Final Project Overview:

For my final project, I chose to analyze an undirected Amazon graph where nodes are products and edges represent if they are often purchased together. I think this is an important graph to analyze because you can get a lot of consumer insight from looking at their spending habits: for example, knowing what is often copurchased together allows you to spend more on marketing those objects together, and if you know certain products are less likely to be bought together, you can tweak your algorithm to not waste time or money marketing those products together. Conversely, if something is not often purchased with something else, maybe you would want your marketing team to work harder on promoting those objects together. All in all, consumer insights can be important to generate more revenue.

My code calculates a few different metrics. One is the average number of connections there are across the graph. This is important if you want to see if your graph is very interconnected or not — more connections means more copurchases, so you could see if your marketing is working or not.

Second, is the maximum number of connections. This is important because if you have a maximum that greatly differs from the average then you know you have a more skewed dataset. Ideally, you should have a graph where all the nodes have a similar amount of copurchases, since that means that all products incentivize consumers to purchase about the same amount. Alternatively, if you have a specific product that seems to really incentivize more people to buy things with it, maybe your company can work harder on marketing that so that more revenue is generated from all the additional purchases.

The third metric takes in user input for a specific node that they want to see more data about. The script then calculates the distance from that specific node to all other nodes using a breadth first search. This is important because if you are interested in looking at the dynamics between specific products, you can see if certain products are more linked than others by inputting the node corresponding to the product you are interested in.

Finally, it calculates the node that is furthest from the node inputted. This is helpful for looking at what product is least likely to be copurchased or associated with a certain product.

Sample Output: this is sample output where the user inputs node 6.

Here's some information about your graph. The average connections a node has is 3, the maximum amount of connections is 549,
and the minimum number of connections is 0
Please input the node number that you want to see the distances to:
** all the distances go here after entering a node, but it’s too much to copy paste into this file**
the maximum distance is 33 from node 150736