



Single Day Event

Joseph Rajan, Pfizer

Open-Source Python in Clinical Trial



Abstract

Popularity

- Most used language
- Usage still growing exponentially

Community

- Vast online community
- Most active contributors
- Thoroughly Documented

Availability

- Huge code repository for python
- Useable libraries are growing
- Most Active libraries in GitHub

Team Player

- Easily integrate with the existing application
- OS independent
- Can be integrated with SAS, R etc.

Scalability

- Data Processing
- Parallel processing
- Load Balancing

Considerations

- Security Features
- Data Isolation
- Accommodate Public Libraries

Migrate

- Easy to migrate from existing process
- Highly skilled programmers available

ML / AI

- Leads in ML and AI libraries
- Statistics Visualization packages
- Secured providers like Amazon, Google.



Python In Statistics

- Huge Libraries of tested packages available to use for statistics.
- Data Sciences models such as BERT, NER, Predictive Model etc readily available to use.
- NLP, Image recognition, Pattern recognition models are will developed in python .

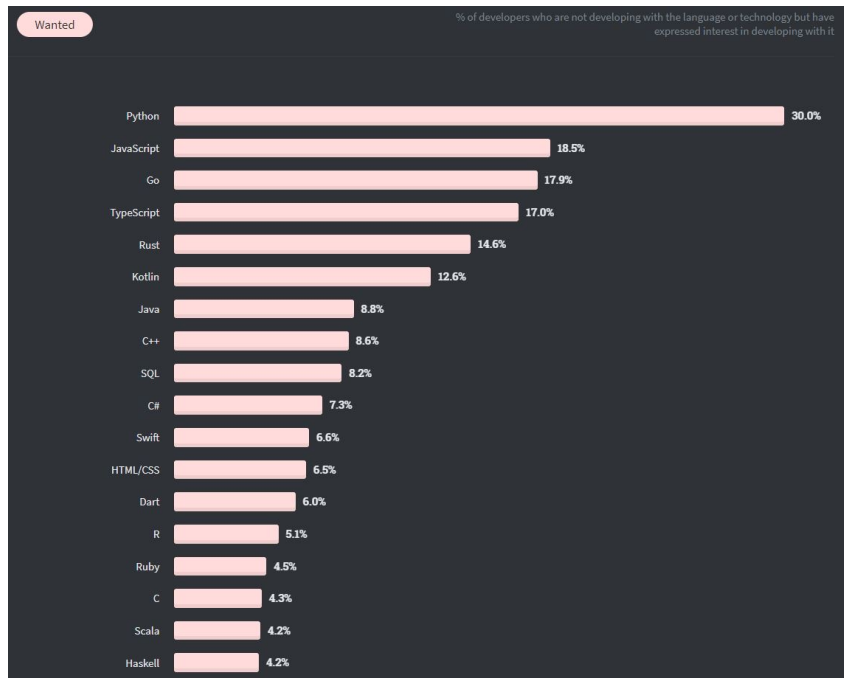
Some Basic Packages:

- NumPy, SciPy, Pandas, Matplotlib most tested and used package for statistical analysis.
- Matplotlib, Seaborn, Plotnine(ggplot), Bokeh, pygal, Plotly, geoplolib, Gleam, missingno, Leather, Altair, Folium for visualization.
- Tensor Flow, Keras, Theano, Scikit-learn for Machine Learning and Artificial Intelligence



Growth

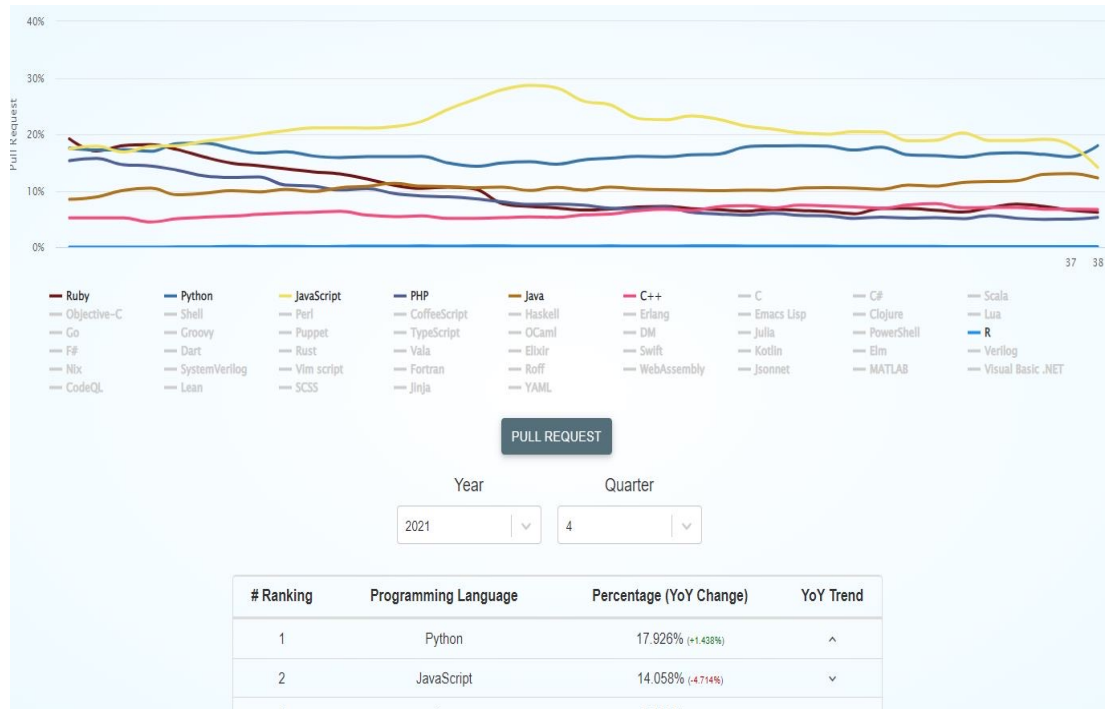
- Most used programming language according to stackoverflow.com
- Most contributed to resolve issues as well.
- Stronger community brings more and better solutions for any given problem statement.
- Millions of blogs to learn from, there by increases python developers.
- Better and structured documentations on packages.





Global Contribution

- Currently the most packages and applications developed for free usage under opensource license is python according to GitHub.
- The growth of application developed is enormous.
- More GitHub contribution implies more developers contributing to resolve problem statements.





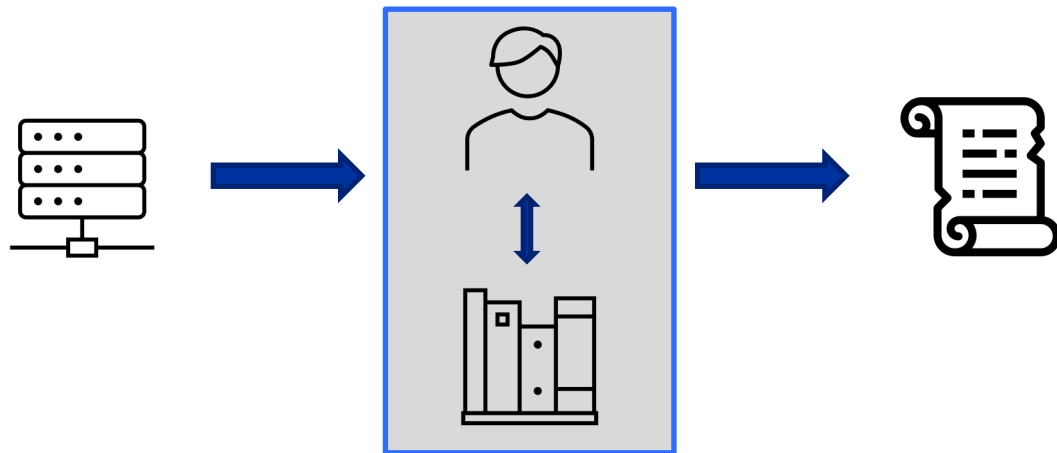
Team Player

- Operating System Independent. Cross OS executable.
- Easily integrate with other programming languages, applications, cloud / Web based as well.
- SAS has SASPy to integrate python and SAS macros and other SAS based applications.
- Can easily implement in Cloud providers like Azure, Amazon, Google etc.
- Every major SDK (Software Development Kit) has a special kit for python.

Migration

- Migration from legacy to python has never been faster.
- Better DevOps management as it complies to programming standards.

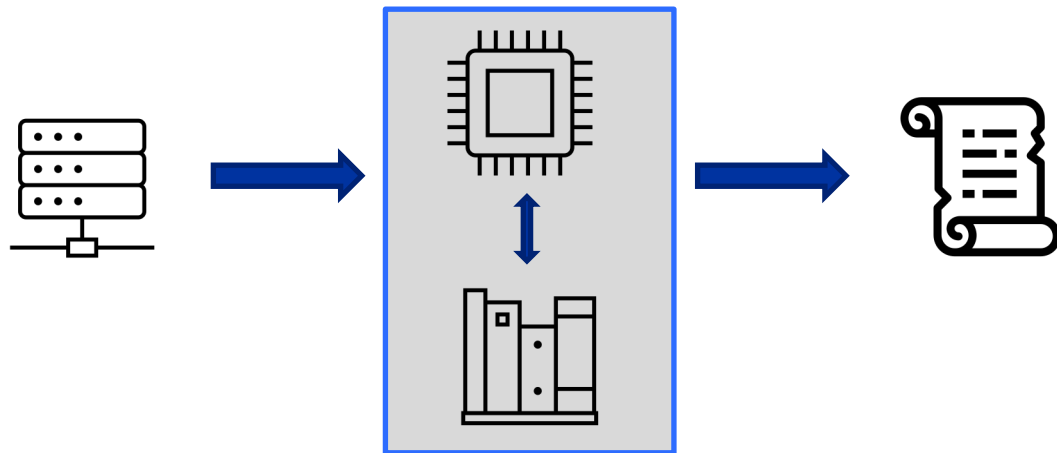
Moving Towards ML / AI



Current process include human's expertise on the subject to structure the Clinical trial data on developing Stats and data processing for submission.

Human's expertise come more from the Clinical literature and learning.

Moving Towards ML / AI



Training Machine to learn from Clinical literature via ML and AI concepts is the future.

Eliminated hours of development, data validation, data structuring as Machine is learned and is a SME.



Machine as SME

- Humans learn from literatures and experience.
- Replicate the same philosophy to machines as well.
- Isolate process and module them.
- Apply AI Models, train the machines with the literatures.
- Apply the trained literature knowledge of the machines to the past historical data and studies.
- Train them again and again for accuracy.
- Now you have well trained model



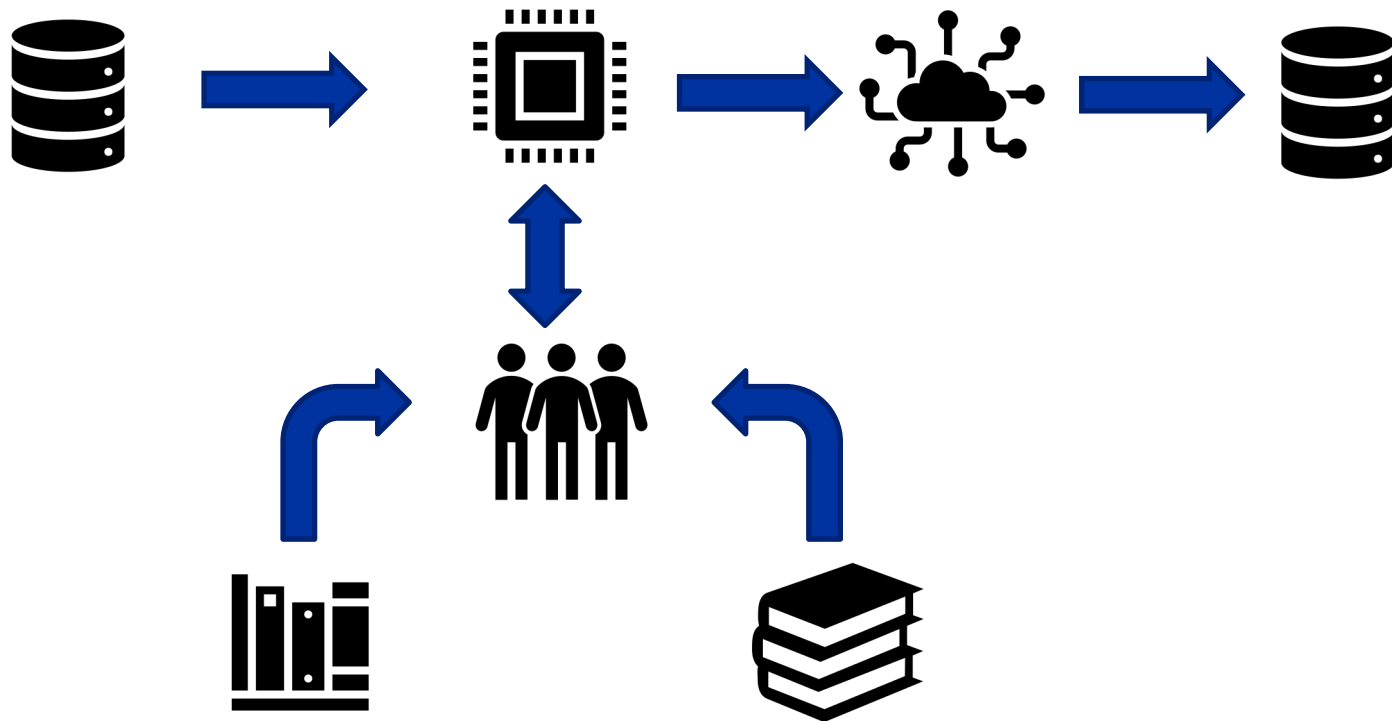
Make them Cognitive

- Design should be cognitive with evolving industry like the Clinical Research.
- Literatures evolve, so should the machine's learning.
- Training should be consistent.
- New scenarios should be validated.
- Machines should be pushed to the limits in terms to cognitive learning.
- Machines are evolving, Evolve them for clinical research.



Business Use Case

SDTM Data Mapping





SDTM Data Mapping

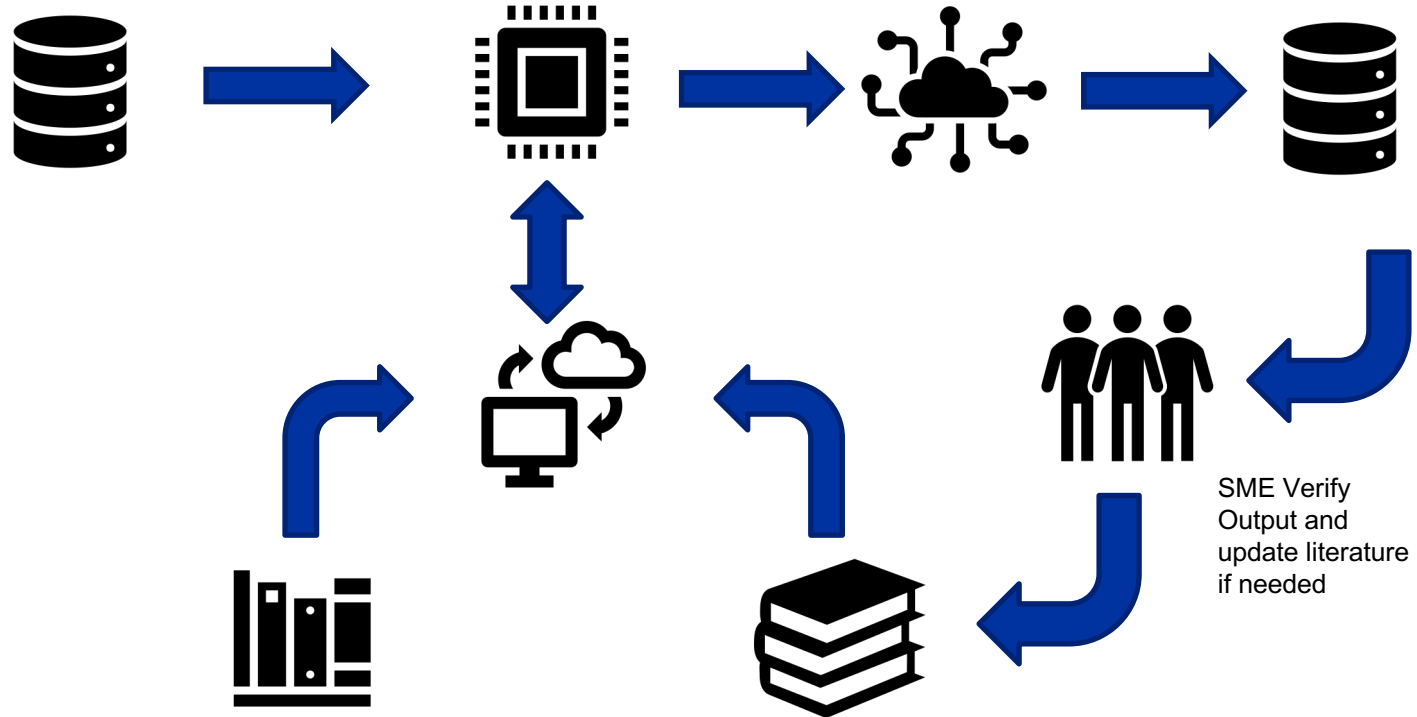
Traditional SDTM Mapping Process

- Data Received from the Upstream.
- SME develop the conventions to identify the source variable.
- SME develop the rules to map the source to the target SDTM Variables.
- Developers utilize the rules and perform the mapping.
- Developers would have to perform n number of tasks before the final output is accurate

Automations SDTM Mapping Process

- Data Received from the Upstream.
- SME develop the conventions to identify the source variable.
- SME develop the rules to map the source to the target SDTM Variables.
- Developers utilize the rules and perform the mapping.
- Developers utilize automated tools and codes to reduce the number of efforts performed in the traditional method.

Python in SDTM Data Mapping





Python in SDTM Data Mapping

AI SDTM Mapping Process

Training

- Develop the appropriate model to teach the machine to learn.
- Train Machine with the existing Clinical Literatures.
- Run test with the existing Datasets.

Cognitive Approach

- Machine developed should accept new parameters
ie, new variables in SDTM or upstream data source.
- If any discrepancies identified in the output there must be a feedback mechanism that will make the machine re-learn the updated literature or identify the uniqueness of the new parameter.
- Apply them for usage.
- Frequent training is required for better accuracy and consistent growth.

Python has already made its way into clinical trials. Leveraging it to the fullest potential is important.

Python could be a key player in clinical research.

Automation, ML AI are revolutionizing logistic, banking, healthcare industries.



The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse