

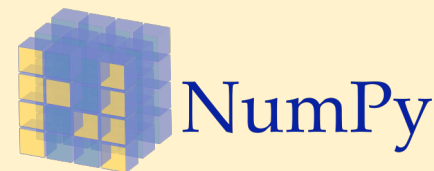
Learn Data Science For
Free



Programming Language

1. Variables, numbers
2. Strings, lists, dictionaries
3. For loop, if condition
4. Functions
5. Read/write file
6. Basics understanding of modules

Data Cleaning and Exploration



1. Series and dataframe basics
2. Create dataframe from csv/excel/json
3. Handle missing data, fillna, dropna
4. groupby
5. merge, concat dataframes



Visual Studio Code



IDE: Integrated Development Environment

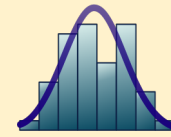
Data Visualization



1. Line chart, bar chart, pie chart
2. Histograms
3. Axes labels, legend, grid
4. Scatter plot



1. Basics of relational database and structured query language (SQL)
2. Query and update data: select, where, insert, update clause
3. Groupby, orderby, joins, views



Math & Statistics

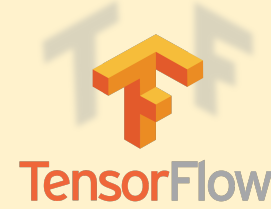
1. Linear algebra
2. Mean, median, normal distribution, standard deviation, mode, variance, range, IQR, correlation
3. Probability, conditional probability, priors, posteriors, maximum likelihood
4. Descriptive statistics, inferential statistics, regression, hypothesis testing, A/B testing

Machine Learning



1. Regression: Linear regression
2. Classification: svm, decision tree, random forest, logistic regression, naïve bayes
3. Unsupervised learning: k means
4. Label and one hot encoding
5. `train_test_split`
6. K fold cross validation
7. `GridSearchCV`

Deep Learning



1. Neural network basics
 - a. Input, hidden, output layers
 - b. Activation function
2. Convolutional neural networks – CNN
3. Recurrent neural networks - RNN

Data Analysis Tool



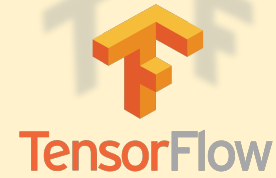
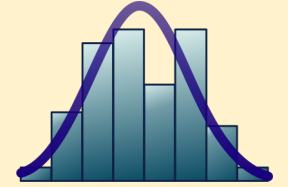
1. Data filters, functions
2. Formulas
3. Charts and plots
4. Pivot table, transpose
5. Vlookup

Allows you to do quick data analysis without having to write code



numpy

Seaborn





Books

- Statistics - Think stats
<http://greenteapress.com/thinkstats2/thinkstats2.pdf>
- Python – Automate Boring Stuff
<https://automatetheboringstuff.com/>
- Python – Data cleaning, exploration, visualization and machine learning
<https://jakevdp.github.io/PythonDataScienceHandbook/>

Online free courses

- <https://www.edx.org/course/python-for-data-science-2>
- <https://www.edx.org/course/data-science-essentials>
- <https://www.coursera.org/specializations/deep-learning>
- <https://cognitiveclass.ai/learn/data-science>
- <http://work.caltech.edu/telecourse>