

Data Mining Assignment Report

Ans 3:

Q1: Is there any correlation present between the day and the transactions encountered?

Explanation:

- The food spread in the mess is usually structured and follows a particular routine. Specific class of items, if not particular items, is made on specific days of the week. For ex. On Tuesday, it's always Chole + either Puri or Bhature.
- It is often observed that some food classes of the mess are universally disliked. When these are prepared, many of the students tend to eat at ANC instead of in the mess.
- Since, the food prepared in the mess and the day of the week are very often correlated, the model prepared for the given question would be able to show which days (and consequently) which food items/class is disliked among students which in turn can help in better increment of ANC prices and also a change in the mess food accordingly.

Solution:

- The dataset given to us only has date for each transaction but no day information. This day information can be easily calculated (using year information) and put into the same file using Python.
-

-
- After obtaining the day information, each transaction would have to be copied its quantity number of times (as was done in Q1).
 - The data can then be fed into IBM-SPSS. The day and itemID attributes would be converted to nominal and then would be set to flag using SetToFlag node.
 - As the final step, rules of the form Day → ItemID can be found.

Q2: Can offers of the form “Buy x number of item_1 get y number of item_1 free” provided? If so, what are the items on which these can be provided?

Explanation:

- Often it is seen that some items are bought in bulk as compared to some items which are always bought once per transactions. For ex. If a person is ordering Naan, it is likely that he won't be ordering just a single one. Similarly, if a person is buying chips, it is likely that he would prefer to get more than a single packet since, chips are usually bought to keep in stock. However, the same argument isn't valid for pizzas and burgers.
- Our model would be able to extract these items and the quantities in which these should be offered for free.
- This might greatly increase the revenue since, such offers often tend to tempt people to buy more of an item even when they don't need it.

Solution:

- Firstly we'd have to club different transactions that happen around the same time and involve the same student ID and item. The quantity column would reflect the change caused because of this aggregation.

-
- Next, we can pass the whole data into SPSS-modeler using the var node. This node can be passed to the filter node where the quantity, price and the itemID can be filtered out.
 - The output of this can be connected to the plot node found in the graph tab. Here, we get the option of plotting the dataset w.r.t. any two of the attributes i.e. the whole data is projected to the plane with the specified axes.
 - The two axes can be selected to be price and quantity. The 3rd dimension, in our case, would represent the itemIDs. Here, we can see the clusters forming and getting projected onto these two axes. From here, we can select the magic wand tool and choose the items that are on the top right side of the graph board. This would correspond to high quantity and high price. However, it is very important to choose only the dense clusters (i.e. clusters with many points) and should be given more priority than the price and quantity values.
 - Using this technique a select node can be generated which would contain only the data points that we selected.
 - Items can be filtered out from the data generated. The price and quantity values can be used to decide the number of free items to be offered through thresholding. The final profit change can be calculated. The threshold of the price and quantity values can be manipulated to get better profit.