Data Analytics Lectureflow

Module1) Introduction to Data Analytics	2

- Difference Between Data Analysis, Data Science, AI, and Business Analysis (New): Understand the distinctions between these fields:
- Data Analysis, Data Science, AI, Business Analysis
- Steps of Data Analysis - Data collection, Data Cleaning, Data Exploration, Data Preprocessing, Data Analysis, Interpretation and Communication

Module 1) Introduction to Statistics (April 2023)	10

- Introduction to Statistics - Understand the fundamental concepts of statistics and its importance in data analysis
- Random Variable - Types of Random Variables - Discrete Random Variable, Continuous Random Variable
- Mean and Variance of Random Variable
- Continuous Distribution -Uniform Distribution • Normal Distribution Standard Normal Distribution • Exponential Distribution • Gamma Function • Chi-square Distribution • t Distribution • F Distribution
- Discrete Distributions: • Uniform Distribution • Bernoulli Distribution • Geometric Distribution • Poisson Distribution
- Confidence Intervals, Sampling, and Statistical Inference
- Hypothesis Testing Sample Size Calculator with Excel

Module 2) Introduction to Excel (April 2023)	14

- Excel Introduction: • Understand the basics of Excel, including worksheets, cells, rows, columns, and formulas
- Excel Functions: • Learn about commonly used Excel functions such as VLOOKUP, XLOOKUP, HLOOKUP, MID, OFFSET, and CHOOSE, and their applications in data manipulation and analysis.
- Text Handling: • Explore techniques for handling text data in Excel, including wrapping text, clearing formatting, and removing duplicates.
- Find and Replace: • Understand how to use the Find and Replace feature to search for and replace specific content within an Excel workbook
- Pivot Tables: • Learn how to create and work with Pivot Tables for summarizing and analyzing large datasets efficiently.
- Calculate Frequency Distribution in Excel
- Descriptive Statistics Using Excel
- Correlation Matrix Using Excel
- Introduction to Power Query: • Understand the basics of Power Query, including installing the Power Query Add-in, overview of the Query Editor, and importing data from various sources

- Importing Data: • Learn how to import data from web sources, text files, CSV files, and external Excel workbooks using Power Query
- Data Manipulation - Explore advanced data manipulation techniques in Power Query, including appending Excel tables, merging tables or queries, combining files from folders, and getting a list of file names from a folder.
- PQ Functions and M Language - Understand the useful text functions available in Power Query, creating IF, OR, and IF AND functions, overview of the M language, inserting comments in M code, and converting queries to functions
- VBA (Visual Basic for Applications) • Introduction to VBA and its role in automating Excel tasks. • Writing and executing VBA macros to perform customized actions in Excel
- Macros • Understanding macros and their significance in Excel automation. • Recording and editing macros to streamline repetitive tasks in Excel.
- Building Excel Dashboards: • Learn how to create interactive and visually appealing dashboards in Excel using Pivot Tables, Pivot Charts, slicers, and other features.
- Overview of the Query Editor
- Import Data from Web
- Import Data from Text Files
- Import Data from CSV Files
- Import Data from an External Excel Workbook
- Import Data from Current Excel Workbook
- Append Excel Tables in the Same Workbook , Merge Different Tables or Queries Append Tables from Different Workbooks into One Table, Combine Excel Files from a folder ,Combine CSV files from a Folder ,Get a List of File Names from a Folder
- Overview of PQ Functions, Useful Text Functions , Creating IF OR and IF AND functions
- Overview of M Language , Inserting Comments in M Code, Convert Query to Function Extract Data from a Table based on Selection Refreshing Queries , Get Files Names from Folder Based on Selection
- Building Excel Dashboards

Module 3) Applied Statistics in Excel (April 2023)	**5**

- Calculate Frequency Distribution in Excel
- Correlation Matrix Using Excel
- Descriptive Statistics Using Excel
- Normal Distribution Using Excel
- Sample Size Calculator with Excel
- Compute Correlation Matrix
- Compute partial correlation matrix

Module 4) Working with Database using SQL (April 2023)	**10**

- Introduction to SQL: • Understand what SQL is, why it's used, and its syntax for querying and managing databases.

- SQL Basic Data Types: • Learn about different data types in SQL, including string, numeric, date, and time data types.
- SQL Operators: • Explore SQL operators, including arithmetic, multiplication, division, modulus, logical, and set operators.
- SELECT Statement: • Understand the SELECT statement and its variations, including SELECT with WHERE, GROUP BY, and HAVING clauses.
- Aggregation Functions: • Learn about aggregation functions in SQL, including COUNT, SUM, and DISTINCT, for summarizing data
- LIMIT Clause: • Understand how to use the LIMIT clause in MySQL to restrict the number of rows returned by a query.
- SELECT statement, SELECT Statement with WHERE clause, SQL SELECT Statement with GROUP BY clause, SQL SELECT Statement with HAVING clause
- SELECT AS: • Learn how to use the AS keyword to alias column names or expressions in SQL queries.
- Joins: • Explore different types of SQL joins, including INNER JOIN, OUTER JOIN, LEFT JOIN, and FULL JOIN, for combining data from multiple tables.
- ORDER BY Clause:Understand how to use the ORDER BY clause to sort query results based on one or more columns.
- Advanced SQL Queries: • Dive into advanced SQL query techniques, such as selecting data from multiple tables, working with dates, handling NULL values, and using logical operators like AND and OR
- Stored Procedures and Views: • Creating and managing stored procedures and views in a database.
- Triggers: • Implementing triggers to automate database actions.
- Normalization: • Applying normalization techniques to eliminate data redundancy and improve database efficiency
- Importing/Exporting Data from Excel: • Exporting database data to Excel for analysis and reporting purposes
- Entity-Relationship (ER) Modelling: • Understanding ER modelling principles for designing database schemas.

Module 5) Creating Dashboard with Visualization Tool (april 2023)	**15**

- Introduction To Power BI , What is Power Bi , Why is Power BI and Power BI Installation and set up
- Power BI Installation and set up , Understanding Power BI Dashboard
- Components Of Power BI
- Power Query (ETL tool) Overview of the Query Editor , Overview of PQ Functions,
- Power Pivot Table
- Power View (Visualization Charts)
- Power BI Services
- Power BI Report View ,Model View , Power BI Table View
- How to make Relations in Two or more tables in Power BI
- Power BI Basic Power Charts (1-Column Chart 2-Stacked Chart 3-Pie Chart 4-Funnel Chart 5-Ribbon Chart
- Types of Data connection power BI

- Format Tools in Power BI for Charts and Visualization
- Create Tables in Power BI
- Data Analysis Expressions DAX Baisc 3 to 4 Examples
- Useful Text Functions
- Creating IF OR and IF AND functions
- Overview of M Language
- How to Change Background in Power BI Map
- How to Create a Map in Power BI
- Subtotal & Total in Matrix
- Cards & Filters in Power BI
- Slicers in Power BI
- Creating Dashboard with POWER BI
- Power BI Dashboard 2

Module 7) DA - Introduction to Python	**14**

- Why Python?, Features of Python Programming, Style Installation, Print Function, Comments
- Variable and data types
- Operators in python
- Arithmetic, Assignment, Logical, Comparison, Identity , Membership
- collections
- List, Tuple, Set, Dictionary
- Conditional Statements
- If, If-else, If-elif-else, Nested If-else
- Looping Statements
- for loop, while Loop, Nested loops, Range Function
- Control Statements
- break , Continue, pass
- Functions
- Definition, Types of Function, Defining a Function, Calling a Function, Function Arguments, Lambda function
- Scope Of Variables
- Global, Local
- Modules
- Introduction, How to import?, Math module, Random Module, Packages
- Input - Output
- Reading Input from Keyboard, Printing Output
- Files and Exceptions Handling
- File Operations: Opening and Closing, Read and Writing , Exceptions: try except finally
- OOPS Concepts
- Class, Objects, Inheritance, Polymorphism, Overloading

Module 8) DA- Working with NumPy (python)	**4**

- Difference Between EDA, AI, ML, and DL
- Math Refresher
- Array Creation: • Learn to create 1D arrays and multi-dimensional arrays using NumPy's array() and ndarray() functions.
- NumPy Functions: • Explore functions like zeros(), ones(), arange(), linspace(), eye(), etc., for array creation and manipulation.
- Array Attributes: • Understand attributes like shape, size, data type, and dimensionality using ndarray.shape, ndarray.size, ndarray.dtype, etc.
- Reshaping and Raveling: • Learn to reshape arrays with reshape() and flatten arrays with ravel().
- Arithmetic Operations: • Explore element-wise arithmetic operations like addition, subtraction, multiplication, and division.
- Broadcasting and Upcasting: • Understand broadcasting rules and upcasting for operations on arrays with different shapes.
- Conditional Operators: • Learn to use conditional operators for element-wise comparisons and boolean masking.
- Array Indexing and Slicing: • Understand indexing and slicing operations to access elements or subarrays.

Module 9) DA - Working with Pandas	**6**

- Handle Missing and Categorical Data, Outliers, Feature Engineering, Model Selection: • Learn techniques for handling missing data, categorical variables, outliers, and performing feature engineering in preparation for model selection
- Working on Series Objects: • Understand the Series data structure in Pandas and learn how to create, manipulate, and perform operations on Series objects.
- Indexing on Series: • Explore different indexing methods for accessing and selecting elements from Series objects based on labels, positions, or boolean conditions.
- Creating DataFrame: • Learn how to create DataFrame objects, which represent two-dimensional labelled data structures, from various data sources such as dictionaries, lists, or arrays.
- Mult indexing in DataFrame: • Understand the concept of multiindexing in DataFrame, which allows hierarchical indexing along multiple dimensions
- Dropping Level, Transposing: • Learn how to drop levels from multiindexing and transpose DataFrame objects to interchange rows and columns.
- Accessing Rows, Adding and Removing Columns: • Explore techniques for accessing and selecting rows from DataFrame objects and learn how to add and remove columns
- Querying and Sorting DataFrame: • Understand how to query and filter DataFrame objects using Boolean conditions and sort DataFrame based on column values.
- Operations on DataFrame: Learn how to perform various operations on DataFrame objects, including arithmetic operations, statistical calculations, and applying functions element-wise.
- Merging and Joining DataFrames (New): • Learn how to merge and join multiple DataFrame objects based on common columns or indices to combine and consolidate data from different sources
- Grouping and Aggregating Data (New): • Learn how to group data in a DataFrame based on one or more columns and perform aggregation operations such as sum, mean, count, etc., on the grouped data.

Module 10) DA – Visualization of Data with Matplotlib and Seaborn (Python)	4

- Creating and Customizing Line Charts using Matplotlib: • Learn how to create line charts to visualize trends and patterns in data using Matplotlib, and explore customization options to enhance the appearance of the charts
- Visualizing Relationships between Variables using Scatter Plots: • Understand how to create scatter plots to visualize the relationship between two or more variables, identify patterns, and detect correlations or clusters in the data.
- Studying Distributions of Variables using Histograms & Bar Charts: • Explore techniques for creating histograms and bar charts to visualize the distribution of a single variable or compare distributions across different categories.
- Visualizing Two-Dimensional Data using Heatmaps: • Learn how to create heatmaps to visualize two-dimensional data, such as correlation matrices or spatial data, using Matplotlib or Seaborn
- Box Plots and Violin Plots for Visualizing Distributions (New): • Explore how to create box plots and violin plots to visualize the distribution of data, including measures of central tendency, variability, and potential outliers.
- Pair Plots and Joint Plots for Exploring Relationships (New): • Learn how to create pair plots and joint plots to explore relationships between multiple variables simultaneously, including scatter plots with marginal histograms or kernel density estimates
- Exploratory Data Analysis (EDA) on a Dataset (New): • Learn the basics and importance of EDA. • Summarize dataset characteristics: size, shape, data types, missing values.

Module 11) DA - Data Scrapping With Python	4

- What is Web Scraping? • Overview of web scraping techniques and tools.
- Introduction to Data Scraping: • Definition and significance of data scraping. • Why data scraping is important for gathering information from the web
- Basic Web Scraping Techniques: • Static vs. dynamic web scraping. • Extracting data from HTML web pages.
- Python Libraries for Web Scraping:
- BeautifulSoup: • Overview of BeautifulSoup for parsing HTML and XML. • Extracting data from HTML tags and attributes
- Extracting Data with BeautifulSoup: • Installing and importing BeautifulSoup. • Parsing HTML content and navigating the parse tree. • Extracting specific elements and attributes from HTML.
- Data Cleaning and Processing: • Cleaning and preprocessing scraped data. • Handling missing values and formatting issues. • Converting scraped data into structured formats like CSV or JSON.
- Web Scraping Applications • Real-world applications of web scraping in data analysis, research, and business intelligence. • Case studies and examples of successful web scraping projects.