Task 4.1

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

Python is very flexible, it is a playground that allows data analyst to be creative. It is also easy to use because of its simplicity and readability. This allows users to learn it quickly making it a great tool for beginners. Python requires fewer lines of code than some other programming languages. Python is free which allows data analyst all over the world to have access. It is also well supported allowing access to lots of libraries and information.

2. 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).

Google, Netflix, IBM, Amazon, Nasa

3. 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 will need to filter some columns and make a quick chart.

   Excel- Excel is simple and however it allows a user to perform task quickly. Excel handles small data well and has filtering features as well as chart features.

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

   SQL- SQL has data bases that are equipped to handle large amounts of data. A user just must write a command to retrieve information.

   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.

   Python- Python can easily handle more than 15000,000 rows and 350 columns. It can perform very advanced analysis when users write codes in programming language to sort data. Excel cannot handle such a large about of data and SQL would not be as fast as python.

4. Download Anaconda.