

Basic AI with Python

DONI RUBIAGATRA
DATA SCIENCE INDONESIA JATIM

About Me

- **Doni Rubiagatra, @rubiagatra**
- B.Sc in **Mathematics** at **ITS Surabaya**
- Co-Founder **surabaya.py**, Internal **DSI Jatim**
- Machine Learning Engineer **Trafizap**



2.2
MILES

300
CAL

2,324
STEPS

3
HOMERUNS

0.300
BATTING
AVERAGE

.520
ON BASE
PERCENTAGE



3
HOMERUNS

0.300
BATTING
AVERAGE

.520
ON BASE
PERCENTAGE





1.5 miles to next turn

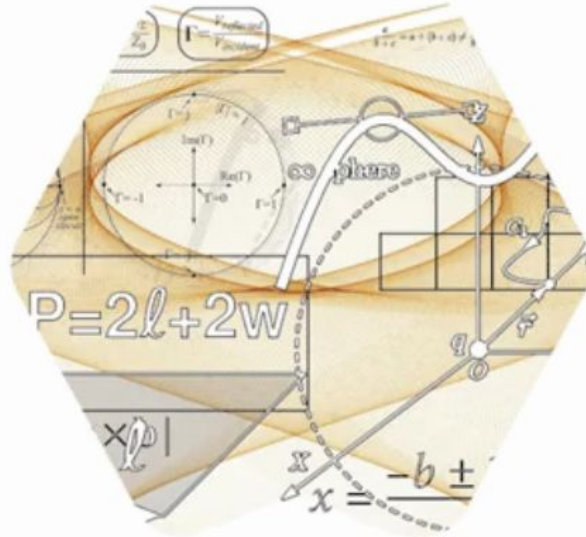
35 miles
to destination

ETA:
11:32 a.m.

What is AI? and Why You Should Care



Lots of data



Complex mathematics in multidimensional spaces



Magical results



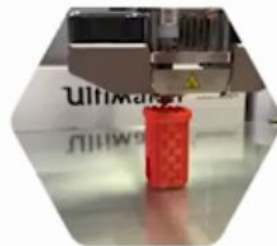
Collect
data



Organize
data



Create
model



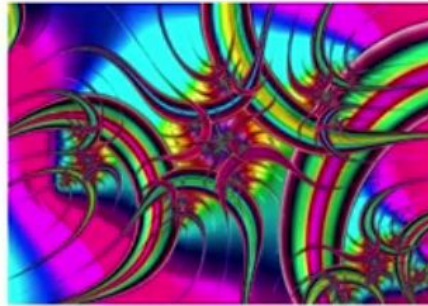
Use machines to
flesh out the
model from data



Deploy fleshed
out model



data



algorithm



insight

PRODUCT

Data Scientist

Full Time

APPLY

Responsibilities

- Analyzing complex, incomplete, and large data, from internal and external databases
- Providing insights based on data, includes data visualization, inference statistics, predictive modeling, machine learning, and presentation
- Extracting data from RDBMS, Non RDBMS, and external data to make dashboards for other divisions
- Improving and monitoring product features based on performance data

Requirements

- Bachelor or Master degree in Mathematics, Statistics, Computer Science, or related fields
- Experienced using RDBMS and Non RDBMS (MySQL, PostgreSQL, MongoDB)
- Experienced using data analysis tools, i.e: Excel (Spreadsheet), R, Python, etc
- Having passion in databases, data analysis, statistics, modeling, and machine learning
- Minimum 2 years experience in the same field are preferred



[← Back to Technology](#)

Data Scientist

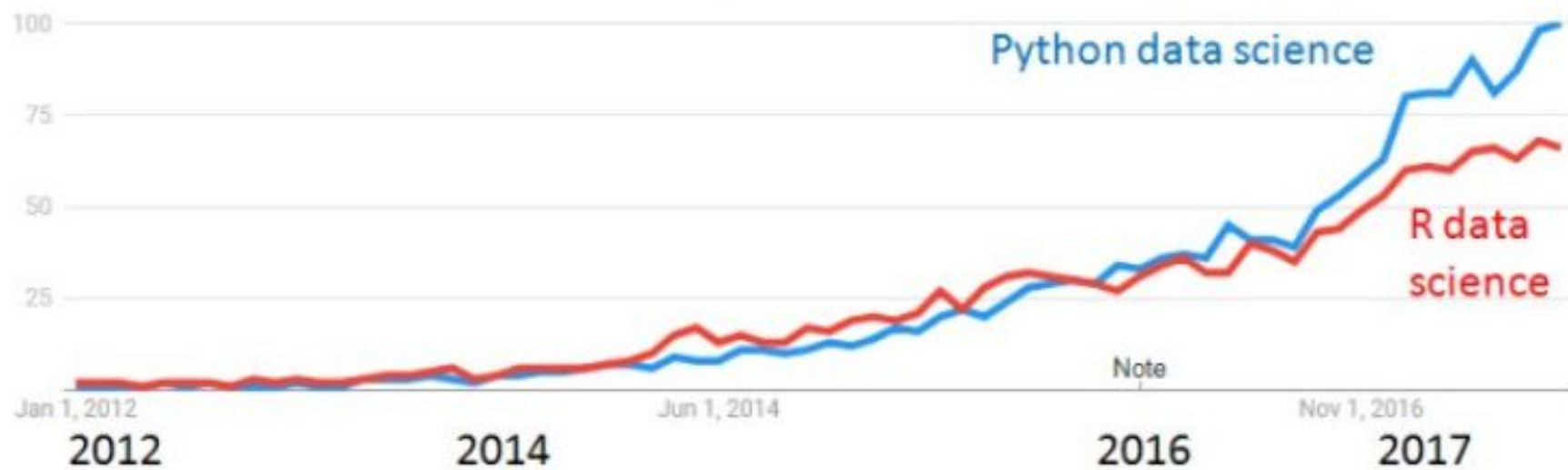
The focus of our data science team is to work on highly scalable data-driven products that have massive impact on the business. If you are passionate in finding patterns in big data and

Requirements

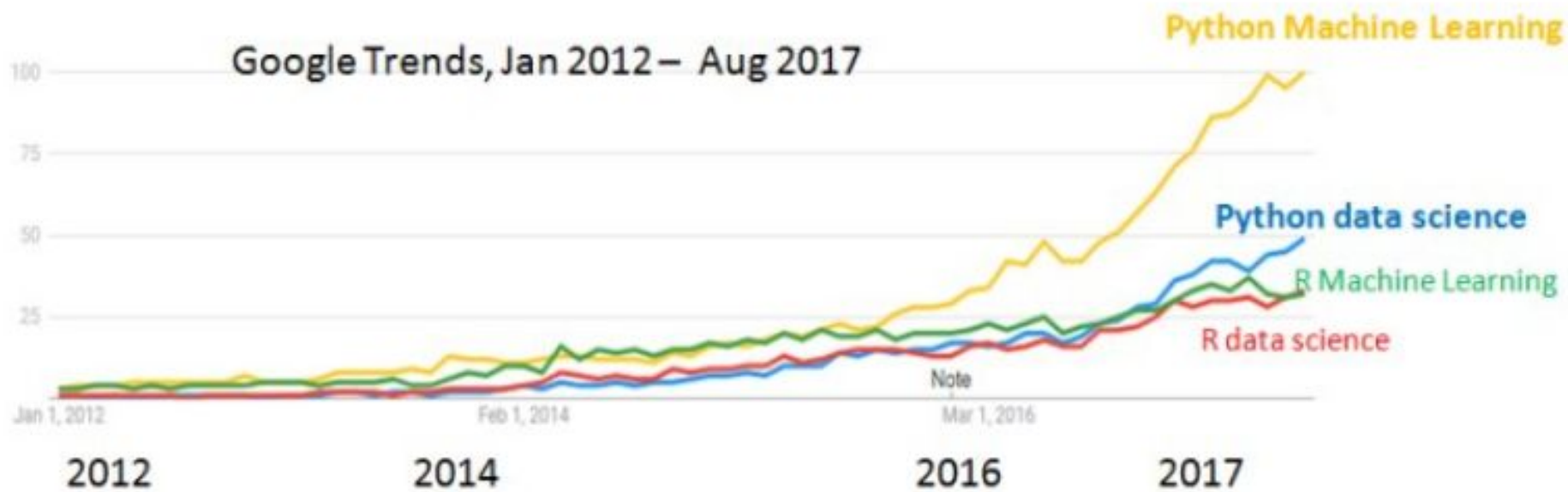
- Minimum Bachelor's degree in quantitative field, e.g. Math, Physics, and Computer Science
- Minimum 1 year of working experience as Data Analyst, Data Scientist, Data Engineer, or similar position
- Have passion in big data, data mining, and machine learning
- Proficient in programming language, i.e : Python, R, Scala, etc
- Have natural curiosity and desire to learn
- Proficient in several modeling techniques including but not limited to: Clustering, SVM, Linear/Non-Linear Modeling, Random Forests, ANN, etc.
- High spirit in model testing and validation to create accurate and consistent models
- Comfortable in mining unstructured dataset and processing it to become usable dataset
- Familiar with relational/NoSQL database, SQL, and ETL process
- Good communication skill for both oral and written
- Have knowledge in big data tools such as Hadoop and Spark is a plus
- Participation in Kaggle is a plus

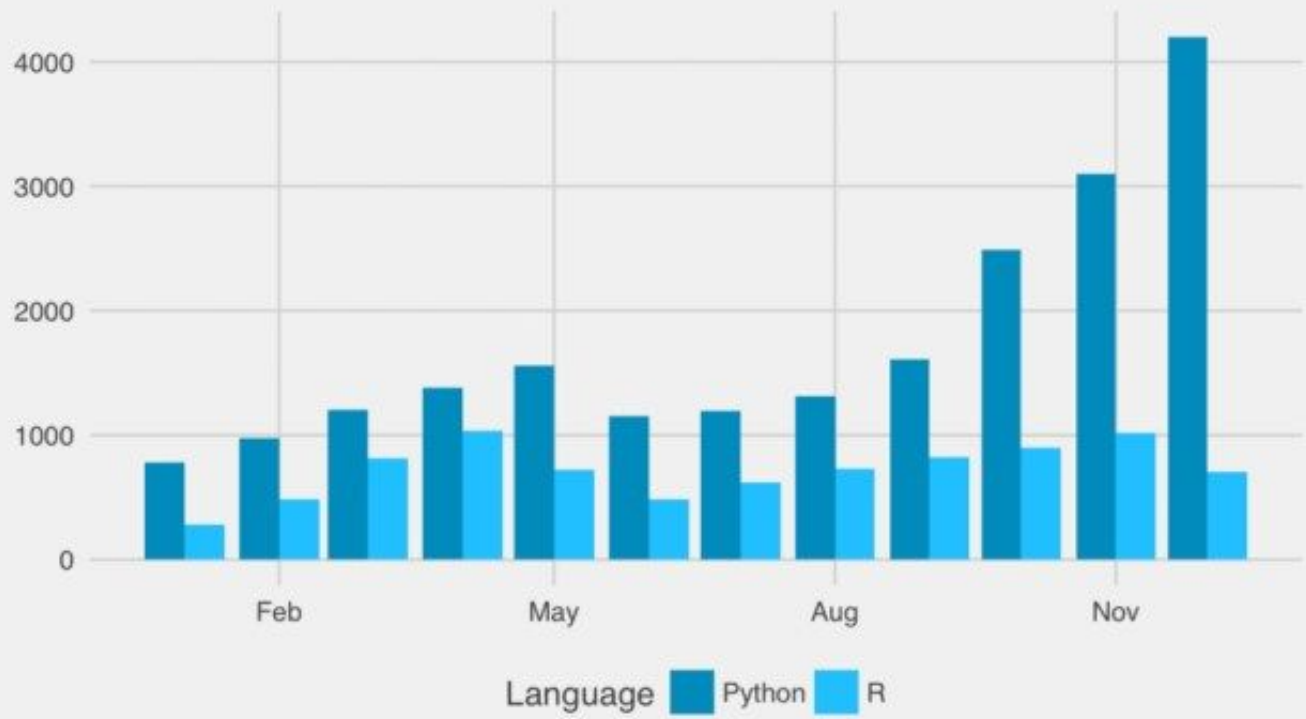
Why Python is the Best Language for AI?

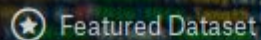
Google Trends, Jan 2012 – Aug 2017



Google Trends, Jan 2012 – Aug 2017







Featured Dataset

509

Kaggle ML and Data Science Survey, 2017

A big picture view of the state of data science and machine learning.



Kaggle · last updated 2 months ago

[Overview](#)

[Data](#)

[Kernels](#)

[Discussion](#)

[Activity](#)

[Download \(4 MB\)](#)

[New Kernel](#)

Tags

[data analysis](#)

[artificial intelligence](#)

[sociology](#)

[employment](#)

[medium](#)

[featured](#)

Top Contributors



Mhamed Jabri

1st



I,Coder

2nd

Kernels



Novice to Grandmaster

run 3 days ago

278

votes

What do Kagglers say about ...

run 4 days ago

263

votes

Discussion



For goodness' sake ...

8 days ago

0

replies

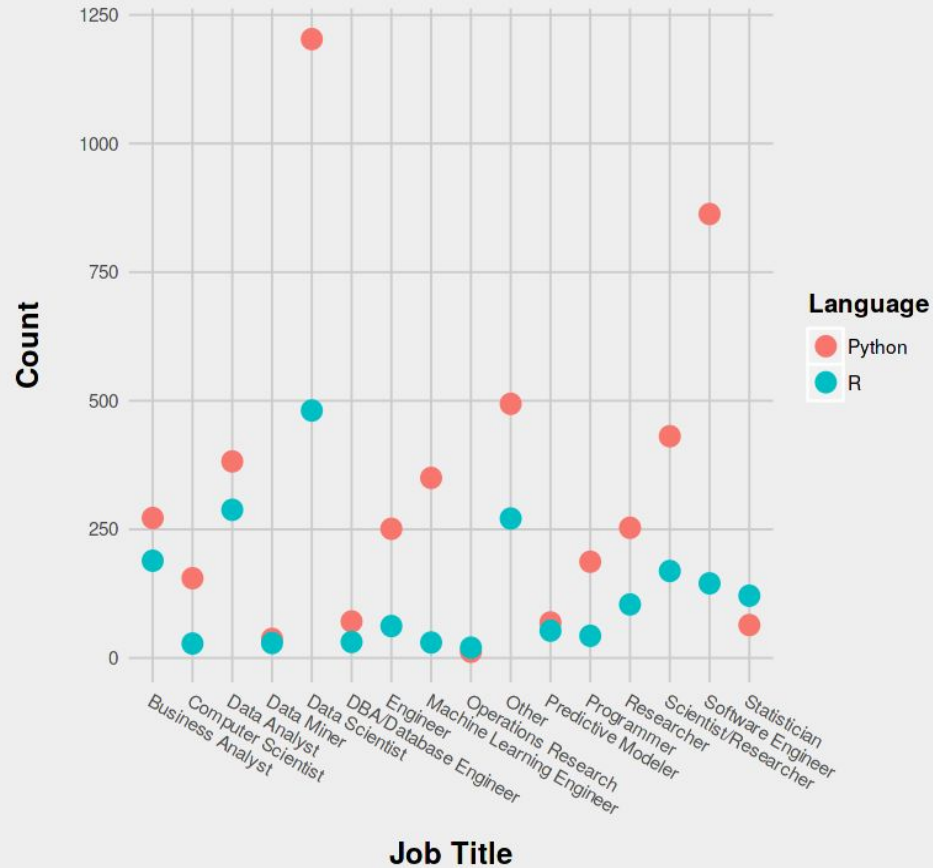
In schema.csv, what's the dif...

15 days ago

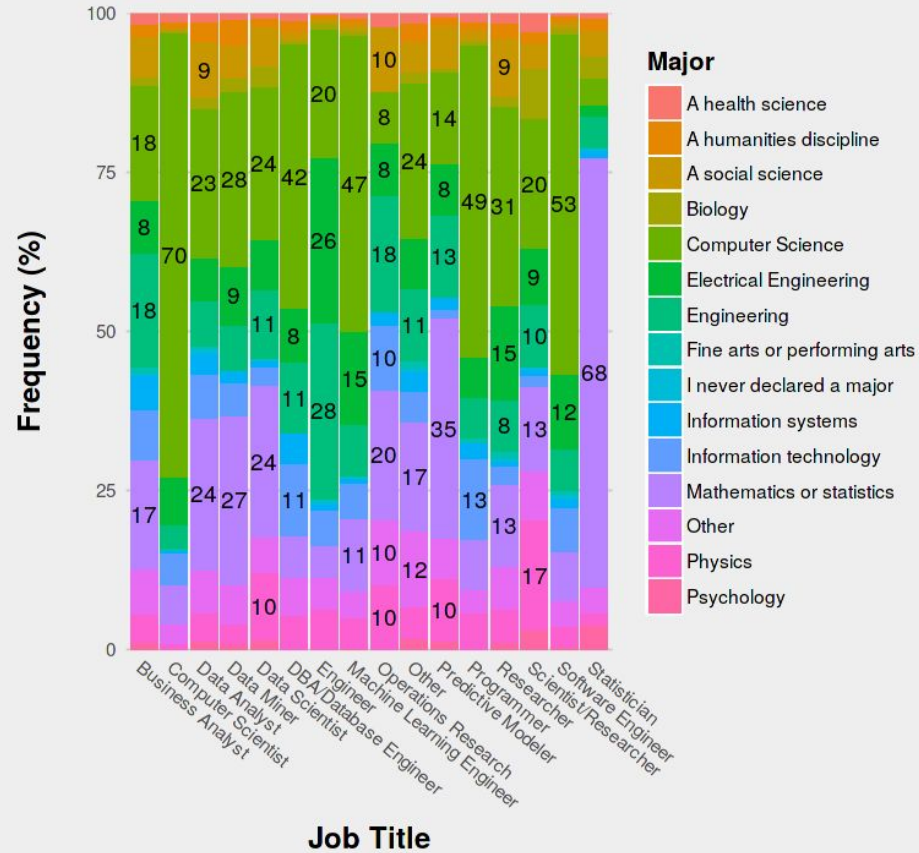
1

reply

Python vs. R by Job Title



Major vs. Job Title



	All	Data Scientist	Software Engineer	Researcher	Machine Learning Engineer
1	Python	Python	Python	Python	Python
2	SQL	SQL	SQL	R	Jupyter
3	R	R	Jupyter	Jupyter	TensorFlow
4	Jupyter	Jupyter	Unix	MATLAB	Unix
5	Unix	Unix	Java	TensorFlow	SQL
6	TensorFlow	AWS	AWS	SQL	R
7	AWS	Spark	NoSQL	Unix	AWS
8	Tableau	TensorFlow	R	C	C
9	Java	Hadoop	TensorFlow	Java	Java
10	NoSQL	Tableau	C	Excel	MATLAB

Python is Dominant in AI

- More **general-purpose** and **practical** than R
- More **productive** and **flexible** than Java and C++
- Better **ecosystem** than Ruby and JavaScript
- Python is a good all-rounder, but it's only #1 in **data science, machine learning** and **AI**

Python's Most Popular

What programming
language do you
mostly use for AI?

78%

thestateofai.com

Most Popular
Technologies per Dev
Type: Math & Data

55.5%

insight.stackoverflow.com

Let's Hands On!