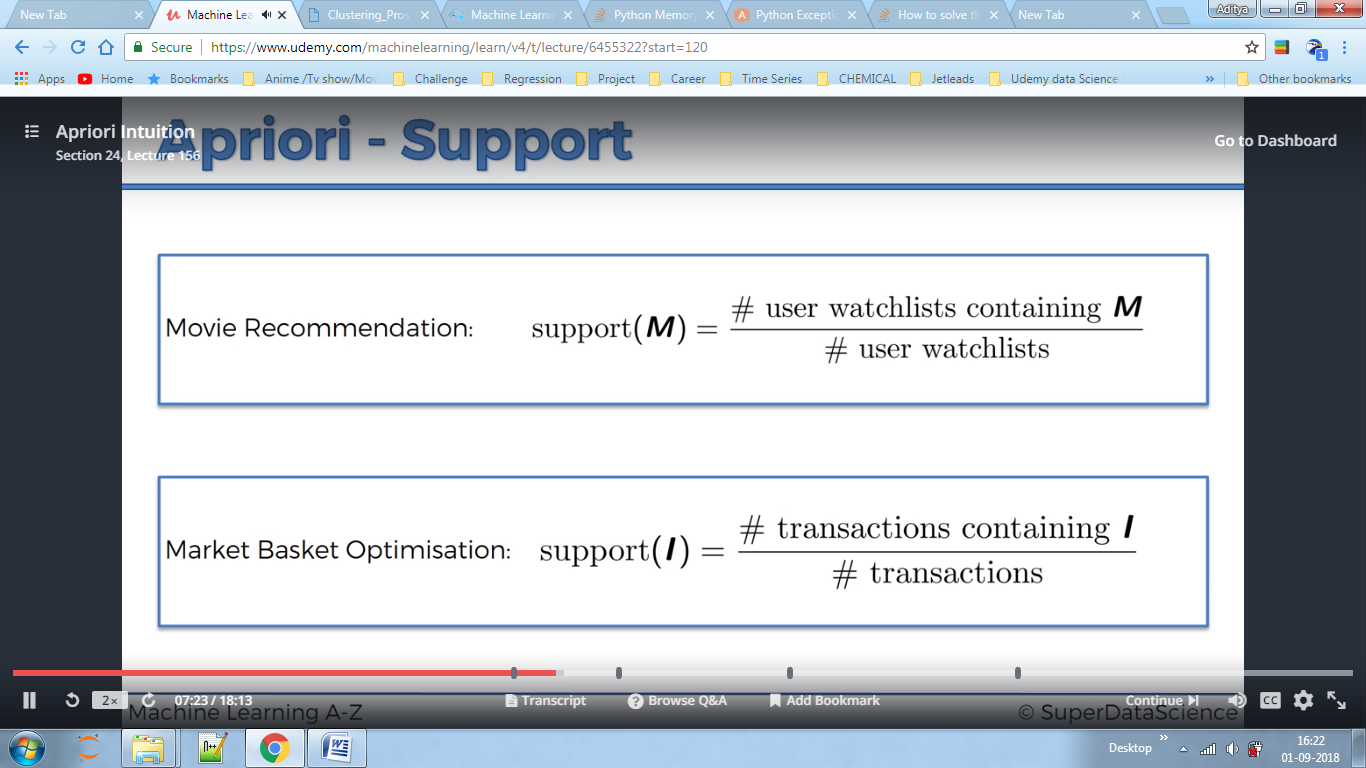
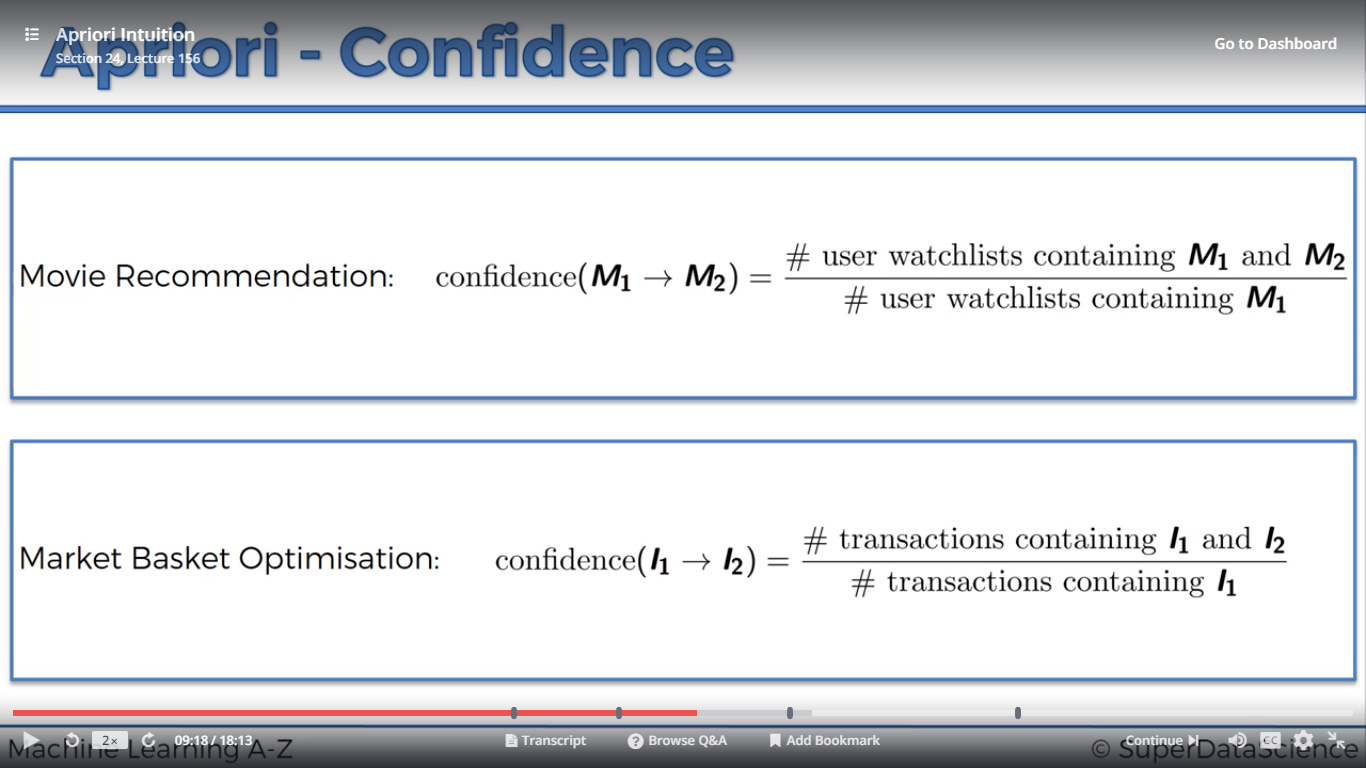


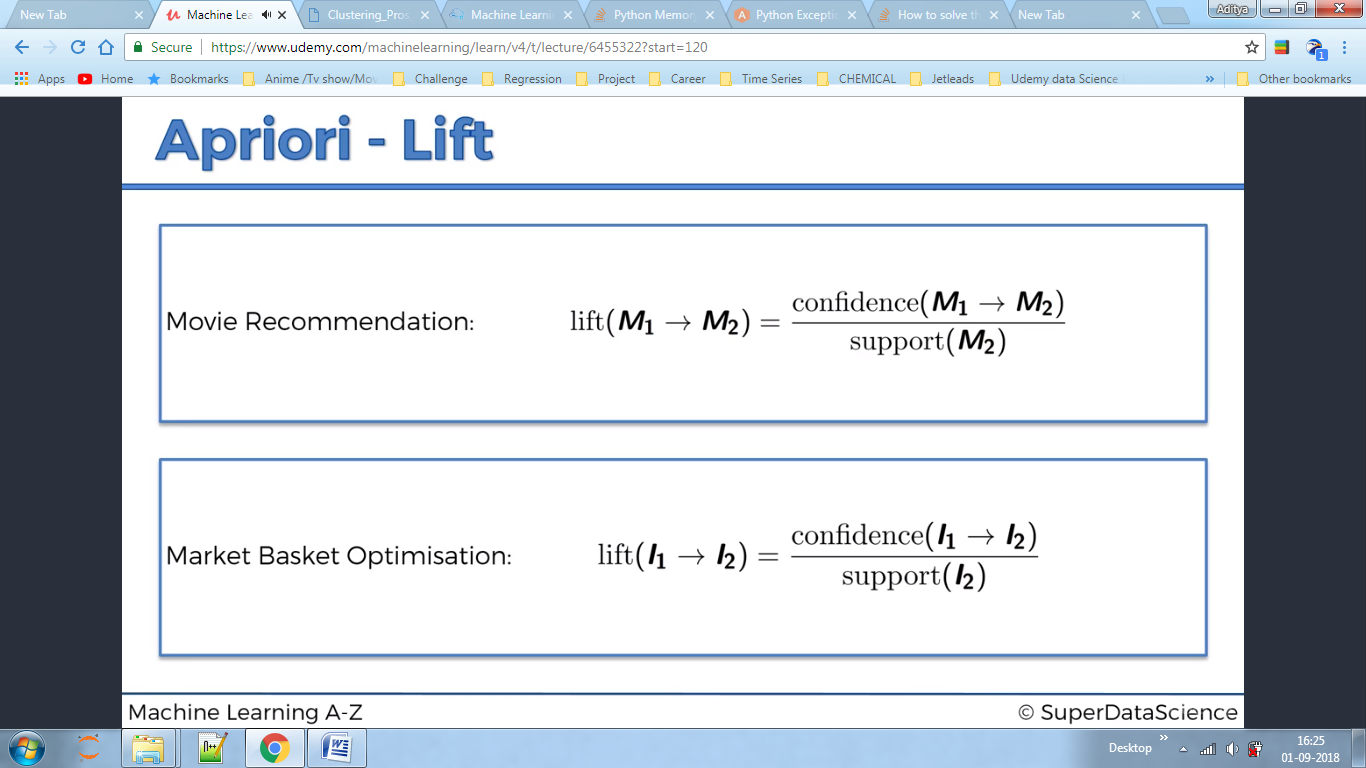
WHY APRIORI?

For Example, In a convenience store we found out bread and milk are very far away ,because if one is taking milk they will also take bread and while walking through the store they will see more and more items so they will buy them.

Apriori has 3 parts support ,confidence and lift!





Lift signifies improvement in your prediction.

Our problem statement is that we have a dataset and we want to predict which products are related to each other, say if person buys A product then he will also buy B.

Apriori accepts list of list hence we need to convert our matrix into lists of lists. The items should be strings.

BELOW IS CODE

from apyori import apriori

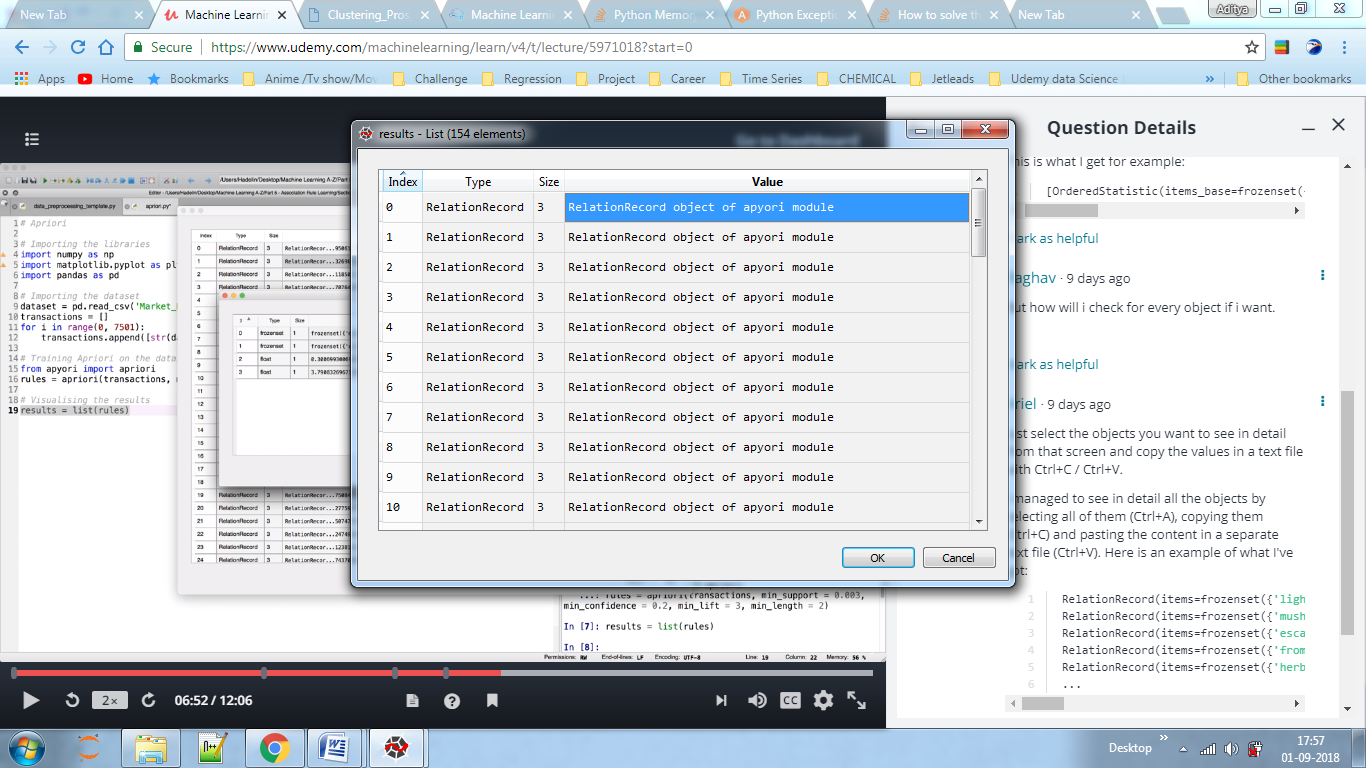
rules = apriori(transactions, min\_support = 0.003, min\_confidence = 0.2, min\_lift = 3, min\_length = 2)

Here min\_support is 3\*7/7501 i.e. the product should be bought atleast 3 times a day for a week,depending on our business value.7501 is the total number of transactions in a week.

First we set min confidence and then we see which rules we got for min\_confidence =0.8(can be any value),then moved down to 0.2 as among them one rule was that customers who eat eggs also buy mineral water but we know this is not true because summers are hot in france and these people eat lots of eggs .hence we can discard this .Here 0.2 means 20 times these rules will be applied on the dataset.

Here lift implies how many rules we want to conclude our statement.

Here 0.5% of transaction will contain these two items in the list.



On clicking on the results and values we will conclude that people who bought 'light cream ‘also bought , 'chicken'

<RelationRecord(items=frozenset({'light cream', 'chicken'}), support=0.004532728969470737, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'light cream'}), items\_add=frozenset({'chicken'}), confidence=0.29059829059829057, lift=4.84395061728395)]) >

Here if they buy light cream they have 29% confidence i.e they will buy chicken 39 out of 100 times,here the lift is also good which makes it relevant rule. If lift is poor then it signifies maybe customers buy both products; mostly not if they bought one they will buy other.

In python this is already sorted, the top rule is at 0 indexes.