Data Analytics With Python

Free Python

Numpy

Pandas

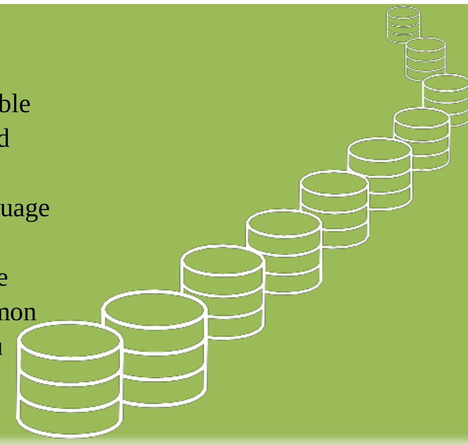
Matplotlib

Statistics

EAGLETFLY SOLUTIONS

981434634 / 9711198265

Data analyst is one of the hottest professions of the time. Learning Python is easy for any IT based student. Here in this article you are going to learn how Python is helpful for data analysis. Python is in trend these days and its community support is tremendous. Once you are a Python expert, you will be able to solve any data analysis problem with an ease. All you need is get complete knowledge of Python and study Python with complete dedication.
Python is gaining interest in IT sector and the top IT students opt to learn Python as their choice of language for learning data analysis. The candidates want to jump into the career of a data analyst must have knowledge about some language and if we compare Python with other languages, Python is much more interesting and easy to learn as compared to other programming languages. Thus, it has become a common language for data analysis. Python is easy to learn and use whether you are new to the language or you are an experienced professional of information technology. Python helps you serve the company as a great data analyst.



Python

1. An introduction to python:

- Brief History
- Why Python
- Where to use

2. Beginning python basics:

- The print statement
- Comments
- Python Data Structures & Data Types
- String Operations in Python
- Simple Input & Output
- Simple Output Formatting

3. Python program flow:

- Indentation
- The If statement and its' related statement
- An example with if and it's related statement
- The while loop
- The for loop
- The range statement
- Break & Continue
- Assert
- Examples for looping

4. Functions & modules:

- Create your own functions
- Functions Parameters
- Variable Arguments
- Scope of a Function
- Function Documentation/Doc strings
- Lambda Functions & map
- An Exercise with functions
- Create a Module
- Standard Modules

5. File handling:

- File Handling Modes
- Reading Files
- Writing & Appending to Files
- Handling File Exceptions
- The with statement

7. Class:

- New Style Classes
- Variable Type
- Static Variable in class
- Creating Classes
- Instance Methods
- Inheritance
- Polymorphism
- Encapsulation
- Scope and Visibility of Variables
- Exception Classes & Custom Exceptions

8. Data structures:

- List Comprehensions
- Nested List Comprehensions
- Dictionary Comprehensions
- Functions
- Default Parameters
- Variable Arguments
- Specialized Sorts
- Integrators
- Generators
- The Functions any and all
- The with Statement
- Data Compression
- Closer
- Decorator

Data Science

1. Numpy Basics

- What is Numpy ?
- Creating arrays from python objects
- Printing arrays
- Universal functions
- Indexing, slicing and selection
- Fancy indexing
- Broadcasting arrays
- arrays from python functions
- Mathematics operations
- Indexing a 2D array
- Practicals
- Finding patterns

2. Importing and Exporting Data using Pandas

- Importing Data from various sources (Csv, txt, excel, access etc)
- Database Input (Connecting to database)
- Viewing Data objects - subsetting, methods
- Exporting Data to various formats
- Important python functions: Pandas

3. Data Manipulation – cleansing – Munging using Python

- Cleansing Data with Python
- Data Manipulation steps(Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc)
- Data manipulation tools(Operators, Functions, Packages, control structures, Loops, arrays etc)
- Python Built-in Functions (Text, numeric, date, utility functions)
- Stripping out extraneous information Normalizing data
- Formatting data
- Important Python modules for data manipulation (Pandas, Numpy, math, string, datetime etc)

4. Data Analysis – Visualization using Python

- Introduction exploratory data analysis
- Descriptive statistics, Frequency Tables and summarization
- Univariate Analysis (Distribution of data & Graphical Analysis)
- Bivariate Analysis(Cross Tabs, Distributions & Relationships, Graphical Analysis)
- Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ density etc)
- Important Packages for Exploratory Analysis(NumPy Arrays, Matplotlib, seaborn, Pandas etc)

Probability and Statistics

1. Plotting for exploratory data analysis (EDA)

- Iris dataset
- Data-point, vector, observation Dataset
- Input variables/features/dimensions/independent variable
- Output Variable/Class Label/ Response Label/ dependent variable
- Scatter-plot: 2D, 3D.
- Pair plots.
- PDF, CDF, Univariate analysis.
- Histogram and PDF
- Univariate analysis using PDFs.
- Cumulative distribution function (CDF)
- Mean , Variance, Std-dev
- Median, Percentiles, Quantiles, IQR, MAD and Outliers.
- Box-plot with whiskers
- Violin plots.
- Summarizing plots.
- Univariate, Bivariate and Multivariate analysis.
- Multivariate probability density, contour plot.
- Exercise: Perform EDA on Iris dataset.

2. Probability and Statistics

- Introduction to Probability and Stats
- Why learn it?
- P(X=x1) , Dice and coin example
- Random variables: discrete and continuous.
- Outliers (or) extreme points.
- Population & Sample.
- Gaussian/Normal Distribution
- Examples: Heights and weights.
- Why learn about distributions.
- Mu, sigma: Parameters
- Symmetric distribution, Skewness and Kurtosis
- Standard normal variate (z) and standardization.
- Kernel density estimation.
- Sampling distribution & Central Limit theorem.
- Q-Q Plot: Is a given random variable Gaussian distributed?

3. Probability - 2.

- Uniform Distribution and random number generators
- Discrete and Continuous Uniform distributions.
- How to randomly sample data points.
- Bernoulli and Binomial distribution
- Log-normal and power law distribution:
- Log-normal: CDF, PDF, Examples.
- Power-law & Pareto distributions: PDF, examples
- Converting power law distributions to normal: Box-Cox/Power transform.

4. Inferential Statistics

- Correlation
- Co-variance
- Pearson Correlation Coefficient
- Spearman Rank Correlation Coefficient
- Correlation vs Causation
- Confidence Intervals
- Confidence Interval vs Point estimate.
- Computing confidence-interval given a distribution.
- For mean of a random variable
- Known Standard-deviation: using CLT
- Unknown Standard-deviation: using t-distribution
- Confidence Interval using empirical bootstrap
- Hypothesis testing
- Hypothesis Testing methodology, Null-hypothesis, test-statistic, p- value.
- Resampling and permutation test.
- K-S Test for similarity of two distributions.

Machine Learning with Scikit Learn

1. Introduction to Predictive Modeling

- Introduction to Predictive Modeling
- Types of Business problems - Mapping of Techniques - Regression vs. classification vs. segmentation vs. Forecasting
- Major Classes of Learning Algorithms -Supervised vs Unsupervised Learning
- Different Phases of Predictive Modeling (Data Pre-processing, Sampling, Model Building, Validation)
- Overfitting (Bias-Variance Trade off) & Performance Metrics
- Feature engineering & dimension reduction
- Concept of optimization & cost function
- Overview of gradient descent algorithm
- Overview of Cross validation(Bootstrapping, K-Fold validation etc)
- Model performance metrics (R-square, Adjusted R-squre, RMSE, MAPE, AUC, ROC curve, recall, precision, sensitivity, specificity, confusion metrics)

2. Data Exploration for modeling

- Need for structured exploratory data
- EDA framework for exploring the data and identifying any problems with the data (Data Audit Report)
- Identify missing data
- Identify outliers data
- Visualize the data trends and patterns
- Consolidation/Aggregation - Outlier treatment - Flat Liners - Missing values- Dummy creation - Variable Reduction
- Variable Reduction Techniques - Factor & PCA Analysis

3. Linear Regression: Solving regression problems

- Introduction - Applications
- Assumptions of Linear Regression
- Building Linear Regression Model
- Understanding standard metrics (Variable significance, R-square/Adjusted R-square, Global hypothesis ,etc)
- Assess the overall effectiveness of the model
- Validation of Models (Re running Vs. Scoring)
- Standard Business Outputs (Decile Analysis, Error distribution (histogram), Model equation, drivers etc.)
- Interpretation of Results - Business Validation - Implementation on new data

4. Logistic Regression: Solving classification problems

- Introduction - Applications
- Linear Regression Vs. Logistic Regression Vs. Generalized Linear Models
- Building Logistic Regression Model (Binary Logistic Model)
- Validation of Logistic Regression Models (Re running Vs. Scoring)
- Standard Business Outputs (Decile Analysis, ROC Curve, Probability Cut-offs, Lift charts, Model equation, Drivers or variable importance, etc)
- Interpretation of Results - Business Validation - Implementation on new data

5. Time Series Forecasting: Solving forecasting problems

- Introduction - Applications
- Time Series Components(Trend, Seasonality, Cyclicity and Level) and Decomposition
  Classification of Techniques(Pattern based - Pattern less) Basic Techniques - Averages, Smoothening, etc
  Advanced Techniques - AR Models, ARIMA, etc
  Understanding Forecasting Accuracy - MAPE, MAD, MSE, etc

8. Keras: A High Level API:

- Introduction to Keras
- Keras vs TFLearn
- Compose models in Keras
- Sequential and functional composition
- Predefined neural network layers
- Using inception v3 predefined model
- Batch normalization: Keras
- Saving and Loading a model with Keras
- Customizing the training process
- Tensor-board: With Keras
- Use-Case implementation with Keras

9. Tuning of Hyperparameters:

- Train, Dev & Test sets
- Bias & Variance
- Regularization & Overfitting
- Why regularization reduces overfitting?
- L1/L2 & Dropout Regularization
- Vanishing / Exploding gradients
- Weight Initialization for Deep Networks
- Numerical approximation of gradients
- Normalizing activations in a network
- Why does Batch Norm work?

10. Miscellaneous

- What are computation graph?
- derivatives with a computation graph
- Broadcasting in Numpy/Tensorflow
- Neural Network Representation
- Computing a Neural Network's Output
- Explanation for Vectorized Implementation
- Parameters vs Hyperparameters
- Forward and Backward Propagation

6. Supervised Learning: Decision Trees

- Decision Trees - Introduction - Applications
- Types of Decision Tree Algorithms
- Construction of Decision Trees through Simplified Examples;
- Choosing the "Best" attribute at each Non-Leaf node; Entropy;
- Information Gain, Gini Index, Chi Square, Regression Trees
- Generalizing Decision Trees;
- Information Content and Gain Ratio;
- Dealing with Numerical Variables;
- other Measures of Randomness
- Pruning a Decision Tree;
- Cost as a consideration; Unwrapping Trees as Rules
- Decision Trees - Validation
- Overfitting - Best Practices to avoid

7. Supervised Learning: Support Vector Machines

- Motivation for Support Vector Machine & Applications
- Support Vector Regression
- Support vector classifier (Linear & Non-Linear)
- Mathematical Intuition (Kernel Methods Revisited, Quadratic Optimization and Soft Constraints)
- Interpretation of Outputs and Fine tune the models with hyper parameters
- Validating SVM models

8. Supervised Learning: KNN

- What is KNN & Applications?
- KNN for missing treatment
- KNN For solving regression problems
- KNN for solving classification problems
- Validating KNN model
- Model fine tuning with hyper parameters

9. Supervised Learning: Ensemble Learning

- Concept of Ensembling
- Manual Ensembling Vs. Automated Ensembling
- Methods of Ensembling (Stacking, Mixture of Experts)
- Bagging (Logic, Practical Applications)
- Random forest (Logic, Practical Applications)
- Boosting (Logic, Practical Applications)
- Ada Boost
- Gradient Boosting Machines (GBM) XGBoost

Fee: 18,500 RS/-
Duration: 3 Months