

# FULLSTACK DATA SCIENCE ENGINEER WITH CLOUD, BIG DATA & DEVOPS



**ACCELERATE YOUR CAREER GEAR TO EXPERIENCE, EXPLORE AND EXCEL THE CUTTING EDGE TECHNOLOGIES OF DEV-OPS, CLOUD AND BIG-DATA IMPLEMENTATION.....**

Just a Click Ahead to Know More About Us

**Why You have To Learn from Inceptez**

- <https://www.shorturl.at/rsL26>

**FAQ about the Course**

- <http://Inceptez.in -> more -> Frequently Asked Questions>
- <https://www.shorturl.at/eflqA>

**Our Other Value Added Services**

- <http://Inceptez.in -> more -> Inceptez Interview & Job Support>
- <https://www.shorturl.at/inpKW>

**ENQUIRE, ANALYSE, COMPARE THEN EXPLORE,  
EXPERIENCE, EVOLVE, ENCHANT, EXCEL & FINALLY  
EXCEED YOUR CAREER GOAL ACCEPTING INCEPTEZ AS  
YOUR CAREER PARTNER**

Inceptez Technologies was founded by a team of Big data Evangelists in 2014 and is one of the leading IT training, Development and staffing company specializing in Big Data, Data Science, Dev-Ops, Cloud Computing and Internet of things (IOT). Inceptez is a non money oriented training center, where we first prioritize Comprehensiveness, Engagement based, Focus based, Competitive model with high Quantity and Quality in all the training as a paramount.

Inceptez Technologies is mastered and administrated by highly skilled industry experts. We are the technology enablers committed to provide comprehensive training to the aspiring professionals in the game changing, high demanding applications such as Hadoop, Spark , Data Analytics, Data Science, DEV-OPS and AWS/Google Cloud Platforms, that are the fastest growing trend setting technologies that provide competitive advantage in the ever changing IT world.

## **ABOUT INCEPTEZ**



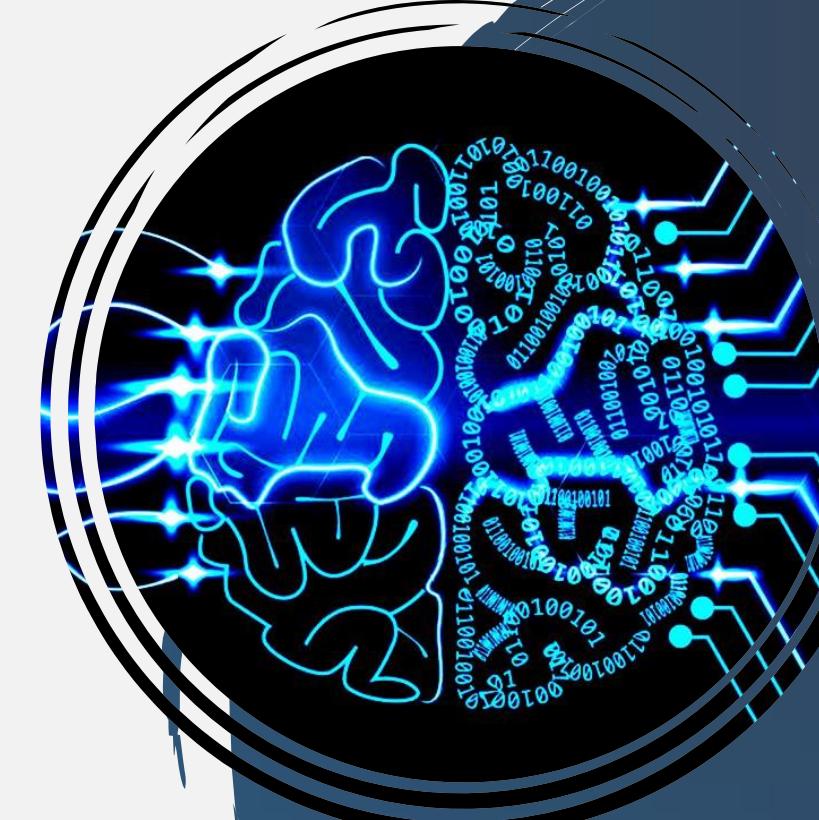
***We Just don't share the Knowledge rather we share  
Knowledge of Experience that Chisel you to feel technically  
Vibrated, Motivated & Overwhelmed***

 <b>TRAINING</b>  We are the technology leaders committed to provide comprehensive training to the aspiring professionals in the game	 <b>DEVELOPMENT</b>  Our software engineering process collects and translates business requirements into imaginative technology solutions that become reality with custom software development.	 <b>SOLUTION</b>  We are the technical leaders expertise in providing end to end solutions for the cutting edge technologies which Industries demands for their business growth and analytics requirements.	 <b>STAFFING</b>  Inceptez Technologies provides flexible, innovative recruitment strategies and technologies to maximize recruiting efficiency and reduce cost.
--	--	--	---

**Data Science is application of analytics of Big data,** that is becoming a strategic weapon for the tremendous growth of the Businesses and Companies of all sizes across all domains that has altered the business models of old industries and enabled the creation of new ones. **Data-driven businesses are worth \$1.2 trillion collectively in 2020**, an increase from \$333 billion in the year 2015.

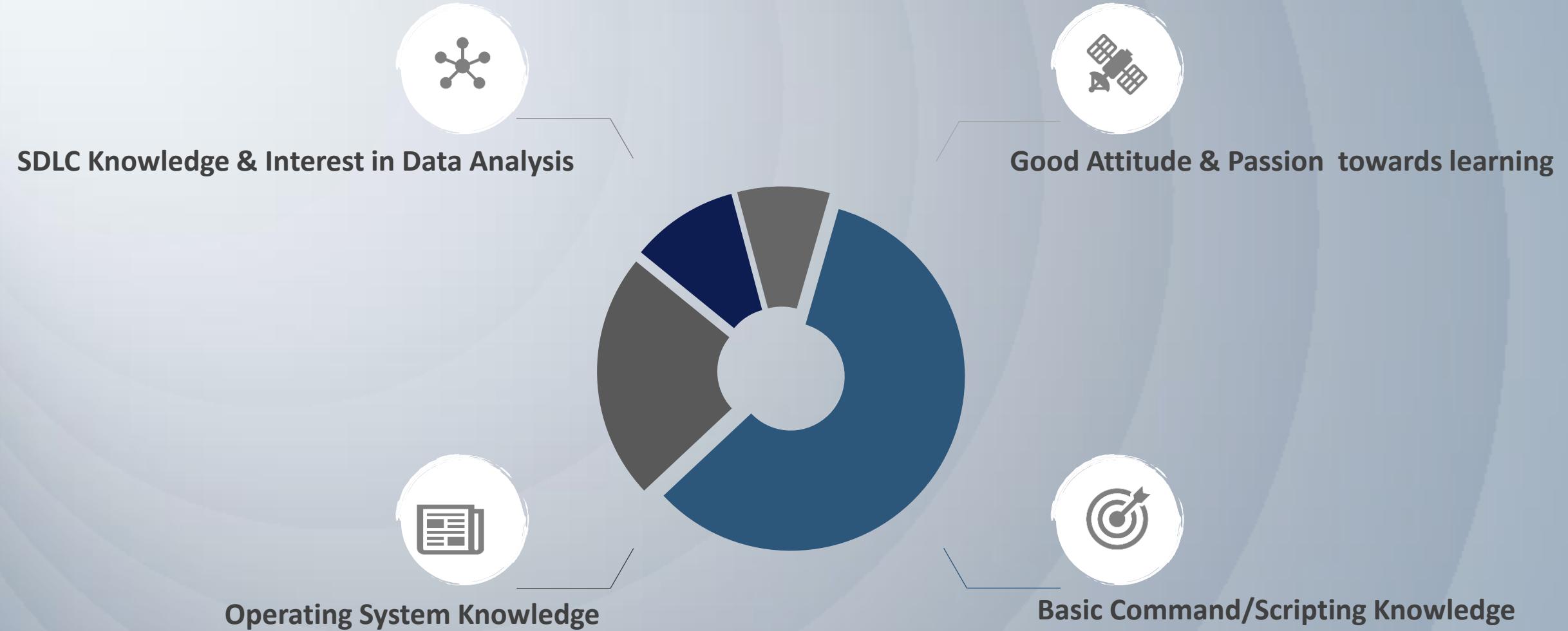
*Data scientists are responsible for breaking down big data into usable information and creating software and algorithms that help companies and organizations use data as a strategic resource for exceeding the growing business need in this competitive World.*

It is a well-proven fact that **effective collection and analytics** of the collected Big Data will place the implying organizations way ahead of their respective competitors. So the growing demand for a **qualified Data Scientists or a Data Science Engineer is multiplying not only in IT but across all Industries** are analyzing their own data. But **in contrast to this demand, there is a shortage of skilled Data Scientists.** And so, many top notching organizations are willing to pay extremely high pay packages for the best-skilled experts.



# About Data Science & Data Analytics

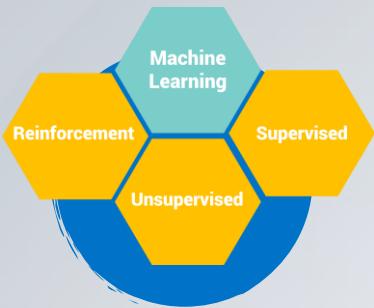
# Prerequisite - Good to Know



# What You will learn in this Unified Data Science Course



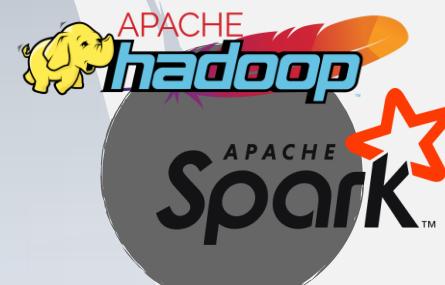
Python, Probability & Statistics



Machine Learning



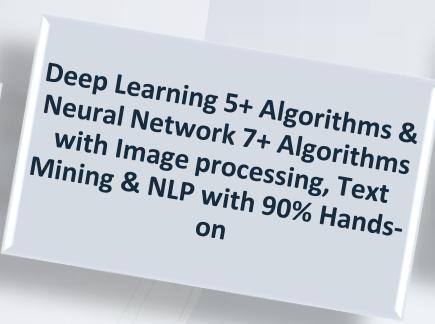
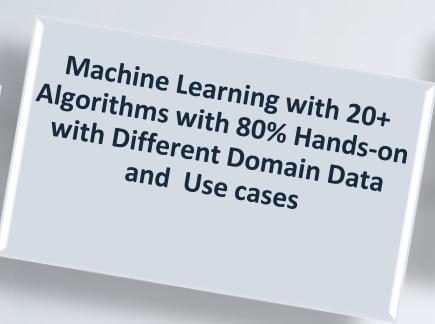
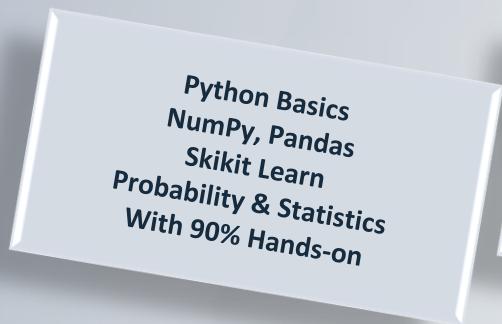
Deep Learning & Text Mining



Hadoop, Spark, Hive, Sqoop



Visualization, Dashboard & Cloud Deployment

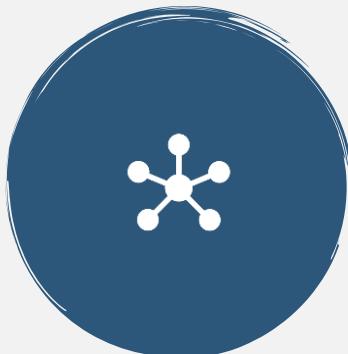




Focussed way of Teaching  
to address People from all  
Background



Competitive  
Teaching &  
Learning Model



Purely Hands-on  
Based Learning



Designed As per the Market  
Standard

# What's Unique in Our Course

**All under one course - Learning, Hands-on, Presentation, Implementation, Interview, Projects, Case Studies, Cloud Deployment etc.**

**Job Oriented Training, Professional Environment**

**All Trainings by Industry Experts, Completely Hands-on Driven,**

**End to End learning model from Data Extraction, Pre Processing, Model implementation, Training, Testing, Packaging & Deployment**

**Extended Training Duration to cover topics Wide and Depth**

**Use cases, Case Studies, Performance Tuning, Best Practices**

**Interview and Job Support with seasoning of Resume with Data Science**

**Provision of End to end simplified & comprehensive learning materials**

**Addresses 360 degree requirements of all students complete the training with Overwhelming Experience**

**Competitive learning, Active, Comparative & Declarative learning**

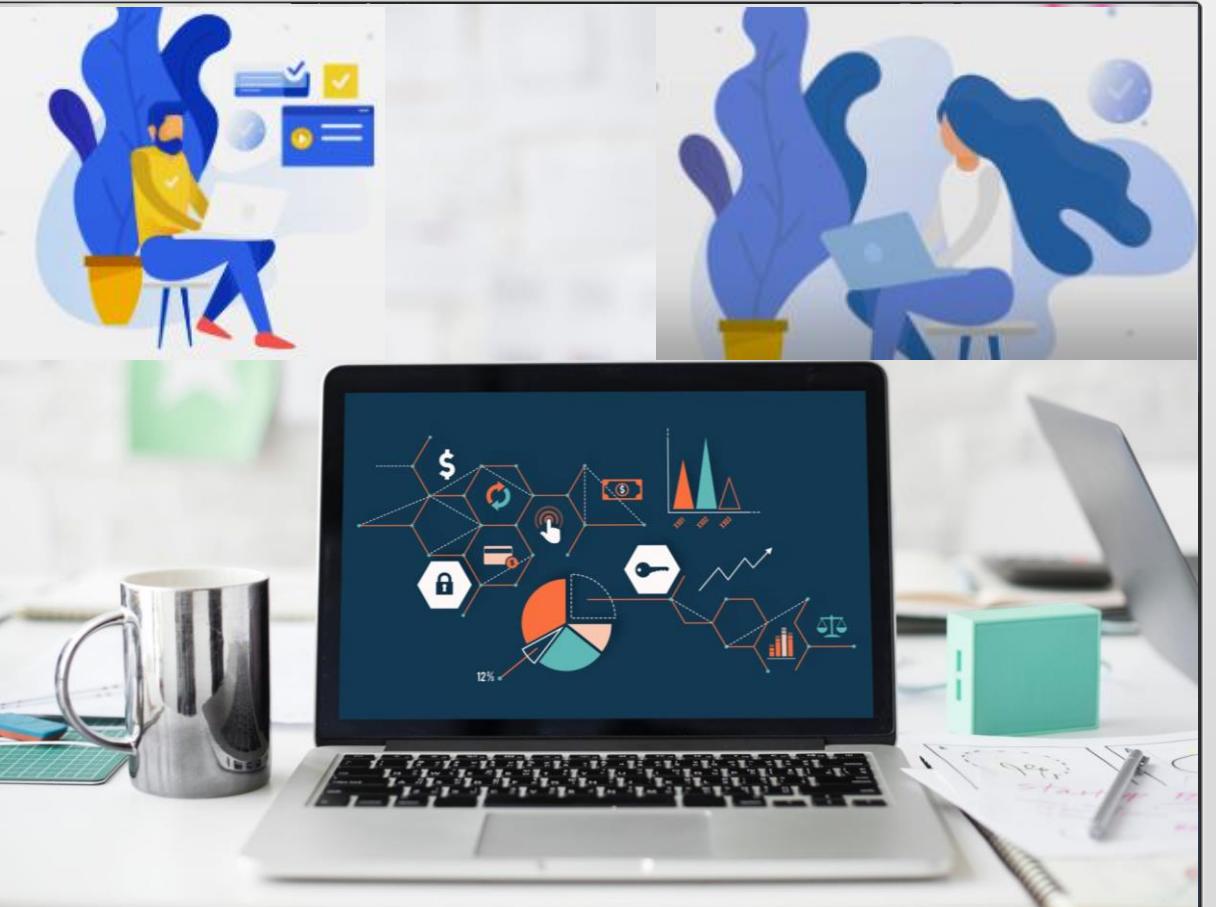
**20+ Use Cases, 4 Realtime Projects for gathering Realtime experience**

**2 Hackathons & 2 Tests with Certification Guidance**

**Packaging & Production Deployment Strategies on Cloud**

**Year 2020 Enriched Content with added Models, Features & Datasets**

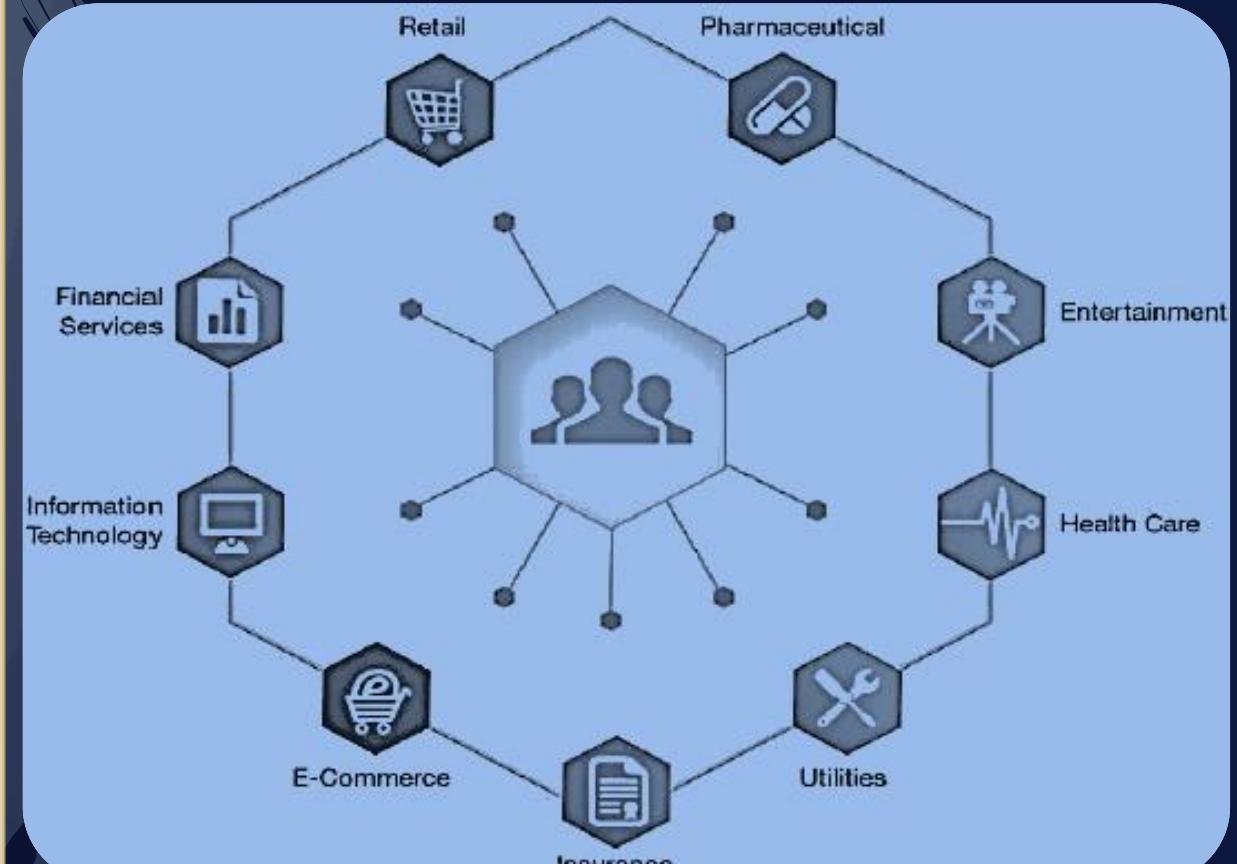
# High Level Course Curriculum



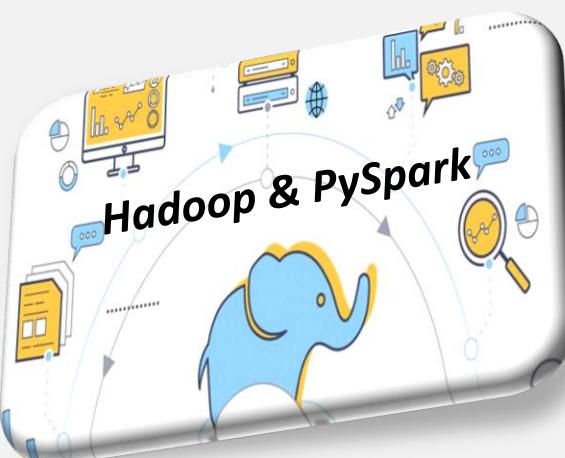
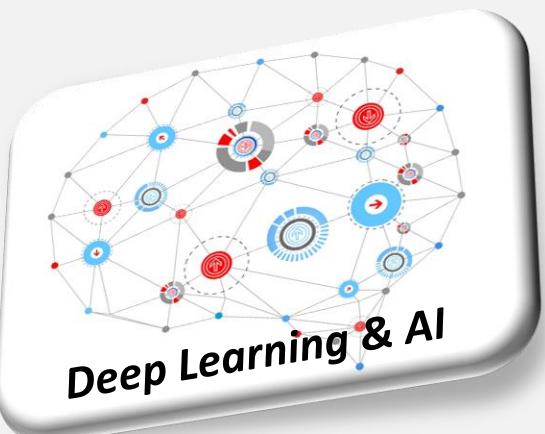
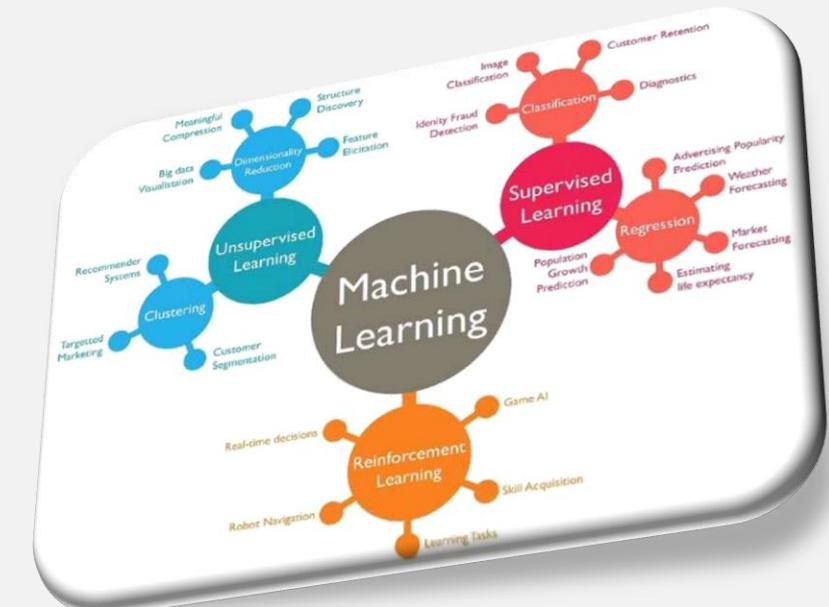
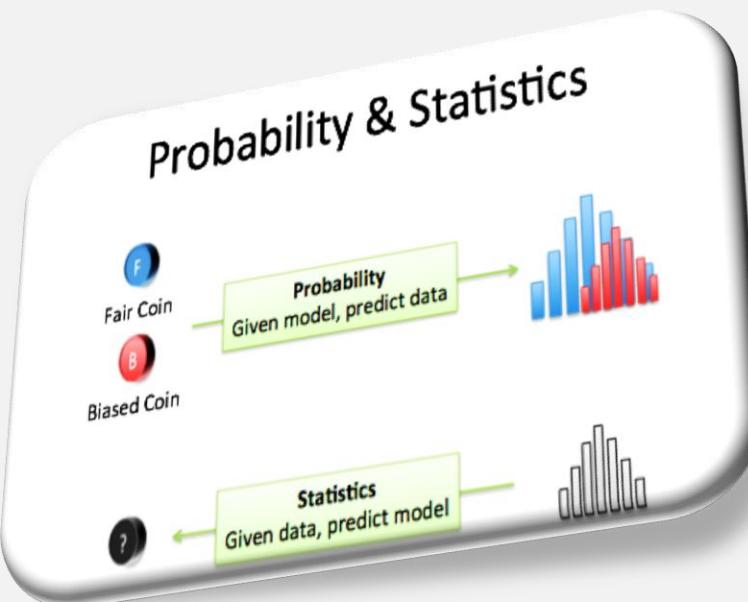
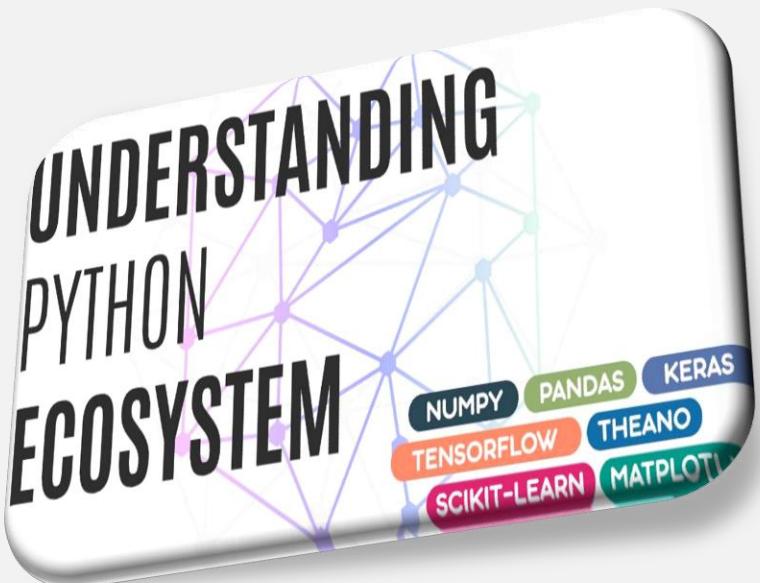
- INTRODUCTION TO DATA SCIENCE AND DATA ANALYTICS
- ESSENTIAL UNDERSTANDING OF PYTHON
- PYTHON BASICS & ADVANCED
- PYTHON ECOSYSTEMS FOR MACHINE LEARNING – NUMPY, PANDAS, SKIKIT
- FOUNDATION OF STATISTICS & FOUNDATION OF PROBABILITY
- OVERVIEW OF DATA SCIENCE
- END TO END PRE PROCESSING TECHNIQUES
- DEEP DIVE IN LINEAR REGRESSION WITH HANDSON & PROJECT
- DEEP DIVE ON LOGISTIC REGRESSION WITH HANDSON & PROJECT
- DEEP DIVE ON KNN ALGORITHM WITH HANDSON & PROJECT
- DEEP DIVE ON NAÏVE BIAS WITH HANDSON & PROJECT
- DEEP DIVE ON TIME SERIES FORECASTING WITH HANDSON & PROJECT
- DEEP DIVE ON SVM WITH HANDSON & PROJECT
- DEEP DIVE ON DECISION TREE WITH HANDSON & PROJECT
- ENSEMBLE TECHNIQUE (RANDOM FOREST, ADABOOST) WITH HANDSON & PROJECT
- UNSUPERVISED ALGORITHMS IN MACHINE LEARNING (K-MEANS, K-MEANS++, HIERARCHICAL CLUSTER) WITH HANDSON & PROJECT
- DEEP LEARNING WITH HANDSON & PROJECT
- NEURAL NETWORKS WITH HANDSON & PROJECT
- TEXT MINING & NLP WITH HANDSON & PROJECT
- DEVELOPMENT & DEPLOYMENT OF MACHINE LEARNING MODEL IN GOOGLE & AWS CLOUD
- DEVOPS – GITHUB WITH HANDSON
- PROJECTS, CASE STUDIES, USE CASES, BEST PRACTICES, OPTIMIZATION TECHNIQUES WITH HANDSON
- HACKATHON & TESTS
- BIGDATA – SQUOOP, HIVE, PYSPARK WITH HANDSON & PROJECT
- VISUALIZATION & DASHBOARD USING TABLEAU WITH HANDSON & PROJECT
- Hands-on from day one
- Exclusive Python coverage
- Insights on business use cases
- Handsome of data sets will be provided for practice
- All the PPTs, PDF and code will be shared
- Post-session support and assistance
- Specific business/domain use cases will be dealt with (based on the availability of the experts)
- Hackathon/Test will be conducted

# Domains Covered

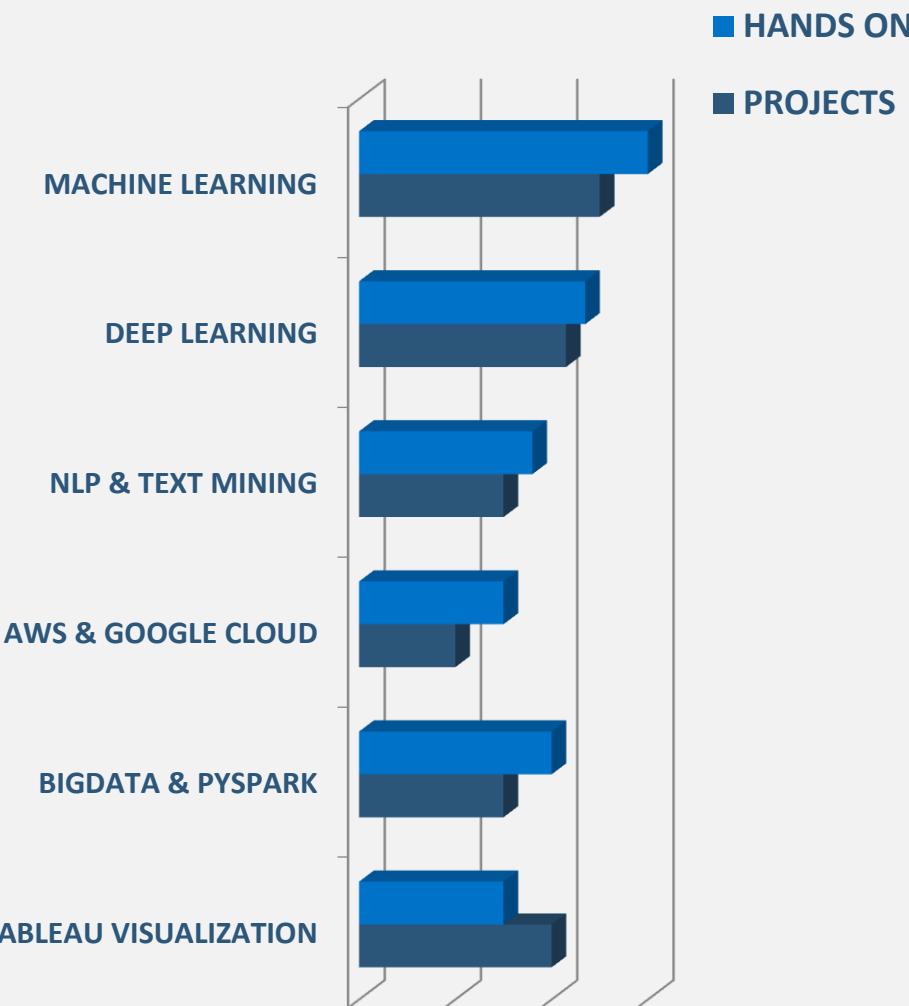
Media, Telecom, Medical, Real estate, retail, Airlines, Finance, Banking, Marketing, Images and Chat Raw text



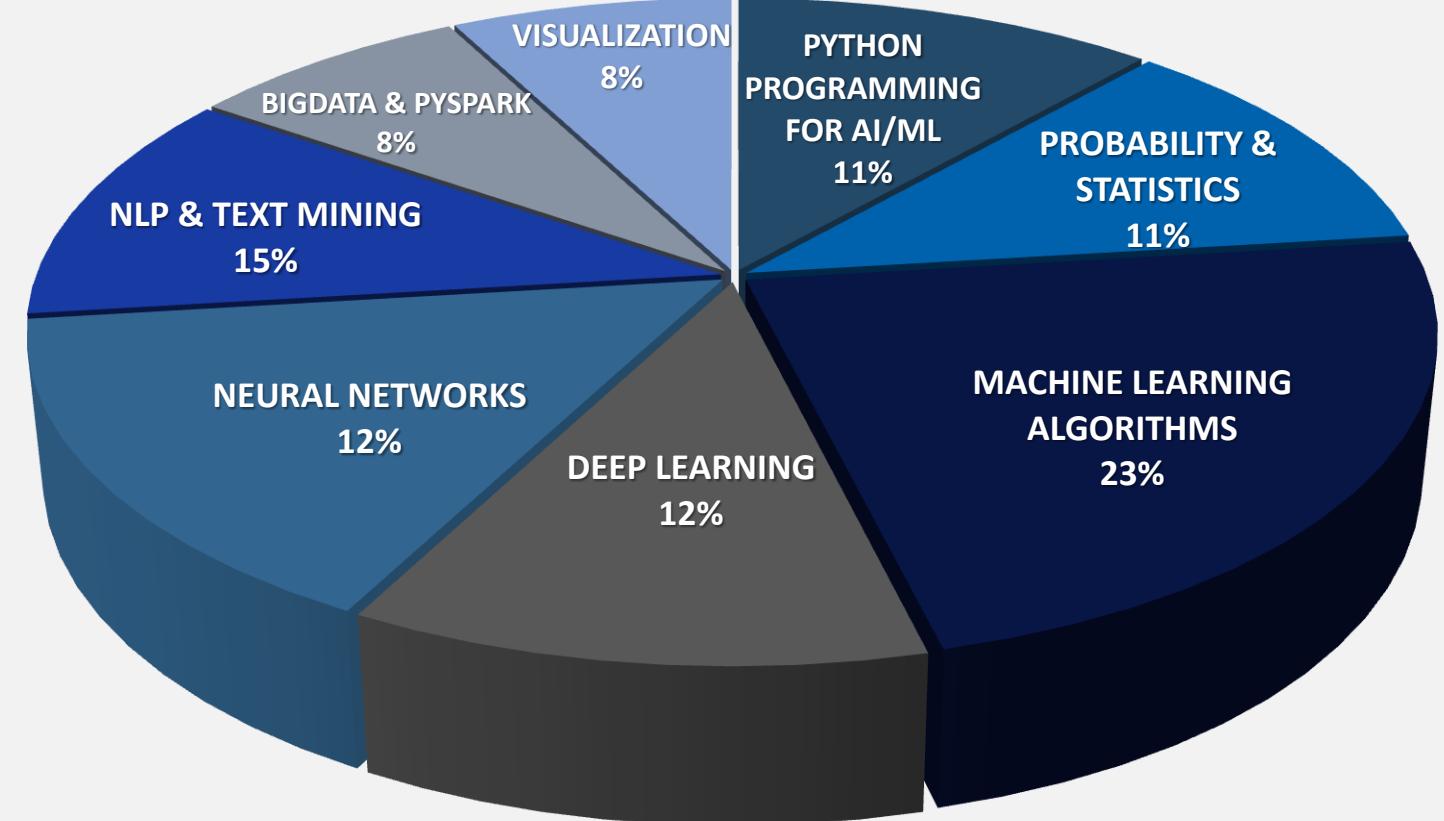
# Single Unified Course on AI & ML Deployed in Cloud with Bigdata & Visualization



## HANDS-ON & PROJECTS



## TECHNOLOGY STACK



# Course Key Highlights – Projects & Use cases

- ✓ Linear Regression on sales by spending their Advertisement on different streams
- ✓ Linear Regression on US Housing Price
- ✓ Logistic Regression
- ✓ Telecom - Churn prediction of customers based on past data.
- ✓ Create a model to predict the expectation of the telecom customer expected to disconnect or leave the service.
- ✓ Logistic Regression to predict the Breast cancer probability in the medical domain
- ✓ Demo on Titanic dataset prediction on who will survive using Logistic Regression
- ✓ Time Series Forecasting
- ✓ Forecasting air carrier traffic in US
- ✓ Forecasting Tractor sales
- ✓ How to Check Stationary of a Time Series?
- ✓ How to make a Time Series Stationary?
- ✓ Forecasting a Time Series
- ✓ Decision Tree - Prediction on balance scale data
- ✓ K-NN Algorithm -
- ✓ Classification
- ✓ Prediction on breast cancer wisconsin data
- ✓ Prediction on fruit data with colors
- ✓ Ensemble technique -
- ✓ Stacking
- ✓ What is an ensemble model?
- ✓ What are bagging, boosting and stacking?

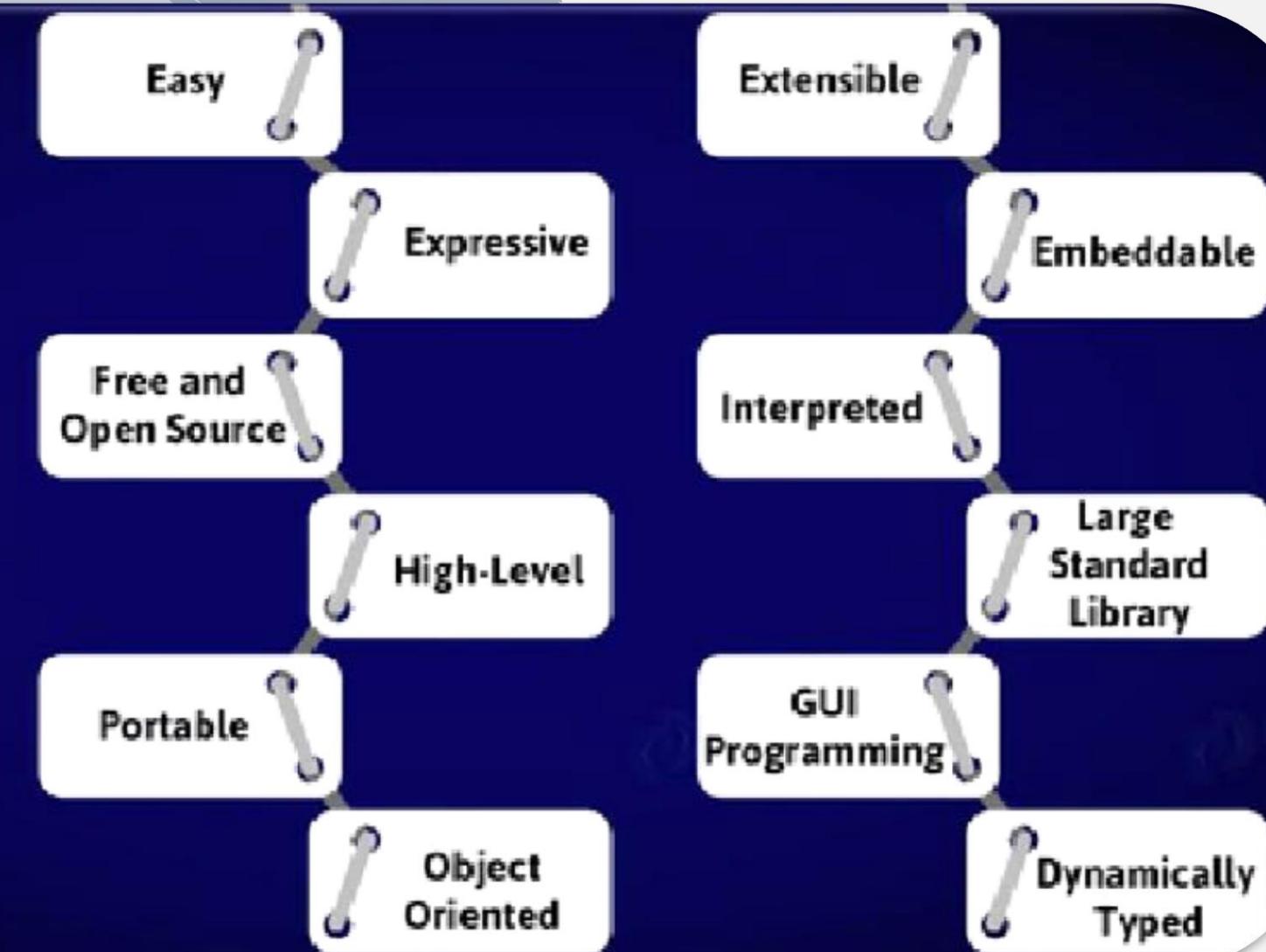
# Course Key Highlights – Projects & Use cases

- ✓ What are the benefits of ensemble model?
- ✓ Random forest Prediction on Lending Club data set
- ✓ AdaBoost Prediction on pima-indians-diabetes.data
- ✓ SVM Prediction with iris dataset
- ✓ Neural Network
- ✓ Toy Example
- ✓ Predicting median value of owner occupied homes
- ✓ Neural networks, a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data
- ✓ Deep learning, a powerful set of techniques for learning in neural networks
- ✓ Text Mining
- ✓ Data Extraction from Shakespeare novel
- ✓ Text mining NLTK and Sci kit learn
- ✓ Unstructured text is very common, and in fact may represent the majority of information available to a particular research or data mining project.
- ✓ Chat bots and NLP with IBM Watson Overview.
- ✓ Hadoop & Spark
- ✓ Customer Transaction batch acquisition and Processing.
- ✓ Twitter Sentiment Analysis.
- ✓ Weblog analysis.
- ✓ Visualization & Dashboard
- ✓ Sales prediction with Exploratory Data Analysis.
- ✓ Chat bot with IBM Watson libs.

## Features of Python



Python



# Data Science, Analytics Intro with Python

## DATA SCIENCE AND DATA ANALYTICS

Understand what is Data Science & Data Analytics at the nut shell before deep diving it

- What is Data Science
- Importance of Data Science
- Analytical Tools
- Profession of the future
- Pillars of Data science
- AREAS & NEED OF DATA ANALYTICS
  - Types of Business Analytics
  - Descriptive Analytics
  - Predictive Analytics
  - Prescriptive Analytics
  - Marketing Analytics
  - Proactive Analytics

## PYTHON FOUNDATION

In this module you will be learning Introduction & Key Components of Linux



- INTRODUCTION TO PYTHON
  - Objectives
  - History and overview of Python
  - Uses of Python
  - Features of Python
  - Install & Configuration of Python
  - Flavors of Python
  - Anaconda
  - Version of Python
- PYTHON HANDSON EXERCISE
  - Getting Started & Basics
  - Keywords & Identifiers
  - Statements & Comments
  - Python Variables

## PYTHON BASICS

Learning programming language basics to become a Python programmer

- PYTHON BASICS
  - Python Data Types
  - Python Type Conversion
  - Python I/O and Import
  - Python Operators
  - Python Namespace
- PYTHON FLOW CONTROL
  - Python if...else
  - Python for Loop
  - while Loop
  - break and continue
  - Pass Statement
  - Looping Techniques



# Python Programming for AI & ML

## PYTHON DATATYPES & FUNCTIONS

Python basic data types, collections, functions, lambda

- DATA TYPES
  - Python Numbers
  - Python List
  - Python Tuple
  - Python String
  - Python Set
  - Python Dictionary
  - Python Arrays
  - Python Matrix
  - List Comprehension

- FUNCTIONS
  - Defining a function
  - Calling a function
  - Types of functions
  - Function Arguments
  - Lambda function
  - Global and local variables



## PYTHON MODULES

Python modules, libraries, exception handling

- MODULES
  - Importing module
  - Math module
  - Random module
  - Packages
- EXCEPTION HANDLING
  - Exceptions type
  - Exception Handling
  - Except clause
  - Try ? finally clause
  - User Defined Exceptions
- PYTHON FOR DATA ANALYTICS
  - Manipulation of Data

## PYTHON FOR MACHINE LEARNING

Understanding and manage the multi dimensional array and matrices using NUMPY library

- NUMPY
  - Why Numpy?
  - Arrays
  - Using array generating functions
  - Diagonal and Zero matrix
  - Array Access
  - Array Slicing
  - Negative indexing
  - Strident Access
  - Array Operation
  - Matrix multiplication
  - Iterating over Array Elements
  - Vectorize
  - Arrays in condition
  - Scipy and Matplotlib
  - Plot visualization

# Python, Statistics & Probability

## PYTHON PANDAS HANDS ON

In this module you will be learning Python Pandas data frame library including data management and wrangling

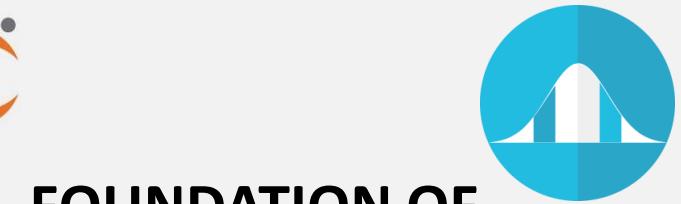
- PANDAS
  - Import modules
  - Create a data frame
  - Write to a csv file
  - Read data from file
  - Get data types
  - Take a look at the data
  - Working on the data
  - Describe the data
  - Add a column
  - Accessing and indexing the data
  - Missing data
  - Query the data
  - Apply a function
  - Grouping the data



## PYTHON PANDAS HACKATHON & TEST

Learn the foundation of Statistics in an interesting way with hands on exercises

- Series
  - Data Frame
  - Re indexing
  - Dropping Entries
  - Indexing, Selecting, Filtering
  - Arithmetic and Data Alignment
  - Function Application and Mapping
  - Sorting and Ranking
  - Axis Indices with Duplicate Values
  - Summarizing and Computing Descriptive Statistics
  - Cleaning Data Input and Output
- 
- [Hands on Exercises by students](#)
  - [Validation of results with Q&A](#)
  - [Writing of Exam in Python](#)



## FOUNDATION OF STATISTICS

Learn the foundation of Statistics in an interesting way with hands on exercises

- FOUNDATION OF STATISTICS
  - Statistical Jargons
  - Understanding the properties of an attribute
  - Central tendencies (Mean, Median, Mode)
  - Measures of spread (Range, Variance, Standard Deviation)
  - Z score
- FOUNDATION OF PROBABILITY
  - Exclusive Event
  - Independent Event
  - Introduction to random variables
  - Joint probability
  - Marginal probability

# Discrete Probability with Hands-on & Test

## FOUNDATION OF PROBABILITY

Learn the discrete probability in an interactive way with real time samples

- Union probability
- Conditional probability
- Probability theory
- Conditional probability
- Most powerful algorithms in probability theory
- Bayes Theorem
- Probability tree
- Confusion Matrix
- **DISCRETE PROBABILITY DISTRIBUTIONS**
  - Bernouli
  - Binomial
  - Geometric



## DISCRETE PROBABILITY DISTRIBUTION

Learn the discrete probability in an interactive way with real time samples

- Poisson and properties of each
- Continuous probability distributions
- Exponential
- Special emphasis on Normal distribution
- t-distribution
- Central Limit Theorem
- Sampling distributions
- Confidence Intervals
- Hypothesis Testing
- Statistical hypothesis
- Chi-square test
- T-test
- Z-test
- F-test & ANOVA



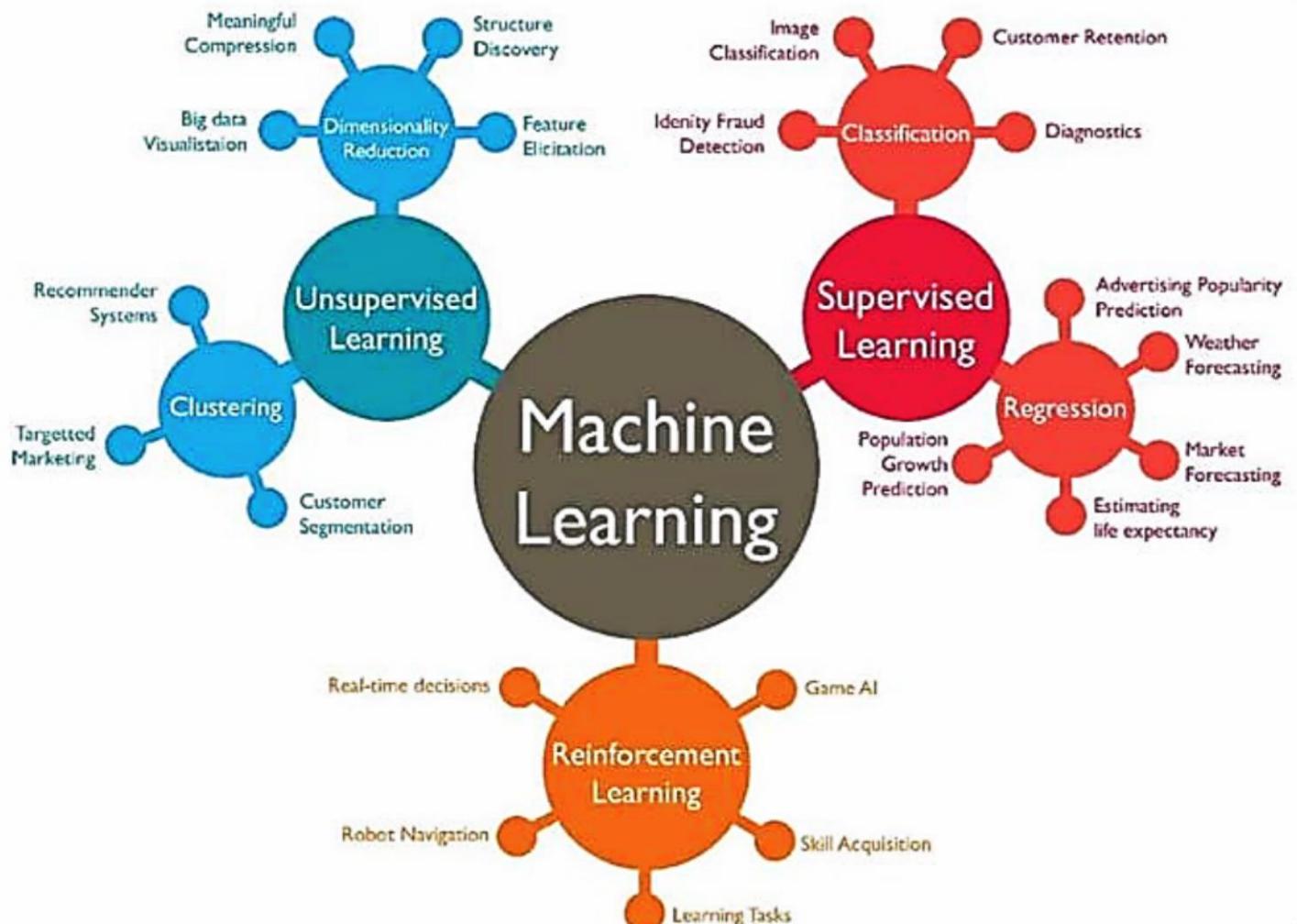
## PROBABILITY & STATISTICS HANDS-ON, TEST & HACKATHON

Solving of Problems of Statistics & Probability with test and hackathon

- Problem Solving in Probability & Statistics
- **Test on Probability**
- **Test on Statistics**
- **Test on Inferential Statistics**
- **Chi-square test**
- **T-test**
- **Z-test**
- **F-test & ANOVA**
- **Hackathon**
- **Q&A**



# Let's Deep Dive into the World of Data Science & Analytics



## Data Science

- based on strict analytical evidence
- deals with structured & unstructured data
- includes various data operations



## Artificial Intelligence

- imparts human intellect to machines
- uses logic and decision trees
- includes machine learning



## Machine Learning

- subset of AI
- uses statistical models
- machines improve with experience



# Deep Dive into Data Science

## ALL ABOUT DATA SCIENCE

Lets Enter into the world of Data  
Science & Analytics

- DATA ANALYTICS OVERVIEW
  - Introducing the world of Data Science
  - Examples of Data science helping up the business
  - Future of Datadriven decisions
  - Analytical Tools
  - Pillars of Data science
  - Understanding Analytics
  - Types of Business Analytics
  - Descriptive Analytics
  - Diagnostic Analytics
  - Predictive Analytics
  - Prescriptive Analytics
  - Real life uses case of Machine Learning

## PRE PROCESSING TECHNIQUES

Foundation of Preprocessing of raw data for data cleaning, transformation, reduction, conversion, feature selection etc.

- Building own use cases of ML (domain specific)
- Supervised Learning
- Unsupervised Learning
- Reinforced Learning
- PRE-PROCESSING
  - Why pre-process the data?
  - Why data is dirty?
  - Why data pre-processing is important?
  - Major task is data pre-processing
  - Data Cleaning
  - Data Integration
  - Data Transformation
  - Data Reduction
  - Feature extraction & Selection

## LINEAR REGRESSION

Analysis of Continuous dataset using Linear Regression using different Algorithms

- DEEP DIVE IN LINEAR REGRESSION
  - Understanding Linear Regression with examples
  - Gradient descent and its parameters
  - Formulae and maths behind this model
  - Multiple Linear Regression
  - Polynomial Regression
  - Categorical Variables in Regression
  - Error metrics to calibrate performance the model
  - End to end Hands-on modeling of real-time problems (Python and Scikit-learn) with domain Dataset
  - Realtime application with Pros & Cons

# Logistics, KNN, Kmeans, Naïve Bias



1. UPLOAD YOUR DATA



2. CREATE A DATASET



3. CREATE A LOGISTIC REGRESSION



4. ANALYZE YOUR RESULTS



5. EVALUATE THE LOGISTIC REGRESSION



6. MAKE PREDICTIONS

## LOGISTICS REGRESSION

Hands on exercise on unbalanced dataset, bias variance, exp functions, logarithms

- DEEP DIVE ON LOGISTIC REGRESSION

- Understanding Logistic Regression with an example
- Sigmoid function
- Formulae and maths behind this model
- Error metrics to calibrate the performance of the model
- End to end Hands-on modeling of real-time problems (using Python and Scikit-learn) with domain Dataset
- Hands on on exponential functions
- Logarithms
- Connecting the Concepts
- Realtime application with Pros & Cons



## KNN, KMEANS ALGORITHMS WITH HANDSON EXERCISE

- DEEP DIVE ON KNN ALGORITHM

- Understanding KNN with examples
- Formulae and maths behind this model
- How to find optimal K value
- Error metrics to calibrate the performance of the model
- End to end Hands-on modelling of real-time problems (Python and scikit-learn) with domain Dataset
- Telesales data analysis hands on
- Old faithful Geyser Data analytics
- Neighbour, metrics, weighed KNN hands on exercise
- Real time application Pros & Cons
- Kmeans Clustering real time hands on



## NAÏVE BIAS ALGORITHM DEEP DIVE

NB - Real time implementation

- DEEP DIVE ON NAÏVE BIAS

- Understanding Bayes Theorem
- Implementation Algorithm with example
- Math Behind the Algorithm
- Error metrics to calibrate the performance of the model
- Likely hood table example
- End to end Hands-on modelling of real-time problems (Python and scikit-learn) with domain Dataset
- Likely hood table management example with hands-on
- Real time application with Pros & Cons

# Time Series, Decision Tree, Hierarchical, Ensemble Algorithms

## TIME SERIES ALGORITHMS

Continuous data analytics of trending data set including smoothing, moving average with ARIMA & Cyclical

- DEEP DIVE ON TIME SERIES FORECASTING
  - Understanding Trend analysis
  - Cyclical
  - Seasonal analysis
  - Smoothing
  - Moving averages
  - Auto-correlation
  - ARIMA Applications of Time Series
  - Visualization of Time series continuous dataset
  - End to end Hands-on modeling using FB Prophet for Time series forecasting (Python) with domain Dataset
  - Auto Arima, Sales regression, Air passenger data analysis etc.,

## HIERARCHIAL CLUSTERING

Lineage and Distance matrix data analytics with hierarchical clustering

- HIERARCHIAL CLUSTERING
  - Divisive method
  - Agglomerative method
  - Linkage or distance matrix
  - Mall customer and shopping cart analysis Hands on
  - Dendograms hands on
  - Real time application with Pros & Cons
- DEEP DIVE ON SVM, STACKING & BOOSTING
  - Understanding SVM with examples
  - Learning about Kernel and Support Vector Machine
  - Formulae and maths behind this model

## SVM, STACKING & BOOSTING

Boosting Techniques with different Algorithms

- Boosting, Ensemble, Ada Boost, Stacking, SVM Boosting Hands-on with Use cases – Competition winning Algorithm
- Error metrics to calibrate the performance of the model
- Hands-on modelling using real-time problem using python
- DEEP DIVE ON DECISION TREE
  - Entropy, Information Gain and Gini Index
  - Formulae and maths behind this model
  - Error metrics to calibrate the performance of the model
  - Understanding Decision Tree with realtime examples and implementation

# Decision Tree, Ensemble, Un Supervised Learning

## DECISION TREE ALGORITHMS

Decision tree algorithms for decisioning a real world problem

- Hands-on modelling of real-time problems (using Python and scikit-learn) with domain Dataset
- Real time application with Pros & Cons
- Implementaion of CART algorithm
- Entropy techniques hands on
- ENSEMBLE TECHNIQUE (RANDOM FOREST, ADABOOST)
  - Bagging, boosting and stacking and its impact
  - Random forest
  - Adaboost
  - Math Behind the model
  - Sequential ensemble methods

## ENSEMBLE TECHNIQUES

Random forest and Ada Boosting with end to end hands on

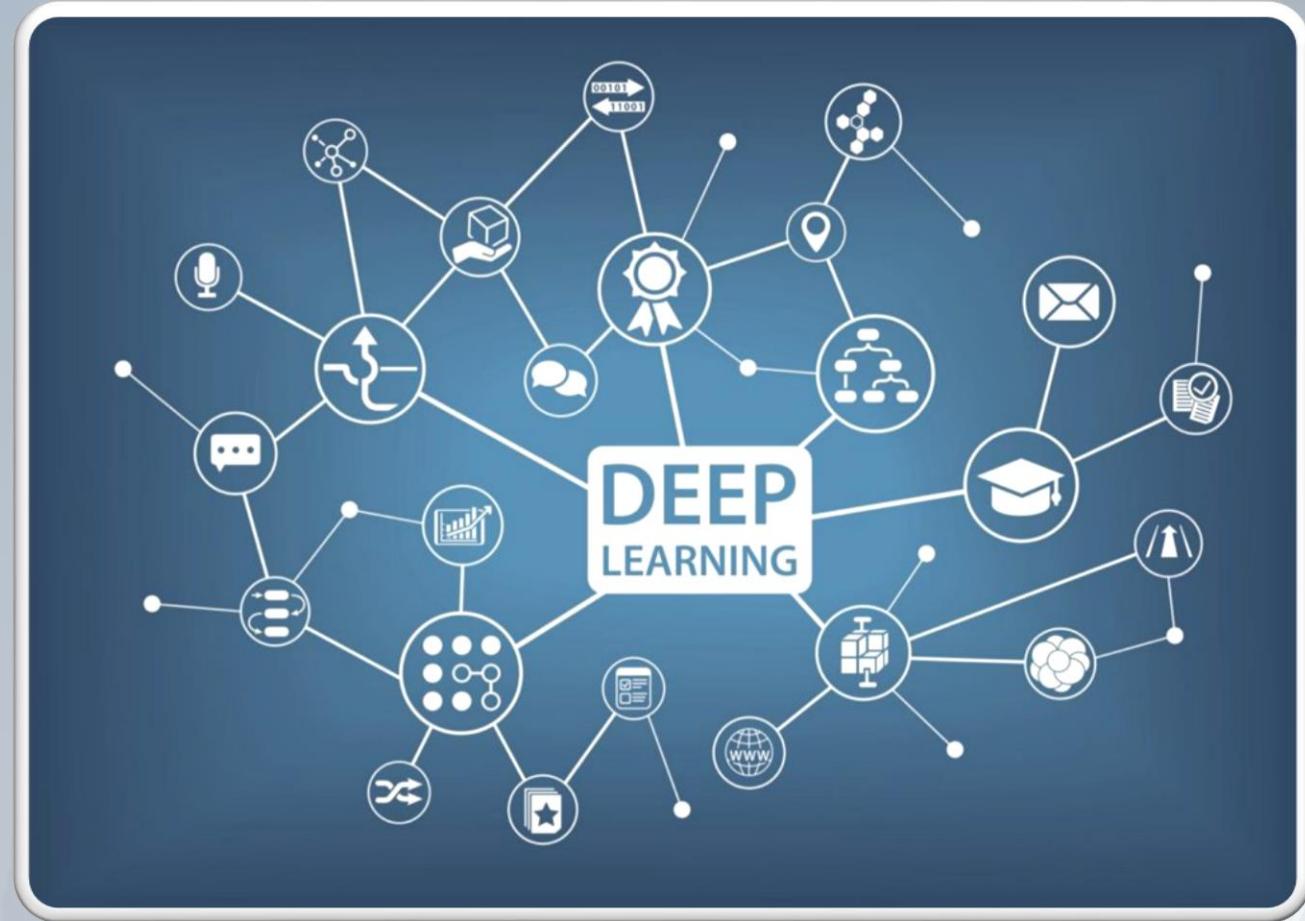
- Random Forest
- Gradient Boosting hands on
- Hands-on modelling of real-time problems (using Python and Scikit-learn)
- Real time application with Pros & Cons
- UNSUPERVISED ALGORITHMS IN MACHINE LEARNING (K-MEANS, K-MEANS++, HIERARCHICAL CLUSTER)
  - Unsupervised Learning: Clustering techniques
  - K means
  - K means++
  - Hierarchical Cluster
  - Math Behind the model

## UNSUPERVISED LEARNING ALGORITHMS + END TO END HANDSON

Usage of K means, ++, Hierarchical clustering for the model with handson

- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Test on every models
- Hackathon including comparision of the accuracy
- Quiz with Q&A
- Execution of Projects using different domain use cases such as ...
- Credit card data, Retail data, Market Data, Telecom Churn data
- Banking transaction data, customer cart data
- Air Carrier data, Medical Data etc.

# Let's Deep Dive into the World of Deep Learning, AI & Text Mining



# Deep Learning And Neural Networks

## DEEP LEARNING ALGORITHMS

Analysis of differential layers in Deep learning with different data set of the real-time domains

- DEEP LEARNING AND NEURAL NETWORKS
  - Demystifying Deep Learning
  - A case study on how DL could change the scope of predictions and decisioning
  - Neural Network Architecture & Basics
  - Output vs hidden layers
  - Linear vs nonlinear networks
  - Learning via gradient descent
  - Recursive chain rule (back propagation)
  - Hand-on Exercise on Churn modeling
  - Project on IMDB movie Classification
  - Diabetes Analytics with TensorFlow

## NEURAL NETWORKS DEEP DIVE

Usage of Differential advanced libraries for the Neural networks with hands-on Exercises

- NEURAL NETWORKS
  - Deep Neural Networks
  - Understanding feature extractions by DL
  - Understanding Feed-forward neural networks
  - Artificial Neural Networks (ANN)
  - Structure of ANN
  - Types of ANN
  - Real time implementation of ANN
  - Convolutional Neural Networks (CNN)
  - Tensor Flow for Neural Networks & Deep Learning
  - All about Recurrent Neural Networks (RNN)
  - RNN Algorithms

## NEURAL NETWORK END TO END PROJECTS

Implementation of the Neural Networks using the differential data sets

- NEURAL NETWORKS
  - RNN real time implementation with samples
  - All about Long Short Term Memory (LSTM)
  - Real time hands-on implementation of LSTM using domain use cases.
  - **Image Processing real time hands on in Deep Learning applied to Images using CNN**
  - **Implementing DL using Keras library in Python**
  - **Object recognition using DL techniques**
  - **AirPassenger Analytics, BitCoin Prediction**
  - **Stock data prediction, Shakespeare data analytics**

# Text Mining & Natural Language Processing

## TEXT MINING ALGORITHMS

Learn the interesting text mining algorithms for chat, voice to text converted datasets

- TEXT MINING
  - Text Transformation
  - Preprocessing
  - Understanding information retrieval
  - Crawling
  - Language modelling
  - Chunking
  - Lexical Chaining
  - Text Indexing
  - Inverted Indexes
  - Relevance Ranking
  - Term Frequency
  - Inverse Document Frequency
  - Chat data analytics hands on
  - Twitter data analytics hands on

## NATURAL LANGUAGE PROCESSING

Execute NLP algorithms with real life examples using different algorithms such as NLTK, TextBlob, Spacy etc.,

- NATURAL LANGUAGE PROCESSING (NLP)
  - Understanding NLP
  - Real-life systems using NLP
  - Parsing and semantic structures
  - Stemming & Lemmatization
  - Part of Speech tagging
  - Named Entity Recognition (NER)
  - Applications of NER
  - Sentiment Analytics
  - Topic Modeling
  - Unsupervised learning in Text analysis
  - **Hands-on activity with NLTK libraries for analyzing customer feedback and chats**

## PRINCIPAL COMPONENT ANALYSIS for Visualization

PCA projections with text mining

- **Hands-on activity with Text Blob libraries and accuracy comparison**
- **UnSupervised learning with feature engineering**
- **LDA Algorithm implementation**
- PRINCIPAL COMPONENT ANALYTICS (PCA)
  - Loading of Data set
  - Standardize the data set
  - PCA Projection to 2D
  - Visualize 2D Projection
  - Explained Variance
  - Limitations of PCA
  - **Handon exercise with Projects, use cases and case study on Text mining**
  - **Test, Quiz with Q&A**

# CLOUD & DEVOPS Implementation of Model

## Model Deployment IN AWS & Google CLOUD

Handson on End to end model building, export, deployment and execution in GCP & AWS Cloud

- **Development and Production Deployment of the ML models (AWS & Google CLOUD)**

- Amazon EC2
  - Purpose of EC2?
  - Key Components in EC2
  - EC2 Key Pair
  - EC2 AMI
  - EC2 EBS
  - EC2 Instance
  - EC2 ELB
  - Do's and Dont's with EC2
  - Introduction of cloud services
  - IAAS, PAAS, SAAS
  - How to create a VM instance
  - How to access server

## DEVOPS Components

This module details Docker to create nodes, swarm clustering, scaling, demolishing with AWS implementation of Docker Containers

- Building a simple REST application using Python FLASK
- Exposing the Linear Regression ML model as REST API using FLASK
- **Exposure of the Cloud deployed application to the Stake holders**
- **Production Deployment Handson on Deploying the ML model in AWS Cloud and consuming it using a sample application**
- **Production Deployment Deploying the ML model in Google Cloud Platform and exposing the model to the Clients and Outside world.**
- **DEVOPS**
  - DevOps Basics
  - Version Control systems

## HACKATHON & PROJECTS

Handson on the Kubernetes orchestration nd networking, Introduction to Jenkins & components

- Create and use a repository
- Cloning of Repositories
- **Hands on on Git commands**
- **Hands on Init, Push, Pull, Commit etc.,**
- **Hands on Start and manage a new branch**
- **Hands on Make changes to a file and push them to GitHub as commits**
- **Hands on Open and merge a pull request**
- **Working in a collaborative environment with the maintenance of versions**
- **Creating public profile in GIT.**
- **HACKATHON & TEST ON MACHINE LEARNING & DEEP LEARNING**

# Let's become a Data Science Engineer with the exposure of Hadoop & Spark



## Benefits of Considering Hadoop



Fast



Flexible



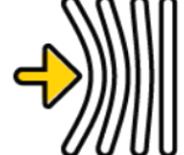
Scalable



Cost-Effective



High Throughput



Resilient to Failure



Cost-Effective



High Throughput



Resilient to Failure

# Big Data Hadoop & Py Spark with Eco Systems

## Big Data Introduction

End to end Big Data Introduction with Sqoop import and export from databases

- INTRODUCTION TO BIGDATA, HADOOP & SPARK
  - Introduction to Big data
  - Classification
  - Characteristics
  - Different methods of analyzing big data
  - Hadoop HDFS & Map reduce overview.
  - **Hadoop HDFS commands Exercise**
- SNOOP – HDFS –DATA EXTRACTION/ACQUISITION
  - Introduction
  - Why Sqoop
  - Sqoop Import Architecture
  - Sqoop Export Architecture

## Exploring Hive an SQL Layer on Hadoop

Perform declarative SQL analysis using Hive Query Language

- Customer & Product Data extraction
- Change Data Capture using Sqoop
- **Export of Complex Event Processed Data using Sqoop Export**

### HIVE – SQL & OLAP Layer on Hadoop

- Introduction
- Architecture
- Hive Vs RDBMS
- Create Tables (Managed, external)
- Managed Vs External tables
- **Hands-on on Hive access through Hive Client**

## PySpark Essentials with Hands On

Explore the Distributed in memory framework on Spark using Python

- **Handson on Load and manage bulk Data**
- **Handson on Hive Query Language**
- **Handson on Analysis of Data using Hive**
- **Handson on Partitioning (static and dynamic)**
- **PySpark Essentials**
  - Overview
  - Daemons
  - A Spark Standalone Cluster
  - Components & Terminologies
  - Workers
  - Driver Programs
  - Tasks, Executors & Cluster Manager

# PySpark Handson with Project

## PySpark Handson

- Py-Spark Hands-on Exercises
  - **RDD Execution model**
  - **Types of RDDs**
  - **RDD Operations**
  - **Transformations and Actions**
  - **Lazy executions**
  - **RDD Lineage**
  - **Loading data in RDD**
  - **Transformations and Actions**
  - **Creating the SparkContext**
  - **REPL Commands**
  - **Different types of basic operations**
  - **Language integrated Query methods**
  - **RDD Operations**
  - **Text, CSV, JSON, Parquet, ORC, Compression Techniques, Spark SQL**

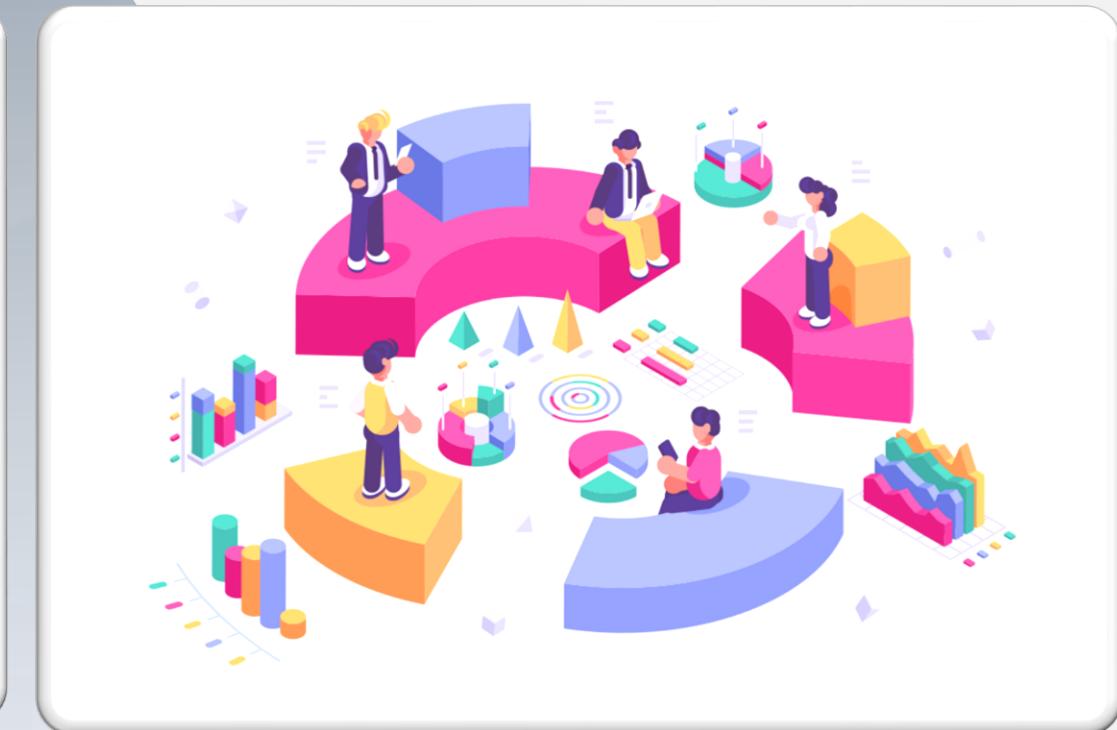
## Project Using PySpark

- End to End Project on PySpark
- Customer and Product dataset
- Serialization and deserialization of datasets
- Project Execution

## Performance Tuning

- Performance tuning with in memory tunables of pyspark
- Interview question discussion
- Test and Hackathon on BigData
- Project Execution

# PRESENTATION LAYER – DATA VISUALIZATION & DASHBOARD



## Visualization Basics

Understand the art of creating visualizations, charts, maps, Dashboards using Tableau

- **VISUALIZATION & DASHBOARD**
  - Art of visualizing Data
  - Types of Visualization
  - How to decide the right visualization
  - Connecting to data
  - Trend lines
  - Reference lines
  - Statistical analysis in tableau
  - Data Strategy
  - Native support for databases
  - Joining multiple tables in tableau
  - Custom SQLs
  - Building of Dashboard invoking the machine learned dataset.

## Visualization End to End Hands on

Learn End to end practices of building Visualizations and Dashboards using Tableau

### • VISUALIZATION HANDS ON

- **How are my Sales spread over geography and which region tops the most?**
- **What are the categories which are giving me good profit? Which subcategory should i be concentrating more?**
- **Which products should i be investing more? Show me the bottom 10 products by Sales**
- **How am I trending over time on Sales and Profit?**
- **Exploratory data analysis with Tableau.**
- **Connecting JSON with Tableau.**
- **Connecting tableau with Oracle and creating dashboards from them**

## Additional Engagements

Get Ready for the job and Interview with extra mile of efforts

- Resume Building
- Mock Interviews
- Project execution
- Presentation
- Attendance of Test
- Quiz Engagements
- Insights on business use cases
- Recap of Components
- Rewards & Awards
- Certifications
- Blogging

# End to End Projects Highlights

- ✓ *End to End integration BigData, Data Analytics & Visualization projects*
- ✓ Linear Regression on sales by spending their Advertisement on different streams
- ✓ Linear Regression on US Housing Price
- ✓ Logistic Regression
- ✓ Telecom - Churn prediction of customers based on past data.
- ✓ Create a model to predict the expectation of the telecom customer expected to disconnect or leave the service.
- ✓ Logistic Regression to predict the Breast cancer probability in the medical domain
- ✓ Demo on Titanic dataset prediction on who will survive using Logistic Regression
- ✓ Time Series Forecasting
- ✓ Forecasting air carrier traffic in US
- ✓ Random forest Prediction on Lending Club data set
- ✓ AdaBoost Prediction on pima-indians-diabetes data
- ✓ Predicting median value of owner occupied homes
- ✓ Chat bots and NLP with IBM Watson Overview.
- ✓ Customer Transaction batch acquisition and Processing.
- ✓ Twitter Sentiment Analysis.
- ✓ Weblog analysis.
- ✓ Sales prediction with Exploratory Data Analysis.



## Projects

Key Stuffs behind the success that provides real experience...



THANK YOU

Instructor Profile

*Noordeen*

(Senior Data Scientist)

<https://www.linkedin.com/in/nursnaaz/>

<https://github.com/nursnaaz>

*Inceptez Technologies Pvt. Ltd.* 

+91 70107 90330 

[info@inceptez.com](mailto:info@inceptez.com) 

[www.inceptez.com](http://www.inceptez.com) 

[www.inceptez.in](http://www.inceptez.in)