**Q1) What is Data Science?**

➔ Data science is an interdisciplinary field that combines elements of statistics, computer science, and domain expertise to extract insights and knowledge from data. The field encompasses a wide range of tasks, including:

1. Data cleaning: Data is often incomplete, inconsistent, or otherwise difficult to work with. Data cleaning is the process of identifying and addressing these issues so that the data is ready for analysis.
2. Data exploration and visualization: Before building models, data scientists will often explore the data to get a sense of its structure and characteristics. This may involve visualizing the data in various ways, such as with histograms, scatter plots, and heatmaps.
3. Feature extraction: Once the data has been cleaned and explored, data scientists will often extract features, or variables, that are relevant for the problem at hand. This step can involve engineering new variables, such as creating a ratio of two existing variables, or selecting a subset of variables from the original dataset.
4. Model building: After the data has been prepared, data scientists will build models to extract insights and make predictions. This step may involve using a variety of techniques, such as linear regression, decision trees, and neural networks.
5. Model evaluation: Once a model has been built, it must be evaluated to determine its performance. This step may involve using techniques such as cross-validation, AUC-ROC Curve, precision, recall and many other techniques to evaluate the model.
6. Model Deployment: After the model is built and evaluated, it may be deployed in a production environment where it can be used to make decisions or predictions.

Data science is used in a wide variety of fields, such as business, finance, healthcare, and government. In business, for example, data science can be used to improve customer retention, optimize pricing, and identify new market opportunities. In healthcare, data science can be used to develop personalized medicine, improve patient outcomes, and reduce costs.

Overall, data science is a rapidly growing field that is essential for organizations that want to make data-driven decisions and gain a competitive advantage.

**Q2) Important Disciplines under Data Science**

→ There are several important disciplines within the field of data science, including:

1. Statistics: used to analyse and interpret data, make predictions and inferences, and develop models for decision-making.
2. Machine Learning: a subset of artificial intelligence that involves using algorithms and statistical models to enable computers to learn from data.
3. Data Mining: the process of extracting useful information and patterns from large datasets.
4. Data Visualization: the use of visual representations, such as charts and graphs, to communicate and interpret data.
5. Data Engineering: the process of designing and building the infrastructure and systems that are needed to collect, store, and process large amounts of data.
6. Database Management: the process of creating, maintaining, and managing databases that store and organize data.

7. Business Intelligence: the use of data, analytics, and visualization to support decision-making in organizations.

These are just some of the many important disciplines within data science, and the field is constantly evolving and expanding.

Q3) What is data? Explain Big Data and Traditional Data.

➔

Data refers to information that is collected, stored, and used. It can be used to make decisions, understand patterns and trends, or gain insights.

Traditional data refers to the data that has been collected and used for many years. It is typically structured and stored in databases or spreadsheets. It is often used for business intelligence and reporting, and is analyzed using tools such as SQL or Excel.

Big data, on the other hand, refers to the large volume of data that is generated and collected from various sources, such as social media, IoT devices, and e-commerce transactions. It is often unstructured, semi-structured and can be in various format. The data is so large and complex that it cannot be processed and analyzed using traditional data processing techniques. To process big data, specialized software and technologies such as Hadoop, Spark, and NoSQL databases are used. The goal of big data is to extract insights and knowledge from the large volume of data.

Big data can be used for a wide variety of applications, such as customer analytics, fraud detection, and predictive maintenance. It allows companies to make data-driven decisions, improve operations, and gain a competitive advantage.

| **Feature** | **Big Data** | **Traditional Data** |
|---|---|---|
| Volume | Extremely large and complex data sets | Smaller and more manageable data sets |
| Variety | Structured, semi-structured, unstructured data | Structured data (e.g. databases, spreadsheets) |
| Velocity | Generated and collected at high speed | Collected at a slower pace |
| Veracity | Uncertain and incomplete | More accurate and reliable |
| Variety | Text, images, video, sensor data, etc. | Mostly structured data |
| Analytics/processing | Requires specialized software and technologies | Can be analyzed using traditional tools |
| Use cases | Predictive analytics, fraud detection, etc. | Business intelligence, reporting, etc. |

**Q4) benefits of big data and traditional data discipline**

→

Here are some benefits of big data and traditional data disciplines:

Big Data:

1. Enables organizations to make data-driven decisions by providing new insights and knowledge.
2. Helps organizations to identify patterns, trends, and relationships in large data sets.
3. Allows organizations to improve operations and gain a competitive advantage by identifying new opportunities for growth and efficiency.
4. Can be used for a wide variety of applications, such as customer analytics, fraud detection, and predictive maintenance.
5. Allows organizations to handle and process a high volume, velocity, and variety of data

Traditional Data:

1. Provides a clear and concise view of data, making it easy to understand and analyze.
2. Enables organizations to make informed decisions by providing accurate and reliable data.
3. Can be used for business intelligence and reporting.
4. The data is usually structured in a way that is easy to search and analyze.
5. The cost of traditional data analysis is usually lower than big data analysis
6. Can be analyzed using traditional tools such as SQL and Excel which are widely known by the professionals.

Both Big Data and Traditional data are important and they both have their own advantages and limitations. Organizations should use a combination of both types of data to make the most informed decisions.

**Q5) Techniques of working with traditional data?**

→

There are several techniques that can be used for working with traditional data. Some of the most common ones include:

1. Data Cleaning: This is the process of removing errors, inconsistencies, and duplicate data from the data set. It is an important step in preparing the data for analysis.
2. Data Transformation: This is the process of converting the data from one format to another, such as from a CSV file to an Excel spreadsheet. This is often necessary when working with data from different sources.
3. Data Aggregation: This is the process of combining data from different sources into a single data set. This can be useful for identifying patterns and trends across different data sets.
4. Data Mining: This is the process of extracting useful information from the data using statistical and machine learning techniques. This can be used to identify patterns, trends, and relationships in the data.
5. Data Visualization: This is the process of creating charts, graphs, and maps to help make the data more understandable and actionable. This can be used to communicate insights and findings to others.

6. SQL: SQL (Structured Query Language) is a widely used language for managing and querying relational databases. It allows for filtering, sorting and joining tables, making it a powerful tool for data manipulation and analysis.
7. Data Warehousing: Data warehousing is the process of collecting, storing and managing data from different sources and making it available for querying and analysis.

These techniques can be used in combination to work effectively with traditional data and gain useful insights from it.

**Q6) Techniques of working with the Big Data?**

→

There are several techniques that can be used for working with big data. Some of the most common ones include:

1. Data Ingestion: This is the process of collecting and importing data from various sources, such as social media, sensors, and web logs. This can be done using tools such as Apache Kafka, Apache Flume, and Apache Nifi.
2. Data Storage: This is the process of storing the data in a distributed file system or a NoSQL database, such as Hadoop Distributed File System (HDFS) or Apache Cassandra. This allows for the data to be easily accessible and scalable.
3. Data Processing: This is the process of analyzing and transforming the data using tools such as Apache Spark, Apache Storm, and Apache Flink. This can be used to identify patterns, trends, and relationships in the data.
4. Data Mining: This is the process of extracting useful information from the data using statistical and machine learning techniques. This can be used to identify patterns, trends, and relationships in the data.
5. Data Visualization: This is the process of creating charts, graphs, and maps to help make the data more understandable and actionable. This can be used to communicate insights and findings to others.
6. Data Governance: This is the process of managing and securing the data, ensuring that it is accurate, complete, and protected from unauthorized access.
7. Cloud computing: This refers to the delivery of computing services, such as servers, storage, and databases, over the internet. Cloud computing allows for the cost-effective and on-demand access to scalable computing resources, which are essential for big data processing.

These techniques can be used in combination to work effectively with big data and gain useful insights from it. It is important to note that there are also specialized technologies such as Apache Hadoop and Apache Hive which are specifically designed for big data processing and analytics.

**Q7)What is Business Intelligence?**

→

Business Intelligence (BI) is a set of tools, techniques, and technologies that are used to collect, store, analyze, and present data in a way that is meaningful and actionable for businesses. The goal of BI is to provide organizations with insights and knowledge that can help them make better decisions and improve their performance.

BI typically involves the use of various data sources, such as transactional databases, data warehouses, and external data sources, and a variety of tools, such as data visualization and reporting tools, data mining and predictive analytics tools, and dashboards and scorecards.

Business Intelligence can be used for a variety of purposes, such as:

- Identifying trends and patterns in customer behavior
- Monitoring performance indicators such as sales and financial metrics
- Improving operational efficiency
- Identifying new business opportunities
- Benchmarking performance against competitors

BI can be used by organizations of all sizes and in all industries. It can be used by businesses to gain a competitive advantage and make better decisions by providing them with real-time access to accurate and relevant information.

**Q8) BI Techniques?**

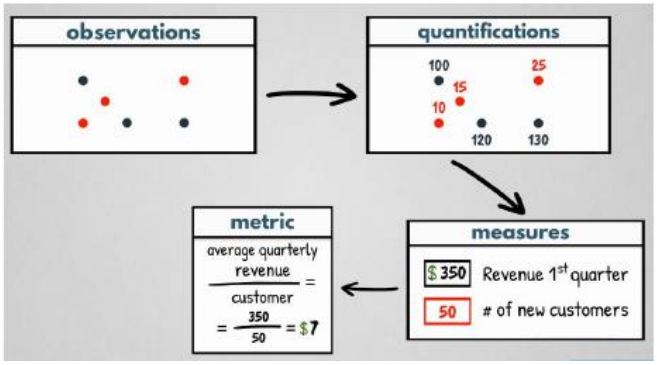
→

There are several techniques that can be used as part of a Business Intelligence (BI) strategy. Some of the most common ones include:

1. Data Warehousing: This is the process of collecting, storing, and managing data from different sources in a central location. Data warehousing allows for data to be easily accessible and queried for analysis.
2. Online Analytical Processing (OLAP): This is a technique that allows for the multidimensional analysis of data. OLAP enables users to view data from different perspectives and drill down into the details to gain insights.
3. Data Mining: This is the process of extracting useful information from the data using statistical and machine learning techniques. Data mining can be used to identify patterns, trends, and relationships in the data.
4. Data Visualization: This is the process of creating charts, graphs, and maps to help make the data more understandable and actionable. Data visualization can be used to communicate insights and findings to others.
5. Dashboards and Scorecards: These are tools that provide a real-time, visual representation of key performance indicators (KPIs) and other important metrics. Dashboards and scorecards can be used to monitor performance and make data-driven decisions.
6. Reporting and Analysis: This is the process of creating reports and analyzing data using tools such as SQL and Excel. Reporting and analysis can be used to identify trends and patterns in the data, and to make informed decisions.
7. Predictive Analytics: This is the use of statistical models and machine learning algorithms to make predictions about future events. Predictive analytics can be used to identify potential

**Q9) Techniques to measure the business performance?**

→




Collecting Observations:

From the above diagram, you can observe variables such as sales volume (marked as blue colour dots) or new customers who have enrolled in your website (marked as red colour dots). Each monthly revenue or each customer is considered a single observation.

Quantification:

However no mathematical manipulations can be applied to these observations. What we must do is quantify that information. It is the process of representing observations as numbers. Consider your revenues from new customers for January, February and March were 100, 120 and $ 130 respectively while the corresponding numbers of new for the same 3 months are 10, 15 and 25.

Measure:

A measure is the accumulations of observations to show some information. For example, if you total the revenues of all 3 months to obtain the value of $350 that would be a measure of the revenue of the first quarter of that year. Similarly add together the number of new customers for the same period and you have another measure.

Metric:

A metric refers to a value that derives from the measure you obtain and aims to gauging business performance or progress to compare. If a measure is related to something like simple descriptive statistics of past performance a metric has a business meaning attached.

EG. If you estimate the average quarterly revenue per customer which equals 350 divided by 50 that is $7. This is a metric.

KPI:

In a real business where the number of observations is significantly larger you can derive thousands of metrics where we can't keep track of all possible metrics we can extract from a dataset. What you need to do is choose the metrics that are tightly aligned with your business objectives. These metrics are called **K.P.I's** Key Performance Indicators. Key because they are related to your main business goals. Performance because they show how successfully you have performed within a specific time frame and indicators because their values or metrics that indicate something related to your business performance.

Q10) Real life examples of Business Intelligence.

→

There are many real-life examples of how businesses use Business Intelligence (BI) to improve their performance. Here are a few examples:

1. Retail companies use BI to analyze customer data to identify buying patterns and preferences. This helps them to personalize their marketing and increase sales.

2. Manufacturing companies use BI to monitor their production processes and identify bottlenecks. This helps them to optimize their operations and improve efficiency.
3. Healthcare organizations use BI to analyze patient data to improve patient outcomes and reduce costs. This can include identifying high-risk patients, tracking disease outbreaks, or monitoring supply chain and inventory.
4. Financial services companies use BI to detect fraudulent activities and monitor customer behavior. This helps them to reduce their losses and improve their risk management.
5. E-commerce companies use BI to track customer behavior and optimize the user experience. This includes analyzing website traffic and customer interactions, identifying customer segments and preferences, and tracking the customer journey.
6. Supply Chain companies use BI to track inventory, logistics and distribution, this helps them to optimize their stock, reduce costs and improve their delivery time.
7. Government agencies use BI to monitor and analyze public data to improve public services and decision-making. This can include monitoring crime trends, tracking public health, or managing public resources.