Task Sheet 4.1

Data Immersion

Ashwani Sherawat

Python's popularity among data analysts stems from its simplicity, versatility, and extensive ecosystem of libraries tailored for data manipulation, analysis, and visualization. Its readable syntax and comprehensive community support make it accessible for both beginners and experienced professionals.

Top global companies utilizing Python include:

1. **Google**: Employs Python for various applications, including web crawling, data analysis, and machine learning

2. **Facebook**: Uses Python for infrastructure management and operational automation.

3. **Instagram**: Relies on Python, particularly the Django framework, to handle its vast user base and data processing

4. **Spotify:** Utilizes Python for back-end services and data analysis to enhance user experience through recommendations.

5. **Netflix:** Leverages Python for data analysis and backend services to optimize streaming and content recommendations.

For the given scenarios:

1. For a small dataset requiring quick tweaks and minor analysis, **Excel** can be an ideal tool. You can easily use Excel’s filters and sorting options to manipulate data, apply conditional formatting for quick insights, and create charts like bar graphs, line charts, or scatter plots directly within Excel. Excel’s pivot tables also allow you to summarize data efficiently for straightforward analysis.

2. Retrieving data from a very large database: Utilize **SQL** for database interaction. SQL is a powerful tool and its unique approach allow us to select precise data from database with Objectional handling from large databases.

3. For a dataset with 15,000,000 rows and 350 columns that needs to be sorted and prepared for advanced analysis, **Python** provides powerful tools and libraries that handle large data efficiently.

Python will allow us to provide insight on big data sets with the help of different libraries e.g numpy, and pandas.

I am using a mac machine so skipping this step.

Installed Jupyter interface Screenshot.

A screenshot of a computer

Description automatically generated