in a way that assists individuals, businesses, and organizations make better decisions.

* Collecting, cleaning, and organizing data from multiple sources.
* Analyzing large amounts of data to identify patterns, trends, and insights.
* Presenting findings and insights to stakeholders in a clear and concise manner.
* Developing and implementing data models to solve complex business problems.
* Collaborating with teams and departments to support data-driven decision-making.
* Building and maintaining data pipelines and systems to ensure accurate and up-to-date data.
* Staying current on new technologies and data analysis techniques.

**5.Write some key skills usually required for a Data Analyst.**

Key skills required for a Data Analyst include:

* Strong analytical and problem-solving skills.
* Advanced proficiency in data manipulation, analysis, and visualization tools (e.g. SQL, Excel, R, Python).
* Good understanding of statistical methods and techniques.
* Knowledge of database management systems and data warehousing.
* Strong communication skills to effectively present findings and insights to stakeholders.
* Experience working with large datasets and data cleaning.
* Ability to interpret and communicate complex data and insights to non-technical stakeholders.
* Attention to detail and ability to work with accuracy and precision.
* Ability to work independently or as part of a team.
* Good organizational skills to manage multiple projects and deadlines.

**6.What are the common problems that data analysts encounter during analysis?**

Data Analytics challenges:

* Bias in data - Data can be biased due to sampling methods, data collection processes, or a lack of diversity in the data source.
* Lack of context - Data may not provide enough context to fully understand the findings and implications.
* Incomplete or missing data - This can be a result of poor data collection or storage processes.
* Dirty data - Data may contain errors, duplicates, or inconsistent formats that need to be cleaned and standardized.
* Outliers or extreme values - These values can have a significant impact on the results of an analysis and need to be handled carefully.
* Large and complex datasets - Analysing large datasets can be time-consuming and challenging, requiring advanced computational resources and skills.
* Difficulty in data visualization - Representing complex data insights in a meaningful and interpretable manner can be challenging.
* Integration of multiple data sources - Integrating data from different sources can require significant effort to standardize and align data formats.

**7.What is the difference between data analytics and data science?**

Data analytics and data science are both fields that deal with analyzing and interpreting data, but they have some key differences in their scope and methods.

Data analytics is focused on the process of analyzing data using statistical and quantitative methods to draw insights and conclusions from data sets. It typically involves working with structured data, such as numerical or categorical data, and using tools like Excel, SQL, or Tableau to perform descriptive and diagnostic analyses. Data analytics is often used in business, marketing, and finance to track key performance indicators, identify trends, and optimize operations.

On the other hand, data science is a broader field that encompasses the entire data pipeline, from data acquisition and cleaning to modeling and deployment. It involves applying advanced statistical and machine learning techniques to large and complex data sets, including unstructured data like text and images. Data scientists often work with programming languages like Python or R and use tools like Hadoop, Spark, or TensorFlow to build predictive models, automate decision-making, and create intelligent systems.