Project module 8

My data is about the pages that users visit on the Microsoft website. The association rules show that if a user has accessed a certain Microsoft page(s) then which other pages that user is likely to access. The original data that I downloaded from the site was not formatted in a way that I could run the apriori algorithm on it so I processed it using java code. The rules shown below (there is also an attached excel document with all the rules), show numbers that correspond to certain web links and a list of that is available in the excel document that I have attached to this post.

When I looked at the page names in the excel sheet, for some of them it wasn’t particularly clear what those pages were about, so I have done the coding part of it and interpreting is where the subject experts come in so for the most part, I can’t particularly comment on the relevance of the rules. However there are some rules that made sense where looking at development related information led a person to try to get a network membership or looking for support led a person to be interested in other more generic user-based pages.

Also, the itemFrequencyPlot function helped me decide a value for the support (I chose 0.01 to include about 100 of the 300 pages in the data). Looking at the range of support values (0.01-0.16) in the image below also meant that my support value wasn’t too high. Sorting the rules by lift assures there are better rules on the top. Some of the lift values obtained are exceptionally high and I wonder if those are the rules that would be totally obvious to a subject expert.



