

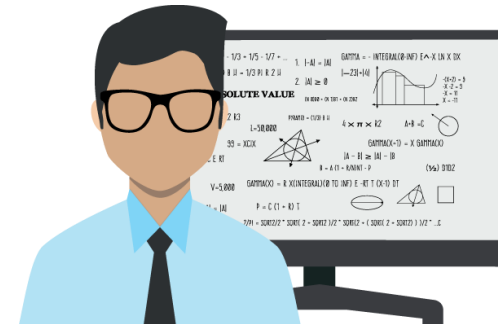
Basic Python

Learn by doing (90% hands on)



About Me

- Name: Shiv
- BSc and MSc in Mathematics from IIT Kharagpur
- PhD in Analytics
- Experience: 19 years (9 Years in Analytics and 10 years in IT)
- Current Role: Chief Data Scientist



Prerequisites

- Passion for learning
- Good Knowledge of any programming technology

Basic Python

Introduction

- Installation
- Environment Walkthrough
- How to get help
- Console and Editor
- History
- Working directory

Data structure

- Various types Variables
- Tensors (Vectors)
- Data frames
- Matrices
- List
- Dictionaries
- Using concat & merge
- Regular Expressions
- Handling dates
- Tuples
- Syntax Rules

Importing data

- Reading/Saving Data files
- Reading/Saving data objects

Manipulating Data

- Selecting rows/observations
- Selecting columns/fields
- Merging data
- Relabeling the column names
- Converting variable types
- Data sorting
- Data aggregation
- Pivot Tables

Using functions

- Mathematical Functions
- Summary Functions
- String Functions
- User defined functions
- Lambda functions

Programming operations

- Truth testing
- Branching
- Looping
- Jump commands
- Arithmetic operations

Charts and Plots

- Box plot
- Histogram
- Line chart
- Scatter plot
- Save images

Basic Statistics

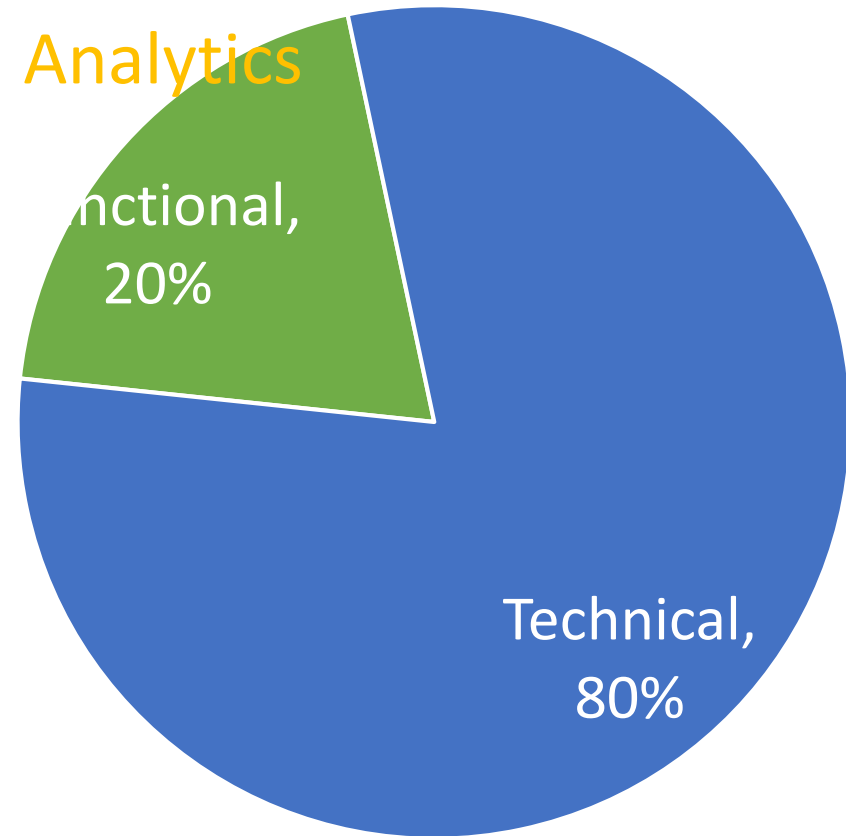
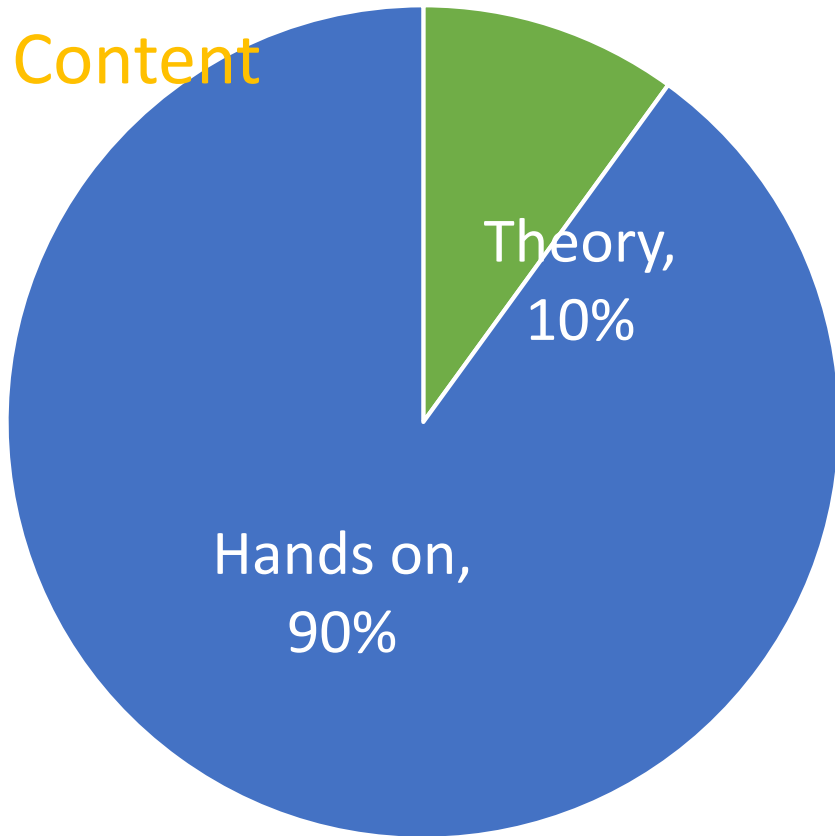
- Measures of Central Tendency
- Measures of Dispersion
- Measure of Shape
- Sampling
- Distributions
- Descriptive Statistics
- Inferential Statistics

Miscellaneous

- Logging the various types of messages
- Exceptional Handling in details
- Error messages
- How to get going in open source
- How to learn new module in open source

All are Hands on topics

Methodology



Introduction

- Installation
 - Software: software_list.txt
 - Libraries: install.txt
- Environment Walkthrough
 - <https://www.anaconda.com/download/>
- Works in RAM
- How to get help

Anaconda

- The open source version of Anaconda is an easy-to-install high performance Python with a package manager, environment manager and collection of 720+ open source packages
- `conda install PACKAGENAME` or `pip install pydotplus`
- Packages
 - i. NumPy | numpy.org: N-dimensional array for numerical computation
 - ii. Pandas | pandas.pydata.org: Powerful Python data analysis toolkit
 - iii. SciPy | scipy.org: Collection of numerical algorithms and toolboxes, including signal processing and optimization
 - iv. Matplotlib | matplotlib.org: Plotting library for Python
 - v. Seaborn | stanford.edu/~mwaskom/software/seaborn/: Statistical data visualization
 - vi. Bokeh | bokeh.pydata.org: Interactive web visualization library
 - vii. SciKit-Learn | scikit-learn.org/stable: Python modules for machine learning and data mining
 - viii. NLTK | nltk.org: Natural language toolkit

Type of Data/Variables

- Numeric data
 - Continuous (measurements)
 - Ratio (interval + clear definition of 0)
 - Interval (difference between two values is meaningful)
 - Discrete (counts)
- Categorical
 - Nominal
 - Ordinal (Likert scale)
 - Dichotomous
- Independent Variables (experimental or predictor)
- Dependent (outcome)

	Nominal	Ordinal	Interval	Ratio
Frequency distribution	Yes	Yes	Yes	Yes
Median and percentiles	No	Yes	Yes	Yes
Add or subtract	No	No	Yes	Yes
mean, std	No	No	Yes	Yes
Ratio	No	No	No	Yes

Basics Hands on

References

- <https://regex101.com/>
- <https://docs.python.org/2/library/re.html>
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.nervanasys.com%2Fwp-content%2Fuploads%2F2016%2F04%2FScreen-Shot-2016-04-27-at-10.59.50-AM.png&imgrefurl=https%3A%2F%2Fwww.nervanasys.com%2Fopenai%2F&docid=GggjYZ3JzfRoVM&tbnid=teq-_Mt_keBQiM%3A&vet=10ahUKEwiDqYXb-7_UAhWLS48KHd2dD00QMwh2KC8wLw..i&w=758&h=423&bih=632&biw=1366&q=Reinforced%20learning&ved=0ahUKEwiDqYXb-7_UAhWLS48KHd2dD00QMwh2KC8wLw&iact=mr&uact=8
- <https://images.google.com/>
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.nervanasys.com%2Fwp-content%2Fuploads%2F2016%2F04%2FScreen-Shot-2016-04-27-at-10.59.50-AM.png&imgrefurl=https%3A%2F%2Fwww.nervanasys.com%2Fopenai%2F&docid=GggjYZ3JzfRoVM&tbnid=teq-_Mt_keBQiM%3A&vet=10ahUKEwiDqYXb-7_UAhWLS48KHd2dD00QMwh2KC8wLw..i&w=758&h=423&bih=632&biw=1366&q=Reinforced%20learning&ved=0ahUKEwiDqYXb-7_UAhWLS48KHd2dD00QMwh2KC8wLw&iact=mr&uact=8
- <https://images.google.com/>
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fcdn.edureka.co%2Fblog%2Fwp-content%2Fuploads%2F2013%2F06%2FData-Scientist.jpg&imgrefurl=https%3A%2F%2Fwww.edureka.co%2Fblog%2Fwho-is-a-data-scientist%2F&docid=q3ij004thhLBIM&tbnid=6_eIIKVKkQtWIM%3A&vet=10ahUKEwi74J_L7sbUAhWHRY8KHycUBbUQMwi8AShFMEU..i&w=601&h=351&bih=632&biw=1366&q=Data%20Science&ved=0ahUKEwi74J_L7sbUAhWHRY8KHycUBbUQMwi8AShFMEU&iact=mr&uact=8
- [https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))
- <http://www.chioka.in/differences-between-roc-auc-and-pr-auc/>
- https://www.google.com/imgres?imgurl=http%3A%2F%2Fstanford.edu%2F~cpiech%2Fcs221%2Fimg%2FkmeansViz.png&imgrefurl=http%3A%2F%2Fstanford.edu%2F~cpiech%2Fcs221%2Fhandouts%2Fkmeans.html&docid=xBG90gMlnM_nKM&tbnid=0dV6dbzMcO1mgM%3A&vet=10ahUKEwiv8PWx49HUAhXKtI8KHV3BigQMwhSKAUwBQ..i&w=501&h=338&bih=632&biw=1366&q=K-means&ved=0ahUKEwiv8PWx49HUAhXKtI8KHV3BigQMwhSKAUwBQ&iact=mr&uact=8
- <https://www.otexts.org/>
- robjhyndman.com/hyndsight/forecasting/
- OTexts.org/fpp
- https://s3.amazonaws.com/MLMastery/MachineLearningAlgorithms.png?_s=s9ac4vp2drr4d1dfcru1
- <http://www.techjini.com/wp-content/uploads/2017/02/mc-learning.jpg>
- <https://machinelearningmastery.com/improve-deep-learning-performance/>

In case you want to learn ML using Python online

Data Science for AI and Machine Learning Using Python

Become Data Science (Machine Learning) professional by learning from Data Science professional

NEW



- Course Name: Data science for AI and Machine learning using Python
- Course Link: <https://www.udemy.com/data-scientist-for-ai-and-machine-learning-using-python/?couponCode=AIMLPY>
- Why this course: Targeted to corporate requirements (different from traditional academic) so that people have the basic understanding of AI and ML using Python
- English Subtitle: Yes
- Cost: \$10 (93% discount with Coupon AIMLPY)
- Price validity: Few days as per TnC of Udemy

Deep Learning by TensorFlow (tf.keras) & Keras using Python

Become Deep Learning professional by learning from Deep Learning professional



- Course Name: Deep Learning by TensorFlow (tf.keras) & Keras using Python
- Course Link: <https://www.udemy.com/deep-learning-by-tensorflow-tfkeras-keras-using-python/?couponCode=DLPY010>
- Why this course: Targeted to corporate requirements (different from traditional academic) so that people have the good understanding of DL using Python
- Cost: \$10 (93% discount with Coupon AIMLPY)
- Price validity: Few days as per TnC of Udemy

Regular expression (<https://docs.python.org/2/library/re.html>)

- The basic flags are I, L, M, S, U, X:
 - re.I: This flag is used for ignoring casing
 - re.M: This flag is useful if you want to find patterns throughout multiple lines
 - re.L: This flag is used to find a local dependent
 - re.S: This flag is used to find dot matches
 - re.U: This flag is used to work for unicode data
 - re.X: This flag is used for writing regex in a more readable format
- re.match(): It finds the pattern at the beginning of the input string
- re.search(): It match the string anywhere and finds all the occurrences of the pattern in the given input string or data.

Regular expression cont

- Find the single occurrence of character a and b: Regex: [ab]
- Find characters except a and b: Regex: [^ab]
- Find the character range of a to z: Regex: [a-z]
- Find range except to z: Regex: [^a-z]
- Find all the characters a to z as well as A to Z: Regex: [a-zA-Z]
- Any single character: Regex: .
- Any whitespace character: Regex: \s
- Any non-whitespace character: Regex: \S
- Any digit: Regex: \d
- Any non-digit: Regex: \D
- Any non-words: Regex: \W
- Any words: Regex: \w

- `r'^(\d{4})'`: find any four digits at the beginning of a string
 - `extr = df['Date'].str.extract(r'^(\d{4})', expand=False)`