Hardik Shahu

Discovering Associations

Overview of the Tasks

- Should we acquire Electronidex?
- Are there any associations between different products?

Gathering the Data

For this project, we were given a dataset containing one month's worth of Electronidex's clientele. Unlike the previous tasks, the data does not contain any attributes as there aren't any columns like we've seen previously. Instead, each row represents a different transaction made by a customer.

Looking at the Data

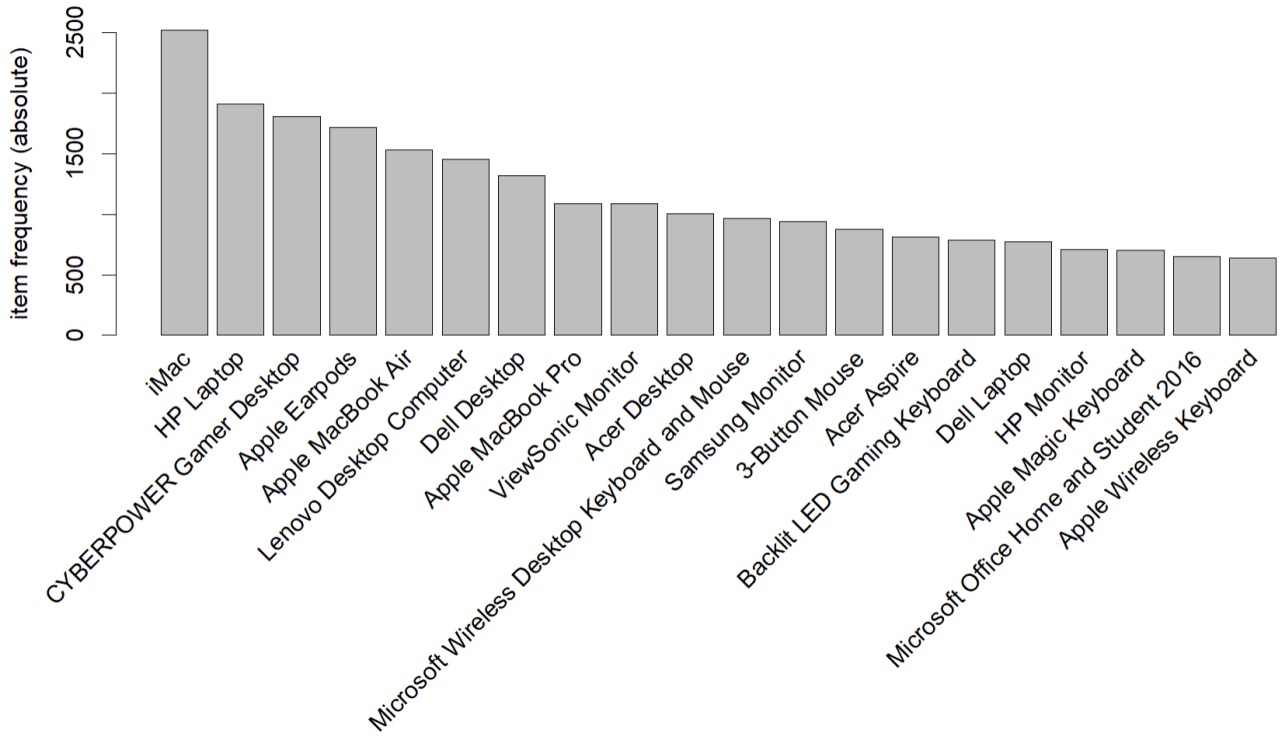
This is a small snipped of the data:

1	Acer Aspir	Brother Pr	Belkin Mo	VGA Monitor Cable					
2	Dell Deskt	Lenovo De	Apple Wireless Keyboard						
3	iMac								
4	Acer Desk	Lenovo De	Intel Desk	XIBERIA Gaming Headset					
5	HP Laptop	iMac	Epson Bla	ASUS Desktop					
6	iMac	ASUS Mor	Lenovo De	Mackie CR	Gaming Mouse Professional				
7	CYBERPOWER Gamer Desktop								
8	HP Laptop	Bose Com	CYBERPOV	Apple Mac	Large Mouse Pad				
9	Logitech Keyboard								
10	iMac	Generic Black 3-Button							
11	Dell Deskt	HP Laptop	AOC Moni	Apple Mac	Dell Laptop				
12	Acer Aspir	Dell Deskt	iMac	ASUS Mor	ASUS 2 M	Lenovo De	Rii LED Ga	Apple Mac	JBL Splashproof Portable Bluetooth Speaker
13	ASUS Chromebook								
14	Acer Moni	CYBERPOV	Apple Earpods						
15	HP Wirele	Dell Desktop							
16	ASUS Mor	Dell Deskt	HP Monitc	Backlit LED Gaming Keyboard					
17	HP Monitor								
18	Logitech Wireless Mouse								
19	Dell Laptop								
20	HP Desktop								
21	HP Laptop								
22	Alienware	3-Button Mouse							
23	iMac								

Finding the most Popular Items

Here we have a chart that shows the top 20 most sought-after items.

We can see that the iMac, HP Laptop and CYBERPOWER Gamer Desktop are the most bought. However, it is interesting to note that the iMac does have a considerable lead meanwhile the rest trail behind each other.



Can we find any Associations?

To do this, we will have to run an Apriori algorithm. Which will analyze the data and tell us if there exists any associations rules between certain items.

For example, it could say something along the lines of: products A, and B are often bought together.

It is important to note that this will not be doing any sort of machine learning tasks or modeling, it will purely find associations without modifying the data.

You might ask, “What is an association?”

In layman’s terms, an association simply means if there we can group certain things or items together. In this context, we want to see if certain products are frequently bought together and the Apriori algorithm will help us in this task.

How to use Apriori Algorithm?

To use this, we take our dataset, and plug in two parameters, support and confidence.

Support is the indication of how often a given item appears in the data and confidence tells us how reliable the association is.

One of the difficult part is finding the right values for parameters as if we don't have the right ones that we won't a good association rules/results.

Both will be a decimal numbers that be a max of 1.00.

How to know if Parameters are well?

When playing around with different values of support and confidence. We see how many rules are created, if its too less or too many then it wont much use for us to analyze. So there needs to be right quantity.

However, another important way is to know if it has a high lift value (Lift is a ratio of the support and confidence).

We consider these to determine if a given support and confidence value is good for a specific dataset.

What are the Optimal Values?

From our experiments, we determined that the optimal values are

- Support: 0.005
- Confidence: 0.6

This yields:

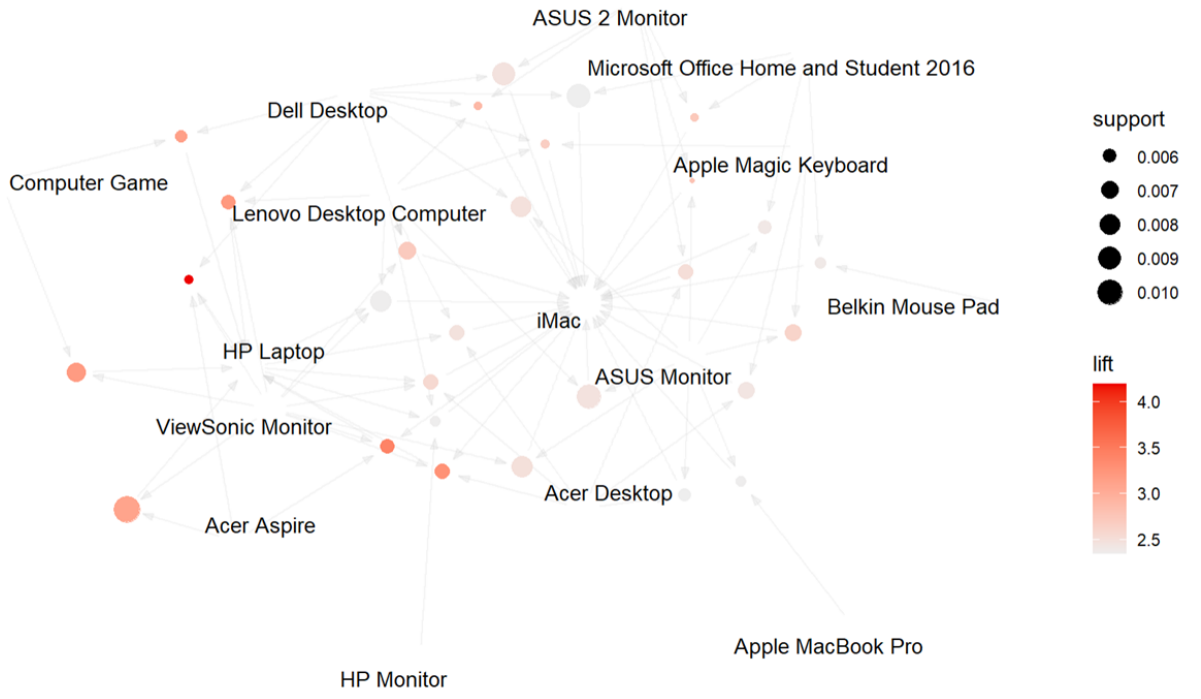
- Number of rules: 54
- Lift(Average): 2.725

What were some of the top Associations?

This figure on the right shows a graphic for the associations found.

In the first place, we had customers that an HP laptop tended to also buy an Acer Aspire, a Dell Desktop and a ViewSonic Monitor.

The next spots all featured an iMac being purchased along side various other computers and accessories.

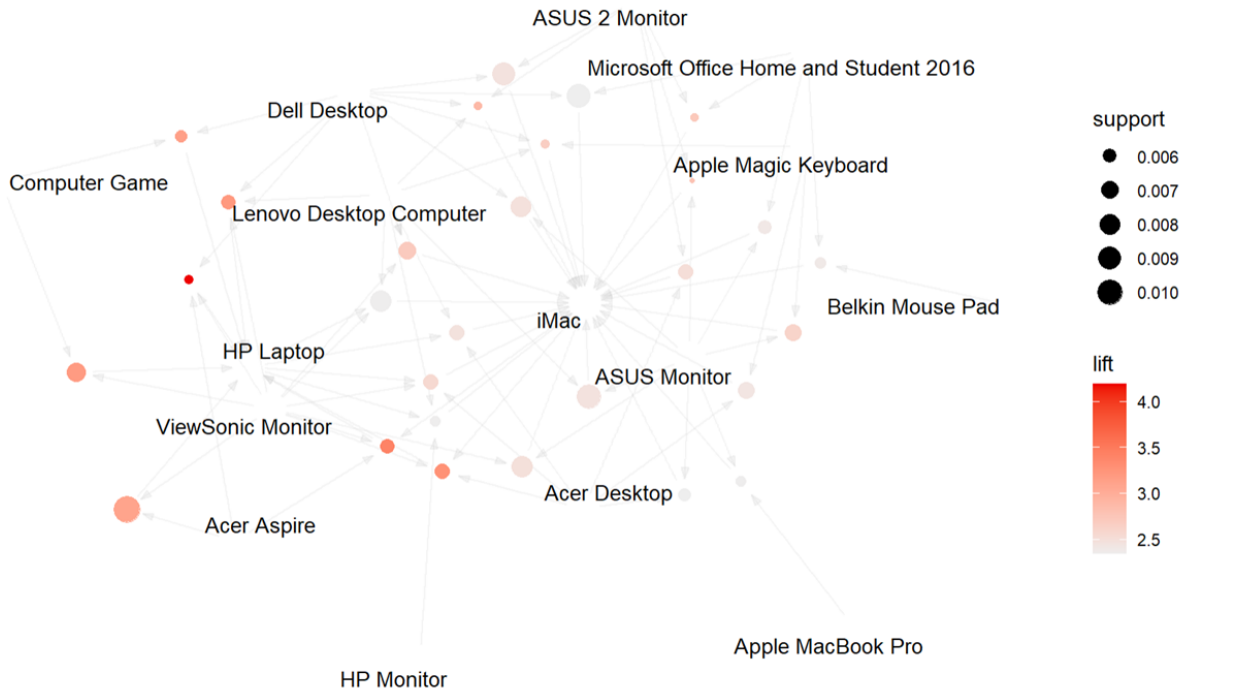


What were some of the top Associations? Cont.

This is very interesting to see as the iMac was the most sold purchased item by a decent margin but when it came to association it was second place.

The reason for this is that we believe is that those that purchased an HP Laptop happened to also buy specific products more often. Whereas with the people that bought an iMac don't have a set of specific products they bought as well.

This can be shown by the figure as you see less arrows connecting to the HP Laptop, but they connect to a higher lift. Compared to the iMac that has more arrows with less lift.



Is this any use for Us?

Yes! We believe if we were to acquire Electronidex, it will likely be a lucrative investment.

One specific way we can increase profits would be by lower the price of HP Laptops and iMac. This will attract even more customers to buy these as they'll be cheaper. Since these had the most associations, people are more likely to also buy other products as well. Therefore, in resulting in us selling more items and making more money.

Conclusion: Should we buy Electronidex?

Yes, We believe that we can acquire Electronidex as there are some associations that we can use to increase sales!

The iMac and HP Laptops can act as a gateway for us to entice customers to buy multiple of the other items and we can sell tons of items that way!