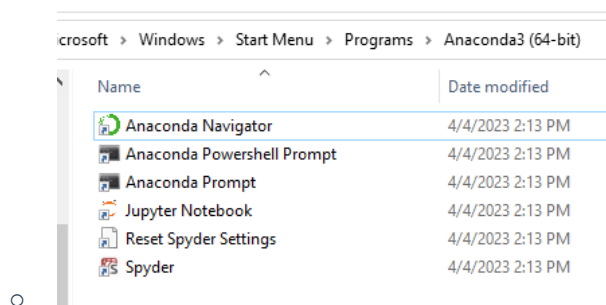


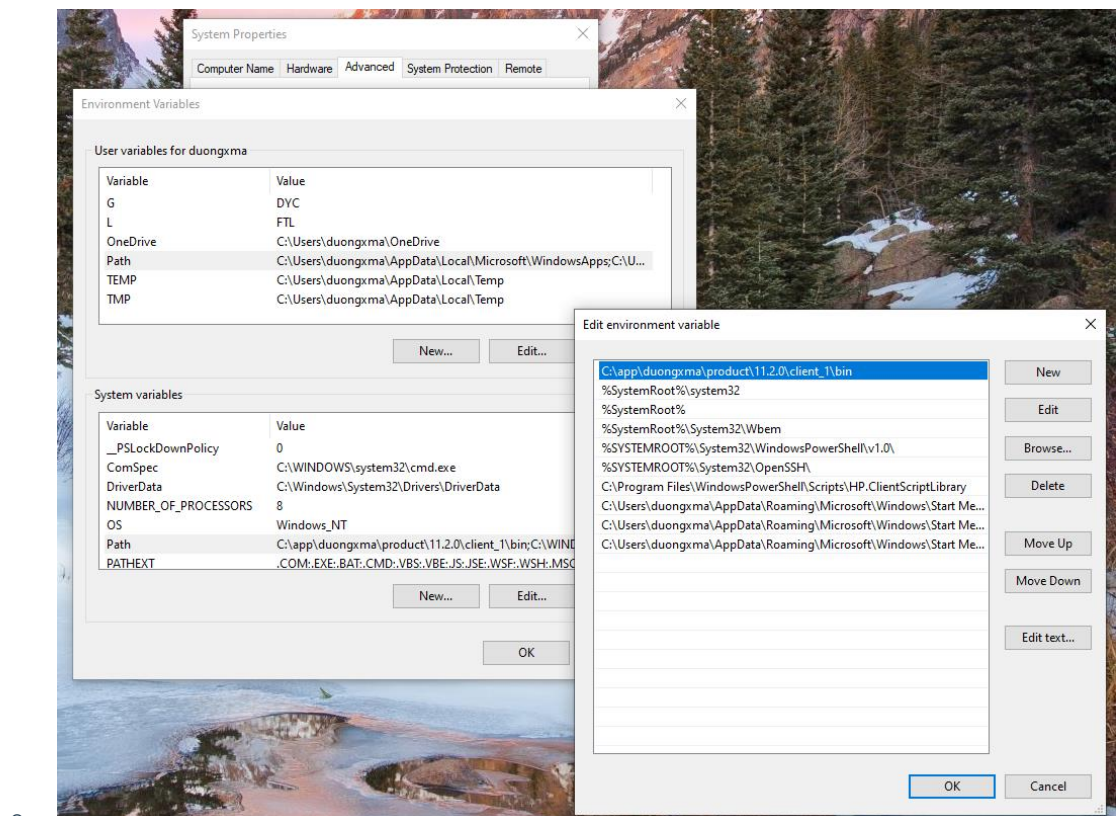
1. Set up a document where you can write your answers to the following questions. You'll want to conduct some research online to help you develop your answers.
2. Write 2 to 3 sentences on why Python is so popular among data analysts.
  - Python is a programming language that can be used for so much more than data analysis, making it versatile and beneficial for different tasks and in different capacities. Python is also known to be an easy-to-learn language comparatively to other programming languages despite being so powerful, making it accessible to beginners in addition to its wide library selection of readily available codes. Since Python is so popular and widely used, having this skill makes data analysts competitive in the job market.
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).
  - "Python is used by Intel, IBM, NASA, Pixar, Netflix, Facebook, JP Morgan Chase, Spotify, and a number of other massive companies. It's one of the four main languages at Google, while Google's YouTube is largely written in Python." Source: [BrainStation](#)
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.
    1. Excel would be the best tool to use to make quick, simple fix for a small dataset. It doesn't make sense to complicate the process more than it needs to be through using more advanced programs or coding. Excel has all we need to filter columns and make a quick chart.
  - You need to retrieve some portion of data from a very large database.
    1. SQL would be best for this whether through PostgreSQL as we learned in the previous achievement, or through Microsoft Access's query tools. Excel wouldn't typically be connected to a database, nor would it be able to store that many data points, so it wouldn't make sense to use it. SQL would be the "vessel" we use to pull the requested data out of the database.

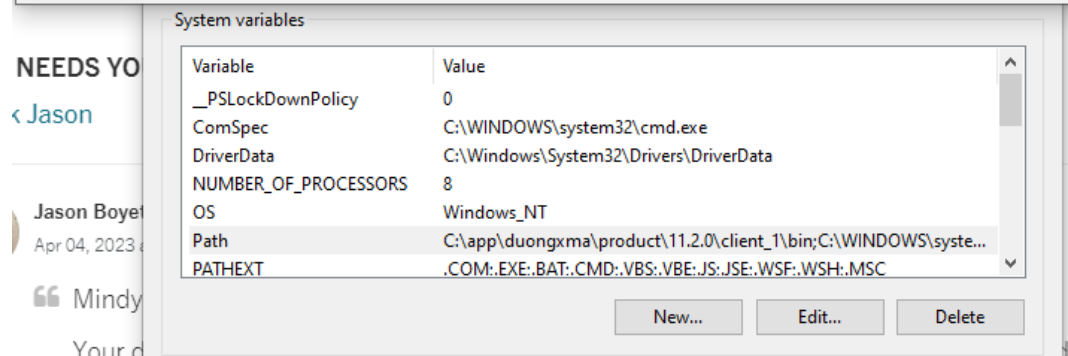
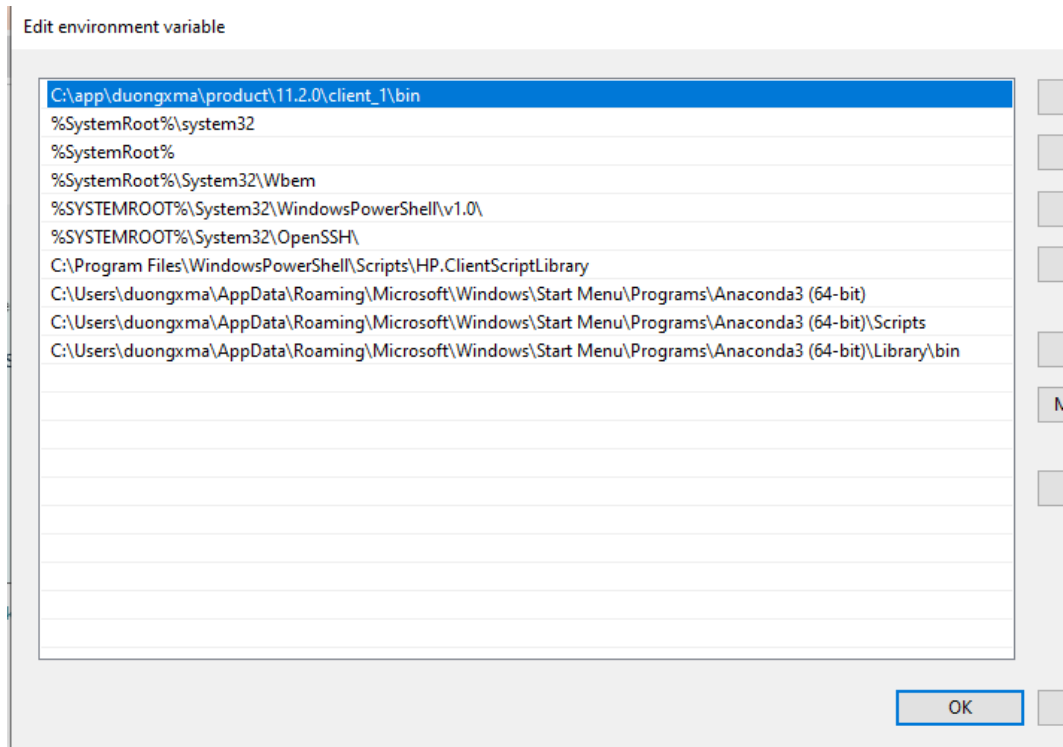
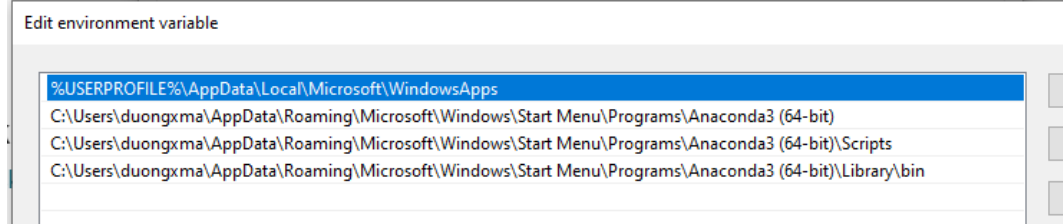
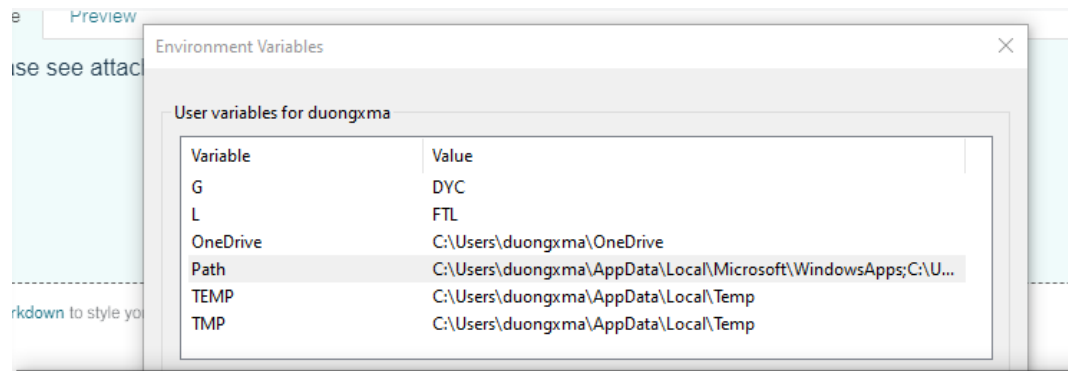
- You have a data set with 15,000,000 rows and 350 columns that needs to be sorted and prepared for a more advanced analysis.
  1. While SQL may be used for this depending on how well-versed you are and how sophisticated your SQL code, it would be better to use Python as it's designed to handle much larger datasets and perform advanced analytics. There's no way Excel would be able to handle this with the sheer number of rows and columns.

## 5. Download Anaconda.



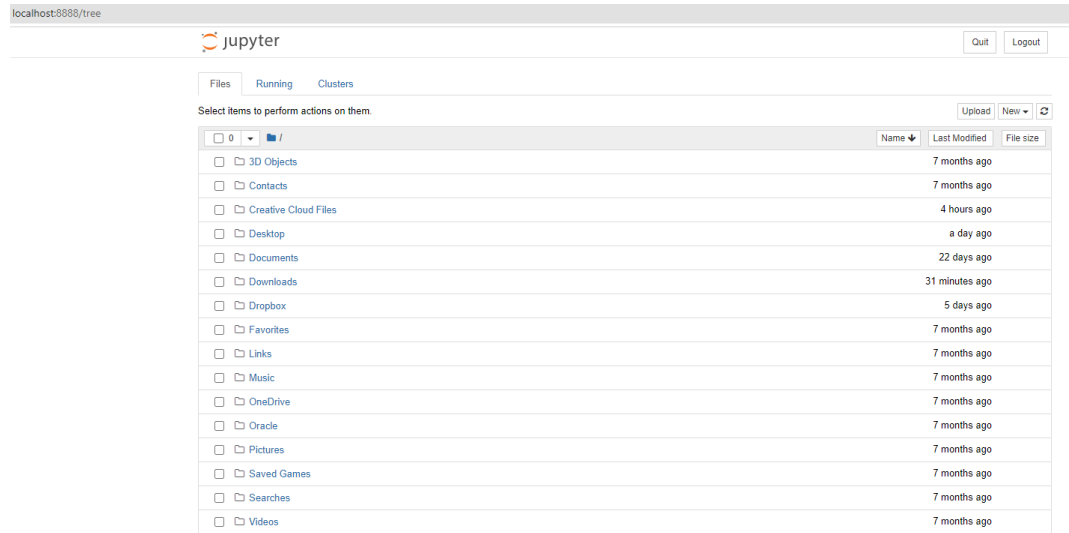
6. Set up the environment variables on your computer and copy them into your document together with your answers to steps 2 through 4.





7. Launch Jupyter.

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



9. Copy-paste this screenshot into your document, export your document as a PDF, then submit the PDF here for your Tutor to review.