

Assignment 2

Question 1

- a. There are 10 dimensions so cuboids in the full data cube: $2^{10}=1024$
- b. Distinct aggregated cells in the complete cube: $3 * (2^{10} - 1) - 2 * 2^7 = 2813$
 $3 * (2^{10} - 1)$ is the number of non-base aggregated cells, $2 * 2^7$ is the number of repeat cells.
- c. Only those cells with the first three dimensions aggregated (i.e. *), e.g., the cell (*, *, *, c4, c5, c9, c10), have count 3. And the number of such cells is $2^7=128$.
- d. There is only one closed cell with count 3, So the number of non-star dimensions is $10 - 3 = 7$.

Question 2

- a. Suppose Q the i th dimension has L_i levels, then the number of cuboids is $\prod_i (L_i + 1)$. So the number is $(2 + 1) (1 + 1) * (1 + 1) * (1 + 1) = 24$.
- b. Use python function `set()` to get the distinct combination of values of city, category, rating and price.
- c. Use python function `set()` to get the distinct combination of values of state, category, rating and price.
- d. Use python function `set()` to get the distinct combination of values of category, rating and price.
- e. Use Pandas slicing to filter rows satisfy the conditions.
- e. Use Pandas slicing to filter rows satisfy the conditions.

Question 3

Import package `apriori` to implement Apriori algorithm.