INTELLIGENCE DIGITALIZED…

info@inceptez.com www.inceptez.com

27, A, Brahmin Street, Dhandeeswaram, Velachery, Chennai 42.

About Us

Inceptez Technologies *was founded by a team of Big data Evangelists in 2014 and is one of the leading IT training and staffing company specializing in Big Data, Data Science, Dev-Ops, Enterprise Mobility, Cloud Computing and Internet of things (IOT), 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 Big Data oriented business applications such as Hadoop , Analytics, Data Science, and Internet of things (IOT) that are the fastest growing trend setting technologies that provide competitive advantage in the ever changing IT world.*

Our training methodology is designed to address each and every audience to excel the technology ***inside out focusing*** *the need of each student from different technologies who wanted to place their foot print in Big Data, Data Science and Internet of things.*

*We have set a clear mission with a **Qualitative and Quantitative** structured training methodology considering the students need in mind. We ensure our training is comprehensive, Focus based and Engagement based by providing **real time oriented training** hence the audience will be able to co-relate with the real scenarios.*

Our Exclusivity

1. In depth advanced and most recent versions **architecture oriented** training.
2. Hands on **exercise and Practicals using real time** use cases.
3. **Focus oriented training** for addressing all candidates having different technical background.
4. Our Significance is to provide comprehensive training coverage on Batch, Interactive, Real time and In-memory data pipeline.
5. Conduct **tests with interview questions** on all topics covered to engage and ensure the learnability of each student.
6. End of course **Project** with end to end integration of the tools learned to acquire **real-time on job knowledge**.
7. Support on building effective **resume** seasoned with bigdata projects and technologies related to Big data-Hadoop for better **job opportunities**.
8. **Certification guidance** will be provided for adding more weightage to the profile.
9. **Lifetime Help and support** through phone or social medias.
10. We provide **job assistance** by posting and guiding on the job requirements from different companies.
11. UNCOMPARABLE COMPLETE COURSE CONTENT ACROSS ANY OTHER TRAINING CENTERS.

What's the focus of this course?

The Data Science with Python and Spark training course has been designed to impart an in-depth knowledge of the various data analytics and data science techniques which can be performed using Python. The course is packed with real-life projects, case studies, lab sessions, Hackathon, Tests and extensive knowledge sharing sessions.

How Inceptez Technologies is Significant and Outstanding?

- ✓ Market standard Contents
- ✓ Professional Environment
- ✓ All Trainings by Industry Techies & Subject Matter Experts
- ✓ Extended Training Duration
- ✓ More Focus on Hands-on
- ✓ Use case Oriented
- ✓ Interview and Job Support
- ✓ Help & Support across different channels like Phone, Whatsapp & Mail
- ✓ End to end documentation
- ✓ Certification Guidance
- ✓ 360 degree Requirements Satisfaction Index
- ✓ Competitive Pricing

Who is the target audience?

- ✓ An aspiring candidates with good data knowledge and business knowledge.
- ✓ Interest in Data Science
- ✓ Wants to be a part of the growing digital community
- ✓ Wants to increases your job prospects and want to be the highly paid professional
- ✓ Wants to improve their data mining skills
- ✓ To improve their statistical modelling skills
- ✓ To improve their data preparation skills
- ✓ Wants to improve their Data Science presentation skills

Extensive Course on Python, Data Science, Data Analytics, Deep Learning, Bigdata & Visualization

Duration: 120 Hours

Level : Expert

An extensive, comprehensive, use case and hands on oriented, Advanced topic covered course designed with End to end Data science, Bigdata Data Engineering with Hadoop and Spark, Visualization and Dashboard.

Hands-on Use Cases and Projects Executed in the complete course on Domains (Media, Telecom, Medical, Real estate, retail, Airlines, Finance, Banking and Raw text) as follows:

- ✓ **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 Stationarity 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

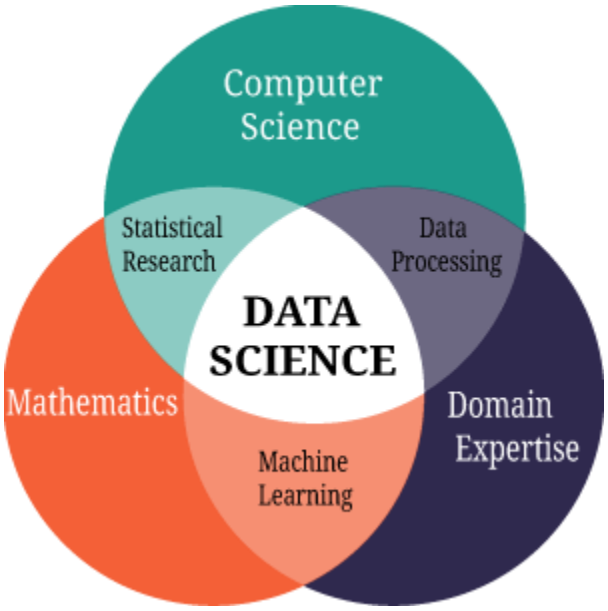
- Stacking
- What is an ensemble model?
- What are bagging, boosting and stacking?
- What are the benefits of ensemble model?

✓ 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 Scikit 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.**
  - ➢ **Chatbot with IBM Watson libs.**

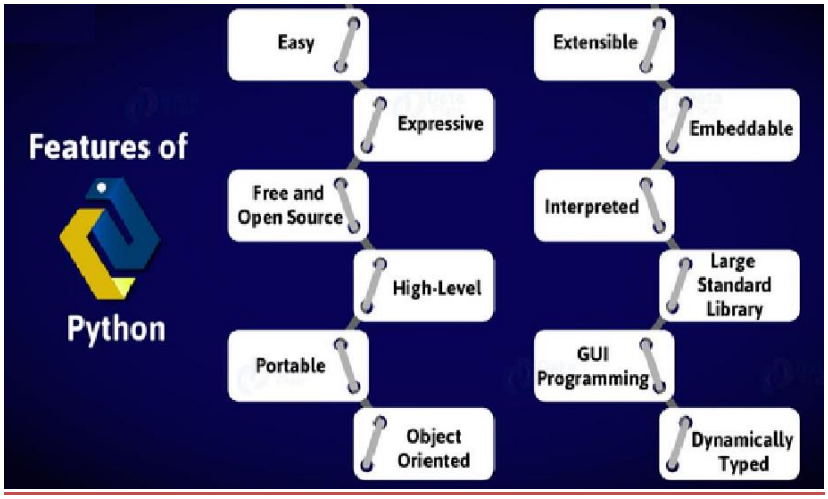
Main Course Content



INTRODUCTION TO DATA SCIENCE AND DATA ANALYTICS

What is Data Science - Importance of Data Science - Analytical Tools - Profession of the future – Pillars of Data science– Areas - Need of Business Analytics - Types of Business Analytics - Descriptive Analytics - Predictive Analytics - Prescriptive Analytics - Marketing Analytics –Proactive Analytics.

PYTHON



ESSENTIAL UNDERSTANDING OF PYTHON

Introduction to Python – Objectives - History and overview of Python - Uses of Python- Features of Python - Install Python - Flavours of Python - Anaconda - Version of Python

PYTHON BASICS

- Getting Started
- Keywords & Identifiers
- Statements & Comments
- Python Variables
- Python Data Types
- Python Type Conversion
- Python I/O and Import
- Python Operators
- Python Namespace

FLOW CONTROL

- Python if...else
- Python for Loop
- while Loop
- break and continue
- Pass Statement
- Looping Technique

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

MODULES

- Importing module
- Math module
- Random module
- Packages

EXCEPTION HANDLING

- Exception
- Exception Handling
- Except clause
- Try ? finally clause
- User Defined Exceptions

PYTHON - MANIPULATING DATA

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.

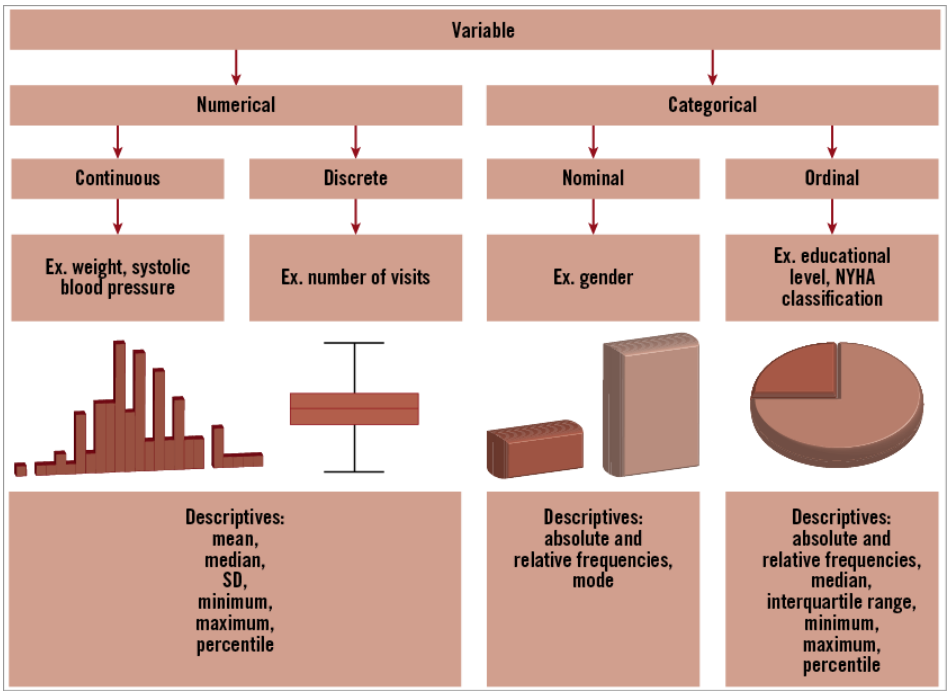
Scipy and Matplotlib

Plot- visualization

PANDAS

Import modules - Create a dataframe - 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

PROBABILITY & STATISTICS



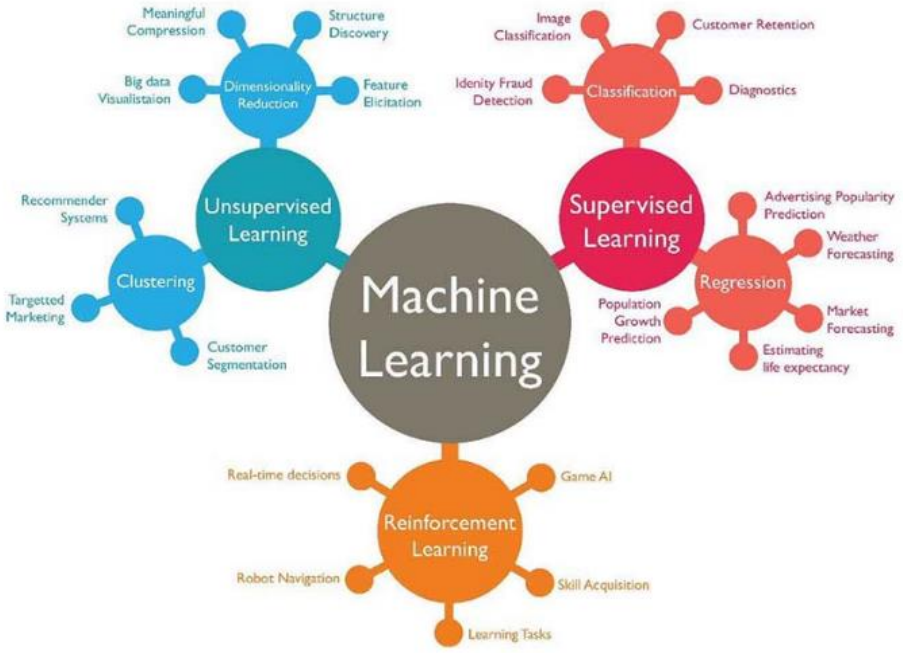
FOUNDATION OF STATISTICS

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 – Union probability – Conditional probability - probability theory, conditional probability, and to a most powerful algorithm in probability theory - Bayes Theorem – Probability tree – Confusion Matrix Discrete probability distributions - Bernoulli, Binomial – Geometric - 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

DATA SCIENCE – AI/ML



OVERVIEW OF DATA SCIENCE

Introducing the world of Data Science -Examples of Data science helping up the business - Future of Data-driven 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 - 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

DEEP DIVE IN LINEAR REGRESSION

- Understanding Linear Regression with an example
- 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
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

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
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

DEEP DIVE ON KNN ALGORITHM

- Understanding KNN with an example
- Formulae and maths behind this model
- How to find optimal K value
- Error metrics to calibrate the performance of the model
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

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
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

DEEP DIVE ON TIME SERIES FORECASTING

- Understanding Trend analysis, Cyclical and Seasonal analysis, Smoothing
- Moving averages, Auto-correlation
- ARIMA Applications of Time Series
- Hands-on modelling using FB Prophet for Time series forecasting (Python)

DEEP DIVE ON SVM

- Understanding SVM with an example
- Learning about Kernel and support vector machines
- Formulae and maths behind this model
- Error metrics to calibrate the performance of the model
- Hands-on modelling of one real-time problem (using Python and sci-kit-learn)
- Pros and cons

DEEP DIVE ON DECISION TREE

- Understanding Decision Tree with an example
- Entropy, Information Gain and Gini Index
- Formulae and maths behind this model
- Error metrics to calibrate the performance of the model
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

ENSEMBLE TECHNIQUE (RANDOM FOREST, ADABOOST)

- Bagging, boosting and stacking and its impact
- Random forest, Adaboost
- Math Behind the model
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and 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
- Hands-on modelling of real-time problems (using Python and scikit-learn)
- Pros and cons

DEEP LEARNING AND NEURAL NETWORKS

- Demystifying DL (with a case study on how DL could change the scope of predictions and decisioning)
- Neural Network Architecture and Basics
- output vs hidden layers; linear vs nonlinear networks
- learning via gradient descent; recursive chain rule (backpropagation)
- Deep Neural Networks
- Understanding feature extractions by DL
- Understanding Feed-forward neural networks
- ANN
- Convolutional Neural Networks
- Deep Learning applied to Images using CNN
- Tensor Flow for Neural Networks & Deep Learning
- Implementing DL using Keras library in Python
- Object recognition using techniques
- Recurrent Neural Networks (RNN) and LSTMs

Text Mining:

- Understanding information retrieval, Crawling and Language modelling
- Text Indexing, Inverted Indexes
- Relevance Ranking TF and IDF

Natural Language Processing:

- Understanding NLP, real-life systems using NLP
- Parsing and semantic structures, Stemming, POS tagging
- Named Entity Recognition and applications of NER
- Sentiment Analysis
- Topic Model – Unsupervised in Text analysis
- Hands-on activity with NLTK, TextBlob

Developing and Deploying ML models (AWS CLOUD):

- Introduction of cloud
- How to create a VM instance
- How to access server
- Building a simple REST application using FLASK
- Exposing the Linear Regression ML model as REST API using FLASK
- Deploying the ML model in AWS and consuming it using a sample application

COURSE COMPLETION PROJECT

- Dataset will be provided
- Expected to prepare and transforms the data to be applicable for modelling
- Model using various algorithms and provide the best prediction results
- Apply Ensemble techniques to improve the prediction accuracy

KEY POINTS

- 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/contest will be conducted

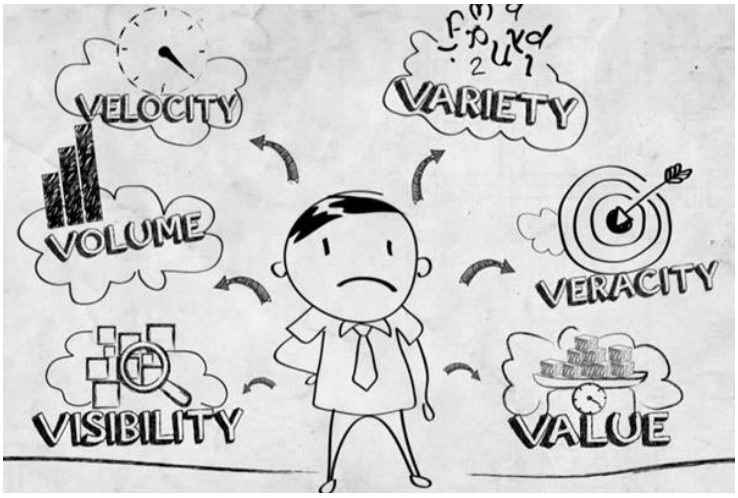
HACKATHON

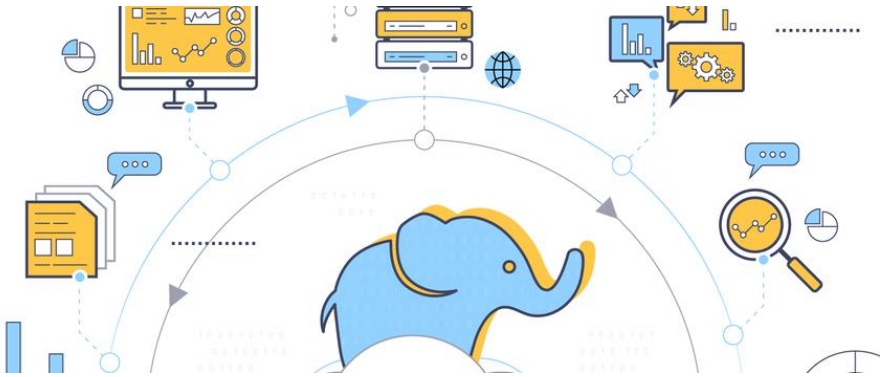
- One day hackathon will be conducted
- Dataset will be given to solve the problem
- Come with business insights
- Presentation should be given to the audience the work
- Based on that scholarship will be awarded

Technical Advisor Profile for Machine Learning

Mohamed Noordeen Alaudeen
https://www.linkedin.com/in/nursnaaz/
https://github.com/nursnaaz

BIGDATA (HADOOP & SPARK)





INTRODUCTION TO BIGDATA, HADOOP & SPARK

- Introduction to Big data
- Classification
- Characteristics
- Different methods of analyzing bigdata
- Hadoop HDFS & Mapreduce overview.

Hadoop HDFS commands Exercise

SQOOP – HDFS – HADOOP DATA EXTRACTION/ACQUISITION

Introduction - Why Sqoop - Import Architecture - Sqoop Export Architecture

Sqoop import/export use cases hands-on.

HIVE – SQL & OLAP Layer on Hadoop

Introduction – Architecture - Hive Vs RDBMS - Create Tables (Managed, external) - Managed Vs External tables

Exercise: Hive access through Hive Client - Partitioning (static and dynamic)

Py-Spark Essentials

Overview – Daemons - A Spark Standalone Cluster - Components & Terminologies - Workers, Driver Programs, Tasks - Executors & Cluster Manager

Working Pyspark (Handson)

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

DATA VISUALIZATION

TABLEAU

- Connecting to data
- Trend lines, reference lines and statistical analysis in tableau
- Data Strategy
- Native support for databases
- Joining multiple tables in tableau
- Custom SQL
- 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

LEAD TO SUCCESS BY JOINING inceptez