DATA ANALYSIS

HANDBOOK

PROSPER EZEKIEL

UPTRAX TECHNOLOGIES LIMITED

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Tel** +234 912 201 9278 | NO. 43 ORO-EKPO STREET  ADA GEORGE, PORT HARCOURT | WWW.UPTRAXTECH.NG |
|  |  |  |

Contents

[**ABOUT US 1**](#_Toc321140622)

**TO OUR STUDENTS**  [**1**](#_Toc321140623)

**OVERVIEW OF THIS HANDBOOK \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

[**DEFINITION AND IMPORTANCE OF DATA ANALYSIS\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**](#_Toc321140624)

**DATA FUNDAMENTALS\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

**DATA ANALYSIS TECHNIQUES­­­­­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

**DATA VISUALIZATION\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

**DATA TOOLS AND TECHNOLOGIES\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

* [**STRUCTURED QUERY LANGUAGE (SQL) 1**](#_Toc321140625)
* [**MICROSOFT EXCEL 1**](#_Toc321140626)
* **POWERBI** [**1**](#_Toc321140627)
* [**TABLEAU 1**](#_Toc321140628)
* **PYTHON \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

**Q&A ABOUT DATA ANALYSIS\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1**

**GLOSSARY AND REFERENCES**

**APPENDICES**

# ABOUT US

“At UPTRAXTECH, building your career in data analytics from the beginner level to advanced level is our priority”

“THANKS FOR CHOOSING UPTRAX”

## UPTRAX TECHNOLOGIES LTD

UPTRAX TECHNOLOGIES LIMITED is an IT-based organization established in Nigeria (RC 1687187), that brings you a comprehensive range of tech courses designed to equip you with the knowledge and skills demanded by today’s digital world.

## OUR GOAL

At UPTRAX TECHNOLOGIES LIMITED, our goal is to create a community where tech enthusiasts learn, grow, and become pros in the tech space. We tend to become the most renowned tech academy in Africa through innovation, creativity, and productivity.

## COURSES WE OFFER

* 21st Century Computer Appreciation Training
* Programming (Web/Mobile)
* UI/UX & Graphic Design
* Cyber Security
* Digital Marketing
* Data Analytics and more.

## WHAT YOU GAIN SIGNING UP WITH US

* 100% Practical ● Internship & Job Opportunities
* Free Internet Access ● Affordable & Flexible Payment
* Conducive Computer Classrooms ● Mentorship & lots more

# TO OUR DATA ANALYSIS STUDENTS

WELCOME ON BOARD!! WELCOME TO UPTRAXTECH ACADEMY!!!

Congratulations on taking the first step towards unlocking the power of data analysis! We are thrilled to have you onboard.

As a data analysis student with UPTRAXTECH, you are about to embark on an exciting journey that will equip you with the skills and knowledge to extract insights from data, drive informed decisions, and propel your career forward.

Our comprehensive data analysis program is designed to provide you with:

* Hands-on experience with cutting-edge tools and technologies
* Expert guidance and experienced instructors
* Real-world case studies and projects to apply your skills
* Opportunities to network with like-minded professionals

Throughout your journey, we will support you with:

* Interactive classes and workshops
* Personalized mentorship and feedback
* Access to our community of data enthusiasts
* Regular updates on industry trends and best practices

We are committed to helping you achieve your goals and succeed in the field of data analysis. If you have any questions or need assistance, please don’t hesitate to reach out.

**Let’s unlock the power of data together!**

**Best Regards,**

**Mr. Valour E. Kooh**

**UptraxTech Limited**

# OVERVIEW OF THIS HANDBOOK

This handbook provides a comprehensive guide to data analysis, covering the fundamental concepts, techniques, and tools necessary to extract insights from data. It is designed for students, professionals, and anyone looking to develop their data analysis skills. The handbook is divided into sections that cover:

- Data fundamentals

- Data visualization

- Data analysis techniques

- Data tools and technologies

- Real-world applications

- Best practices and ethics

- Uptraxtech students projects

**OBJECTIVES OF THIS HANDBOOK**

**The objectives of this handbook are to:**

- Equip you with a solid understanding of data analysis concepts and techniques

- Provide practical guidance on using data analysis tools and technologies

- Develop your skills in extracting insights and telling stories with data

- Enable you to apply data analysis in real-world scenarios

- Foster a community of data analysts who can collaborate and learn from each other.

# DATA ANALYSIS; WHAT IT IS.

Data analysis is a comprehensive method of inspecting, cleansing, transforming, and modeling data to discover useful information, draw conclusions, and support decision-making. It is a multifaceted process involving various techniques and methodologies to interpret data from various sources in different formats, both structured and unstructured.

Data analysis is not just a mere process; it's a tool that empowers organizations to make informed decisions, predict trends, and improve operational efficiency. It's the backbone of strategic planning in businesses, governments, and other organizations.

Consider the example of a leading e-commerce company. Through data analysis, they can understand their customers' buying behavior, preferences, and patterns. They can then use this information to personalize customer experiences, forecast sales, and optimize marketing strategies, ultimately driving business growth and customer satisfaction.

## The Importance of Data Analysis in 2025

In the era of digital transformation, data analysis has become more critical than ever. The explosion of data generated by digital technologies has led to the advent of what we now call 'big data.' This vast amount of data, if analyzed correctly, can provide invaluable insights that can revolutionize businesses. [Statista](https://www.statista.com/statistics/590054/worldwide-business-analytics-software-vendor-market/) estimates that 'the market size for business intelligence and analytics software applications is forecast to increase worldwide over the next few years from 15.3 billion U.S. dollars in 2021 to more than 18 billion in 2026'. This huge growth shows just how important data and analytics have become.

Data analysis is the key to unlocking the potential of big data. It helps organizations to make sense of this data, turning it into actionable insights. These insights can be used to improve products and services, enhance experiences, streamline operations, and increase profitability.

A good example is the [healthcare industry](https://www.datacamp.com/business/healthcare-industry). Through data analysis, healthcare providers can predict disease outbreaks, improve patient care, and make informed decisions about treatment strategies. Similarly, in the finance sector, data analysis can help in risk assessment, fraud detection, and investment decision-making. We've also seen the impact of [AI in healthcare](https://www.datacamp.com/blog/ai-in-healthcare), demonstrating the rapidly changing environment and the need for ongoing analysis.

## The Data Analysis Process: A Step-by-Step Guide

The process of data analysis is a systematic approach that involves several stages, each crucial to ensuring the accuracy and usefulness of the results. Here, we'll walk you through each step, from defining objectives to data storytelling.



### Step 1: Defining objectives and questions

The first step in the data analysis process is to define the objectives and formulate clear, specific questions that your analysis aims to answer. This step is crucial as it sets the direction for the entire process. It involves understanding the problem or situation at hand, identifying the data needed to address it, and defining the metrics or indicators to measure the outcomes.

### Step 2: Data collection

Once the objectives and questions are defined, the next step is to collect the relevant data. This can be done through various methods such as surveys, interviews, observations, or extracting from existing databases. The data collected can be quantitative (numerical) or qualitative (non-numerical), depending on the nature of the problem and the questions being asked.

### Step 3: Data cleaning

Data cleaning, also known as data cleansing, is a critical step in the data analysis process. It involves checking the data for errors and inconsistencies, and correcting or removing them. This step ensures the quality and reliability of the data, which is crucial for obtaining accurate and meaningful results from the analysis.

### Step 4: Data analysis

Once the data is cleaned, it's time for the actual analysis. This involves applying statistical or mathematical techniques to the data to discover patterns, relationships, or trends. There are various tools and software available for this purpose, such as Python, R, Excel, and specialized software like SPSS and SAS.

### Step 5: Data interpretation and visualization

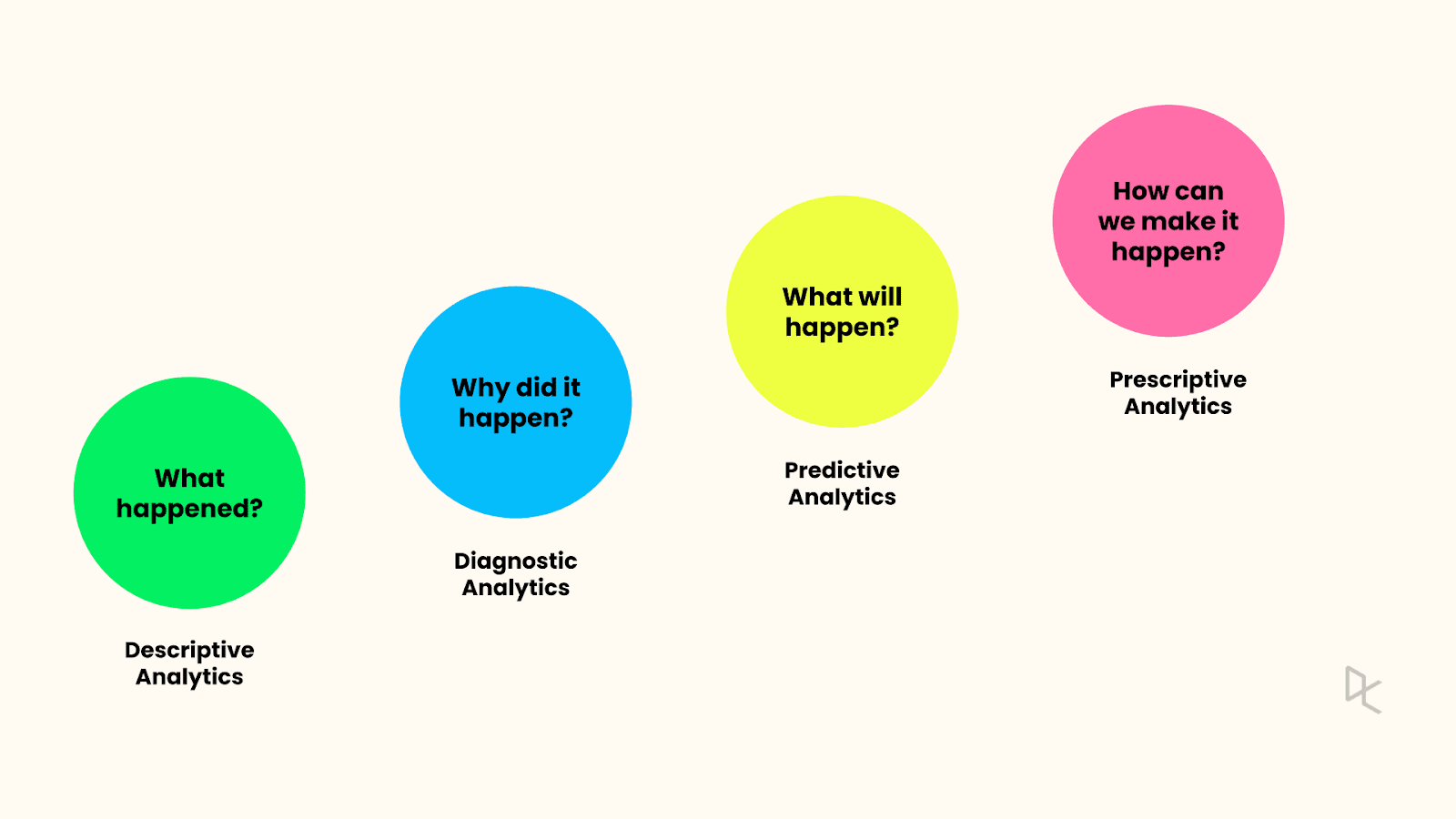
After the data is analyzed, the next step is to interpret the results and visualize them in a way that is easy to understand. This could involve creating charts, graphs, or other visual representations of the data. Data visualization helps to make complex data more understandable and provides a clear picture of the findings.

### Step 6: Data storytelling

The final step in the data analysis process is data storytelling. This involves presenting the findings of the analysis in a narrative form that is engaging and easy to understand. Data storytelling is crucial for communicating the results to non-technical audiences and for making data-driven decisions.

## The Types of Data Analysis

Data analysis can be categorized into four main types, each serving a unique purpose and providing different insights. These are descriptive, diagnostic, predictive, and prescriptive analyses.



### Descriptive analysis

[Descriptive analysis](https://www.datacamp.com/blog/data-demystified-an-overview-of-descriptive-statistics), as the name suggests, describes or summarizes raw data and makes it interpretable. It involves analyzing historical data to understand what has happened in the past.

This type of analysis is used to identify patterns and trends over time.

For example, a business might use descriptive analysis to understand the average monthly sales for the past year.

### Diagnostic analysis

[Diagnostic analysis](https://campus.datacamp.com/courses/introduction-to-data-literacy/working-with-and-analyzing-data?ex=5#:~:text=Diagnostic%20analytics%20can%20be%20used,based%20on%20the%20identified%20causes.) goes a step further than descriptive analysis by determining why something happened. It involves more detailed data exploration and comparing different data sets to understand the cause of a particular outcome.

For instance, if a company's sales dropped in a particular month, diagnostic analysis could be used to find out why.

### Predictive analysis

[Predictive analysis](https://www.datacamp.com/blog/predictive-analytics-guide) uses statistical models and forecasting techniques to understand the future. It involves using data from the past to predict what could happen in the future. This type of analysis is often used in risk assessment, marketing, and sales forecasting.

For example, a company might use predictive analysis to forecast the next quarter's sales based on historical data.

### Prescriptive analysis

[Prescriptive analysis](https://campus.datacamp.com/courses/introduction-to-data-literacy/working-with-and-analyzing-data?ex=13) is the most advanced type of data analysis. It not only predicts future outcomes but also suggests actions to benefit from these predictions. It uses sophisticated tools and technologies like machine learning and artificial intelligence to recommend decisions.

For example, a prescriptive analysis might suggest the best marketing strategies to increase future sales.

## Data Analysis Techniques

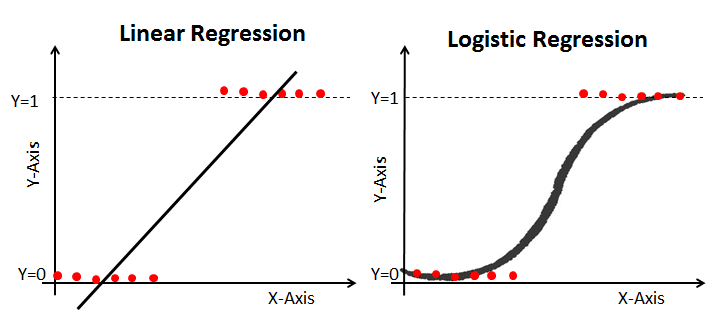
There are numerous techniques used in data analysis, each with its unique purpose and application. Here, we will discuss some of the most commonly used techniques, including exploratory analysis, regression analysis, Monte Carlo simulation, factor analysis, cohort analysis, cluster analysis, time series analysis, and sentiment analysis.

### Exploratory analysis

Exploratory analysis is used to understand the main characteristics of a data set. It is often used at the beginning of a data analysis process to summarize the main aspects of the data, check for missing data, and test assumptions. This technique involves visual methods such as scatter plots, histograms, and box plots.

### Regression analysis

Regression analysis is a statistical method used to understand the relationship between a dependent variable and one or more independent variables. It is commonly used for forecasting, time series modeling, and finding the causal effect relationships between variables.



LINEAR AND LOGISTIC REGRESSION

### Factor analysis

Factor analysis is a technique used to reduce a large number of variables into fewer factors. The factors are constructed in such a way that they capture the maximum possible information from the original variables. This technique is often used in market research, customer segmentation, and image recognition.

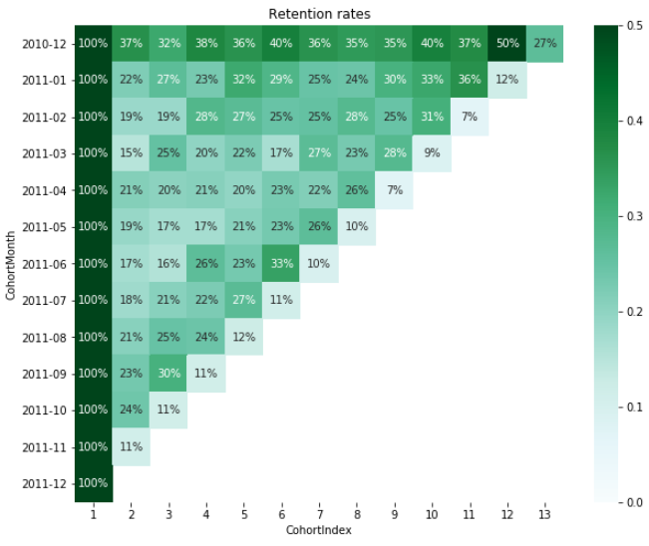
### Cluster analysis

Cluster analysis is a technique used to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups. It is often used in market segmentation, image segmentation, and recommendation systems.

You can explore a range of clustering techniques, including hierarchical clustering and k-means clustering.

### Cohort analysis

Cohort analysis is a subset of behavioral analytics that takes data from a given dataset and groups it into related groups for analysis. These related groups, or cohorts, usually share common characteristics within a defined time span. This technique is often used in marketing, user engagement, and customer lifecycle analysis.



### Time series analysis

Time series analysis is a statistical technique that deals with time series data, or trend analysis. It is used to analyze the sequence of data points to extract meaningful statistics and other characteristics of the data. This technique is often used in sales forecasting, economic forecasting, and weather forecasting.

Our [Time Series with Python](https://www.datacamp.com/tracks/time-series-with-python) skill track takes you through how to manipulate and analyze time series data, working with a variety of Python libraries.

### Sentiment analysis

Sentiment analysis, also known as opinion mining, uses natural language processing, text analysis, and computational linguistics to identify and extract subjective information from source materials. It is often used in social media monitoring, brand monitoring, and understanding customer feedback.

## Data Analysis Tools

In the realm of data analysis, various tools are available that cater to different needs, complexities, and levels of expertise. These tools range from programming languages like Python and R to visualization software like Power BI and Tableau. Let's delve into some of these tools.

### Python

Python is a high-level, general-purpose programming language that has become a favorite among data analysts and data scientists. Its simplicity and readability, coupled with a wide range of libraries like [pandas](https://www.datacamp.com/tutorial/pandas), [NumPy](https://www.datacamp.com/courses/introduction-to-numpy), and [Matplotlib](https://www.datacamp.com/courses/introduction-to-data-visualization-with-matplotlib), make it an excellent tool for data analysis and data visualization

### R

R is a programming language and free software environment specifically designed for statistical computing and graphics. It is widely used among statisticians and data miners for developing statistical software and data analysis. R provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, and more.

### SQL

SQL (Structured Query Language) is a standard language for managing and manipulating databases. It is used to retrieve and manipulate data stored in relational databases. SQL is essential for tasks that involve data management or manipulation within databases.

### Power BI

[Power BI is a business analytics tool](https://www.datacamp.com/blog/all-about-power-bi) developed by Microsoft. It provides interactive visualizations with self-service business intelligence capabilities. Power BI is used to transform raw data into meaningful insights through easy-to-understand dashboards and reports.

### Tableau

Tableau is a powerful data visualization tool used in the Business Intelligence industry. It allows you to create interactive and shareable dashboards, which depict trends, variations, and density of the data in the form of charts and graphs.

### Excel

Microsoft Excel is one of the most widely used tools for data analysis. It offers a range of features for data manipulation, statistical analysis, and visualization. Excel's simplicity and versatility make it a great tool for both simple and complex data analysis tasks.

## Understanding the Impact of Data Analysis

Data analysis, whether on a small or large scale, can have a profound impact on business performance. It can drive significant changes, leading to improved efficiency, increased profitability, and a deeper understanding of market trends and customer behavior.

### Informed decision-making

Data analysis allows businesses to make informed decisions based on facts, figures, and trends, rather than relying on guesswork or intuition. It provides a solid foundation for strategic planning and policy-making, ensuring that resources are allocated effectively and that efforts are directed towards areas that will yield the most benefit.

### Impact on small businesses

For small businesses, even simple data analysis can lead to significant improvements. For example, analyzing sales data can help identify which products are performing well and which are not. This information can then be used to adjust marketing strategies, pricing, and inventory management, leading to increased sales and profitability.

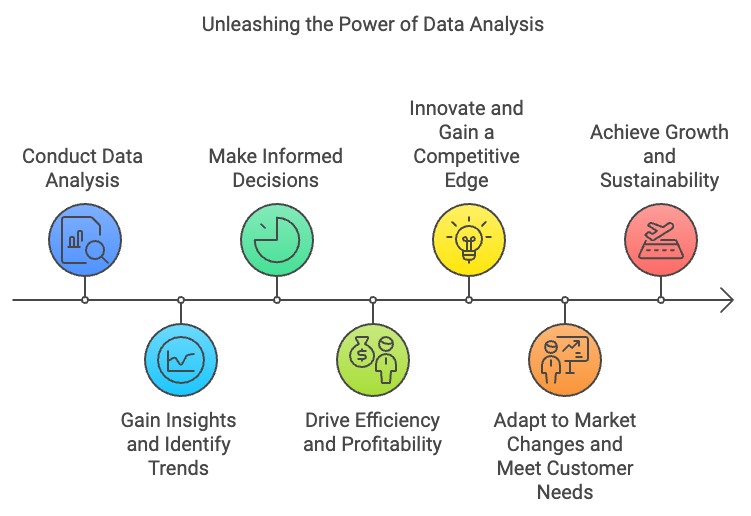
### Impact on large businesses

For larger businesses, the impact of data analysis can be even more profound. Big data analysis can uncover complex patterns and trends that would be impossible to detect otherwise. This can lead to breakthrough insights, driving innovation and giving the business a competitive edge.

For example, a large retailer might use data analysis to optimize its supply chain, reducing costs and improving efficiency. Or a tech company might use data analysis to understand user behavior, leading to improved product design and better user engagement.

### **The critical role of data analysis**

In today's data-driven world, the ability to analyze and interpret data is a critical skill. Businesses that can harness the power of data analysis are better positioned to adapt to changing market conditions, meet customer needs, and drive growth and profitability.



## Top Careers in Data Analysis in 2025

In the era of Big Data, careers in data analysis are flourishing. With the increasing demand for data-driven insights, these professions offer promising prospects. Here, we will discuss some of the top careers in data analysis in 2025.

### 1. Data scientist

Data scientists are the detectives of the data world, uncovering patterns, insights, and trends from vast amounts of information. They use a combination of programming, statistical skills, and machine learning to make sense of complex data sets. Data scientists not only analyze data but also use their insights to influence strategic decisions within their organization.

**Key skills**:

* Proficiency in programming languages like Python or R
* Strong knowledge of statistics and probability
* Familiarity with machine learning algorithms
* Data wrangling and data cleaning skills
* Ability to communicate complex data insights in a clear and understandable manner

**Essential tools**:

* Python, R
* SQL
* Jupyter Notebook
* Machine learning libraries like Scikit-learn, TensorFlow
* Data visualization libraries like Matplotlib, Seaborn

### 2. Business Intelligence Analyst

Business intelligence analysts are responsible for providing a clear picture of a business's performance by analyzing data related to market trends, business processes, and industry competition. They use tools and software to convert complex data into digestible reports and dashboards, helping decision-makers to understand the business's position and make informed decisions.

**Key skills**:

* Strong analytical skills
* Proficiency in SQL and other database technologies
* Understanding of data warehousing and ETL processes
* Ability to create clear visualizations and reports
* Business acumen

**Essential tools**:

* SQL
* Power BI, Tableau
* Excel
* Python, R

### 3. Data engineer

Data engineers are the builders and maintainers of the data pipeline. They design, construct, install, test, and maintain highly scalable data management systems. They also ensure that data is clean, reliable, and preprocessed for data scientists to perform analysis.

**Key skills**:

* Proficiency in SQL and NoSQL databases
* Knowledge of distributed systems and data architecture
* Familiarity with ETL tools and processes
* Programming skills, particularly in Python and Java
* Understanding of machine learning algorithms

**Essential tools**:

* SQL, NoSQL
* Hadoop, Spark
* Python, Java
* ETL tools

### 4. Business analyst

Business analysts are the bridge between IT and business stakeholders. They use data to assess processes, determine requirements, and deliver data-driven recommendations and reports to executives and stakeholders. They are involved in strategic planning, business model analysis, process design, and system analysis.

**Key skills**:

* Strong analytical skills
* Understanding of business processes and strategies
* Proficiency in SQL
* Ability to communicate effectively with both IT and business stakeholders
* Project management skills

**Essential tools**:

* SQL
* Excel
* Power BI, Tableau
* Python, R

|  |  |  |
| --- | --- | --- |
| **Career** | **Key Skills** | **Essential Tools** |
| **Data Scientist** | Proficiency in programming, strong statistical knowledge, familiarity with machine learning, data wrangling skills, and effective communication. | Python, R, SQL, Scikit-learn, TensorFlow, Matplotlib, Seaborn |
| **Business Intelligence Analyst** | Strong analytical skills, proficiency in SQL, understanding of data warehousing and ETL, ability to create visualizations and reports, and business acumen. | SQL, Power BI, Tableau, Excel, Python |
| **Data Engineer** | Proficiency in SQL and NoSQL, knowledge of distributed systems and data architecture, familiarity with ETL, programming skills, and understanding of machine learning. | SQL, NoSQL, Hadoop, Spark, Python, Java, ETL tools |
| **Business Analyst** | Strong analytical skills, understanding of business processes, proficiency in SQL, effective communication, and project management skills. | SQL, Excel, Power BI, Tableau, Python |

# DATA FUNDAMENTALS

Data fundamentals are the building blocks of data analysis. Understanding these concepts is crucial for working with data effectively and extracting meaningful insights.

**Data Types**

Data comes in different forms, including:

- Quantitative data: Numerical data that can be measured and analyzed using mathematical operations (e.g., sales figures, temperatures).

- Qualitative data: Non-numerical data that provides descriptive information (e.g., customer feedback, survey responses).

- Categorical data: Data that falls into categories or groups (e.g., gender, product categories).

**Data Sources**

Data can come from various sources, including:

- Internal data: Data generated within an organization (e.g., sales data, customer information)

- External data: Data obtained from outside sources (e.g., market research, social media).

- Primary data: Original data collected directly from sources (e.g., surveys, experiments).

- Secondary data: Existing data collected by others (e.g., publicly available datasets).

**Data Quality**

Data quality refers to the accuracy, completeness, and reliability of data. Ensuring high data quality is essential for accurate analysis and decision-making.

**Data Preprocessing**

Data preprocessing involves preparing data for analysis by:

- Cleaning: Identifying and correcting errors or inconsistencies.

- Transforming: Converting data into suitable formats for analysis.

- Reducing: Selecting relevant data to reduce complexity.

**Data Sampling**

Data sampling involves selecting a representative subset of data for analysis. This is useful when working with large datasets or to reduce analysis time.

**Data Summarization**

Data summarization involves condensing data into summary statistics and visualizations to understand key trends and patterns.

# DATA ANALYSIS TECHNIQUES

**Data analysis techniques are methods used to extract insights and meaning from data. These techniques help to identify patterns, trends, and correlations, and to draw conclusions and make predictions.**

**DESCRIPTIVE STATISTICS**

Descriptive statistics provide a summary of the basic features of the data, including:

- Measures of central tendency: Mean, median, mode

- Measures of variability: Range, variance, standard deviation

- Data distribution: Shape, skewness, kurtosis

**INFERENTIAL STATISTICS**

Inferential statistics involve using sample data to make conclusions about a larger population. Techniques include:

- Hypothesis testing: Testing statements about the population based on sample data

- Confidence intervals: Estimating population parameters with a range of values

- Regression analysis: Modeling the relationship between variables

**DATA MINING AND MACHINE LEARNING**

Data mining and machine learning involve using algorithms and statistical models to discover patterns and make predictions. Techniques include:

- Clustering: Grouping similar data points into clusters

- Decision trees: Creating a tree-based model to classify data

- Neural networks: Modeling complex relationships using artificial neural networks

**DATA REDUCTION AND DIMENSIONALITY REDUCTION**

Data reduction and dimensionality reduction involve reducing the size and complexity of the data while preserving the most important information. Techniques include:

- Principal component analysis (PCA): Reducing dimensionality while preserving variance

- Factor analysis: Identifying underlying factors that explain the data

- Data compression: Reducing data size while preserving key information

# DATA VISUALIZATION

Data visualization is the process of creating graphical representations of data to better understand and communicate insights. Effective data visualization helps to:

- Identify patterns and trends

- Spot outliers and anomalies

- Communicate complex data insights to non-technical audiences

- Facilitate data-driven decision-making

**TYPES OF DATA VISUALIZATION**

- Charts and Graphs: Used to display numerical data, such as:

- Line charts

- Bar charts

- Scatter plots

- Heatmaps and Tree-maps: Used to display categorical data, such as:

- Heatmaps for geographic data

- Tree-maps for hierarchical data

- Interactive Visualizations: Used to explore data in real-time, such as:

- Dashboards

- Interactive charts and graphs

**BEST PRACTICES FOR DATA VISUALIZATION**

- Keep it simple and clear: Avoid clutter and focus on key insights

- Choose the right visualization: Select the appropriate visualization type for the data

- Use color effectively: Use color to convey meaning and draw attention

- Label and annotate: Provide context and explanations for the visualization

**DATA VISUALIZATION TOOLS**

- Tableau: A popular data visualization platform for interactive dashboards

- Power BI: A business analytics service for data visualization and reporting

- Matplotlib and Seaborn: Python libraries for creating static and interactive visualizations

- D3.js: A JavaScript library for creating interactive, web-based visualizations

# KNOW YOUR COMPUTER SHORT-CUTS

**CTRL + A = All Select**

**CTRL + B = Bold**

**CTRL + C = Copy**

**CTRL + D = Duplicate**

**CTRL + E = Alignment Center Side**

**CTRL + F = Find**

**CTRL + G = Group & Go To**

**CTRL + H = Replace**

**CTRL + I = Italic**

**CTRL + J = Justify**

**CTRL + K = Hyperlink**

**CTRL** **+** **A** **TO** **Z**

**CTRL + L = Alignment Left Side**

**CTRL + M = New Slide**

**CTRL + N = New Page**

**CTRL + O = Open**

**CTRL + P = Print**

**CTRL + Q = Quit**

**CTRL + R = Alignment Right Side**

**CTRL + S = Save**

**CTRL + T = New Tab**

**CTRL + U = Underline**

**CTRL + V = Paste**

**CTRL + W = Close the Window**

**CTRL + X = Cut**

**CTRL + Y = Redo**

**CTRL + Z = Undo**

# 1STRUCTURED QUERY LANGUAGE (SQL)

**SQL is used to communicate with a database. SQL statements are used to perform tasks such as update data on a database, or retrieve data from a database. In this article, we will be looking at the following:**

* SQL DATABASE
* DATABASE MANAGEMENT SYSTEMS (DBMS)
* SQL; WHAT IT IS
* SQL COMMANDS, KEYWORDS AND FUNCTIONS
* SQL PRACTICES
* ASSIGNMENTS, TESTS AND PROJECTS.

**WHAT IS DATABASE?**

A database is an organized collection of data, stored and retrieved digitally from a remote or local computer system.

Databases can be vast and complex, and such databases are developed using fixed design and modelling approaches.

**WHAT IS DBMS?**

DBMS stands for Database Management System. DBMS is a system database software responsible for the creation, retrieval, updating, and management of the database. It ensures that our data is consistent, organized and easily accessible by serving as an interface between the database and its end-users or application software.

**WHAT IS RDBMS? How is it different from DBMS?**

RDBMS stands for Relational Database Management System. The key difference here compared to the DBMS, is that RDBMS stores data in the form of a collection of tables, and relations can be defined between the common fields of these tables. Most modern DBMS like MySQL, Microsoft SQL server, Oracle, IBM DB2, and Amazon Redshift are based on RDBMS.

**SQL; WHAT IT IS.**

SQL stands for Structured Query Language. It is the standard language used to maintain the relational database and perform many different data manipulation operations on the data. SQL was initially invented in 1970. It is a database language used for database creation, deletion, fetching and modifying rows, etc. Sometimes, it is pronounced as “SEQUEL”. We can also use it to handle organized data comprised of entities (variables) and relations between different entities of the data.

**WHAT IS THE DIFFERENCE BETWEEN SQL AND MYSQL**

|  |  |
| --- | --- |
| SQL | MYSQL |
| SQL is a standard language which stands for (Structured Query Language) based on the English Language. | **MySQL** is a database management system. |
| SQL is the core of the relational database which is used for accessing and managing databases. | **MySQL** is an RDBMS (Relational Database Management System) such as SQL server, Informix, etc. |

**WHAT ARE THE USAGES OF SQL?**

SQL is responsible for maintaining the relational data and the data structures present in the database.

Some of common usages are given below;

* To execute queries against a database.
* To retrieve data from a database.
* To insert records in a database
* To Update records from a database.
* To delete records from a database.
* To create new databases.
* To create new tables databases.
* To create views in a database.
* To perform complex operations on a database.

**WHAT ARE TABLES AND FIELDS?**

TABLES: A table is an organized collection of data stored in the form of rows and columns. Columns can be categorized as vertical and rows as horizontal.

FIELDS: The columns in a table are called FIELDS while the rows can be referred to as RECORDS.

**SQL COMMANDS,** **KEYWORDS AND FUNCTIONS.**

**SQL COMMANDS;**

SELECT – Extract data from database. The data returned is stored in a result table called result-set.

**SELECT SYNTAX:**

**SELECT column1, column2…….**

**FROM table\_name;**

**Note:** All SQL queries closes with a semi-colon.

From the above select syntax. The column1, column2… are field names of the table you want to select data from.

If you want to select all the fields available in the table, use the following syntax:

**SELECT \***

**FROM table\_name;**

Many queries and commands will be looked at in the Practice section of this chapter.

**KEYWORDS AND FUNCTIONS;**

**SELECT –** retrieves data from a database

**INSERT –** inserts new data into a database

**UPDATE –** updates existing data in a database

**DELETE –** deletes data from a database

**CREATE DATABASE –** creates a new table in a database

**ALTER TABLE –** modifies an existing table structure

**DROP TABLE –** deletes a table from a database

**TRUNCATE TABLE –** removes all records from a table

**CREATE INDEX –** creates an index on a table

**DROP INDEX –** deletes an index from a table

**JOIN –** combines rows from two or more tables based on a related column

**INNER JOIN –** returns rows when there is a match in both columns

**LEFT JOIN –** returns all rows from the left table, and the matched rows from the right table

**RIGHT JOIN –** returns all rows from the right table and the matched rows from the left table

**FULL JOIN –** returns rows when there is a match in one of the tables

**UNION –** combines the results of two or more SELECT statements

**UNION ALL –** combines the results of two or more SLECT statements, including duplicates

**GROUP BY –** groups rows that have the same values into summary rows.

**HAVING –** filters records based on a specified condition

**ORDER BY -** sorts the result set in ascending or descending order

**COUNT –** returns the number of rows that satisfies a condition

**SUM –** calculates the sum of a set of values

**AVG –** calculates the average (arithmetic mean) of a set of values

**MIN –** returns the smallest value in a set of values

**MAX -** returns the largest value in a set of values

**DISTINCT –** selects unique values from a column

**WHERE –** filters records based on a specified condition

**AND –** combines multiple conditions in a WHERE clause

**OR –** specifies multiple alternative conditions in a WHERE clause

**NOT –** negates a condition in a WHERE clause

**BETWEEN –** selects values within a specified range

**IN –** specifies multiple values for a column

**LIKE –** selects rows that match a specified pattern

**IS NULL –** checks for null values in a column

**IS NOT NULL –** checks for non-null values in a column

**EXISTS –** tests for the existence of any record in a subquery

**CASE -** performs conditional logic in SQL statements

**WHEN –** specifies condition in a CASE statements

**THEN -** specifies the result if a condition is true in a CASE statements

**ELSE -** specifies the result if no condition is true in a CASE statements

**END –** ends the CASE statement

**PRIMARY KEY –** uniquely identifies each record in a table

**FOREIGN KEY –** establishes a relationship between tables

**CONSTRAINT –** enforces rules for data in a table

**DEFAULT –** specifies a default value for a column

**NOT NULL –** ensures that a column cannot contain NULL values

**UNIQUE -** ensures that all values in a column are unique

**CHECK –** enforces a condition on the values in a column

**CASCADE ­–** automatically performs a specified action on related records

**SET NULL –** sets the value of foreign key columns to NULL value when a referenced record is deleted

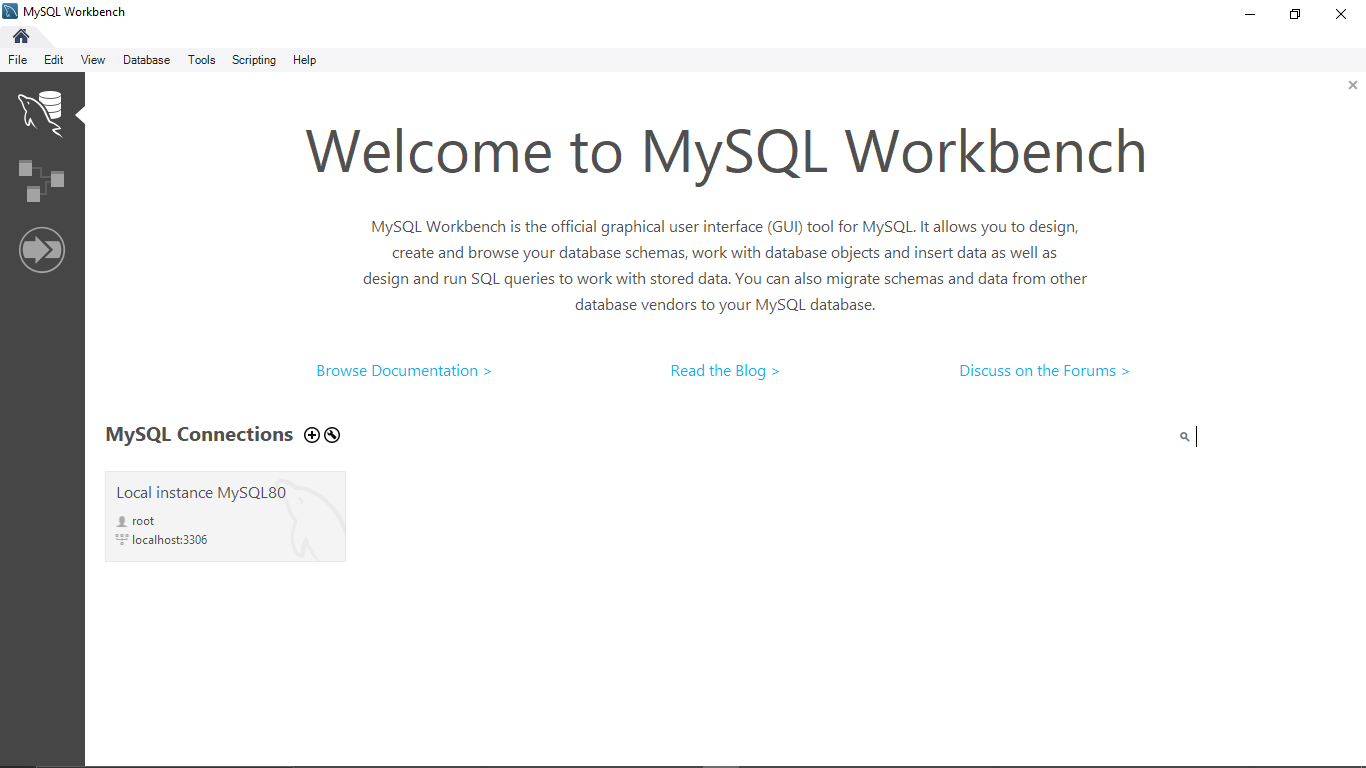
**SET DEFAULT –** sets the value of foreign key columns to their default value when a referenced record is deleted

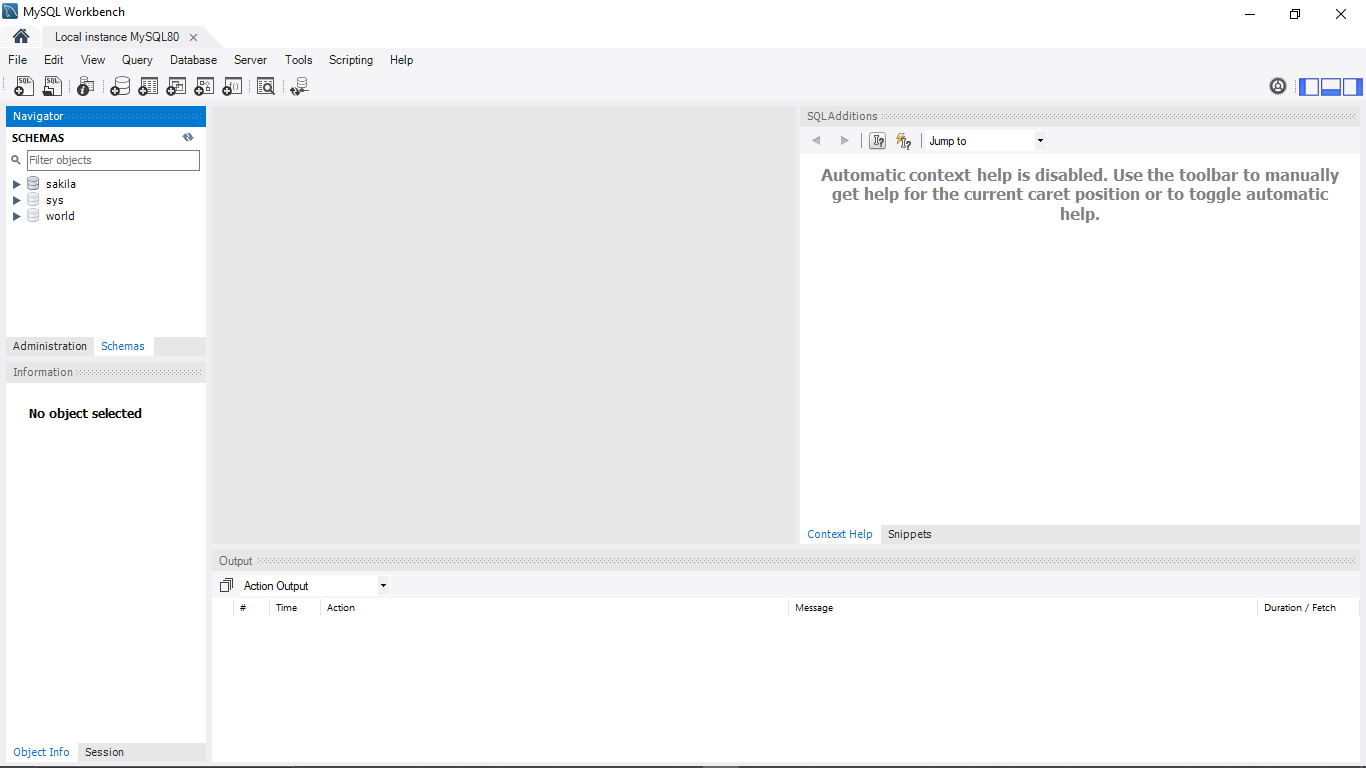
**NO ACTION –** specifies that no action should be taken on related records when a referenced record is deleted.

**SQL PRACTICES**

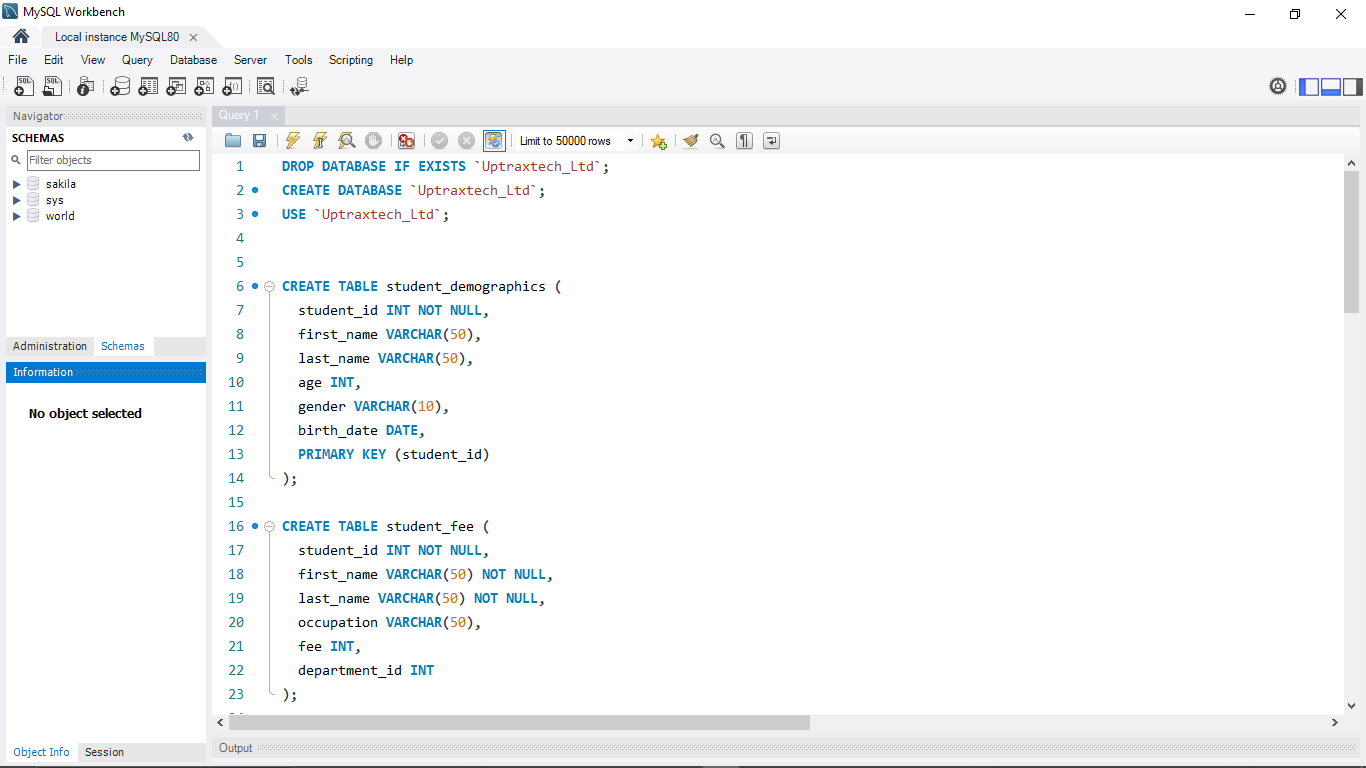
In this series, we will be using the MySQL Workbench.

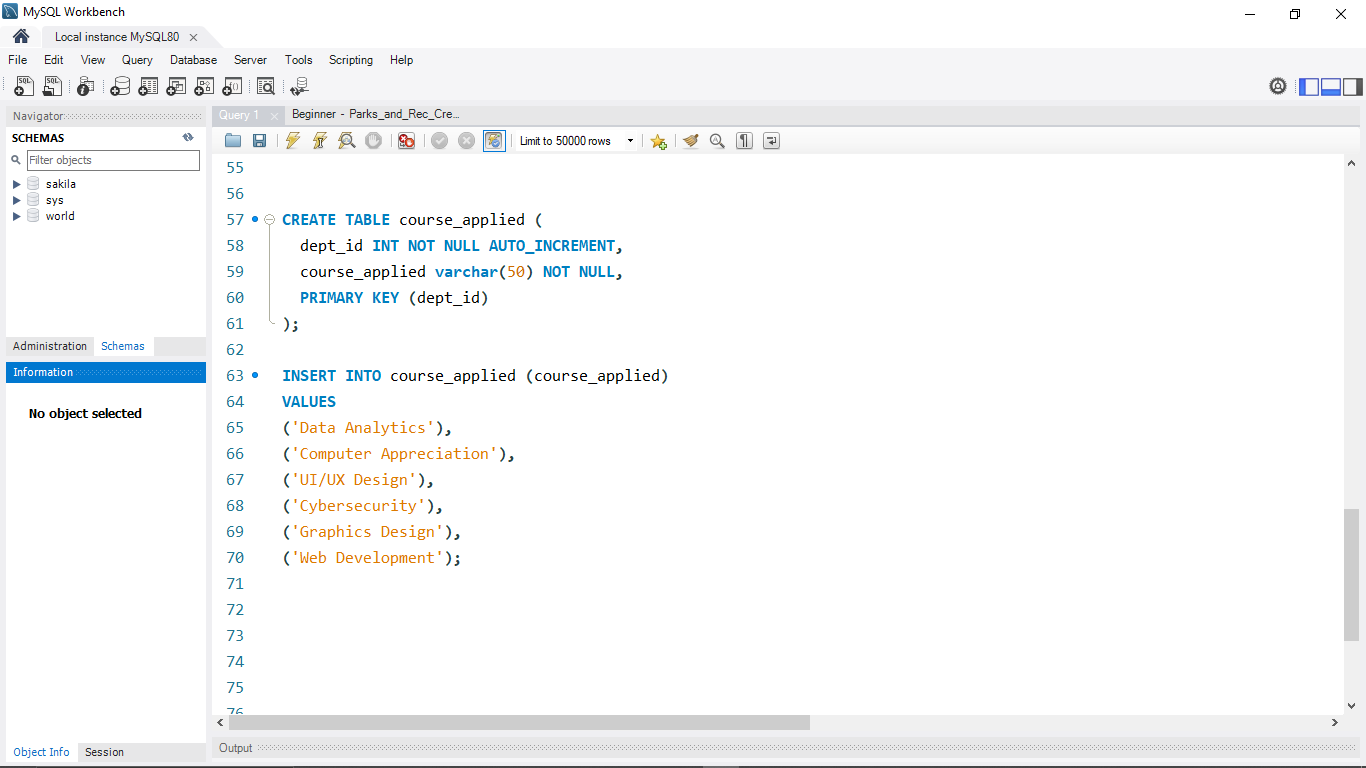
MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to design, create, and browse your database schemas with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

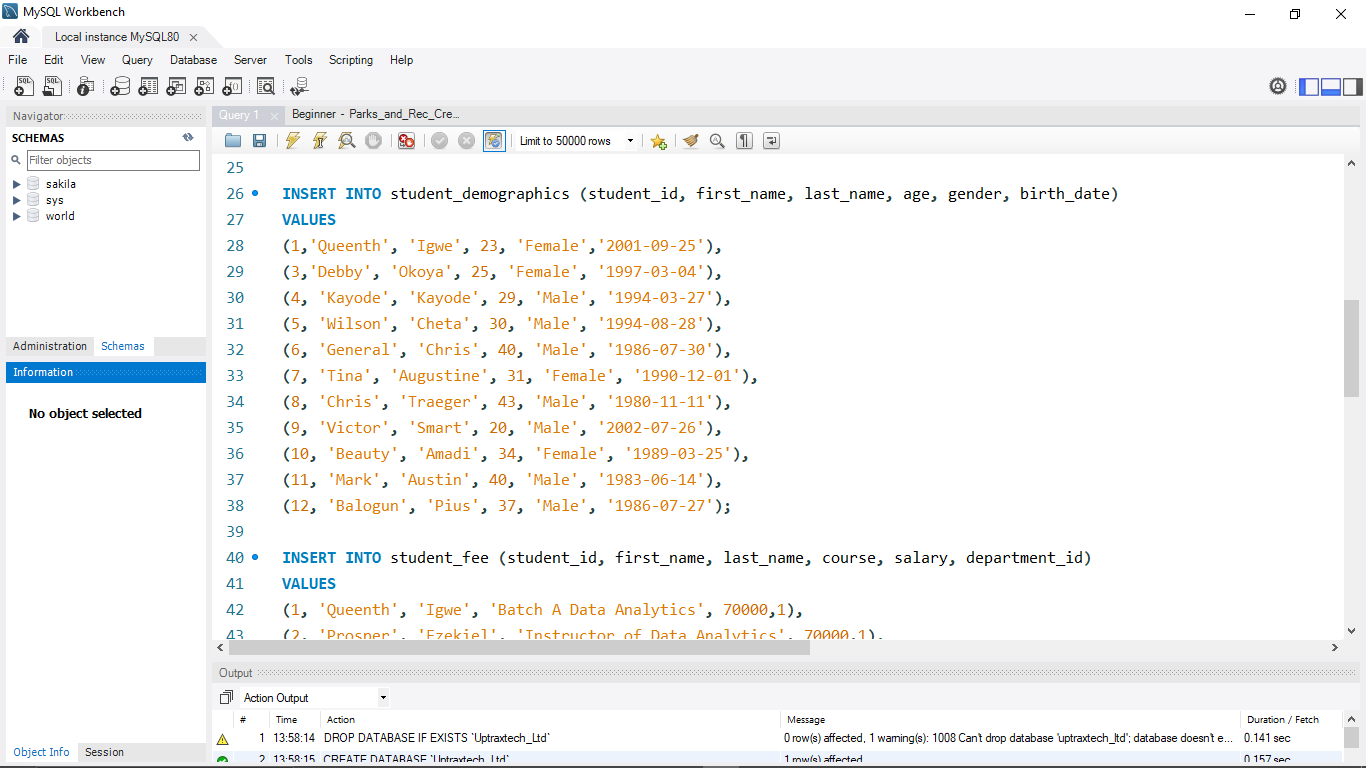
****

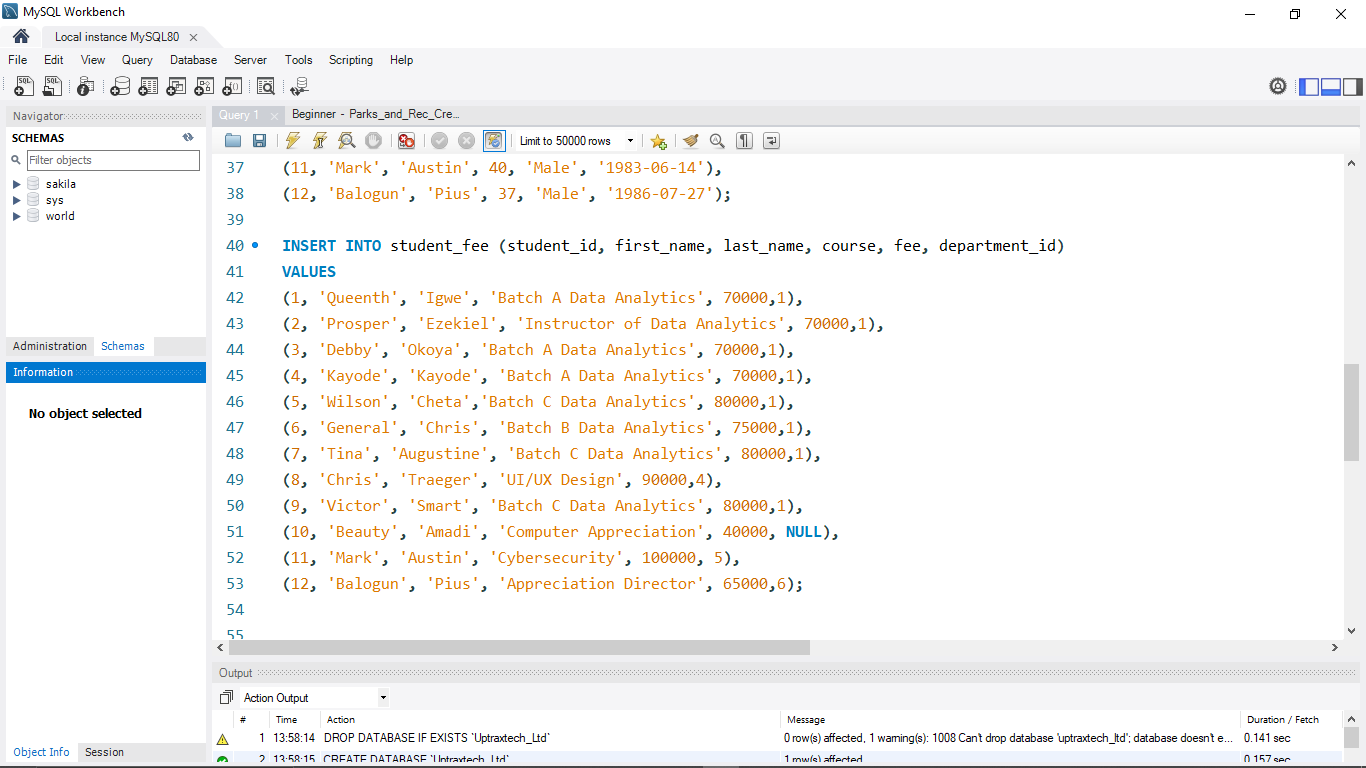
****

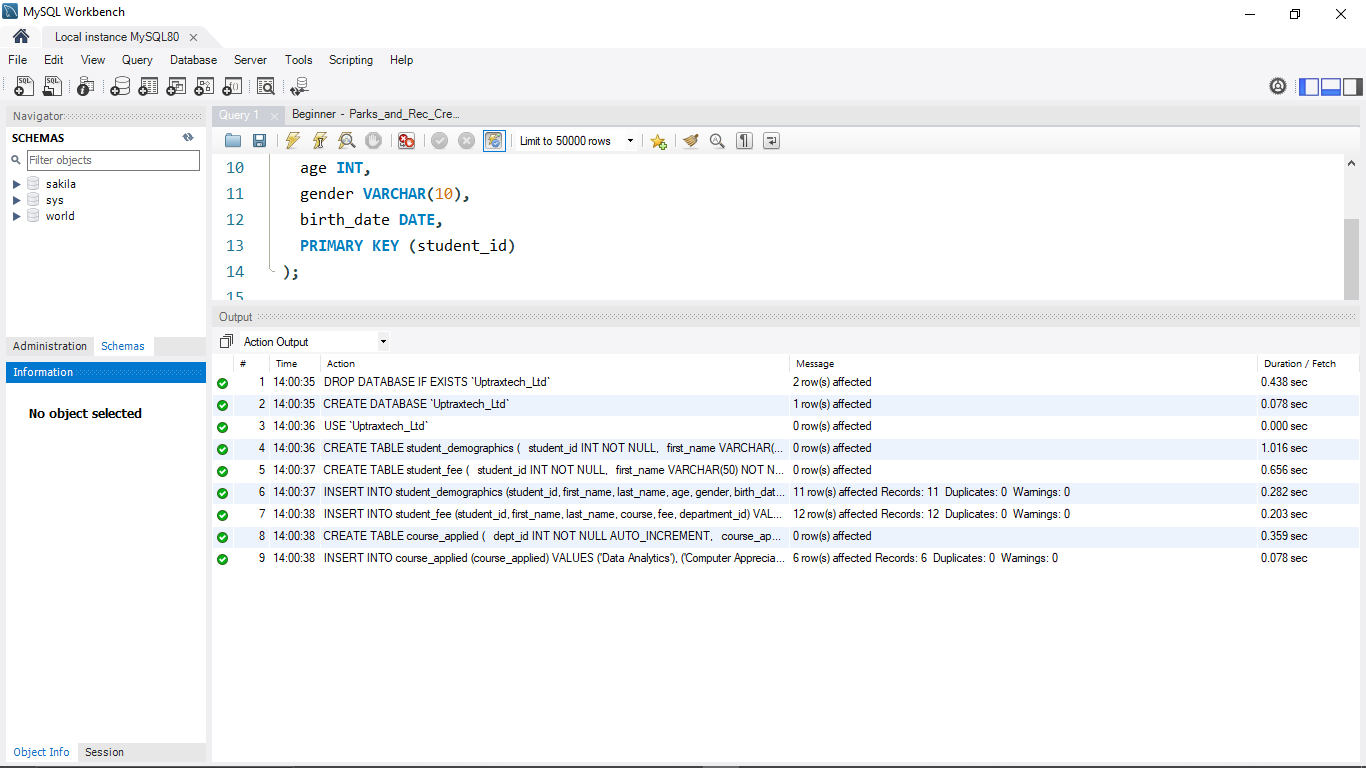
Students should create the following Database.

****

****

****

****

****

# 2 MICROSOFT EXCEL

**Excel is a powerful tool for data analysis that enable users to manipulate, analyze, and visualize large and complex amounts of data quickly. In this article, we will be looking at the following:**

* FORMULAS AND FUNCTIONS
* PIVOT TABLE
* CONDITIONAL FORMATTING
* XLOOKUP/VLOOKUP
* PIVOT CHARTS
* CUBE FUNCTIONS
* PROJECTS/DASHBOARDS
* TESTS/ASSIGNMNETS/MINI-PROJECTS

# 3 POWER BI

**Microsoft Power BI is used to find insights within an organization’s data. Power BI can help connect disparate datasets, transform, and clean data into data model and create charts and graphs to provide visuals of the data. In this article, we will be looking at the following:**

* POWER BI INSTALLATION
* POWER QUERY
* DATA ANALYSIS EXPRESSION (DAX)
* CREATING RELATIONSHIP
* CONDITIONAL FORMATTING
* BINS & LIST
* VISUALIZATIONS IN POWER BI
* DRILL DOWN & DRILL UP
* PROJECTS/DASHBOARDS
* TEST/ASSIGNMNETS/MINI-PROJECTS

# 4TABLEAU

Tableau is a powerful data visualization and business intelligence tool that helps users see and understand their data. It is widely used for transforming raw data into an understandable format through visualizations like graphs, charts, and dashboards, facilitating data-driven decision-making.

**In this article, we will be looking at the following:**

* TABLEAU INSTALLATION
* BINS/CALCULATED FIELDS
* JOINS
* TABLEAU CHARTS
* PROJECTS/TESTS/ASSIGNMENTS

### Key Features of Tableau

1. \*\*Data Visualization\*\*:

- Tableau excels in creating a wide range of visualizations, including bar charts, line charts, pie charts, scatter plots, histograms, and geographical maps.

- It provides interactive dashboards that allow users to drill down into the data, filter views, and discover insights.

2. \*\*Ease of Use\*\*:

- Tableau is designed for users of all technical levels. It uses a drag-and-drop interface that simplifies the process of creating visualizations.

- It supports real-time data analytics and can handle large volumes of data efficiently.

3. \*\*Integration and Connectivity\*\*:

- Tableau can connect to numerous data sources, including spreadsheets, databases, big data, cloud services, and other applications.

- It supports data blending and data joins, enabling users to combine data from different sources for comprehensive analysis.

4. \*\*Data Sharing and Collaboration\*\*:

- Users can share their visualizations and dashboards with others via Tableau Server, Tableau Online, or Tableau Public.

- Tableau supports collaboration, allowing multiple users to work on the same dashboard simultaneously.

5. \*\*Advanced Analytics\*\*:

- It offers robust analytical capabilities, including trend analysis, forecasting, statistical analysis, and what-if analysis.

- Tableau integrates with R and Python for advanced data science and machine learning tasks.

6. \*\*Customization and Flexibility\*\*:

- Tableau allows extensive customization of dashboards and reports to meet specific business needs.

- Users can create calculated fields, use parameters, and apply custom formatting to their visualizations.

### Tableau Products

1. \*\*Tableau Desktop\*\*:

- A development tool used to create and publish dashboards and visualizations.

- Provides a wide range of features for data preparation, analysis, and visualization.

2. \*\*Tableau Server\*\*:

- An enterprise-level product that allows users to share visualizations and dashboards across an organization.

- Provides centralized management, security, and collaboration features.

3. \*\*Tableau Online\*\*:

- A cloud-based version of Tableau Server, enabling users to share, collaborate, and manage content without the need for on-premise infrastructure.

4. \*\*Tableau Public\*\*:

- A free version of Tableau for creating and sharing public visualizations and dashboards.

- Best suited for individuals and small organizations with a focus on public data.

5. \*\*Tableau Prep\*\*:

- A tool for preparing, cleaning, and shaping data before analysis.

- Provides an intuitive interface for combining, reshaping, and cleaning data from various sources.

### Use Cases of Tableau

- \*\*Business Intelligence\*\*: Organizations use Tableau to track key performance indicators (KPIs), sales performance, and financial metrics.

- \*\*Market Research\*\*: Tableau helps in analyzing market trends, customer behavior, and competitive analysis.

- \*\*Healthcare\*\*: Used for patient data analysis, hospital performance tracking, and research.

- \*\*Education\*\*: Academic institutions use Tableau for analyzing student performance, enrollment trends, and research data.

- \*\*Public Sector\*\*: Governments and non-profits use Tableau to analyze and visualize public data, improve transparency, and drive policy decisions.

### Benefits of Using Tableau

- \*\*Speed and Efficiency\*\*: Quickly connects to data sources and generates visualizations.

- \*\*User-Friendly\*\*: Intuitive interface that requires minimal technical expertise.

- \*\*Powerful Insights\*\*: Helps uncover trends and patterns through interactive visualizations.

- \*\*Scalability\*\*: Can handle small datasets to large-scale enterprise data.

- \*\*Collaboration\*\*: Facilitates sharing and collaborative analysis across teams and organizations.

Tableau is widely regarded as one of the leading tools in the business intelligence and data visualization space, empowering users to make data-driven decisions with greater confidence and clarity.

## What is Data Analysis?