

Foundation Data Analytics – 6 Modules

Module – 1 : Excel Analytics

Module – 2 : Sql

Module – 3: PowerBi

Module – 4: Business Statistics & Visualization

Module – 5: Python for Analytics

Module – 6: Hands-On-Project (Analytics)

Specialization Data Science – Machine Learning & Artificial Intelligence 4 Modules

Module – 7 : Machine Learning

Module – 8 : Artificial Intelligence /
Deep Learning

Module – 9 : Deploying Models on Production

Module – 10 : Hands-On-Project (ML & AI)

1. Introduction to Data Analytics

- a. What is Data Analytics ? Its importance in business and decision-making process ?
- b. Skills required to become Data Analyst , Business Analyst, Business Intelligence Analyst, Marketing Analyst, Research Analyst, Reporting Analyst
- c. Difference between Data Analyst, Business Analyst, Business Intelligence Analyst
- d. Tools Covered: Excel, Sql, PowerBi, Python
- e. Career Opportunities, Roles, Jobs Description & Pay Packages for Data Analyst ?
- f. Difference between Data Analytics, Business Analytics & Data Science

2. Introduction to Data Science

- a. What is Data Science?
- b. Difference between Data Science, Artificial Intelligence, Machine Learning, Deep Learning, Generative AI
- c. Who are Data Scientists? Roles, Job Descriptions, Pay Scales ?
- d. Real-Time Data Science Project Life Cycle
- e. Data Science Project Architecture

Module - 1

Excel

2.

- a. Creating a New Workbook
- b. Navigating in Excel
- c. Moving the Cell Pointer
- d. Using Excel Menus
- e. Using Excel Toolbars: Hiding, Displaying, and Moving Toolbars
- f. Entering Values in a Worksheet and Selecting a Cell Range
- g. Previewing and Printing a Worksheet
- h. Switching Between Sheets in a Workbook
- i. Inserting and Deleting Worksheets
- j. Renaming and Moving Worksheets
- k. Hiding Columns, Rows and Sheets
- l. Splitting and Freezing a Window
- m. Editing, Clearing, and Replacing Cell Contents
- n. Cutting, Copying, and Pasting Cells
- o. Moving and Copying Cells with Drag and Drop
- p. Collecting and Pasting Multiple Items Using the Paste Special Command
- q. Inserting and Deleting Cells, Rows, and Columns Using Undo, Redo
- r. Checking Your Spelling Finding and Replacing Information
- s. Inserting Cell Comments
- t. Sort

u. Filter

3. **Formulas**

- a. Creating a Basic Formula
- b. IF, Nested Ifs
- c. IFNA, IF Error
- d. Vlookup
- e. HLookup
- f. XLookup
- g. Sumif & Sumifs
- h. Countif & Countifs
- i. Averageif & Averageifs
- j. Match & Index
- k. Text Functions
- l. Data Functions
- m. Editing & Copying Formulas Fixing Errors in Your Formulas
- n. Formulas with Several Operators and Cell Ranges

4. **Data Analysis & Pivot Tables**

- a. Creating a PivotTable
- b. Specifying the Data, a PivotTable Analyzes
- c. Changing a PivotTable's Calculation
- d. Selecting What Appears in a PivotTable
- e. Grouping Dates in a PivotTable
- f. PivotTable Formatting
- g. Pivot Chart
- h. Slicing
- i. Conditional-Formatting
- j. Data Validation
- k. Data Consolidation
- l. Text To Columns
- m. Remove Duplicates
- n. Record Macros

5. **Data Visualization Using EXCEL**

- a. Creating a Chart
- b. Chart Formatting
- c. Changing a Chart Type and Working with Pie Charts
- d. Adding Titles, Gridlines

6. Conclusion

- a. Dashboard Building Tips & Tricks
- b. Data Cleaning Demo
- c. Sample Project Demo
- d. Review and Recap
- e. Final Project: Call Centre Analysis & Apparel Manufacture Analysis

Module - 2 SQL

7. Introduction to SQL

- a. What is SQL
- b. Overview of the SQL Language
- c. SQL Database Structure
- d. Advantages of SQL

8. Data Definition Language (DDL)

- a. Introduction to DDL
- b. Creating Database and Tables
- c. Modifying and Deleting Tables

9. Data Manipulation Language (DML)

- a. Introduction to DML
- b. Retrieving Data using Select Statement
- c. Inserting, Updating and Deleting data

10. Basics

- a. Syntax of Queries
- b. Create Database, Tables
- c. Drop Database, Tables
- d. Insert Into, Update, Alter, Truncate, Delete
- e. Constraints, Primary Key, Foreign Key
- f. Select Statement
- g. Commenting Query
- h. Filtering, Grouping, Sorting and Limiting.

11. Intermediate

- a. Aggregation
- b. Group by
- c. Having
- d. Case, Coalesce
- e. Text Functions
- f. Date & Time Functions
- g. Order Of Execution & Order of Writing

12. Advanced SQL

- a. Joins
- b. Subqueries
- c. Common Table Expressions (CTE)
- d. Views
- e. Temporary Tables
- f. Stored Procedures and Triggers
- g. Advanced Query Optimization
- h. Window Functions

Module - 3 PowerBI

13. Introducing Power-BI Desktop

- a. Overview of Power-BI
- b. Power-BI Platform Types
- c. Why Power-BI
- d. Exploring Power-BI

- e. Power-BI vs Excel
- f. Three Components of Power-BI

14. Connecting & Shaping Data

- a. Data Connectors
- b. Connection Modes
- c. Intro to Power Query
- d. The Query Editor
- e. Data QA & Profiling
- f. Data Cleaning

15. Table Transformations

- a. Calendar Tools
- b. Combine & Append Queries
- c. M Language

16. Creating Data Model

- a. Data Modeling 101
- b. Normalization
- c. Facts & Dimensions
- d. Primary & Foreign Keys
- e. Cardinality
- f. Filter Flow
- g. Star & Snowflake Schemas
- h. Hierarchies
- i. Managing Relationships

17. Calculated Fields With DAX

- a. DAX 101
- b. Columns & Measures
- c. Row & Filter Context
- d. DAX Syntax
- e. Common Functions
- f. Text, Date Functions
- g. Calculate
- h. Iterators
- i. Time Intelligence

18. Visualising Data

- a. Data Viz Best Practices
- b. Formatting & Filtering

- c. Bookmarks
- d. Report Interactions
- e. User Roles
- f. Parameters
- g. Custom Tooltips
- h. Mobile Layouts
- i. Slicer panels
- j. Report navigation
- k. Row-level security
- l. Optimize reports
- m. Filtering
- n. Interaction Controls
- o. Top N Cards

19. Advance Topics

- a. Smart Narrative
- b. Q&A Visual
- c. Decomposition Tree
- d. Key Influencers
- e. Top Segments
- f. Optimize Ribbon
- g. Pause Visuals
- h. Optimization Presets
- i. Apply all Slicers
- j. Performance Analyzer
- k. External Tools

20. Hands-On Project

- a. Excel, Sql, PowerBI, Statistics, Python

3. Business Statistics

- a. Qualitative & Quantitative
- b. Scales of Measurement – Nominal / Ordinal / Interval / Ratio
- c. Types of Statistics
 - Descriptive Statistics
 - Inferential Statistics
- d. **Descriptive Statistics**
 - Measures of Central Tendency or Location
 - Mean
 - Median
 - Mode
 - Percentiles
 - Quartiles (Q1 – 25th Percentile, Q2 – 50th Percentile/Median, Q3 – 75th Percentile)
- e. Measures of Dispersion or Spread or Variability
 - Variance
 - Standard Deviation
 - Interquartile Range
 - Range
 - Coefficient of Variation
- f. Measures of Distribution – Shape
 - Distribution Shape
 - Histogram (Summarizing Quantitative Data)
 - Skewness – Left & Right Skewed
 - Normal Distribution / Normal Curve / Normal Probability Distribution
 - Bell Shaped Curved
 - Understanding Properties of Normal Distribution
 - Area Under the Curve for any Normal Distribution
 - Normal Probability Density Function
 - Standard Normal Distribution
 - Z – scores – Standard Normalization (computation)
- g. **Inferential Statistics**
 - Point Estimation
 - Population & Sample

- ✓ Population Parameter
- ✓ Sample Statistics
- ✓ Difference between Population Parameter & Sample Statistic?
- ✓ Why Sample when Population data available?
- ✓ Sample Size to select
- ✓ Determining sample size for a hypothesis test about a population mean
- ✓ Sampling Methods or Distributions
 - Simple Random Sample
 - Systematic Sampling
 - Stratified Sampling
- Hypothesis Testing and Decision Making
 - What is hypothetical question (more than one answer)
 - Developing Null and Alternate Hypothesis
 - Alternate Hypothesis as a Research Hypothesis and we need evidence to believe
 - Null Hypothesis is an assumption to be challenged
 - Confidence Level
 - Type I & Type II Errors
 - T-Test (comparison of means)
 - Comparison between two groups
 - ✓ One Sample t-test
 - One Sided/Tailed
 - Two Sided/Tailed
 - ✓ Two Sample t-test
 - One Sided/Tailed
 - Two Sided/Tailed
 - ✓ Paired t-test
 - Steps for Evaluating Hypothesis – One Sample & Two Samples
 - ✓ Compute confidence level with t-test
 - ✓ Finding out t-critical values from t-distribution table
 - ✓ Degrees of Freedom
 - ✓ Calculating p-value (probability of making Type I Error)
- h. Measures of Association between Two Continuous Variables
 - Covariance
 - Correlation Coefficient (Measure of Relationship)
 - Interpretation of Covariance & Correlation Coefficient
 - Cause & Effect relationship

- Difference between spurious correlation & correlation
- i. Probability
 - Sample Space
 - Event
 - How do we calculate Probability?
 - Classical approach
 - Relative frequency approach
 - Conditional Probability
 - Bayesian Conditional Probability
- j. Discrete Distributions
 - Ratio & Proportions
 - Odds & Odds Ratio
 - Conditional Odds Ratio
 - Discrete Distributions on Binary Variable
 - Binomial Distribution
 - Poisson Distribution
 - Uniform Distribution
 - Test of Independence
 - Chi-Square test for association (Measure of Relationship between two categorical/binary variables)
 - Chi-Square distribution
 - Degrees of Freedom in Chi-Square statistics
 - Chi-Square statistic formula
 - Steps for hypothesis testing using chi-square test
- k. ANOVA – Analysis of Variance
 - F-Test
 - Compute F-Statistics
 - F-Statistics Formula (between groups and within the group)
 - Compute F-Critical value
 - Degrees of Freedom
 - Confidence
 - Steps for hypothesis testing using F-statistics and F-critical values for more than two groups

4. Data Pre Processing

- a. Cleaning data with python
- b. Data Type Conversions
- c. Encoding categorical data

- d. Binning and Normalization
- e. Feature Scaling & Standardizing Data
- f. Handling missing values – Imputation

5. Exploratory Data Analysis (EDA) & Visualization

- a. Summary Statistics
- b. Charts & Graphs
 - One Dimensional Charts
 - Histogram / BarChart
 - Two Dimensional Charts
 - Bar Charts (Stack & Dodge)
 - Box Plots
 - Scatter Plots
 - Multi-Dimensional Plots
 - Fancy charts – Bubble Charts, Word Clouds
 - Outlier Detection & Management
 - Variable Selection / Variable Transformation

Module – 5

Python for Data Analytics

1. Introduction to Python

- a. Understanding the Reason for Python’s popularity
- b. Different IDE, Anaconda and Jupyter Notebook
- c. **Basic, Core & Advanced Python:**
 - Variables & Data Types in Python
 - Functions
 - Local & Global Statements
 - Data Structures: List, Tuple, Set, Dictionaries
 - Operators, Loops, Functions, Dictionaries
 - Numeric & String related functions
 - Object Oriented Programming (OOP)
- d. **Scientific Python:**
 - Numpy
 - Pandas
 - Matplotlib
 - Seaborn
 - Sci-kit Learn

Module - 7

Machine Learning

6. Machine Learning Introduction

- a. Introduction to Machine Learning
- b. Types of Machine Learning Algorithms:
 - Supervised Learning
 - Unsupervised Learning
- c. Selection of algorithm based on problem statement
- d. Difference between Regression & Classification
- e. Splitting data into
 - Train, Test & Validation dataset
- f. Measures of Performance
 - R-Square
 - RMSE for Regression
 - Confusion Matrix
 - Accuracy, Precision & Recall
 - F-1 Score
 - Sensitivity & Specificity
 - Roc & Auc

Supervised Learning

7. Linear Regression (Simple & Multiple)

- a. Introduction to Linear Regression
 - Simple Linear Regression
 - Multiple Linear Regression
- b. Model development, improvement and interpretation
- c. Sum of Least Squares
- d. Cost Function
- e. Feature Selection
- f. Model validation & diagnostics – Tests to validate assumptions
- g. Disadvantages of linear models
- h. Regularization Techniques
 - Ridge Regression (L2 Regularization)
 - Lasso Regression (L1 Regularization)
- i. Non-Linear Regression

8. Logistic Regression

- a. Variable Selection Methods
 - ✓ Forward, Backward & Stepwise
- b. Need for logistic regression
- c. Logit link function
- d. Maximum Likelihood Estimation
- e. Model Development And Validation
- f. Measures of Performance – Confusion Matrix / Roc Curve / RMSE / MAPE / Misclassification Rate / AUC / Precision & Recall / Decile / Lift & Gain Charts
- g. Advantages and disadvantages of logistic regression models

9. Decision Trees

- a. Decision Trees Types
 - Regression Tree
 - Classification Tree
- b. C5.0
- c. Classification and Regression Trees (CART)
- d. Rule based learning
- e. Construction of Rules
- f. Choosing variables for decision nodes / Feature Importance
- g. Over-fitting and Pruning
- h. Setting constraints – Number of Samples, Features & Trees / Size / Depth of Tree etc.
- i. Entropy / Gini / Chi-Square / Information Gain
- j. Advantages of tree based models

10. Bagging & Random Forest

- a. Bagging Technique and need of Bagging
- b. Re-sampling Methods with and without Replacement
- c. Hyper Parameter Tuning

11. Boosting

- a. Introduction to Boosting Techniques
- b. Types of Boosting
 - Gradient Boosting Machine (GBM)
 - Extreme Gradient Boosting Machine (XGBM)
 - AdaBoost
- c. Application of Boosting in Classification and Regression Models

12. Cross Validation

- a. Leave One Out Cross Validation
- b. K-Fold Cross Validation
- c. Bias & Variance

13. Forecasting – Time Series Analysis

- a. Components of Time Series
 - Trend / Seasonality / Randomness
- b. Additive & Multiplicative
- c. Simple moving averages
- d. Exponential smoothing
 - Dousie & Triple Exponential Smoothing
- e. Time series decomposition
- f. Holt Winters
- g. ARIMA & ARIMAX

14. Support Vector Machines (SVM)

- a. Maximum Margin Classifier
- b. Support Vector Classifier
- c. Kernels – Linear and Non Linear

UnSupervised Learning

15. Clustering (Segmentation)

- a. Hierarchical Clustering
- b. K-Means Clustering
- c. Cluster Profiling
- d. KNN

16. Dimensionality Reduction Techniques

- a. Principal Component Analysis (PCA)

17. Association Rules (APRIORI Algorithm)

18. Recommender Systems

- a. Collaborative Filtering
- b. Non-Negative Matrix Factorization
- c. Anomaly Detection

Module - 8

Artificial Intelligence - Deep Learning

19. Text Mining

- a. Text Analytics
- b. Unstructured Data
- c. Natural Language Processing (NLP)
- d. Cleaning Text Data: Tokenization, Pre-Processing
- e. Sentiment Analysis
- f. Text Classification & TFIDF
- g. POS Tagging
- h. Topic Modeling
- i. Feature Extraction

20. Probabilistic Methods Introduction

- a. Naïve Bayes
- b. Joint & Condition Probabilities
- c. Classification Using Naïve Bayes Approach

21. Artificial Neural Networks (ANN)

- a. Perceptron
- b. Single Layer Perceptron
- c. Multi Layered Perceptron
- d. Activation Functions & Types
 - ReLU / Sigmoid / Tanh
- e. Forward & Backward Propagation
- f. Training & Epochs
- g. Dropout & Drop connect
- h. Tools: TensorFlow & Keras

22. Convolution Neural Networks (CNN)

- a. CNN for Computer Vision
- b. CNN for Image Classification
- c. CNN
 - Convolution Layers
 - Pooling Layers

- Filters & Feature Maps
- Padding & Strides
- Training of CNN

23. Recurrent Neural Networks (RNN)

- a. Sequential Networks
- b. RNN Structure
- c. Introduction to LSTM
- d. Word Embedding
 - Glove
 - Word2Vec

Module - 10

Hands - On Projects (ML & AI-DL)

24. Data Science Project Life Cycle

- a. Proof of Concept (POC)

Key Takeaways

- a. Data Analytics Hands-On-Project delivered using Tools like: Excel ,Sql, PowerBi & Python
- b. Data Science Hands-On-Project delivered on Machine Learning & Deep Learning using Tools like: Python, Keras & Tensorflow
- c. Resume preparation
- d. All topics explained with REAL WORLD projects only
- e. Projects End to End Lifecycle explanation
- f. Mock Interviews & Test for Data Science Interview preparation
- g. Detailed assistance in Resume preparation. Special attention for experienced people on previous experience