Module - 2

# Data Analytics

SAP

edunet foundation

code unnati

## Units for Discussion

Fundamentals of Data Analytics

**Unit - 1**

Data Analysis using NumPy

**Unit - 2**

Data Analysis using Pandas

**Unit - 3**

Data Visualization using Matplotlib

**Unit - 4**

Data Visualization using Seaborn

**Unit - 5**

Unit - 1

# Fundamentals of Data Analytics

# DISCLAIMER

The content is curated from online/offline resources and used for educational purpose only.

DMart

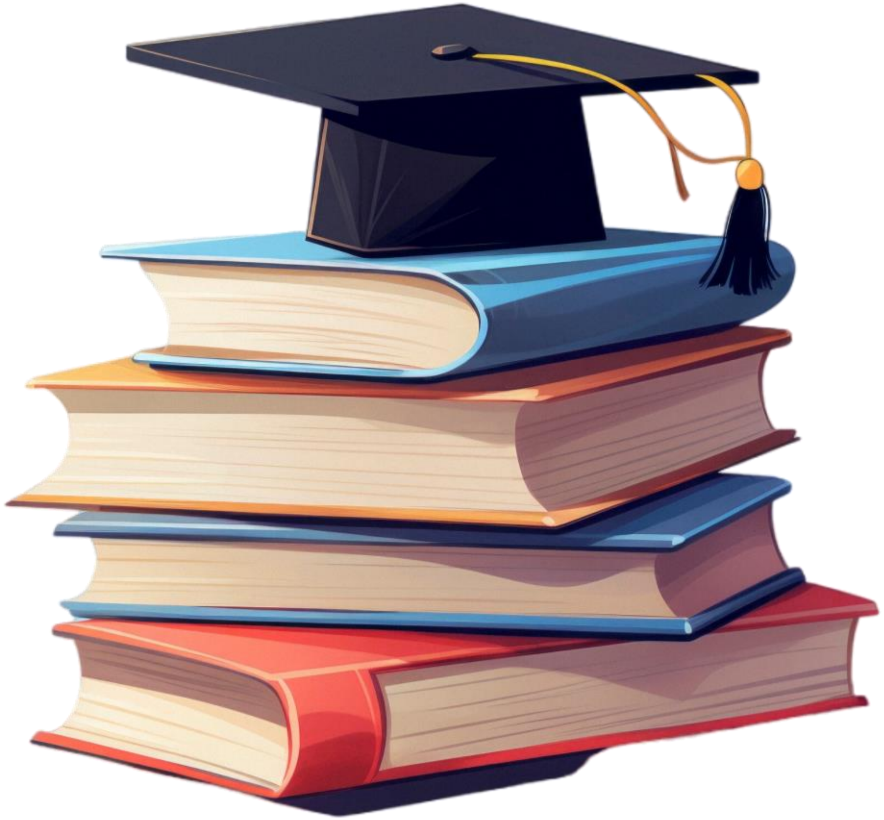
Walmart

SWIGGY

zomato

# Learning Objectives

- What is Data?
- What is Data Analytics?
- Why we need to learn Data Analytics?
- Types of Data Analytics
- Data Ecosystem
- What is Data Analysis?
- Key Steps in a Data Analysis Process
- Data Analytics vs. Data Analysis
- Data Warehouse
- NoSQL Database
- CRUD operation in MongoDB
- Pymongo



**Source :**

Reference

# What is Data?

- Data can be defined as a collection of facts and figures that are used for analysis or a survey.
- It is a series of representations of various values of that quantity.
- When data has been organized in a structured manner, it is referred to as information.

Source :

Reference

## Sources of Data

Relational Databases

APIs and Web Services

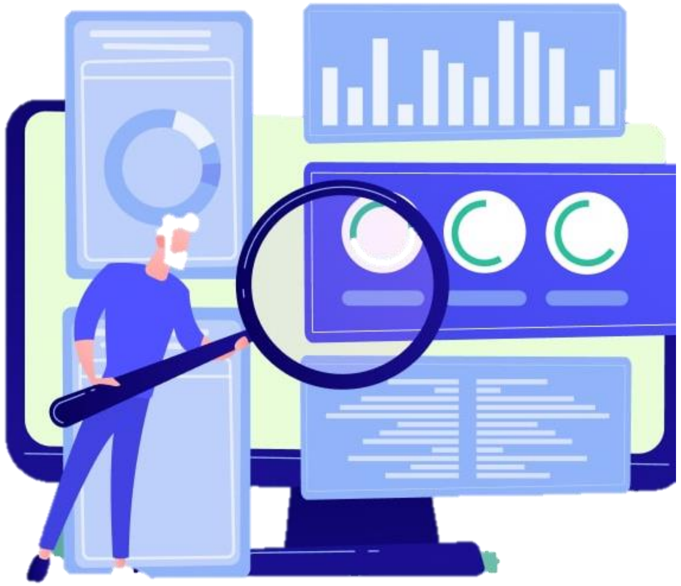
Data Streams, and Feeds.

Flat files and XML Datasets

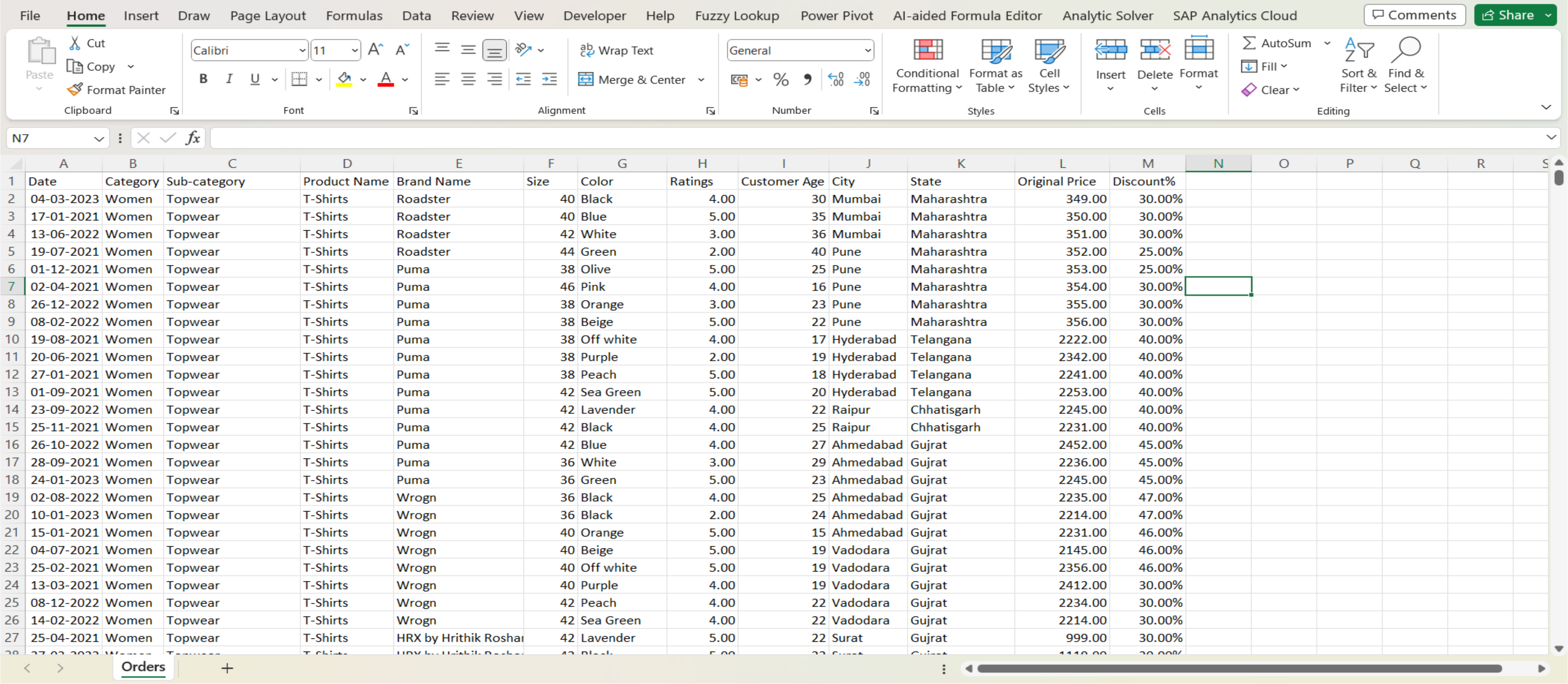
Web Scraping

# What is Data Analytics ?

- The term "data analytics" describes the methods used to analyze data in order to increase productivity and financial gain.
- It is the procedure of deriving valuable, useful insights from the analysis of raw data, which are then used to guide and inform wise business decisions.
- In order to analyze different behavioral patterns, data is extracted from a variety of sources, cleaned up, and categories.

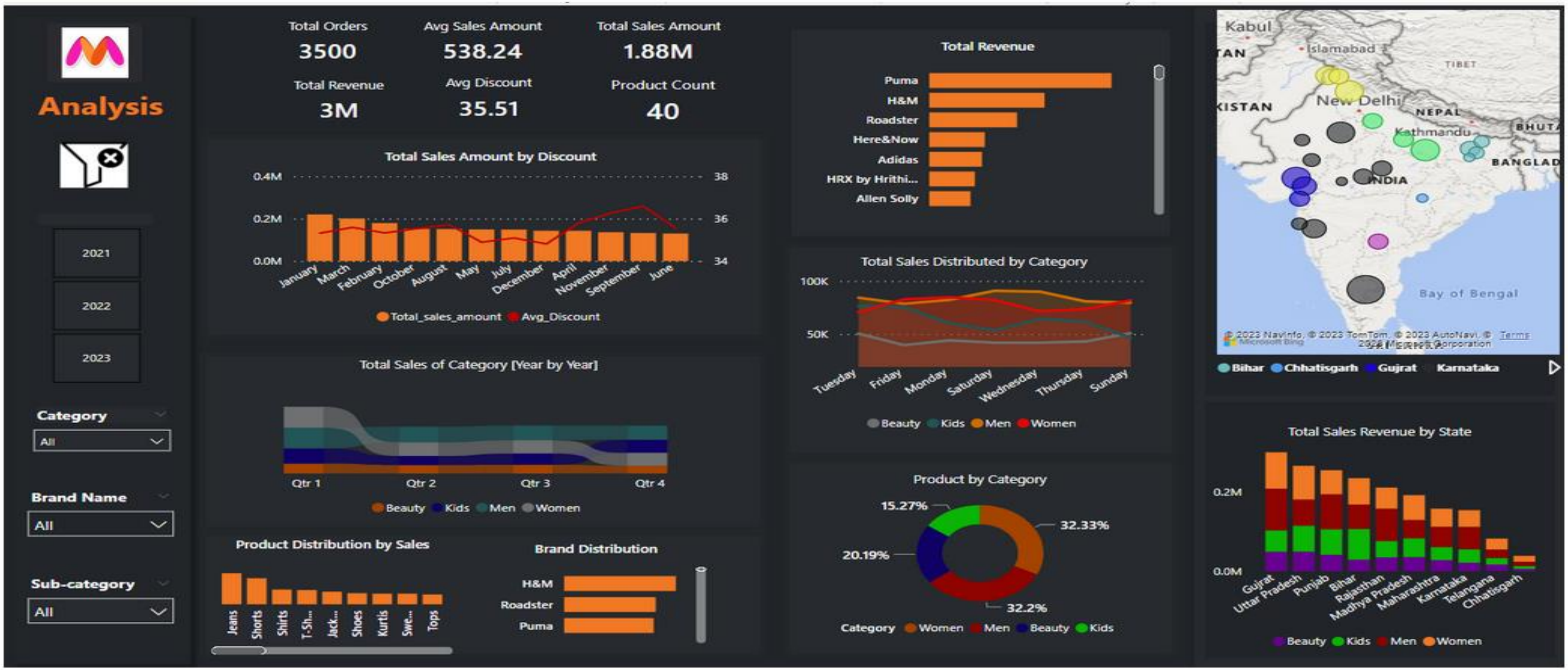


# Why we need to learn Data Analytics ?



| | Date | Category | Sub-category | Product Name | Brand Name | Size | Color | Ratings | Customer Age | City | State | Original Price | Discount% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 04-03-2023 | Women | Topwear | T-Shirts | Roadster | 40 | Black | 4.00 | 30 | Mumbai | Maharashtra | 349.00 | 30.00% |
| 3 | 17-01-2021 | Women | Topwear | T-Shirts | Roadster | 40 | Blue | 5.00 | 35 | Mumbai | Maharashtra | 350.00 | 30.00% |
| 4 | 13-06-2022 | Women | Topwear | T-Shirts | Roadster | 42 | White | 3.00 | 36 | Mumbai | Maharashtra | 351.00 | 30.00% |
| 5 | 19-07-2021 | Women | Topwear | T-Shirts | Roadster | 44 | Green | 2.00 | 40 | Pune | Maharashtra | 352.00 | 25.00% |
| 6 | 01-12-2021 | Women | Topwear | T-Shirts | Puma | 38 | Olive | 5.00 | 25 | Pune | Maharashtra | 353.00 | 25.00% |
| 7 | 02-04-2021 | Women | Topwear | T-Shirts | Puma | 46 | Pink | 4.00 | 16 | Pune | Maharashtra | 354.00 | 30.00% |
| 8 | 26-12-2022 | Women | Topwear | T-Shirts | Puma | 38 | Orange | 3.00 | 23 | Pune | Maharashtra | 355.00 | 30.00% |
| 9 | 08-02-2022 | Women | Topwear | T-Shirts | Puma | 38 | Beige | 5.00 | 22 | Pune | Maharashtra | 356.00 | 30.00% |
| 10 | 19-08-2021 | Women | Topwear | T-Shirts | Puma | 38 | Off white | 4.00 | 17 | Hyderabad | Telangana | 2222.00 | 40.00% |
| 11 | 20-06-2021 | Women | Topwear | T-Shirts | Puma | 38 | Purple | 2.00 | 19 | Hyderabad | Telangana | 2342.00 | 40.00% |
| 12 | 27-01-2021 | Women | Topwear | T-Shirts | Puma | 38 | Peach | 5.00 | 18 | Hyderabad | Telangana | 2241.00 | 40.00% |
| 13 | 01-09-2021 | Women | Topwear | T-Shirts | Puma | 42 | Sea Green | 5.00 | 20 | Hyderabad | Telangana | 2253.00 | 40.00% |
| 14 | 23-09-2022 | Women | Topwear | T-Shirts | Puma | 42 | Lavender | 4.00 | 22 | Raipur | Chhatisgarh | 2245.00 | 40.00% |
| 15 | 25-11-2021 | Women | Topwear | T-Shirts | Puma | 42 | Black | 4.00 | 25 | Raipur | Chhatisgarh | 2231.00 | 40.00% |
| 16 | 26-10-2022 | Women | Topwear | T-Shirts | Puma | 42 | Blue | 4.00 | 27 | Ahmedabad | Gujrat | 2452.00 | 45.00% |
| 17 | 28-09-2021 | Women | Topwear | T-Shirts | Puma | 36 | White | 3.00 | 29 | Ahmedabad | Gujrat | 2236.00 | 45.00% |
| 18 | 24-01-2023 | Women | Topwear | T-Shirts | Puma | 36 | Green | 5.00 | 23 | Ahmedabad | Gujrat | 2245.00 | 45.00% |
| 19 | 02-08-2022 | Women | Topwear | T-Shirts | Wrogn | 36 | Black | 4.00 | 25 | Ahmedabad | Gujrat | 2235.00 | 47.00% |
| 20 | 10-01-2023 | Women | Topwear | T-Shirts | Wrogn | 36 | Black | 2.00 | 24 | Ahmedabad | Gujrat | 2214.00 | 47.00% |
| 21 | 15-01-2021 | Women | Topwear | T-Shirts | Wrogn | 40 | Orange | 5.00 | 15 | Ahmedabad | Gujrat | 2231.00 | 46.00% |
| 22 | 04-07-2021 | Women | Topwear | T-Shirts | Wrogn | 40 | Beige | 5.00 | 19 | Vadodara | Gujrat | 2145.00 | 46.00% |
| 23 | 25-02-2021 | Women | Topwear | T-Shirts | Wrogn | 40 | Off white | 5.00 | 19 | Vadodara | Gujrat | 2356.00 | 46.00% |
| 24 | 13-03-2021 | Women | Topwear | T-Shirts | Wrogn | 40 | Purple | 4.00 | 19 | Vadodara | Gujrat | 2412.00 | 30.00% |
| 25 | 08-12-2022 | Women | Topwear | T-Shirts | Wrogn | 42 | Peach | 4.00 | 22 | Vadodara | Gujrat | 2234.00 | 30.00% |
| 26 | 14-02-2022 | Women | Topwear | T-Shirts | Wrogn | 42 | Sea Green | 4.00 | 22 | Vadodara | Gujrat | 2214.00 | 30.00% |
| 27 | 25-04-2021 | Women | Topwear | T-Shirts | HRX by Hrithik Roshai | 42 | Lavender | 5.00 | 22 | Surat | Gujrat | 999.00 | 30.00% |

# Why we need to learn Data Analytics ?



Analysis

Total Orders 3500

Avg Sales Amount 538.24

Total Sales Amount 1.88M

Total Revenue 3M

Avg Discount 35.51

Product Count 40

2021

2022

2023

Category: All

Brand Name: All

Sub-category: All

# Types of  Data Analytics

Descriptive Analytics

Diagnostic Analytics

Predictive Analytics

Prescriptive Analytics

Source :

Reference

# Descriptive Analytics

- The simplest type of analytics is descriptive analytics, which serves as the base for all other types.
- It enables you to quickly summarise what occurred or is happening by drawing trends from the raw data.
- When used in business, it gives the analyst a view of important metrics and measurements used by the company.

**Source :**

Reference

## Diagnostic Analytics

- This type compares concurrent trends or movements, reveals correlations between variables, and, when possible, establishes causal relationships.
- The question of why something happened can now be answered by comparing historical data to other data.
- Diagnostic analytics offers in-depth perceptions into a specific issue.

**Source :**

Reference

# Predictive Analytics

- The use of predictive analytics provides an answer to the question, "What might happen in the future?" by making predictions about present trends or events.
- It makes use of the outcomes of descriptive and diagnostic analytics to find clusters and exceptions as well as make predictions about future trends.

**Source :**

[Reference]

## Prescriptive Analytics

- Prescriptive analytics assists in providing recommendations for action.
- Prescriptive analytics helps by making suggestions for taking action.
- Prescriptive analytics is a type of advanced analytics that offers many benefits, such as sophisticated analysis based on deep learning or machine learning.

Source :

Reference

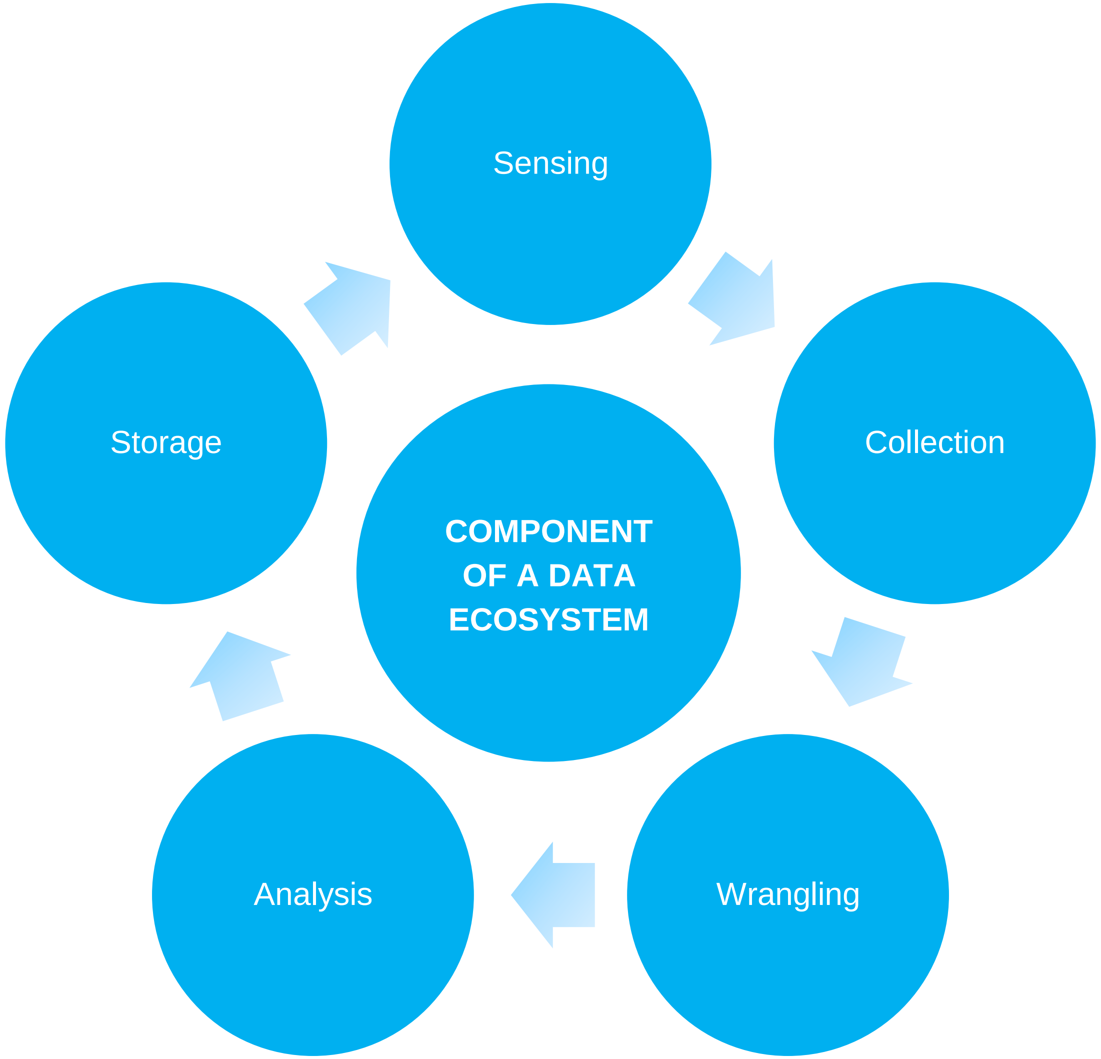
## Data Ecosystem

- A data ecosystem is a collection of business applications and infrastructure that is used to gather and analyze data.
- It enables businesses to develop better marketing, pricing, and business plans by helping them better understand their customers.
- All the programming languages, algorithms, applications, and underlying infrastructure that are used to gather, process, and store data are collectively referred to as the "data ecosystem".

**Source :**

www.freepik.com/

## Different Components of a Modern Data Ecosystem



# Key Players in the Data Ecosystem

Data Engineer

Data Scientists

Business Intelligence Analysts

Data Analysts

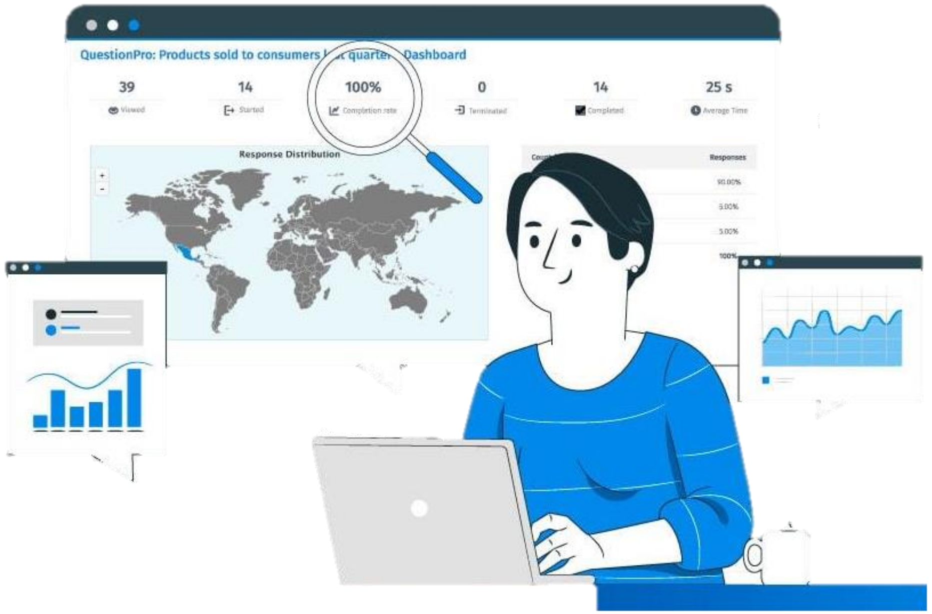
Business Analysts

**Source :**

[www.freepik.com/](www.freepik.com/)

# What is Data Analysis?

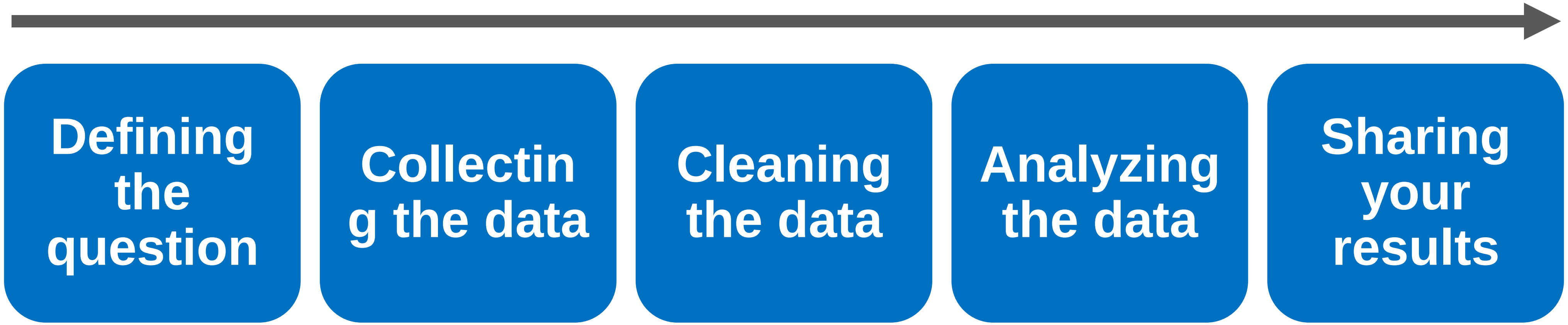
- Data analysis is the process of cleaning, transforming, and processing raw data in-order to extract actionable, relevant information that assists businesses in making informed decisions. The procedure reduces the risks associated with decision-making by providing useful insights and statistics, which are frequently presented in charts, images, tables, and graphs.



**Source :**

Reference

# Key Steps in a Data Analysis Process



Defining the question

Collecting the data

Cleaning the data

Analyzing the data

Sharing your results

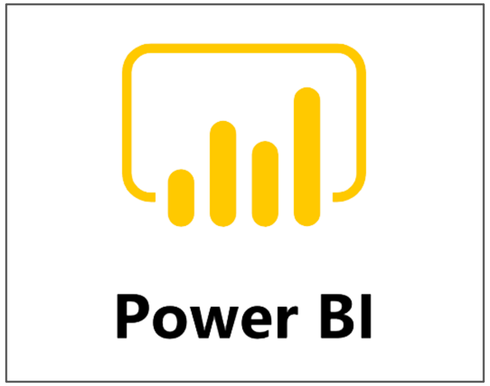
# Data Analytics vs. Data Analysis

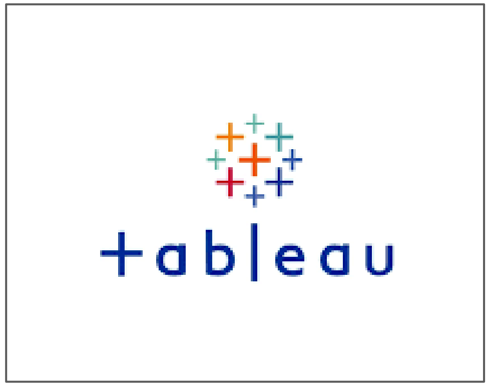
| Data Analytics | Data Analysis |
|---|---|
| Data analytics is a type of traditional or generic analytics that is used in businesses to make data-driven decisions. | Data analysis is a subset of analytics that is used in businesses to evaluate data and gain insights. |
| It has one or more users and is typically used for data collection and inspection. | It entailed defining the data, investigating it, cleaning it up, and modifying it to produce a useful result. |
| It is divided into several stages, such as data collection and business data inspection. | In order to process data, raw data must first be defined in a meaningful way before relevant information can be extracted from it via data cleaning and conversion. |
| It uses a variety of technologies to process data, including Tableau, Python, Excel, Google Analytics, and others. | It analyses data using a variety of tools, including SPARK, Google Fusion tables, Node XL, Excel, and others. |

**Source :**

Reference

## Popular Data Analysis Tools





# Database

- Database is the collection of organized data that is structured and stored electronically on a computer system.
- Databases can store data in the form of tables depending upon the type of database. The database's primary goal is to store a huge amount of data.

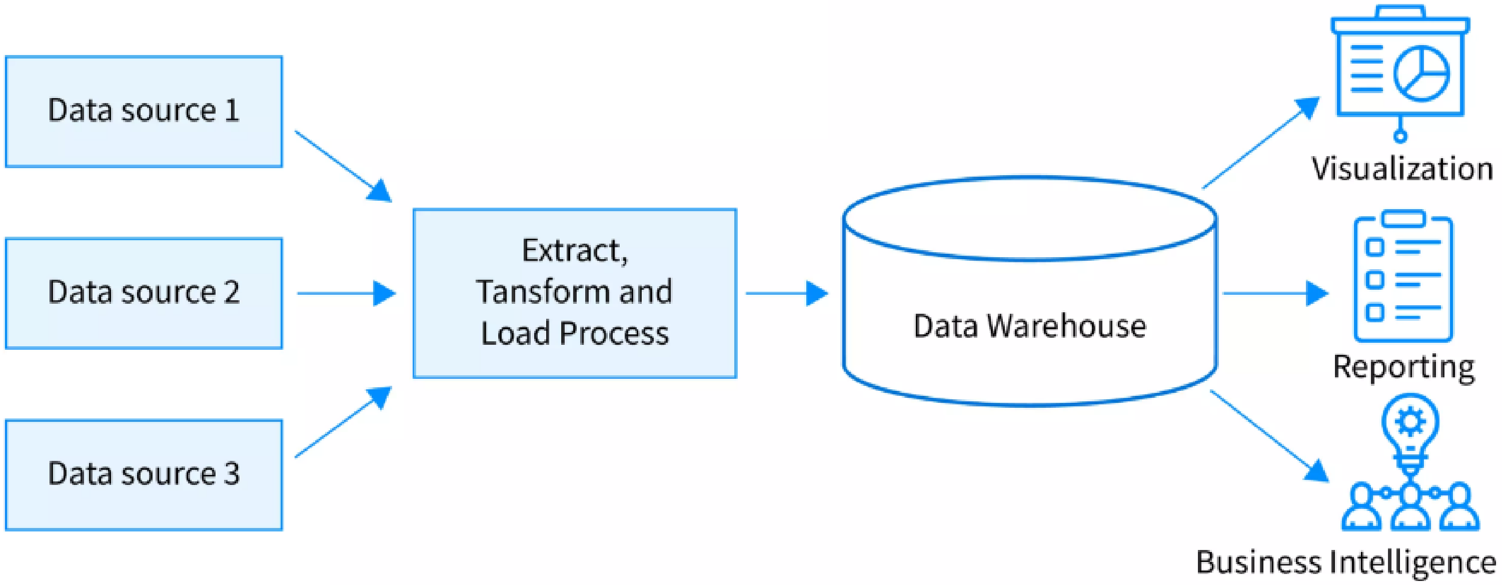
**Examples of some databases: MySQL, Oracle, MongoDB, PostgreSQL, SQL Server, etc.**

**Source :**

Reference

# What is  Data Warehouse?

- A data warehouse is a data management system that stores current and historical data from multiple sources in a business-friendly manner for easier insights and reporting.
- Data warehouses are typically used for business intelligence (BI), reporting and data analysis.

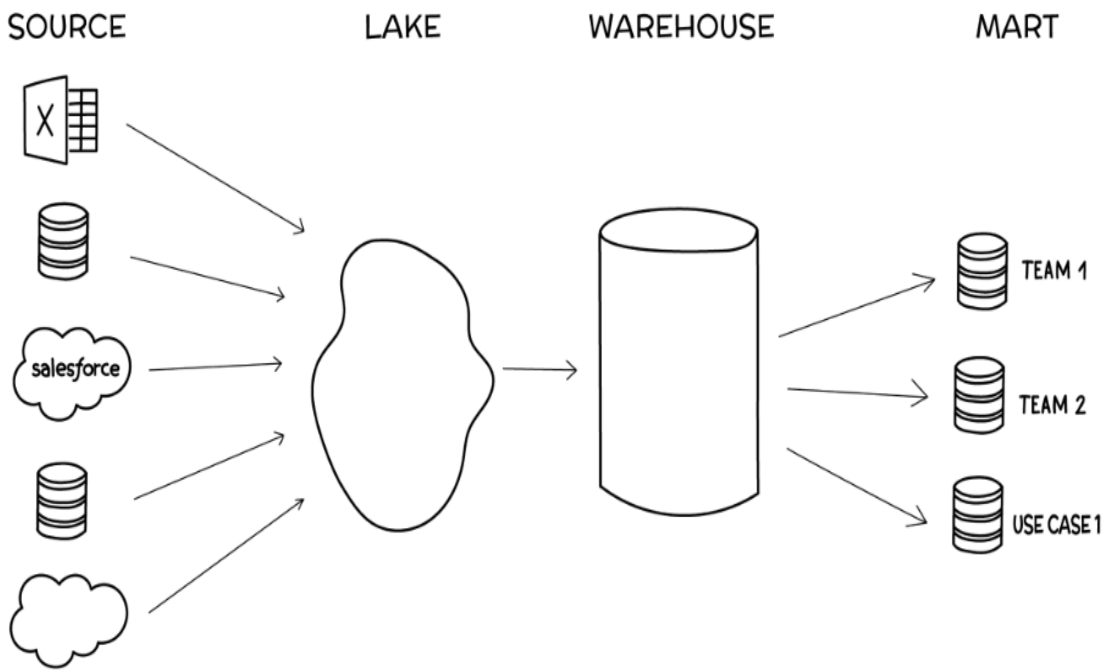


**Source :**

Reference

# Key Components of data warehouse?

- Source data is any individual set of data like databases, Excel spreadsheets, individual application reports, etc.
- A data lake serves as a central repository for all raw, unstructured (i.e., not organized) data.
- Like a data lake, a data warehouse centralizes your data, but it's well-organized and set up for efficient analysis.
- Data marts are curated data sets created for specific use cases.



Source :

[Reference](#)

# Introduction to NoSQL

- The term NoSQL, short for “not only SQL,” refers to non-relational databases that use a non-tabular format to store data, rather than in rule-based, relational tables like relational databases.
- NoSQL databases use a flexible schema model that supports a wide variety of unstructured data, such as documents, key-value, wide columns, graphs, and more.
- Organizations choose NoSQL databases for their flexibility, horizontal scalability, and ease of development.

**Source :**

Reference

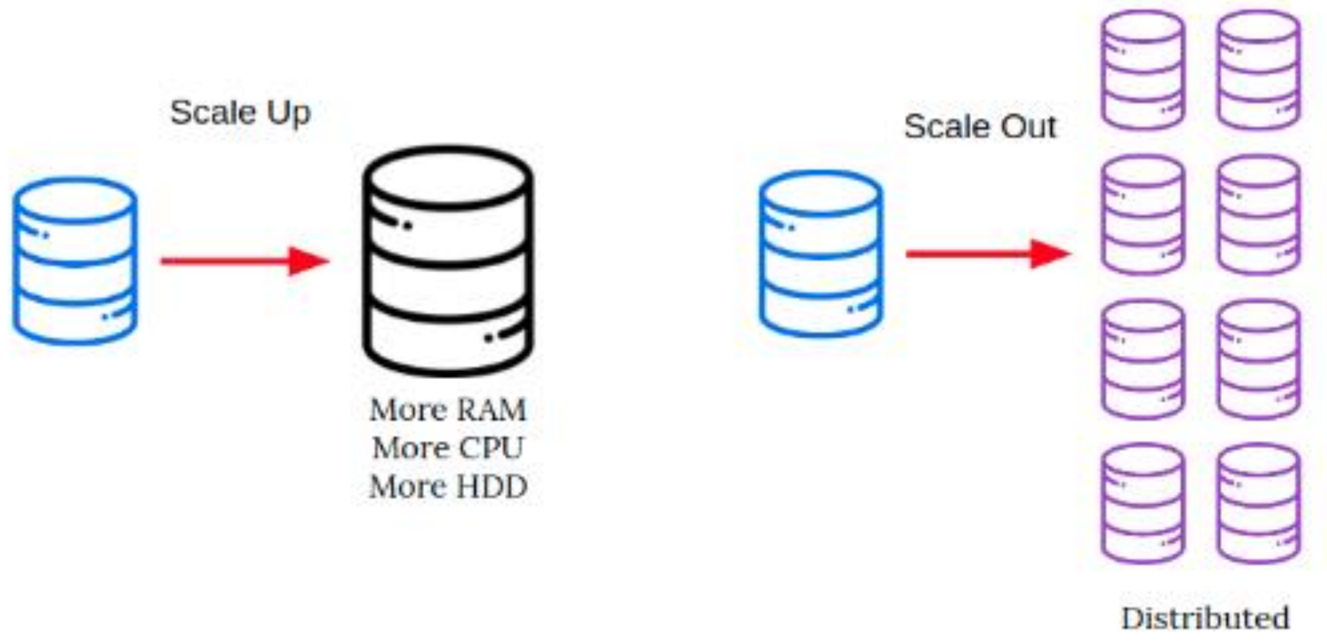
# SQL vs NoSQL

| Property | SQL Databases | NoSQL Databases |
|---|---|---|
| **Data model** | Relational | Nonrelational |
| **Structure** | Table-based, with columns and rows | Document based, key-value pairs, graph, or wide-column |
| **Query language** | Structured Query Language (SQL) | Varies from database to database |
| **ACID transactions** | Supported | Supported, depending on the specific NoSQL database |
| **Ability to add new properties** | Need to alter the schema first | Possible without disturbing anything |

# Features of NoSQL

Below are some of the important features of NoSQL:

- Dynamic Schemas
- Replication
- Auto Sharding
- Eliminated Downtime
- Distributed
- Simple API

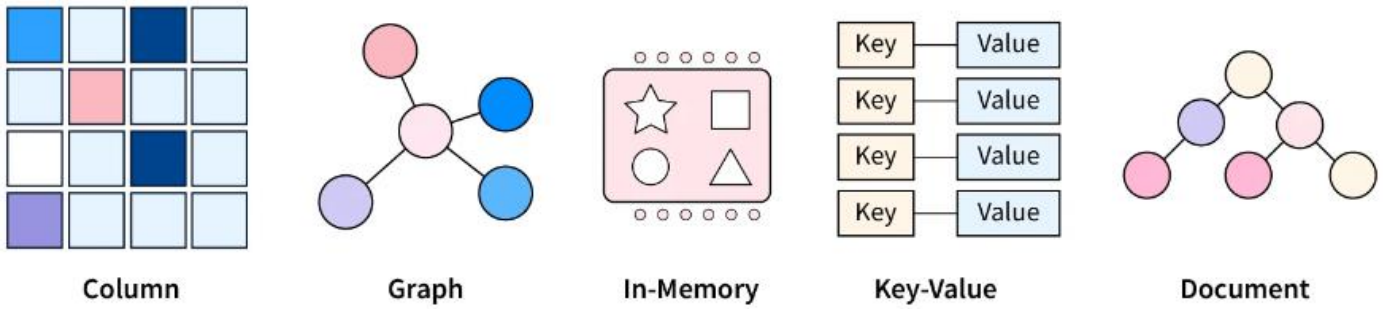


**Source :**

Reference

# Different Types of NoSQL Databases

- NoSQL databases use a different approach. Based on a approach there are five main types of NoSQL databases:
- Column  databases - Amazon SimpleDB, HBase, and Apache Cassandra are examples column stores.
- Document databases - MongoDB, Couchbase, and CouchDB are examples of document databases.
- Key-Value databases - Examples of key-value stores include Redis, Riak, and Amazon DynamoDB
- Graph database - Neo4j, OrientDB, and ArangoDB etc.
- In-Memory databases- Redis, Memcached, and Apache Ignite are few examples.



**Source :**

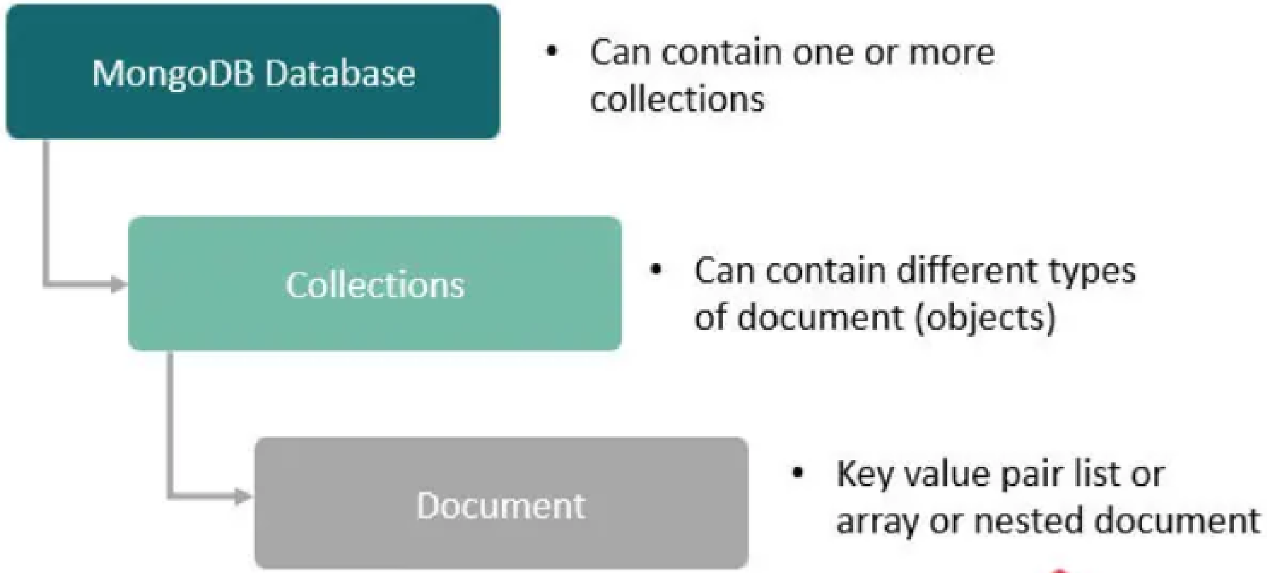
Reference

# Introduction to MongoDB

- MongoDB is a NoSQL, open-source document database that is created using C++.
- It is used to  store high-volume data.
- It stores data in a type of JSON format called BSON.
- MongoDB Atlas is a cloud database solution for contemporary applications that is available globally.
- It is a highly scalable, flexible, and distributed NoSQL database.

# Core MongoDB Concept

- MongoDB stores data records as documents (specifically BSON documents) which are gathered together in collections. A database stores one or more collections of documents.
- Field is a key value pair in a document. A document can have zero or more fields, which are like columns in relational databases.



MongoDB Database

- Can contain one or more collections

Collections

- Can contain different types of document (objects)

Document

- Key value pair list or array or nested document

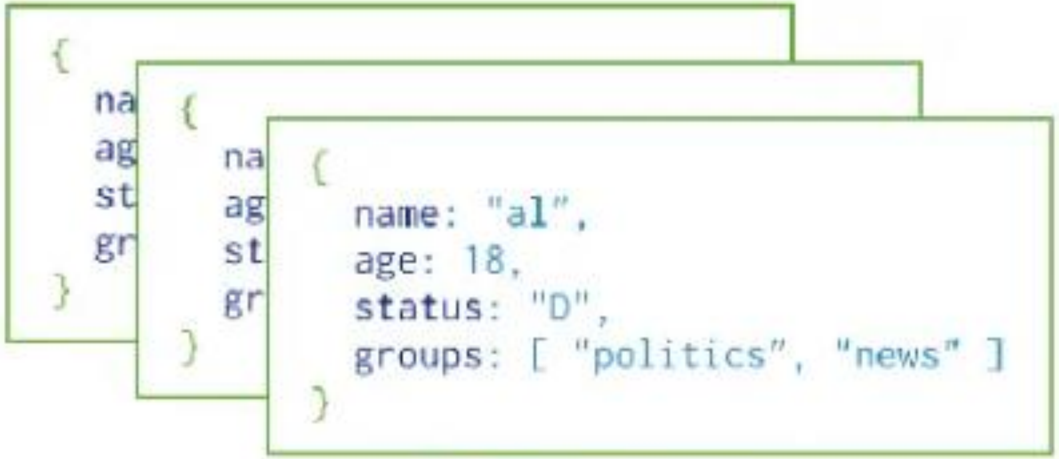
**Source :**

[Reference](#)

# Core MongoDB Concept

## What is Collection ?

- A collection in MongoDB is a group of MongoDB documents.
- Collections are just like tables in relational databases, they also store data, but in the form of documents. A single database is allowed to store multiple collections.
- An example of a collection is shown below:



```
{
   name: "al",
   age: 18,
   status: "D",
   groups: [ "politics", "news" ]
}
```

Collection

## Core MongoDB Concept

### What is Document ?

- In MongoDB, all data within databases and collections are stored in documents.
- MongoDB documents are made up of key-value or field-value pairs which are just like JSON. But these documents are stored in BSON (Binary JSON) format in MongoDB.
- Here is an example of a document that contains data about an employee.

```
{
    "employee_name":  “Amar",
     "employee_skills":  "UI Design",
     "employee_salary":  40000,
      "employee_status":  true,
}
```

# MongoDB Datatypes

In MongoDB, the documents are stores in BSON. BSON data format supports various data-types. The following are some of the most often used data types in MongoDB.

- String
- Integer
- Double
- Boolean
- Array
- Object
- Timestamp
- Date
- Binary
- Null

## MongoDB Datatypes

- **String -** String type is used to simply represent text. The strings in BSON are UTF-8 which allow us to represent most international characters in BSON strings without any problems.
- **Integer -** The integer data type is used to store numeric values. It can store 32-bit or 64-bit integers which depends on the server.
- **Double -** The double datatype is used to store numeric values with 8 bytes floating-point.
- **Boolean -** The boolean datatype is used to store boolean (true or false) values.

```
{
    "employee_name":  “Amar ",
     "employee_salary":  40000,
     “employee_score”:  85.44,
      "employee_status":  true,
}
```

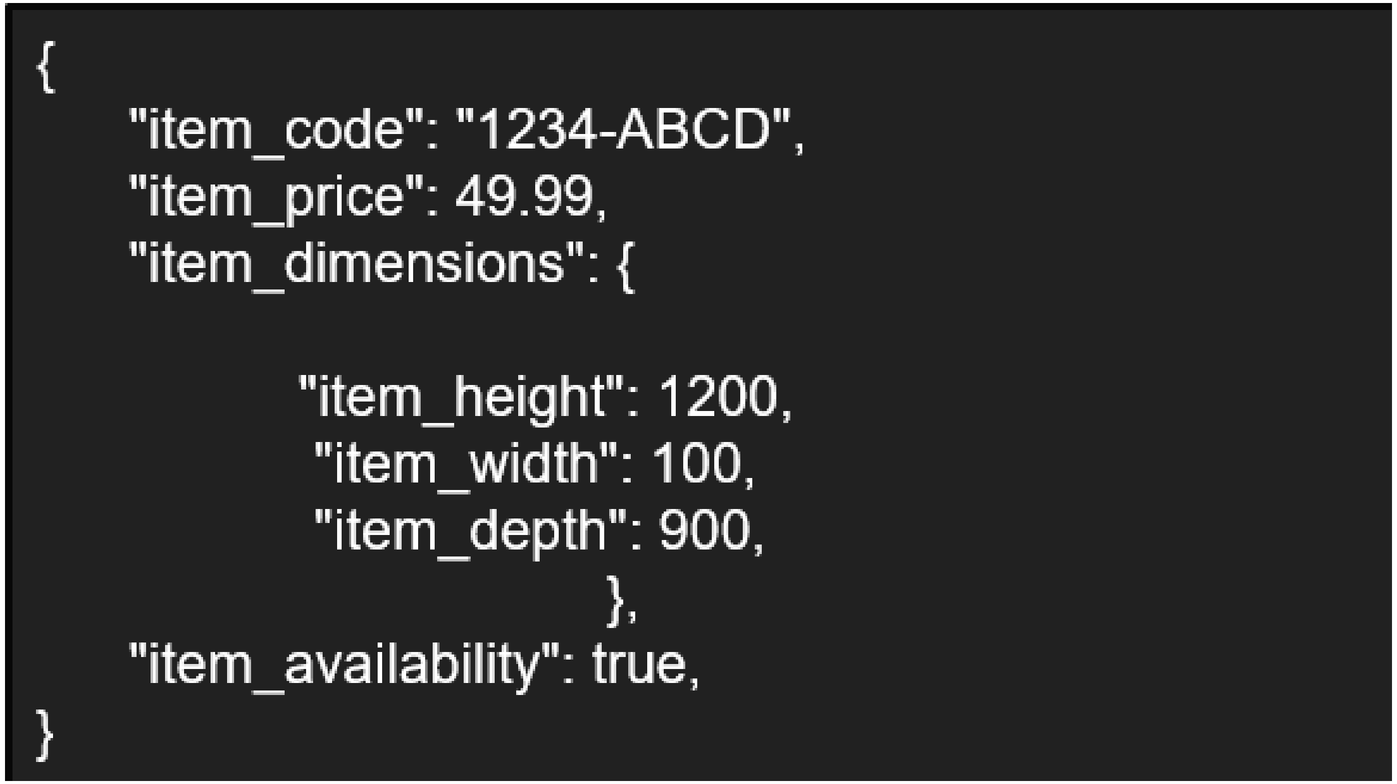
## MongoDB Datatypes

- **Arrays -** Arrays are used to store multiple values of the same type under a single key. For example employee_skills.
- **Date -** The date datatype is used to store date and time in Unix-time format. Unix timestamps can be easily converted to and from the JavaScript Date object.
- **Timestamp -** The timestamp type is a special type for internal MongoDB use and is not associated with the regular Date type.

```
{
    "employee_name":  “Amar ",
    "employee_skills":  ["UI Design", "Graphic Design", "2D Animation"],
     “employee_dob": ISODate("2006-02-10T10:50:42.389Z"),
     "item_created": Timestamp(1412180887, 1),
```

## MongoDB Datatypes

- **Object -** This data type is used to store embedded documents within a document. An embedded document is a series of nested documents in key: value pair format.
- In the given example, the item_dimensions field is an embedded document as it contains its own set of key-value pairs.



```
{
      "item_code": "1234-ABCD",
      "item_price": 49.99,
      "item_dimensions": {

                 "item_height": 1200,
                  "item_width": 100,
                  "item_depth": 900,
                                   },
      "item_availability": true,
}
```

Lab - 1

# MongoDB Installation in Windows

# Basic Commands in MongoDB

## Creating Database

- At the top of the hierarchy, in the MongoDB server, is database entity. There can be multiple databases on a single server.
- To create a database, MongoDB provides the command:

```
use DATABASE_NAME
```

This command:

- is used when a user wants to switch from one database to another.
- will create a new database if it doesn’t exist; otherwise, it will select the existing database.

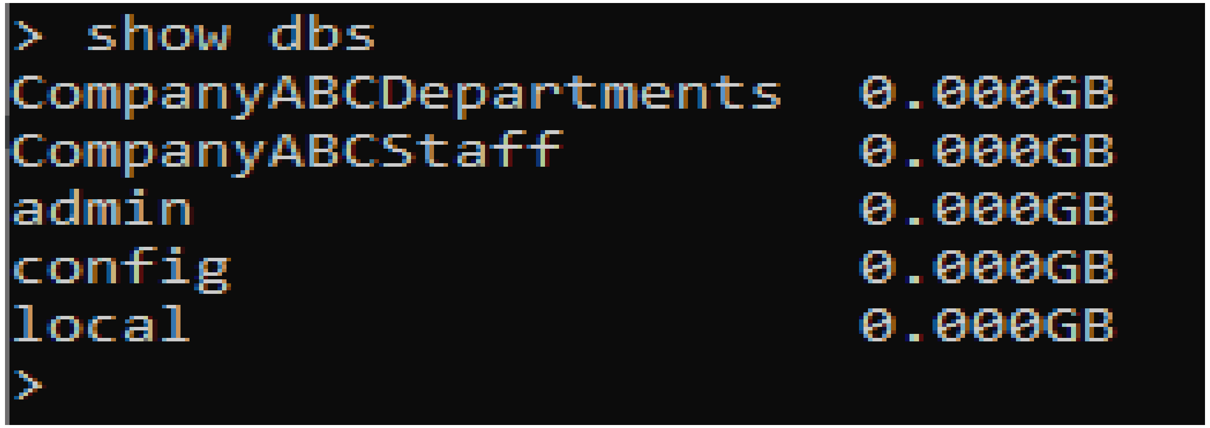
# Basic Command in MongoDB

## Listing Database

- To enlist all databases in MongoDB Server, MongoDB provides the command:

```
show dbs
```

Output:



```
> show dbs
CompanyABCDepartments  0.000GB
CompanyABCStaff        0.000GB
admin                  0.000GB
config                 0.000GB
local                  0.000GB
>
```

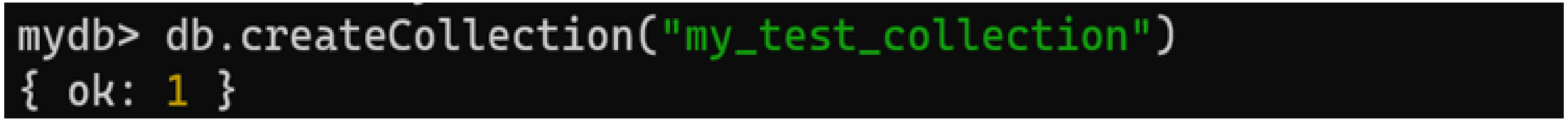
## Basic Command in MongoDB

### Creating Collections inside Database

- MongoDB provides the db.createCollection() method to create collection inside database.
- Syntax to create a collection:

```
db.createCollection(COLLECTION_NAME, OPTIONS)
```

- The first parameter is the name of the collection which is a string and the other is an options object which is used to configure the collection.



```
mydb> db.createCollection("my_test_collection")
{ ok: 1 }
```

# Basic Command in MongoDB

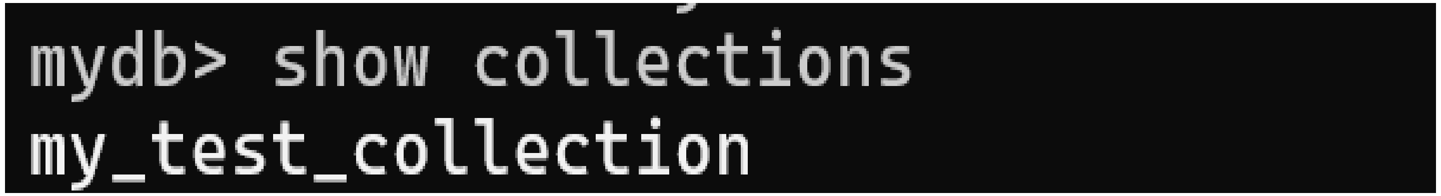
## Showing collections inside MongoDB

MongoDB provides show collections method to enlist all the collections inside database.

- Syntax to show collection:

```
show collections
```

- For example:



```
mydb> show collections
my_test_collection
```

- You can see my_test_collection in mydb database.

## MongoDB CRUD Operations

CRUD is an acronym for Create, Read, Update, and Delete. CRUD operations in MongoDB are used to manipulate data in databases.

- To add new documents to a collection, use the Create operation.
- Data from a collection is retrieved using the Read operation.
- The Update operation is used to edit existing documents in a collection.
- A collection of documents can be deleted using the Delete procedure.

| | |
|---|---|
| **C** | **Insert()** |
| R | find() |
| U | update() |
| D | remove() |

# MongoDB CRUD Operations

## Create

Create (or insert) operations add new documents to a collection. There are two ways to add new documents to a collection:

- db.collection.insertOne()
- db.collection.insertMany()

insertOne() operation allows us to create individual documents in a collection, while the insertMany() operation is used to create multiple documents in a single operation

# MongoDB CRUD Operations

## Read

- Read operations retrieve documents from a collection. Here is the method in Mongodb to retrieve information:

```
db.collection.find()
```

- find() operation will return everything from a collection if you call it without any parameters. On the other hand, we can specify any filter or criteria to retrieve information from a collection using:

```
db.collection.find(query)
```

# MongoDB CRUD Operations

## Update

- In MongoDB, the "update" operation is used to modify existing documents in a collection.
- The updateOne() method is used to update a single document that matches a specified filter.

```
db.collectionName.updateOne(filter,
update)
```

- The updateMany() method is used to update multiple documents that match a specified filter.

```
db.collection.updateMany(filter, update)
```

# MongoDB CRUD Operations

## Delete

- In MongoDB, the "delete" operation is used to remove documents from a collection.
- The deleteOne() method is used to remove a single document that matches a specified filter.

```
db.collectionName.deleteOne(filter,
options)
```

- The deleteMany() method is used to remove multiple documents that match a specified filter.

```
db.collectionName.deleteMany(filter,
options)
```

# Comparison operators in MongoDB

- Some of the comparison operator are given below:

| Name | Description |
|---|---|
| **$eq** | Matches values that are equal to a specified value. |
| **$gt** | Matches values that are greater than a specified value. |
| **$gte** | Matches values that are greater than or equal to a specified value. |
| **$in** | Matches any of the values specified in an array. |
| **$lt** | Matches values that are less than a specified value. |
| **$lte** | Matches values that are less than or equal to a specified value. |
| **$ne** | Matches all values that are not equal to a specified value. |
| **$nin** | Matches none of the values specified in an array. |

## Logical operators in MongoDB

- Some of the Logical operators are given below:

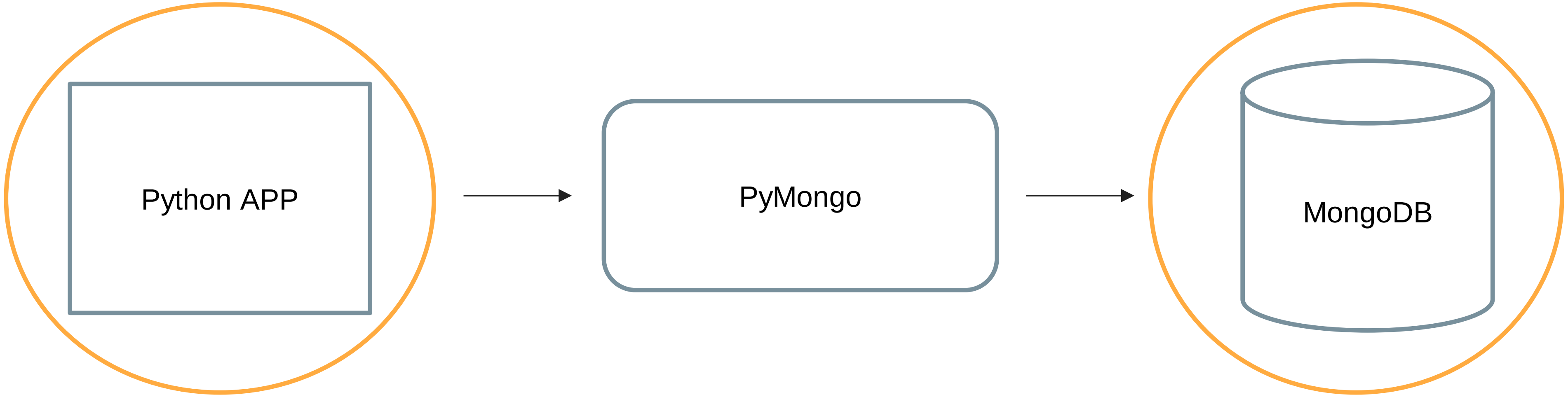
| Name | Description |
|---|---|
| **$and** | We use this operator to join query clauses with a logical AND and return the documents matching the given conditions of both clauses. |
| **$or** | We use the $or operator to join query clauses with a logical OR and return the documents matching the given conditions of either clause. |
| **$nor** | We use the $nor operator to join query clauses with a logical NOR and return the documents which fail to match both the clauses. |
| **$not** | We use the $not operator to invert the effect of the query expressions and return documents that do not match the query expression. |

Lab - 2

# CRUD Operations in MongoDB

## PyMongo

- PyMongo is a Python library containing tools for working with MongoDB, and is the recommended way to work with MongoDB from Python.
- PyMongo create a connection to the MongoDB using Python.
- Perform all operation like create, read, update and delete (CRUD).



# Installing PyMongo

- PyMongo can be installed using pip:

```
pip install pymongo
```

**Prerequisites**

- Ensure that you have installed the latest version of Python on your machine.
- MongoDB installed on your local machine.

- From the command-line tool, install the PyMongo by typing the command given below.

```
python -m pip install pymongo
```

Lab - 3

# CRUD Operations in MongoDB Using Python

# Conclusion

- Implementing data analytics into the business model means companies can help reduce costs by identifying more efficient ways of doing business. A company can also use data analytics to make better business decisions.
- NoSQL database are non-tabular databases and stores data differently than relational database.
- MongoDB is a document database. It stores data in a type of  json format called BSON.
- MongoDB Query API used to perform CRUD operation.
- PyMongo is the official MongoDB driver for synchronous Python applications. You have  learn how to connect and use MongoDB from your Python application and performed various crud operations.

**Source :**

Reference

# References

- https://bootcamp.berkeley.edu/resources/coding/learn-data-analytics/introduction-to-data-analytics/
- https://data-flair.training/blogs/power-bi-tutorial/
- https://www.w3schools.com/python/python_mysql_getstarted.asp
- https://www.tutorialspoint.com/python/python_database_access.htm
- https://intellipaat.com/blog/tutorial/data-analytics-tutorial/data-analytics-lifecycle/#no3
- https://online.hbs.edu/blog/post/types-of-data-analysis
- https://www.tutorialspoint.com/power_bi/power_bi_introduction.htm

Let’s Start

## Quiz

1. **To get insights from the data, many analysts and data scientists rely on ____________________.**

a) Data warehouse

b) Data visualization

c) Data mining

d) All of the above

**Answer: B**

Data visualization

## Quiz

**2. Data Analysis is a process of____________________**

a) Data Cleaning

b) Transforming of data

c) Inspecting data

d) All of the above

**Answer: D**

All of the above

## Quiz

**3. Which is not a type analytics**

a) Predictive

b) Diagnostic

c) Prescriptive

d) Perspective

**Answer: D**

Perspective

## Quiz

**4. Data Analytics uses ___ to get insights from data.**

a) Statistical figures

b) Numerical aspects

c) Statistical methods

d) None of the mentioned above

**Answer: C**

Statistical methods

## Quiz

**5.  What type of database is MongoDB?.**

a) Relational

b) Document Oriented

c) Key value

d) Graph

**Answer: B**

Document Oriented

# Thank You