CURRICULUM

FOUNDATION

Basics of MS Excel

- Intro to Excel
- Importing data
- Formatting in Excel
- Excel Formulas

Calculations using Excel

- Data Validation
- Calculations
- Lookup and Reference

Reporting using Excel

- Pivot Tables
- Charts
- What-if Analysis
- Intro to Macros

SQL

Basic SQL-1

- Introduction to SQL
- DDL Statements

Basic SQL-2

- DML Statements
- DQL Statements

Advanced SQL - Part 1

- AggregateFunctions
- Date functions Union
- Union All & Intersect Operators

Advanced SQL - Part 2

- Joins
- Views
- Indexes

Advanced SQL - Part 3

- Sub-Queries
- Exercise on SQL

CORE TRACK

Python Programming

Intro to Python

- Introduction to Python
- Variables
- Functions Python
- Operators

Programming Constructs in Python

- PythonFlow Controls
- Conditional Statements
- Loops

Python Objects

- Python Collection Objects
- Strings
- List
- Tuple
- Dictionary

List Comprehension + Functions

- List Comprehension
- User defined Functions
- Function Arguments
- Lambda Functions

Numpy

- Introduction to Numpy
- NumPy Array
- Creating NumPy Array
- Array Attributes
- Array Methods
- Array Indexing
- Slicing Arrays
- Array Operation
- Iteration through Arrays

Pandas

- Introduction to Pandas
- Pandas Series
- Creating Pandas Series
- Accessing Series
- Elements Filtering a Series
- Arithmetic Operations
- Series Ranking and Sorting
- Checking Null Values
- Concatenate a Series

Data Frame Manipulation

- Pandas Dataframe - Introduction
- Dataframe Creation
- Reading Data from Various Files
- Understanding Data Accessing
- Dataframe - elements using Indexing
- Dataframe Sorting
- Ranking in Dataframe
- Dataframe Concatenation
- Dataframe Joins
- Dataframe Merge Reshaping
- Dataframe Pivot Tables Cross
- Tables Dataframe Operations
- Checking Duplicates Dropping Rows and Columns Replacing Values
- Grouping Dataframe Missing Value
- Analysis & Treatment

Python Programming

Visualization – Part 1

- Visualization using Matplotlib
- Plot Styles & Settings
- Line Plot Multiline
- Plot Matplotlib
- Subplots
- Histogram
- Boxplot
- Pie Chart
- Scatter Plot

Visualization – Part 2

- Visualization using Seaborn
- Strip plot
- Distribution plot
- Joint plot
- Violin plot
- Swarm plot
- Pair plot
- Count plot
- Heatmap

EDA

- Summary Statistics
- Missing Value
- Treatment Dataframe analysis using groupby()
- Advanced-Data Explorations

Statistics & Probability

Introduction to Statistics

- Introduction to Statistics
- Random Variables
- Descriptive Statistics
- Measure of Central Tendency
- Measure of Dispersion
- Skewness and Kurtosis
- Covariance and Correlation

Probability Theory

- What is Probability?
- Events and Types of Events
- Sets in Probability
- Probability Basics using Python
- Conditional Probability
- Expectation and Variance

Probability Distributions

- Discrete Distributions -
  - Uniform
  - Bernoulli
  - Binomial
  - Poisson
- Continuous Distributions
  - Uniform
  - Normal
- Probability Distributions using Python

Statistics & Probability

Hypothesis Testing

- Introduction to HypothesisTesting
- Terminologies used in Hypothesis Testing
- Procedure for testing a Hypothesis
- Test for Population Mean
- Small Sample Tests
- Large Sample Tests
- Test for Normality

Statistical Tests

- One way ANOVA
- Assumptions
- ANOVA Hypothesis
- Post Hoc Test
- Chi-Square Test
- Chi-Square Test Steps
- Chi-Square Example

Machine Learning

Introduction to Machine Learning

- Machine Learning Modelling Flow
- Parametric and Non-parametric algorithms
- Types of Machine Learning

Linear Regression using OLS

- Introductionof LinearRegression
- Types of Linear Regression OLS Model
- Math behind Linear Regression
- Decom- position Variability
- Metrics to Evaluate Model
- Feature Scaling
- Feature Selection
- Regularisation Techniques

Project on Linear Regression

- Class Assessment on Linear Regression
- Project - Property Price Prediction
- Class Assessment on Linear Regression

Logistic Regression

- Intro to Logistic Regression
- Maximum Likelihood Estimation
- Performance Metrics

Model Tuning Techniques

- Performance Measure
- Bias-Variance Tradeoff
- Overfitting and Underfitting Problems
- Cross Validation

Project on Logistic Regression

- Project - VaccineUsagePrediction
- Home Assignment on Logistic Regression

Machine Learning

Decision Tree

- Introduction to Decision Tree
- Entropy
- Information Gain
- Greedy Algorithm
- Decision Tree: Regression
- Gini Index
- Tuning of Decision Tree-Pruning

Project on Decision Tree

- Project - Heart Disease Prediction

Random Forest

- Introduction to Random Forests
- Averaging
- Bagging
- Random Forest – Why & How?
- Feature Importance
- Advantages & Disadvantages

Project on Random Forest

- Project - Taxi Fare Prediction
- Class Assessment on Classification

K-means Clustering

- What is Clustering?
- Prerequisites
- Cluster Analysis
- K-means
- Implementation of K-means
- Pros and Cons of K-means
- Application of K-means

Project on K-means Clustering

- Project - E-commerceCustomerSegmentation

Hierarchical Clustering

- Introduction to Hierarchical Clustering
- Types of Hierarchical Clustering
- Dendrogram
- Pros and Cons of Hierarchical Clustering

Project on Hierarchical Clustering

- Project - TravelReview Segmentation
- Home Assignment on Clustering

Project on PCA

- Project - RealEstateData Analysis using PCA

Machine Learning

Principal Components Analysis

- Prerequisites
- Introduction to PCA
- Principal Component
- Implementation of PCA
- Case study
- Applications of PCA

Time Series Modelling

- Understand Time Series Data
- Visualizing Time Series Components
- Exponential Smoothing
- Holt's Model
- Holt-Winter's Model
- ARIMA

Project on Forecasting

- Project - ForecastingtheSalesofa Furniture Store

Data Visualisation with Tableau and Power BI

Tableau – Part 1

- Introduction to Tableau
- Data Connection
- Tableau Interface and Basic Chart Types
- Workingwith Metadata
- Visual Analytics

Tableau – Part 2

- Mapping
- Calculations
- Dashboard and Stories

Project Mortgage Analysis

Power BI – Part 1

- Introduction
- Interface
- Data Connections
- Data Transformation
- Advance Data Transformation

Project Stock Data Analysis

Capstone Project 1 Allocation – Based on ML

Deployment on Cloud

Cloud Basics

- WhatisCloud Computing?
- Advantages & Disadvantages of Cloud Computing
- Infrastructure as a Service (IAAS)
- Platform as a Service (PAAS)
- Software as a Service (SAAS) Major
- Players in Cloud Computing

AWS and SageMaker

- Introduction to AWS
- AWS Building Blocks
- AWS well architected framework – six pillars
- AWS Cloud Tour
- Build a Simple Regression Model

HTML, VS, Flask

- Steps to install VisualStudio
- Writing HTML Script
- Web Application through Flask

Deployment on Cloud

- What is Amazon EC2
- Getting started with EC2
- Step by step deployment

Advanced Track

Ensemble Modelling Techniques

- Ensemble Techniques
- What is Ensembling?
- Bootstrap Method
- Bagging
- Boosting
- XGBoost
- AdaBoost

Analytics in Healthcare/Finance

- Detailed project based on Classification

Association Rule Mining

- AssociationRules - Apriorialgo

Analytics in Retail

- MarketBasketAnalysis - Problem

Analytics in E-Commerce

- Some detailed project based on RFM model

Analytics in Marketing/Sales

- Some detailed project based on Regression