Problem 3

Q1. The record of cash payments when quantified, showed us the results that there were high number of such transitions in the month of August, which was because the ID cards take a couple of weeks to be distributed. The month of September sees an increase of payments because of the presence of fests like Junoon and BOSM. There is a significant decline of such transactions during October. Still the number remains high, due to transactions during the fest: Oasis. November sees a plummeting decline, where the cash payments fall to more than half the October sales. This can be credited to the busy schedule of the students owing to more evaluatives and end semester preparations and no fests. December further sees transactions falling to half, due to the period of comprehensive examinations going on. The ANC committee may decide to set up automatic cash counters, where you can select your order and pay and get bill and change. For this the ANC committee may want to know at what time of the year to set up the cash counters.

Solution: For this problem, clustering(DBSCAN- since the clusters can be non globular) can be done on each month's data. The proximity measure will have the student segment, selling date in it. So the clusters having segment as F0, we can find the ranges of dates on which cash transactions are high. So on those dates, a cash counter will be set up.

The **performance metrics** which can be used to evaluate the clusters is:
Similarity oriented measures : The class of a data point is the student segment. So we can create the ideal cluster similarity matrix and ideal class similarity matrix and find the correlation between them. The more correlated the two matrices are, the better the clustering.

Q2. The frequency of purchase of certain items vary seasonally. With the transition of weather conditions from summers to winters, sale of items like Tea has seen an increase, and hence a legitimate increase in its price also. The price of the item samosa has witnessed a subsequent increase also.For some items, the sale has decreased from summer to winter. For analyzing, which items are sold more in which season and which items are not sold at all(so that their production can be stopped), data mining techniques can be used.

Solution: Here too clustering can be used to visualize and see which items are sold in which month. Items having the same ID are in the same cluster. We can set a threshold, if number of items in a cluster are below that threshold, then we can say that item is sold less that month. For clusters containing number of data points greater than another threshold, we can say that that item is sold in large quantities that month, so the ANC committee may decide to increase the prices of the items sold in large amounts and discontinue the items sold less.

The **performance metrics** which can be used to evaluate the clusters is:
Similarity oriented measures : The class of a data point is the student segment. So we can create the ideal cluster similarity matrix and ideal class similarity matrix and find the correlation between them. The more correlated the two matrices are, the better the clustering.