Full Stack Data Scientist

1. Excel Fundamentals

- **Introduction to Excel**
  - Spreadsheet Environment
  - Absolute and Relative Reference
  - Data Sorting and Filtering
  - Basic and Advance Conditional Formatting
- **Excel Formulas and Functions**
  - Basics Data Manipulation
  - Basic Function
  - Basic and Advance Date Functions
  - Basic and Advance Logical Functions
  - Lookup and Reference Functions
- **Data Wrangling on real-world Dataset**
- **Excel Pivot Tables**
  - Intro to Tables
  - Pivot Charts
  - DAX functions
- **Excel Visualization**
  - Intro to Different charts
  - Dynamic charts
  - Views for a worksheet
- **Excel Dashboarding**
  - What is a Dashboard?
  - How to build a Dashboard?
  - Build a Dashboard
  - Dynamic Dashboard
  - Overview of the Examples of attractive Dashboard
- **VBA and Macros**
  - Difference between VBA, Macros, VB, VBS
  - Macros and VBE
  - VBA Object Model
  - Referencing range, workbook, worksheets with VBA
  - Working with Variables
  - Looping through Collection and Making Decision
  - Practical Exercise of VBA and Macros

2. Data Visualization using Tableau

- **Data Visualization Fundamentals**
  - Basics of Data Visualization
  - Exploratory v/s Explanatory Visualization

- Excel vs SPSS vs R vs Tableau

- **Introduction to Tableau**
  - Tableau Environment Overview
  - Tableau Desktop and Public Overview
  - Connecting with Data Sources
- **Managing, Organizing and Enhancing Data in Tableau**
  - Basic Data Wrangling in Tableau
  - Creating Sets, Calculated fields, Parameters and Joins
  - Use of Data Filtering and Data Blending
- **Interactive Graphing and Charts**
  - Types of Charts in Tableau
  - Waterfall Chart, Bump chart, Dual Axis chart and Scatterplots etc.
  - Usage and filtration of data with charts
- **Storytelling and Dashboarding in Tableau**
  - Designing Dashboard
  - Creating Stories in Tableau
  - Dashboard Layout and formatting

3. Introduction to SQL

- **Introduction to Databases**
  - What is Database
  - Introduction to MySQL and NoSQL
  - DDL v/s DML v/s DCL v/s TCL
  - Datatypes in SQL
- **Basics of SQL**
  - Basic SQL statements (SELECT, DELETE and UPDATE)
  - How to convert data into tables
  - COMMIT and ROLLBACK statements
- **Filtering Data using SQL**
  - Filter Data using the WHERE and ORDER BY Clause
  - Using of Filtering Operators – IN, NOT IN, IS NULL, BETWEEN
  - Regular Expression for Filtering
- **Functions in Database**
  - Basics of Function
  - Boolean Expressions and Concatenation
  - String Function
  - Grouping Function
- **Displaying Data from Multiple tables**
  - Introduction to Joins and its types
  - Using UNION, UNION ALL and EXPERT Clause
  - Views, Sequences and Indexes in SQL
- **Grouping Data and Computing Aggregates**

- Introduction to Grouping
- Using GROUP BY & HAVING

- **Subqueries and Nested queries in SQL**
  - Single-Row, Multiple-Row Subqueries
  - Subqueries with ANY and ALL Operators
  - Conditional Expressions using CASE Clause
  - Correlated Subqueries

4. Statistics and Probability

- **Fundamental of Statistics and its types**
  - Basics of Statistics
  - Types of Data
  - Descriptive Statistics
  - Inferential Statistics
- **Hypothesis Testing**
  - Null and Alternative Hypothesis
  - Type I and Type II error
  - p-value
- **Fundamentals of Probability**
  - Basics of Probability
  - Probability Theory
  - Probability Distribution
  - Random and Multivariate Variables
  - Bayes Theorem
  - Central Tendencies
  - Correlation coefficient
  - T-Test, F Test and Z Test
- **Linear Algebra**
  - Matrix Multiplication
  - CRISP Data framework
  - Factorization
- **Analysis of Variance and Covariance**
  - One-way Analysis of Variance
  - ANOVA Assumption
  - Two-Way Analysis of Variance
  - Analysis of Covariance

5. Python Basics

- **Introduction to Data Science Tools**
  - Intro to Jupyter Notebook
  - Notebook Type Interface for Data Science
  - Google Colab and Kaggle Kernels
  - Introduction to Kaggle and GitHub
  - Overview of the Projects in Python

- **Python Fundamentals**
  - Why Python and how it is different from the R programming language
  - Variables, Identifiers and Keywords in Python
  - Data Structures in Python
  - Strings, Array, Lists, Tuples, Set and Dictionaries
- **Python Conditionals and Loops**
  - If, Nested If, Indentations
  - Loops in Python
- **Basic Operation and Operator in Python**
  - Python Functions and Classes
  - Functions and its types
  - Classes in Python
  - Type Conversion
  - Function Arguments, Recursion and creating and Lambda Functions
- **Solving Basic Programs in Python**
  - Fibonacci Sequences
  - Prime Numbers
  - 0/1 Knapsack Problem

6. Advance Python for Data Scientist

- **Numpy for Data Engineers**
  - Introduction to NumPy Fundamentals
  - Arithmetic with NumPy Arrays
  - Broadcasting NumPy Arrays
  - Datasets and Boolean Indexing
  - NumPy Data Types
- **Pandas for Data Engineers**
  - Optimizing a DataFrame Memory Footprint
  - Processing in DataFrame in Chunks
  - Practice Optimizing DataFrames and Processing and Chunks
  - Augmenting Pandas with SQLite

7. Machine Learning

- **Introduction to Machine Learning**
  - What and why Machine Learning
  - Classification and Regression
  - Application of ML in real-world
- **Types of Machine Learning**
  - Supervised v/s Unsupervised v/s Reinforcement v/s Semi-Supervised Learning
  - Parametric and Non-Parametric ML Algorithm
  - Optimization Techniques
- **Supervised Learning**
  - Introduction to scikit learn Library

- Linear Models: Linear and Logistic Regression with Stochastic Gradient Descent
- Probability-Based Models: Naïve Bayes
- Tree-Based Model: Decision Tree and Random Forest
- Ensemble Methods: Bagging
- Proximity Based Models: K Nearest neighbor

- **Unsupervised Learning**
  - K-Means
  - Clustering
  - LDA and PCA
  - Hyper Parameter Tuning

8. Deep Learning and Introduction to Tensorflow

- Introduction to Deep learning
- Artificial neural network
- Gradient Descent and variants
- Backpropagation
- Intro to Tensorflow and keras
- A first artificial neural network with Sequential API

9. Deep Learning and Introduction to Tensorflow

- Introduction to Deep learning
- Artificial neural network
- Gradient Descent and variants
- Backpropagation
- Intro to Tensorflow and keras
- A first artificial neural network with Sequential API

10. Natural language Processing

- **Introduction to Natural Language Processing**
  - Intro to NLP
  - Bag of words, tf-idf
  - Sentiment analysis
  - POS Tagging
  - Named Entity Recognition
- **NLP with Deep Learning**
  - Deep Learning in NLP
  - Intro to RNN, LSTM
  - Word2vec
  - Language models
- **Advanced NLP**

- Transformer architecture
- Transfer Learning in NLP
- Intro to BERT
- Variants of BERT
- Intro to Hugging Face library

11. Deploy Model on Cloud

- **Introduction to Heroku**
  - Introduction to Streamlit
  - Intro to Heroku
  - Deploying the model on Heroku
  - Sentiment analysis

12. Big Data with Hadoop & Spark

- **Big Data with Hadoop**
  - Introduction to Big Data and Hadoop
  - Hadoop Architecture, Distributed Storage (HDFS)
  - Data Ingestion into Big Data Systems and ETL
  - Distributed Processing Map Reduce Framework and Pig
  - Apache Hive
  - NoSQL Database
- **Big Data with Spark**
  - Basics of Functional Programming and Scala
  - Apache Spark Next Generation
  - Spark Core Processing
  - Spark SQL – Processing Data Frames
  - Stream Processing Frameworks and Spark Streaming
- **Big Data with PySpark**
  - Introduction to PySpark
  - Resilient Distributed Datasets
  - Data frames and Transformations
  - Data Processing with Spark Data Frames
  - Sorting Techniques