

---

## 1. Introduction of Course

1. What is Data Science
2. How it is different from Big Data and Data Analytics
3. Data Driven decision making
4. Purpose and Business problems
5. How Data Scientist work
6. Skills of a data scientist
7. Different sectors using Data science
8. Real World Applications
9. Future of AI and how the world is changing

## 2. Python Programming

- Why python for data analysis
- how to install Anaconda
- Running few simple programs using python
- Python objects
  - ✓ Lists
  - ✓ Strings
  - ✓ sets
  - ✓ file objects
  - ✓ Tuples
  - ✓ Dictionaries
  - ✓ Arrays, Data frames in python

## 3. Python Libraries

- ✓ NumPy
- ✓ SciPy
- ✓ Matplotlib
- ✓ Pandas
- ✓ Scikit Learn
- ✓ Seaborn
- ✓ os
- ✓ regular expressions
- Introduction to Series and Data frames
- Visualization on dataset using python
- Distribution analysis in python
- Box plot in python
- Comments in python
- Functions in python
- conversion functions

- **Math functions**
- **User defined Functions**
- **Parameters and arguments of functions**
- **Range functions python**
- **Recursive function and its examples**
- **Conditionals in python**
  - ✓ **If loop**
  - ✓ **elif**
  - ✓ **if elif else**
  - ✓ **Loops in python**
  - ✓ **for loop**
  - ✓ **while loop**
  
- **What is pandas?**
- **Benefits of using pandas**
- **Broadcasting in Python**
- **Array shape manipulations**
- **Data structures in pandas**
  - ✓ **Series**
  - ✓ **Data frame**
  - ✓ **Panel**
- **Various Data Frame Operations**
  - ✓ **Selection**
  - ✓ **Deletion etc.**
  - ✓ **"Grouping, Merging, and Reshaping of Data**
  - ✓ **Group by**
  - ✓ **Aggregate**
  - ✓ **Transform**
  - ✓ **Filtering**
  - ✓ **Merging and joining (concat and append)**
  - ✓ **Drop "**
- **Apply functions in pandas**
- **Accessing the objects in python by index**
- **Creating matrixes using NumPy**
- **Statistical operators using NumPy**

#### **4. Statistics / Maths**

- **Introduction to Statistics**
  - ✓ **Statistical and Non-Statistical Analysis**
  - ✓ **Major categories of statistics – Frequency and Bayesian**
  - ✓ **Difference between Statistics and Probabilities**
  - ✓ **Statistical terms**
  - ✓ **Difference between Descriptive Statistics and Inferential Statistics**
  - ✓ **Understanding of Population and Samples**
- **Descriptive Statistics**
- **Inferential Statistics**

- **Central Limit Theorem**
- **Types of variables**
  - ✓ **Nominal/Categorical**
  - ✓ **Ordinal**
  - ✓ **Interval/Ratio**
  - ✓ **Continuous, Time Series**
- **Central Tendency**
  - ✓ **Mean**
  - ✓ **Weighted mean**
  - ✓ **Trimmed mean/Truncated Mean**
  - ✓ **Interquartile mean**
  - ✓ **Trimmed Mean**
  - ✓ **Median**
  - ✓ **Mode**
- **Measure of Statistical dispersions**
  - ✓ **Variance and Bessel correction**
  - ✓ **Standard Deviation**
  - ✓ **Standard Error**
  - ✓ **Margin of Error**
  - ✓ **IQR**
  - ✓ **Range**
  - ✓ **Mean absolute difference**
  - ✓ **median absolute deviation**
  - ✓ **Coefficient of variance**
  - ✓ **Skewness**
  - ✓ **Kurtosis**
  - ✓ **Degrees of freedom**
  - ✓ **Law of Large Numbers**
  - ✓ **Confidence Level & Interval**
  - ✓ **P value and its interpretation**
  - ✓ **Correlation and auto correlation & correlation matrix**
  - ✓ **Correlation ratio**
  
- **Sampling Techniques**
- **Sampling errors**
- **Sample size estimation**
- **Point estimation & margin of error**
- **Multi Collinearity**
- **Co-variance and correlation**
- **P- value and critical value approach**
- **T-Distribution and T-Statistics**
- **Hypothesis testing's**
  - ✓ **What is Hypothesis Testing**
  - ✓ **Different types of Errors (Type I and Type II Errors)**
  - ✓ **Z-test**
  - ✓ **T-test**

- ✓ Chi-square test
- ✓ Power test, Beta test
- ✓ ANOVA (one way and two way)
- ✓ F-test & f score
- ✓ P-Value & Significance Level

## Probability

- Probability
- Venn diagram
- counting (permutation & combination)
- Expectation
- Conditional probability
- Joint Probability
- Marginal Probability
- Mutually exclusive events and Rules of Independence
- Rules of Probabilities
- Bayesian Network
- Random Variables and Expected Values

## 5. Machine Learning

- Supervised Learning
- Unsupervised Learning
- Difference between Classification and Regression
- Data pre-processing
  - What is data set.
  - What is training set
  - What is test set and need for test set
  - Missing values treatment
  - Expectation-Maximization technique for missing value
  - using Gradient
  - Feature scaling
  - Feature transformation
  - binning
  - one hot encoding
  - Feature engineering
  - Outliers' treatment
  - Bias and Variance trade off
  - Over fitting and Under fitting
- Exploratory Data analysis (EDA)
  - Univariate analysis
    - Continuous variable
    - Categorical variable
  - Bivariate Analysis
    - Continuous - Continuous
    - Categorical and Categorical
    - Categorical and Continuous "

- **Feature Engineering**
- **Variable transformation**
- **Variable /Feature Creation**
- **Project**
- **Supervised Regression Algorithms**
  - **Simple Linear Regression**
  - **Multiple Linear Regression**
  - **Ordinary Least Square (OLS)**
  - **Decision tree Regression**
  - **Random Forest Regression**
  - **GLM (Poisson regression, spline)**
  - **Support Vector Machines Regression**
  - **Error and Accuracy**
  - **Gradient Descent**
  - **Regularization Techniques**
  - **Maximum Likelihood estimation (MLE)**
  - **Probabilistic diagnosis of outliers**
  - **L2 and L1 Norms**
  - **Ridge Regression**
  - **Lasso Regression and Elastic Net**
  - **Project**
- **Supervised Classification Algorithms**
  - **Logistic regression classification**
  - **Multiclass Classification using Logistic Regression**
  - **Decision tree Classification**
  - **Random Forest classification**
  - **Support Vector Machines classification**
  - **What is Naïve Bayes theorem and the limitation**
  - **Naïve Bayes Classification**
  - **Ada boost/ Adaptive - Boosting Algorithm**
  - **GBM**
  - **Probability in Classification**
  - **Creating the log loss formula with entropy**
  - **SoftMax Function**
  - **MLE in classification**
  - **Understanding the Neural Networks**
  - **SVM**
  - **Gradient Boosting**
  - **XG Boost (Extreme Gradient Boosting)**
  - **Project**
- **Unsupervised Algorithms**
  - **K-means Clustering**
  - **Hierarchical clustering**
  - **Association Rule Mining**
  - **KNN Classifier**
  - **PCA**
  - **Project**

- **Model Evaluation Metrics**
  - **ROC Curves**
  - **Confusion matrix**
  - **Accuracy**
  - **Recall & Precision**
  - **Specificity & Sensitivity**
  - **Receiver Operating Characteristic (ROC) curve**
  - **Area Under Curve (AUC)**
  - **F1-Score**
  - **AIC & BIC Scores**
  - **R squared & Adjusted R squared**
  - **RMSE, MSE**
  
- **Model selection Techniques**
  - **Cross validation**
  - **Boot strap**
  - **Model selection using Statistical tests**
  - **Grid search**
  - **Evaluation Matrix**
  
- **Natural Language Processing (NLP)**
  - **What is NLP**
  - **Cleaning Text**
  - **Tokenization**
  - **Term Frequency (TF)**
  - **Term Frequency – Inverse Document Frequency (TF-IDF)**
  - **Document Term Matrix**

## 6. SQL

### Introduction to RDBMS

- What is Data Base
- Importance of Data Base
- Working RDBMS
- Why is Data Base need in Industries
- What is the use of Data Engineering in DB

### Sub Language Commands

- Data Definition Language (DDL)
- Data Retrieval Language (DRL)
- Data Manipulation Language (DML)
- Transaction Control Language (TCL)
- Database Security and Privileges (DCL)
- Oracle Pre Defined Datatypes
- DDL Commands
- Create, Alter (add, modify, rename, drop) Columns, Rename, truncate, drop
- DML- Insert, update, delete
- DQL-SELECT Statements using WHERE clause
- Comparison and Conditional Operators
- Arithmetic and Logical Operators
- Special Operators – IN (NOT IN), BETWEEN (NOT BETWEEN), LIKE (NOT LIKE), IS NULL (IS NOT NULL)
- Transactional commands
- Rollback
- Commit
- Save Point
- Working with joins
- inner join
- outer join
- cross join
- full join
- Working with Subquery
- Single-row subquery
- Multiple row subquery
- Correlated subquery
- Nested subquery

# I. DATA SCIENCE

## a. Intro to Neural Network & Deep Learning

- Intro
- Deep Learning Importance [Strength & Limitations]
- SP MLP
- Neural Network Overview
- Neural Network Representation
- Activation Function
- Loss Function
- Importance of Non-linear Activation Function
- Gradient Descent for Neural Network

### Parameter & Hyper parameter

- Train, Test & Validation Set
- Vanishing & Exploding Gradient
- Dropout
- Regularization-Optimization algo
- Learning Rate
- Tuning
- Softmax

### CNN

- CNN
- Deep Convolution Model
- Detection Algorithm
- Face Recognition

### RNN

- GAN
- RNN
- LSTM
- Bi Directional LSTM
- ANN

## b. NLP (Natural Language Processing)

- Introduction to NLP
- Stop Words
- Tokenization
- Stemming and Lemmatization
- Bag of Words Model
- Word Vectorizer
- TF-IDF
- POS Tagging
- Named Entity Recognition
- Introduction to Sequential data
- RNNs and its Mechanisms
- Vanishing & Exploding gradients in RNNS



- **LSTMS-Long short-term memory**
- **GRUS - Gated Recurrent Unit**
- **LSTMs Applications**
- **Time Series Analysis**
- **LSTMs with Attention Mechanism**
- **Neural Machine Translation**
- **Advanced Language Models: Transformers, BERT, XLNet**

### **c. Computer Vision**

- **Introduction to Convolutional Neural Networks**
- **Introduction to Images**
- **Convolution, Pooling, Padding & its Mechanisms**
- **Forward Propagation & Backpropagation for CNNs**
- **CNN architectures like AlexNet, VGGNet, InceptionNet & ResNet**
- **Transfer Learning**
- **Object Detection**
- **YOLO, R-CNN, SSD**
- **Semantic Segmentation**
- **U-Net**
- **Face Recognition using Siamese Networks**
- **Instance Segmentation**

### **d. Artificial Intelligence (AI)**

- **AI vs ML vs DL vs GenAI**
- **Supervised vs Unsupervised Learning**
- **Discriminative vs Generative AI**
- **A Brief Timeline of GenAI**
- **Basics of Generative Models**
- **Large Language Models**
- **Word Vectors**
- **Attention Mechanism**
- **Business Applications of ML, DL and GenAI**
- **Hands-on Bing Images and ChatGPT**

## **II. Data Analytics**

### **a. Excel / Adv Excel**

- **Introduction to Excel**
- **Columns & Rows\**
- **Functionality Using Ranges**
- **Excel Basic & Advanced Formulae's**
- **Data Analysis Using Excel**
- **Pivot & Power Pivot**
- **Spreadsheet Tools**
- **Data Cleaning**
- **Data Validation**
- **Data Visualization Using Excel**

## b. Power BI

- Power BI Introduction
- Data Visualization, Reporting
- Business Intelligence (BI), Traditional BI, Self-Serviced BI
- Cloud Based BI, On Premise BI
- Power BI Products
- Power BI Desktop (Power Query, Power Pivot, Power View)
- Flow of Work in Power BI Desktop
- Power BI Report Server, Power BI Service, Power BI Mobile Flow
- Power Query
- Data Transformation, Benefits of Data Transformation
- Shape or Transform Data using Power Query
- Overview of Power Query / Query Editor, Query Editor User Interface
- Ribbon (Home, Transform, Add Column, View Tabs)
- The Queries Pane, The Data View / Results Pane, The Query
- Settings Pane, FormulaBar
- Saving the Work
- Datatypes, Changing the Datatype of a Column Filter in Power Query
- Auto Filter/ Basic Filtering
- Filter a Column using Text Filters
- Filter a Column using Number Filters
- Filter a Column using Date Filters
- Filter Multiple Columns
- Remove Columns / Remove Other Columns
- Name / Rename a Column
- Reorder Columns or Sort Columns
- Add Column / Custom Column SplitColumns
- Merge Columns
- PIVOT, UNPIVOT Columns
- Transpose Columns
- Header Row or Use First Row as Headers
- Keep Top Rows, Keep Bottom Rows Keep Range of Rows
- Keep Duplicates, Keep Errors
- Remove Top Rows, Remove Bottom Rows, Remove Alternative Rows
- Remove Duplicates, Remove Blank Rows, Remove Errors
- Group Rows / Group By
- Data Modeling
- Data Modeling Introduction
- Relationship, Need of Relationship
- Relationship Types / Cardinality in General
- One-to-One, One-to-Many (or Many-to-One), Many-to-Many Auto
- Detect the relationship, Create a new relationship, Edit existing relationships
- Make Relationship Active or Inactive
- Delete a relationship
- DAX
- What is DAX, Calculated Column, Measures
- DAX Table and Column Name Syntax
- Creating Calculated Columns, Creating Measures Calculated
- Columns Vs Measures
- DAX Syntax & Operators

- **DAX Operators**
- **Types of Operators**
- **Arithmetic Operators, Comparison Operators, Text Concatenation Operator, Logical Operators**
- **DAX Functions Types**
- **Date and Time Functions**
- **Text Functions**
- **Logical Functions**
- **Math & Statistical Functions**
- **Filter Functions**
- **Time Intelligence Functions**
- **Date and Time Functions**
- **YEAR, MONTH, DAY**
- **WEEKDAY, WEEK**
- **NUMFORMAT (Text Function) à Month Name, Weekday Name**
- **DATE, TODAY, NOW**
- **HOUR, MINUTE, SECOND, TIME**
- **DATEDIFF, CALENDAR**
- **Creating Date Dimension Table**
- **Text Functions**
- **LEN, CONCATENATE (&)**
- **LEFT, RIGHT, MID UPPER, LOWER**
- **TRIM, SUBSTITUTE, BLANK**
- **Logical Functions**
- **IF TRUE, FALSE NOT, OR, IN, AND**
- **IFERROR SWITCH**
- **Math & Statistical Functions**
- **INT ROUND, ROUNDUP, ROUND DOWN**
- **DIVIDE**
- **EVEN, ODD**
- **POWER, SIGN**
- **SQRT, FACT**
- **SUM, SUMX**
- **MIN, MINX**
- **MAX, MAXX**
- **COUNT, COUNTX**
- **AVERAGE, AVERAGEX**
- **COUNTROWS, COUNT BLANK**
- **Filter Functions**
- **CALCULATE**
- **ALL RELATED**
- **Report View**
- **Report View User Interface**
- **Fields Pane, Visualizations pane, Ribbon, Views, Pages Tab, Canvas**
- **Visual**
- **Interactions**
- **Interaction Type (Filter, Highlight, None)**
- **Visual Interactions Default Behavior, Changing the Interaction**

- **Grouping and Binning Introduction**
- **Using grouping, Creating Groups on Text Columns**
- **Using binning, Creating Bins on Number Column and Date Columns**
- **Sorting Data in Visuals**
- **Changing the Sort Column, Changing the Sort Order**
- **Sort using column that is not used in the Visualization**
- **Sort using the Sort by Column button**
- **Hierarchy Introduction, Default Date Hierarchy**
- **Creating Hierarchy, Creating Custom Date Hierarchy**
- **Change Hierarchy Levels**
- **Drill-Up and Drill-Down Reports**
- **Data Actions, Drill Down, Drill Up, Show Next Level**
- **Visualizations**
- **Visualizing Data, Why Visualizations**
- **Visualization types, Create and Format Bar and Column Charts**
- **Create and Format Stacked Bar Chart Stacked Column Chart Create and**
- **Format Clustered Bar Chart, Clustered Column Chart**
- **Create and Format 100% Stacked Bar Chart, 100% Stacked Column**
- **Chart Create and Format Pie and Donut Charts**
- **Create and Format Scatter Charts**
- **Create and Format Table Visual, Matrix Visualization**
- **Line and Area Charts**
- **Create and Format Line Chart, Area Chart, Stacked Area Chart**
- **Combo Charts**
- **Create and Format Line and Stacked Column Chart, Line and**
- **Clustered Column Chart**
- **Create and Format Ribbon Chart, Waterfall Chart, Funnel Chart**
- **Power BI Service**
- **Power BI Service Introduction, Power BI Cloud Architecture**
- **Creating Power BI Service Account, SIGN IN to Power BI Service Account**
- **Publishing Reports to the Power BI service, Import / Getting the Report to PBI Service My**
- **Workspace / App Workspaces Tabs**
- **DATASETS, WORKBOOKS, REPORTS, DASHBOARDS**

### **c. Tableau**

- **Tableau Architecture**
- **Tableau introduction and overview Workbook, Stories,**
- **Dashboards and Worksheets**
- **Sources and Data Connection Types**
- **Graphs/Maps/Charts**
- **Define bar charts and line charts**
- **Define individual axis, blended axis and dual axis**
- **Define filled maps and symbol maps**
- **Define heat maps and tree maps**
- **Define pie charts**
- **Define Gantt charts.**
- **Filters and Sort**
- **filters - Filter on Measures and Dimension.**
- **Global filters**
- **Context filters**

- filtering at source
- Explain sort
- Sets and Groups
- Define Groups
- Define Sets
- Join Data using Sets
- Calculated Fields
- Introduction to Calculated Fields
- Define Table Calculations
- Define Date Calculations
- Define String Calculations
- Define Logical Functions
- Define Aggregation and Deaggregation
- Data Blending and Data Refresh
- Define Data blending
- Define Data Refresh
- Explain joins
- aspects of Data Refresh
- Actions
- Explain Actions
- Actions using - Filter, Highlight and URL
- Dashboards
- Explain dashboards
- How to create a dashboard
- Dashboard formatting
- Actions in Dash Boards
- How to perform navigation in dashboards.
- Parameters and Custom Fields
- Introduction to Parameters
- Define Calculated fields with Parameters
- Interactive features using Parameter driven fields
- Report Migration Process
- Introduction to Environments
- Report Migration Checklist
- Source Data and Data Refresh Needs

#### **d. R Programming**

- R Basics
- Numbers, Attributes
- Creating Vector
- Mixing Objects
- Explicit Coercion
- Formatting Data Values
- Matrices, List, Factors, Data Frames
- Missing Values, Names
- Reading and Writing Data
- Using Dput/DDump
- Interface to the Outside world
- Sub setting R objects
- Vectorized Operations
- Dates and Times

- **Managing Data Frames with the DPLYR**
- **package**
- **Control Structures**
- **Functions**
- **Lexical/Dynamic Scoping**
- **Loop Functions**
- **Debugging**
- **Data Visualization in R**
- **Storytelling with Data**
- **Principle tenets**
- **Elements of Data Visualization**
- **Infographics vs Data Visualization**
- **Data Visualization & Graphical functions in R**
- **Plotting Graphs**
- **Customizing Graphical Parameters to improvise the plots**
- **Various GUIs**
- **Spatial Analysis**
- **Other Visualization concepts**

**Additional Support – Interview Questions, Sample Resumes, Resume Building Assistance and Live Projects**

