

# DataScience Training

Build your own predictive models in 45 days with zero prior knowledge



## Course Details

“Machine Learning gives computer the ability to learn without being explicitly programmed” ~I.Samuel

In this course, you will learn how to apply Data Science through seven pragmatic steps - Frame, Acquire, Refine, Transform, Explore, Model, and Insight - to any business problem. The focus will be to learn the principles through an applied case study and by actually coding in **Python** to solve this.

## Objective

- ▶ Learn how to employ statistical and machine learning algorithms to solve real life problems by working on real time Projects.
- ▶ Develop proficiency in using python and its libraries like Pandas, numPy, Seaborn,

## Project-1

**Sale Prediction**



In this project ,we will build a predictive model to find out the sales of each product at a particular store

## Project-2

**Predict Taxi Destination**



In this project, we will build a predictive framework that is able to infer the final destination of taxi rides based on their (initial) partial trajectories. The output of such a framework will be the final trip's destination employee's attributes change over time.

## Approach

- ▶ Interactive and live coding session
- ▶ Taught by Real time Practitioners



# Module-1(Python Basics)

## Welcome To The Course

- ▶ Introduction To DataScience
- ▶ Real Time UseCases Of DataScience
- ▶ Who is a DataScientist??
- ▶ Github Tutorial
- ▶ Skillsets needed for DataScientist
- ▶ 6 Steps to take in 3 Months for a complete transformation to DataScience from any other domain
- ▶ Machine Learning-Giving Computers The ability to learn from data
- ▶ Supervised vs Unsupervised
- ▶ DeepLearning vs Machine Learning
- ▶ Link to get Free Data to Practice?
- ▶ Some Great self Learning DataScience Resources(Books,Tutorials,Vedios,Papers)

## Python Fundamentals

Python Fundamentals begins with acquiring an in-depth knowledge of the Python programming language. By the end of the week, students will be expected to program intermediate level scripts in Python

- ▶ Software Installation
- ▶ Introduction To Python
- ▶ "Hello Python Program" in IDLE
- ▶ Jupyter Notebook Tutorial
- ▶ Spyder Tutorial
- ▶ Introduction to Python
- ▶ Variable,Operators,DataTypes
- ▶ If Else,For and While Loops
- ▶ Functions
- ▶ Lambda Expression
- ▶ Filter, Map,Reduce
- ▶ Taking input from keyboard
- ▶ **HANDS ON-**
- ▶ **INTERVIEW QUESTION DISCUSSION**



# Module-2(Python Advance)

## NumPy

- ▶ Create Arrays
- ▶ Array Item Selection and Indexing
- ▶ Array Mathematics
- ▶ Array Operation
- ▶ HANDS ON

## Pandas

- ▶ Introduction to Pandas
- ▶ Series
- ▶ Series indexing and Selection
- ▶ Series Operation
- ▶ Introduction to Pandas
- ▶ Data Frames
- ▶ Data Collection from csv,json,html,excel
- ▶ Data Merging,Concatenation,join
- ▶ Group By and Aggregate Function
- ▶ Order By
- ▶ Missing Value Treatment
- ▶ Outlier Detection and Removal
- ▶ Pandas builtin Data Visualisation
- ▶ HANDS ON
- ▶ INTERVIEW QUESTION DISCUSSION



# Module-3( Visualisation)

## Visualisation- matplotlib,seaborn

we'll begin curriculum focused on various data visualization techniques and how they can help us engage and learn from our data using [Matplotlib](#), [Seaborn](#), [ggplot](#)

- ▶ Line Plots
- ▶ Scatter Plots
- ▶ Pair Plots
- ▶ Histograms
- ▶ Heat Maps
- ▶ Bar Plots
- ▶ Count Plots
- ▶ Factor Plots
- ▶ Box Plots
- ▶ Violin Plots
- ▶ Swarm Plots
- ▶ Strip Plots
- ▶ Pandas Builtin Visualisation Library
- ▶ **HANDS ON**
- ▶ **INTERVIEW QUESTION DISCUSSION**

## Project-1

Prctatice , Practice and Practice!!!!!!! Implement what you have learnt so far by working in a real time Project.....

Pandas  
Numpy  
Seaborn  
MatplotLib



# Module-4 (Statistics)

## Statistics

This session is dedicated to creating a deep understanding of mathematical concepts we'll later see in topics like machine learning and statistical analysis. Contrary to the traditional mathematics course, students will learn statistics and linear algebra in a programmatic way to fit a problem's needs.

- ▶ Descriptive vs Inferential Statistics
- ▶ Mean, Median, Mode, Variance, Std. dev
- ▶ Central Limit Theorem
- ▶ Co-Variance
- ▶ Pearson's Product Moment Correlation
- ▶ R - Square
- ▶ Adjusted R-Square
- ▶ Spearman's. Rank order Coefficient
- ▶ Sample vs Population
- ▶ Standardizing Data(Z-score)
- ▶ Hypothesis Testing
- ▶ Normal Distribution
- ▶ Bias Variance Tradeoff
- ▶ Skewness
- ▶ P Value
- ▶ Z-test vs T-test
- ▶ The F distribution
- ▶ The chi-Square test of Independence
- ▶ Type-1 and Type-2 errors
- ▶ Annova
- ▶ HANDS ON
- ▶ INTERVIEW QUESTION DISCUSSION



# Module-5 (Intro to ML)

## Introduction to Machine Learning

- ▶ Introduction to Machine Learning
- ▶ Machine Learning Usecases
- ▶ Supervised vs Unsupervised vs Semi-Supervised
- ▶ Machine Learning process Workflow
- ▶ Training a model
- ▶ Validating results
- ▶ Overfitting vs Underfitting
- ▶ Ordinal vs Nominal data
- ▶ Structured vs unstructured vs semi-structured data
- ▶ Intro to scikitLearn
- ▶ **HANDS ON**



# Module-6 (Supervised)

## Regression

- ▶ Regression Vs Classification
- ▶ Linear regression
- ▶ Multivariate regression
- ▶ Polynomial regression
- ▶ Multi-Collinearity,
- ▶ Auto correlation
- ▶ Heteroscedascity
- ▶ [Hands On](#)

## Classification

- ▶ KNN
- ▶ Svm
- ▶ Decision Tree
- ▶ Random Forest
- ▶ Performance tuning of Random Forest
- ▶ Naive Bayse
- ▶ Overfitting Vs Underfitting
- ▶ [Hands On](#)

## Model Validation

- ▶ Classification Report
- ▶ Confusion Report
- ▶ ROC
- ▶ RMSE
- ▶ MSE
- ▶ Cross validation
- ▶ [Hands On](#)



## Module-7 (Unsupervised)

Clustering & PCA

- ▶ Kmeans
- ▶ How to choose number of K in KMeans
- ▶ Hands on
- ▶ PCA
- ▶ Hands on



## Module-8 (Ensemble)

Ensemble Methods

- ▶ What is Ensembling
- ▶ Types of Ensembling
- ▶ Bagging
- ▶ Boosting
- ▶ Stacking
- ▶ Random Forest
- ▶ Important Feature Extraction
- ▶ XGBoost
- ▶ HANDS ON





## Module-9 (NLP)

NLP

- ▶ Tokenizer
- ▶ Stop Word Removal
- ▶ Tf-idf
- ▶ Document similarity
- ▶ Word2vec Model
- ▶ t-SNE visualisation
- ▶ Sentiment Analysis
- ▶ **HANDS ON**



## Module-10 (Deep Learning)

Deep Learning

- ▶ Basic of Neural Network
- ▶ Type of NN
- ▶ Cost Function
- ▶ Tensorflow Basics
- ▶ Hands on Simple NN with Tensorflow
- ▶ Image classification using CNN
- ▶ **HANDS ON**