Why Python for Data Science

DONI RUBIAGATRA
DATA SCIENCE INDONESIA JATIM

About Me

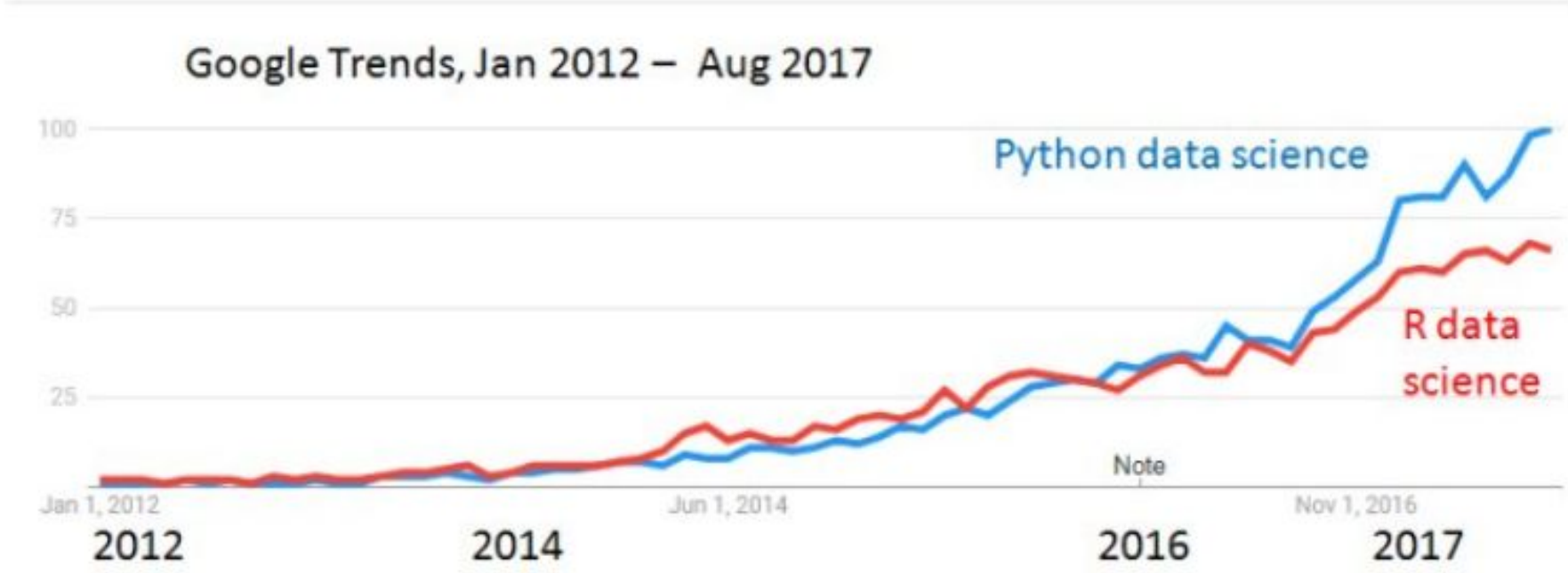
DATA SCIENCE INDONESIA

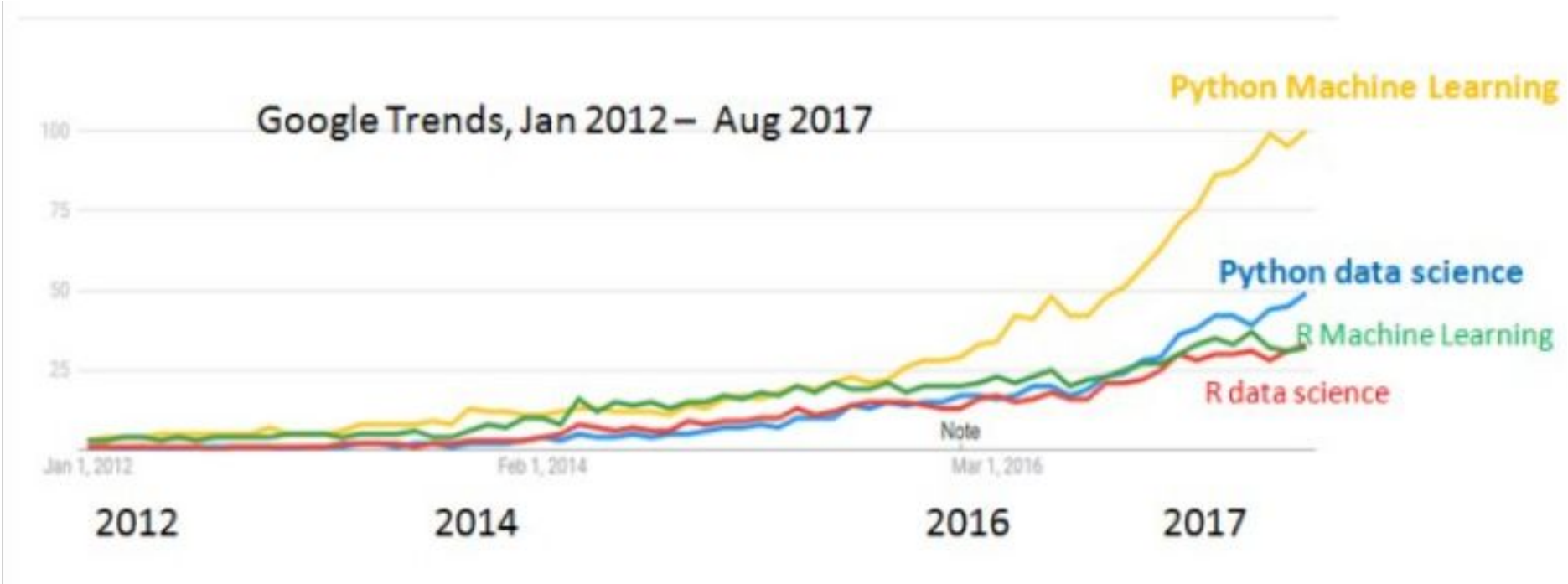
- **Doni Rubiagatra,** @rubiagatra
- B.Sc in **Mathematics** at **ITS Surabaya**
- Co-Founder **surabaya.py,** Internal **DSI Jatim,** Co-Organizer Golang Surabaya
- Back End Engineer **kumparan**
- **Pycon Korea, Malaysia, Indonesia, APAC**

DATA SCIENCE INDONESIA

Why Python is the Best Language for Data Science?

DATA SCIENCE INDONESIA




Google Trends, Jan 2012 – Aug 2017
Python Machine Learning
Python data science
R Machine Learning
R data science
100
75
50
25
Note
Jan 1, 2012
Feb 1, 2014
Mar 1, 2016
2012
2014
2016
2017

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

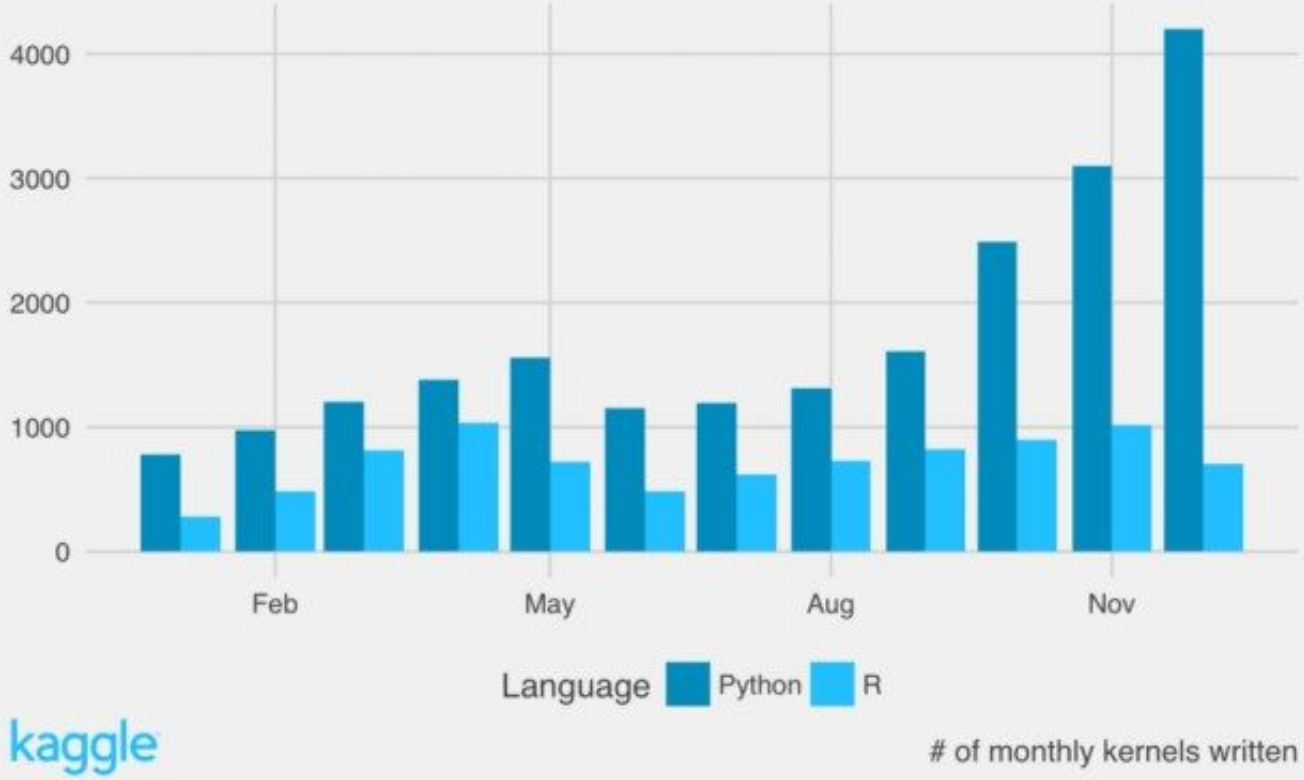
Data Scientist

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

DATA SCIENCE INDONESIA



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

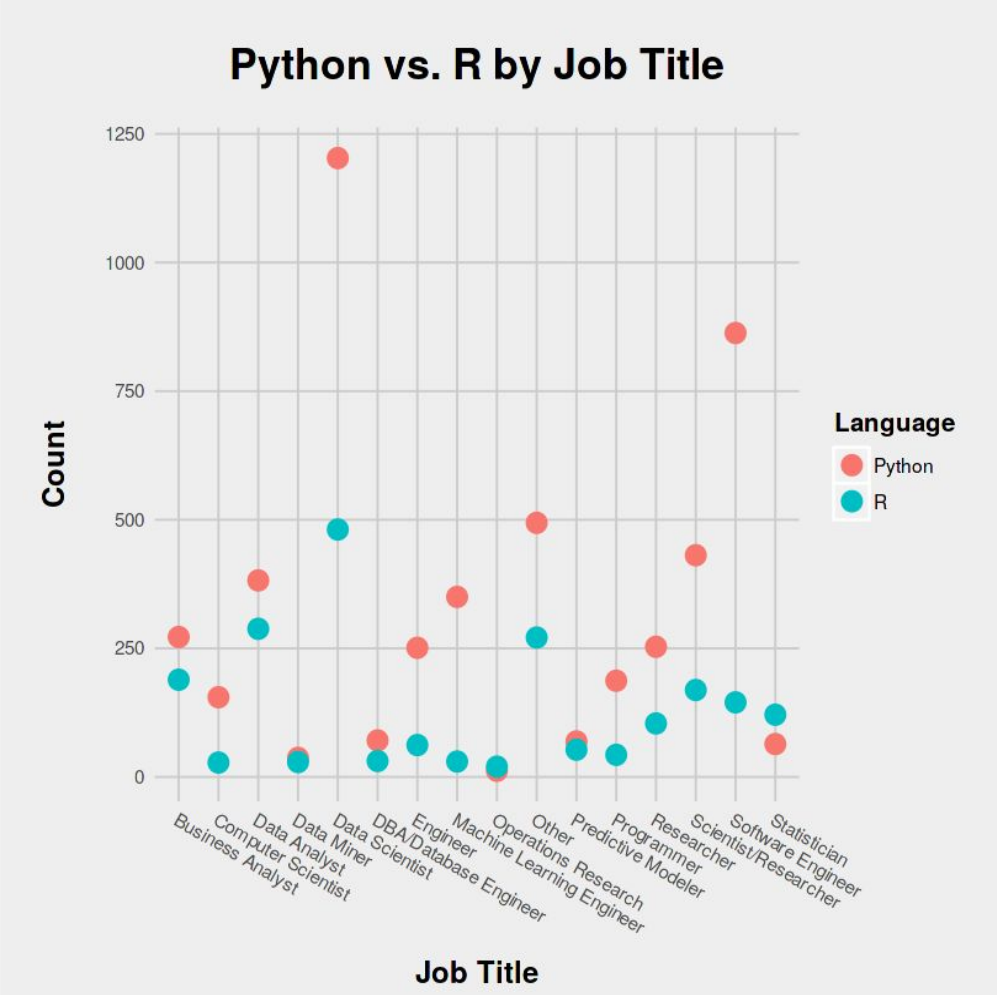
Novice to Grandmaster — 278 votes
run 3 days ago

What do Kagglers say about ... — 263 votes
run 4 days ago

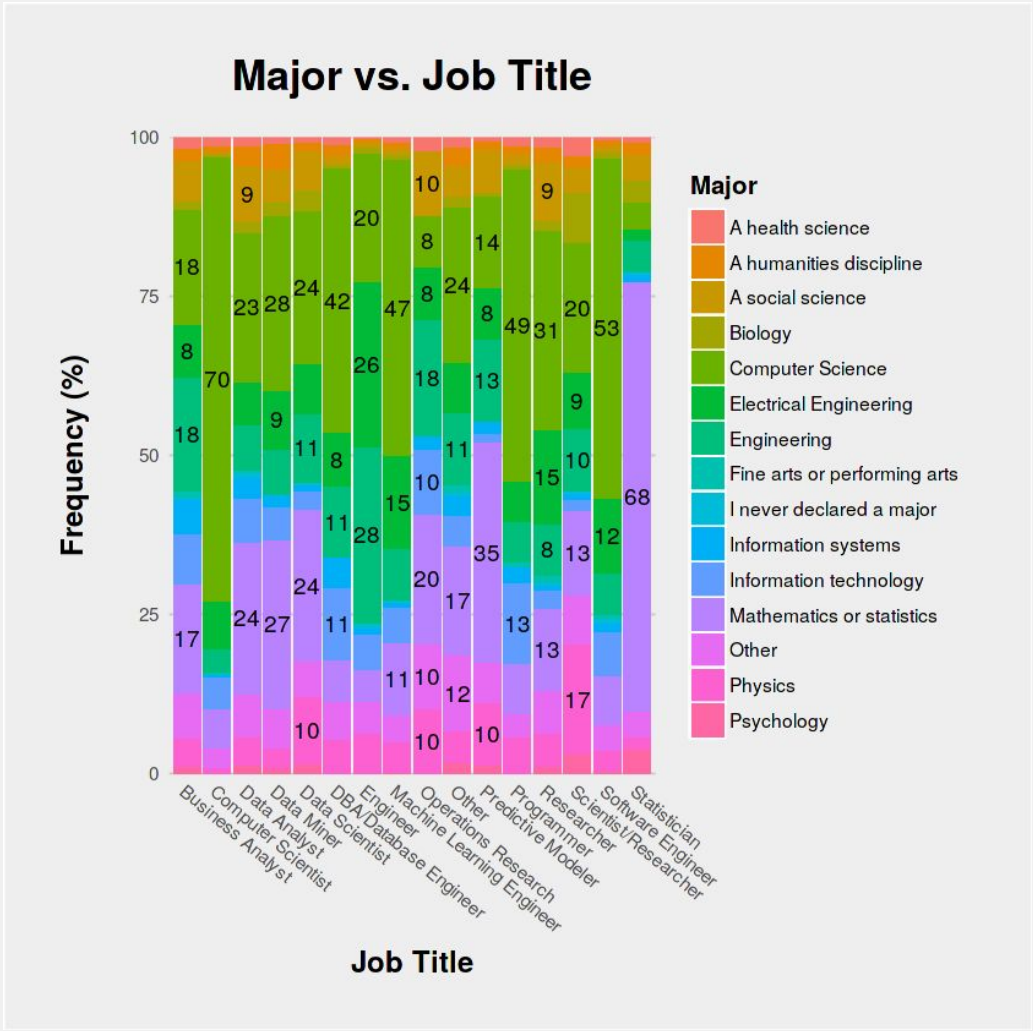
Discussion

For goodness' sake ... — 0 replies
8 days ago

In schema.csv, what's the dif... — 1 reply
15 days ago



DATA SCIENCE INDONESIA


Major vs. Job Title
Frequency (%)
100
75
50
25
0
Business Analyst
Computer Scientist
Data Analyst
Data Miner
Data Scientist
DBA/Database Engineer
Machine Learning Engineer
Operations Research
Other
Predictive Modeler
Programmer
Researcher
Scientist/Researcher
Software Engineer
Statistician
Job Title
Major
A health science
A humanities discipline
A social science
Biology
Computer Science
Electrical Engineering
Engineering
Fine arts or performing arts
I never declared a major
Information systems
Information technology
Mathematics or statistics
Other
Physics
Psychology

DATA SCIENCE INDONESIA

	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 DS

DATA SCIENCE INDONESIA

- 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

DATA SCIENCE INDONESIA

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!