Data Analytics Workshop
Course Content

Techfest is the annual science and technology festival of IIT Bombay.
Following is the basic outline of the **in-person** workshop that would be happening at **Techfest in IIT Bombay**

About

The purpose of this document is to provide the details content outline of 2 days Data Analytics workshop conducted by Techobytes. The intent of this document provides organizers with a clear picture of the topics that we are going to cover in the workshop

Training details
Prerequisites: knowledge of basic computers.
Delivery Method: In-Person classroom session

Course structure:
Day 1 Session 1:

Getting started

- Getting Started:
- Installing Anaconda
- Introduction to Jupyter notebook
- Importing important libraries
- Popular keywords and shortcuts in jupyter

Basics of Python

- Intro to the Python Crash Course
- Basic Data Types
- Operators
- Variables
- Declare Variable
- Coding Exercise Solution: Declare Variables
- Built-in Functions
- Custom Functions
- String Methods
- Lists

- Creating Lists
- Coding Exercise Solution: Creating Lists
- Index Positions and Slicing
- Coding Exercise Solution: Index Positions and Slicing
- Dictionaries
- Creating Dictionaries

Day 1 Session 2
STATISTICAL INFERENCE

- Normal Distribution, Central Limit Theorem, and Confidence Intervals
- Skewness in data
- Understanding Syntax, Creating and Updating Columns
- Chaining, Functions, and .SD
- Fast Loops with set(), Keys, and Joins

ACCESSING/IMPORTING AND EXPORTING DATA USING PYTHON PACKAGES

- Importing Data from various sources (CSV, txt, excel, access, etc)
- Database Input (Connecting to the database)
- Viewing Data objects - subsetting, methods
- Exporting Data to various formats

DATA MANIPULATION – CLEANSING

- Cleansing Data with Python Programming
- Data Manipulation steps(Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc)
- Data manipulation tools(Operators, Functions, Packages, control structures, Loops, arrays etc)
- Scaling and Normalizing data
- Pre-processing and Formatting data
- Feature selection – Correlation, P Values, Multi-Collinearity, etc.

Day 2 Session 1
BASIC STATISTICS & IMPLEMENT
REGRESSION ANALYSIS

- Overview
- Introduction to Regression Analysis
- Types of Regression Analysis Models
- Linear Regression

- Model
- Model statistics
- Gradient Descent Algorithm
- Demo: Simple Linear Regression
- Demo: Regression Analysis with Multiple Variables
- Cross Validation
- Factor Analysis
- Fitting model and Predictions

CLASSIFICATION ANALYSIS

- Decision Tree Classification
- Entropy & Gini Index
- Classification and Regression Trees
- Decision Tree Statistics
- Decision Tree
- Demo: Decision Tree Classification
- Random Forest Classification
- Evaluating Classifier Models
- K-Fold Cross Validation

CLUSTERING

- Overview
- Introduction to Clustering
- Clustering Example
- Clustering Methods: Prototype Based Clustering
- Centroids and Means
- Euclidean Distance Formula
- Elbow Method – Picking values of K
- K-means Clustering

Day 2 Session 2
Visualization in Python

- Intro to Visualization Section
- Use the plot Method to Render a Line Chart
- Modifying Plot Aesthetics with matplotlib Templates
- Creating Bar Graphs to Show Counts
- Creating Pie Charts to Represent Proportions
- Visualization
- Creating Dashboards from available data