

Answers 4.1

2.

There are a lot of articles online describing the benefits of knowing Python and benefits of using the language in the field of data analysis. The most mentioned attributes of the language are:

- Simple, easy to learn, and more readable than other programming languages.
- It is well supported with a large Python community.
- It is open source, which makes it cheaper than paid statistical programs.
- It is flexible; with a large collection of libraries, it can handle any type of challenge the coder is facing.

All these important benefits of Python make it one of the must-know languages for a professional data analyst.

3.

1. Amazon-the global eCommerce platform utilizes Python in the recommendation system, employing AI and ML to analyze customer buying habits and suggest relevant products.
2. Facebook-Python language is the core of Facebook: it makes up 21% of the whole codebase. It is a key component in Facebook's backend system, ensuring that the platform is up and running.
3. Spotify-the world's largest music streaming service with more than 280 million monthly users. This is the company using Python for data analysis and backend services.
4. Google-Google, the world's leading search engine, is a well-known Python software company. Python has been a major language at Google since the company's inception and remains a vital technology alongside C++ and Java.
5. Netflix-the leading video streaming service with over 230 million subscribers globally. To improve user experience and retention, Netflix uses Python to analyze user viewing history and provide personalized suggestions. Netflix engineers also use it to forecast audience size, viewership, and other metrics for platform development.

4.

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.

For this scenario I would use excel to achieve results. Since the data set is small and only minor analysis is needed with the ability to make a quick chart, I think excel will be ideal for this task.

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

For this task a programming language will need to be used. Either SQL or Python can help generate results. Since the dataset is very large and only retrieval of data is needed for the task, SQL might be the best way to go about it.

For this assignment I think utilizing the functionality of Python programming will be ideal. Being able to manipulate a very large amount of data in a short period of time is one of the key functions offered by Python, which is exactly what is needed for this assignment.

The screenshot shows the 'System Properties' dialog box with the 'Environment Variables' tab selected. The 'User variables for Sam' section contains the following variables:

Variable	Value
anaconda	C:\Users\Sam\anaconda3
anaconda library	C:\Users\Sam\anaconda3\Library\bin
anaconda scripts	C:\Users\Sam\anaconda3\Scripts
OneDrive	C:\Users\Sam\OneDrive
Path	C:\Users\Sam\AppData\Local\Programs\Python\Python312\S...
TEMP	C:\Users\Sam\AppData\Local\Temp
TMP	C:\Users\Sam\AppData\Local\Temp

The 'System variables' section contains the following variables:

Variable	Value
anaconda	C:\Users\Sam\anaconda3
anaconda library	C:\Users\Sam\anaconda3\Library\bin
anaconda scripts	C:\Users\Sam\anaconda3\Scripts
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
NUMBER_OF_PROCESSORS	8
OnlineServices	Online Services

The screenshot shows the JupyterLab web interface. At the top, the browser address bar indicates the URL is `localhost:8888/tree`. Below the browser window, the JupyterLab header features the Jupyter logo and buttons for 'Quit' and 'Logout'. The interface has three main tabs: 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser view. A table lists various files and folders, including 'anaconda3', 'Contacts', 'Desktop', 'Documents', 'Downloads', 'Favorites', 'HP', 'Links', 'Music', 'OneDrive', 'Pictures', 'Saved Games', 'Searches', 'Videos', and 'bash_profile'. The table columns are 'Name', 'Last Modified', and 'File size'. The status bar at the bottom of the browser shows the system time as 11:18 AM on 2/23/2024.