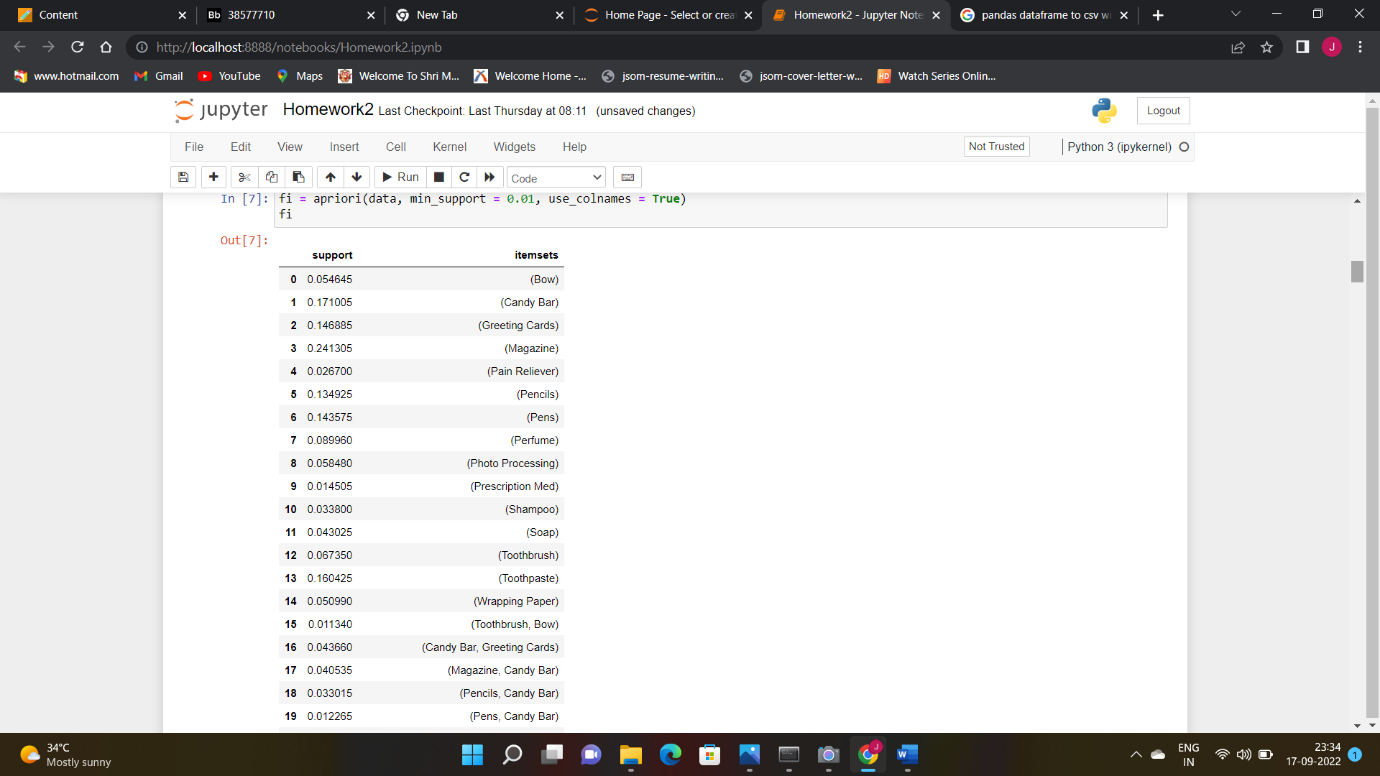
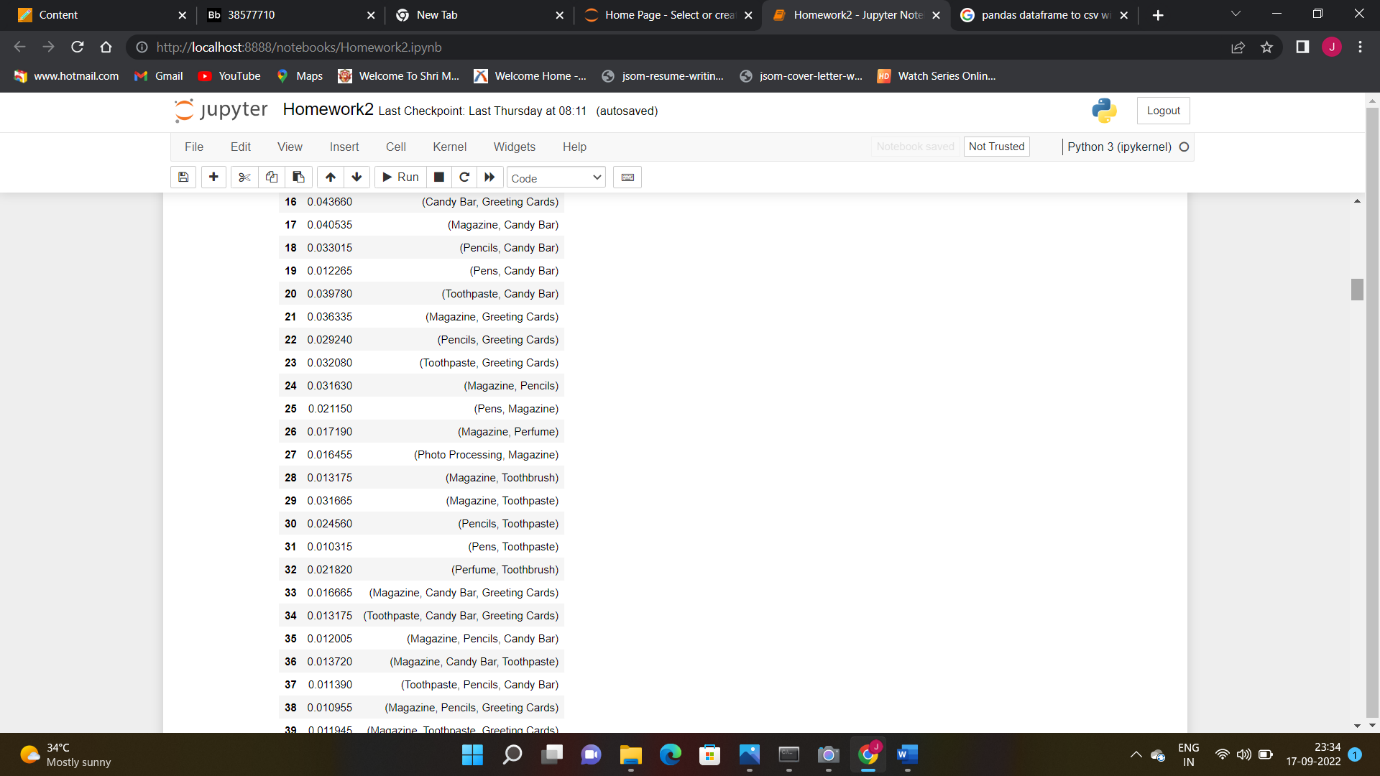
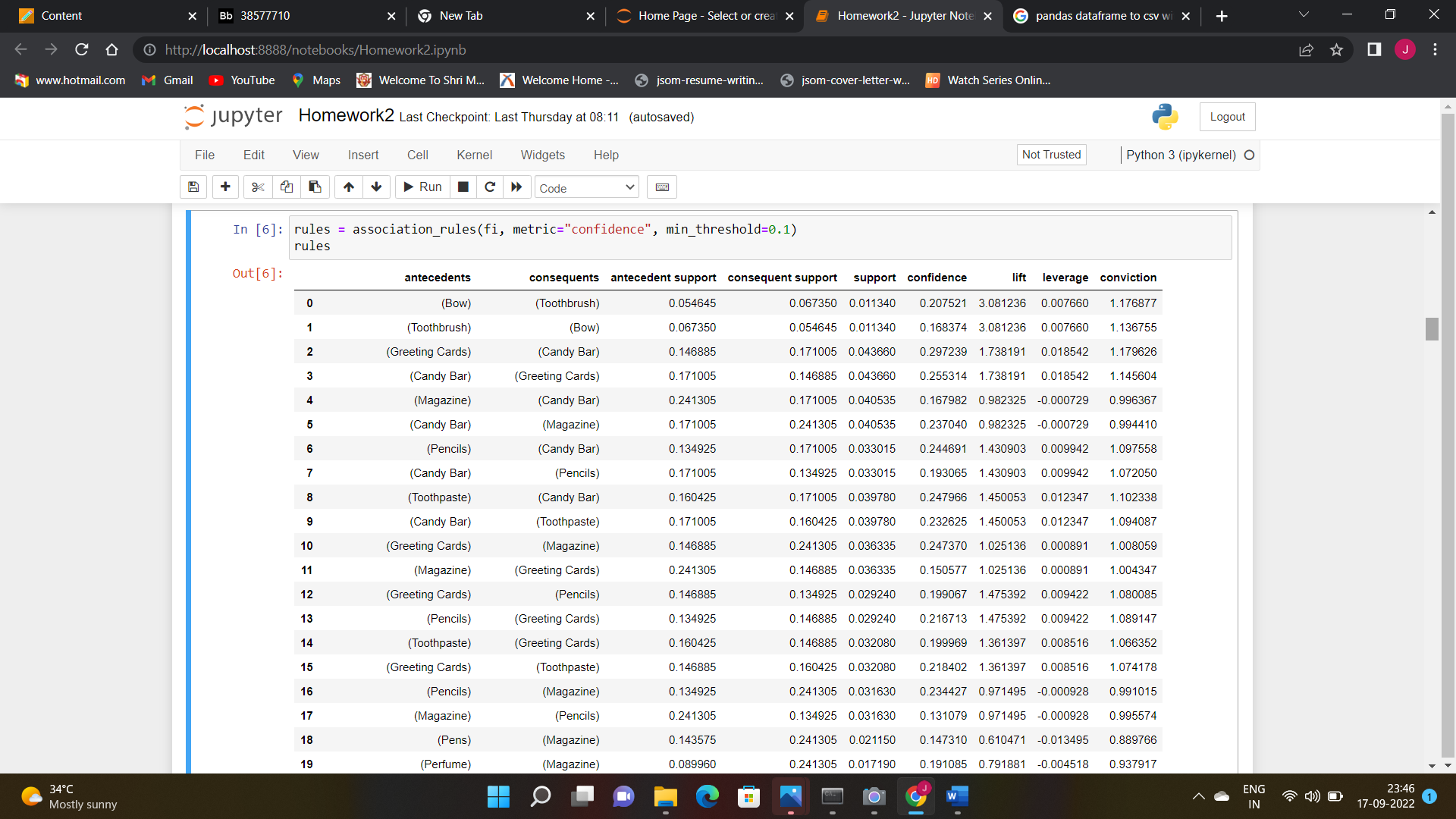
2. Identified the frequent item sets with minimum support threshold of 5%.

There are 39 frequent itemsets





3) The rules were generated with 10% minimum confidence. There were 49 rules

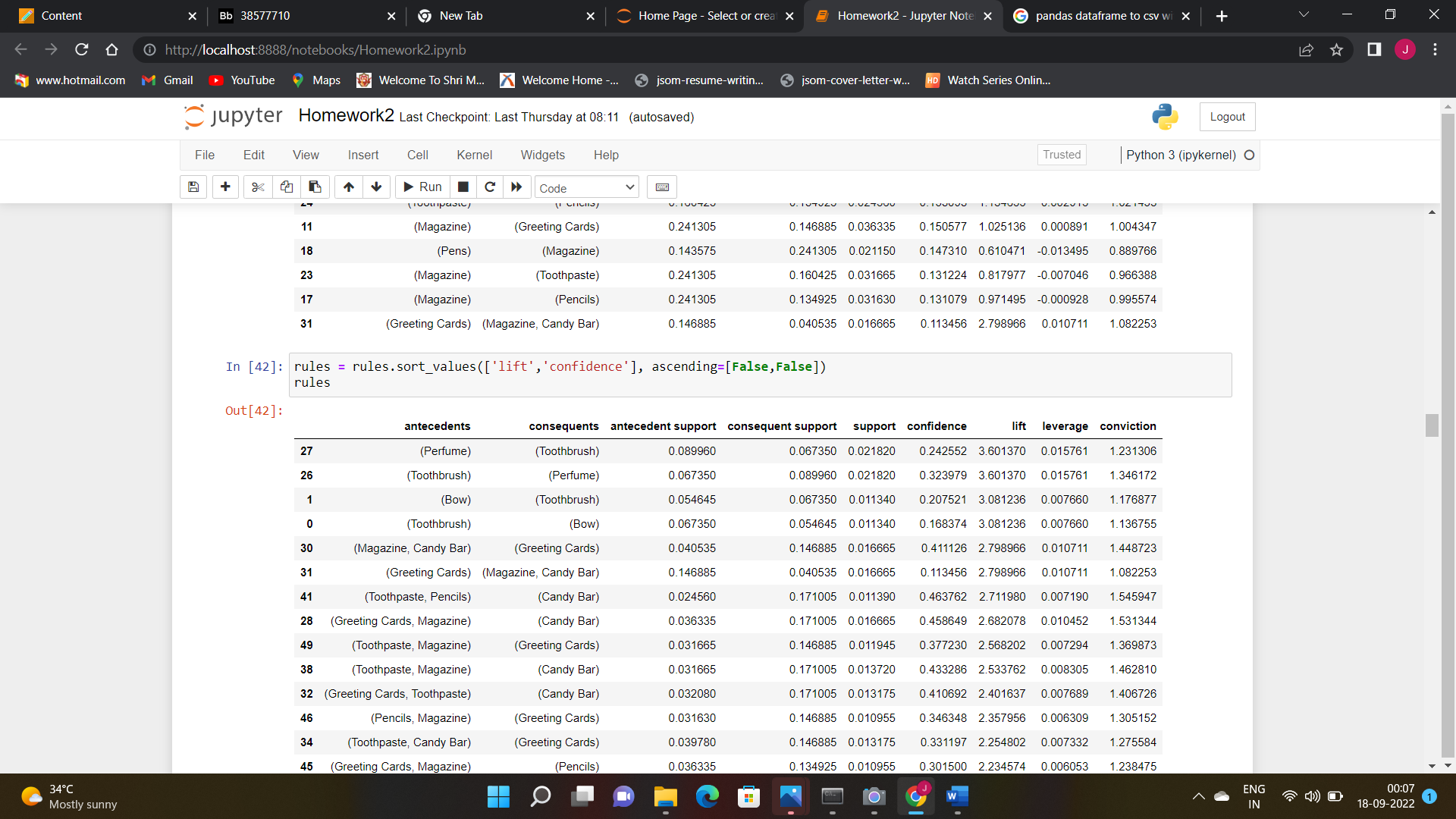


4) The rules with highest lift were {Perfume} -> {Toothbrush}. So from the previous result it can be observed that Support for Perfume is 0.0899 and for toothbrush is 0.0673 and support for Perfume and Toothbrush is 0.0218. So Lift is basically how much more left hand item in our case Perfume is found with right that is toothbrush than without toothbrush. One way of calculating lift is Sup(Perfume and Toothbrush)/ Sup(Perfume) X Sup(Toothbrush)

= 0.0218/ 0.0899\*0.0673

= 3.601

Which we got from the result



5)

Leverage :- So Leverage is basically used to measure the correlation between item sets considering support under independence assumption . It can also be interpreted as measure of deviation from independence

So leverage is calculated for {Perfume} -> {Toothbrush} as

Support(Perfume and Toothbrush) - Support(Perfume) X Support(Toothbrush)

= 0.0218 – (0.0899\*0.0673)

= 0.0157

So for leverage if it is >1 then it is positively correlated

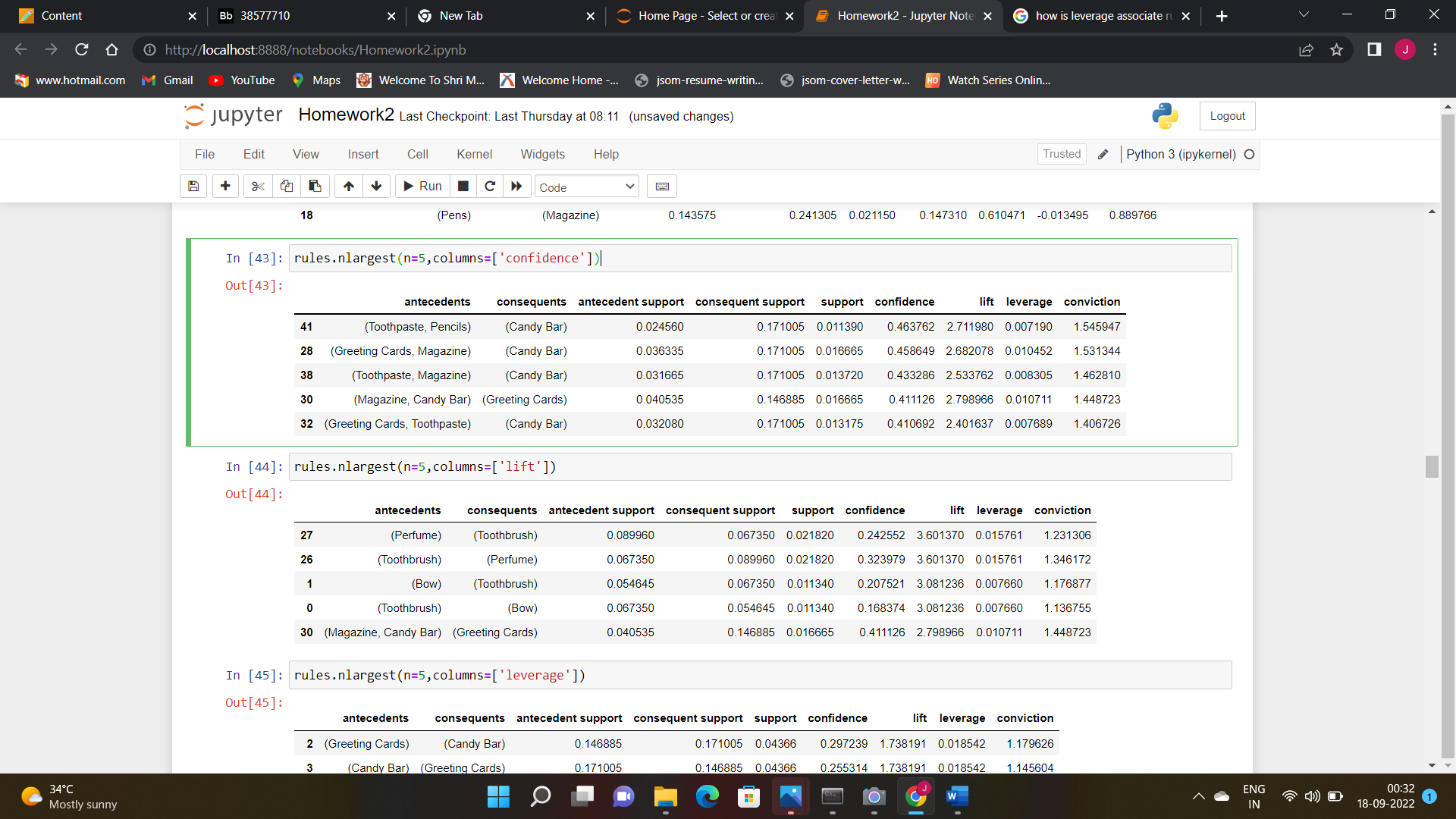
Leverage<1 then it is negatively correlated

And If leverage = 0 both antecedent and consequent are independent.

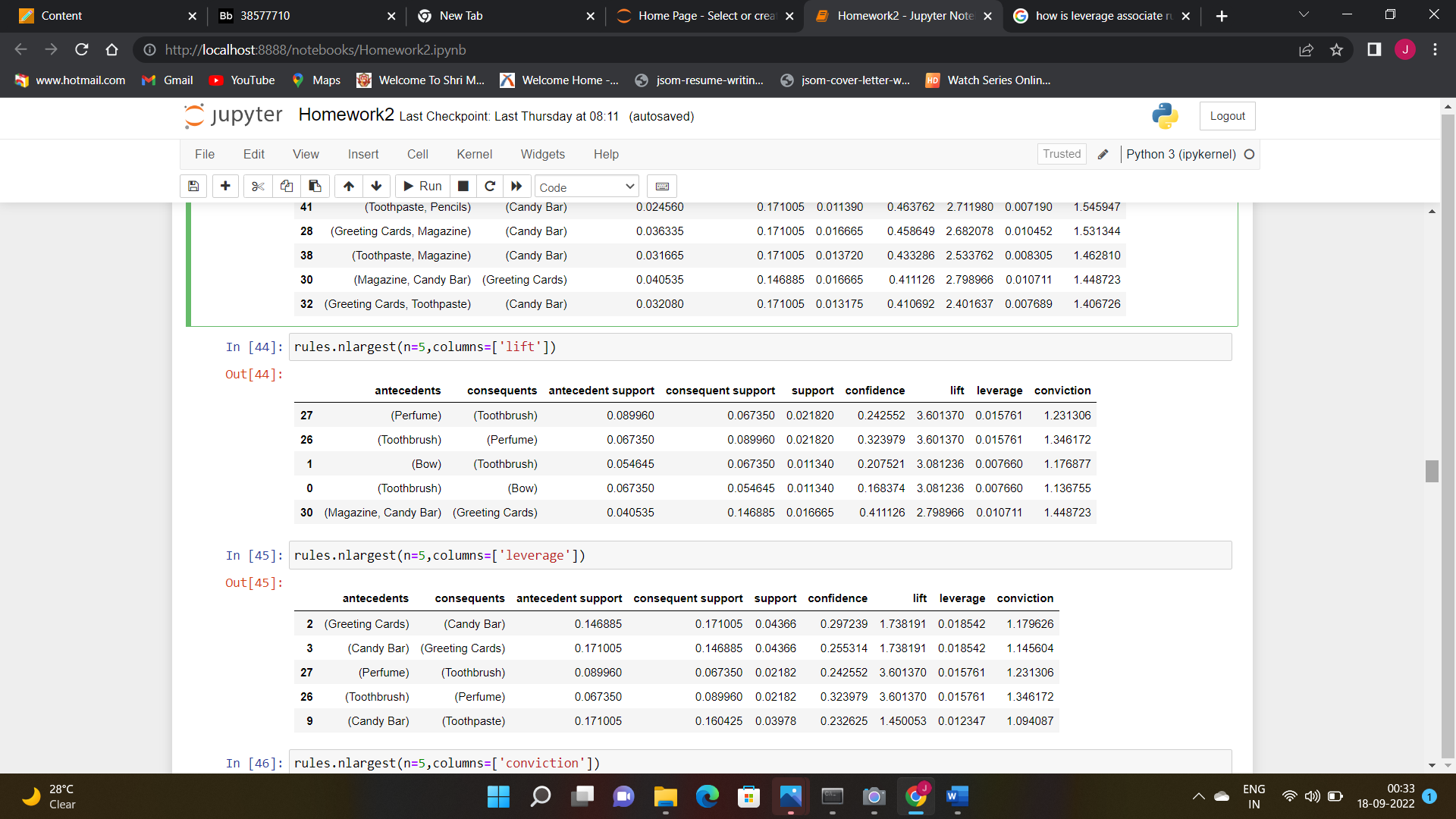
Conviction:-

6)

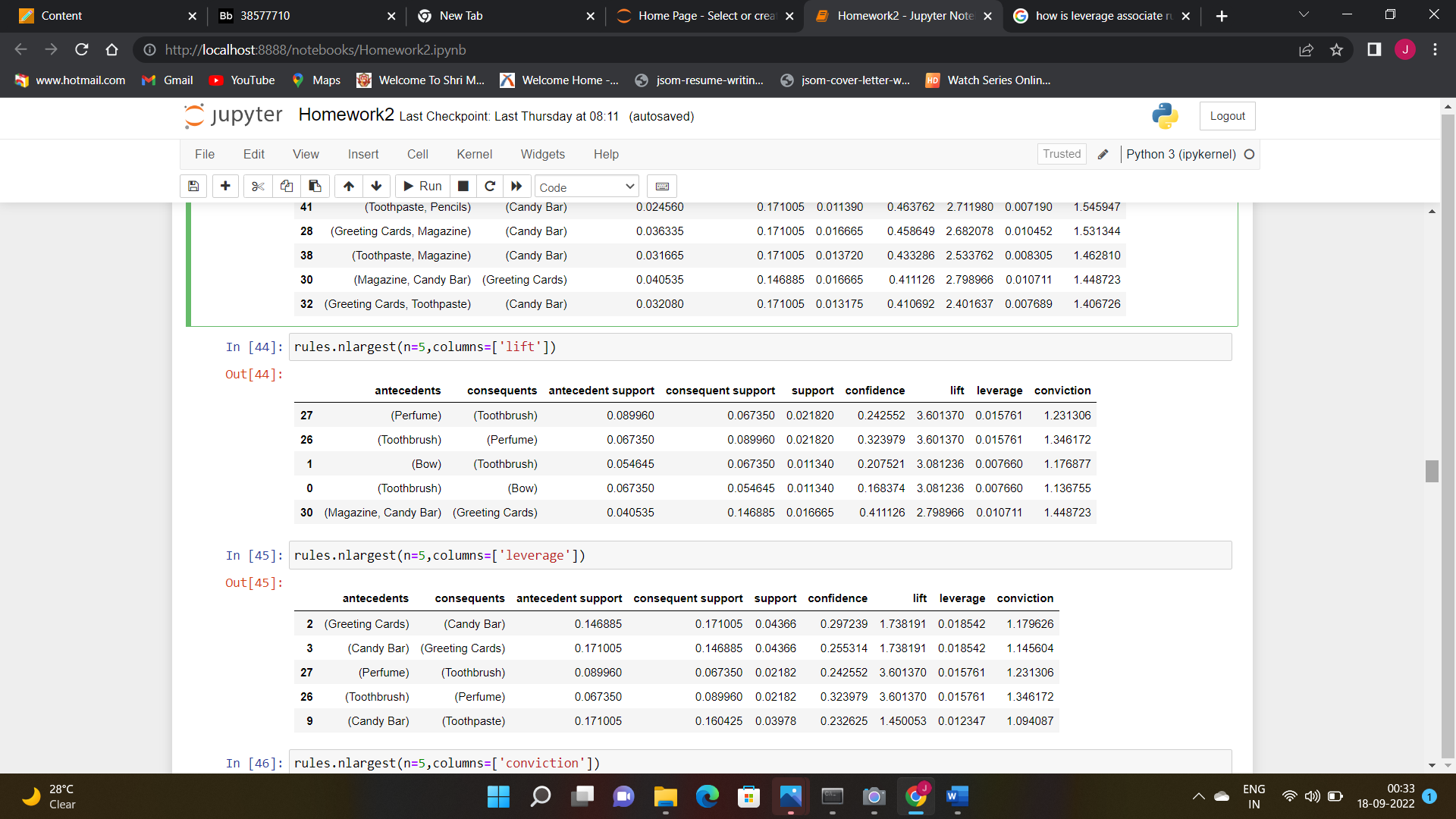
a) 5 rules with highest confidence



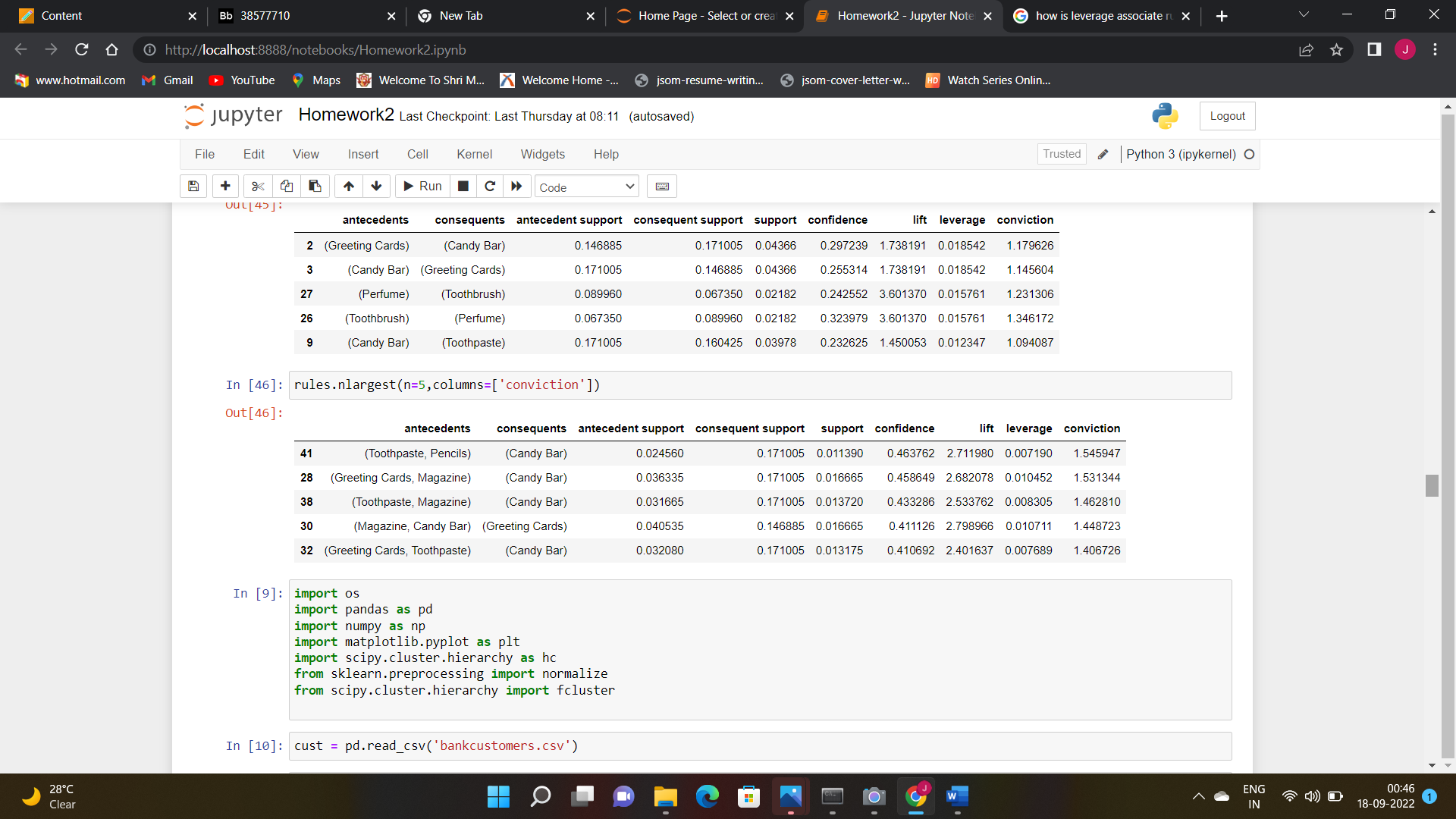
b) 5 rules with highest lift



c) 5 rules with highest leverage

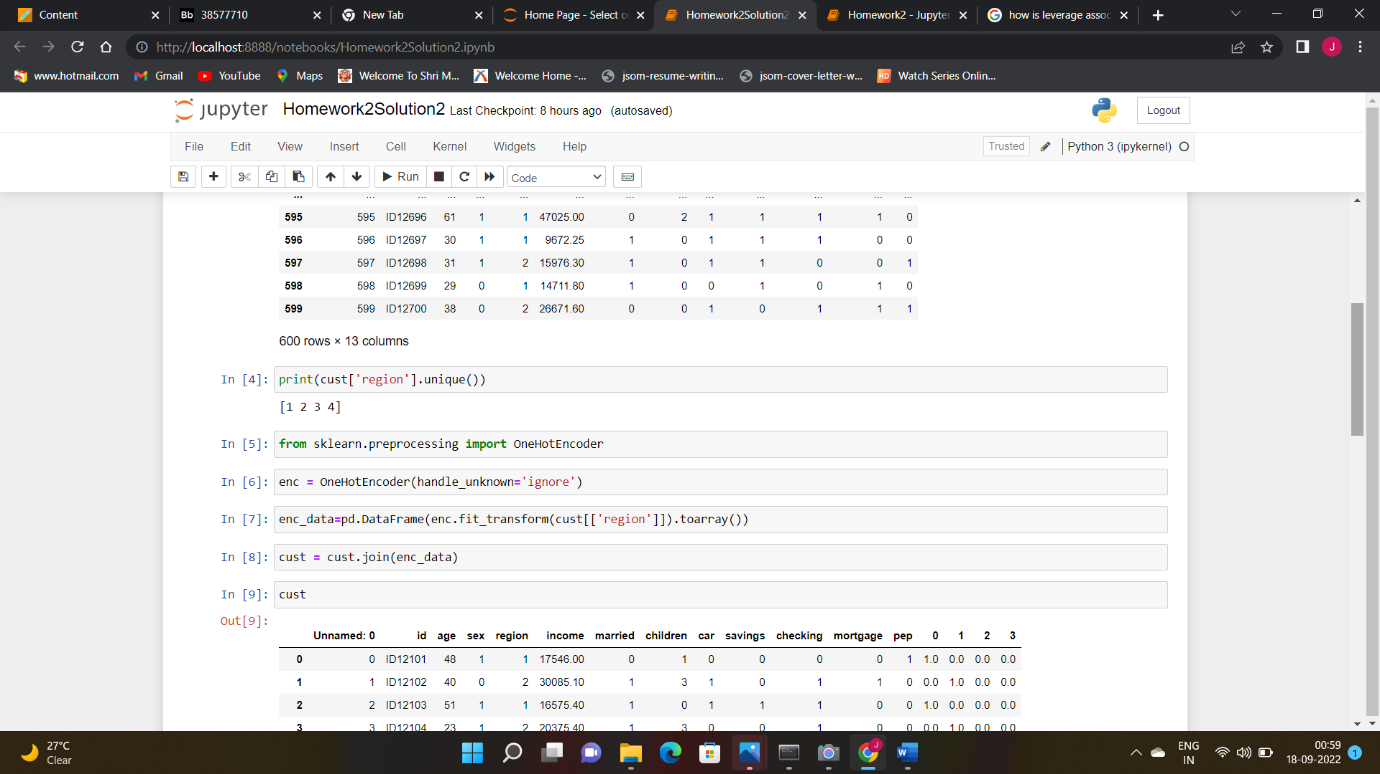


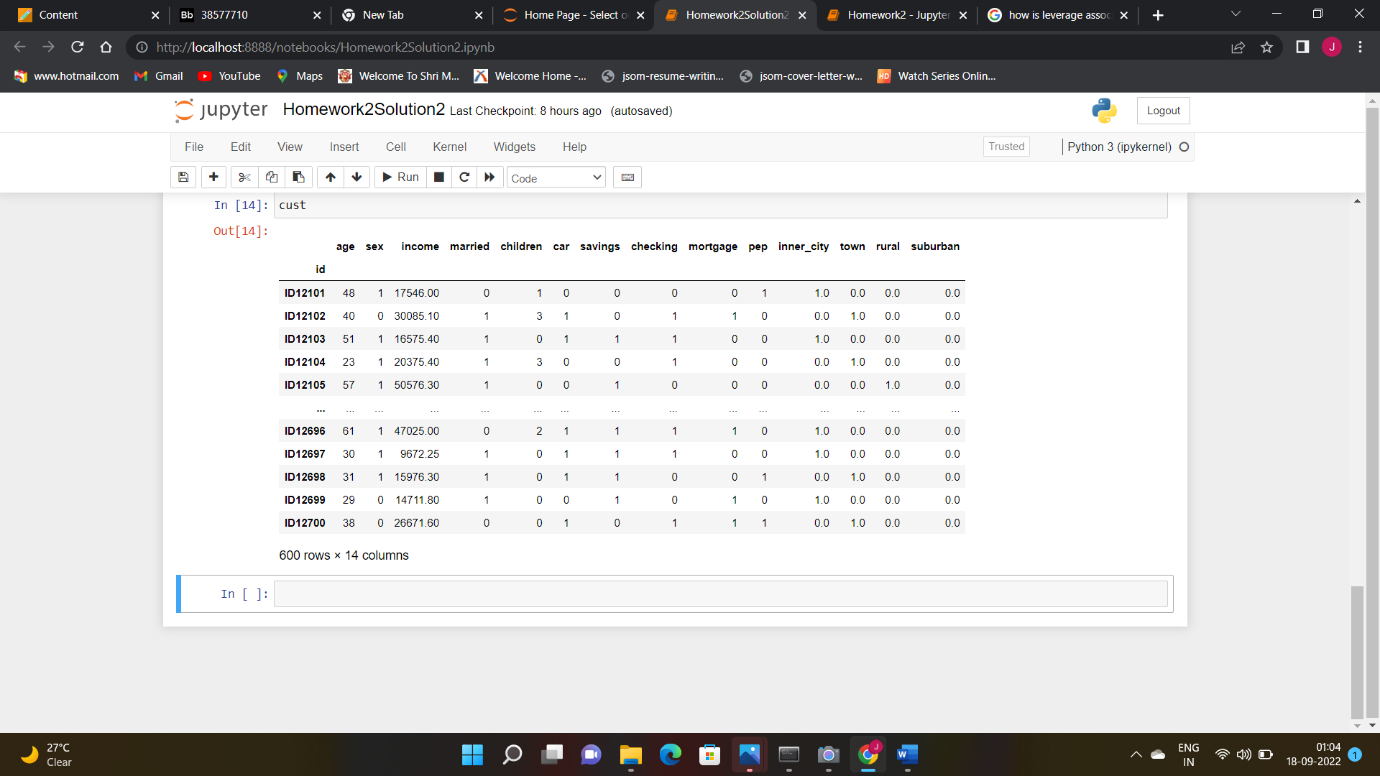
d) 5 rules with highest conviction



Part II

1. Perfomed OneHotEncoder and converted categorical region into inner\_city , town , rural, suburban.





1. After Applying Hierarchical Clustering and using Centroid Linkage method it can be inferred that 3 clear clusters can be seen.

