**What is Data Analytics? Explain in detail.**

The process of analyzing raw data to find trends and answer questions

The data analytics process has some components that can help a variety of initiatives. By combining these components, a successful data analytics initiative will provide a clear picture of where you are, where you have been and where you should go.

## Types of Data Analytics

There are four primary types of data analytics: descriptive, diagnostic, predictive and prescriptive analytics.

## Descriptive analytics

helps answer questions about what happened. These techniques summarize large datasets to describe outcomes to stakeholders. By developing key performance indicators (KPIs,) these strategies can help track successes or failures.

This process requires the collection of relevant data, processing of the data, data analysis and data visualization. This process provides essential insight into past performance.

## Diagnostic analytics

helps answer questions about why things happened.

They take the findings from descriptive analytics and dig deeper to find the cause. The performance indicators are further investigated to discover why they got better or worse. This generally occurs in three steps:

Identify anomalies in the data. These may be unexpected changes in a metric or a particular market.

Data that is related to these anomalies is collected.

Statistical techniques are used to find relationships and trends that explain these anomalies.

## Predictive analytics

helps answer questions about what will happen in the future Predictive analytical tools provide valuable insight into what may happen in the future and its techniques include a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression.

## Prescriptive analytics

helps answer questions about what should be done. By using insights from predictive analytics, data-driven decisions can be made

Prescriptive analytics techniques rely on machine learning strategies that can find patterns in

large datasets. By analyzing past decisions and events, the likelihood of different outcomes can be estimated.

These types of data analytics provide the insight that businesses need to make effective and efficient decisions. Used in combination they provide a well-rounded understanding of a company’s needs and opportunities.

**The Role of Data Analytics**

The primary goal of a data analyst is to increase efficiency and improve performance by discovering patterns in data.

The primary steps in the data analytics process are data mining, data management, statistical analysis, and data presentation. The importance and balance of these steps depend on the data being used and the goal of the analysis.

Data mining is an essential process for many data analytics tasks. This involves extracting data from unstructured data sources.

These may include written text, large complex databases, or raw sensor data.

The key steps in this process are to extract, transform, and load data (often called ETL.) These steps convert raw data into a useful and manageable format. Data mining is generally the most time-intensive step in the data analysis pipeline.

Data management or data warehousing is another key aspect of a data analyst’s job. Data warehousing involves designing and implementing databases that allow easy access to the results of data mining. This step generally involves creating and managing SQL databases. Non- relational and NoSQL databases are becoming more common as well.

Statistical analysis allows analysts to create insights from data. Both statistics and machine learning techniques are used to analyze data.

The final step in most data analytics processes is data presentation. This step allows insights to be shared with stakeholders. Data visualization is often the most important tool in data presentation.

 **Importance of Data Analytics**

Data analytics has an important role in the banking and finance industries, used to predict market trends and assess risk.

Credit scores are an example of data analytics that affects everyone.

Data analytics is also used to detect and prevent fraud to improve efficiency and reduce risk for financial institutions.

Data analytics can provide critical information for healthcare (health informatics), crime prevention, and environmental protection. These applications of data analytics use these techniques to improve our world.

**Explain about different data analytics Tools and Environment.**

There are many tools that are used for deriving useful insights from the given data. Some are programming based and others are non-programming based. Some of the most popular tools are:

SAS

Microsoft Excel R

Python Tableau RapidMiner KNIME SAS:

SAS was a programming language developed by the SAS Institute for performed advanced analytics, multivariate analyses, business intelligence, data management and predictive analytics.

Its programming language is considered to be high level thus making it easier to learn.

It is used by many companies such as Google, Facebook, Twitter, Netflix and Accenture.

SAS brought to the market a huge set of products in 2011 for customer intelligence and various SAS modules for web, social media and marketing analytics used largely for profiling customers and gaining insights about prospective customers.

Even though it is under attack by upcoming languages such as R, Python, SAS still continues to develop in order to prove that it is still a major stakeholder in the data analytics market.

1. Easy to learn 2 Easy to debug 3 Data Security

Disadvantages of SAS Cost

Lack of graphic representation Difficult than R

Microsoft Excel :

It is an important spreadsheet application that can be useful for recording expenses, charting data and performing easy manipulation and lookup and or generating pivot tables to provide the desired summarized reports of large datasets that contain significant data findings.

It is written in C#, C++ and .NET Framework and its stable version were released in 2016.

It has various built-in functions to satisfy the various statistical, financial and engineering needs.

It is relatively useful for performing somewhat complex analyses of data when compared to other tools such as R or python. It is a common tool among financial analysts and sales managers to solve complex business problems.

Advantage:

* Organizing Data
* Formula and Calculation Features DISADVANTAGES
* Calculation Errors
* Time Consuming

R :

It is a free and open-source language that can be run on various UNIX platforms, Windows and MacOS. It also has a command line interface which is easy to use. However, it is tough to learn especially for people who do not have prior knowledge about programming. However, it is very useful for building statistical software and is very useful for performing complex analyses. It has more than 11, 000 packages and we can browse the packages category-wise.

It is a powerful high-level programming language that is used for general purpose programming. ADVANTAGES

Open Source

The Array of Packages Platform Independent DISADVANTAGES

Basic Security Lesser Speed

Complicated Language

Python

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

It is used for:

* web development (server-side),
* software development,
* mathematics,
* system scripting.

Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc). Python has a simple syntax similar to the English language.

Python has syntax that allows developers to write programs with fewer lines than some other programming languages.

Python can be used on a server to create web applications.

Python can be used to handle big data and perform complex mathematics.

supports both structured and functional programming methods. It’s an extensive collection of libraries make it very useful in data analysis.

It is easy to learn compared to R and can be assembled onto any platform such as MongoDB or SQL server.

It is very useful for big data analysis and can also be used to extract data from the web. It can also handle text data very well.

ADVANTAGES

Easy to Read, Learn and Write Free and Open-Source Portability DISADVANTAGES

Slow Speed

Not Memory Efficient Database Access Tableau Public :

Tableau Public is free software developed by the public company “Tableau Software” that allows users to connect to any spreadsheet or file and create interactive data visualizations.

It can also be used to create maps, dashboards along with real-time updation for easy presentation on the web.

The results can be shared through social media sites or directly with the client making it very convenient to use.

The resultant files can also be downloaded in different formats. This software can connect to any type of data source, be it a data warehouse or an Excel application or some sort of web-based data.

Approximately 446 companies use this software for operational purposes and some of the companies that are currently using this software include SoFi, The Sentinel and Visa.

ADVANTAGES

Ease of Use

High Performance Mobile-Friendliness DISADVANTAGES

High Cost Security Issues

Time- and Resource-Intensive Staff Training

RapidMiner :

RapidMiner is an extremely versatile data science platform developed by “RapidMiner Inc”. The software emphasizes lightning fast data science capabilities and provides an integrated

environment for preparation of data and application of machine learning, deep learning, text mining and predictive analytical techniques.

It can also work with many data source types including Access, SQL, Excel, Tera data, Sybase, Oracle, MySQL and Dbase.

Here we can control the data sets and formats for predictive analysis.

Approximately 774 companies use RapidMiner and most of these are US-based. Some of the esteemed companies on that list include the Boston Consulting Group and Dominos Pizza Inc.

ADVANTAGES

No coding skills needed! Rapidminer is a GUI tool that you can connect boxes on a canvas to conduct data anlysis, this serves as a great introduction to data analytics. Free for students! You can get a provisional liscence with a dot edu account. This is a great perk of the software. Data analytics and data visualization tools are available within the software with a plethora of other features!

DISADVANTAGES

Very buggy! The software tends to crash often, this is especially more common with things such as neural networks etc. Limitations of some versions! Even with the student version there is a limit of 10,000 rows of output, so if you are trying to do analysis on a 12,000 point data set , 2000 points will randomly be omitted.

Knime :

Knime, the Konstanz Information Miner is a free and open-source data analytics software. It is also used as a reporting and integration platform.

It involves the integration of various components for Machine Learning and data mining through the modular data-pipe lining.

It is written in Java and developed by KNIME.com AG.

It can be operated in various operating systems such as Linux, OS X and Windows.

More than 500 companies are currently using this software for operational purposes and some of them include Aptus Data Labs and Continental AG.

ADVANTAGES

Easy to learn as you can be freer and do not have to follow programming standards

* Allows you to focus on innovation / product vs. programming concepts
* Once you learn node-based development it will allow you to operate multiple platforms
* You can deploy advanced functionality like machine learning without any deep understanding of it

DISADVANTAGES

* Smaller network and therefore less education materials and training available
* You will not actually learn the underlying fundamentals as you are leveraging existing code (nodes) instead of building it yourself
* Smaller skill demand vs actual language and inability to work as part of a team (dev team)

**Explain about Application of Modeling in Business.**

## Business Model

A Business Model can be defined as a representation of a business or solution that often include a graphic component along with supporting text and relationships to other components. For example, if we have to understand a company’s business model, then we would like to study the following areas like −

Core values of the company What it serves?

What is sets apart? Its key resources Major relationships Its delivery channels

With the help of modelling techniques, we can create a complete description of existing and proposed organizational structures, processes, and information used by the enterprise.

Business Model is a structured model, just like a blueprint for the final product to be developed. It gives structure and dynamics for planning. It also provides the foundation for the final product.

Purpose of Business Modelling

Business modelling is used to design current and future state of an enterprise. This model is used by the Business Analyst and the stakeholders to ensure that they have an accurate understanding of the current “As-Is” model of the enterprise

What are the applications of modeling in business?

[https://docs.servicenow.com/bundle/rome-it-business-management/page/product/application-](https://docs.servicenow.com/bundle/rome-it-business-management/page/product/application-portfolio-management/concept/application-portfolio-management.html) [portfolio-management/concept/application-portfolio-management.html](https://docs.servicenow.com/bundle/rome-it-business-management/page/product/application-portfolio-management/concept/application-portfolio-management.html)

## 7 Applications of the Business Model Canvas

1. Application Portfolio Management (APM)

application to gain a comprehensive understanding of the applications used in your organization so you can identify redundancies, and decrease budgetary costs. By consolidating applications within the same business function, you can identify applications to modernize or upgrade.

APM also helps you decide whether to invest, sustain, or replace applications based on the business need aligned towards the organization goal

You can address business challenges such as:

* Redundant applications for similar functions.
* Increasing cost of owning and maintaining applications.
* Increasing demand to upgrade the existing applications for new functions.
* Conflict between in-house legacy applications and that of the vendors.
* Inadequate performance because of outdated applications.

Overcome these challenges, improve business functions in an efficient and smooth manner, and optimize cost with these APM processes:

Identify

Build a comprehensive inventory of the business applications in use. Measure

Determine the usage of the applications by gathering metrics such as cost, usage, risk, and quality.

Evaluate

Assess the value of each application by evaluating their significance and usage based on the metrics.

Decide

Decide whether to invest in the application, maintain it as it is or replace it with another application, or retire the application.

Take action

Implement your decision by retaining, upgrading, or retiring the applications. APM uses the following key solution components:

Application Classification

Focuses on Enterprise Business Applications, which can also include functional modules part of a larger ERP suite.

Provides additional attributes to classify applications in a new CI class, Business application,

which extends the base Configuration Management Database (CMDB) configuration item.

The configuration items used in APM are related by establishing a CMDB relationship with each other.



Business Architecture.

A business architecture model concerns itself with providing a conceptual model of how business architecture components come together to provide a cohesive and coherent view of an enterprise. Each model is important and complementary but not the same.



Business Model Management.

What is a business management model?

A business model is a company's core strategy for profitably doing business. Models generally include information like products or services the business plans to sell, target markets, and any anticipated expenses. The two levers of a business model are pricing and costs.

Business Process Management.

Business process management (BPM) is a discipline that uses various methods to discover, model, analyze, measure, improve and optimize business processes. A business process coordinates the behavior of people, systems, information and things to produce business outcomes in support of a business strategy.

What is an example of business process management? Image result for Business Process Management.

This can mean onboarding a new employee, for example, or shipping a product to a customer. “Processes” shouldn't be confused with either projects or tasks, though. A project is usually a one-time thing. You could be, for example, creating a marketing strategy for a client.

Business Process Modeling.

Business process modeling is the graphical representation of a company's business processes or workflows, as a means of identifying potential improvements. This is usually done through different graphing methods, such as the flowchart, data-flow diagram, etc.

What is the importance of business process modeling?

Business process modeling facilitates this by helping managers and executives ensure that business processes are consistent, and enable business execution toward achieving the overall strategy and goals of the organization.

Enterprise Architecture.

Enterprise architecture (EA) is a discipline for proactively and holistically leading enterprise responses to disruptive forces by identifying and analyzing the execution of change toward desired business vision and outcomes.

What are the six basic elements of enterprise architecture?

Architecture management. Each enterprise will need an oversight team for the architecture. ...

Architecture framework. ...

What is a Architecture Framework?

An architecture framework provides the minimum set. of principles, practices, and requirements for creating and using artifacts describing the. system's architecture.10,11. The Infrastructure Architecture Framework (IAF) is primarily an architectural descriptive exercise, rather than a system engineering process.

What are the four main enterprise architecture frameworks?

Image result for Architecture framework. ...

In particular, the article focuses on the four most widely known EA frameworks: the Zachman Framework, FEAF, DoDAF and TOGAF.

Implementation methodology. ...

The implementation methodology is the method by which the projects are technically and operationally implemented in the field, most often by using contractors or subcontractors. Typical implementation models are Energy Performance Contracting, Energy Supply Contracting and Separate Contractor Based.

What are the different implementation methodologies? Implementation Methodologies

Direct cutover. In the direct-cutover implementation methodology, the organization selects a particular date that the old system will not be used anymore. ...

Pilot implementation. ...

A pilot implementation allows an organization to validate its approach for full application deployment. Executing an application pilot can uncover operability issues associated with production-like conditions and provide an opportunity to address these issues before full application roll out.

Parallel operation. ...

Parallel operation is the connec- tion of two or more power sources of the same size and output volt- age to obtain a higher output cur- rent. When connecting in parallel the outputs must be connected together keeping polarity in mind. The output voltage remains the same, but the current increases.

Phased implementation.

As the term suggest, “phased implementation” is the project planning strategy where anything new, such as a software solution, is rolled out in stages rather than all at once. The decision to purchase new software or a new system for your organization is significant.

Documentation artifacts. ...

What is a document artifact?

A Document Artifact is similar to a word processor file that can be created and stored directly inside the model to record structured documentation. It is a light-weight and convenient alternative to creating a document in your favorite word processor.

Architecture repository. ...

The Architecture Repository is a software tool that stores the important architectural input and output, including Architectures themselves, the elements of which they are composed, standards,

references, principles and the Governance Register.

Associated best practices.

1. Business model innovation must come from the leadership team.
2. Business model innovation must be separated from the current business.
3. Business model innovation must be human-centered.
4. Business model innovation requires a design approach.
5. Business model prototypes must be tested in the real world.
6. Business model innovation requires different metrics.
7. Tell the story.

Enterprise Architecture Software.

What is enterprise architecture software?

An enterprise architecture (EA) is a conceptual blueprint that defines the structure and operation of organizations. The intent of enterprise architecture is to determine how an organization can effectively achieve its current and future objectives.

Enterprise Portfolio Management.

Enterprise Portfolio Management (EPM) is an integrated portfolio management approach that tightly manages strategic planning against the various portfolios of interdependent assets, like product portfolios and project portfolios.

It is used to verify if, stakeholders have a shared understanding of the proposed “To-be of the solution.



Analyzing requirements is a part of business modelling process and it forms the core focus area. Functional Requirements are gathered during the “Current state”. These requirements are provided by the stakeholders regarding the business processes, data, and business rules that describe the desired functionality which will be designed in the Future State.

**What is database? Explain about different databases.**

A database is an organized collection of data, so that it can be easily accessed and managed.

You can organize data into tables, rows, columns, and index it to make it easier to find relevant information.

Database handlers create a database in such a way that only one set of software program provides access of data to all the users.

The main purpose of the database is to operate a large amount of information by storing, retrieving, and managing data.

There are many dynamic websites on the World Wide Web nowadays which are handled through databases. For example, a model that checks the availability of rooms in a hotel. It is an example of a dynamic website that uses a database.

There are many databases available like MySQL, Sybase, Oracle, MongoDB, Informix, PostgreSQL, SQL Server, etc.

Modern databases are managed by the database management system (DBMS).

SQL or Structured Query Language is used to operate on the data stored in a database. SQL depends on relational algebra and tuple relational calculus.

A cylindrical structure is used to display the image of a database.

## Network data model

Charles Bachman developed the first DBMS at Honeywell called Integrated Data Store (IDS). It was developed in the early 1960s, but it was standardized in 1971 by the CODASYL group (Conference on Data Systems Languages).

In this model, files are related as owners and members, like to the common network model. Network data model identified the following components:

Network schema (Database organization) Sub-schema (views of database per user) Data management language (procedural)

This model also had some limitations like system complexity and difficult to design and maintain.

## Relational Database

1970 - Present: It is the era of Relational Database and Database Management. In 1970, the relational model was proposed by E.F. Codd.

Relational database model has two main terminologies called instance and schema. The instance is a table with rows or columns

Schema specifies the structure like name of the relation, type of each column and name. This model uses some mathematical concept like set theory and predicate logic.

The first internet database application had been created in 1995.

During the era of the relational database, many more models had introduced like object-oriented model, object-relational model, etc.

## Cloud database

Cloud database facilitates you to store, manage, and retrieve their structured, unstructured data via a cloud platform. This data is accessible over the Internet. Cloud databases are also called a database as service (DBaaS) because they are offered as a managed service.

Some best cloud options are:

AWS (Amazon Web Services) Snowflake Computing

Oracle Database Cloud Services Microsoft SQL server

Google cloud spanner Advantages of cloud database Lower costs

Generally, company provider does not have to invest in databases. It can maintain and support one or more data centers.

Automated

Cloud databases are enriched with a variety of automated processes such as recovery, failover, and auto-scaling.

Increased accessibility

You can access your cloud-based database from any location, anytime. All you need is just an internet connection.

## NoSQL Database

A NoSQL database is an approach to design such databases that can accommodate a wide variety of data models. NoSQL stands for "not only SQL." It is an alternative to traditional relational databases in which data is placed in tables, and data schema is perfectly designed before the database is built.

NoSQL databases are useful for a large set of distributed data. Some examples of NoSQL database system with their category are:

MongoDB, CouchDB, Cloudant (Document-based) Memcached, Redis, Coherence (key-value store) HBase, Big Table, Accumulo (Tabular)

Advantage of NoSQL High Scalability

NoSQL can handle an extensive amount of data because of scalability. If the data grows, NoSQL database scale it to handle that data in an efficient manner.

High Availability

NoSQL supports auto replication. Auto replication makes it highly available because, in case of any failure, data replicates itself to the previous consistent state.

Disadvantage of NoSQL Open source

NoSQL is an open-source database, so there is no reliable standard for NoSQL yet. Management challenge

Data management in NoSQL is much more complicated than relational databases. It is very challenging to install and even more hectic to manage daily.

GUI is not available

GUI tools for NoSQL database are not easily available in the market. Backup

Backup is a great weak point for NoSQL databases. Some databases, like MongoDB, have no powerful approaches for data backup.

## The Object-Oriented Databases

The object-oriented databases contain data in the form of object and classes. Objects are the real- world entity, and types are the collection of objects. An object-oriented database is a combination of relational model features with objects oriented principles. It is an alternative implementation

to that of the relational model.

Object-oriented databases hold the rules of object-oriented programming. An object-oriented database management system is a hybrid application.

The object-oriented database model contains the following properties. Object-oriented programming properties

Objects Classes Inheritance Polymorphism Encapsulation

Relational database properties Atomicity

Consistency Integrity Durability Concurrency Query processing Graph Databases

A graph database is a NoSQL database. It is a graphical representation of data. It contains nodes and edges. A node represents an entity, and each edge represents a relationship between two edges. Every node in a graph database represents a unique identifier.

Graph databases are beneficial for searching the relationship between data because they highlight the relationship between relevant data.

Graph databases are very useful when the database contains a complex relationship and dynamic schema.

It is mostly used in supply chain management, identifying the source of IP telephony. DBMS (Data Base Management System)

Database management System is software which is used to store and retrieve the database. For example, Oracle, MySQL, etc.; these are some popular DBMS tools.

DBMS provides the interface to perform the various operations like creation, deletion,

modification, etc.

DBMS allows the user to create their databases as per their requirement.

DBMS accepts the request from the application and provides specific data through the operating system.

DBMS contains the group of programs which acts according to the user instruction. It provides security to the database.

Advantage of DBMS Controls redundancy

It stores all the data in a single database file, so it can control data redundancy. Data sharing

An authorized user can share the data among multiple users. Backup

It providesBackup and recovery subsystem. This recovery system creates automatic data from system failure and restores data if required.

Multiple user interfaces

It provides a different type of user interfaces like GUI, application interfaces. Disadvantage of DBMS

Size

It occupies large disk space and large memory to run efficiently. Cost

DBMS requires a high-speed data processor and larger memory to run DBMS software, so it is costly.

Complexity

DBMS creates additional complexity and requirements.



1. Centralized Database

It is the type of database that stores data at a centralized database system. It comforts the users to access the stored data from different locations through several applications. These applications contain the authentication process to let users access data securely. An example of a Centralized database can be Central Library that carries a central database of each library in a college/university.

Advantages of Centralized Database

It has decreased the risk of data management, i.e., manipulation of data will not affect the core data.

Data consistency is maintained as it manages data in a central repository.

It provides better data quality, which enables organizations to establish data standards. It is less costly because fewer vendors are required to handle the data sets.

Disadvantages of Centralized Database

The size of the centralized database is large, which increases the response time for fetching the data.

It is not easy to update such an extensive database system.

If any server failure occurs, entire data will be lost, which could be a huge loss.

1. Distributed Database

Unlike a centralized database system, in distributed systems, data is distributed among different database systems of an organization. These database systems are connected via communication links. Such links help the end-users to access the data easily. Examples of the Distributed database are Apache Cassandra, HBase, Ignite, etc.



We can further divide a distributed database system into:

Types of Databases

Homogeneous DDB: Those database systems which execute on the same operating system and use the same application process and carry the same hardware devices.

Heterogeneous DDB: Those database systems which execute on different operating systems under different application procedures, and carries different hardware devices.

Advantages of Distributed Database

Modular development is possible in a distributed database, i.e., the system can be expanded by including new computers and connecting them to the distributed system.

One server failure will not affect the entire data set.

1. Relational Database

This database is based on the relational data model, which stores data in the form of rows(tuple) and columns(attributes), and together forms a table(relation). A relational database uses SQL for storing, manipulating, as well as maintaining the data. E.F. Codd invented the database in 1970. Each table in the database carries a key that makes the data unique from others. Examples of Relational databases are MySQL, Microsoft SQL Server, Oracle, etc.

Properties of Relational Database

There are following four commonly known properties of a relational model known as ACID properties, where:

A means Atomicity: This ensures the data operation will complete either with success or with failure. It follows the 'all or nothing' strategy. For example, a transaction will either be committed or will abort.

C means Consistency: If we perform any operation over the data, its value before and after the operation should be preserved. For example, the account balance before and after the transaction should be correct, i.e., it should remain conserved.

I means Isolation: There can be concurrent users for accessing data at the same time from the database. Thus, isolation between the data should remain isolated. For example, when multiple transactions occur at the same time, one transaction effects should not be visible to the other tranactions in the database.

D means Durability: It ensures that once it completes the operation and commits the data, data changes should remain permanent.

1. NoSQL Database

Non-SQL/Not Only SQL is a type of database that is used for storing a wide range of data sets. It is not a relational database as it stores data not only in tabular form but in several different ways. It came into existence when the demand for building modern applications increased. Thus, NoSQL presented a wide variety of database technologies in response to the demands. We can further divide a NoSQL database into the following four types:



Key-value storage: It is the simplest type of database storage where it stores every single item as a key (or attribute name) holding its value, together.

Document-oriented Database: A type of database used to store data as JSON-like document. It helps developers in storing data by using the same document-model format as used in the application code.

Graph Databases: It is used for storing vast amounts of data in a graph-like structure. Most commonly, social networking websites use the graph database.

Wide-column stores: It is similar to the data represented in relational databases. Here, data is stored in large columns together, instead of storing in rows.

Advantages of NoSQL Database

It enables good productivity in the application development as it is not required to store data in a structured format.

It is a better option for managing and handling large data sets. It provides high scalability.

Users can quickly access data from the database through key-value.

1. Cloud Database

A type of database where data is stored in a virtual environment and executes over the cloud computing platform. It provides users with various cloud computing services (SaaS, PaaS, IaaS, etc.) for accessing the database. There are numerous cloud platforms, but the best options are:

Amazon Web Services(AWS) Microsoft Azure

Kamatera PhonixNAP ScienceSoft

Google Cloud SQL, etc.

1. Object-oriented Databases

The type of database that uses the object-based data model approach for storing data in the database system. The data is represented and stored as objects which are similar to the objects used in the object-oriented programming language.

1. Hierarchical Databases

It is the type of database that stores data in the form of parent-children relationship nodes. Here, it organizes data in a tree-like structure.



Data get stored in the form of records that are connected via links. Each child record in the tree will contain only one parent. On the other hand, each parent record can have multiple child records.

1. Network Databases

It is the database that typically follows the network data model. Here, the representation of data is in the form of nodes connected via links between them. Unlike the hierarchical database, it allows each record to have multiple children and parent nodes to form a generalized graph structure.

1. Personal Database

Collecting and storing data on the user's system defines a Personal Database. This database is basically designed for a single user.

Advantage of Personal Database It is simple and easy to handle.

It occupies less storage space as it is small in size.

1. Operational Database

The type of database which creates and updates the database in real-time. It is basically designed for executing and handling the daily data operations in several businesses. For example, An

organization uses operational databases for managing per day transactions.

1. Enterprise Database

Large organizations or enterprises use this database for managing a massive amount of data. It helps organizations to increase and improve their efficiency. Such a database allows simultaneous access to users.

Advantages of Enterprise Database:

Multi processes are supportable over the Enterprise database. It allows executing parallel queries on the system.

**What are different types of data? Explain with examples**

Data has so much importance in our life, so it is important to properly store and process this without any error. When dealing with datasets, the category of data plays an important role to determine which preprocessing strategy would work for a particular set to get the right results or which type of statistical analysis should be applied for the best results. Let’s dive into some of the commonly used categories of data.

Qualitative Data Type

Qualitative or Categorical Data describes the object under consideration using a finite set of discrete classes. It means that this type of data can’t be counted or measured easily using numbers and therefore divided into categories. The gender of a person (male, female, or others) is a good example of this data type.

These are usually extracted from audio, images, or text medium. Another example can be of a smartphone brand that provides information about the current rating, the color of the phone, category of the phone, and so on. All this information can be categorized as Qualitative data. There are two subcategories under this:

Nominal

These are the set of values that don’t possess a natural ordering. Let’s understand this with some examples. The color of a smartphone can be considered as a nominal data type as we can’t compare one color with others.

It is not possible to state that ‘Red’ is greater than ‘Blue’. The gender of a person is another one where we can’t differentiate between male, female, or others. Mobile phone categories whether it is midrange, budget segment, or premium smartphone is also nominal data type.

Ordinal

These types of values have a natural ordering while maintaining their class of values. If we consider the size of a clothing brand then we can easily sort them according to their name tag in the order of small < medium < large. The grading system while marking candidates in a test can

also be considered as an ordinal data type where A+ is definitely better than B grade.

These categories help us deciding which encoding strategy can be applied to which type of data. Data encoding for Qualitative data is important because machine learning models can’t handle these values directly and needed to be converted to numerical types as the models are mathematical in nature.

For nominal data type where there is no comparison among the categories, one-hot encoding can be applied which is similar to binary coding considering there are in less number and for the ordinal data type, label encoding can be applied which is a form of integer encoding.

Quantitative Data Type

This data type tries to quantify things and it does by considering numerical values that make it countable in nature. The price of a smartphone, discount offered, number of ratings on a product, the frequency of processor of a smartphone, or ram of that particular phone, all these things fall under the category of Quantitative data types.

The key thing is that there can be an infinite number of values a feature can take. For instance, the price of a smartphone can vary from x amount to any value and it can be further broken down based on fractional values. The two subcategories which describe them clearly are:

Discrete

The numerical values which fall under are integers or whole numbers are placed under this category. The number of speakers in the phone, cameras, cores in the processor, the number of sims supported all these are some of the examples of the discrete data type.

Continuous

The fractional numbers are considered as continuous values. These can take the form of the operating frequency of the processors, the android version of the phone, wifi frequency, temperature of the cores, and so on.

**Explain about different types of variables.**

There exist 4 types of variables in the field of Data Science which are listed below: Numerical

Categorical DateTime Mixed

Understanding the type of variable is very much important to choose the corresponding processing technique for that in Data Science.

Now, that being said, let’s proceed towards the explanation of each type of variable. Numerical Variables

This is the category of variables that deals with the numbers only. This category can be further divided into 2 sub-categories.

Discrete Continuous

Based on the value of the numerical variable, it can be sub-categorized into Discrete or continuous.

Discrete Numerical Variable

This is the category of the variables that only contain a discrete quantity, that is integers. No fractions are allowed in this category.

Few examples of this category are:

Whole numbers or Natural Numbers

The number of laptops you own, (here you will definitely have an integer as an answer, no one can say that it has 1.1 laptops, it will be obviously 1, 2, 3, etc.)

The number of shirts you have. The number of houses you own. The number of vehicles you own.

The number of children in the family. The number of pets in the family.

The number of bank accounts one has. 💰 The number of relationships someone has. and many more…

Continuous Numerical Variables

This is the category of the numerical variables that deal with the continuous quantities or fractional quantities.

Few examples of this category are:

The amount of your postpaid mobile bill. (Generally, it is not an integer, it is like 500.50, etc) Total time spent on watching a web series (120.5 seconds, etc)

Total amount spent on ordering food online (1500.57 INR, etc) The interest rate on loan.

Weight of an individual. (It is not always an integer, mostly it is like 90.5 kg, etc) and many more…

These are the type of variables that falls under the category of the Numerical variables. Categorical Variables

This is the category of the Variables that deals with the categories. It is also divided into 2 subcategories.

Ordinal Nominal

Based on the value of the categorical variable, it can be sub-categorized into Nominal or Ordinal. Ordinal Categorical Variable

Those variables that have an order associated with them are known as Ordinal Categorical variables or categorized into Ordinal Categorical Variables. The order is just been observed in the values that a particular variable is holding.

Few examples of this category are:

Days of the week (here there is a specific order => Monday always comes before Tuesday and so on…)

A particular Car series of any most popular companies like Audi, etc. Here, the car model is associated with a specific price, therefore based on the model of the car, we can categorize the car series in an order of either ascending or descending price bracket.

Educational Degrees (Ph.D. always comes after master's degree & so on…) Grades obtained in college.

Steps to create a mobile phone. (This process definitely has ordered steps that cannot be altered.) and many more…

Nominal Categorical variables

Those variables that do not have any order associated with them are known as Nominal Categorical variables or categorized into Nominal Categorical Variables.

Few examples of this category are: Country Names

City Names Sex / Gender

Operating Systems

Display Types (Retina, IPS, AMOLED, etc) and many more…

These are the type of variables that falls under the category of the Categorical variables.

Note (Insight): There are very few people who know that categorical variables can also contain numbers, they do not always contain strings. For example, The class of plane or train in which you are traveling. There can be first class, second class, etc. of AC in the train, whereas different classes like economy, business, etc for plain can be first encoded & then they will be in numerical order & still be a categorical variable (ordinal categorical variable to be precise).

DateTime Variable

This category of variable deals with the date & time aspects. This category can contain the type of values mentioned below:

Only having a date. Only having time.

Having both date & time.

Few examples of this category are:

Birthdate

Time of boarding the plane Timestamp

Time on which the log of the system is generated. Date of Application.

Date of Graduation.

Date of ordering a product online. and many more…

These are the type of variables that falls under the category of the DateTime variables. Mixed Variables

This category of variables deals with a collection of multiple values for the multiple observations of a specific variable. This category can also be divided into 2 different categories mentioned

below:

Numbers or labels/strings in different observations. Numbers & Labels/strings in the same observation. Examples of category 1:

The number of credit cards a person owns ( The value of this variable can be either a number for some observations, & also it can be “unknown” for some other observations due to any reason, therefore, it includes numbers and strings for the different observations).

Performance of a student (Here someone can mention the CGPA, percentage, or grades also, which includes numbers as well as labels/strings).

and many more… Examples of category 2:

Cabin Number (A1,A3, E5, etc)

Vehicle Registration number

What are data modeling techniques?

Image result for data modelling techniques in data analytics - geeksforgeeks

Data Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database.

Data modeling comprises the methodologies of creating data representations for data visualization, which allows users to better understand the global values and associations that create the data’s potential underlying value.

Data modeling is used to define and analyze the data requirements to support data mining and data analytics. The data modeling process involves professional data modelers working closely with business stakeholders as well as potential users of a system.

**What is Data Modeling? Write about different data modeling techniques.**

A data model is a visual representation of data elements and the relations between them. It is the fundamental method used to leverage abstraction in an information system. Data models define the logical structure of data, how they are connected and how the data are processed and stored in information systems.

Data models provide the conceptual tools for describing the model of an information system at the data abstraction level. It enables users to decide how data will be stored, leveraged, updated, accessed, and shared across an organization.

Data models may also provide a portrait of the final system, and how it will look after implementation. It helps in the development of effective information systems by supporting the definition and structure of data on behalf of relevant business processes. It facilitates the communication of business and technical needs for the development of an action plan.

Earlier data models could be “flat data models,” in which data was displayed in the same plane and was therefore limited; flat models could introduce duplications and anomalies. Now, data models are more likely 3-D, and are extremely effective and useful to the development of business and IT strategy.

Conceptual Schema

A conceptual data model or conceptual schema is a high-level description of information used in developing an information system, such as database structures. It is a map of concepts and the relationships between them, typically including only the main concepts and the main relationships.

The conceptual schema describes the semantics of an organization and represents a series of assertions. It may exist on various levels of abstraction and hides the internal details of physical storage structures and instead focuses on describing entities, data types, relationships, and constraints. The conceptual schema design process takes information requirements for an application as input and produces a schema that is expressed in a form of conceptual modeling notation. Below is an example of a conceptual schema:



Logical Schema

A logical data model or logical schema is a representation of the abstract structure of the information domain that defines all the logical constraints applied to the data stored. A specific problem domain expresses information system management or storage technology independently and defines views, tables, and integrity constraints. A logical schema defines the design of the information system at its logical level.

Software developers, as well as administrators, tend to work at this level. Although the data can be described as data records that are stored in the form of data structures, the data structure implementation and other internal details are hidden at this level. Below is an example of a logical schema:



Physical Schema

A physical data model or physical schema is a representation of an implementation design; it defines data abstraction within physical parameters.

A complete physical schema includes all the information system artifacts required to achieve performance goals or create relationships between data, such as indexes, linking tables, and constraint definitions. Analysts can use a physical schema to calculate storage estimates, and this may include specific storage allocation details for an information system.



**What are Data Modeling Techniques?**

There are various techniques to achieve data modeling successfully, though the basic concepts remain the same across techniques. Some popular data modeling techniques include Hierarchical, Relational, Network, Entity-relationship, and Object-oriented.

Hierarchical Technique

The Hierarchical data modeling technique follows a tree-like structure where its nodes are sorted in a particular order. A hierarchy is an arrangement of items represented as “above,” “below,” or “at the same level as” each other. Hierarchical data modeling technique was implemented in the IBM Information Management System (IMS) and was introduced in 1966.

It was a popular concept in a wide variety of fields, including computer science, mathematics, design, architecture, systematic biology, philosophy, and social sciences. But it is rarely used now due to the difficulties of retrieving and accessing data.

Relational Technique

The relational data modeling technique is used to describe different relationships between entities, which reduces the complexity and provides a clear overview. The relational model was first proposed as an alternative to the hierarchical model by IBM researcher Edgar F. Codd in 1969. It has four different sets of relations between the entities: one to one, one to many, many to one, and many to many.

Network Technique

The network data modeling technique is a flexible way to represent objects and underlying

relationships between entities, where the objects are represented inside nodes and the relationships between the nodes is illustrated as an edge. It was inspired by the hierarchical technique and was originally introduced by Charles Bachman in 1969.

The network data modeling technique makes it easier to convey complex relationships as records and can be linked to multiple parent records.

Entity-relationship technique

The entity-relationship (ER) data modeling technique represents entities and relationships between them in a graphical format consisting of Entities, Attributes, and Relationships. The entities can be anything, such as an object, a concept, or a piece of data. The entity-relationship data modeling technique was developed for databases and introduced by Peter Chen in 1976. It is a high-level relational model that is used to define data elements and relationships in a sophisticated information system.

Object-Oriented Technique

The object-oriented data modeling technique is a construction of objects based on real-life scenarios, which are represented as objects. The object-oriented methodologies were introduced in the early 1990s’ and were inspired by a large group of leading data scientists.

It is a collection of objects that contain stored values, in which the values are nothing but objects. The objects have similar functionalities and are linked to other objects.



Data Modeling: An Integrated View

Data modeling is an essential technology for understanding relationships between data sets. The integrated view of conceptual, logical, and physical data models helps users to understand the information and ensure the right information is used across an entire enterprise.

Although data modeling can take time to perform effectively, it can save significant time and money by identifying errors before they occur. Sometimes a small change in structure may require modification of an entire application.

**What is Missing Imputation? Different methods to impute missing data.**

**Imputation**

What is Imputation? Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data/information of the dataset.

6 popular ways for data imputation for cross-sectional datasets ( Time-series dataset is a different story ).

1- Do Nothing:

That’s an easy one. You just let the algorithm handle the missing data. Some algorithms can factor in the missing values and learn the best imputation values for the missing data based on the training loss reduction (ie. XGBoost). Some others have the option to just ignore them (ie. LightGBM — use\_missing=false). However, other algorithms will panic and throw an error complaining about the missing values (ie. Scikit learn — LinearRegression). In that case, you will need to handle the missing data and clean it before feeding it to the algorithm.

Let’s see some other ways to impute the missing values before training:

Note: All the examples below use the California Housing Dataset from Scikit-learn. 2- Imputation Using (Mean/Median) Values:

This works by calculating the mean/median of the non-missing values in a column and then replacing the missing values within each column separately and independently from the others. It can only be used with numeric data.



Mean Imputation Pros:

Easy and fast.

Works well with small numerical datasets. Cons:

Doesn’t factor the correlations between features. It only works on the column level.

Will give poor results on encoded categorical features (do NOT use it on categorical features). Not very accurate.

Doesn’t account for the uncertainty in the imputations. Mean/Median Imputation

1. Imputation Using (Most Frequent) or (Zero/Constant) Values:

Most Frequent is another statistical strategy to impute missing values and YES!! It works with categorical features (strings or numerical representations) by replacing missing data with the most frequent values within each column.

Pros:

Works well with categorical features. Cons:

It also doesn’t factor the correlations between features. It can introduce bias in the data.

Most Frequent Imputation

Zero or Constant imputation — as the name suggests — it replaces the missing values with either zero or any constant value you specify



1. Imputation Using k-NN:

The k nearest neighbours is an algorithm that is used for simple classification. The algorithm uses ‘feature similarity’ to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. This can be very useful in making predictions about the missing values by finding the k’s closest neighbours to the observation with missing data and then imputing them based on the non- missing values in the neighbourhood. Let’s see some example code using Impyute library which provides a simple and easy way to use KNN for imputation:

KNN Imputation for California Housing Dataset How does it work?

It creates a basic mean impute then uses the resulting complete list to construct a KDTree. Then, it uses the resulting KDTree to compute nearest neighbours (NN). After it finds the k-NNs, it takes the weighted average of them.



Pros:

Can be much more accurate than the mean, median or most frequent imputation methods (It depends on the dataset).

Cons:

Computationally expensive. KNN works by storing the whole training dataset in memory. K-NN is quite sensitive to outliers in the data (unlike SVM)

1. Imputation Using Multivariate Imputation by Chained Equation (MICE)



Main steps used in multiple imputations [1]

This type of imputation works by filling the missing data multiple times. Multiple Imputations (MIs) are much better than a single imputation as it measures the uncertainty of the missing values in a better way. The chained equations approach is also very flexible and can handle different variables of different data types (ie., continuous or binary) as well as complexities such as bounds or survey skip patterns. For more information on the algorithm mechanics, you can refer to the Research Paper



MICE imputation using impyute

1. Imputation Using Deep Learning (Datawig):

This method works very well with categorical and non-numerical features. It is a library that learns Machine Learning models using Deep Neural Networks to impute missing values in a dataframe. It also supports both CPU and GPU for training.

Imputation using Datawig Pros:

Quite accurate compared to other methods.

It has some functions that can handle categorical data (Feature Encoder). It supports CPUs and GPUs.

Cons:

Single Column imputation.

Can be quite slow with large datasets.

You have to specify the columns that contain information about the target column that will be imputed.

Other Imputation Methods:

Stochastic regression imputation:

It is quite similar to regression imputation which tries to predict the missing values by regressing it from other related variables in the same dataset plus some random residual value.

Extrapolation and Interpolation:

It tries to estimate values from other observations within the range of a discrete set of known data points.

Hot-Deck imputation:

Works by randomly choosing the missing value from a set of related and similar variables.

In conclusion, there is no perfect way to compensate for the missing values in a dataset. Each strategy can perform better for certain datasets and missing data types but may perform much worse on other types of datasets. There are some set rules to decide which strategy to use for particular types of missing values, but beyond that, you should experiment and check which model works best for your dataset.

**Explain about need for Business Modeling.**

Every business or companies makes a plan for generating profit. They create a model for identifying products and services to sell, the market they want to target and also take into account anticipated expenses. This is known as business models.

Even if the business is already established or even if it is a new business, plan needs to be made. Businesses need to regularly update their plans and strategy as they need to take into accounts the challenges and trends for the future models.

## Importance of Business Models:

The business model helps to target the customer base for the company. It helps in making marketing strategies, projection of revenues and expenses taking into account the type of Business models and clienteles.

Every investor needs to review the business model in order to get [knowledge about the](https://www.elearnmarkets.com/blog/how-investors-evaluate-a-company?utm_campaign=blog_CTA&utm_medium=blogpage&utm_source=elearnmarkets_blog) [company’s competitive edge.](https://www.elearnmarkets.com/blog/how-investors-evaluate-a-company?utm_campaign=blog_CTA&utm_medium=blogpage&utm_source=elearnmarkets_blog) Understanding the business model helps the investors to have a better sense of financial data.

Evaluating the business model helps the investors to get the overall view about the company’s products, its business strategies and future prospects.

## Types of Business Models:

We will discuss here about 4 types of business models:

## Business -To- Business Models (B2B):

When the dealings or the transactions take place between two companies or the business then this type of business model is known as business to business models.

It has good [market predictability and more market stability.](https://www.elearnmarkets.com/blog/market-volatility-vs-market-risk?utm_campaign=blog_CTA&utm_medium=blogpage&utm_source=elearnmarkets_blog) Since under B2B sale is made in bulk amount this model leads to lower cost for the businesses.

The best example of this type of business model in India is IndiaMart InterMesh which is a wholesale B2B marketplace. It offers millions of products to its customers which includes consumer electronics, machinery, apparel and many more.

## Business -To-Consumer Models (B2C):

Business-2-consumer business model is a model that refers to businesses that sell their services or the products directly to the consumer who are the end users of the products or services.

There is an ongoing demand for the products as it provides the essential items. This thus eliminates the risk of fluctuation in demand and helps in maintaining consistency in the business. Since direct contact is there with the customer’s so information is shared with them directly and easily.

Customers are given products at a low price compared to its competitors for the business to run smoothly.

Example of business to consumer model is Avenue Supermart which provides goods directly to its customers.

## Subscription Based Models:

Any application based businesses or software companies have subscription based business models.They offer their product as a onetime purchase, in return company earns monthly or annual revenues.

This type of business model allows the company to earn regular income by giving the client the opportunity to pay for the cost of the purchase in 12 equal payments rather asking them to pay the wholesome amount at one go.

One of the leading examples is Infoedge for this type of business model.

## On-DEMAND BUSINESS MODEL

It is the most recent form of model which is made out on the need by answering immediately. Under this type of business model is prepared in such a way where all the questions will be answered by just a click of a button in seconds.

It is very much convenient and easy for customers as even before customers have visited the particular city they get their hotels or places booked.

One of the example is make my trip which allows the customers to plan the holidays and make the bookings in advance.

## Advantages of Business Models

* A good business models gives the company a competitive edge in the industry.
* A strong business model provides the company good reputation in the market place encouraging the investors to remain invested in the company.
* Making the business model strong leads to an ongoing business profit leading to increase in cash reserve and new investments.
* Proven business model brings a financial stability in the organization. Business models have disadvantages as well.

## Disadvantages of Business Model:

* Once a business model is created, then it restricts to implement new ideas for the product.
* Creating a business model is time consuming as lot of factors needs to be considered.
* There might be a chance that business model may turn out to be inaccurate.

Apart from the disadvantages, business model is mandatory to be prepared before starting of a new project.

## What is a good business model?

A good business model is the one that provides the company a competitive edge in the industry leading to good business profits.

## Why is a business model important?

Business model is important because it provides the investors the knowledge about the competitive edge of the company and provides better insight into working of the company. A strong business model leads to cash generation and future expansion.

## How do you create a business model?

Business model is created by identifying the products and services that will be sold in the market to be targeted like B2B, B2C, subscription based model or on demand market.

## What are the components of business model?

Business model includes information about company’s products, its target market and its future prospect related to its business type.