Intro to Programming for Data Analysts

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

Python is a popular choice for data analysts due to its open-source nature and free availability, avoiding the costly licensing fees of statistical software like SPSS, Stata, and SAS. The reproducibility of Python code, ability to handle massive amounts of data beyond Excel's limits, and faster data processing capabilities further contribute to its widespread adoption among data analysts.

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).

Based on my research, here are 5 top companies in the world that use Python, either for software engineering or analytics:

1. Google Google makes extensive use of Python for various purposes, including components of the Google Search engine, YouTube video management, and many other products and services. Python is one of the official server-side languages at Google.
2. Netflix Python plays a crucial role in Netflix's data pipeline, recommendation engines, and content delivery networks. The company also uses Python for its machine learning and data analysis applications.
3. Spotify Spotify relies on Python for its core data pipelines, data analysis, and machine learning models that power music recommendations and personalization features.
4. Facebook Facebook uses Python for various purposes, including data analysis, content management, and infrastructure engineering. The company's Augmented Traffic Control tool, which manages network traffic, is written in Python.
5. Dropbox Dropbox has integrated Python into many aspects of its operations, including the Dropbox desktop client, server backend, and

data analysis tools. Python's simplicity and scalability make it a good fit for Dropbox's needs.

These companies have embraced Python for its versatility, extensive libraries, and suitability for various applications, ranging from web development and data processing to machine learning and scientific computing.

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.**

  For this scenario, I would use Microsoft Excel. It is a powerful spreadsheet application that is well-suited for small to medium-sized datasets. Excel provides easy-to-use functions for filtering, sorting, and basic data manipulation. Additionally, it offers built-in charting and visualization capabilities, making it convenient to create quick charts from the data**.**

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

  In this scenario, I would use SQL as it is excellent tools for working with large databases and retrieving specific subsets of data. It is a domain-specific language designed specifically for managing and querying relational databases.

- **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**.

  For a dataset of this size and complexity, I would use a programming language like Python or R. Both languages are well-suited for handling large datasets and performing advanced data manipulation and analysis tasks.

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