2. Write 2 to 3 sentences on why Python is so popular among data analysts.

Python is open source, free to use, and has a solid and active developer community.

Thanks a lot Python is better suited for solving analytical problems: you can work with an unlimited amount of data in it and write a couple of lines of code for a complex operation.

Python can be used to work with multiple data sources. If Excel is both a storage and a computing engine, then Python is completely independent. If you can find a way to read data in Python, you should be able to use it. Python has a lot of libraries, so you can analyze data from different sources, be it CSV, Excel, JSON, or SQL.

Analysts also value Python for its built-in interpreter that allows you to code on the go. In Data Science, this is relevant for testing hypotheses interactively.
Python supports visual elements such as charts and graphs, allowing data analysts to work with a wide variety of data. Data visualization methods are easy to understand and remember, making them ideal for professionals working with large amounts of data.

3. After doing some research, name the 5 top companies in the world that use Python (either as a tool for software engineering or for analytics).

Google uses Python for many of its most famous apps, including Youtube, which you're probably familiar with.

The use of Python on Facebook's back-end is mostly for production engineering, a specific sector of development at Facebook that combines systems and software engineering.

Day by day, PayPal is helping people keep their money safe through convenient and secure transactions via the PayPal platform. PayPal uses Python across its entire infrastructure and one of the most shared articles on Python, "10 Myths of Enterprise Python" is credited to the official PayPal Engineering writers on Medium.

Spotify uses Python for its back-end services and for data analysis.

Python is a facet of many internal structures of Netflix, including but not limited to:
Open Connect, a content delivery network (CDN)
Demand Engineering, a component of Netflix's network availability that works to prevent latencies and improve user experience (UX)
Experimentation, a platform to perform AB tests and other experiments

CORE, a team that uses data analysis to alert relevant parties about system issues
Animation & VFX, a subsection of Netflix for creating animated content and video effects

4. For each of the following scenarios, explain what tool you would use and why

- You have a small data set that needs some quick tweaks and minor analysis. You'll need to filter some columns and make a quick chart.

Excel is great for this task, because its functionality will be enough to filter some columns and make a quick chart

- You need to retrieve some portion of data from a very large database.

If the database is very large, Python must be used because Excel can only serve 1,048,576 rows of observations.

- You have a data set with 15,000,000 rows and 350 columns that needs to be sorted and prepared for a more advanced analysis.

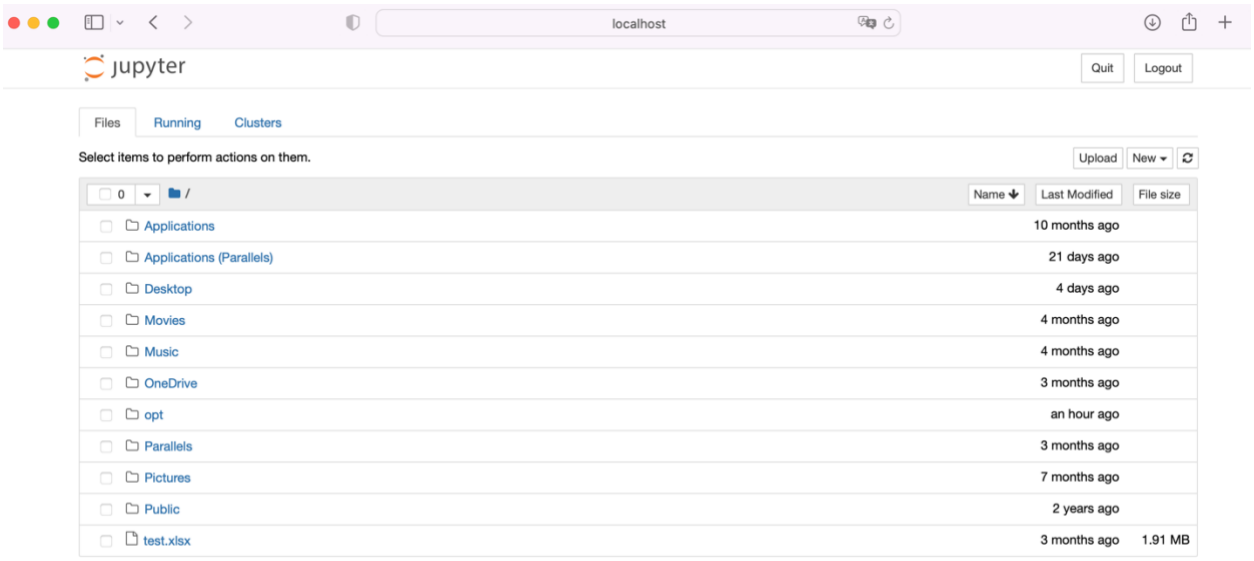
If the database is very large, Python must be used because Excel can only serve 1,048,576 rows of observations.

5. Download Anaconda.

6. Set up the environment variables on your computer and copy them into your document together with your answers to steps 2 through 4.

7. Launch Jupyter.

8. Take a screenshot of the page that opens in your browser upon launching Jupyter.


localhost
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
0
/
Name
Last Modified
File size
Applications 10 months ago
Applications (Parallels) 21 days ago
Desktop 4 days ago
Movies 4 months ago
Music 4 months ago
OneDrive 3 months ago
opt an hour ago
Parallels 3 months ago
Pictures 7 months ago
Public 2 years ago
test.xlsx 3 months ago 1.91 MB