**CERTIFIED DATA SCIENTIST PROFESSIONAL FULL – CDSPF**

**92 Hrs (4Hrs/day).**

**Training Program Curriculum**

1. **Python 3**

**Day 1**

* **Introduction to Computer science**
* **Environment Setup (Anaconda)**
* **Command Line**
* **Conda & pip package managers**

**Day 2**

* **Jupyter Notebook**
* **Input & Output**
* **Variables**
* **Data types**
  + Numbers & Math
  + Boolean & Comparison and Logic
  + Strings
  + Lists
  + Tuples
  + Sets
  + Dictionaries

**Day 3**

* **File Handling**
* **If Conditions**
* **For Loops**
* **Built-in functions & Operators (zip, enumerate, range, …)**
* **List Comprehensions**

**Day 4**

* **Functions**
* **Lambda Expressions**
* **Map, Filter, Reduce**
* **Modules & Packages**

**Day 5**

* **Git & GitHub**
* **Project #1 (Thanos.py)**

**Day 6**

* **Object-Oriented Programming (OOP)**
  + Classes & Objects
  + Data Hiding and Encapsulation
  + Inheritance
  + Exceptions
  + **Project #2 (Library System using OOP)**

1. **Mathematics For AI**

**Day 7 & 8 & 9**

* **Calculus**
  + Rate of Change
  + First order and second order derivatives
  + Partial Derivatives
  + Gradient Descent
  + Chain Rule
  + Integration
* **Linear Algebra**
  + Vector’s operations
  + Matrix operations
* **Probability**
  + Probability Basics
  + Combinatorics
  + Bayes Rules
  + Conditional Probability
* **Statistics**
  + Central Tendency
  + Measures of Dispersions
  + Data Visualization
  + Probability Density Function and Distributions
  + Normal Distributions
  + Standard Normal Distributions
  + Correlation and co-variance
  + Sample Distribution
  + Central Limit Theorem
  + Confidence Interval
  + Statistical Significance
  + Hypothesis Testing

1. **Web Scraping & Web Services**

**Day 10 & 11**

* **Network Topologies**
* **Internet and Web Servers**
* **HTTP Request/Response Cycle**
* **HTML**
* **CSS**
* **Scrapping Concept**
* **Beautiful Soap Library**
* **Web Services & JSON**
* **Project #3 (Weather Logs data collecting system)**
* **Project #4 (Employee’s data collecting using web services)**

1. **Databases & SQL**

**Day 12 & 13**

* **RDBMS.**
* **Tables, Columns and Data types**
* **How to design a database.**
* **One-To-Many & Many-To-Many Relationships.**
* **Project #5 (Design Database systems like Facebook, Souq, YouTube)**
* **SQL**
* **CRUD**
* **Selecting data**
* **Filtering data**
* **Ordering data**
* **Limiting data**
* **Aggregate Functions**
* **Joining tables**
* **Grouping data**
* **Subqueries**
* **Inserting new data**
* **Updating data**
* **Deleting data**
* **Python and SQLite**
* **DB Browser for SQLite**
* **Project #6 (Weather Logs data collecting system using database)**

1. **Exploratory Data Analysis with NumPy & Pandas**

**Day 14**

* **NumPy**
* **Pandas**
* **Project #7 (Analyze SF Salaries dataset from Kaggle)**
* **Project #8 (Analyze Ecommerce Purchase dataset from Kaggle)**

1. **Data Visualization with Matplotlib & Seaborn**

**Day 15**

* **Data Visualization**
* **Project #9 (Titanic Analysis Project)**
* **Project #10 (911 calls dataset from Kaggle analysis)**

1. **Data Preprocessing & ETL**

**Day 16**

* **Feature Engineering and Extraction**
  + Domain knowledge features
  + Date and Time features
  + String operations
  + Web Data
  + Geospatial features
  + Work with Text

**Day 17**

* **Feature Transformations**
  + Data Cleaning or Cleansing
  + Work with Missing data
  + Work with Categorical data
  + Detect and Handle Outliers
  + Deal with Imbalanced classes
  + Split data to Train and Test Sets
  + Feature Scaling
  + **Project #11 (Preprocess Loan data)**

1. **Machine Learning**

**Day 18 & 19 & 20**

* **Supervised Learning**
  + **Regression**
    - Simple Linear Regression
    - Multiple Linear Regression
    - Other Regression Methods
    - Evaluating Model Performance
    - **Project #12 (Ecommerce Expenses Prediction)**
    - **Project #13 (Kaggle Bike Demand Predictions)**
    - **Project #14 (Kaggle Black Friday Purchase Predictions)**
  + **Classification**
    - Logistic Regression
    - K-Nearest Neighbors (KNN)
    - Naive Bayes
    - Decision Trees
    - SVM
    - Ensemble Methods
    - Bagging & Boosting
    - Random Forests
    - XGBoost
    - Evaluating Model Performance
    - **Project #15 (Predict Loan Approval Problem)**
    - **Project #16 (Advertising Problem)**
    - **Project #17 (Sentiment Analysis Problem)**
  + **Intro Deep Learning**

**Day 21**

* **Unsupervised Learning**
  + **Clustering**
    - K-Means
    - Hierarchical Clustering
    - Density Based Clustering - DBSCAN
    - **Project #18 (Mall Problem)**
    - **Project #19 (University Problem)**
  + **Dimension Reduction**
    - PCA
* **Model Selection & Evaluation**
  + Cross Validation
  + Hyperparameter Tuning
    - Grid Search
    - Randomized Search

1. **Software Engineering & Model Deployment**

**Day 22 & 23**

* **What is Internet and Web Servers**
* **HTTP Request/Response Cycle**
* **HTML**
* **CSS**
* **Python as a backend language**
* **Flask**
* **Work with requests**
* **Work with templates**
* **Integrate machine learning model**
* **Deploy app to Heroku**
* **Capstone Project #20 (Bike demand predictor web application deployment on Heroku)**

**Final Graduation Day**

1. **Advance your Career**

* **Boost your Profile on Kaggle**
* **Build up your online presence**
  + Medium Blog
  + YouTube Channel
  + Contribute to Open Source Community on GitHub
* **Build your Resume**
* **LinkedIn and Networking**
* **Learn how to seek a job**