

Data Science Curriculum & Data Analytics Curriculum

# 👀 Data Analytics and Data Science Introduction

Introduction to Data Analytics Introduction to Data Science

Applications and impact in various industries

💭 **Statistics**

### About Data

Variables in Statistics

Qualitative and Quantitative Population and Sampling Sampling Techniques

### Descriptive Statistics

Measure Of Central Tendency Measure of Dispersion

Normal Distribution Skewness and Kurtosis

### Inferential Statistic

Hypothesis Testing Methods

T-Test | Z-Test | Chi-Square Test ANOVA

### Probabilities

Probability basics

Common probability distributions Sampling and the Central Limit Theorem

🛫 **Python**

### Introduction to Python

**Basic**

Introduction to Python

Syntax Understanding in Python Variables in python

Operators in python

Built-in keywords and Methods in Python Data Types in Python

Data Structures in Python

### Intermediate

Flow Control Statements If Else Statements

Loops

Defining a Function

Lambda, Map, Reduce and Filter functions

### Advance

Classes and Objects (OOPs) File Handling and Exceptions

# 📈 Exploratory Data Analytics (EDA)

Introduction to Exploratory Data Analytics Libraries and Modules used for EDA

Import and Export Data.

Getting Data from Web Sources

# NUMPY

Introduction to NumPy and its importance in EDA Introduction to Arrayʼs

Numpy arrays: Creating, indexing, and slicing

**Data Manipulation with NumPy** Array reshaping and dimensions Concatenating and splitting arrays

Array filtering and boolean indexing Array sorting and searching

Working with missing data

### Numerical Computations

Basic statistical functions in NumPy

Linear algebra operations with NumPy Random number generation with NumPy

# Pandas

Introduction to Pandas and its role in EDA Series and DataFrame data structures

Data indexing, selection, and filtering

Basic operations on Series and DataFrames

### Data Cleaning and Preprocessing

Handling missing data in Pandas

Data deduplication and data type conversion Merging and joining DataFrames

Working with time series data

### Data Analysis with Pandas

Descriptive statistics and data summarization Grouping and aggregation in Pandas

Pivot tables with pandas

Working with text data in Pandas

# Data Visualization

Introduction to data visualization and its importance in EDA Basic Visualizations concepts (figure, axes, plots)

Understanding Line plots, scatter plots, and bar plots etc.

### Matplotlib

Basic Plotting with Matplotlib Introduction to Matplotlib Create Basic Plots

Customization of Plots Plotʼs Formatting

Adding Annotations

Creating Subplots and Multiple Plots Creating Subplots

Adding multiple plots

Grid and Complex Layouts

### Seaborn

Understanding Seaborn Library Default Statistical Plots

Line Plot, Scatter Plot, Bar Plot, Histograms, Box Plots, Violin Plots, Heatmaps etc.

Customized Plots Categorical Data Plotting

Distribution Plots dist Plots

Kernel Distribution estimators plots (KDE) Regression Plots

### Plotly

Concept of Plotly

Understanding Plotly with its benifits Plots using Plotlu

Line, Scatter, Bar plot, Pir Chart, Heatmaps, Area Charts etc.

3D Plots with Plotly 3D Scatter Plos, 3D Line charts

Annotations and Text Addition Geographical Plotting with Plotly

Visualization with Maps Geo Spatial Data Mapping

## CAPSTONE PROJECT FOR DATA ANALYTICS

There will be at least 10 Data Analytics Capstone Projects available according to the different industries:

BFSI - 2

FMCG - 2

Healthcare - 2

Automobile - 2

Entertainment - 2

Student Suppose to Complete any 3 capstone projects to move forward for Certification or Data Science Course.

Student will be provided with 2 weeks to complete 3 Capstone till we can resume with our add on courses like MySQL and Power BI/Tableau.

## 💾 My SQL (Sequel)

#### Introduction to Databases and SQL

What is a database?

Introduction to SQL (Structured Query Language) Key database concepts (tables, rows, columns) Setting up MySQL (or MySQL Workbench)

Basic SQL queries (SELECT, FROM, WHERE)

#### Data Retrieval with SQL

Inserting data into tables (INSERT INTO) Updating existing data (UPDATE)

Deleting data (DELETE FROM)

Creating and modifying tables (CREATE TABLE, ALTER TABLE)

#### Data Aggregation and Grouping

Aggregating data using functions (SUM, AVG, COUNT, etc.)

GROUP BY clause for data grouping HAVING clause for filtering grouped data

#### Data Joins and Relationships

Understanding table relationships (foreign keys) INNER JOIN, LEFT JOIN, RIGHT JOIN

Using subqueries for complex queries

💹 **Tableau**

**Introduction to Tableau and Data Sources** Overview of Tableau and its components Installing Tableau

Connecting to various data sources (Excel, databases, web data, etc.) Loading and transforming data in the Tableau

#### Data Modeling in Tableau

Introduction to data modeling concepts Creating relationships between tables Work with filter, parameters, and sets Measures and calculated columns

#### Data Visualization with Tableau

Creating different types of visualizations (bar charts, line charts, pie charts, etc.)

Customizing visual elements (colors, fonts, titles) Interactivity with slicers and filters

Drill-through and drill-down

#### Tableau for Reports and Dashboards

Building interactive reports and dashboards Layout and design considerations

Publishing and sharing reports on the Tableau Public

# Data Science

## Foundations of Data Science

### Introduction to Data Science

Data Science vs. Data Analytics The data science workflow

Tools and environments (Python, Jupyter) Concepts of Regression and Classification

### Data Acquisition and Cleaning

Data sources and collection

Data cleaning and preprocessing Handling missing data

## Predictive Analytics

### Regression

What is regression analysis?

Types of regression (linear, multiple, polynomial, etc.)

Use cases and applications of regression in data science

#### Linear Regression

Simple linear regression Multiple linear regression

Assumptions of linear regression Model interpretation and coefficients

Model evaluation metrics (R-squared, MSE, MAE)

#### Logistic Regression

Introduction to logistic regression

Logistic regression vs. linear regression Binary and multinomial logistic regression Odds ratio and log-odds interpretation Model evaluation for classification

Assessment for Linear and Logistic Regression

## Machine Learning

### Machine Learning Fundamentals

Understanding machine learning

Types of machine learning (supervised, unsupervised, reinforcement) Machine learning workflow

### Data Preprocessing for Machine Learning

Data cleaning and transformation Data scaling and normalization Handling missing data

### Model Evaluation and Validation

Cross-validation and train-test split

Evaluation metrics (accuracy, precision, recall, F1-score) Bias-variance trade-off

### Supervised Learning Algorithms Decision Tree

What is a decision tree? and How decision trees work. Decision Tree Splitting Criteria

Dealing with categorical features.

Advantages and Disadvantages of Using a Decision tree Algorithm.

### Random Forest

What is a Random Forest?

The concept of bagging and boosting

Random Forest Features and Hyperparameters Handling Imbalanced Data

Feature Importance and Interpretability

### XGBoost (Extreme Gradient Boosting)

Introduction to Gradient Boosting

Understanding boosting and the concept of weak learners.

How XGBoost improves upon traditional gradient boosting. XGBoost Hyperparameters.

### Naive Bayes Algorithm

Introduction to Naive Bayes Types of Naive Bayes

Probability Distributions Training and Classification

### SVM (Support Vector Machine)

Introduction to Support Vector Machines Support Vector Classification (SVC)

Support Vector Machines for Regression (SVR) Kernel Trick and Non-Linear SVM

Assessment for each Algorithm of Machine Learning

### Unsupervised Learning Algorithms

**K-Means Algorithm**

Introduction to Clustering K-Means Algorithm

Objective Function and Optimization Challenges and Limitations

### K-Nearest Neighbors (K-NN)

Introduction to K-NN Distance Metrics

Hyperparameter K

Decision Boundary and Majority Voting

Assessment for each Algorithm of Unsupervised Learning

💡

Capstone Projects for Machine Learning, where Data Analytics

Problems solving is Mandatory. Capstone Projects will be distributed accordingly (Each Student will get different Capstone Project — Not Similar — )

### Deep Learning

Introduction to Deep Learning and TensorFlow.

Convolutional Neural Networks (CNNs) with TensorFlow. Recurrent Neural Networks (RNNs) and Sequence Models. Advanced Deep Learning with TensorFlow and Keras.

### NLP (Natural Language Programming)

#### Introduction to NLP

What is NLP?

Applications of NLP

History and evolution of NLP Challenges and limitations in NLP

#### Text Preprocessing

Tokenization

Stop words removal Stemming and Lemmatization

Text cleaning and normalization

#### Text Representation

Bag of Words (BoW)

Term Frequency-Inverse Document Frequency (TF-IDF) Word embeddings (Word2Vec, GloVe)

Document-term matrices

#### Sentiment Analysis and Text Classification

Sentiment analysis techniques

Binary and multi-class classification Building a sentiment analysis model