Clustering

Features:

New:

1. `mean_family_price`
2. `family_size`
3. `freq`

sku	bought_by_customers_number	bought_in_different_hour_number	mean_price	mean_family_price	family_size	freq
0	1	1	2.272133	2.128901	22	1.000000
1	5	3	1.898885	2.128901	22	1.000000
2	1	1	2.461702	2.128901	22	0.997260
3	17	2	2.155855	2.128901	22	0.323288
4	3	2	2.155855	2.128901	22	0.323288

Features description:

1. `mean_family_price` - mean price over item family
2. `family_size` - family size of the item
3. `freq` - frequency customers by item (number of days bough / all days)

Cluster's information:

Number of clusters = 4

Sizes of clusters:

1. 187
2. 104
3. 253
4. 76

Clusters interpretation:

1. Cluster of items users buy the most time with the almost biggest frequency over the all clusters (top2 `freq`). This cluster may represent every-day things people always buy despite the situation (coffee and sandwiches).
2. Medium size cluster with the high number of customers bought with the cheapest items over all clusters and highest family size.

3. The largest cluster over another clusters. Contains items that not all people buy, but those who buy do it almost every day. Most probably these are items that people enjoy every day (like snacks and fast food).
4. The smallest cluster that contains items, that are not frequently bough, but the number of customers who bought this items is high. Most probably, these are things not for every day consuming.

Visualization