Python in Healthcare: Coding the Future of Medicine

Swetha Mandava
swetha.mandava@bayer.com

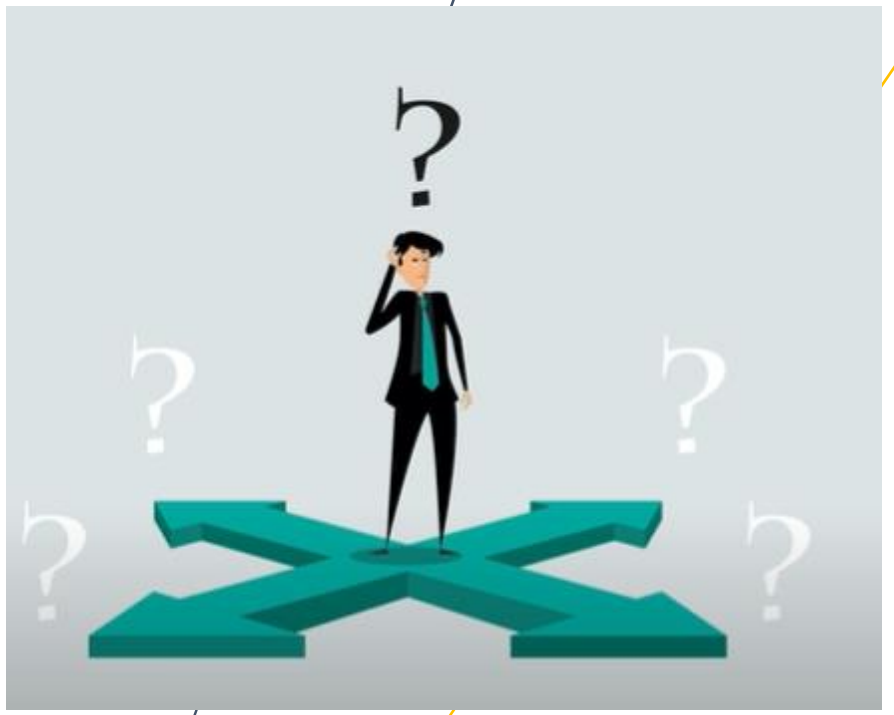
Disclaimer

The content and conclusions of this presentation are the works of the author. They do not necessarily represent the opinions, recommendations, or practices of Bayer Pharmaceuticals Pvt. Ltd.

Agenda

- Top 10 reasons to learn Python
- Why Python in healthcare?
- Applications of Python and R
- Comparison of Python and R
- Disadvantages of Python
- R vs Python vs SAS
- Conclusion


Top 10 Reasons To Learn Python



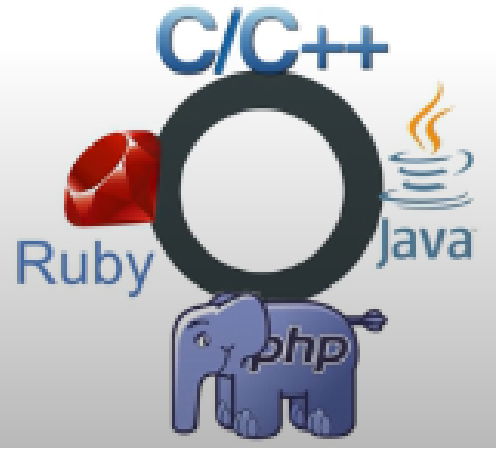
10. Simple & Easy To Learn

Java

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world");
    }
}
```

Python

```
print("Hello, world")
```



It's that **SIMPLE**!

9. Portable & Extensible

8. Web Development

- Develop web applications
- Scrape websites

Frameworks

django | Flask | Pylons | WEB2PY

7. Artificial Intelligence

Libraries

- Scikit-learn
- Keras
- Tensorflow
- Opencv

6. Computer Graphics

- Graphical User Interface
- Desktop applications
- Game development

Libraries

Tkinter

Jython

wxPython

pygame

5. Testing Frameworks

- Python supports testing with cross-platform & cross-browser
- Built in testing framework which covers debugging time and fastest workflows

Tools

Splinter

Framework

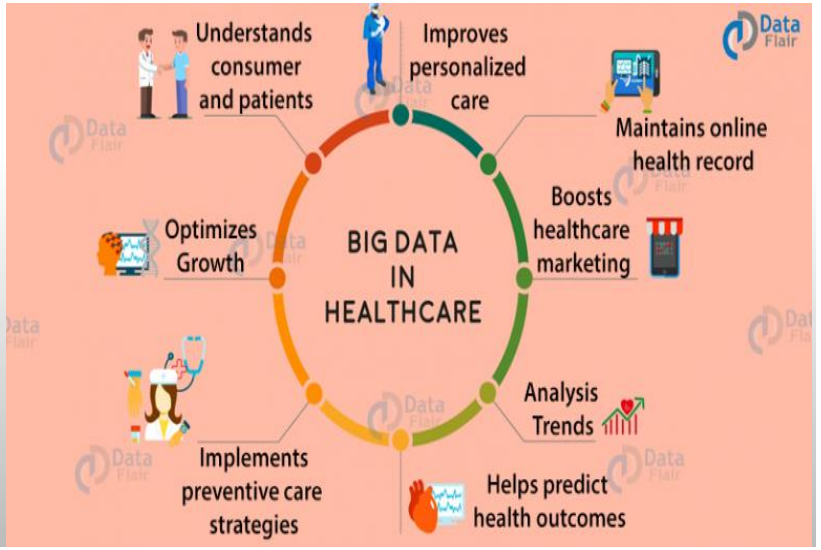
pytest

4. Big Data

- Python handles **BIG DATA!**
- Python supports parallel computing
- You can write MapReduce codes in Python

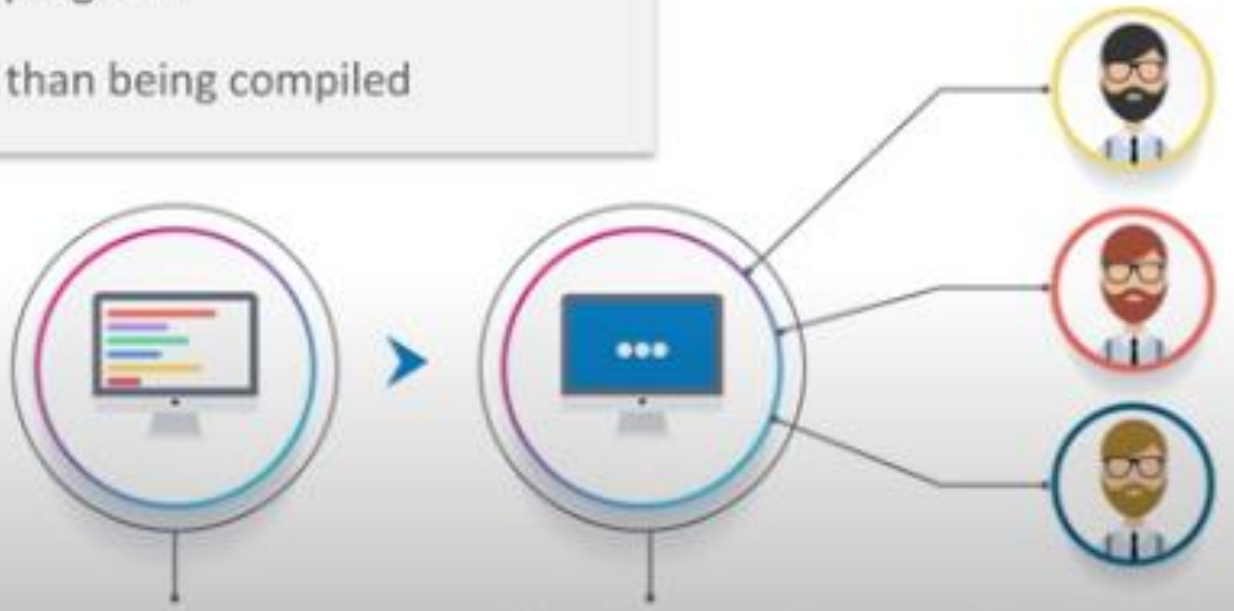
Libraries

PYDOOP | DASK | PySpark



3. Scripting: Automation

- It is the most popular scripting language in the industry
- Automate certain tasks in a program
- They are interpreted rather than being compiled



2. Data Science

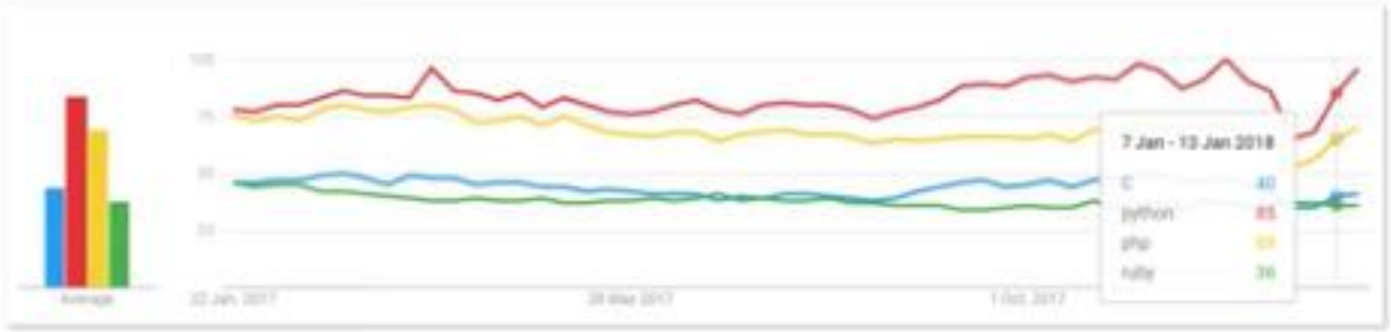
- Well-suited for data manipulation & analysis
- Deals with tabular data with heterogeneously-typed columns
- Arbitrary matrix data
- Observational/ statistical datasets

Libraries

NumPy | Pandas | matplotlib | seaborn

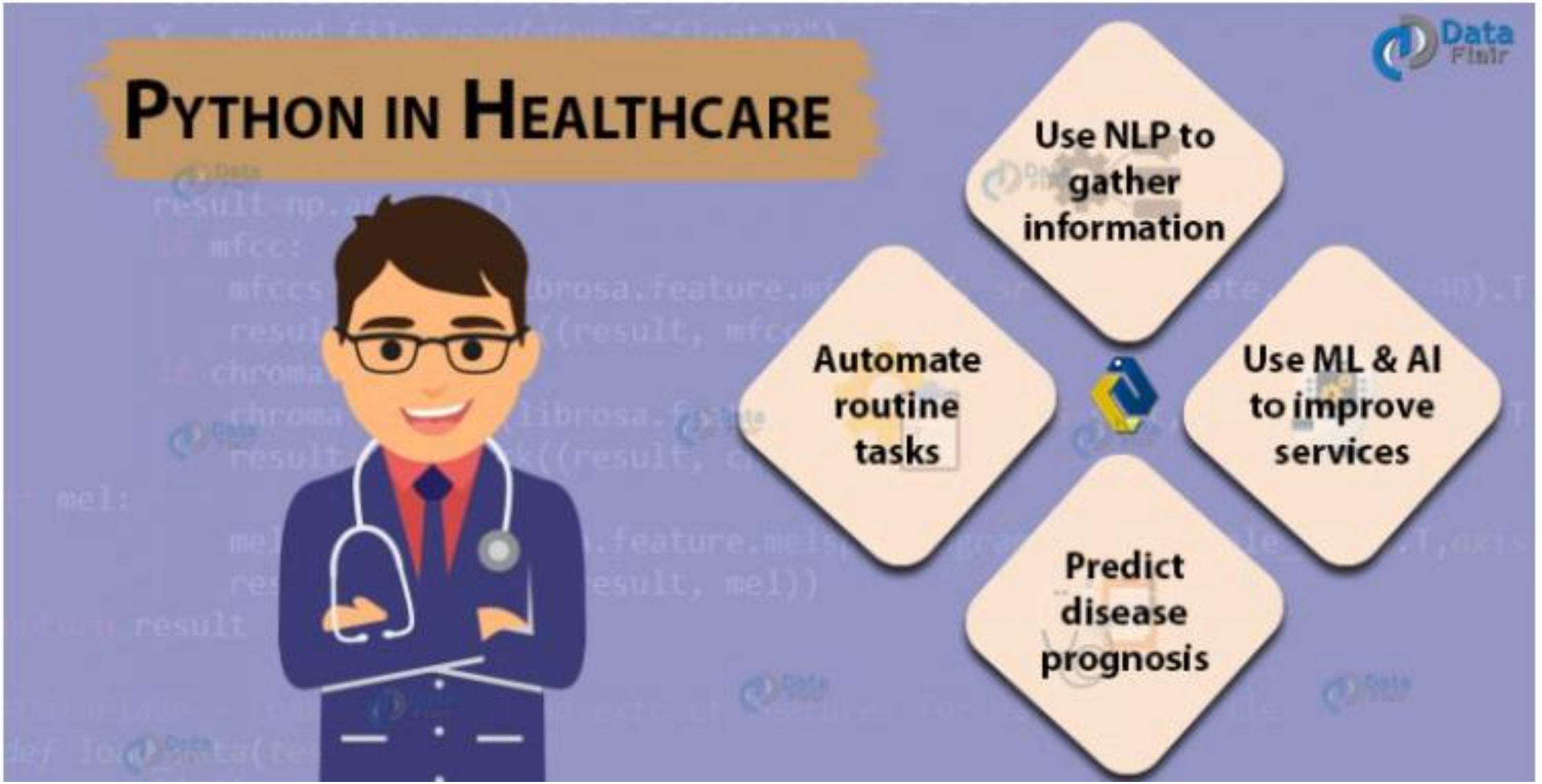
1. Popularity & High Salary

USD $116,028



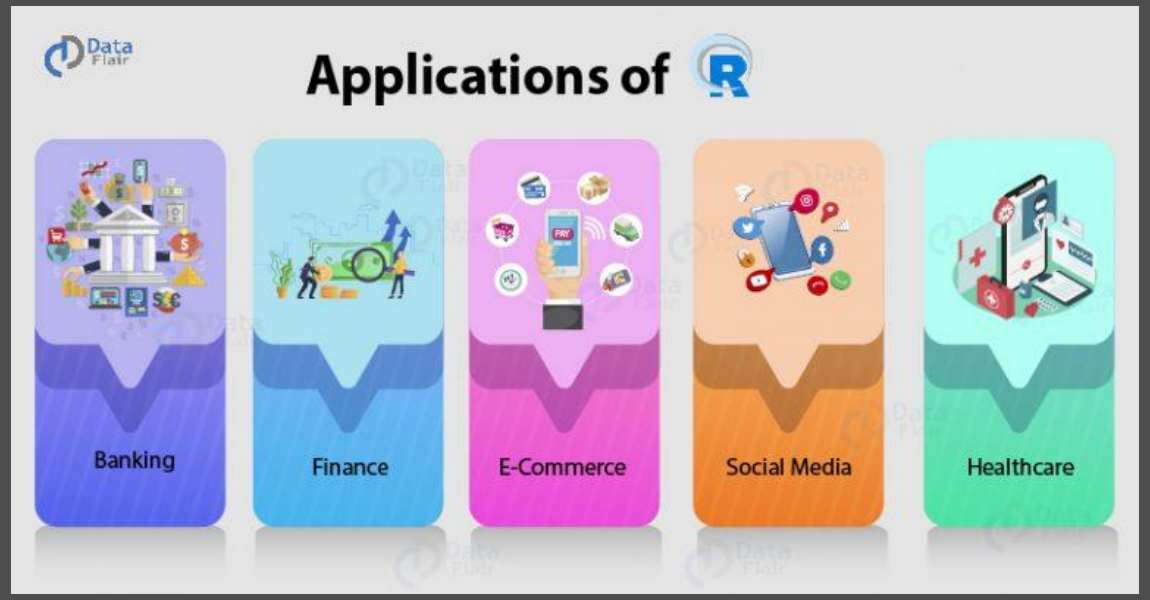
Big Giants

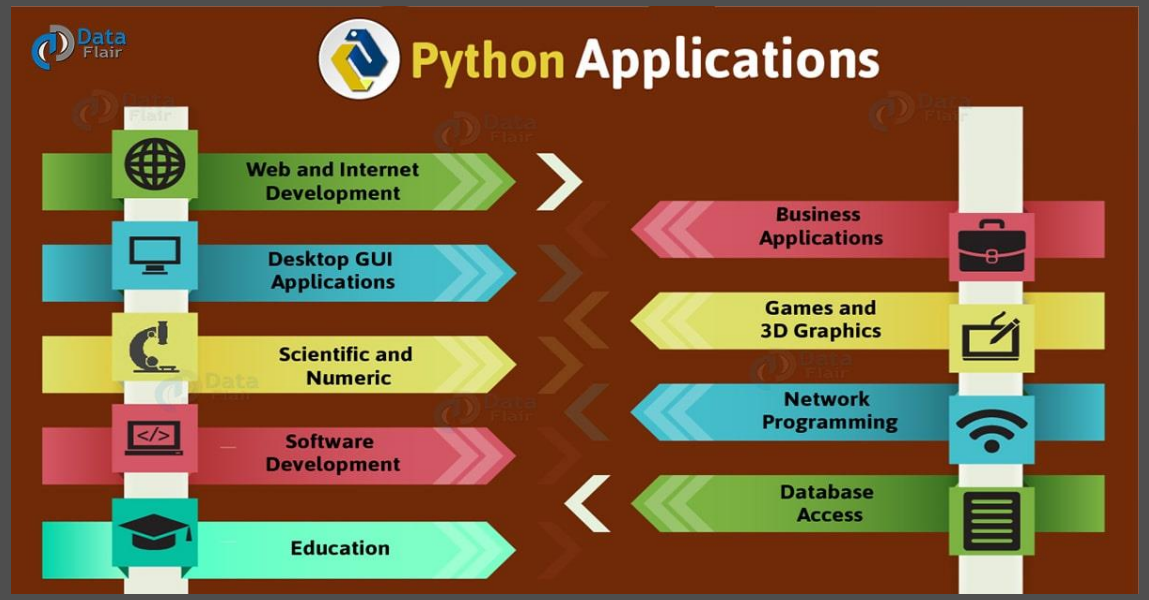
Google YouTube NASA IBM RaspberryPi facebook. Dropbox


PYTHON IN HEALTHCARE
Data Flair
Use NLP to gather information
Automate routine tasks
Use ML & AI to improve services
Predict disease prognosis

Python in Healthcare

- **AiCure :** It is an NIH and VC-funded healthcare startup in New York. AiCure automates the process of **ensuring that patients are taking their medicines**, that too at their assigned time. It has combined artificial intelligence and mobile technology together.
- **Berg :** Boston-based biopharma company Berg is using AI to research and **develop diagnostics and therapeutic treatments** in multiple areas, including oncology. Current research projects underway include dosage trials for intravenous tumor treatment and detection and management of prostate cancer.
- **Fathom Health :** Again, a Healthcare startup with deep learning NLP system for reading and **understanding electronic health records**.
- **IBM Watson Oncology :** It is a leading institution at the forefront of driving change in treatment decisions, using patient medical information and history to optimize the **selection of treatment options**
- **Success Story :** PyDrone(Drug from Astro Zeneca) took about 3 months of development time, another 3 months of QA, and 3 weeks of documentation time to produce about 5,600 lines of finished Python code . Here is the success story : https://www.python.org/about/success/astra/


Data Flair
Applications of
Banking
Finance
E-Commerce
Social Media
Healthcare


Data Flair
Python Applications
Web and Internet Development
Desktop GUI Applications
Scientific and Numeric
Software Development
Education
Business Applications
Games and 3D Graphics
Network Programming
Database Access

Difference Between R and Python

Features	R	Python
Scope	Used mainly for statistical modeling	Used for a variety of purposes like web-application development and data analysis
Used By	Statisticians, Analyst & Data Scientist	Developer, Data Engineers & Data Scientist
Suitable For	People with no prior experience in programming	Newbies to experienced IT professionals
Package Distribution	CRAN	PyPi
Visualization Tools	ggplot2, plotly, ggiraph	Matplotlib, bokkeh, seaborn

Parameter	R	Python
Objective	Data Analysis and Statistical Modeling	Data Science, Web Development, Embedded Systems
Workability	Consists of many easy to use packages	Can easily perform matrix computation as well as optimization
Integration	Locally Run Programs	Programs integrated with web-app for easy deployment
Database Handling Capacity	Poses problem for handling large dataset	Can handle large data easily without any fault
IDE	Rstudio, R GUI	Spyder, IPython, Jupyter Notebook
Essential Packages and library	ggplot2, tidyverse, caret	Numpy, pandas, scipy, scikit-learn, TensorFlow

Comparison between R Programming and Python

Disadvantages of Python

Resource requirements

- When software or tools used during analysis are improperly joined with existing data sources, organizations may experience several side effects like loss of regulatory compliance and Increased strain on IT resources that can cause harm to data integrity and clinical trial timelines.
- Python lags R when it comes to statistical analysis. Python though may have improved a lot, but it still lacks certain statistical packages as compared to R.

Error prone

- The dynamically typed nature of Python makes it vulnerable to runtime errors. Tasks that require heavy memory suffer from Python.
- The flexible data-types in Python contribute towards its high memory consumption.

Disadvantages of Python

Data Integrity

- Python, when compared with JDBC has an underdeveloped database access layer which may create problems to get approval from regulatory authorities.
- When a new data system or analysis tool is introduced to a data environment, there are several potential roadblocks that may appear that prevent proper integration and implementation. Such roadblocks include:
  - An increase of regulations surrounding clinical trial data technologies
  - Lack of integration experience relating to modern software, tools, or data sources
  - An inability to provide required security on data access
  - The use of custom, legacy tools and software that do not support full integration

R, Python or SAS – Which is the Best Tool for Data Science Learning

- **Choosing the right learning curve:**
  - R will take some time for us to gain proficiency whereas Python provides an easier learning curve.
  - When it comes to SAS, it is easy to implement complex statistical thinking efficiently and with ease.

- **Cost and expenses of the tools:**
  - Both Python and R are open-source.
  - SAS is a closed-source proprietary tool that is highly expensive.

R, Python or SAS – Which is the Best Tool for Data Science Learning

- **Libraries and Support tools**
  - Both Python and R enjoy a wide range of packages.
  - Python is famous for its wide variety of packages on machine learning. It also provides versatile packages of web-development, GUI programming and much more.
  - R is limited only to packages of statistical modeling. However, the visualization packages of R such as ggplot2, Lattice, RGIS are much more diverse and visually aesthetic.
  - SAS on the other hand, provides a wide variety of Business Intelligence, Statistical, and analytics tools. However, it still lags in more advanced tools of machine learning and data visualization.

R, Python or SAS – Which is the Best Tool for Data Science Learning

- **What do industries require?**
  - Industries have long trusted SAS as their primary tool for data analytics and business intelligence. This is due to high reliance, sophistication and stability that SAS provides to its clients.
  - However, gradually, the trend is shifting to Python, R and other open-source libraries that provide far more powerful features than SAS. While SAS may be ideal for large scale industries who have not adapted open-source as their primary tool, it is still not flexible as other free alternatives.
- **Tool for the right need :**
  - In the end, the choice of learning Python, R and SAS depend on their usage and where you need to apply them.
  - For data scientists, seeking careers in the field of natural language processing, visual computing, and big data, Python and R are the ideal programming languages.

Conclusion

- **Python** can be chosen by **beginners** who want to have an in-depth knowledge of data science.
- **R** is befitting for beginners in data science who have **experience in statistics**, as R will also introduce them to several aspects of a programming language. Nevertheless, R is a must have tool for aspiring data scientists even if you start with Python.
- SAS is customized for business requirements and is used heavily by large scale companies. This makes **SAS** a specific language for **business intelligence needs**. Also, the high costs make it an unaffordable tool for many.
- Therefore, we conclude Python and R to be the best tools for aspiring data scientists.

References

- https://data-flair.training/blogs/python-in-healthcare/
- https://www.edureka.co/blog/10-reasons-why-you-should-learn-python
- https://www.educba.com/sas-vs-r-vs-python

Thank you!