**New Age University**

Data Analytics and Business Intelligence Course

E-Book

**Course Description**

This comprehensive 8-week course aims to equip students with the essential skills and knowledge required to excel in the fields of Data Analytics (DA) and Business Intelligence (BI). The course is designed to cover both fundamental and advanced topics, ranging from data collection and cleaning to advanced analytics and visualization. Students will gain hands-on experience with popular tools and software used in the industry, including Excel, SQL, Python, Tableau, and Power BI. The course will culminate in a Capstone Project that integrates all learned concepts, preparing students for real-world challenges.

**Learning Objectives**

* Understand the core principles and differences between Data Analytics and Business Intelligence.
* Develop proficiency in data collection, cleaning, and preparation techniques.
* Gain expertise in using Excel for data analysis and visualization.
* Master SQL for database querying and manipulation.
* Learn Python programming for advanced data analysis and visualization.
* Create interactive and insightful visualizations using Tableau and Power BI.
* Apply statistical and machine learning techniques for advanced analytics.
* Complete a Capstone Project that demonstrates the practical application of course concepts.

# Week 1: Introduction to Data Analytics and Business Intelligence

**Overview:** An introduction to the foundational concepts of Data Analytics and Business Intelligence, focusing on their definitions, differences, and the importance of each in modern business environments.

**Topics Covered:**

1. Definitions and Key Roles
2. Data Types and Collection Methods
3. Data Quality and Cleaning
4. Introduction to Tools and Software
5. Real-World Applications

## 1. Definitions and Key Roles

### 1.1 Introduction to Data Analytics (DA)

Data Analytics (DA) involves examining datasets to draw meaningful conclusions about the information they contain. This encompasses a wide range of techniques from simple descriptive statistics to complex predictive modeling. The goal of DA is to uncover patterns, correlations, and trends within data that can inform business decisions and strategies. Data Analytics is a critical component of modern business operations, providing insights that can lead to improved efficiency, increased revenue, and competitive advantages.

Importance:

In today’s data-driven world, the ability to analyze and interpret data is invaluable. Businesses generate massive amounts of data through various channels such as transactions, customer interactions, and social media. Data Analytics enables organizations to convert this raw data into actionable insights, helping to understand customer behavior, optimize operations, predict future trends, and make informed decisions. For instance, through predictive analytics, businesses can forecast sales, manage inventory, and tailor marketing efforts to target specific customer segments effectively.

Key Techniques:

- Descriptive Analytics: Summarizes historical data to understand what has happened in the past. This includes techniques like data aggregation and data mining.

- Diagnostic Analytics: Examines historical data to determine why something happened. This often involves data discovery and drill-down analysis.

- Predictive Analytics: Uses statistical models and machine learning techniques to predict future outcomes based on historical data. Examples include regression analysis and time series forecasting.

- Prescriptive Analytics: Recommends actions to achieve desired outcomes. This involves the use of optimization and simulation algorithms to suggest the best course of action based on predictive models.

### 1.2 Introduction to Business Intelligence (BI)

Business Intelligence (BI) refers to the technologies, applications, and practices used for the collection, integration, analysis, and presentation of business information. BI is focused on providing comprehensive, up-to-date, and accurate information to support decision-making processes. BI systems compile data from various sources into a coherent format, making it accessible and useful for business analysis.

Importance:

BI tools and systems are essential for understanding the historical, current, and future state of business operations. They provide insights into various aspects such as sales performance, customer behavior, market trends, and operational efficiency. By leveraging BI, organizations can make data-driven decisions that enhance strategic planning, improve operational efficiency, and gain a competitive edge. For example, a BI dashboard might allow a retail company to track real-time sales data, monitor inventory levels, and analyze customer buying patterns to optimize stock levels and improve customer satisfaction.

Key Components:

- Data Warehousing: The centralized storage of data from different sources, making it easy to access and analyze.

- Data Mining: The process of discovering patterns and relationships in large datasets.

- Reporting and Query Tools: Tools that generate reports and enable users to query data to answer specific business questions.

- Dashboard and Visualization Tools: Applications that present data in visual formats like charts and graphs for easy interpretation.

- Performance Metrics and Benchmarking: Tools that measure performance against benchmarks to assess and improve business operations.

### 1.3 Differences between DA and BI

|  |  |  |
| --- | --- | --- |
|  | **Data Analysis** | **Business Intelligence** |
| Scope and Focus | - DA: Encompasses a broad range of techniques, including descriptive, diagnostic, predictive, and prescriptive analytics. It is more focused on extracting insights from data to forecast future trends and recommend actions. | - BI: Primarily focuses on descriptive and diagnostic analytics to understand historical and current data. It is more about reporting and querying to provide a snapshot of business performance. |
| Outcome | - DA: Produces models and simulations to predict future outcomes and suggest strategic actions. | - BI: Produces dashboards, reports, and visualizations that summarize past and present data to support decision-making. |
| Tools and Techniques | - DA: Utilizes statistical analysis, machine learning algorithms, and advanced modeling techniques (e.g., Python, R, Scikit-learn). | - BI: Uses tools like dashboards and reporting software (e.g., Power BI, Tableau) primarily for visualization and summary. |

### 1.4 Similarities between DA and BI

Data-Driven:

* Both fields rely on data to derive insights and support decision-making processes. They involve collecting, processing, and analyzing data to generate meaningful information.

Tools and Technologies:

* Both DA and BI use similar tools and technologies for data processing and visualization. For instance, SQL can be used in both fields to manage and query data, while tools like Tableau and Power BI are used for data visualization.

Goal:

* The ultimate goal of both DA and BI is to improve business performance by providing actionable insights. They help organizations understand their operations better, identify opportunities for improvement, and make informed decisions.

Activity

* Discussion: Engage in a class discussion about how DA and BI are used in various industries. Share examples and personal experiences.

Reading Materials

- Article: “The Evolution of Data Analytics and Business Intelligence”

- Book Chapter: “Introduction to Data Analytics” from “Data Analytics Made Accessible” by Anil Maheshwari

## 2. Data Types and Data Collection Methods

### 2.1 Data Types

Structured Data:

Structured data is highly organized and easily searchable in databases by simple, straightforward search engine algorithms or other search operations. This type of data resides in fixed fields within a record or file, such as databases or spreadsheets. Examples include names, dates, addresses, and credit card numbers. Structured data is typically stored in relational databases and can be queried using SQL. Because it is organized and formatted, structured data is easier to analyze and integrate into BI tools and processes.

Unstructured Data:

Unstructured data does not have a predefined format or organization, making it more complex to process and analyze. This type of data includes text documents, emails, videos, images, social media posts, and more. Unstructured data is often stored in its raw form and requires more advanced tools and techniques, such as natural language processing (NLP) and machine learning, to extract meaningful information. Despite its complexity, unstructured data holds valuable insights and is increasingly important in areas like sentiment analysis and customer feedback.

Semi-Structured Data:

Semi-structured data is a hybrid that contains elements of both structured and unstructured data. It does not conform to a rigid structure like relational databases but has some organizational properties that make it easier to analyze than completely unstructured data. Examples include JSON, XML, and HTML documents. Semi-structured data is often used in web development and data integration scenarios where flexible data formats are needed.

### 2.2 Data Collection Methods

Surveys:

Surveys are a common method of collecting structured data directly from individuals. They can be conducted online, via phone, in person, or through mailed questionnaires. Surveys are useful for gathering specific information from a large number of respondents. However, they can be prone to biases, such as self-selection and response bias.

Transaction Logs:

Transaction logs record the details of every transaction made within a system. These logs can provide a wealth of structured data, including timestamps, transaction amounts, and user actions. Transaction logs are particularly valuable in retail and financial services for tracking sales, purchases, and financial transactions.

Social Media:

Social media platforms generate vast amounts of unstructured data in the form of posts, comments, likes, and shares. Collecting data from social media requires specialized tools that can handle large volumes of text and multimedia content. This data is valuable for understanding customer sentiment, brand perception, and market trends.

Sensors:

Sensors embedded in various devices collect data related to physical conditions like temperature, motion, humidity, and more. This data is often semi-structured and is used in industries such as manufacturing, healthcare, and transportation to monitor and control processes in real-time.

Public Data Sources:

Public data sources include data released by government agencies, research institutions, and other organizations. This data can be both structured and unstructured and covers a wide range of topics such as demographics, economic indicators, and scientific research. Public data is valuable for benchmarking and augmenting internal data sources.

Activity

* Exercise: Identify and categorize examples of structured, unstructured, and semi-structured data from a provided dataset.

Reading Materials

- Research Paper: “Data Collection Methods: Challenges and Solutions”

- Book Chapter: “Data Types and Data Sources” from “Data Science for Business” by Foster Provost and Tom Fawcett

## 3. Data Quality and Data Cleaning

### 3.1 Importance of Data Quality

Impact on Decision-Making:

High-quality data is essential for accurate and reliable decision-making. Poor data quality can lead to incorrect insights, faulty strategies, and missed opportunities. Ensuring data quality means that the data is accurate, complete, consistent, timely, and relevant; which in turn enhances the reliability of analyses and business decisions.

### 3.2 Aspects of Data Quality:

- Accuracy: Data should be correct and free from errors.

- Completeness: All required data should be present.

- Consistency: Data should be the same across different datasets and systems.

- Timeliness: Data should be up-to-date and available when needed.

- Uniqueness: There should be no duplicate records within the dataset.

### 3.3 Common Data Quality Issues

Missing Values:1

Missing data occurs when no value is stored for a data point. This can happen due to various reasons such as data entry errors, system issues, or respondents skipping questions in surveys. Missing values can lead to biased analyses and incomplete conclusions if not properly handled.

Duplicate Records:

Duplicate records refer to multiple entries that represent the same entity in a dataset. These can occur due to data entry errors, merging datasets from different sources, or system glitches. Duplicates can distort statistical analyses and lead to inaccurate results.

Inconsistent Data:

Inconsistent data arises when the same entity has different values in different datasets or within the same dataset. This can be due to various reasons like different formats, data entry mistakes, or discrepancies in data collection methods. Inconsistencies need to be identified and resolved to ensure data integrity.

Outliers and Anomalies:

Outliers are data points that significantly differ from other observations. While some outliers can provide valuable insights, others may be the result of data entry errors or measurement issues. Identifying and handling outliers is crucial for accurate data analysis.

### 3.4 Data Cleaning Techniques

Handling Missing Values:

- Imputation: Replacing missing values with substituted values. This can be done using the mean, median, mode, or using more sophisticated methods like regression or machine learning algorithms.

- Deletion: Removing records with missing values. This approach is viable if the proportion of missing data is relatively small.

Removing Duplicates:

- Identifying Duplicates: Use functions in data processing tools (e.g., Excel, SQL) to identify and flag duplicate records.

- Merging Duplicates: Combine information from duplicate records to create a single, comprehensive record.

Standardizing Data Formats:

- Consistent Formatting: Ensure consistency in data formats across the dataset. For example, dates should be in the same format (e.g., YYYY-MM-DD), and units of measurement should be standardized.

- Normalization: Adjusting the values in the dataset to a common scale without distorting differences in the range of values.

Outlier Detection and Treatment:

- Statistical Methods: Use statistical techniques like z-scores or IQR (Interquartile Range) to identify outliers.

- Treatment: Depending on the context, outliers can be removed, transformed, or investigated further for insights.

Activity

* Practical Exercise: Clean a sample dataset by identifying and addressing missing values, duplicates, inconsistencies, and outliers using data processing tools.

Reading Materials

- Article: “Data Quality and Its Impact on Business Intelligence”

- Video: “

- Book Chapter: “Data Cleaning Techniques” from “Data Science from Scratch” by Joel Grus

## 4. Introduction to Tools and Software

### 4.1 Overview of Popular Tools

Excel:

Excel is a powerful spreadsheet tool used for data analysis, visualization, and reporting. It offers a wide range of functions, formulas, pivot tables, and charting capabilities. Excel is widely used due to its accessibility and versatility, making it a fundamental tool for data analysts.

SQL:

SQL (Structured Query Language) is the standard language for managing and manipulating relational databases. It allows users to query, update, and manage data efficiently. SQL is essential for working with large datasets stored in relational databases, making it a crucial skill for data analysts and BI professionals.

Python:

Python is a versatile programming language widely used in data analytics and data science. It offers powerful libraries such as Pandas for data manipulation, Matplotlib and Seaborn for data visualization, and Scikit-learn for machine learning. Python is favored for its simplicity, readability, and extensive community support.

R:

R is a programming language and software environment specifically designed for statistical computing and graphics. It is widely used in academia and industry for data analysis, visualization, and modeling. R provides numerous packages and functions for statistical analysis, making it a popular choice for data scientists.

Tableau:

Tableau is a leading BI tool known for its data visualization capabilities. It allows users to create interactive and shareable dashboards. Tableau connects to various data sources and provides a user-friendly interface for building complex visualizations, making data insights more accessible to non-technical users.

Power BI:

Power BI is a BI and data visualization tool developed by Microsoft. It enables users to create interactive dashboards and reports, connecting to a wide range of data sources. Power BI integrates seamlessly with other Microsoft products and offers powerful features for data modeling and visualization.

### 4.2 Tool Selection Criteria

When choosing tools for DA and BI, consider the following factors:

- Ease of Use: The tool should be user-friendly and easy to learn, especially for those with limited technical skills.

- Scalability: The tool should be able to handle large volumes of data and scale as the organization’s data needs grow.

- Integration: The tool should integrate smoothly with other software and data sources used by the organization.

- Cost: Consider the total cost of ownership, including licensing fees, training, and maintenance.

- Community and Support: A strong community and support system can be invaluable for troubleshooting and learning best practices.

Activity

* Tool Demo: Demonstrate the basic functionalities of Excel, SQL, Python, R, Tableau, and Power BI. Discuss the pros and cons of each tool and scenarios where they are most effective.

Reading Materials

- Article: “Choosing the Right Data Analysis Tool”

- Book Chapter: “Introduction to Data Analysis Tools” from “Data Science for Dummies” by Lillian Pierson

## 5. Real-World Applications

### 5.1 Case Studies and Real-World Applications

Case Study 1: Retail Analytics

Background:

A major retail chain implemented a data analytics program to understand customer buying behavior and optimize inventory management. By analyzing transaction logs, customer demographics, and social media data, the company aimed to improve sales and customer satisfaction.

Implementation:

The retail chain used a combination of Excel for initial data analysis, SQL for managing and querying large datasets, and Tableau for visualizing sales trends and customer segments. Python was used for predictive analytics to forecast demand and manage inventory levels.

Results:

The data analytics program led to a 15% increase in sales and a 20% reduction in stockouts. Customer satisfaction scores improved significantly as the company was able to stock products that better met customer preferences and demand patterns.

### 5.2 Case Study 2: Healthcare Analytics

Background:

A healthcare provider sought to improve patient outcomes and operational efficiency through data analytics. By integrating electronic health records (HER), patient feedback, and operational data, the provider aimed to identify areas for improvement and optimize resource allocation.

Implementation:

The healthcare provider used R for statistical analysis and data visualization, along with Power BI for creating interactive dashboards to monitor key performance indicators (KPIs). Machine learning models in Python were used to predict patient readmissions and optimize staffing levels.

Results:

The analytics initiatives resulted in a 10% reduction in patient readmissions and a 15% improvement in resource utilization. The provider was able to allocate staff more effectively, reducing wait times and improving patient care.

### 5.3 Case Study 3: Financial Services Analytics

Background:

A financial services firm aimed to enhance its risk management processes and improve customer retention through data analytics. By analyzing transaction data, customer profiles, and market trends, the firm sought to develop more effective risk assessment models and personalized customer engagement strategies.

Implementation:

The firm used SQL for data extraction and management, Excel for initial data analysis, and Tableau for creating detailed visualizations of risk factors and customer segments. Python was employed for building predictive models to assess credit risk and identify at-risk customers.

Results:

The firm experienced a 25% reduction in loan defaults and a 20% increase in customer retention. The predictive models allowed for more accurate risk assessments and targeted customer retention efforts, leading to improved financial performance.

Activity

* Group Discussion: Analyze the provided case studies and discuss the methodologies, tools, and outcomes. Identify key takeaways and how similar strategies could be applied in different industries.

Reading Materials

- Case Study Report: “How Data Analytics Transformed Retail Operations”

- Case Study Report: “Improving Healthcare with Data Analytics”

- Case Study Report: “Enhancing Financial Services with Predictive Analytics”

## Additional Resources for Week 1

Books

- “Data Science for Business” by Foster Provost and Tom Fawcett

- “Data Analytics Made Accessible” by Anil Maheshwari

Articles

- “The Evolution of Data Analytics and Business Intelligence”

- “Data Quality and Its Impact on Business Intelligence”

# Week 2: Data Analysis with Excel

**Overview**: Hands-on training with Excel to perform basic and advanced data analysis tasks, focusing on data manipulation, visualization, and automation techniques.

**Topics Covered:**

1. Excel Basics: Functions, Formulas, and Data Manipulation
2. Advanced Excel Techniques: Solver, Analysis ToolPak, Macros, and VBA
3. Data Visualization with Excel: Charts, Graphs, and Dashboards
4. Practical Project

## 1. Excel Basics

### 2.1 Introduction to Excel:

Excel is a spreadsheet program developed by Microsoft, widely used for data analysis, visualization, and reporting. Its grid of rows and columns allows users to organize and manipulate data efficiently.

### Functions and Formulas

Excel functions are predefined formulas that perform specific calculations using the data in your worksheet. Common functions include:

SUM: Adds a range of cells.

AVERAGE: Calculates the average of a range of cells.

IF: Performs logical tests and returns values based on the results.

VLOOKUP: Searches for a value in the first column of a range and returns a value in the same row from another column.

### Data Manipulation

Excel provides various tools for data manipulation, including sorting, filtering, and conditional formatting. These tools help in organizing and highlighting data to facilitate analysis.

### Pivot Tables:

Pivot tables are powerful tools for summarizing, analyzing, exploring, and presenting data. They allow users to transform large datasets into meaningful summaries and insights.

### Charts and Graphs:

Excel offers a variety of chart types, including bar charts, line charts, pie charts, and scatter plots. These visual representations help in understanding data trends and patterns.

### Practical Exercise

Hands-On Practice: Perform basic data manipulation, create pivot tables, and generate charts using a sample dataset.

## 2. Advanced Excel Techniques

### Data Analysis Tools:

Excel provides several built-in tools for data analysis:

Solver: An optimization tool that finds the best solution for decision problems.

Analysis ToolPak: A collection of data analysis tools for statistical and engineering analysis.

### Macros:

Macros are sequences of instructions that automate repetitive tasks in Excel. They are written in Visual Basic for Applications (VBA), allowing users to perform complex operations with a single click.

### VBA Basics:

VBA is a programming language used to create macros in Excel. It allows users to automate tasks, create custom functions, and interact with Excel objects programmatically.

### Advanced Formulas:

Advanced formulas include array formulas, nested functions, and custom formulas using VBA. These techniques enable more complex and powerful data analysis.

### Practical Exercise

Hands-On Practice: Use the Solver and Analysis ToolPak for data analysis. Create and run macros to automate tasks and write simple VBA scripts.

## 3. Data Visualization with Excel

### Creating Effective Charts:

Choosing the right chart type and design is crucial for effective data visualization. Best practices include:

Clear Titles and Labels: Ensure that charts have clear and descriptive titles and axis labels.

Consistent Colors and Styles: Use consistent color schemes and styles to enhance readability.

Avoiding Clutter: Keep charts simple and avoid unnecessary elements that can distract from the data.

### Dashboards:

Dashboards are collections of charts, tables, and other visual elements that provide an overview of key metrics and trends. They should be interactive, easy to read, and tailored to the audience's needs.

### Dynamic Visualizations:

Dynamic visualizations update automatically based on changes

in the underlying data. Techniques include using slicers, data validation, and linking charts to dynamic data ranges.

### Practical Exercise

Hands-On Practice: Create charts, design a dashboard, and implement dynamic visualizations using a sample dataset.

## 4. Excel Project

Project Description

### Objective:

Apply the skills learned throughout the week to analyze a real-world dataset, create visualizations, and present findings in a comprehensive report.

### Dataset:

Use a provided dataset (e.g., sales data, customer feedback data, financial data) for analysis.

### Tasks:

Data Cleaning and Preparation: Clean and prepare the dataset for analysis.

Data Analysis: Perform data analysis using pivot tables, formulas, and advanced techniques.

Data Visualization: Create effective charts and design a dashboard to present findings.

Report Writing: Write a report summarizing the analysis process, results, and insights.

### Presentation:

Prepare a presentation to share the findings with peers, demonstrating the use of Excel for data analysis and visualization.

## Additional Resources for Week 2

Books

"Excel 2019 Bible" by Michael Alexander, Richard Kusleika, and John Walkenbach

"Excel Data Analysis For Dummies" by Stephen L. Nelson and E. C. Nelson

Articles

"Top Excel Data Analysis Tools" by DataCamp

"Best Practices for Excel Dashboards" by Excel Campus

# Week 3: Introduction to SQL

**Overview:** Students will be introduced to SQL (Structured Query Language), a fundamental tool for managing and manipulating relational databases. The focus will be on understanding the basics of SQL, mastering advanced queries, learning data manipulation techniques, and applying these skills to practical projects.

**Learning Objectives**

- Understand the basics of relational databases and SQL.

- Master basic and advanced SQL queries.

- Learn data manipulation techniques using SQL.

- Optimize SQL queries for better performance.

- Apply SQL skills to real-world database projects.

## Introduction to Relational Databases:

Relational databases store data in tables, which consist of rows and columns. Each table represents a different entity, and relationships between tables are defined using primary and foreign keys.

### SQL Overview:

SQL is a standard language for querying and manipulating databases. It allows users to perform various operations, such as retrieving data, inserting new records, updating existing records, and deleting records.

### Basic SQL Queries:

- SELECT: Retrieves data from one or more tables.

```sql

SELECT column1, column2 FROM table\_name;

```

- FROM: Specifies the table from which to retrieve data.

- WHERE: Filters records based on specified conditions.

```sql

SELECT FROM customers WHERE country = ‘USA’;

```

### Data Types:

SQL supports various data types, including integers, floats, strings, dates, and more. Understanding data types is crucial for designing tables and writing accurate queries.

### Practical Exercise

- Hands-On Practice: Write basic SQL queries to retrieve data from sample databases. Practice using the SELECT, FROM, and WHERE clauses to filter and sort data.

## 2. Advanced SQL Queries

### JOINs:

JOINs are used to combine rows from two or more tables based on related columns. Common types of JOINs include:

- INNER JOIN: Returns only the matching rows from both tables.

```sql

SELECT orders.order\_id, customers.customer\_name

FROM orders

INNER JOIN customers ON orders.customer\_id = customers.customer\_id;

```

- LEFT JOIN (or LEFT OUTER JOIN): Returns all rows from the left table and the matched rows from the right table. If no match, NULL values are returned.

```sql

SELECT orders.order\_id, customers.customer\_name

FROM orders

LEFT JOIN customers ON orders.customer\_id = customers.customer\_id;

```

- RIGHT JOIN (or RIGHT OUTER JOIN): Returns all rows from the right table and the matched rows from the left table. If no match, NULL values are returned.

- FULL JOIN (or FULL OUTER JOIN): Returns rows when there is a match in one of the tables. If there is no match, NULL values are returned for columns from the table that lacks the match.

### Subqueries:

Subqueries are nested queries used within another SQL query. They are useful for performing complex filtering and calculations.

```sql

SELECT customer\_name

FROM customers

WHERE customer\_id IN (SELECT customer\_id FROM orders WHERE order\_date > ‘2024-01-01’);

```

### Aggregations:

Aggregation functions perform calculations on a set of values and return a single value. Common aggregation functions include:

- COUNT: Counts the number of rows.

```sql

SELECT COUNT() FROM orders;

```

- SUM: Calculates the sum of a numeric column.

```sql

SELECT SUM(total\_amount) FROM orders;

```

- AVG: Calculates the average value of a numeric column.

```sql

SELECT AVG(total\_amount) FROM orders;

```

- MAX: Returns the maximum value.

```sql

SELECT MAX(total\_amount) FROM orders;

```

- MIN: Returns the minimum value.

```sql

SELECT MIN(total\_amount) FROM orders;

```

### Practical Exercise

- Hands-On Practice: Write advanced SQL queries using JOINs, subqueries, and aggregation functions. Practice combining multiple tables and performing complex data retrievals.

## 3. SQL Data Manipulation

### INSERT Statement:

The INSERT statement is used to add new rows to a table.

```sql

INSERT INTO customers (customer\_name, contact\_name, country)

VALUES (‘John Doe’, ‘John’, ‘USA’);

```

### UPDATE Statement:

The UPDATE statement modifies existing rows in a table.

```sql

UPDATE customers

SET contact\_name = ‘Jane’

WHERE customer\_name = ‘John Doe’;

```

### DELETE Statement:

The DELETE statement removes rows from a table.

```sql

DELETE FROM customers

WHERE customer\_name = ‘John Doe’;

```

### Data Integrity:

Maintaining data integrity involves ensuring the accuracy and consistency of data within a database. Key concepts include:

- Primary Keys: Unique identifiers for each row in a table.

- Foreign Keys: Columns that create a relationship between tables by referencing primary keys.

- Constraints: Rules applied to columns to enforce data integrity (e.g., NOT NULL, UNIQUE, CHECK).

### Transactions:

Transactions ensure that a series of SQL operations are executed as a single unit, maintaining data integrity. If any operation fails, the entire transaction is rolled back.

```sql

BEGIN TRANSACTION;

UPDATE accounts SET balance = balance – 100 WHERE account\_id = 1;

UPDATE accounts SET balance = balance + 100 WHERE account\_id = 2;

COMMIT;

```

### Practical Exercise

- Hands-On Practice: Write SQL statements to insert, update, and delete data in sample databases. Practice using primary keys, foreign keys, and constraints to maintain data integrity. Implement transactions to ensure data consistency.

## 4. SQL Performance Tuning

### Indexes:

Indexes improve the speed of data retrieval operations. They can be created on one or more columns of a table.

```sql

CREATE INDEX idx\_customer\_name ON customers (customer\_name);

```

### Query Optimization:

Optimizing SQL queries involves writing efficient queries to minimize execution time and resource usage. Best practices include:

- Using Indexes: Ensure that columns used in WHERE clauses and JOIN conditions are indexed.

- Avoiding Unnecessary Columns: Select only the necessary columns to reduce data transfer.

- Limiting Results: Use LIMIT or TOP to restrict the number of rows returned.

```sql

SELECT customer\_name FROM customers LIMIT 10;

```

### Explain Plan:

The EXPLAIN plan shows the execution strategy of a SQL query, helping to identify performance bottlenecks.

```sql

EXPLAIN SELECT FROM customers WHERE country = ‘USA’;

```

### Practical Exercise

- Hands-On Practice: Create indexes on sample tables and write optimized SQL queries. Use EXPLAIN plans to analyze query performance and make necessary adjustments.

## 5. SQL Project

Project Description

### Objective:

Apply the skills learned throughout the week to design, implement, and query a relational database. The project involves creating tables, inserting data, performing complex queries, and optimizing performance.

### Dataset:

Use a provided dataset (e.g., sales data, customer feedback data, financial data) for the project.

### Tasks:

- Database Design: Design the schema, including tables, columns, primary keys, and foreign keys.

- Data Insertion: Populate the tables with sample data using INSERT statements.

- Complex Queries: Write and execute advanced SQL queries to retrieve and analyze data.

- Performance Tuning: Optimize the queries for better performance using indexes and query optimization techniques.

### Report Writing:

Write a report summarizing the database design, data manipulation, queries executed, and performance optimization strategies.

### Presentation:

Prepare a presentation to share the findings with peers, demonstrating the use of SQL for database management and data analysis.

## 6. Additional Resources for Week 3

Books

- “SQL for Data Scientists: A Beginner’s Guide for Building Datasets for Analysis” by Renee M. P. Teate

- “Learning SQL: Generate, Manipulate, and Retrieve Data” by Alan Beaulieu

Articles

- “SQL Basics for Beginners” by Codecademy

- “Advanced SQL Techniques: Tips and Tricks” by DataCamp

# Week 4: Data Analysis with Python

**Overview:** In Week 4, students will explore data analysis using Python, a versatile programming language widely used in data science and analytics. The focus will be on understanding Python fundamentals, data manipulation with Pandas, data visualization with Matplotlib and Seaborn, and applying these skills to practical projects.

**Learning Objectives**

- Understand Python basics and its role in data analysis.

- Master data manipulation techniques using Pandas.

- Learn data visualization using Matplotlib and Seaborn.

- Apply Python skills to real-world data analysis projects.Topics Covered:

1. Definitions and Key Roles
2. Data Types and Collection Methods
3. Data Quality and Cleaning
4. Introduction to Tools and Software
5. Real-World Applications

## 1. Introduction to Python for Data Analysis

### Python Basics:

Python is a high-level programming language known for its simplicity and readability. It offers rich libraries and tools for data analysis, making it popular among data scientists and analysts.

### Setting up the Environment:

Setting up Python for data analysis involves installing Anaconda, a distribution of Python that includes data science packages such as Pandas, NumPy, Matplotlib, and more. Jupyter Notebooks are commonly used for interactive data analysis and visualization.

### Syntax and Data Structures:

- Variables and Data Types: Integers, floats, strings, lists, tuples, dictionaries.

- Control Flow: Conditional statements (if, else), loops (for, while).

- Functions: Defining and calling functions to reuse code.

### Practical Exercise

- Hands-On Practice: Install Anaconda, create a Jupyter Notebook, and write Python code to perform basic operations (e.g., arithmetic operations, string manipulation).

## 2. Data Manipulation with Pandas

### Introduction to Pandas:

Pandas is a powerful library for data manipulation and analysis in Python. It provides data structures like Series (1-dimensional) and DataFrame (2-dimensional), which are used to handle and analyze tabular data.

### DataFrames and Series:

- Creating DataFrames: Loading data from files (e.g., CSV, Excel), creating DataFrames from dictionaries, lists, or NumPy arrays.

- Indexing and Selecting Data: Accessing rows, columns, and individual elements using labels or positions.

### Data Cleaning and Transformation:

- Handling Missing Data: Removing or imputing missing values using fillna(), dropna().

- Data Manipulation: Applying functions, sorting, merging, and reshaping data.

### Practical Exercise

- Hands-On Practice: Use Pandas to load a dataset, perform data cleaning (e.g., handling missing values, removing duplicates), and transform the data (e.g., filtering, sorting).

## 3. Data Visualization with Matplotlib and Seaborn

### Introduction to Matplotlib:

Matplotlib is a plotting library for Python, used to create static, animated, and interactive visualizations. It provides a MATLAB-like interface for plotting.

### Basic Plots with Matplotlib:

- Line Plot: Visualizing trends over time.

- Bar Plot: Comparing categorical data.

- Scatter Plot: Examining relationships between variables.

### Introduction to Seaborn:

Seaborn is a statistical data visualization library based on Matplotlib. It provides a higher-level interface for drawing attractive and informative statistical graphics.

### Advanced Visualizations with Seaborn:

- Histograms and Density Plots: Visualizing distributions of data.

- Pair Plots and Heatmaps: Exploring relationships and correlations between variables.

### Practical Exercise

- Hands-On Practice: Use Matplotlib and Seaborn to create various types of plots (e.g., line plots, bar plots, scatter plots, histograms) using sample datasets. Customize plots with titles, labels, legends, and color schemes.

## 4. Python Project

Project Description

### Objective:

Apply the skills learned throughout the week to analyze a real-world dataset using Python. The project involves loading data, performing data manipulation and analysis with Pandas, creating visualizations with Matplotlib and Seaborn, and presenting findings in a report.

### Dataset:

Use a provided dataset (e.g., housing prices, stock market data, weather data) for analysis.

### Tasks:

- Data Loading and Cleaning: Load the dataset into a Pandas DataFrame, clean and prepare the data for analysis.

- Data Analysis: Use Pandas to perform exploratory data analysis (EDA), calculate summary statistics, and extract insights.

- Data Visualization: Create meaningful visualizations (e.g., line charts, scatter plots, histograms) to illustrate trends and patterns in the data.

- Report Writing: Write a report summarizing the analysis process, key findings, and visualizations.

### Presentation:

Prepare a presentation to showcase the analysis and findings, demonstrating proficiency in Python for data analysis and visualization.

## Additional Resources for Week 4

Books

- “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython” by Wes McKinney

- “Python Data Visualization Cookbook” by Igor Milovanović

Articles

- “Data Visualization with Matplotlib and Seaborn” by Real Python

- “Pandas Documentation” by Pandas Development Team

# Week 5: Business Intelligence with Tableau

**Overview:** Week 5 focuses on Business Intelligence (BI) using Tableau, a powerful data visualization tool that allows users to create interactive and shareable dashboards. Participants will learn to connect to data sources, create visualizations, and design effective dashboards to derive insights and support decision-making.

**Learning Objectives**

- Understand the role of Business Intelligence (BI) in data-driven decision-making.

- Master Tableau basics, including connecting to data sources and creating visualizations.

- Learn advanced Tableau features such as calculations, parameters, and filters.

- Design interactive and informative dashboards using best practices in data visualization.

**Overview of Business Intelligence (BI)**

Business Intelligence involves technologies and strategies for collecting, analyzing, and presenting business information to support decision-making processes.

## 1. Introduction to Tableau

### Introduction to Tableau:

Tableau is a leading BI tool known for its intuitive drag-and-drop interface and powerful visualization capabilities. It allows users to connect to various data sources, create dynamic visualizations, and share insights across organizations.

### Tableau Interface Overview:

- Data Source Connection: Connecting to data sources such as Excel, CSV files, databases, and web data connectors.

- Worksheet vs. Dashboard: Understanding the difference between worksheets (individual visualizations) and dashboards (composite views).

### Practical Exercise

- Hands-On Practice: Install Tableau Public or Tableau Desktop (trial version), connect to a sample dataset, and explore the Tableau interface. Create basic visualizations (e.g., bar charts, line charts) using sample data.

## 2. Creating Visualizations in Tableau

### Charts and Graphs in Tableau:

Tableau offers a wide range of visualization types to represent data effectively:

- Bar Charts, Line Charts, Area Charts: Visualizing trends and comparisons.

- Pie Charts, Scatter Plots: Showing distributions and relationships.

- Maps: Geographical data visualization using maps.

### Formatting and Customization:

- Color, Size, Shape: Customizing visual elements to enhance clarity and emphasis.

- Labels, Tooltips: Adding annotations and interactive elements to visualizations.

### Practical Exercise

- Hands-On Practice: Create various types of visualizations in Tableau using sample datasets. Customize visualizations by adjusting colors, labels, and adding interactive features (e.g., tooltips, filters).

## 3. Advanced Tableau Features

### Calculations and Parameters:

- Calculations: Creating calculated fields to perform complex calculations (e.g., aggregations, conditional logic).

- Parameters: Using parameters to allow users to control aspects of the visualization dynamically.

### Filters and Sets:

- Filters: Applying filters to focus on specific subsets of data.

- Sets: Creating sets to group and analyze data based on common attributes or conditions.

### Practical Exercise

- Hands-On Practice: Explore advanced Tableau features by creating calculated fields, applying filters, and using parameters to build interactive dashboards. Use sets to segment data and perform comparative analysis.

## 4. Tableau Best Practices

### Designing Effective Dashboards:

- Layout and Structure: Organizing visualizations for clarity and flow.

- Storytelling: Using dashboards to tell a compelling data-driven story.

- Performance Optimization: Improving dashboard performance by optimizing data queries and visual elements.

### Interactivity and User Experience:

- Actions and Filters: Adding interactivity through actions like dashboard filters and highlight actions.

- User Interface (UI) Design: Designing intuitive user interfaces for seamless navigation.

### Practical Exercise

- Hands-On Practice: Design and develop a comprehensive dashboard in Tableau. Incorporate best practices for layout, storytelling, and interactivity. Optimize dashboard performance and user experience.

## 5. Tableau Project

### Objective:

Apply Tableau skills learned throughout the week to create a business intelligence dashboard using real-world data. The project involves connecting to data sources, designing visualizations, building interactive dashboards, and presenting insights.

### Dataset:

Use a provided dataset (e.g., sales performance, customer demographics, marketing campaign data) for the project.

### Tasks:

- Data Connection: Connect Tableau to the dataset and perform data preprocessing if necessary.

- Visualization Design: Create multiple visualizations (e.g., charts, graphs, maps) to explore different aspects of the data.

- Dashboard Creation: Build a cohesive dashboard with interactive elements (e.g., filters, parameters) to convey insights effectively.

- Presentation: Prepare a presentation to showcase the Tableau dashboard, highlighting key findings and actionable insights.

### Report Writing:

Write a report documenting the dashboard design process, data analysis techniques applied, and recommendations based on the insights gained.

## Additional Resources for Week 5

Books

- “Learning Tableau 2023: Master Tableau for Business Intelligence, Data Visualization, and Data Science” by Joshua N. Milligan

- “Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software” by Daniel G. Murray

Articles

- “Getting Started with Tableau” by Tableau Software

- “Tableau Tutorials” by VizWiz

# Week 6: Business Intelligence with Power BI

**Overview:** Week 6 focuses on Business Intelligence (BI) using Power BI, a leading data visualization and business analytics tool by Microsoft. Participants will learn to connect to data sources, create interactive visualizations, and design effective dashboards to derive insights and support decision-making.

**Learning Objectives**

- Understand the role of Business Intelligence (BI) in data-driven decision-making.

- Master Power BI basics, including connecting to data sources and creating visualizations.

- Learn advanced Power BI features such as DAX functions and data modeling.

- Design interactive and informative dashboards using best practices in data visualization.

## 1. Introduction to Power BI

### Overview of Business Intelligence (BI):

Business Intelligence involves technologies and strategies for collecting, analyzing, and presenting business information to support decision-making processes.

### Introduction to Power BI:

Power BI is a powerful suite of business analytics tools developed by Microsoft. It allows users to connect to a wide variety of data sources, create interactive visualizations, and share insights across organizations.

### Power BI Interface Overview:

- Data Source Connection: Connecting to data sources such as Excel files, databases (SQL Server, MySQL), cloud services (Azure, Google Analytics), and web data connectors.

- Data Model: Understanding Power BI’s data model and relationships between tables.

### Practical Exercise

- Hands-On Practice: Install Power BI Desktop (free version), connect to a sample dataset, and explore the Power BI interface. Create basic visualizations (e.g., bar charts, line charts) using sample data.

## 2. Creating Visualizations in Power BI

### Charts and Graphs in Power BI:

Power BI offers a variety of visualization types to represent data effectively:

- Bar Charts, Line Charts, Area Charts: Visualizing trends and comparisons.

- Pie Charts, Scatter Plots: Showing distributions and relationships.

- Maps: Geographical data visualization using maps.

### Formatting and Customization:

- Themes and Styles: Applying predefined themes and customizing visual elements (e.g., colors, fonts).

- Interaction and Drill-Down: Adding interactivity through drill-down features and cross-filtering.

### Practical Exercise

- Hands-On Practice: Create various types of visualizations in Power BI using sample datasets. Customize visualizations by adjusting formatting, applying themes, and adding interactive features (e.g., tooltips, drill-through).

## 3. Advanced Power BI Features

### DAX Functions (Data Analysis Expressions):

- Introduction to DAX: Understanding the role of DAX in Power BI for creating calculated columns, measures, and calculated tables.

- Common DAX Functions: Aggregate functions (SUM, AVERAGE), time intelligence functions (DATESYTD, SAMEPERIODLASTYEAR), logical functions (IF, SWITCH).

### Data Modeling in Power BI:

- Relationships: Defining and managing relationships between tables in Power BI’s data model.

- Hierarchies and Filters: Creating hierarchies for organizational data and applying filters for dynamic data analysis.

### Practical Exercise

- Hands-On Practice: Explore advanced Power BI features by creating DAX calculations (e.g., year-to-date sales, cumulative totals), optimizing data models with relationships, and using hierarchies and filters to perform in-depth data analysis.

## 4. Power BI Best Practices

### Designing Effective Dashboards:

- Layout and Structure: Organizing visualizations for clarity and flow in Power BI dashboards.

- Data Storytelling: Using Power BI to tell a compelling data-driven story and convey insights effectively.

- Performance Optimization: Improving dashboard performance by optimizing data models and queries.

### Collaboration and Sharing:

- Publishing and Sharing: Publishing Power BI reports to Power BI Service and sharing them with stakeholders. Using Power BI Apps for controlled distribution.

### Practical Exercise

- Hands-On Practice: Design and develop a comprehensive dashboard in Power BI. Incorporate best practices for layout, storytelling, and performance optimization. Publish the dashboard to Power BI Service and share it with peers for feedback.

## 5. Power BI Project

Project Description

### Objective:

Apply Power BI skills learned throughout the week to create a business intelligence dashboard using real-world data. The project involves connecting to data sources, designing visualizations, building interactive dashboards, and presenting insights.

### Dataset:

Use a provided dataset (e.g., sales performance, customer demographics, marketing campaign data) for the project.

### Tasks:

- Data Connection: Connect Power BI to the dataset and perform data preprocessing if necessary.

- Visualization Design: Create multiple visualizations (e.g., charts, graphs, maps) to explore different aspects of the data.

- Dashboard Creation: Build a cohesive dashboard with interactive elements (e.g., slicers, filters) to convey insights effectively.

- Presentation: Prepare a presentation to showcase the Power BI dashboard, highlighting key findings and actionable insights.

### Report Writing:

Write a report documenting the dashboard design process, data analysis techniques applied, and recommendations based on the insights gained.

## 6. Additional Resources for Week 6

Books

- “Power BI: The Ultimate Guide to Mastering Power BI for Business Intelligence and Data Visualization” by Matthew Adams

- “Mastering Microsoft Power BI: Expert Techniques for Effective Data Analytics and Business Intelligence” by Brett Powell

Articles

- “Getting Started with Power BI” by Microsoft Power BI Documentation

- “Advanced Data Modeling with Power BI” by SQLBI

# Week 7: Advanced Analytics Techniques

**Overview**

Week 7 delves into advanced analytics techniques, including statistical analysis and machine learning basics. Participants will learn to apply statistical methods, explore predictive analytics, and understand the fundamentals of machine learning using Python. These skills are essential for extracting deeper insights from data and making informed decisions.

**Learning Objectives**

- Master statistical analysis techniques for descriptive and inferential purposes.

- Understand the basics of predictive analytics and its applications in business contexts.

- Gain foundational knowledge of machine learning concepts and algorithms.

- Develop hands-on experience with Python libraries for data analysis and machine learning.

## 1. Statistical Analysis

### Descriptive Statistics:

- Measures of Central Tendency: Mean, median, mode.

- Measures of Dispersion: Variance, standard deviation, range.

- Data Visualization for Descriptive Analysis: Histograms, box plots, scatter plots.

### Inferential Statistics:

- Sampling Techniques: Random sampling, stratified sampling.

- Hypothesis Testing: Null and alternative hypotheses, p-values, significance levels.

- Statistical Tests: T-tests, ANOVA, Chi-square tests.

### Practical Exercise

- Hands-On Practice: Perform descriptive statistical analysis using Python libraries (e.g., NumPy, Pandas, Matplotlib). Calculate measures of central tendency and dispersion, visualize data distributions, and interpret results.

## 2. Machine Learning Basics

### Introduction to Machine Learning:

- Definition and Scope of Machine Learning: Supervised, unsupervised, and reinforcement learning.

- Applications of Machine Learning in Business: Predictive modeling, customer segmentation, recommendation systems.

### Supervised Learning Algorithms:

- Regression: Linear regression, logistic regression.

- Classification: Decision trees, random forests, support vector machines.

### Practical Exercise

- Hands-On Practice: Implement supervised learning algorithms using Python’s Scikit-learn library. Build regression and classification models, evaluate model performance using metrics like accuracy and RMSE, and interpret model outputs.

## 3. Machine Learning with Python

### Data Preparation and Preprocessing:

- Feature Engineering: Feature scaling, encoding categorical variables.

- Handling Missing Data: Imputation techniques (mean, median, mode).

### Model Evaluation and Validation:

- Cross-Validation: K-fold cross-validation, stratified cross-validation.

- Overfitting and Underfitting: Techniques to mitigate overfitting (e.g., regularization).

### Practical Exercise

- Hands-On Practice: Preprocess data for machine learning tasks, including feature engineering and handling missing values. Implement cross-validation techniques to assess model performance and optimize hyperparameters.

## 4. Machine Learning Project

Project Description

### Objective:

Apply machine learning concepts and techniques learned throughout the week to solve a real-world business problem. The project involves data preprocessing, model selection, training, evaluation, and interpretation of results.

### Dataset:

Use a provided dataset (e.g., customer churn data, sales forecasting data) for the project.

### Tasks:

- Data Exploration: Explore and visualize the dataset to understand distributions, correlations, and outliers.

- Model Selection: Choose appropriate machine learning algorithms based on the nature of the problem (classification or regression).

- Model Training: Train machine learning models on the dataset, using appropriate techniques for feature selection and model optimization.

- Evaluation and Interpretation: Evaluate model performance using relevant metrics (e.g., accuracy, precision, recall). Interpret results and draw actionable insights for business stakeholders.

### Report Writing:

Write a comprehensive report documenting the machine learning project, including data preprocessing steps, model selection rationale, evaluation metrics, and business recommendations based on the findings.

## 5. Additional Resources for Week 7

Books

- “Introduction to Machine Learning with Python: A Guide for Data Scientists” by Andreas C. Müller and Sarah Guido

- “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems” by Aurélien Géron

Articles

- “Statistical Analysis in Python” by DataCamp

- “Introduction to Machine Learning” by Towards Data Science

# Week 8: Capstone Project, Final Review & Feedback

**Overview:** Week 8 is dedicated to the Capstone Project, where participants apply all the skills and knowledge acquired throughout the course to solve a comprehensive data analytics and business intelligence challenge. The project culminates in a final presentation and review session, where participants showcase their work and receive feedback from peers and instructors.

**Learning Objectives**

- Integrate and apply data analytics and business intelligence techniques learned throughout the course.

- Demonstrate proficiency in data collection, cleaning, analysis, visualization, and interpretation.

- Develop effective communication skills for presenting insights and recommendations to stakeholders.

## Capstone Project

Project Description

### Objective:

The Capstone Project challenges students to tackle a complex business problem using data analytics and business intelligence techniques. Students will work with a real-world dataset (provided or sourced) to perform end-to-end data analysis, develop actionable insights, and present findings in a compelling manner.

### Tasks:

- Problem Definition: Define a clear business problem or research question that can be addressed through data analysis.

- Data Collection and Preparation: Gather and preprocess data from relevant sources, ensuring data quality and consistency.

- Exploratory Data Analysis (EDA): Explore the dataset to uncover patterns, trends, and relationships using descriptive statistics and visualizations.

- Advanced Analytics: Apply advanced analytics techniques such as statistical analysis, machine learning, or predictive modeling to derive insights.

- Data Visualization: Create informative and visually appealing dashboards or reports to communicate findings effectively.

- Presentation Preparation: Prepare a professional presentation summarizing the project scope, methodology, key findings, and actionable recommendations.

### Report Writing:

Write a detailed report documenting the entire project process, from problem definition to data analysis outcomes and recommendations. Include visualizations, statistical summaries, and interpretations to support findings.

## 2. Project Presentation and Review

Presentation Session

### Presentation Structure:

- Introduction: Introduce the project scope, objectives, and methodology.

- Data Analysis: Present key findings, insights, and data-driven recommendations.

- Visualization Showcase: Demonstrate dashboards or visualizations created during the project.

- Conclusion and Next Steps: Summarize findings, discuss implications, and propose next steps for further analysis or action.

### Peer and Instructor Feedback:

Receive constructive feedback from peers and instructors on project methodology, analysis techniques, presentation clarity, and overall project execution.

## 3. Final Review and Next Steps

Review Session

### Review of Key Concepts:

- Recap key concepts, techniques, and tools covered throughout the course.

- Discuss challenges faced during the Capstone Project and lessons learned.

### Career Advice and Next Steps:

- Provide guidance on career paths in data analytics and business intelligence.

- Discuss opportunities for further learning, certifications, or specialization areas in the field.

## 4. Additional Resources for Week 8

Books

- “Storytelling with Data: A Data Visualization Guide for Business Professionals” by Cole Nussbaumer Knaflic

- “Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking” by Foster Provost and Tom Fawcett

Articles

- “How to Give a Killer Presentation” by Harvard Business Review

- “Effective Data Storytelling: The Ultimate Guide” by Towards Data Science

# Resources

**Here are content links to resources structured weekly.**

Week 1: Introduction to Data Analytics and Business Intelligence

- Overview: Introduce Data Analytics and BI, their importance in business.

- Read More: [Data Analytics Overview](https://www.investopedia.com/terms/d/data-analytics.asp)

- Roles in Data Analytics and BI: Data Analyst, BI Analyst, Data Scientist, Data Engineer, BI Developer.

- Read More: [Roles in Data Analytics](https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-roles-and-responsibilities)

- Types of Data: Structured, unstructured, semi-structured data.

- Read More: [Types of Data](https://www.dataversity.net/the-importance-of-structured-data-unstructured-data-and-semi-structured-data/)

- Data Collection Methods: Surveys, transaction logs, social media.

- Read More: [Data Collection Methods](https://research-methodology.net/research-methods/data-collection/)

- Data Quality and Cleaning: Importance of data quality, cleaning techniques.

- Read More: [Data Cleaning](https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4)

- Tools and Software Overview: Excel, SQL, Python, Tableau, Power BI.

- Read More: [Popular Data Analysis Tools](https://www.datamation.com/big-data/best-data-analytics-tools/)

- Real-World Applications: Healthcare, retail, finance.

- Read More: [Applications of Data Analytics](https://www.dataversity.net/5-ways-data-analytics-is-transforming-industries/)

Week 2: Data Analysis with Excel

- Excel Basics: Functions and formulas, data manipulation, pivot tables.

- Read More: [Excel Functions](https://exceljet.net/excel-functions) | [Pivot Tables](https://www.excel-easy.com/data-analysis/pivot-tables.html)

- Advanced Excel Techniques: Solver, Analysis ToolPak, macros, and VBA.

- Read More: [Excel Solver](https://www.excel-easy.com/examples/solver.html) | [VBA for Beginners](https://www.excel-easy.com/vba.html)

- Data Visualization: Charts and graphs, dashboards.

- Read More: [Excel Charts](https://www.excel-easy.com/examples/charts.html) | [Excel Dashboards](https://chandoo.org/wp/excel-dashboards/)

- Practical Project: Analyze a dataset, create visualizations, build a dashboard.

- Project Link: [Excel Project Example](https://www.datapine.com/blog/excel-dashboards-templates-examples/)

Week 3: Introduction to SQL

- SQL Basics: Relational databases, basic queries (SELECT, FROM, WHERE).

- Read More: [SQL Basics](https://www.w3schools.com/sql/sql\_intro.asp)

- Advanced SQL Queries: JOINs, subqueries, aggregations.

- Read More: [SQL JOINs](https://www.sqlservertutorial.net/sql-server-basics/sql-server-joins/)

- Data Manipulation: INSERT, UPDATE, DELETE.

- Read More: [SQL Data Manipulation](https://www.tutorialspoint.com/sql/sql-insert-query.htm)

- Performance Tuning: Indexes, query optimization.

- Read More: [SQL Performance Tuning](https://www.datadoghq.com/blog/sql-server-performance-tuning/)

- Practical Project: Create a database, write queries, optimize performance.

- Project Link: [SQL Project Example](https://mode.com/sql-tutorial/introduction-to-sql/)

Week 4: Data Analysis with Python

- Python Basics: Introduction to Python, environment setup.

- Read More: [Python Basics](https://www.learnpython.org/) | [Setting Up Python Environment](https://realpython.com/python-environment/)

- Data Manipulation with Pandas: DataFrames, data import/export, data cleaning.

- Read More: [Pandas Tutorial](https://pandas.pydata.org/pandas-docs/stable/getting\_started/intro\_tutorials/index.html)

- Data Visualization with Matplotlib and Seaborn: Basic plots, advanced visualizations.

- Read More: [Matplotlib Tutorial](https://matplotlib.org/stable/tutorials/index.html) | [Seaborn Tutorial](https://seaborn.pydata.org/tutorial.html)

- Practical Project: Analyze a dataset, clean and transform data, create visualizations.

- Project Link: [Python Data Analysis Project](https://www.datacamp.com/community/tutorials/pandas-tutorial-dataframe-python)

Week 5: Business Intelligence with Tableau

- Introduction to Tableau: Overview, connecting to data sources.

- Read More: [Getting Started with Tableau](https://www.tableau.com/learn/training)

- Creating Visualizations: Basic charts, advanced visualizations.

- Read More: [Tableau Visualization Guide](https://www.tableau.com/learn/articles/data-visualization)

- Building Dashboards: Interactive dashboards, storytelling with data.

- Read More: [Tableau Dashboards](https://interworks.com/blog/rcurtis/2018/07/16/building-your-first-dashboard-in-tableau/)

- Practical Project: Develop a comprehensive dashboard with interactive elements.

- Project Link: [Tableau Project Example](https://www.datapine.com/blog/tableau-dashboard-examples/)

Week 6: Business Intelligence with Power BI

- Introduction to Power BI: Overview, connecting to data sources.

- Read More: [Power BI Getting Started](https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started)

- Creating Visualizations: Basic charts, advanced visualizations.

- Read More: [Power BI Visualizations](https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a)

- Building Dashboards: Interactive dashboards, data modeling, DAX.

- Read More: [Power BI Dashboards](https://docs.microsoft.com/en-us/power-bi/create-reports/service-dashboard-create)

- Practical Project: Create an interactive dashboard with various visualizations and data models.

- Project Link: [Power BI Project Example](https://www.mssqltips.com/sqlservertip/6832/power-bi-dashboard-examples/)

Week 7: Advanced Analytics Techniques

- Statistical Analysis: Descriptive and inferential statistics.

- Read More: [Statistics Tutorial](https://www.statisticshowto.com/)

- Machine Learning Basics: Introduction to machine learning, supervised and unsupervised learning.

- Read More: [Introduction to Machine Learning](https://www.coursera.org/learn/machine-learning)

- Practical Project: Apply statistical analysis and machine learning techniques to a dataset.

- Project Link: [Machine Learning Project Example](https://www.kaggle.com/learn/intro-to-machine-learning)

Week 8: Capstone Project

- Capstone Project: Integrate knowledge from previous weeks into a comprehensive project.

- Project Guide: [Capstone Project Guide](https://www.coursera.org/learn/data-science-project)

- Presentation Skills: Techniques for presenting data insights effectively.

- Read More: [Effective Data Presentation](https://hbr.org/2013/04/how-to-tell-a-story-with-data)

- Final Submission: Compile all work and present findings.

- Submission Guide: [Project Submission Guidelines](https://www.kdnuggets.com/2021/09/tips-final-year-data-science-project.html)