

Data Analytics Model Questions and Answers

Module -1

1. What Is Data Analytics?

Ans. Organisations are becoming data-driven as data becomes more prominent by the minute, which implies adopting strategies to capture more data. This data is subsequently sorted, saved, and analysed to provide logical and useful information. The technique is made feasible through data analytics.

Data analytics is the process of gaining valuable insights and information from data via the use of quantitative and qualitative methodologies. It is beneficial to businesses and even science; for example, researchers utilise it to validate their hypotheses.

There are three main types.

- **First-party data:** Information gathered by a corporation about its consumers.
- **Second-party data:** Information obtained by a corporation from a known organisation that initially gathered it.
- **Aggregated data:** Information obtained from a marketplace by a corporation.

2. What Are The Types of Data Analytics?

Ans. Data analytics is classified into three types: descriptive, diagnostic, predictive, and prescriptive. Each has its own set of objectives and duties to play in the data analytics process.

- Descriptive Analytics:** In the data analytics process, descriptive analytics addresses the "what" inquiries. It aids stakeholders in comprehending huge datasets by summarising them. The descriptive analysis examines the previous performance of the organisation. It consists of the following steps:
 - Data gathering
 - Processing of data
 - Data visualisation and analysis.
- Diagnostic Analytics:** Diagnostic analytics provides solutions to the "why" questions that arise during the data analytics process. It analyses the descriptive analysis data and then further assesses them to determine the reason. The diagnostic analysis procedure is divided into three steps:
 - Detecting any unusual changes in the data.
 - Data pertaining to the modifications is gathered.
 - Statistical approaches aid in the discovery of linkages and patterns associated to changes.
- Predictive Analytics:** The goal of predictive analytics is to provide answers to queries regarding the future of data analytics. The patterns are identified using historical data. Statistical and machine learning approaches are employed in the process. Neural networks, decision trees, and regression are a few examples.
- Prescriptive Analytics:** Prescriptive analysis assists firms in making well-informed decisions and forecasting analytics. Machine learning algorithms are used in this form of data analytics to uncover patterns in massive datasets.

Data Analytics Model Questions and Answers

3. What is the Role of Data Analytics in Business?

Ans. As previously said, data analysis serves as the foundation for decision-making. Beyond the basics, businesses now utilise data to maintain the seamless operation of many other business activities, as well as to research their performance, find commercial prospects, and develop new strategies.

There are so many uses for data analysis in the business environment that it is impossible to mention all of them. However, here are some of the most frequent among companies:

- Interpretation of the business reality
- Acquisition of a global, comprehensive and reliable vision of the business and all its areas
- Critical thinking capabilities
- Obtaining insights and valuable information
- Enhancing the monitoring of the business activity
- Fostering inter-departmental cooperation
- Optimising processes
- Streamlining work routines
- Identification of errors, weaknesses and potential areas for improvement
- Prediction of future scenarios (forecasting)
- Increased knowledge of customers
- Definition of target customers and buyer persona
- Identification of business opportunities and generation of actions towards the market
- Improving the customer experience
- Optimisation of existing markets
- Reorientation of business strategies
- Adaptation to an uncertain and unstable market
- Increasing Return on Investment (ROI)
- Risk reduction

4. What are the tools using data analytics?

Ans.

- R and Python
- Microsoft Excel
- Tableau
- RapidMiner
- KNIME
- Power BI
- Apache Spark
- QlikView
- Talend
- Splunk

5. What are the applications used in data analytics?

Data Analytics Model Questions and Answers

Ans. Organisations that have implemented the concepts and principles of business analytics have seen substantial improvements and remarkable growth. Business analytics may be used in a variety of sectors.

Business analytics is used in many industries, including agriculture, medicine, production and development, human resources, finance, and many more, to help firms expand and keep their customers satisfied.

Here are some examples of business analytics to show how it is used in different industries.

i. Finance:

- Business Analytics assists financial managers in managing their finances optimally and then taking relevant measures. Implementing business analytics in various sectors of finance (such as investment banking and budgeting) can prove to be highly fruitful for the finance industry.
- It helps in building future strategies for a new product by observing similar products and methodologies.
- In addition to this, business analytics can also be used to predict future loan defaulters.

ii. HUMAN RESOURCES MANAGEMENT (HRM):

- Human Resource Management is the process or practice of managing, hiring, organizing, training, and directing people in an organization in a strategic manner. Human Resources (or HR) professionals use business analytics in several ways.
- It helps them in analyzing large amounts of data to understand employees' needs and grievances and therefore assist them accordingly.
- Business analytics can be used by HR in determining the right candidates, the expected salaries as well as the trending retention rates in the industries.
- Moreover, HR professionals can leverage business analytics to forecast the trajectory of the organization and thus efficiently design appropriate training and development programs for trainees or employees.

iii. PRODUCTION AND INVENTORY MANAGEMENT:

- Management is a key element in every organization. It aims to enhance the profits and productivity of an organization all the while trying to reduce overall costs.
- Business Analytics serves as a great tool for management and manufacturing. It is involved in every phase of product development. It supports analyzing the inventory measures and designing business solutions that are most suitable for products.
- It can help determine the costs and gauge the expected sales of products. This way the organizations can adapt to the latest styles and opportunities in the industry.
- Hence, business analytics stands as a boon for the diverse sectors of management, be it inventory management or product management.

iv. CUSTOMER RELATIONSHIP MANAGEMENT (CRM):

- Customer Relationship Management or CRM is the process of building and managing the organization's relationships as well as interactions with customers.
- Business analytics can be used in customer relationship management to understand the customer base better and therefore, implement corresponding strategies. This helps significantly drive sales and amplifies the organization's profits.
- Customers' purchasing patterns, needs, buying behaviors, issues, feedback, and all the other indicators can be obtained and analyzed through business analytics methodologies. These indicators can then be used to foster long-lasting and loyal relationships between clients and the organization.

v. MARKETING:

Data Analytics Model Questions and Answers

- Marketing, when combined with business analytics can prove to be one of the best strategies an organization can implement.
- Business analytics helps the organization to know its users, their needs, behaviors, and purchasing styles to design and modify suitable plans and schemes.
- Sales can be optimized and user experience can be enhanced. Business analytics can help marketers know their target audience and their interest.
- It can also be used to evaluate and determine how well a product or a marketing strategy is performing in the market. Considering these factors, organizations can modify their strategies and implement better planning.

Module – 2

6. Data Collection and Data Pre-Processing Data Collection Strategies?

Ans. Preparing data for use in machine learning models and deep learning.

The data preparation stage is the starting point for each ML project's development cycle.

Data preparation comes before the training and learning step of any ML model, as seen in the rudimentary ML development process below. As a result, it is critical to carry out this stage appropriately from the start.

Within the data preparation stage are the data collection and data pre-processing stages.

i. Data collection:

- Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. Following are some of the problems that can arise in data collection:
- Inaccurate data. The collected data could be unrelated to the problem statement.
- Missing data. Sub-data could be missing. That could take the form of empty values in columns or missing images for some class of prediction.
- Data imbalance. Some classes or categories in the data may have a disproportionately high or low number of corresponding samples. As a result, they risk being under-represented in the model.
- Data bias. Depending on how the data, subjects and labels themselves are chosen, the model could propagate inherent biases on gender, politics, age or region, for example. Data bias is difficult to detect and remove.

Several techniques can be applied to address those problems:

- Pre-cleaned, freely available datasets. If the problem statement (for example, image classification, object recognition) aligns with a clean, pre-existing, properly formulated dataset, then take advantage of existing, open-source expertise.
- Web crawling and scraping. Automated tools, bots and headless browsers can crawl and scrape websites for data.
- Private data. ML engineers can create their own data. This is helpful when the amount of data required to train the model is small and the problem statement is too specific to generalize over an open-source dataset.

Data Analytics Model Questions and Answers

- Custom data. Agencies can create or crowdsource the data for a fee.
- ii. Data pre-processing:

Real-world raw data and images are often incomplete, inconsistent and lacking in certain behaviors or trends. They are also likely to contain many errors. So, once collected, they are pre-processed into a format the machine learning algorithm can use for the model.

Pre-processing includes a number of techniques and actions:

- Data cleaning. These techniques, manual and automated, remove data incorrectly added or classified.
- Data imputations. Most ML frameworks include methods and APIs for balancing or filling in missing data. Techniques generally include imputing missing values with standard deviation, mean, median and k-nearest neighbors (k-NN) of the data in the given field.
- Oversampling. Bias or imbalance in the dataset can be corrected by generating more observations/samples with methods like repetition, bootstrapping or Synthetic Minority Over-Sampling Technique (SMOTE), and then adding them to the under-represented classes.
- Data integration. Combining multiple datasets to get a large corpus can overcome incompleteness in a single dataset.
- Data normalization. The size of a dataset affects the memory and processing required for iterations during training. Normalization reduces the size by reducing the order and magnitude of data.

7. What is data preprocessing?

Ans. Data preprocessing, a subset of data preparation, refers to any sort of processing performed on raw data in order to prepare it for further processing. It has typically been a critical first stage in the data mining process. Data preparation approaches have lately been modified for training machine learning and AI models as well as conducting inferences against them.

Data preprocessing converts data into a format that can be handled more readily and efficiently in data mining, machine learning, and other data science operations. To achieve reliable findings, the approaches are often utilised at the early phases of the machine learning and AI development pipeline.

There are several different tools and methods used for preprocessing data, including the following:

- sampling, which selects a representative subset from a large population of data;
- transformation, which manipulates raw data to produce a single input;
- denoising, which removes noise from data;
- imputation, which synthesizes statistically relevant data for missing values;
- normalization, which organizes data for more efficient access; and
- feature extraction, which pulls out a relevant feature subset that is significant in a particular context.

8. Why is data preprocessing important?

Ans. Virtually any type of data analysis, data science or AI development requires some type of data preprocessing to provide reliable, precise and robust results for enterprise applications.

Real-world data is messy and is often created, processed and stored by a variety of humans, business processes and applications. As a result, a data set may be missing individual fields, contain manual input errors, or have duplicate data or different names to describe the same thing. Humans can often identify and

Data Analytics Model Questions and Answers

rectify these problems in the data they use in the line of business, but data used to train machine learning or deep learning algorithms needs to be automatically preprocessed.

Machine learning and deep learning algorithms work best when data is presented in a format that highlights the relevant aspects required to solve a problem. Feature engineering practices that involve data wrangling, data transformation, data reduction, feature selection and feature scaling help restructure raw data into a form suited for particular types of algorithms. This can significantly reduce the processing power and time required to train a new machine learning or AI algorithm or run an inference against it.

One caution that should be observed in preprocessing data: the potential for reencoding bias into the data set. Identifying and correcting bias is critical for applications that help make decisions that affect people, such as loan approvals. Although data scientists may deliberately ignore variables like gender, race or religion, these traits may be correlated with other variables like zip codes or schools attended, generating biased results.

Most modern data science packages and services now include various preprocessing libraries that help to automate many of these tasks.

9. What are the key steps in data preprocessing?

Ans.

1. Data profiling. Data profiling is the process of examining, analyzing and reviewing data to collect statistics about its quality. It starts with a survey of existing data and its characteristics. Data scientists identify data sets that are pertinent to the problem at hand, inventory its significant attributes, and form a hypothesis of features that might be relevant for the proposed analytics or machine learning task. They also relate data sources to the relevant business concepts and consider which preprocessing libraries could be used.
2. Data cleansing. The aim here is to find the easiest way to rectify quality issues, such as eliminating bad data, filling in missing data or otherwise ensuring the raw data is suitable for feature engineering.
3. Data reduction. Raw data sets often include redundant data that arise from characterizing phenomena in different ways or data that is not relevant to a particular ML, AI or analytics task. Data reduction uses techniques like principal component analysis to transform the raw data into a simpler form suitable for particular use cases.
4. Data transformation. Here, data scientists think about how different aspects of the data need to be organized to make the most sense for the goal. This could include things like structuring unstructured data, combining salient variables when it makes sense or identifying important ranges to focus on.
5. Data enrichment. In this step, data scientists apply the various feature engineering libraries to the data to effect the desired transformations. The result should be a data set organized to achieve the optimal balance between the training time for a new model and the required compute.

Data Analytics Model Questions and Answers

6. Data validation. At this stage, the data is split into two sets. The first set is used to train a machine learning or deep learning model. The second set is the testing data that is used to gauge the accuracy and robustness of the resulting model. This second step helps identify any problems in the hypothesis used in the cleaning and feature engineering of the data. If the data scientists are satisfied with the results, they can push the preprocessing task to a data engineer who figures out how to scale it for production. If not, the data scientists can go back and make changes to the way they implemented the data cleansing and feature engineering steps.

10. What are the data preprocessing techniques and cleansing?

Ans.

Data preprocessing techniques

There are two main categories of preprocessing -- data cleansing and feature engineering. Each includes a variety of techniques, as detailed below.

Data cleansing

Techniques for cleaning up messy data include the following:

Identify and sort out missing data. There are a variety of reasons a data set might be missing individual fields of data. Data scientists need to decide whether it is better to discard records with missing fields, ignore them or fill them in with a probable value. For example, in an IoT application that records temperature, adding in a missing average temperature between the previous and subsequent record might be a safe fix.

Reduce noisy data. Real-world data is often noisy, which can distort an analytic or AI model. For example, a temperature sensor that consistently reported a temperature of 75 degrees Fahrenheit might erroneously report a temperature as 250 degrees. A variety of statistical approaches can be used to reduce the noise, including binning, regression and clustering.

Identify and remove duplicates. When two records seem to repeat, an algorithm needs to determine if the same measurement was recorded twice, or the records represent different events. In some cases, there may be slight differences in a record because one field was recorded incorrectly. In other cases, records that seem to be duplicates might indeed be different, as in a father and son with the same name who are living in the same house but should be represented as separate individuals. Techniques for identifying and removing or joining duplicates can help to automatically address these types of problems.

Feature engineering

Feature engineering, as noted, involves techniques used by data scientists to organize the data in ways that make it more efficient to train data models and run inferences against them. These techniques include the following:

Feature scaling or normalization. Often, multiple variables change over different scales, or one will change linearly while another will change exponentially. For example, salary might be measured in thousands of

Data Analytics Model Questions and Answers

dollars, while age is represented in double digits. Scaling helps to transform the data in a way that makes it easier for algorithms to tease apart a meaningful relationship between variables.

Data reduction. Data scientists often need to combine a variety of data sources to create a new AI or analytics model. Some of the variables may not be correlated with a given outcome and can be safely discarded. Other variables might be relevant, but only in terms of relationship -- such as the ratio of debt to credit in the case of a model predicting the likelihood of a loan repayment; they may be combined into a single variable. Techniques like principal component analysis play a key role in reducing the number of dimensions in the training data set into a more efficient representation.

Discretization. It's often useful to lump raw numbers into discrete intervals. For example, income might be broken into five ranges that are representative of people who typically apply for a given type of loan. This can reduce the overhead of training a model or running inferences against it.

Feature encoding. Another aspect of feature engineering involves organizing unstructured data into a structured format. Unstructured data formats can include text, audio and video. For example, the process of developing natural language processing algorithms typically starts by using data transformation algorithms like Word2vec to translate words into numerical vectors. This makes it easy to represent to the algorithm that words like "mail" and "parcel" are similar, while a word like "house" is completely different. Similarly, a facial recognition algorithm might reencode raw pixel data into vectors representing the distances between parts of the face.

11. How is data preprocessing used?

Ans. Data preprocessing plays a key role in earlier stages of machine learning and AI application development, as noted earlier. In an AI context, data preprocessing is used to improve the way data is cleansed, transformed and structured to improve the accuracy of a new model, while reducing the amount of compute required.

A good data preprocessing pipeline can create reusable components that make it easier to test out various ideas for streamlining business processes or improving customer satisfaction. For example, preprocessing can improve the way data is organized for a recommendation engine by improving the age ranges used for categorizing customers.

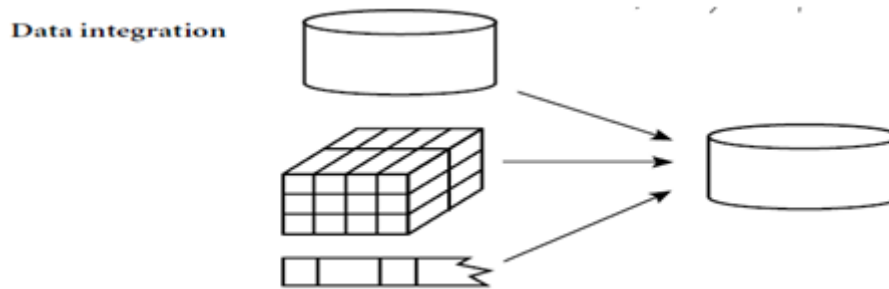
Preprocessing can also simplify the work of creating and modifying data for more accurate and targeted business intelligence insights. For example, customers of different sizes, categories or regions may exhibit different behaviors across regions. Preprocessing the data into the appropriate forms could help BI teams weave these insights into BI dashboards.

In a customer relationship management (CRM) context, data preprocessing is a component of web mining. Web usage logs may be preprocessed to extract meaningful sets of data called user transactions, which consist of groups of URL references. User sessions may be tracked to identify the user, the websites requested and their order, and the length of time spent on each one. Once these have been pulled out of the raw data, they yield more useful information that can be applied, for example, to consumer research, marketing or personalization.

Data Analytics Model Questions and Answers

12. What is Data Integration and Transformation?

Ans. Data integration is one of the steps of data pre-processing that involves combining data residing in different sources and providing users with a unified view of these data.



- It merges the data from multiple data stores (data sources)
- It includes multiple databases, data cubes or flat files.
- Metadata, Correlation analysis, data conflict detection, and resolution of semantic heterogeneity contribute towards smooth data integration.
- There are mainly 2 major approaches for data integration - commonly known as "tight coupling approach" and "loose coupling approach".

Tight Coupling

o Here data is pulled over from different sources into a single physical location through the process of ETL - Extraction, Transformation and Loading.

o The single physical location provides an uniform interface for querying the data.

ETL layer helps to map the data from the sources so as to provide a uniform data warehouse. This approach is called tight coupling since in this approach the data is tightly coupled with the physical repository at the time of query.

ADVANTAGES:

- Independence (Lesser dependency to source systems since data is physically copied over)
- Faster query processing
- Complex query processing
- Advanced data summarization and storage possible
- High Volume data processing

DISADVANTAGES:

- Latency (since data needs to be loaded using ETL)
- Costlier (data localization, infrastructure, security)

Data Analytics Model Questions and Answers

Loose Coupling

- Here a virtual mediated schema provides an interface that takes the query from the user, transforms it in a way the source database can understand and then sends the query directly to the source databases to obtain the result.
- In this approach, the data only remains in the actual source databases.
- However, mediated schema contains several "adapters" or "wrappers" that can connect back to the source systems in order to bring the data to the front end.

ADVANTAGES:

- Data Freshness (low latency - almost real time)
- Higher Agility (when a new source system comes or existing source system changes - only the corresponding adapter is created or changed - largely not affecting the other parts of the system)
- Less costlier (Lot of infrastructure cost can be saved since data localization not required)

DISADVANTAGES:

- Semantic conflicts
- Slower query response
- High order dependency to the data sources

For example, let's imagine that an electronics company is preparing to roll out a new mobile device. The marketing department might want to retrieve customer information from a sales department database and compare it to information from the product department to create a targeted sales list. A good data integration system would let the marketing department view information from both sources in a unified way, leaving out any information that didn't apply to the search.

DATA TRANSFORMATION:

Data transformation $-2, 32, 100, 59, 48 \longrightarrow -0.02, 0.32, 1.00, 0.59, 0.48$

In data mining pre-processes and especially in metadata and data warehouse, we use data transformation in order to convert data from a source data format into destination data.

• **Data Mapping:**

It maps the data elements from the source to the destination and captures any transformation that must occur.

• **Code Generation:**

It creates the actual transformation program.

Data transformation:

- Here the data are transformed or consolidated into forms appropriate for mining.
- Data transformation can involve the following:

♣ **Smoothing:**

- It works to remove noise from the data.

Data Analytics Model Questions and Answers

- It is a form of data cleaning where users specify transformations to correct data inconsistencies.
- Such techniques include binning, regression, and clustering.

♣ Aggregation:

- Here summary or aggregation operations are applied to the data.
- This step is typically used in constructing a data cube for analysis of the data at multiple granularities.
- Aggregation is a form of data reduction.

♣ Generalization :

- Here low-level or “primitive” (raw) data are replaced by higher-level concepts through the use of concept hierarchies.
- For example, attributes, like age, may be mapped to higher-level concepts, like youth, middle-aged, and senior.
- Generalization is a form of data reduction.

♣ Normalization:

- Here the attribute data are scaled so as to fall within a small specified range, such as 1:0 to 1:0, or 0:0 to 1:0.
- Normalization is particularly useful for classification algorithms involving neural networks, or distance measurements such as nearest-neighbor classification and clustering
- For distance-based methods, normalization helps prevent attributes with initially large ranges (e.g., income).
- There are three methods for data normalization:

1. min-max normalization :

- o performs a linear transformation on the original data
- o Suppose that $\min A$ and $\max A$ are the minimum and maximum values of an attribute, A .
- o Min-max normalization maps a value, v , of A to v_0 in the range $[\text{new_min}A; \text{new_max}A]$ by computing
- o Min-max normalization preserves the relationships among the original data values.

$$V_0 = \frac{v - \min A}{\max A - \min A} (\text{new_max}A - \text{new_min}A) + \text{new_min}A$$

1. z-score normalization

- o Here the values for an attribute, A , are normalized based on the mean and standard deviation of A .

o Value, v of A is normalized to v_0 by computing $V_0 = \frac{v - \bar{A}}{\sigma A}$, where \bar{A} and σA are the mean and standard deviation, respectively.

- o This method of normalization is useful when the actual minimum and maximum of attribute A are unknown, or when there are outliers that dominate the min-max normalization.

1. normalization by decimal scaling:

- o Here the normalization is done by moving the decimal point of values of attribute A .

Data Analytics Model Questions and Answers

o The number of decimal points moved depends on the maximum absolute value of A.

o Value, v of A is normalized to v_0 by computing $v_0 = \frac{v}{10^{\text{to the power of } |j|}}$ where j is the smallest integer such that $\text{Max}(|v_0|) < 1$.

• Attribute construction:

o Here new attributes are constructed and added from the given set of attributes to help the mining process.

o Attribute construction helps to improve the accuracy and understanding of structure in high-dimensional data.

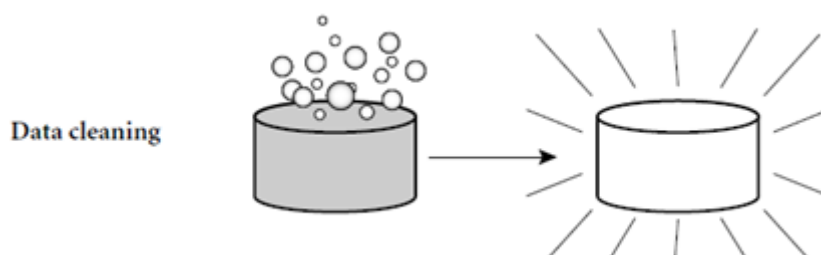
o By combining attributes, attribute construction can discover missing information about the relationships between data attributes that can be useful for knowledge discovery.

EG: The structure of stored data may vary between applications, requiring semantic mapping prior to the transformation process. For instance, two applications might store the same customer credit card information using slightly different structures:

<u>APPLICATION A</u>	<u>EXAMPLE</u>	<u>APPLICATION B</u>	<u>EXAMPLE</u>
Cardholder First Name	JOHN	Cardholder Name	JOHN <u>DOE</u>
Cardholder Last Name	<u>DOE</u>	Card Type	VISA
Card Type and Card Number	VISA 0123 4567 8910 1112	Card Number	0123 4567 8910 1112
Expiration Date	05/2012	Expiration Date	05/2012

To ensure that critical data isn't lost when the two applications are integrated, information from Application A needs to be reorganized to fit the data structure of Application B.

III. DATA CLEANING:



• Real-world data tend to be incomplete, noisy, and inconsistent.

• Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data.

Data Analytics Model Questions and Answers

• Data cleaning tasks include:

→ Fill in missing values

1. Ignore the tuple: This is usually done when the class label is missing
2. Fill in the missing value manually: this approach is time-consuming and may not be feasible given a large data set with many missing values.
3. Use a global constant to fill in the missing value: Replace all missing attribute values by the same constant
4. Use the attribute mean to fill in the missing value: Use a particular value to replace the missing value for an attribute.
5. Use the attribute mean for all samples belonging to the same class as the given tuple: replace the missing value with the average value of the attribute for the given tuple.
6. Use the most probable value to fill in the missing value: This may be determined with regression, inference-based tools using a Bayesian formalism, or decision tree induction.

→ Identify outliers and smooth out noisy data

• Noise is a random error or variance in a measured variable. And can be smoothed using the following steps:

1. Binning: Binning methods smooth a sorted data value by consulting its “neighborhood,” that is, the values around it.
2. Regression: Data can be smoothed by fitting the data to a function, such as with regression. Linear regression involves finding the “best” line to fit two attributes (or variables), so that one attribute can be used to predict the other. Multiple linear regression is an extension of linear regression, where more than two attributes are involved and the data are fit to a multidimensional surface.
3. Clustering: Outliers may be detected by clustering, where similar values are organized into groups, or “clusters.”

→ Correct inconsistent data

→ Resolve redundancy caused by data integration

Module – 6

13. What is the STP Marketing Model?

Ans. The STP model is useful when creating marketing communications plans since it helps marketers to prioritize propositions and then develop and deliver personalized and relevant messages to engage with different audiences. The three-step funnel consists of market segmentation, market targeting, and product positioning.

Within your research-based market segmentation phase, you are aiming to identify a basis for the segmentation of your target customers, and determine important characteristics to differentiate each market segment.

When creating your targeting and positioning strategy, you must evaluate the potential and commercial attractiveness of each segment, and then develop detailed product positioning for each selected segment, including a tailored marketing mix based on your knowledge of that segment.

Data Analytics Model Questions and Answers

Market Segmentation

- Identify basis for segmentation
- Determine important characteristics of each market segment



Market Targeting

- Evaluate potential and commercial attractiveness of each segments
- Select one or more segments



Product Positioning

- Develop detailed product positioning for selected segments
- Develop a marketing mix for each selected segment

14. What are STP Marketing Planning Tools?

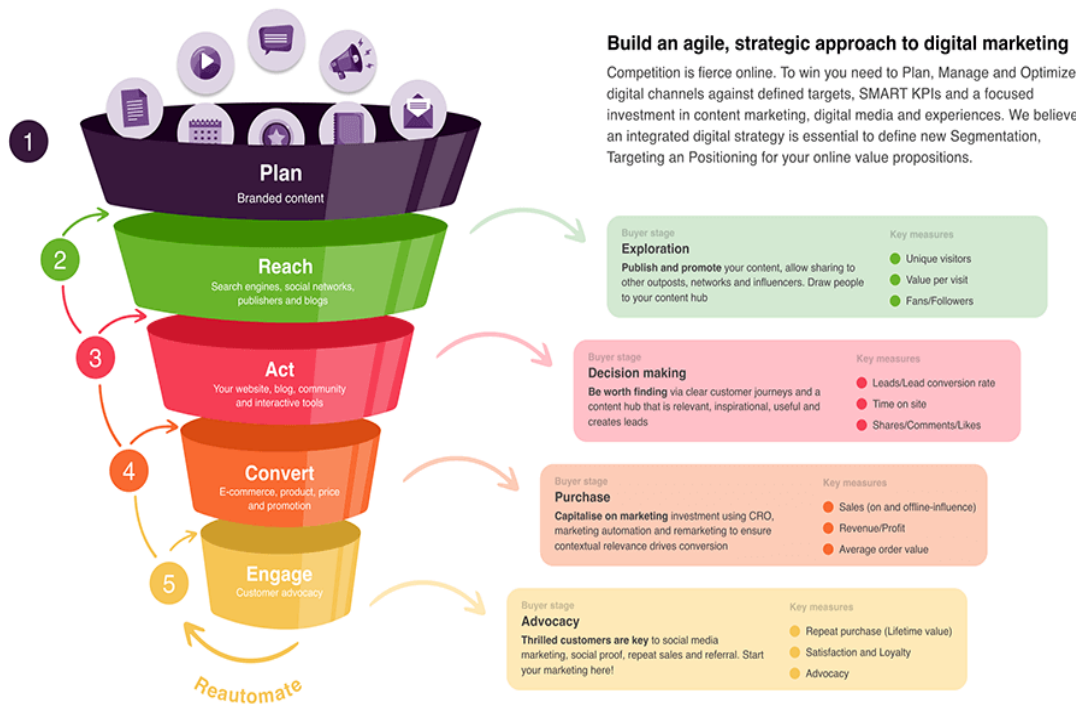
Ans. Segmentation, targeting, and positioning is an audience-focused rather than product-focused approach to marketing communications which helps deliver more relevant messages to commercially appealing audiences.

STP is a critical strategy and planning tool, featured in our RACE Growth System. Our RACE Growth System supports marketers, managers, and business owners to create a 90-day marketing plan across each stage of their marketing funnel.

So, while STP sits within the planning activities, the benefits of effective segmentation, targeting and positioning can be felt across the types of customers you reach, interact with, convert, and engage.

Data Analytics Model Questions and Answers

The Smart Insights RACE Planning Framework



Let us know what you think @smartinsights and www.smartinsights.com



15. What is STP Marketing?

Ans. Segmentation, targeting, and positioning (STP) is a marketing model that redefines whom you market your products to, and how. It makes your marketing communications more focused, relevant, and personalised for your customers.

In short, STP is a marketing approach where you segment your audience, target the best-fit audience segments for your product, and position your product to capture your target segment effectively.

16. What is the Formula of STP Marketing?

Ans. The easiest way to remember the STP model is through the STEP formula, which is

$$\text{Segmentation} + \text{Targeting} = \text{Positioning}$$

A closer look at this formula tells us that the product positioning for each target segment is different. This forms the essence of the STP (Segmentation, Targeting, and Positioning) marketing model.

Let's take a closer look at each of these parts of segmentation, targeting and positioning.

17. What are the Segmentation in Marketing?

Ans. When you start creating a GTM strategy for your product, you have an idea of who your audience is. You can target the entire group that fits the broad definition of your audience, but chances are a generic message may fail to resonate with a huge chunk of that group.

Data Analytics Model Questions and Answers

Segmenting the audience into smaller groups based on specific attributes gives you better clarity on who benefits the most out of your product and how. With this clarity, you can make your messages more focused and relevant to target groups.

While you can segment your audience using any criteria that best suits your business, the below criteria are commonly used:



Geography (where):

segmentation based on audience location.
It can be country, region, or even city-specific.



Demography (what):

segmentation based on audience demographics,
such as age, gender, economic status,
profession, etc.



Behaviour (how):

segmentation based on audience buying patterns,
preferred communication channels, browsing
habits, brand loyalty, etc.



Psychography (who):

segmentation based on audience personality
traits- such as their lifestyle, hobbies,
interests, etc.

Segmentation with an example

Suppose your product is plant-based milk. Your general audience is people who want to move away from dairy-based products. You can segment this audience into two categories:

Segment A: people who are looking at dairy-free alternatives for lifestyle purposes, typically high-income groups.

Segment B: lactose-intolerant people looking for other options.

The message you use for these two segments is obviously going to be different from each other. Using tools like Data Studio, you can further segment the above two segments into groups that already use a competitor product and those that don't. You can then hone your messaging according to it.

Data Analytics Model Questions and Answers

18. What is Targeting in STP Model?

Ans.

The next step in the STP model is targeting. This is the stage where you decide which segments you created during the segmentation phase are worth pursuing. You should ideally consider the below criteria to choose your targetable segments:

Size: Your audience segments must have enough potential customers to be worth marketing to. If your segments are too small, you may not get enough conversions to justify your marketing efforts.

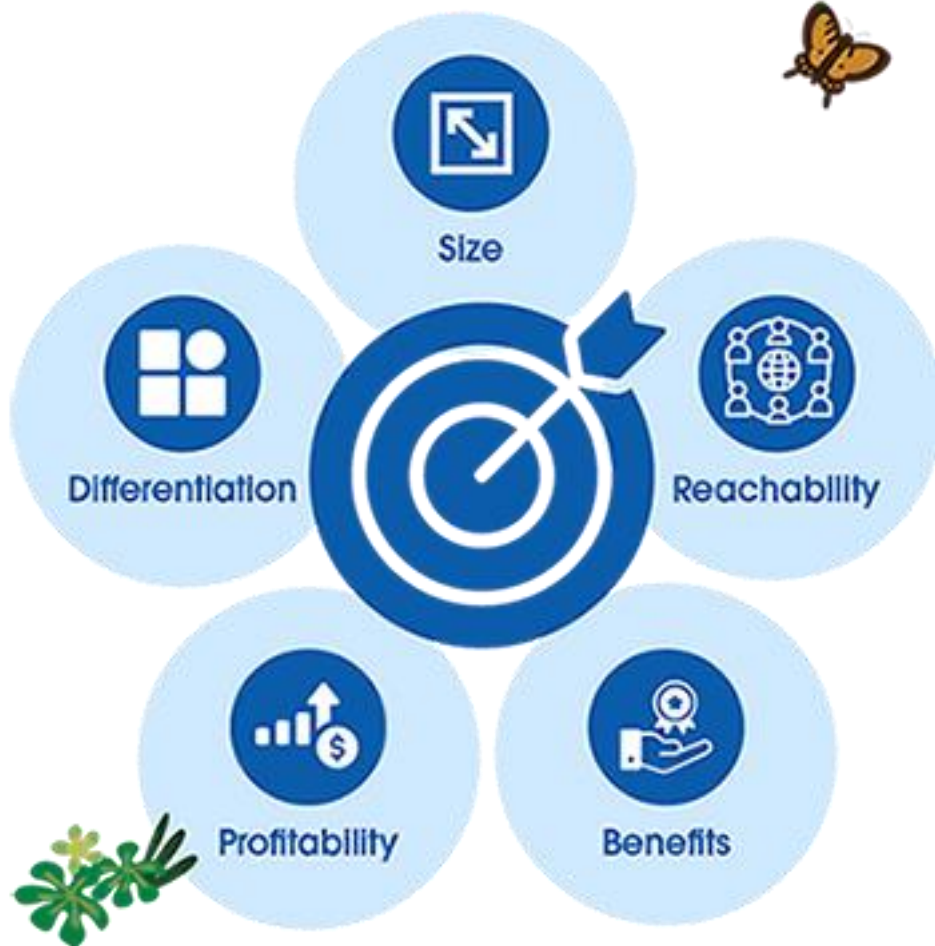
Difference: There should be a measurable difference between any two segments. The lack of it leads to unnecessary duplication of efforts.

Reachability: The segments should be accessible to your sales and marketing teams and not be marred by technical or legal complications.

Profitability: The segment should have a low-to-medium customer acquisition cost (CAC) while bringing in high returns, i.e., the audience must be willing to spend money on your product.

Benefits: Different benefits attract different segments. In our plant-based milk example, Segment A would go for cruelty-free while Segment B for dairy-free.

Criteria to Consider When Choosing Target Segments



Data Analytics Model Questions and Answers

Knowing which audience segments to target comes from having all-around visibility of those segments in one place. This makes comparing segments and weighing the pros and cons of targeting some segments over others easier.

In our example of plant-based milk, you've determined through research that veganism is all the rage, and roughly 60% of the people are searching for dairy-free alternatives. You also discover that approximately 80% of the people in your chosen demographic are lactose intolerant. Though the audience size is more significant in the second segment, you're likely to get more returns when you go after the first segment as it consists of high-income groups who are ready to pay a premium for quality lifestyle-changing products.

19. What is Positioning in STP Marketing?

Ans. The final stage of the STP model, positioning, is where you use the insights gained from segmentation and targeting to decide how you're going to communicate your product to chosen audience segments.

While segmentation and targeting are about customers, positioning is about your product from the customer's perspective. You can consider positioning as the bridge that connects your product with the audience. This is the stage where you perform competitor analysis, figure out your value proposition, and communicate that to your customers.

Based on what your brand stands for, you can position your product in several ways. If you're in the luxury market, you can appeal to the 'desire for prestige' among customers by positioning yourself as a status symbol. Or, if you fall in the budget category, you could differentiate yourself by offering more benefits to your target at a lower cost than your competitors.

The best way to approach positioning is by drawing a Product Positioning Map that has two key market attributes as its axes and plotting your competitors and you in it. This will give you a clear picture of how you stack up against your competition and where you should place your product to maximise profits.

qualtrics.^{XM}

Brand Positioning

The winning space must be unique to competitors, motivating to consumers and ownable for the brand



20. What are the Benefits of STP Marketing?

Data Analytics Model Questions and Answers

Ans.

- Improved engagement: Because you're targeting precise audience segments with personalised messages, your audience finds you relevant and is more likely to engage and convert.
- Reduced marketing costs: Since you're going after only those segments with a high potential return on investment, you're no longer wasting your budget on channels and segments that don't work.
- More robust product: Because you know precisely whom you're pitching your product to, you can make improvements based on feedback from that audience segment, fostering focused product innovation.

21. What are the Value Generation in STP Model?

Ans. We've covered the basics of the STP marketing model with benefits and examples. Now, it's time to get down to the brass tacks; that is, see how you can implement a segmentation, targeting, and positioning model for your business step-by-step.

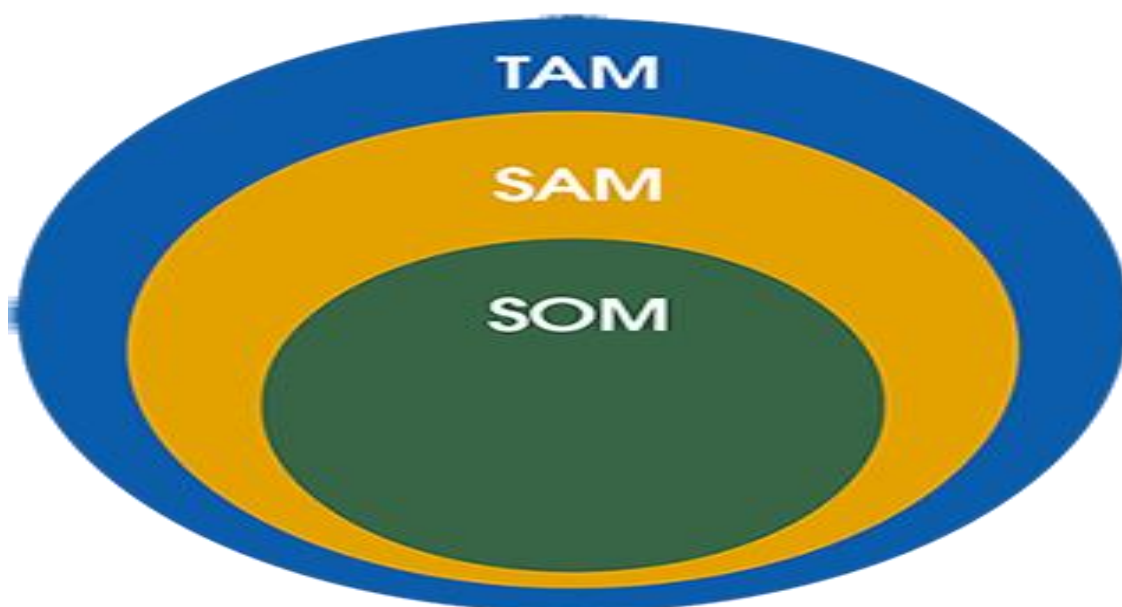
Step 1: Define your market

The world may be your market, but breaking it down into manageable segments is how you conquer it. To know the market segment in which you can hit the bullseye, you start by defining your Total Available Market (TAM), Serviceable Available Market (SAM), and Serviceable Obtainable Market (SOM). Let's look at what each of these is:

Total Available Market (TAM): TAM is the total market demand for a product or service. In other words, it's the biggest available market for the brand. TAM is the maximum revenue that a business can generate if it achieves 100% of its market share.

Serviceable Available Market (SAM): SAM is a subset of TAM, that is, a portion of the total available market that fits your product or service. You can define SAM by geographical or product specialisation constraints.

Serviceable Obtainable Market (SOM): SOM is a subset of SAM, that is, the segment of the serviceable available market that you can realistically reach after considering factors like product differentiation, budget, and competition.



Step 2: Create audience segments

Data Analytics Model Questions and Answers

Now that you know your market definition, you can segment the audience within that definition.

You can segment the audience based on geography, demography, behaviour, or psychography, but ideally, a mix of all four can help you achieve clearly differentiated segments. The more segmentation layers or variables you add, the more delineated your segments would be.

For example, suppose you're selling a luxury makeup product. In that case, you can target high-income working women (demographics) in India (geography), who follow makeup handles on social media (behaviour) and are willing to spend money on premium makeup products (psychography).

This kind of repeated layering and segmentation creates focused audience groups that you can target with hyper-personalised messages. Research by McKinsey found that companies that excel at personalisation generate 40% more revenue from those activities than average players.

Step 3: Identify the more attractive segments

Have all the segments data in one place and evaluate the attractiveness of each segment. You can use metrics like return on investment, segment size, and growth potential in your evaluation. Again, solutions like Salesforce Data Studio and CDP help gather data and get comprehensive visibility into different audience segments, improving segmentation.

Step 4: Evaluate your competition

With your audience segmentation sorted, it's now time to look at your product and determine how it stacks up against your competition. Prepare a table that lists down all of your product capabilities and your competition's, do a SWOT analysis, identify gaps, and figure out the most viable entry point into your desired customer segment.

Step 5: Fix your positioning

The groundwork on segmentation and targeting is now out of your way, so you can focus on positioning your product to grab the lion's share of the market. You can follow any one or a mix of the following positioning strategies:

- Competitor-based positioning: where you show in what aspects better than your competitor.
- Consumer-based positioning: how well your product aligns with consumer needs.
- Price-based positioning: how you're competitively priced and give customers more value for their money
- Benefit-based positioning: how your customers benefit from buying your product, either individually or over your competition.
- Attribute-based positioning: what your unique selling point or value proposition is, above and beyond benefits and price.
- Prestige-based positioning: how customers get a status boost from buying your product.

Step 6: Determine your marketing mix

The final step of the segmentation, targeting, and positioning model is to choose your 'marketing mix' that helps reinforce your positioning. The marketing mix consists of four Ps - Product, Price, Placement, and Promotions.

Data Analytics Model Questions and Answers

- Product represents factors like quality, benefits, features, design, services, support, availability, and edge over the competition.
- Price reflects what customers are willing to pay for the product. It covers list price, discounts, payment methods, etc. Pricing your product much lower than your competitor might fetch you immediate benefits but will be detrimental to revenue in the long run.
- Placement covers "where" your product is available. It includes ecommerce, physical stores, inventory, logistics, trade channels, etc.
- Promotion takes into account "how" your product reaches your customer. It covers marketing campaigns, advertising, public relations, sales promotions, word of mouth, influencer marketing, and so on.

22. How could you manage the market segmentation process?

Ans. Segmentation refers to the process of creating small segments within a broad market to select the right target market for various brands. Market segmentation helps the marketers to devise and implement relevant strategies to promote their products amongst the target market.

A market segment consists of individuals who have similar choices, interests and preferences. They generally think on the same lines and are inclined towards similar products. Once the organizations decide on their target market, they can easily formulate strategies and plans to make their brands popular amongst the consumers.

Steps in Market Segmentation

Identify the target market

The first and foremost step is to identify the target market. The marketers must be very clear about who all should be included in a common segment. Make sure the individuals have something in common. A male and a female can't be included in one segment as they have different needs and expectations.

Burberry stocks separate merchandise for both men and women. The management is very clear on the target market and has separate strategies for product promotion amongst both the segments.

A Garnier men's deodorant would obviously not sell if the company uses a female model to create awareness.

Segmentation helps the organizations decide on the marketing strategies and promotional schemes.

Maruti Suzuki has adopted a focused approach and wisely created segments within a large market to promote their cars.

- Lower Income Group - Maruti 800, Alto
- Middle Income Group - Wagon R, Swift, Swift Dzire, Ritz
- High Income Group - Maruti Suzuki Kizashi, Suzuki Grand Vitara, Suzuki Grand Vitara would obviously have no takers amongst the lower income group.

The target market for Rado, Omega or Tag Heuer is the premium segment as compared to Maxima or a Sonata watch.

Data Analytics Model Questions and Answers

Identify expectations of Target Audience

Once the target market is decided, it is essential to find out the needs of the target audience. The product must meet the expectations of the individuals. The marketer must interact with the target audience to know more about their interests and demands.

Kellogg's K special was launched specifically for the individuals who wanted to cut down on their calorie intake.

Marketing professionals or individuals exposed to sun rays for a long duration need something which would protect their skin from the harmful effects of sun rays. Keeping this in mind, many organizations came with the concept of sunscreen lotions and creams with a sun protection factor especially for men.

Create Subgroups

The organizations should ensure their target market is well defined. Create subgroups within groups for effective results.

- Cosmetics for females now come in various categories.
- Creams and Lotions for girls between 20-25 years would focus more on fairness.
- Creams and lotions for girls between 25 to 35 years promise to reduce the signs of ageing.

Review the needs of the target audience

It is essential for the marketer to review the needs and preferences of individuals belonging to each segment and sub-segment. The consumers of a particular segment must respond to similar fluctuations in the market and similar marketing strategies.

Name your market Segment

Give an appropriate name to each segment. It makes implementation of strategies easier. A kids section can have various segments namely new born, infants, toddlers and so on.

Marketing Strategies

Devise relevant strategies to promote brands amongst each segment. Remember you can't afford to have same strategies for all the segments. Make sure there is a connect between the product and the target audience. Advertisements promoting female toiletries can't afford to have a male model, else the purpose gets nullified.

A model promoting a sunscreen lotion has to be shown roaming or working in sun for the desired impact.

Review the behavior

Review the behavior of the target audience frequently. It is not necessary individuals would have the same requirement (demand) all through the year. Demands vary, perceptions change and interests differ. A detailed study of the target audience is essential.

Data Analytics Model Questions and Answers

Size of the Target Market

It is essential to know the target market size. Collect necessary data for the same. It helps in sales planning and forecasting.

23. What is Cluster Analysis?

Ans. Cluster analysis is a data analysis technique that identifies meaningful, naturally occurring groups within a dataset and distinguishes them as clusters. It is used to discover hidden relationships in data based on specific characteristics.

Can you manually segment your customers and market? Sure, you can, but the manual method is limiting and only truly effective with a small number of characteristics or attributes. It simply doesn't scale well. Cluster analysis with advanced analytics and machine learning can quickly scale to a high number of attributes. It's also completely data-driven, using an unsupervised model (the algorithm learns patterns without tagged data), which makes it more accurate and credible.

24. What is Hierarchical Cluster Analysis?

Ans. The hierarchical clustering technique is a commonly used, easy clustering technique. It is divided into two types:

Agglomerative hierarchical clustering begins by considering each data point as its own cluster. Subsequent iterations merge similar clusters until one cluster or K clusters are established.

The basic algorithm does the following:

- Compute proximity matrix
- Identify each point as a cluster
- Repeat and merge the two closest clusters and update the proximity matrix
- Repeat until you reach a single cluster

Divisive hierarchical clustering is the opposite of agglomerative hierarchical clustering in that all data points are considered together as a single cluster. Each iteration separates points that are not similar. Those dissimilar data points are considered individual clusters. When completed, the result is n clusters, with each data point in its own cluster.

The basic algorithm does the following:

- Consider all data points as a single cluster
- Iterate and separate data points from the cluster which are dissimilar
- Repeat until each data point is separate and considered an individual cluster.

25. What is K-Means Clustering?

Data Analytics Model Questions and Answers

Ans. K-Means clustering is an unsupervised learning algorithm. There is no labeled data for this clustering, unlike in supervised learning. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects belonging to another cluster.

The term 'K' is a number. You need to tell the system how many clusters you need to create. For example, K = 2 refers to two clusters. There is a way of finding out what is the best or optimum value of K for a given data.

For a better understanding of k-means, let's take an example from cricket. Imagine you received data on a lot of cricket players from all over the world, which gives information on the runs scored by the player and the wickets taken by them in the last ten matches. Based on this information, we need to group the data into two clusters, namely batsman and bowlers.

<https://shorturl.at/oDOWY>

26. What is Market Segmentation?

Ans. Both cluster analysis, and market segmentation involve grouping customer segments based on similarities. While segmentation is based on human input, cluster analysis is driven by machine learning. Cluster analysis provides insights that allow businesses to drill down into the needs and wants of each market segment, allowing them to offer more personalized products and messaging. Using cluster analysis, you can identify new target market segments as well as ones to avoid.

Your clusters in market segmentation will usually have a heavier emphasis on geographic information, such as metro areas, states, countries, regions, etc., and demographics, such as age, income, gender, etc.

27. What is Customer Segmentation?

Ans. Cluster analysis in customer segmentation is used to create homogeneous groups of customers. In general, customer segmentation is used to identify behaviors and attitudes of the groups you've segmented by market. Cluster analysis will reveal clusters based on these characteristics.

28. What is K-Means Cluster Analysis?

Ans. The K-means algorithm divides a single cluster into K different clusters. It does this by finding organically similar data points and assigning each one to a cluster with similar characteristics. K-means clustering works by constantly trying to find a centroid (a data point that represents the mean or centre of the cluster). The end clusters will each have a centroid and data points that are closer to the centroid compared to the other centroids.

There are pros and cons to using K-means cluster analysis:

Pros

- Simple, popular method
- Guarantees convergence
- Offers a good estimate of centroids' initial positions

Cons

- You must specify the number of clusters

Data Analytics Model Questions and Answers

- Depends on random initial values, so it may be inconsistent in different runs
- Data may need to be scaled before clustering

K-means cluster analysis is widely used across several verticals, from determining urban traffic patterns for Uber drivers to segmenting customers based on interests, purchase history, or buying behaviors.

29. What is Discriminant Analysis and What are the Two Groups?

Ans. Discriminant analysis (DA) is a multivariate technique used to separate two or more groups of observations (individuals) based on variables measured on each experimental unit (sample) and find the contribution of each variable in separating the groups. DA works by finding one or more linear combinations of the selected variables. Furthermore, prediction or allocation of new observations to previously defined groups can be investigated with a linear or quadratic function to assign each individual to one of the predefined groups. For example, this method could be used to separate four types of flour prepared from green and ripe Cavendish bananas based on physicochemical properties (green peel (Gpe), ripe peel (Rpe), green pulp (Gpu), and ripe pulp (Rpu)). DA has been widely used for analyzing food science data to separate different groups.

DA is typically used when the groups are already defined prior to the study.

The end result of DA is a model that can be used for the prediction of group memberships. This model allows us to understand the relationship between the set of selected variables and the observations. Furthermore, this model will enable one to assess the contributions of different variables.

Discriminant analysis is a technique that is used by the researcher to analyze the research data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature. The term categorical variable means that the dependent variable is divided into a number of categories. For example, three brands of computers, Computer A, Computer B and Computer C can be the categorical dependent variable.

The objective of discriminant analysis is to develop discriminant functions that are nothing but the linear combination of independent variables that will discriminate between the categories of the dependent variable in a perfect manner. It enables the researcher to examine whether significant differences exist among the groups, in terms of the predictor variables. It also evaluates the accuracy of the classification.

Discriminant analysis is described by the number of categories that is possessed by the dependent variable.

As in statistics, everything is assumed up until infinity, so in this case, when the dependent variable has two categories, then the type used is two-group discriminant analysis. If the dependent variable has three or more than three categories, then the type used is multiple discriminant analysis. The major distinction to the types of discriminant analysis is that for a two group, it is possible to derive only one discriminant function. On the other hand, in the case of multiple discriminant analysis, more than one discriminant function can be computed.

There are many examples that can explain when discriminant analysis fits. It can be used to know whether heavy, medium and light users of soft drinks are different in terms of their consumption of frozen foods. In the field of psychology, it can be used to differentiate between the price sensitive and non price sensitive buyers of groceries in terms of their psychological attributes or characteristics. In the field of business, it can

Data Analytics Model Questions and Answers

be used to understand the characteristics or the attributes of a customer possessing store loyalty and a customer who does not have store loyalty.

Module – 7

Prof. Tapas Kumar Mondal