**Data Science and Machine Learning**

**Week 1 – Python Fundamentals**

**You will be introduced to the Python programming language. You will learn its basics, then implement control flow and functions.**

**Day 1 Presentation: Python Basics**

* Basics of Python
* Founded by Guido Van Rossum
* Monty python flying circus

Features of python

* Easy to learn : Simple learning curve – more like English language
* Multipurpose
* Open source
* Community and support
* Wide range of libraries available

**A character set** is a set of valid characters acceptable by a programming language in scripting. In this case, we are talking about the Python programming language. So, the Python character set is a valid set of characters recognized by the Python language. These are the characters we can use during writing a script in Python. Python supports all ASCII / Unicode characters that include:

* **Alphabets:**All capital (A-Z) and small (a-z) alphabets.
* **Digits:**All digits 0-9.
* **Special Symbols:**Python supports all kind of special symbols like, ” ‘ l ; : ! ~ @ # $ % ^ ` & \* ( ) \_ + – = { } [ ] \ .
* **White Spaces:**White spaces like tab space, blank space, newline, and carriage return.
* **Other:**All ASCII and UNICODE characters are supported by Python that constitutes the Python character set.
  + Tokens : A token is the smallest individual unit in a python program. All statements and instructions in a program are built with tokens.
  + Tokens are ( Keywords, Identifiers, Literals )
  + Keywords
  + Keywords are words that have some special meaning or significance in a programming language. They can’t be used as variable names, function names, or any other random purpose. They are used for their special features. In Python we have 33 keywords some of them are: *try, False, True, class, break, continue, and, as, assert, while, for, in, raise, except, or, not, if, elif, print, import,*etc.
  + Literals ( String, character, number {int, float, complex}, Special )
  + Literal collections ( List, Tuple, Set, Dict )
  + Identifiers

Identifiers are the names given to any variable, function, class, list, methods, etc. for their identification. Python is a case-sensitive language and it has some rules and regulations to name an identifier. Here are some rules to name an identifier:-

* As stated above, Python is case-sensitive. So case matters in naming identifiers. And hence **greeks**and **Greeks**are two different identifiers.
* Identifier starts with a capital letter (A-Z) , a small letter (a-z) or an underscore( \_ ). It can’t start with any other character.
* Except for letters and underscore, digits can also be a part of identifier but can’t be the first character of it.
* Any other special characters or whitespaces are strictly prohibited in an identifier.
* An identifier can’t be a keyword.
* **Operators ( w3 )**
* Arithmetic operators
* Assignment operators
* Comparison operators
* Logical operators ( and or not )
* Identity operators ( is , isnot )
* Membership operators ( in notin )
* Variables
* Python Data Types
  + Numbers
  + Strings
  + Lists : Dynamically sized array, Need not be homogeneous and they are mutable
  + Tuples : Collection of python objects separated by , and are immutable
  + Sets : A set is a collection which is *unordered*, *unchangeable\**, and *unindexed*.
  + **\* Note:** Set *items* are unchangeable, but you can remove items and add new items.
  + Dictionaries : Data structure with Key value pair within {}
* Here are the differences between List, Tuple, Set, and Dictionary in Python:



| **Parameters** | **List** | **Tuple** | **Set** | **Dictionary** |
| --- | --- | --- | --- | --- |
| Basics | A list is basically like a dynamically sized array. | The tuples refer to the collections of various objects of Python separated by commas between them. | The sets are an unordered collection of data types. These are mutable, iterable, and do not consist of any duplicate elements. | In Python, the dictionary refers to a collection (unordered) of various data types. We use these for storing data values such as maps, and unlike other data types capable of holding only one value in the form of an element, a dictionary can hold the key: value pair. |
| Homogeneity | A list refers to a data structure of a non-homogenous type that functions to store various elements in columns, multiple rows, and single rows. | A tuple also refers to a data structure of the non-homogenous type that functions to store various elements in columns, multiple rows, and single rows. | A set also refers to a data structure of the non-homogenous type, but it stores various elements in a single row. | A dictionary also refers to a data structure of the non-homogenous type that functions to store key-value pairs. |
| Representation | We can represent a List by [ ] | We can represent a Tuple by ( ) | We can represent a Set by { } | We can represent a Dictionary by {K:V, K:V, K:V,} |
| Duplicate elements | It allows various duplicate elements. | It allows various duplicate elements. | It does not allow any duplicate elements. | The keys are not at all duplicated. |
| Nested Among All | It can be utilized in a List. | It can be utilized in a Tuple. | It can be utilized in a Set. | It can be utilized in a Dictionary. |
| Example | [6, 7, 8, 9, 10] | (6, 7, 8, 9, 10) | {6, 7, 8, 9, 10} | {A:6, B:7, C:8} |
| Function for Creation | We can create a list using the **list()** function. | We can create a tuple using the **tuple()** function. | We can create a set using the **set()** function. | We can create a dictionary using the **dict()** function. |
| Mutation | It is mutable. It means that a user can make any changes to a list. | It is immutable. It means that a user can’t make any changes to a tuple. | It is mutable. It means that a user can make any changes to a set. | It is mutable, but the keys are not at all duplicated. |
| Order | It is ordered in nature. | It is ordered in nature. | It is unordered in nature. | It is ordered in nature. |
| Empty Elements | If we want to create an empty list, we use:  l=[] | If we want to create an empty tuple, we use:  t=() | If we want to create an empty set, we use:  a=set()  b=set(a) | If we want to create an empty dictionary, we use:  d={} |

**Operations on List Ref** : https://docs.python.org/3/tutorial/datastructures.html

Ref Code notebook for detailed explanation and code .

**Day 2 – Control Flow and Functions in Python**

* Control Flow in Python
  + Conditional Statements
  + Loops
  + Nested Loops
  + Loop Control Statements
* Functions in Python
  + Types of Arguments
  + Range Function
  + Lambda Function
* Packages and Modules in Python
* File Handling

**Week 2 – Array Manipulation Using NumPy**

**You will be introduced to the NumPy library in Python and perform array manipulation using this library.**

**Day 3 – NumPy Arrays and Functions**

* What is NumPy?
* Installing NumPy
* N-Dimensional Array
* Array Creation Routines
  + Arrays of ones and zeroes
  + From existing data
  + Using numerical ranges
* Basic Operations
* Arithmetic Operators
  + Single Dimensional Arrays
  + Multi-dimensional Arrays
* Matrix Product
* NumPy Functions
  + Universal Functions
  + Aggregate Functions
  + Logic Functions

**Day 4 – Array Manipulation and File Handling**

* Indexing
* Fancy Indexing
* Slicing
* Iterating in a NumPy Array
* Array Manipulation
  + Changing Array Shape
  + Transpose-like Operations
  + Joining Arrays
  + Splitting Arrays
* File Handling using NumPy
  + Loading & Saving Data in Binary File
  + Loading & Saving Data in Text File
* NumPy Case Study – Movie Rating System

**Week 3 – Data Pre-processing and Manipulation Using Pandas**

**You will be introduced to the Pandas library in Python and perform data pre-processing, cleaning, and transformation using this library.**

**Day 5 – Data Structures in Pandas**

* What is Pandas?
* Significance of Pandas in Python
* Installing pandas
* Pandas Data Structures
  + Pandas Series
  + Pandas DataFrame
* Importing and Exporting Data
  + Reading Data from CSV, Excel and JSON
  + Writing DataFrame Data to CSV and Excel
* Functionality of Pandas Series
* Functionality Of Pandas DataFrame
  + Selecting Columns and Rows of a DataFrame
  + Adding and Removing Columns
  + Removing Rows
  + Updating a cell’s value
  + Filtering a DataFrame

**Day 6 – Date Pre-processing Using Pandas**

* Combining Data using Pandas
  + Concatenate Rows
  + Concatenate Columns
  + Append Rows
  + Merge Operations
* Data Cleaning
  + Check for missing values
  + Handling missing values
  + Removing duplicates
  + Replacing values
* Grouping Data using Pandas
  + Groups and Aggregation
  + Group By

**Week 4 – Data Visualization Using Matplotlib and Seaborn**

**You will be introduced to the Matplotlib and Seaborn libraries in Python and learn how to visualize your data using these libraries.**

**Day 7 – Data Visualization Using Matplotlib**

* Introduction to Data Visualization
* The Matplotlib Library
* Types of Plots and Charts
  + Line plot
  + Bar plot
  + Histogram
  + Pie chart
  + Scatter plot
  + Boxplot
* Customizing Visualizations
* Saving Chats and Plots
* Grid and Subplots

**Day 8 – Data Visualization Using Seaborn and Plotly**

* Use Case 1: Global Superstore Sales Analysis
* Difference between Matplotlib and Seaborn
* Use Case 2: IBM Employee Attrition
* The Seaborn Library
* Types of Plots and Charts
* The Plotly Library
* Plots and Graphs in Plotly

**Week 5 – Probability and Statistics**

**You will be introduced to the concept of descriptive statistics and perform exploratory data analysis.**

**Day 9 – Descriptive Statistics**

* Data and Information
* Introduction to Statistics
* Population and Sample
* Statistical Analysis Division
* Importance of Statistical Analysis in Data Science
* Introduction to Data Types
  + Categorical Data
  + Numerical Data
* Descriptive Statistics
  + Measures of Central Tendency
  + Measures of Dispersion
  + Measures of Position
* Case Study – Super Market

**Day 10 – Exploratory Data Analysis**

* What is Exploratory Data Analysis?
* EDA Techniques
* EDA Classification
* Python Data Visualization Libraries
* Univariate Non-graphical EDA
  + Categorical Data
  + Numerical Data
* Univariate Graphical EDA
* Pie charts
* Bar Graphs
* Histograms
* Line graphs
* Box plots
* Dot plots
* Bivariate Non-graphical EDA
  + Cross-tabulation
  + Covariance and Correlation
* Bivariate Graphical EDA
  + Scatter plots
* Heat Maps
* Summarizing Graphical EDA Techniques

**Week 6 – Probability and Statistics**

**You will recall the concept of probability. You will also perform hypothesis testing after learning inferential statistics.**

**Day 11 – Probability**

* Introduction to Probability
* Define Probability
* Common terminology of Probability
* Events
  + Union
  + Intersection
  + Types of Events
* Rules of Probability
* Types of Probability
  + Marginal
  + Joint
  + Conditional
* Bayes Theorem

**Day 12 – Inferential Statistics**

* What is Inferential Statistics?
* Forms of Inferential Statistics
* Point and Interval Estimation
* Confidence Intervals
* Statistical Hypothesis
* Hypothesis Testing
* Decision Errors in Hypothesis Testing
* Decision Rules
* Statistical Test Interpretations
* Two-tailed and One-tailed Tests
  + Hypothesis Test for Single Population Mean: Z-test
  + Hypothesis Test for Single Population Mean: T-test
  + Hypothesis Test for Single Population Proportion
* ANOVA Test
* Chi-square Test

**Week 7 – Supervised Machine Learning**

**You will be introduced to Machine Learning, its types, and applications. You will also learn about Regression in Machine Learning.**

**Day 13 – Introduction to Machine Learning**

* What is Machine Learning?
* AI vs Machine Learning vs Deep Learning
* Significance of Machine Learning
* Applications of Machine Learning
* Types of Machine Learning
  + Supervised Learning
  + Unsupervised Learning
  + Reinforcement Learning
* Applications of Machine Learning
* Types of Machine Learning
  + Supervised Learning
  + Unsupervised Learning
  + Reinforcement Learning
* Data Pre-processing Techniques
  + Imputing Missing Values
  + Handling Categorical Values
  + Scaling the Data
  + Normalization
  + Feature Selection
* Train/Test split method

**Day 14 – Regression**

* What is Supervised Learning?
* Introduction to Regression
* What is Linear Regression?
* Assumptions of Linear Regression
* Types of Linear Regression
* OLS Regression Results Summary
* Calculation of R2
* Gradient Descent
* Regularization Techniques
  + Ridge Regression
  + Lasso Regression
  + Elastic net Regression

**Week 8 – Supervised Machine Learning**

**You will learn about the various Classification models in Machine Learning.**

**Day 15 – Classification**

* What is Classification?
* Classification vs Regression
* Types of classification
* Logistic Regression
  + What Is Logistic Regression?
  + Log Odds
  + Logistic Regression Cost Function
  + Maximum Likelihood
  + Evaluation Parameters
* Decision Tree
  + Common Terminologies
  + Decision Tree using CART Algorithm
  + Decision Tree using ID3 Algorithm
  + Attribute Selection
* Random Forest

**Day 16 – Mathematical and Bayesian Model**

* Naïve Bayes Classification
  + Revisiting Bayes’ Theorem
  + Likelihood
* K-Nearest Neighbors (KNN)
  + Distance Metric
  + Standardization (Normalization, Z-score)
  + Choosing K
* Support Vector Machines (SVM)
  + Linear SVM Classification
  + Non-Linear SVM Classification
  + SVM Regression
  + Kernel SVM
* Case Study – Credit Card Payment

**Week 9 – Unsupervised Machine Learning**

**You will learn about dimensionality and implement Unsupervised Learning using Clustering in Machine Learning.**

**Day 17 – Dimensionality Reduction**

* Curse of Dimensionality
* What is Dimensionality Reduction?
* Why Dimensionality Reduction?
* Feature Selection and Extraction
* Principal Component Analysis (PCA)
* Eigen Vector/Singular Vector
* Eigen Value/Singular Value
* Scree Plot
* Linear Discriminant Analysis (LDA)
* Other Dimensionality Techniques

**Day 18 – Clustering**

* What is Unsupervised Learning?
* What is Clustering?
* Types of Clustering
* Hierarchical Clustering
* Agglomerative Clustering
* Division Clustering
* K-Means Clustering
  + Euclidean Distance
  + Elbow Method
* Fuzzy C-Means Clustering
* DBSCAN Clustering

**Week 10 – Unsupervised Machine Learning**

**You will learn about Association Rule Mining and Time Series Analysis in Machine Learning.**

**Day 19 – Association Rule Mining**

* Association Rule Mining
  + Support
  + Confidence
  + Lift
* Apriori Algorithm
* Market Basket Analysis
* Recommendation Engine
  + User-Based Collaborative Filtering (UBCF)
  + Content-Based Filtering (CBF)

**Day 20 – Time Series Analysis**

* Time Series Analysis
* Components of Time Series
* Types of Data
  + Stationary Data
  + Non-Stationary Data
* Checks for Stationarity of Data
  + Augmented Dicky Fuller Test
* Convert Non-Stationary Data to Stationary Data
  + Differencing
  + Seasonal Differencing
  + Transformation
* Time Series Analysis Model
  + AR -> MA -> ARMA
  + ARIMA

**Week 11 – Model Evaluation and Optimization**

**You will learn how to evaluate and your Machine Learning model and optimize it to produce the best results.**

**Day 21 – Model Evaluation and Hyperparameter Tuning**

* Model Selection
* Resampling Techniques
  + K-Fold Cross-Validation
  + LOOCV
  + Bootstrapping
* Need for Model Evaluation
  + Underfit
  + Goodfit
  + Overfit
* Metrics for evaluating Regression Models
  + MAPE
  + RMSE
  + RMSLE
* Metrics for evaluating Classification Models
  + Confusion Matrix
  + ROC & AUC
  + F1 Score
* Hyperparameter Tuning
  + Manual Search
  + Grid Search
  + Random Search

**Day 22 – Model Boosting and Optimization**

* Ensemble Learning
* Bagging
* Boosting
  + AdaBoost
  + Gradient Boosting
  + XGBoost
  + CatBoost
* Model Optimization
  + Elements of Optimization
* Linear Programming
  + Examples
  + Applications
  + Formulating Optimization
  + Accelerated Gradient Methods
* Second-order Methods

**Week 12 – Neural Networks with TensorFlow 2.x**

**You will be introduced to Deep Learning and Neural Networks with TensorFlow.**

**Day 23 – Introduction to Neural Networks**

* Introduction to Deep Learning
* Artificial Neural Network
* The Perceptron
  + Forward Propagation
* Activation Functions
  + Sigmoid
  + Tanh
  + Rectified Linear Unit (ReLu)
  + Softmax
* Multilayer Perceptron
* Evaluating a Neural Network
  + Cost Function
  + Gradient Descent

**Day 24 – Deep Learning with TensorFlow**

* Back Propagation
* TensorFlow 2.x and Keras
* Project: MNIST Digit Classification using TensorFlow 2.x
  + Defining the model
  + Compiling the model
  + Training the model
  + Evaluating the model

**Week 13 – Deep Learning Using CNN**

**You will learn how Convolutional Neural Networks are used for image recognition in Deep Learning.**

**Day 25 – Deep Learning for Images**

* Limitations of Multilayer Perceptron
* Importance of CNN
* Layers in CNN
  + Convolutional
  + ReLU
  + Pooling
  + Fully connected

**Day 26 – CNN Implementation**

* Use Case – Predict Zero Using MNIST Dataset
* Step by step implementation of CNN
* Why is ReLU layer used?
* Image recognition

**Week 14 – Deep Learning Using RNN**

**You will learn how Recurrent Neural Networks are used for sequences in Deep Learning.**

**Day 27 – Deep Learning for Sequences**

* Issues with Feed Forward Network
* Recurrent Neural Network (RNN)
* Architecture of RNN
* Applications of RNN
* Vectors
* Training RNN

**Day 28 – Long Short-Term Memory**

* Drawback of Backpropagation
* Vanishing Gradient
* Exploding Gradient
* Long Short-Term Memory (LSTM)
* Gates
* Working of LSTM

**Week 15 – Natural Language Processing**

**You will learn about text pre-processing and classification using NLP.**

**Day 29 – Text Pre-processing**

* Need for Text Mining
* NLTK Library
* Applications of NLP
* Text Reading Procedures
* Tokenization
* Bigrams, Trigrams & Ngrams
* Stemming
* Lemmatization
* StopWords Removal
* POS Tagging
* Named Entity Recognition (NER)

**Day 30 – Text Classification**

* Bag of Words
* Count Vectorizer
* Term Frequency (TF)
* Inverse Document Frequency (IDF)
* Converting Text to Features and Labels
* Multinomial Naive Bayes Classifier
* Leveraging Confusion Matrix

**Week 16 – Data Connection and Visualization in Tableau**

**You will learn how to use Tableau as a tool for data visualization and perform various operations on your visualizations.**

**Day 31 – Data Visualization Using Tableau**

* Data Visualization
* Business Intelligence tools
* VizQL Technology
* Connect to data from File
* Connect to data from Database or server
* Basic Charts
  + Bar Chart
  + Line Chart
  + Pie Chart
* Chart Operations
  + Hierarchies
  + Data Granularity
  + Highlighting
  + Sorting
  + Grouping
  + Filtering

**Day 32 – Calculations in Tableau**

* Combining Data
  + Data Blending
  + Joins
* Built-in Functions
* Table calculations
* Parameters and Input Controls
* Level Of Detail (LOD) Calculations
  + Include
  + Exclude
  + Fixed

**Week 17 – Advanced Visualizations**

**You will learn how to use Tableau to create advanced visualizations and maps.**

**Day 33 – Forecasting and Clustering**

* Trend lines
* Reference lines
* Forecasting
* Clustering

**Day 34 – Maps in Tableau**

* Geographic Maps
  + Types of Maps
  + Spatial Files
  + Web Map Services

**Week 18 – Sharing Insights through Dashboards**

**You will learn how to build dashboards and create story points with Tableau.**

**Day 35 – Building Dashboards in Tableau**

* Using charts effectively
* Dashboards
  + Dashboard Objects
  + Building a Dashboard

**Day 36 – Dashboard Layouts and Story Points**

* Dashboard Layouts and Formatting
* Interactive Dashboards with Actions
* Dashboards for Devices
* Story Points
* Visual Best Practices
* Publish to Tableau Online