

**Course Name****Data Science****Course Objective**

1. To make the learner identify potential zones of uses of Data Science.
2. Providing experience of working with real time applications of Data Science and Data Analysis.
3. Make a learner easily land up to a job role of either Data Scientist, Data Analyst, Business Analyst in IT Industry.

**Course Overview**

The Course covers –

1. Statistical Analysis
2. Hypothesis Testing
3. Probabilistic Analysis
4. R Programming
5. Data Import/Export and Data Manipulation
6. Predictive Analysis
7. Regression and Classification
8. Time Series Forecasting
9. Anomaly Detection
10. NLP – Sentiment Analysis

**Course Outcome**

After completion of this course –

1. The learner will be able to land up in a job role related to Data Analyst and Data Science.
2. The Learner can also easily switch from existing job role with around of 20% hike from the current salary switch to any of the field where Data Science is being used.
3. The learner will become capable of handling any project relevant to Data Science in a proper way.

**Course Code (to be filled by TechTrunk Ventures)**

TTV/IND/00034

**Duration**

90 Hours for online Live Training

**Modules**

22 Modules (4 Hours Each for online Training)

**Prerequisite**

Basic Understanding of mathematics and programming

**Machine Requirement**

Windows Machine (Windows 7 or Above) /Linux Machine  
Only 64 Bit  
4 GB RAM (8 GB Recommended)

<b>Software used</b>	R 3.5 Python 3.x
<b>Software Free/Licensed</b>	FREE
<b>If licensed, Is demo version available</b>	FREE
<b>Download link</b>	R 3.5 can be downloaded from <a href="https://cran.r-project.org/bin/windows/base/R-3.5.1-win.exe">https://cran.r-project.org/bin/windows/base/R-3.5.1-win.exe</a>  R studio can be downloaded from <a href="https://download1.rstudio.org/RStudio-1.1.456.exe">https://download1.rstudio.org/RStudio-1.1.456.exe</a>  Anaconda for Python 3.x <a href="https://www.python.org/ftp/python/3.7.0/python-3.7.0.exe">https://www.python.org/ftp/python/3.7.0/python-3.7.0.exe</a>
<b>Any extra hardware other than PC required (If Yes kindly mention the list of hardware components required)</b>	NO
<b>Hands on</b>	80%
<b>Projects Covered</b>	5
<b>10 Possible Project (Number of projects covered will be the count mentioned in above)</b>	<ol style="list-style-type: none"> <li>1. Exploratory Data Analysis on USA Bachelors Dataset</li> <li>2. Exploratory Data Analysis on Hyderabad Air Traffic Dataset from data.gov.in</li> <li>3. Hypothesis testing on sales data, sales pattern extraction</li> <li>4. Exploratory Data Analysis on IPL Dataset</li> <li>5. Churn Prediction for an Enterprise</li> <li>6. Credit Card Fraud Detection</li> <li>7. Hypothesis testing on the female literacy and fertility dataset, working with correlation analysis, visualizing the data.</li> <li>8. Fetching data related to number of Pinterest accounts maintained by IS state Department and embassies and missions and performing Exploratory Data Analysis</li> <li>9. Social Media Data Analysis</li> <li>10. Predictive Analysis on Credit Default Dataset</li> </ol>
<b>Study Material</b>	<ol style="list-style-type: none"> <li>1. PPTs</li> <li>2. Practice Examples</li> </ol>

	<ol style="list-style-type: none"><li>3. Reading Material in softcopy</li><li>4. Project Codes</li></ol>
<b>Suggested relevant courses after taking this course:</b>	<ol style="list-style-type: none"><li>1. Artificial Intelligence and Machine Learning</li><li>2. Deep Learning</li><li>3. Application Development using Python</li></ol>
<b>Suggested Job Profile after taking this course:</b>	<ol style="list-style-type: none"><li>1. Data Scientist</li><li>2. Data Analyst</li><li>3. Business Analyst</li><li>4. Data Science Associate</li><li>5. Chief Data Scientist</li><li>6. BI Professional</li><li>7. R &amp; D Professional</li></ol>
<b>Any other relevant information</b>	<ol style="list-style-type: none"><li>1. Life time access to LMS</li><li>2. 24*7 Technical Support</li><li>3. Python course will be complementary</li></ol>

TechTrunk

Detailed Content:

<p><b>Module 1</b> Introduction to Data Science  Duration: 4 Hours</p>	<p>Data Science Data Scientists Data Mining Introduction Examples of Data Science Data Science case studies for Banking &amp; Finance, Manufacturing, Healthcare, Retail and Supply Chain AI v/s ML v/s DL and Data Science What makes a Data Scientist? Introduction to Statistics Statistical and Non-Statistical Analysis Some Common Terms Used in Statistics Hypothesis Testing and its interpretation Inferential Statistics Case Studies</p>
<p><b>Module 2</b> Statistical Analysis  Duration: 4 Hours</p>	<p>Types of variable Categorical and Continuous Data Ratio and Interval Nominal and Ordinal Data Grouped and Ungrouped Data Measure of Central Tendency – Mean, Mode and Median Percentile and Quartile Measure of Spread – IQR, Variance and Standard Deviation Empirical Rule Chebyshev's Theorem Z score Coefficient of Variation Kurtosis and Skewness</p>
<p>Assignment 1</p>	
<p><b>Module 3</b> Hypothesis Testing Analysing Continuous Data  Duration: 4 Hours</p>	<p><b>Analyzing the Continuous Data</b> Sampling Sample Size How to select sample P value Z Test One Sample T-Test Two Independent Samples Tests Paired T-test Wilcoxon Test</p>

<p><b>Module 4</b> Hypothesis Testing Analysing Categorical Data</p> <p>Duration: 4 Hours</p>	<p>ANOVA – for One independent Variable ANOVA – for 2 independent Variables Covariance Correlation</p> <p><b>Analyzing the categorical Data</b> Proportional Test Chi Square Test Fisher’s Exact Test</p>
<p><b>Module 5</b> Probabilistic Analysis</p> <p>Duration: 4 Hours</p>	<p><b>Probabilistic Analysis</b> Random Variable Continuous and Discrete Random Variables Expectation Mean and Variance of a Random Variable Probability basic rules Conditional Probability and Independence Bayes Theorem Central Limit Theorem Discrete Data Distribution</p> <ul style="list-style-type: none"> <li>- Binomial Distribution</li> <li>- Poison Distribution</li> <li>- Hypergeometric Distribution</li> <li>- Bernoulli Distribution</li> </ul> <p>Continuous Data Distribution</p> <ul style="list-style-type: none"> <li>- Uniform Distribution</li> <li>- Normal Distribution</li> <li>- Exponential Distribution</li> <li>- T- Distribution</li> <li>- F-Distribution</li> <li>- Chi-Square Distribution</li> </ul>
<p>Assignment 2</p>	
<p><b>Module 6</b> R Programming Introduction</p> <p>Duration: 4 Hours</p>	<p><b>What is R?</b> Installing R For Windows, Mac OS, and Linux R environment How to get help in R</p> <p><b>R Studio Overview</b> Understanding R data structure Variables in R Scalars</p>

	<p><b>Vectors</b>  Matrices and List  Data frames  Cbind,Rbind, attach and detach functions in R</p> <p><b>Factors</b>  Getting a subset of Data  Missing values  Converting between vector types</p>
<p><b>Module 7</b>  Data Import/ Export and  Data Manipulation using  R</p> <p>Duration: 4 Hours</p>	<p><b>Importing data</b>  Reading Tabular Data files  Reading CSV files  Importing data from excel  Loading and storing data with clipboard  Accessing database  Saving in R data  Loading R data objects  Writing data to file  Writing text and output from analyses to file</p> <p><b>Manipulating Data</b>  Selecting rows/observations  Rounding Number  Creating string from variable  Search and Replace a string or Number  Selecting columns/fields  Merging data  Relabeling the column names  Data sorting and Data aggregation  Finding and removing duplicate records</p>
<p><b>Module 8</b>  R Programming &amp;  Functions</p> <p>Duration: 4 Hours</p>	<p><b>Using functions in R</b>  Apply Function Family  Commonly used Mathematical Functions  Commonly used Summary Functions  Commonly used String Functions  User defined functions  local and global variable  Working with dates</p> <p><b>R Programming</b>  While loop  If loop  For loop</p>

### Assignment 3

<p><b>Module 9</b> Data Visualization using R Programming</p> <p>Duration: 4 Hours</p>	<p>Installing visualization packages Plotting graphs Line Chart and Scatter Plot Customizing the figure windows Labels, legends, titles, axes xlim, ylim, color and basic properties of graph Bar Charts and Dot Charts Box plot Histogram Pie graph Quick plots (qplot function) Building graphics by pieces (ggplot function) low level graphics functions Adding to plots and setting graphical parameters Exporting graphics to different formats Developing graphs</p>
<p><b>Module 10</b> Linear Regression</p> <p>Duration: 4 Hours</p>	<p>Regression Problem Analysis Mathematical modelling of Regression Model OLS method for Linear Regression Finding the coefficients and intercept Programming Process Flow Use cases Implementing a linear Regression model using R Bifurcate Data into Training / Testing Data set Build Model on Training Data Set Predict using Testing Data Set Validate the Model Performance Building simple Univariate Linear Regression Model</p> <p>Multivariate Regression Model Correlation Analysis – Analyzing the dependence of variables Apply Data Transformations L1 &amp; L2 Regularization Identify Multicollinearity in Data Treatment on Data Identify Heteroscedasticity Modelling of Data Variable Significance Identification Model Significance Test R<sup>2</sup>, MAPE, RMSE Project: Predictive Analysis using Linear Regression</p>

<p><b>Module 11</b> Logistic Regression</p> <p>Duration: 4 Hours</p>	<p>Classification Problem Analysis Variable and Model Significance Sigmoidal Function Maximum Likelihood Concept Null Vs Residual Deviance Cost Function Formation Mathematical Modelling</p> <p>Model Parameter Significance Evaluation Accuracy, recall, precision and F1 Score Drawing the ROC Curve Estimating the Classification Model Hit Ratio Isolating the Classifier for Optimum Results Project: Predictive Analysis using Logistic Regression</p>
--	---

#### Assignment 4

<p><b>Module 12</b> KNN and Decision Tree</p> <p>Duration: 4 Hours</p>	<p><b>K Nearest Neighbour</b> Understanding the KNN Distance metrics Case Study on KNN Example with Python</p> <p><b>Decision Trees</b> Forming Decision Tree Components of Decision Tree Mathematics of Decision Tree Variance – Decision Tree for Regression Gini Impurity, Chi Square – Decision Tree for Classification Decision Tree Evaluation</p>
--	--

<p><b>Module 13</b> Decision Tree and Random Forest</p> <p>Duration: 4 Hours</p>	<p><b>Decision Tree</b> Practical Examples &amp; Case Study Project: Financial Prediction with Decision Tree</p> <p><b>Random Forest</b> Bag of Trees Random Forest Mathematics Examples &amp; use cases using Random Forests Case Study: Bank Marketing Analysis Customer Churn Analysis</p>
--	---



## Assignment 5

<b>Module 14</b> Time Series Prediction  Duration: 4 Hours	Definition of Time Series Time Series Decomposition Simple Moving Average Method Weighted Moving Average Method Single Exponential Smoothing Method Double Exponential Smoothing Method Triple Exponential Smoothing Method Stationarity of Data ARIMA Models
<b>Module 15</b> Unsupervised Learning – Clustering  Duration: 4 Hours	<b>Clustering</b> Application of clustering DBSCAN Hierarchical Clustering K Means Clustering Use Cases for K Means Clustering Programming for K Means using Python Image Color Quantization using K Means Clustering Technique Customer segmentation using KMeans Cluster Size Optimization vs Definition Optimization Projects & Case Studies
<b>Module 16</b> Principal Component Analysis and Anomaly Detection  Duration: 4 Hours	<b>Principal Component Analysis</b> Dimensionality Reduction, Data Compression Curse of dimensionality Multicollinearity Factor Analysis Concept and Mathematical modelling Use Cases Programming using Python  <b>Anomaly Detection</b> Moving Average Filtering Mean, Standard Deviation Statistical approach for Anomaly Detection Isolation Forest for Anomaly Detection Hands on project on Anomaly Detection Do's and Don'ts for Anomaly Detection

## Assignment 6

### Module 17

Introduction to Python programming

Duration: 4 Hours

Introduction to Python Programming  
What is Python?  
Understanding the Spyder Integrated Development Environment (IDE)  
Python basics and string manipulation  
lists, tuples, dictionaries, variables  
Control Structure – If loop, For loop and while Loop  
Single line loops  
Writing user defined functions  
Object oriented programming with Python

## Assignment 7

### Module 18

Python for Data handling – numpy and Pandas

Duration: 4 Hours

#### Mathematical Computing with Numpy

NumPy Overview  
Properties, Purpose, and Types of ndarray  
Class and Attributes of ndarray Object  
Basic Operations: Concept and Examples  
Accessing Array Elements: Indexing, Slicing, Iteration, Indexing with Boolean Arrays  
Copy and Views  
Universal Functions (ufunc)  
Shape Manipulation & Broadcasting  
Linear Algebra using numpy  
Stacking and resizing the array

#### Introduction to Pandas

Data Structures  
Series, DataFrame & Panel  
DataFrame basic properties  
Importing excel sheets, csv files, executing sql queries  
Importing and exporting json files  
Selection of columns  
Filtering Dataframes  
Handling Missing Values  
Finding unique values and deleting duplicates  
Descriptive Analysis with pandas  
Creating new categorical features from continuous variable  
groupby operations  
Apply method  
String Manipulation

<p><b>Module 19</b> Data Visualization with matplotlib and seaborn</p> <p>Duration: 4 Hours</p>	<p><b>Introduction to Data Visualization</b></p> <p>Matplotlib Features: Line Properties Plot with (x, y) Controlling Line Patterns and Colors Set Axis, Labels, and Legend Properties Alpha and Annotation Multiple Plots Subplots Types of Plots and Seaborn Boxplots Distribution Plots Clustermaps Heatmaps Violin plots Swarmplots and countplots</p>
<p><b>Module 20</b> Natural Language Processing</p> <p>Duration: 4 Hours</p>	<p>Natural Language Processing &amp; Generation Semantic Analysis and Syntactic Analysis Text Cleaning and Preprocessing using Regex Using NLTK &amp; Textblob Basic Text data processing Tokenization, Stemming and Lemmatization Pos Tagging Tf-IDF, count vector and Word2vec Sentiment Analysis Using Google, Bing and IBM Speech to Text APIs Project: Streaming live tweets and Sentiment Analysis Wordcloud Project: Building an Email Classification Model Chatbots Building Chatbots using Dialog Flow and Facebook Messenger Facebook Messenger API Integration Project: Building a utility based chatbot</p>
<p><b>Assignment 8</b></p>	
<p><b>Module 21 &amp; 22</b> Projects</p> <p>Duration: 8 Hours</p>	<p>Working Final Project Do's and Don'ts with Data Science Productization of Data Science Application</p>

*Thank you for query*

For any query please feel free to reach us

[contact@techtrunk.in](mailto:contact@techtrunk.in), [www.techtrunk.in](http://www.techtrunk.in)

Call/WhatsApp: +91-9182275802

TechTrunk