Q.1 (i)

There are 11 columns in the dataset. The definition of each column are as follow –

**Month** –

**Town** – Town refers to the name of the town in Singapore where the flat is located.

**Flat\_type** – This column contains types of the resale flat which is 2 to 5 Room and Executive

**Block –** This column contains the block number of the flat.

**Street\_name –** This column contains the street name where the flat is located.

**Storey\_range –** this column contains the range of the storey where the flat is located.

**Floor\_area\_sqm –** this column describes the floor area by square meter.

**Flat\_model –** this column describes the model of the flat.

**lease\_commence\_date –** this column describes the started lease date of the flat.

**Remaining\_lease –** this column describes the remaining lease duration of the flat in years and months

**resale\_price –** the column describes the resale price of the flat

Q.1 (ii)

Similarity and differences between association rules mining and cluster analysis.

Both, cluster analysis and association rule mining, are in the field of unsupervised machine learning. They can find the similarity and relationship between sets of items and the variables.

Association rules mining is a method for discovering interesting relations between variables in large datasets. This can also be used to find out the frequency of the variables in the large dataset. We can also find the frequent pattern of the variables using this method.

Cluster analysis is to create different grouping which is called clusters based on their similarity to identify discrete groups of customers, sales transactions, or other types of behaviors and things.

Q.1(iv)

a. Tools/Software we used

we used python as the programming language and visual studio code as the code editor and Microsoft excel for the dataset.

We used **Apriori Algorithm** to find out the relationship between flat type and Town.

b. Comment on the results of our findings

* We used python to perform association rule mining to find out the association between town and flat type
* We increased each threshold’s value bit by bit to narrow down the result of the each rule to detect the most related association between town and flat\_type
* Our findings are as follow based on the parameters of  min\_support of 0.01, min\_confidence of 0.4 and min\_lift=1.8
* The most popular association will be the ANG MO KIO -> 3 ROOM according to their confidence of 0.5359520639147803 and likelihood of being 3 rooms flat in ANG MO KIO is 2.28 times more than other flat types.
* Follow by the CLEMENTI -> 3 ROOM of 0.47068747363981445 confidence and 1.99979100824268 lift value and KALLANG/WHAMPOA of 0.44598713173044363 confidence and 1.8948476553446194 lift value respectively
* ANG MO KIO area have 3 ROOM flat type more than other flat type and CLEMENTI area also have the 3 ROOM flat types more than other flat types.