

Data Science with Python and Friends

Douglas Starnes

DevSpace

October 14, 2016

Why Data Science?

An interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured.

The study of where information comes from, what it represents and how it can be turned into a valuable resource in the creation of business and IT strategies

Data science involves using automated methods to analyze massive amounts of data and to extract knowledge from them.

Data Science Skills

- Software development
- Databases
- Math
- Business / Domain
- Storytelling
- Aesthetic
- Soft Skills
- Etc.

Ask 10 different data
scientists to define 'data
science' and ...

(at least!)
You'll get [^]20 different
answers.

Why Python?

Easy to learn and simple to use

General purpose

Large number of practitioners are not computer scientists

Cross platform

Active community

Open source

Tooling

Friends of Python and of Data Science

numpy

pandas

matplotlib / seaborn

IPython / Jupyter

scikit-learn

Julia

vim / emacs / Atom / PyCharm

PostgreSQL / MongoDB / Hadoop / Redis

Spark

How do you get it?

python.org

Package manager

Windows: chocolatey

macOS: home-brew

Linux: apt, yum, etc.

Anaconda

Docker container

```
docker pull continuumio/anaconda
```

APP	Notebooks Embeddable Dashboards Data Services Visual Apps							
VIZ	Plots	Interactive Viz	Big Data	Maps & GIS	3D	Streaming	Graphs	
STORYBOARD	Notebooks	Interactive Exploration		Visual Programming		Data IDEs		
ANALYTICS	Data Prep	Stats	ML & Ensembles		Deep Learning		Simulation & Optimization	
	Geospatial	Text & NLP		Video/Image/Audio Mining		Graph & Network		
DATA	Hadoop & Hive		Spark	NoSQL	DW & SQL	Files & Web Services		
HW	Servers	Clusters	GPUs & High End Workstations					

DEMOS

Thank You!

douglas@douglasstarnes.com
@poweredbyaltnet
douglasstarnes.com