

Natalie Aguilar
Exercise 4.1
January 2, 2025

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

- Python is popular among data analysts due to its simplicity, readability, and extensive library ecosystem, simplifying complex data manipulation and analysis tasks. Libraries like Pandas, NumPy, and Matplotlib make data cleaning, statistical analysis, and visualization efficient and intuitive. Additionally, Python's versatility allows seamless integration with machine learning and automation tools, making it a comprehensive choice for modern data workflows.

3. Name the 5 top companies in the world that use Python (either as a tool for software engineering or for analytics).

- Google: Python is a core programming language used by Google to develop tools, systems, and data analytics solutions.
- Facebook (Meta): Utilizes Python for data analysis, backend services, and machine learning applications.
- Amazon: Employs Python for automation, data processing, and AI-driven recommendations on its e-commerce platform.
- Netflix: Uses Python for content recommendation algorithms, data analytics, and backend services.
- Spotify: Relies on Python for backend development, music recommendation systems, and data analytics.

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

o 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.

Tool: Microsoft Excel

- Excel is perfect for small-scale data manipulation and quick visualizations.
- Offers intuitive filtering, sorting, and charting features without requiring coding skills.

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

Tool: SQL

- SQL (Structured Query Language) is optimized for efficiently querying large relational databases.
- It allows precise data retrieval with filtering, joining, and aggregation capabilities.
- Most large databases are designed for SQL-based interaction, making it the natural choice.

o 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.

Tool: Python

- For datasets of this size, Python is excellent for data cleaning, sorting, and transformation, leveraging efficient memory management.
- The tools are versatile, scalable, and equipped with advanced data manipulation capabilities.

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

Edit environment variable



%USERPROFILE%\AppData\Local\Microsoft\WindowsApps

C:\Users\krist\anaconda3

C:\Users\krist\anaconda3\Scripts

C:\Users\krist\anaconda3\Library\bin

New

Edit

Browse...

Delete

Move Up

Move Down

Edit text...

OK

Cancel

Home

+

http://localhost:8889/tree?

Relaunch to update

jupyter

File View Settings Help

Files

Running

Open

Download

Rename

Duplicate

Delete

New

Upload

/

Name	Last Modified	File Size
dgyan	10 months ago	
Documents	2 months ago	
Downloads	1 hour ago	
eclipse	5 years ago	
eclipse-workspace	5 years ago	
flutter	4 years ago	
git	5 years ago	
GNS3	last month	
iCloud Drive (Archive)	4 years ago	
iCloud Drive (Archive) - 1	4 years ago	
iCloud Drive (Archive) - 2	4 years ago	
knime-workspace	10 months ago	
Library	21 hours ago	
Movies	2 years ago	
Music	last month	
myenv	last year	
node_modules	3 years ago	
PhpstormProjects	6 years ago	
Pictures	4 years ago	
Public	6 years ago	
PycharmProjects	5 years ago	