**D212 Data Mining II**

Nicole Reiswig

College of Information Technology, Western Governors University

Master of Science in Data Analytics

Kesselly Kamara

February 21, 2024

**Table of Contents**

**Part I: Research Question**

1. Question 3
2. Goal 3

**Part II: Market Basket Justification**

1. Method 3
2. Transaction example 4
3. Assumption 5

**Part III: Data Preparation and Analysis**

1. Clean data 5
2. Code execution 5
3. Rules table 6
4. Three rules 7

**Part IV: Analysis**

1. Summary 12
2. Practical significance 13
3. Recommendation 13

**Part V: References 13**

# Principal Component Analysis

**Part I: Research Question**

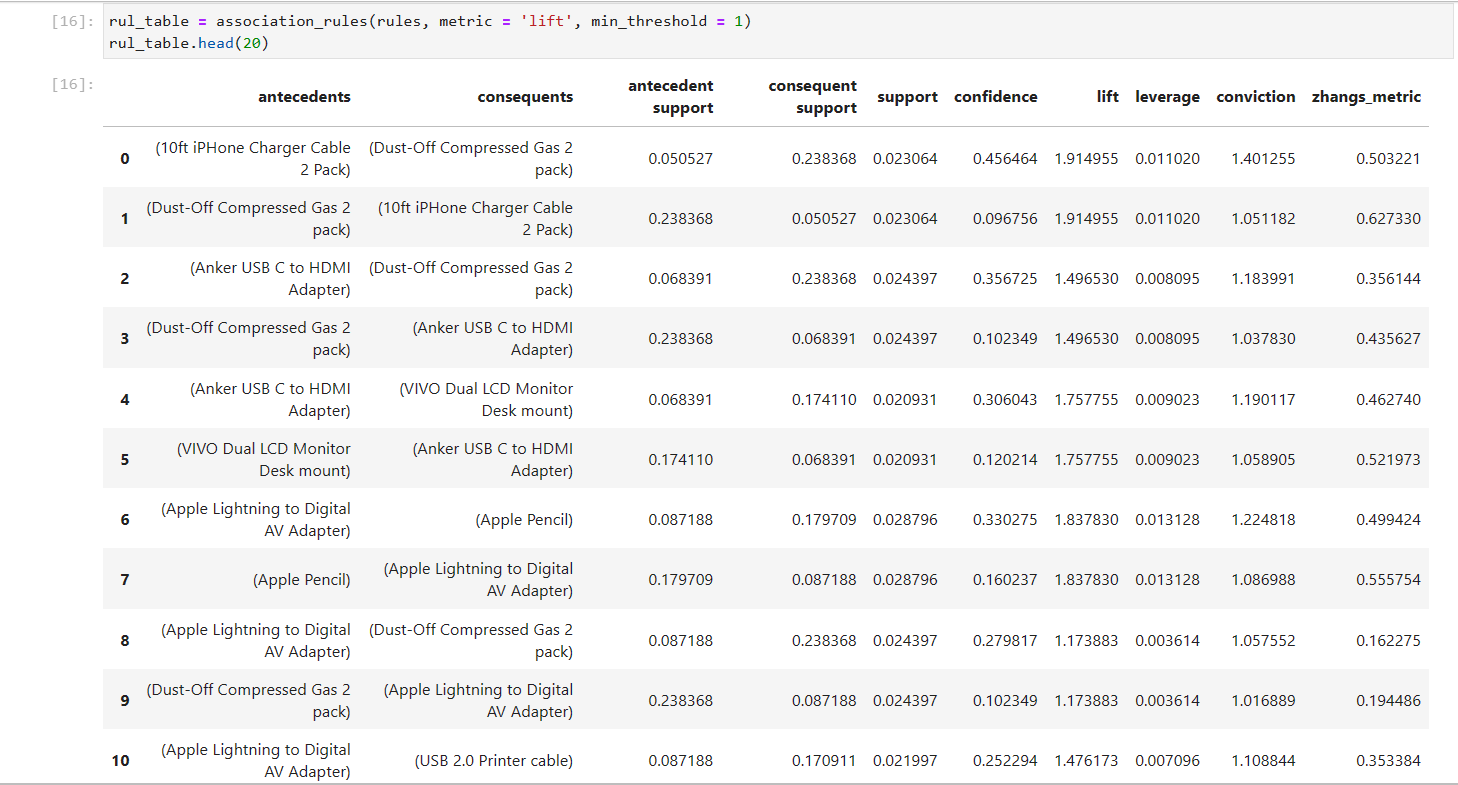
1. The research question that will be answered is, “Which products are most frequently purchased together?” We will identify characteristics of our customers' spending habits to understand better who our customers are and how to best provide products to them. We will identify groups of customers with similar purchasing habits ultimately enabling better business and strategic decision-making. Market basket analysis is the best approach for this research question.
2. Market basket analysis aims to answer our research question, “Which products are most frequently purchased together?” We will utilize this method to identify customers' spending habits and identify which items are frequently purchased together. The goal is to make better business decisions and with this information, we can use the information to create more strategic marketing goals such as product placement, pricing, cross-selling, or upselling.

**Part II: Technique Justification**

1. Market basket analysis is a data mining approach that analyzes the data by finding products that are most likely to be purchased or grouped. Market basket analysis is common in the retail industry to boost sales and understand buying habits. An example of this we are all familiar with is when shopping during any given holiday. Many times, common foods purchased for Thanksgiving will be grouped at the grocery store. You will find green beans, cream of mushroom soup, and french-fried onions on the same aisle when they typically would be in different sections during the off-season. Another example would be Valentine's Day, to increase spending they place all the items purchased frequently together near that holiday such as chocolates, stuffed animals, flowers, etc on display when you first enter the store. The expected outcome is that customers will spend more money on these items. This is accomplished by evaluating big data spending patterns and making it easier for customers to accomplish. It has to do with if and then. If a customer needs X the antecedent, then they also likely will need Y the consequent.

In addition to the if and then analysis there are rules that can be put in place to analyze further. The rules are support, confidence and lift.

1. An example transaction from the data set is below:

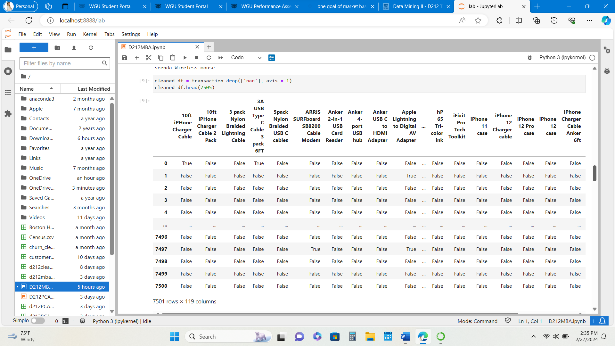
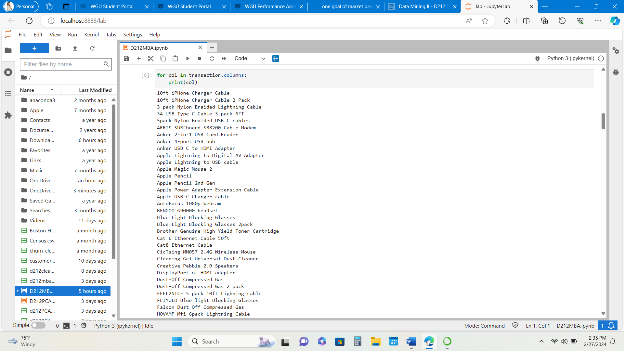
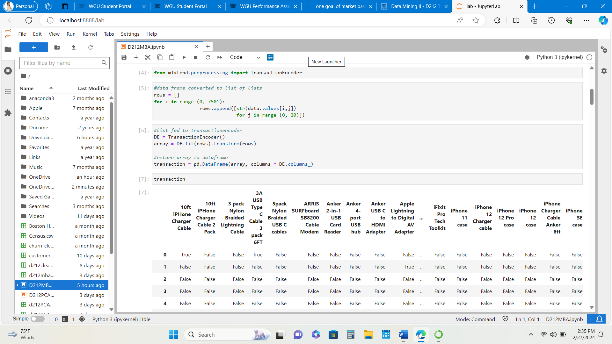
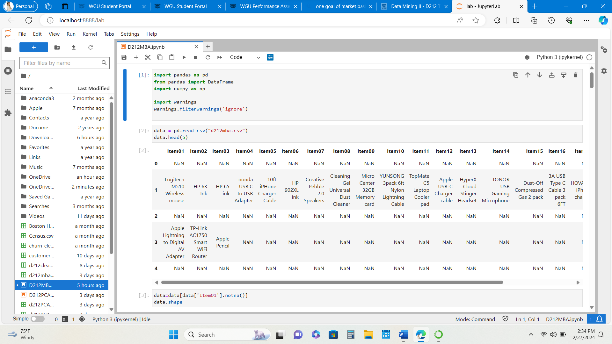


Here you can see the lift rule table. This table lists 20 transactions (10 of which are pictured) and tells you which product the antecedent and which product is the consequent to that antecedent purchase. For purchase 0 a 10 ft iPhone charger cable 2 pack was purchased, and the consequent was a 2 pack of dust-off compressed gas. In purchase 1, the same two items were purchased together but in the opposite order, the dust-off gas was the antecedent and the charger was the consequent. Market basket analysis looks for associations and patterns such as these to make better-informed business decisions.

1. Market basket analysis assumes that if products are frequently purchased together, they are compliments of one another.

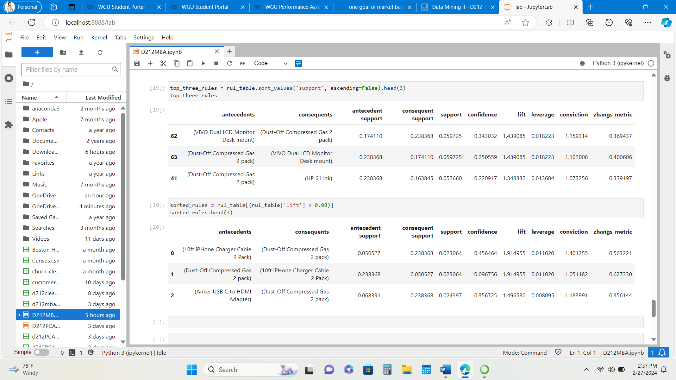
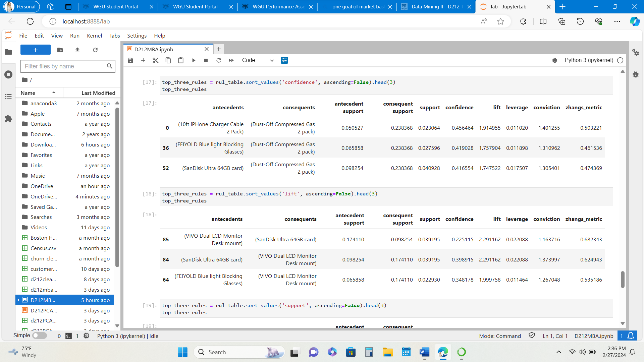
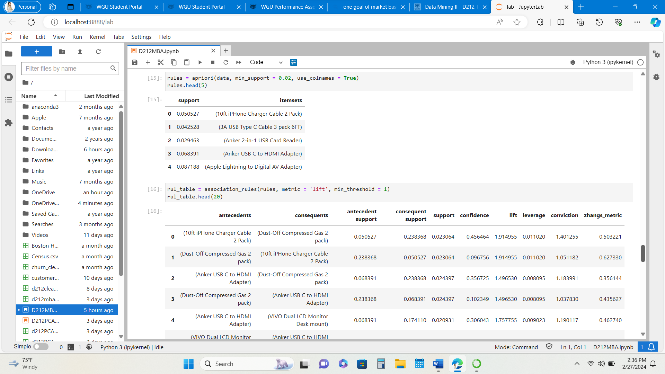
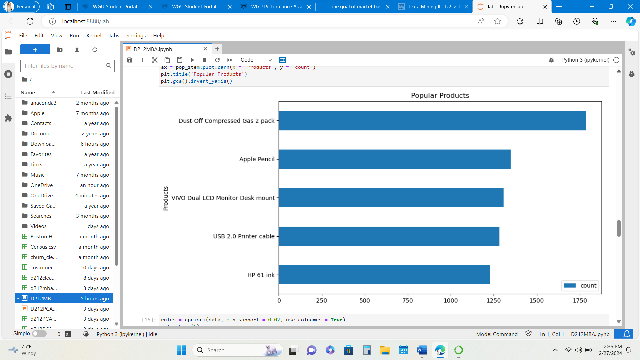
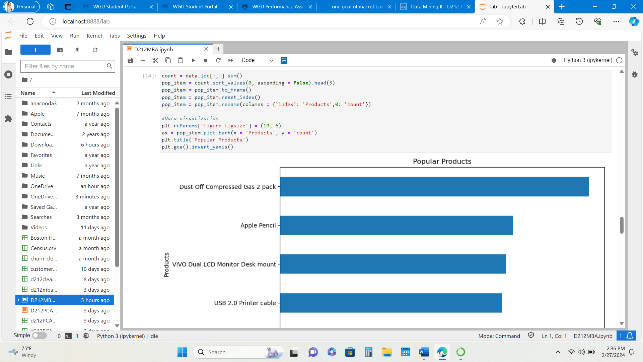
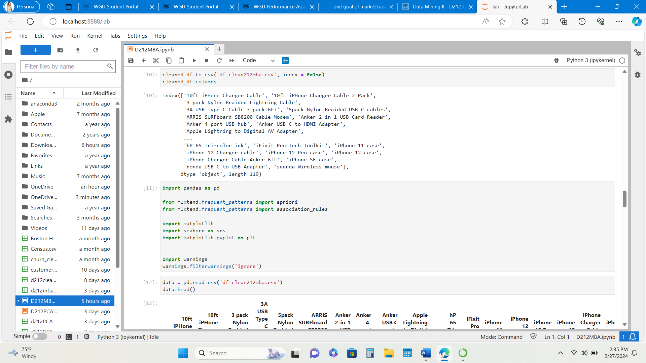
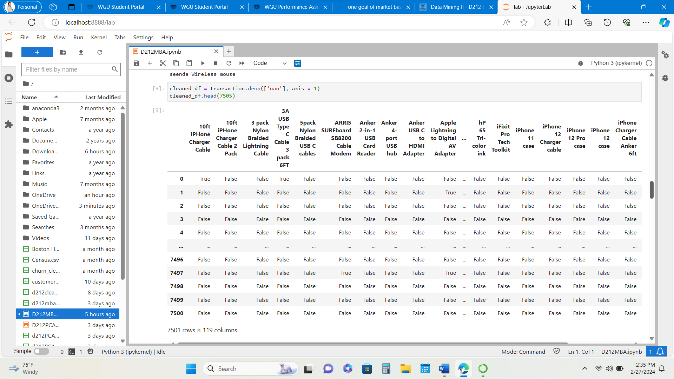
**Part III: Data Preparation and Analysis**

1. Please see the attached cleaned data set.

A screenshot of a computer

Description automatically generated

1. See the attached executable code.



1. The values for support, lift and confidence are included in the below rules table:

A screenshot of a computer

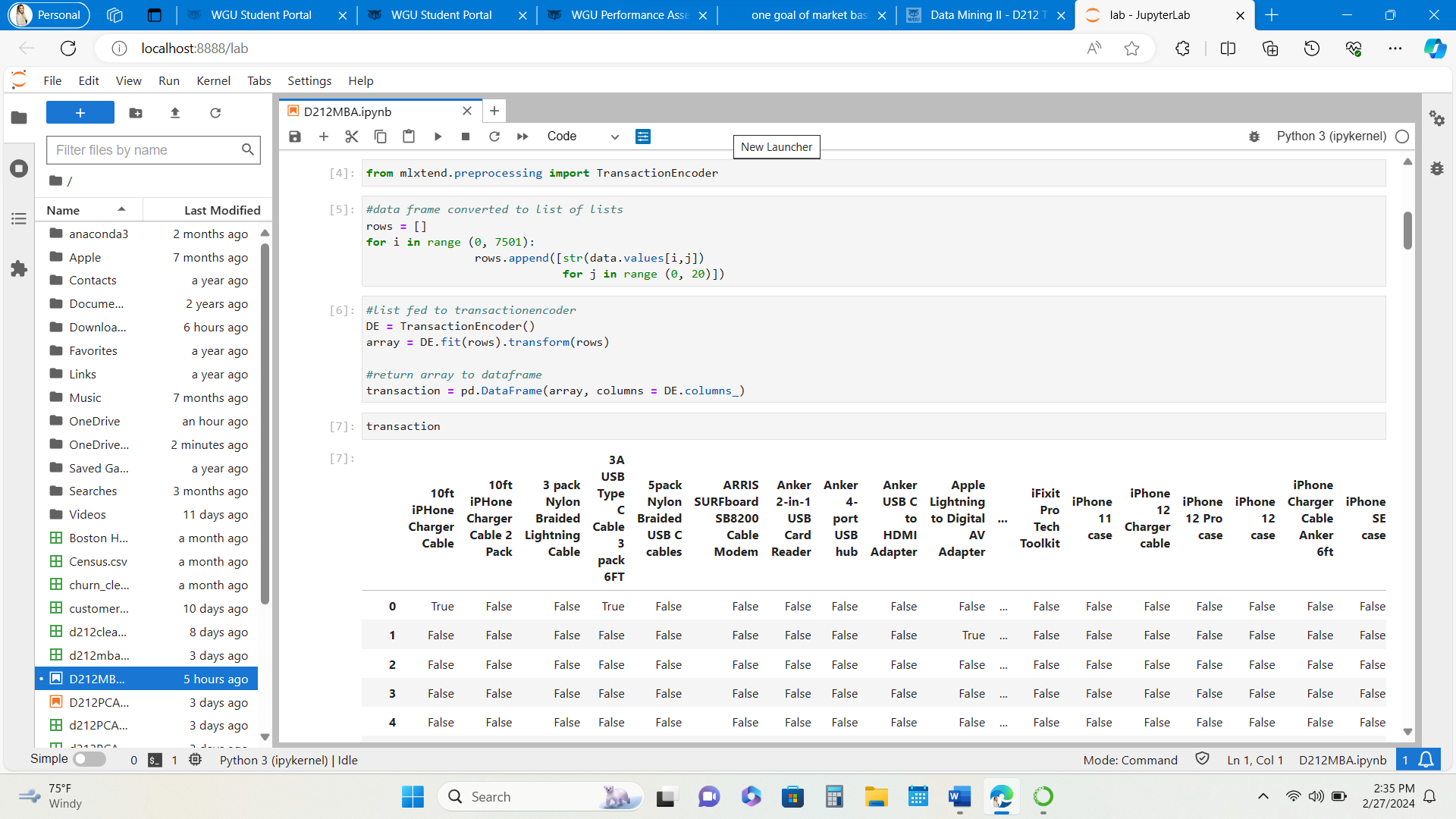
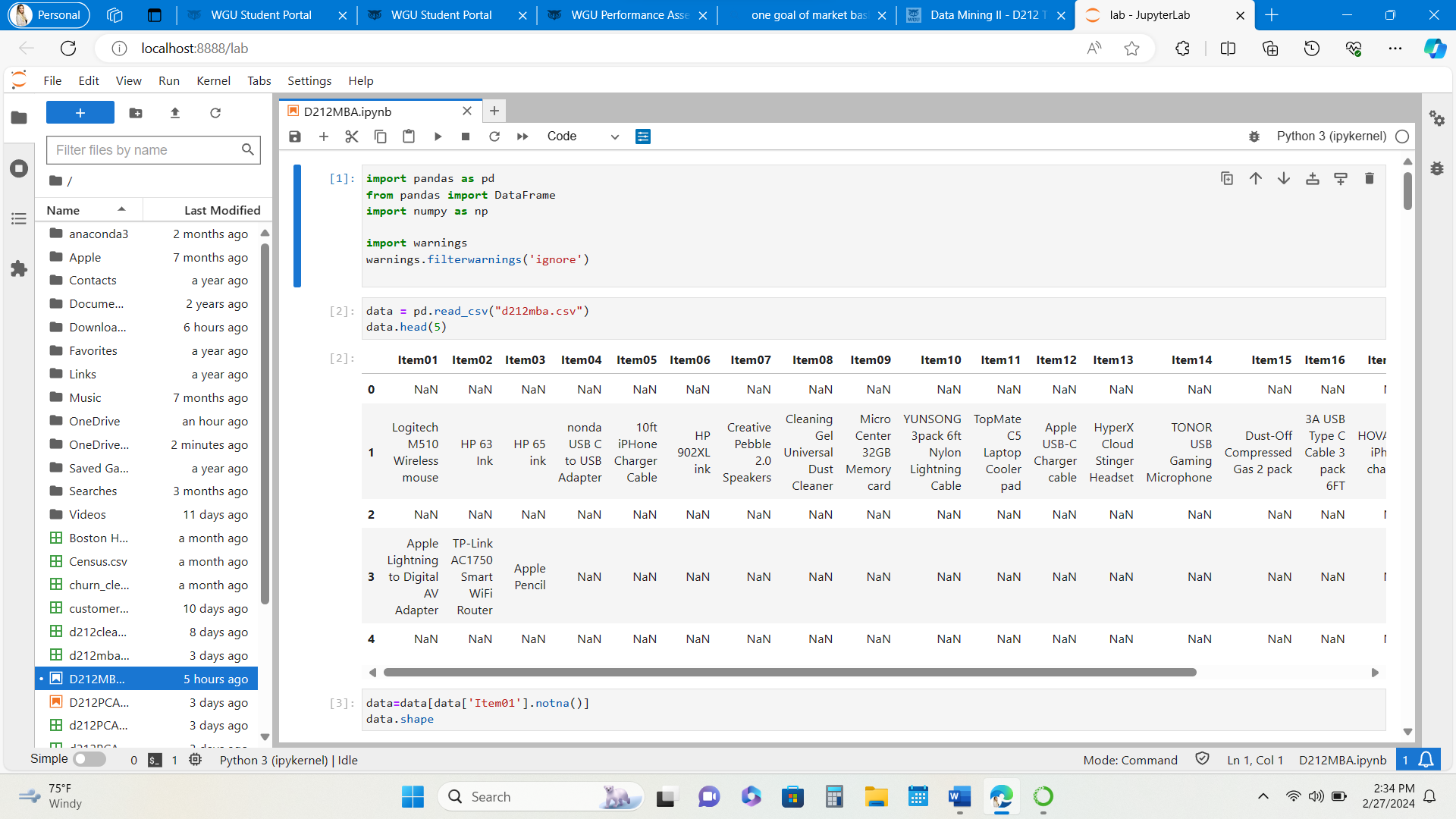
Description automatically generated

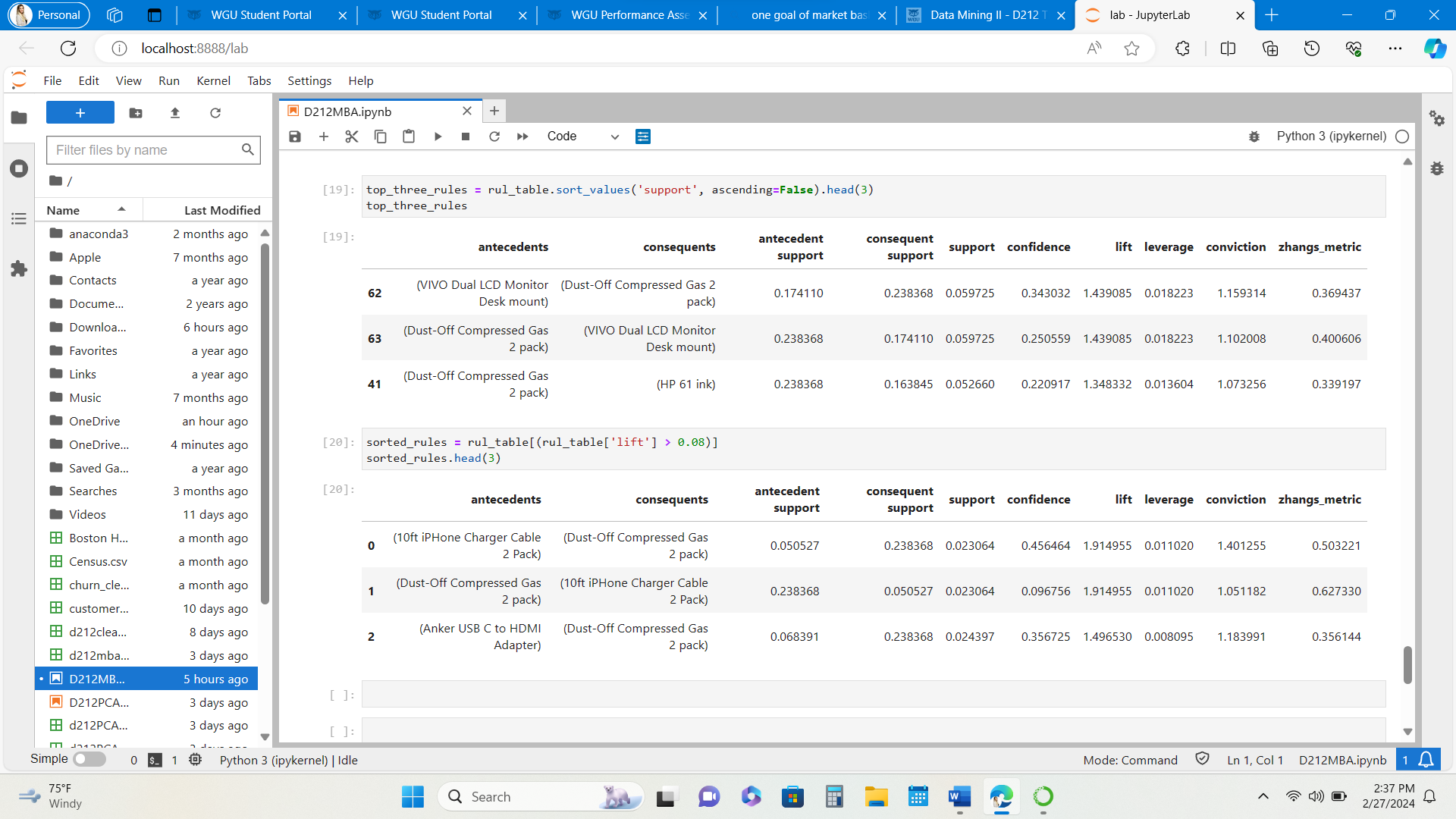
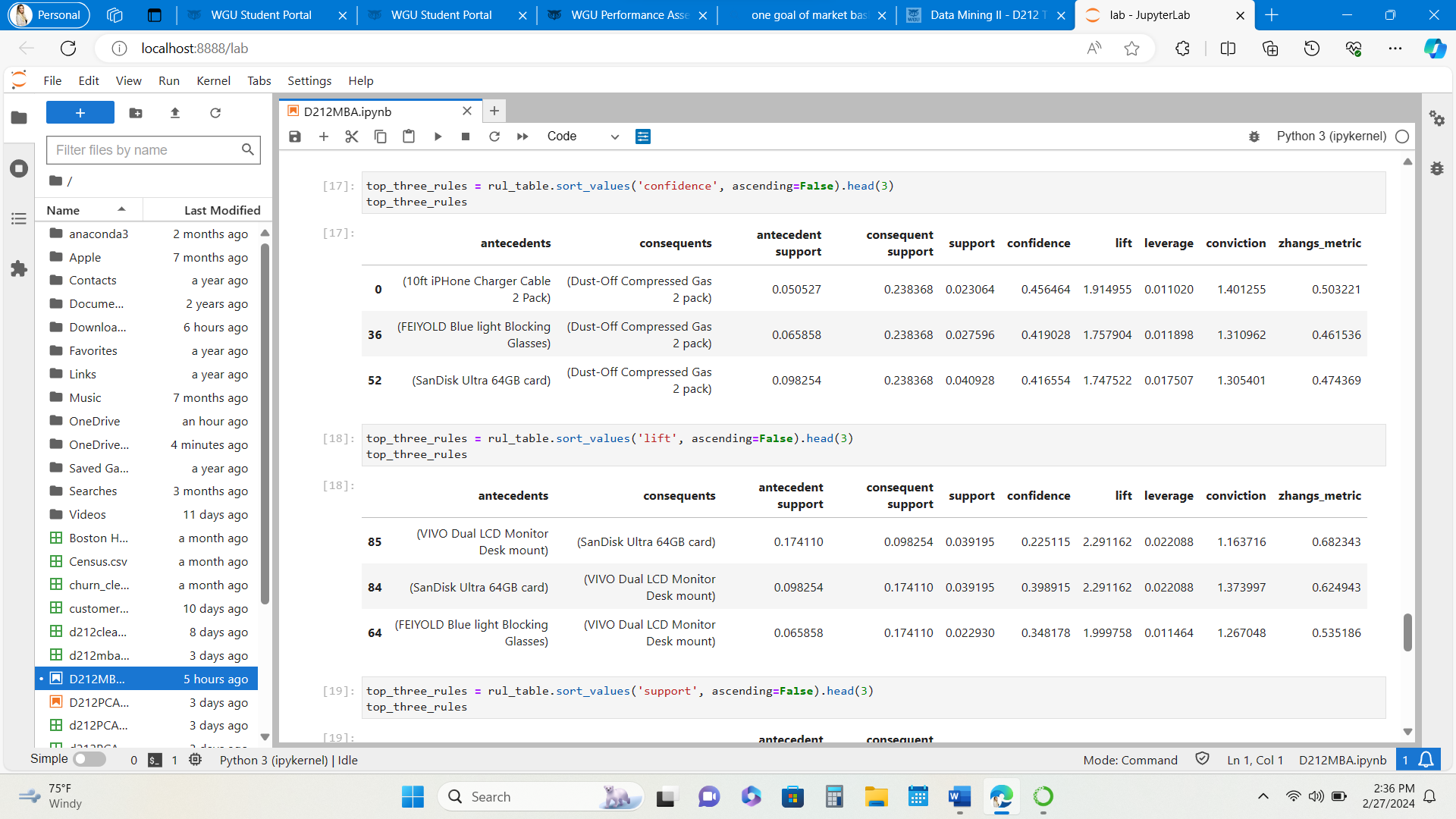
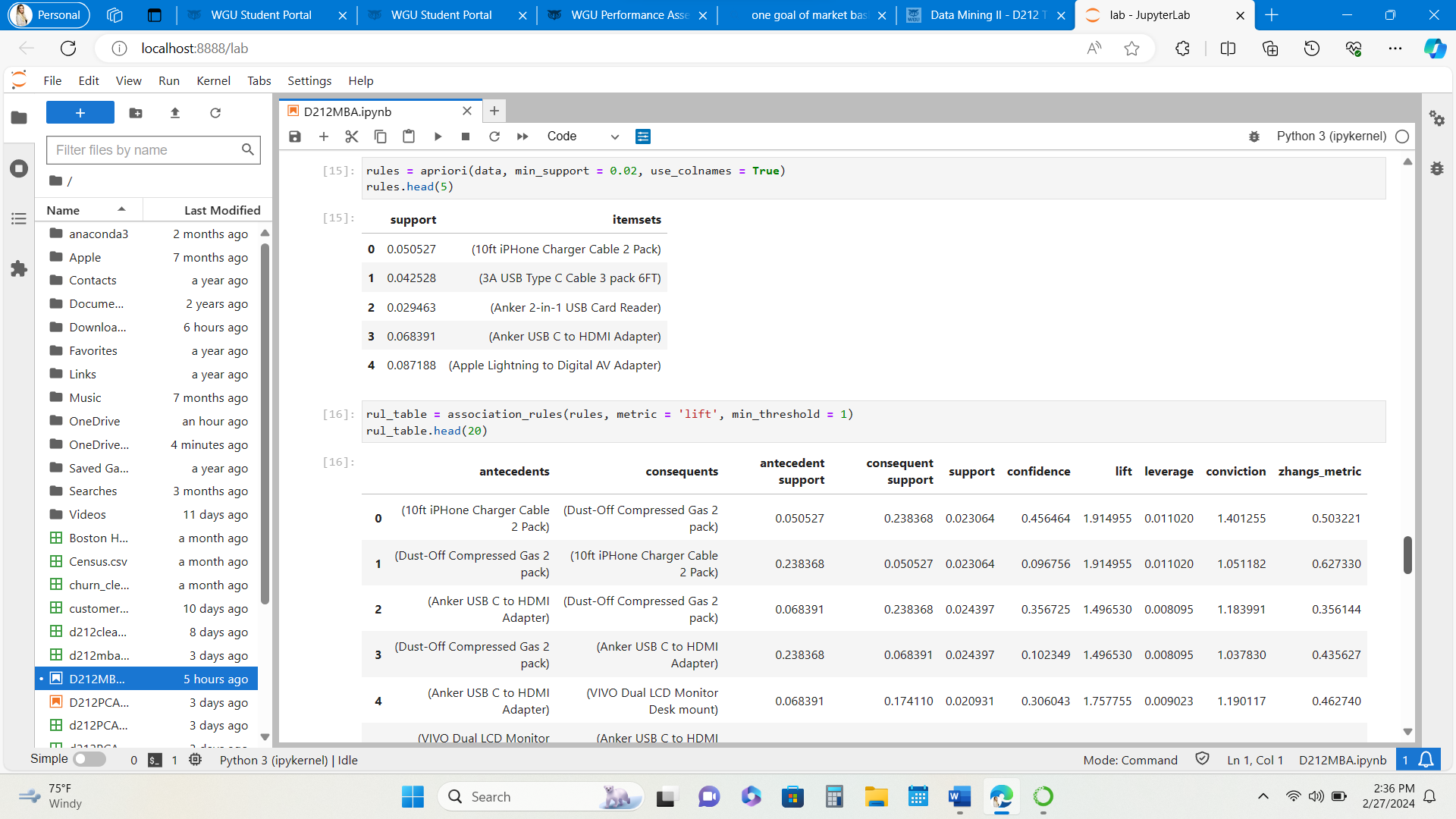
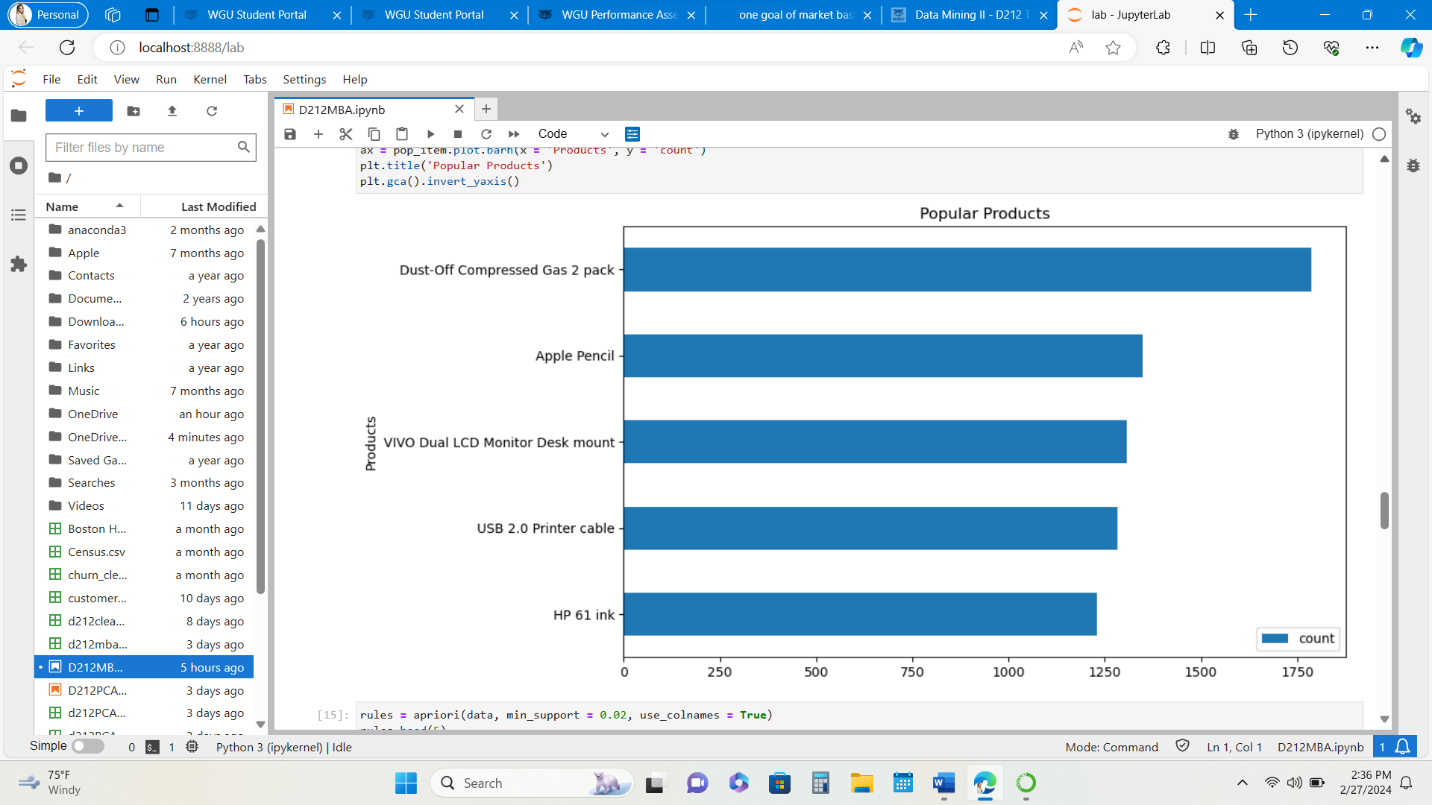
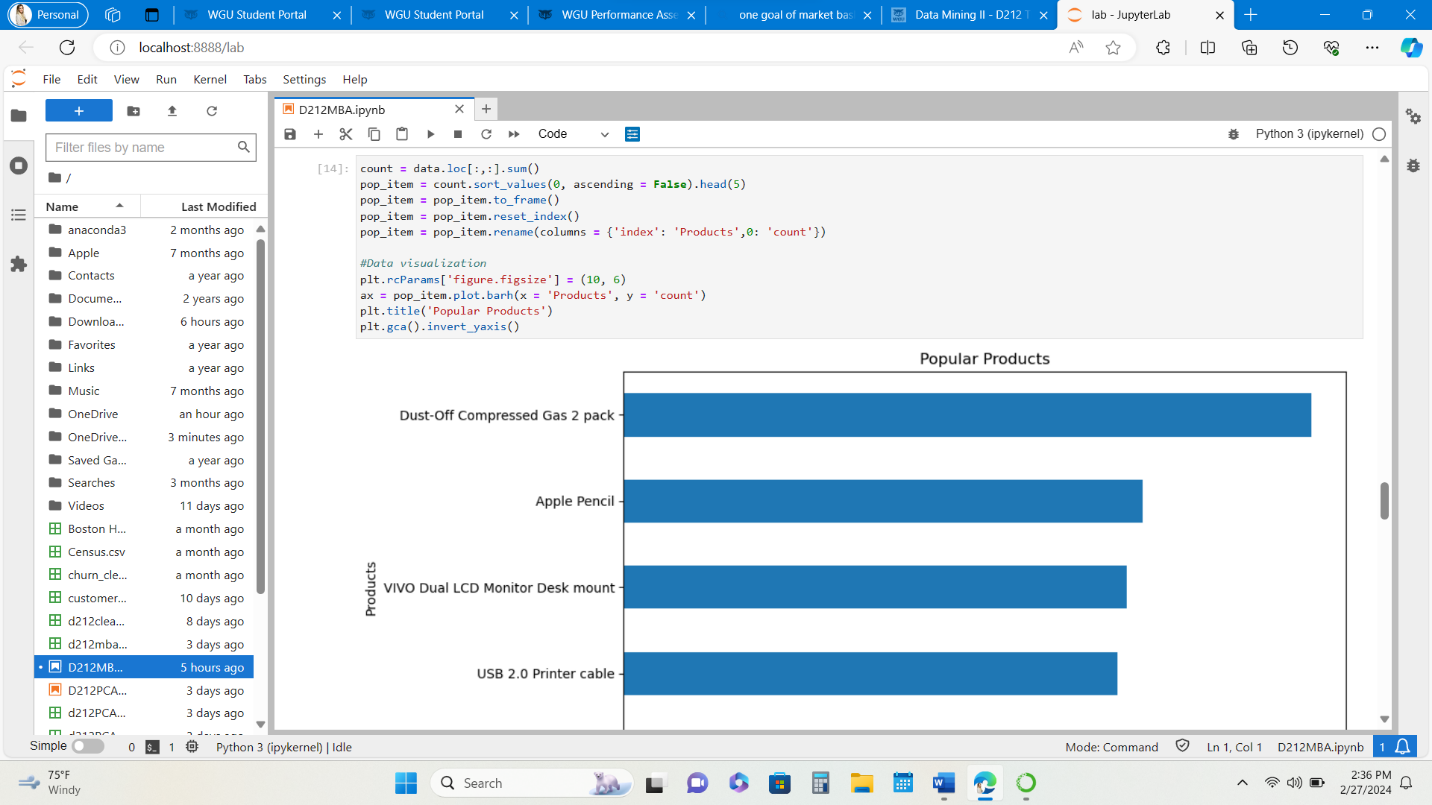
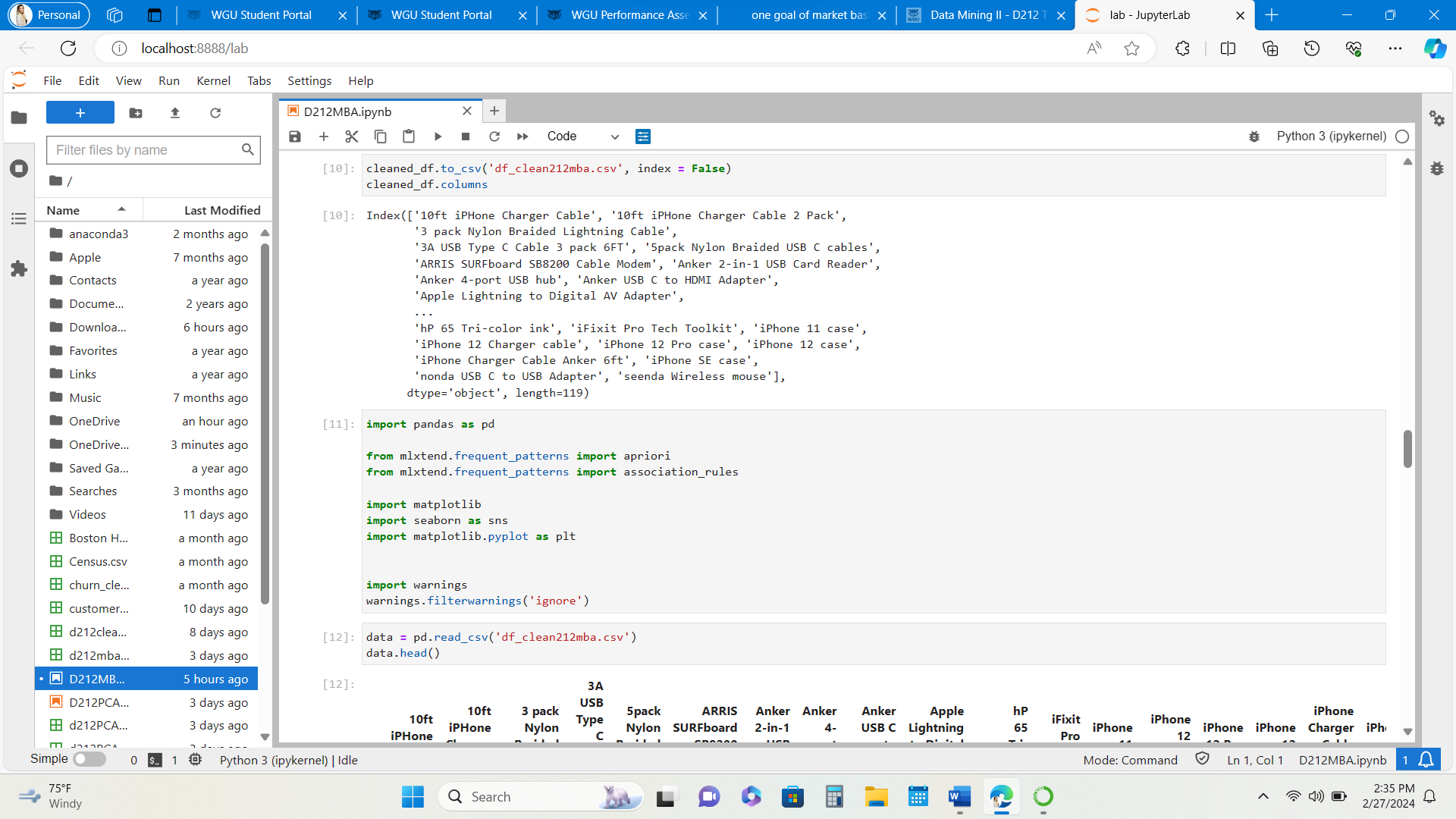
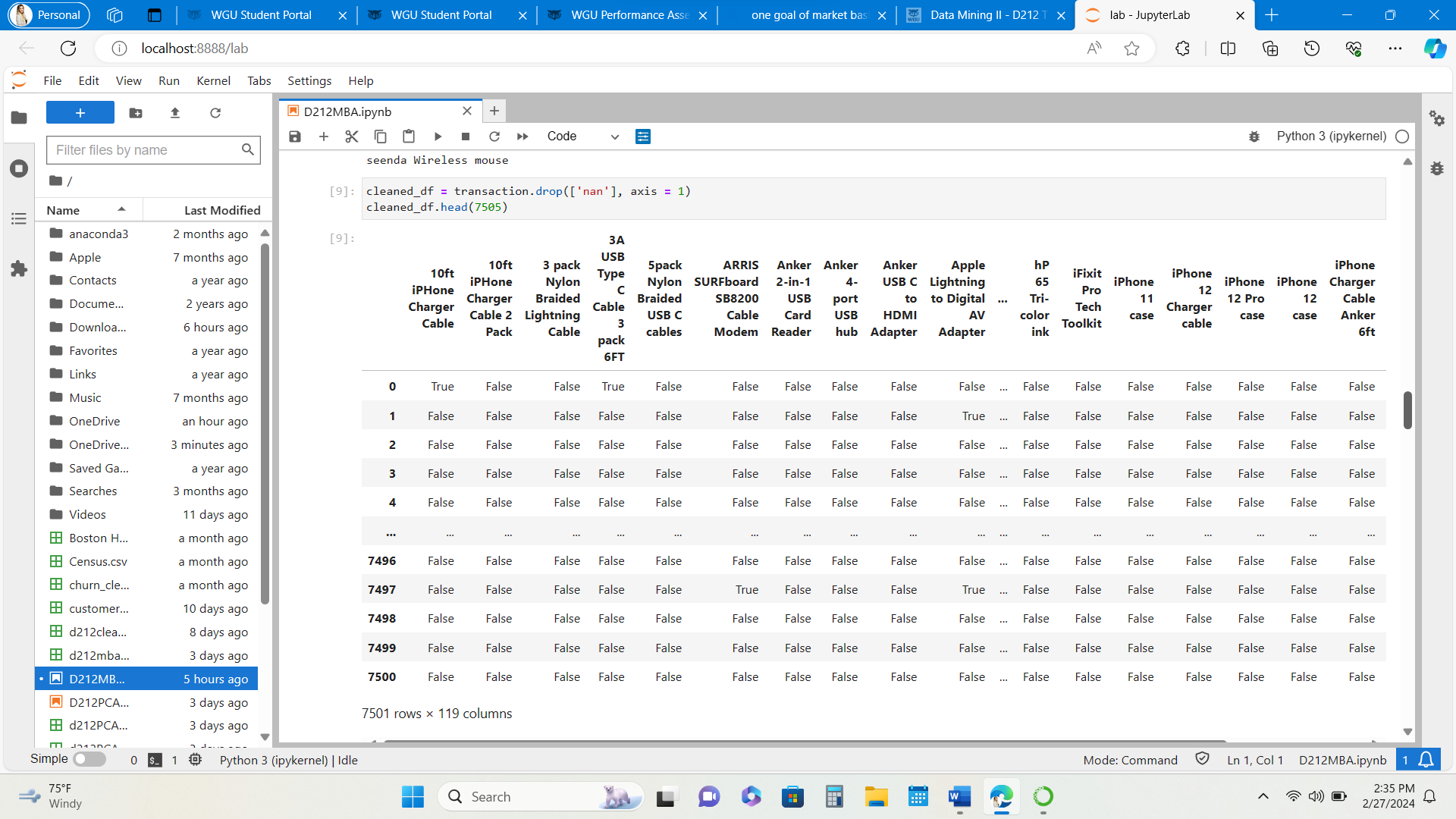
1. The top three relevant rules are in the table below:

A screenshot of a computer

Description automatically generated

The lift score is defined by an association rule. If greater than 1 then the antecedent and consequent are dependent on one another. The confidence score is also defined by an association rule and indicates how likely it is that the products are purchased together. The support is the probability that an item set is present in a transaction.

A screenshot of a computer

Description automatically generated

**Part IV: Analysis**

1. In summary, the top three rules table, line 0 has a support of .023, a confidence of .456, and a lift of 1.91. Line 1 has a support of .023, a confidence of .096, and a lift of 1.91. Line 2 has a support of .238, a confidence of .356, and a lift of 1.49. The lift score is defined by an association rule. If greater than 1 then the antecedent and consequent are dependent on one another. The confidence score is also defined by an association rule and indicates how likely it is that the products are purchased together. The support is the probability that an item set is present in a transaction.
2. The results are practically significant. The items that are frequently purchased together are items, that have a high demand. Technology is utilized by the majority of the population. It’s not infrequent to misplace a charger or to need an adapter. We rely heavily on electronics and these items are necessary for proper functioning. I do see these items in the checkout lines at stores and near each other in the electronic section which indicates other stores may have similar data and analysis.
3. The recommended course of action is to place the iPhone charger, dust off compressed gas, and the USB to HDMI adapter next to each other at the store. These could be placed when you first enter the store and/or at the check-out lines. This will provide optimal location, convenience, and mind prompting to customers shopping in the store and increase sales.

**References:**

I utilized Dr. Kesselly Kamaras' WGU instructor videos to reference segments of code to perform this analysis and D206 data cleaning course materials.

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9541a29b-2f14-4c5d-9d86-af030005bcf6>

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=db85c4f1-0da5-4bde-a1a4-b07c0019d46d>