



Data Science Engineering Toolset



Roman Golovnya

28 April 2018



Agenda:

- Programming Languages:
- Integrated Development Environments (IDE)
- Command Line / Bash Scripting
- Python virtual environments
- Version control & repository systems
- Virtualization - Containerization
- Cloud Resources
- Collaboration tools
- Data Science Engineering Fundamentals program

Programming Languages for Data Science:

Python

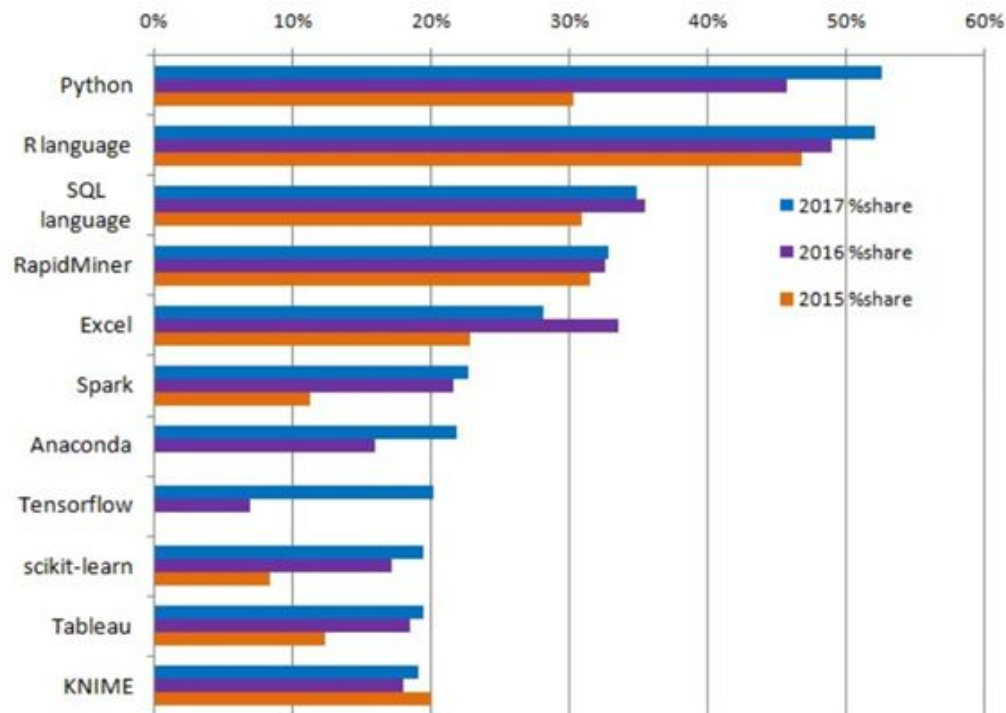
R

SQL

Scala

Javascript for web app development and visualisations.

KDnuggets Analytics, Data Science, Machine Learning Software Poll, top tools share, 2015-2017



<https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html>

Programming Languages for Data Science Engineering:



Commercial Software



Gartner 2018 Magic Quadrant for DS and ML

Integrated Development Environments (IDE)

Jupyter notebook, JupyterHub, JupyterLab <http://jupyter.org/>
<https://nbviewer.jupyter.org/>

Apache Zeppelin notebook <https://zeppelin.apache.org/>

PyCharm <https://www.jetbrains.com/pycharm/>

Atom <https://atom.io/>

VIM <https://www.vim.org/>

RStudio <https://www.rstudio.com/>

Command Line / Bash Scripting is a way of interacting with your computer operating system using text-based commands.

Command Line

<https://www.computerhope.com/issues/chusedos.htm>

Bash Scripting

<https://linuxconfig.org/bash-scripting-tutorial>

Python virtual environments and package management system

Virtualenv vs Conda

<https://virtualenv.pypa.io/en/stable/>

<https://packaging.python.org/guides/installing-using-pip-and-virtualenv/>

<https://conda.io/docs/>

Version control & repository systems :

- Git <https://git-scm.com/>
- Github <https://github.com/>
- Gitlab <https://gitlab.com/>
- Bitbucket <https://bitbucket.org>

Virtualization - Containerization

VM VirtualBox Oracle

Docker

Cloud Resources:

Amazon Web Services (AWS)

Microsoft Azure

Google Cloud Platform

IBM Cloud

Collaboration tools

Slack

<https://dseclub.slack.com>

https://join.slack.com/t/dseclub/shared_invite/enQtMjQxNTI1NDI3MjQ4LTFIZTA0NjVINzNiMzI0MmY0YWJkNjMwYTZiY2NmMGJkMWM5NDExZTgyZTdhYzQ4NWZlYzFkNDI3ZmZjZTAyMDA

Zoom

<https://www.zoom.us/>

Skype

<https://www.skype.com/en/>

Data Science Engineering Fundamentals

Part 1 - Foundation

- DSE Toolset, Git, Docker
- Data Wrangling - Programming
- Intro to Cloud Computing for DS
- Applied Math, Probability, Statistics for DS
- Exploratory Data Analysis & Visualization
- Data Management Systems

Part 2 - Specialization

- Big Data Analytics - Spark
- Data Engineering
- Building Web Applications
- Advanced Visualisations
- Machine Learning
- Natural Language Processing
- Deep Learning
- Projects

Resources:

<https://www.ngdata.com/top-tools-for-data-scientists/>

<https://www.coursera.org/>

<https://eu.udacity.com/>

<https://www.datacamp.com/>

<https://www.edx.org/>

<http://www.deeplearningbook.org/>

<https://www.kdnuggets.com/>

<https://www.datasciencecentral.com/>

<https://medium.com/>

We need
you to join
our crew!



Contact

Via meetup.com/messages/

<https://dseclub.slack.com>

roman.golovnya@gmail.com



Questions?

Thank You!