Task 4.1:
Intro to Programming for Data Analysts

1. Python is immensely popular among data analysts due to its rich ecosystem of libraries and tools Specifically tailored for data manipulation, analysis, and visualization. Libraries like Pandas, NumPy, and matplotlib provide powerful data manipulation capabilities, statistical functions, and data visualization tools, enabling analysts to efficiently clean, process, and visualize data. Pythons' intuitive syntax and active community supports make it accessible for professionals from diverse backgrounds, further contributing to its widespread adoptions in the data analysis field.

2. **Companies that Use Python:** several top companies use Python extensively for software engineering and data analytics 5 prominent companies that are known to use Python are followimg:

- **Google:** Python is widely used at Google for various purposes, including web development, automation, and scripting. Google also maintains open-source Python projects, such as TensorFlow and Colab, which are popular in the machine learning community.

- **Facebook:** Python is used at Facebook for a range of tasks, including web development, infrastructure management, and data analysis. The popularity of the Django web framework within Facebook is one example of Python's usage.

- **Amazon:** Python is employed at Amazon for various backend services, scripting, and automation. The AWS SDK for Python (Boto3) allows developers to interact with Amazon Web Services, making Python a valuable tool for cloud computing.

- **Microsoft**: Python is increasingly integrated into Microsoft's ecosystem, with support in products like Azure, Visual Studio Code, and Power BI. Python is used for web development, scripting, and data analysis within the company.

- **Netflix:** Python is used at Netflix for various purposes, including data analysis, recommendation algorithms, and content delivery.

Python's simplicity and versatility make it valuable for rapid development in Netflix's dynamic environment.

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

- For this scenario where I have a small data set that needs quick tweaks and minor analysis, I would use a spreadsheet software like Microsoft Excel or Google Sheets. These tools allow me to easily filter columns make quick data manipulations and create basic charts without the need for complex programming or data processing.

- When I need to retrieve a portion of data from a very large database, I would use query language like SQL (Structured Query Language). SQL allows efficiently retrieve specific subsets of data from large databases by writing queries tailored to the needs, minimizing the amount of data transferred and improving performance.

- Dealing with a large data set of 15,000,000 rows and 350 columns that need sorting and preparation for advance analyzes would require a data processing tool such as Python with libraries like “Pandas” and “NumPy”. These libraries provide powerful data manipulation and analysis capabilities, enabling efficiently sort, filter, and process large data sets, while also preparing them for more complex analytical tasks.

4.