

# Data Science Course Content :

## **1:Introduction Data science & Business Analytics**

- Data Science and Business analytics
- Introduction to Advanced Data Analytics
- Charts for Data Science and Business Analytics üHadoop for Data Science

## **2:Descriptive Statistics**

- Descriptive Statistical
- Inferential Statistics
- Types of Variables
- Measures of central tendency
- Data Viability Dispersion
- Five number Summary Analysis
- Data Distribution Techniques
- Exploration Techniques for Numerical and Character data
- Summary and Visualization Exploration

## **3.Basic Probability for Business issues**

- Simple
- Marginal
- Joint
- Conditional
- Bayes' Theorem

## **4:Basic Distributions**

- Discrete
- Binomial

- Hyper geometric
- Poisson
- Continuous
- Normal
- Standardized

### **5. Sampling Technique Big Data**

- Sampling Distributions
- Simple Random
- Systematic Sample
- Cluster Sample
- Standard Error of the Mean
- Skewed Std. Error
- Kurtosis Std. Error
- Sampling from Infinity
- Sampling Distributions for Mean
- Sampling Distributions for proportions Theorem's

### **6: Data Validation & Data Normality**

- Stem and leaf analysis
- Univariate normality techniques
- Multivariate techniques
- Q-Q probability plots
- Cumulative frequency
- Explorer analysis
- Histogram
- Box plot

- Scores for Normality Check

- Testing

### **7: Data cleaning process Quality check**

- PCA for Big Data Analysis or Unsupervised data üPCA Regression  
Scores for Supervised data üNoise Data detecting
- Data cleaning with Regression Residual üData scrubbing with statistical sense

### **8: Data Imputation and outlier treatment**

- Outlier treatment with central tendency Mean
- Outlier with Min Max
- Outlier Detection
- Visualize Outlier Treatment
- Summarized Outlier Treatment
- Outlier with Residual Analysis
- Outlier Detection with PCA Analysis
- Data Imputation with series Central Tendency

### **9: Test of Hypothesis**

- Null Hypothesis formulation
- Alternative Hypothesis
- Type I and Type II errors
- Power Value
- One tail and two tail
- T-TEST's
- ANOVA
- MANOVA
- Chi Square Test

- Kendall Chi Square
- Kruskal-Wallis Rank Test Chi Square
- Mann-Whitney, Chi Square
- Wilcoxon, Chi Square

### **10: Data Transformation**

- Log, Arcsine, Box- Cox, Square root Inverse and Data normalization

### **11: Predictive modeling & Diagnostics**

- Correlation ü Regression
- Examination Residual analysis ü Auto Correlation
- Test of ANOVA Significant ü Homoscedasticity ü Heteroskedasticity  
ü Multicollinearity
- Cross validation
- Check prediction accuracy.

### **12: Logistic Regression Analysis**

- Logistic Regression
- Discriminate Regression Analysis Multiple Discriminate Analysis Stepwise  
Discriminate Analysis Logic function
- Test of Associations
- Chi-square strength of association, Binary Regression Analysis
- Estimation of probability using logistic regression, Hosmer Lemeshow
- nagelkerke R square
- Pseudo R square
- Model Fit
- Model cross validation
- Discrimination functions

### **13: Big Data Analytics**

- Introduction to Factor Analysis
- Principle component analysis
- Reliability Test
- KMO MSA tests, etc..
- Rotation and Extraction steps
- Conformity Factor Analysis
- Exploratory Factor Analysis
- Factor Score for Regression

#### **14:Cluster Analysis and Methods**

- Introduction to Cluster Techniques
- Hierarchical clustering
- K Means clustering
- Wards Methods
- Agglomerative Clustering
- Variation Methods
- Maximum distance Linkage Methods
- Centroid distance Methods
- Minimum distance Linkage Method
- Cluster Dendrogram
- Euclidean distance

#### **15:Data Mining Machine Learning and Artificial Intelligence**

- Prediction
- Support Vector Machines
- Gaussian Models
- Neural Network

- Classification Models
- Ordinal Regression
- Multinomial Regression
- Discriminate analysis
- Simple Cluster
- Hierarchical Cluster

### **16:Time series**

- Auto Regression, Moving Average, Multiplicative, ARMA, Additive Model

### **17:Model Validation and Testing**

- AIC, BIC, Kappa Statistics, ROC, APE, MAPE, Lift Curve, Errors

### **18: Hadoop Ecosystem**

- Pig,Hive,Map Reduce,NoSQL,etc