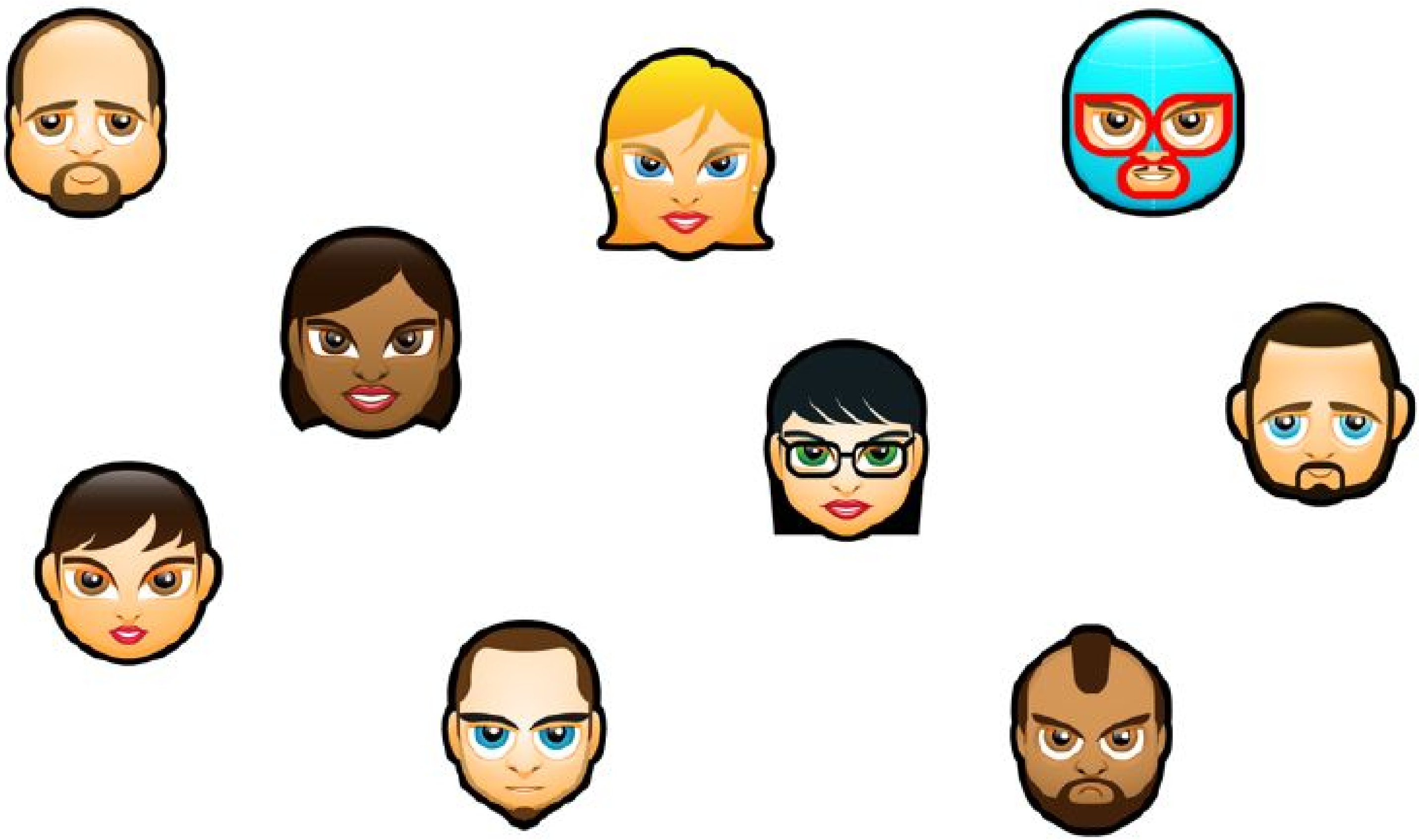
Customer Segmentation for Retail Marketing

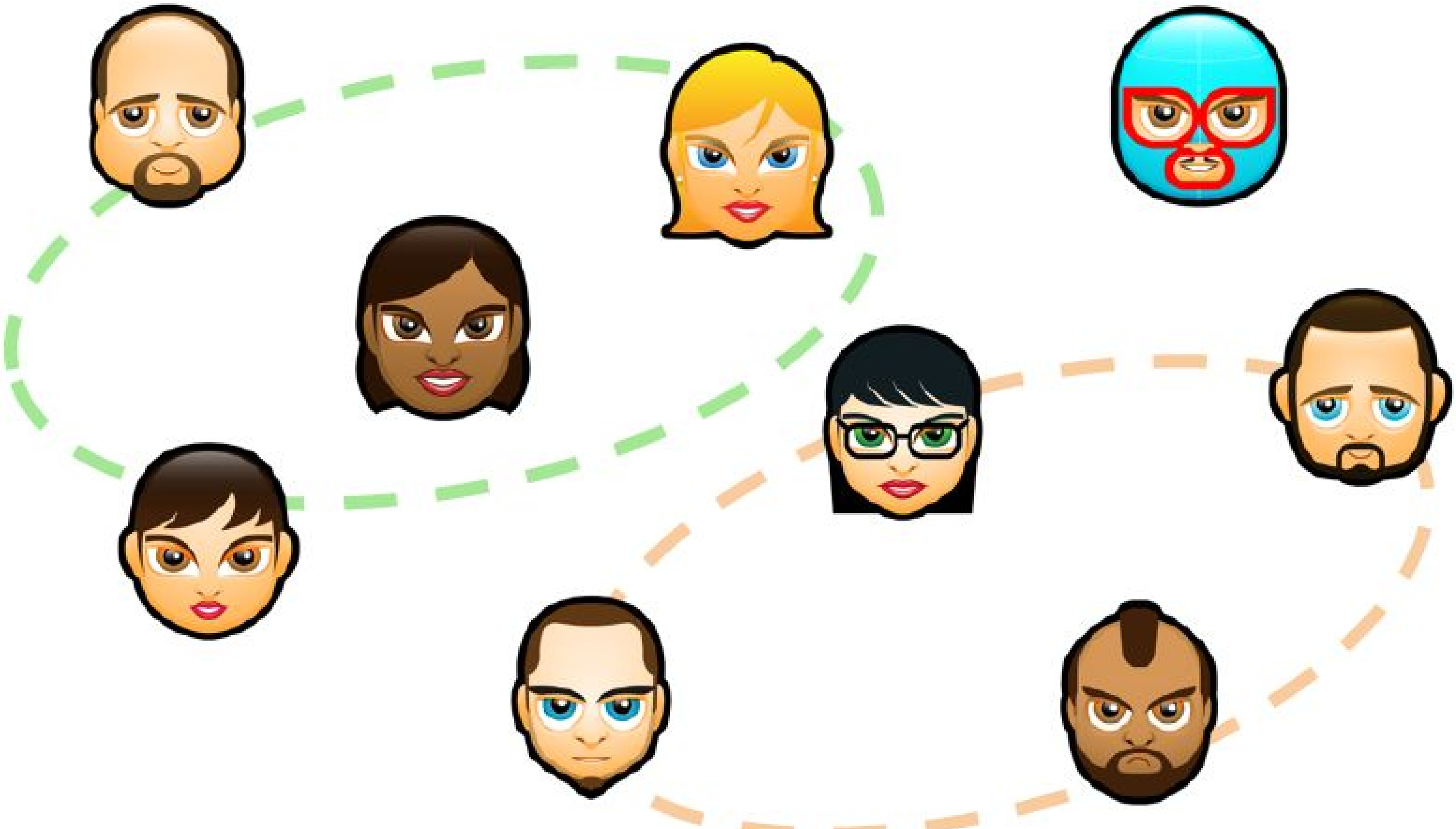
By:
Sachin Kumar (15BM6JP38)
Sanjeev Kumar (15BM6JP41)
Under the supervision of Dr. Sujoy Bhattacharya

Segmenation ??

“Who’s shopping at my stores and how can I market to them



Let’s try to group them by similar shopping behavior



Benefits: customer profiling enables targeted marketing and operations

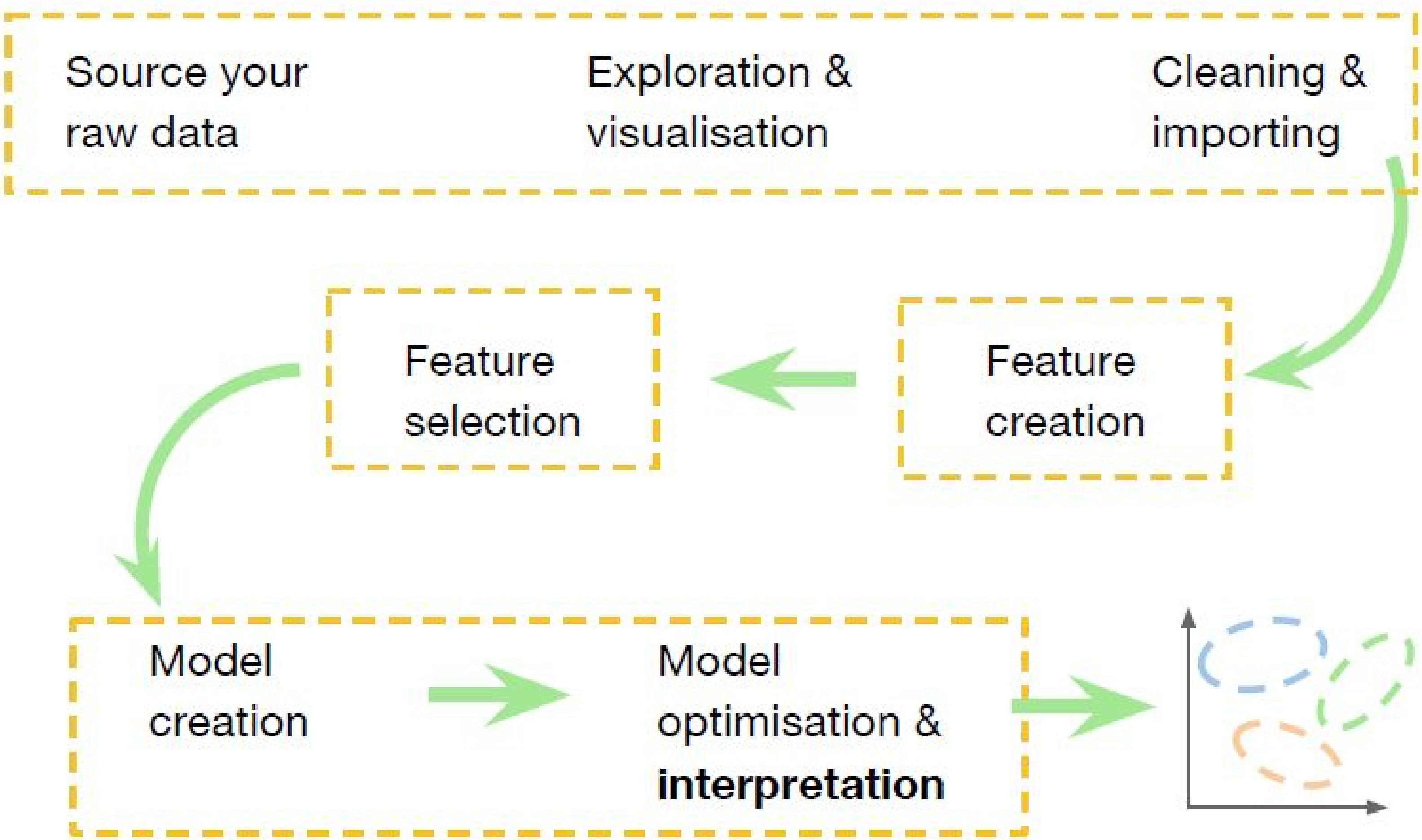
Bob
Lives "SW15"
Age "40"

Bob
Lives "SW15"
Age "40"

Type: Family First

Retention offers
Product promotions
Loyalty rewards
Optimise stock levels & store layout

Methodology



Data

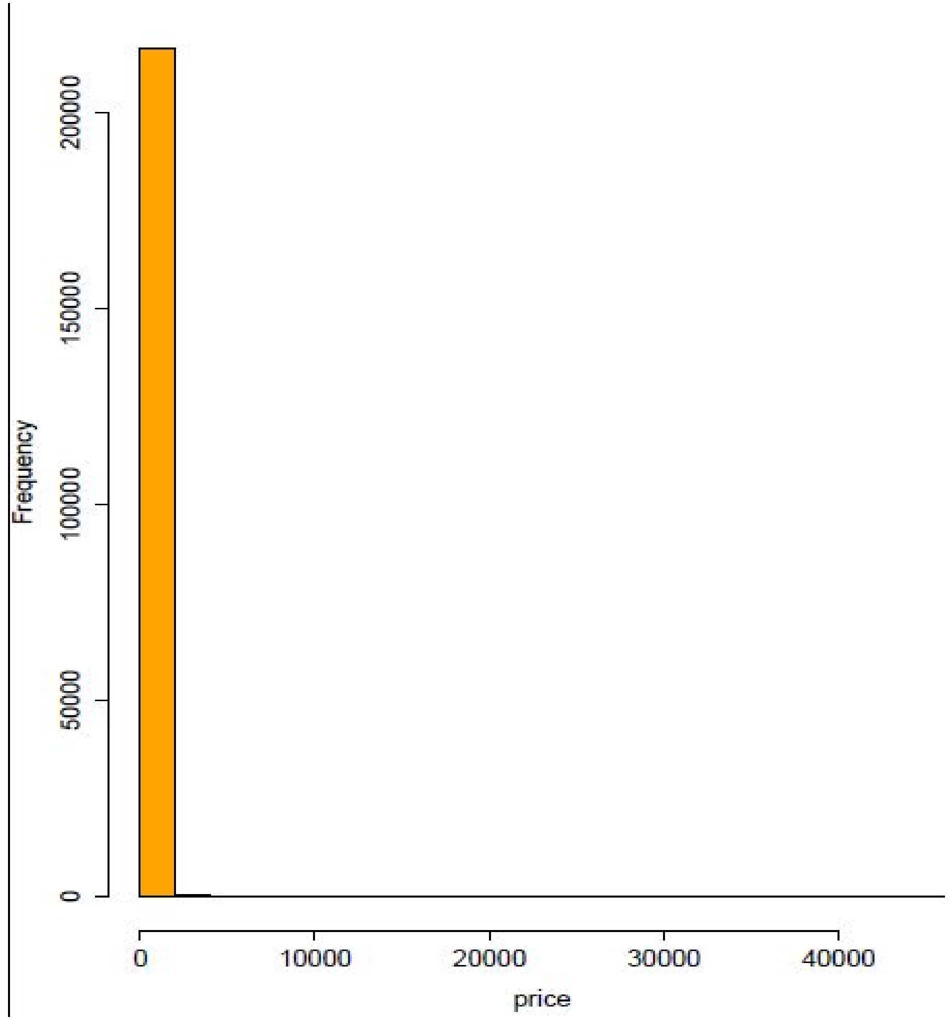
- Grocery shopping dataset
- 200,000 transactions
- 16,000 customer ids
- 17,000 product ids
- 1 month period

p1.R | train | counts | counts2 | counts3 | sales_data1

Filter

	date	cust_id	age_group	address	product_subclass	product_id	quantity	asset	price
1	2001-01-01	141833	F	F	130207	4.710105e+12	2	44	52
2	2001-01-01	1376753	E	E	110217	4.710266e+12	1	150	129
3	2001-01-01	1603071	E	G	100201	4.712019e+12	1	35	39
4	2001-01-01	1738667	E	F	530105	4.710169e+12	1	94	119
5	2001-01-01	2141497	A	B	320407	4.710431e+12	1	100	159
6	2001-01-01	1868685	J	E	110109	4.710044e+12	1	144	190
7	2001-01-02	1101270	D	C	730303	4.714903e+12	1	740	969
8	2001-01-02	1754698	H	A	560402	4.710499e+12	1	676	849
9	2001-01-02	1027365	F	C	530404	9.555009e+12	1	170	219
10	2001-01-03	956710	E	E	500303	4.710367e+12	1	36	59
11	2001-01-04	477796	E	H	100108	5.085399e+07	2	220	270
12	2001-01-05	1267471	C	F	500804	9.310023e+12	1	185	218
13	2001-01-06	904391	[illegible]	[illegible]	110109	4.716782e+12	1	80	89

Exploration and Visualization

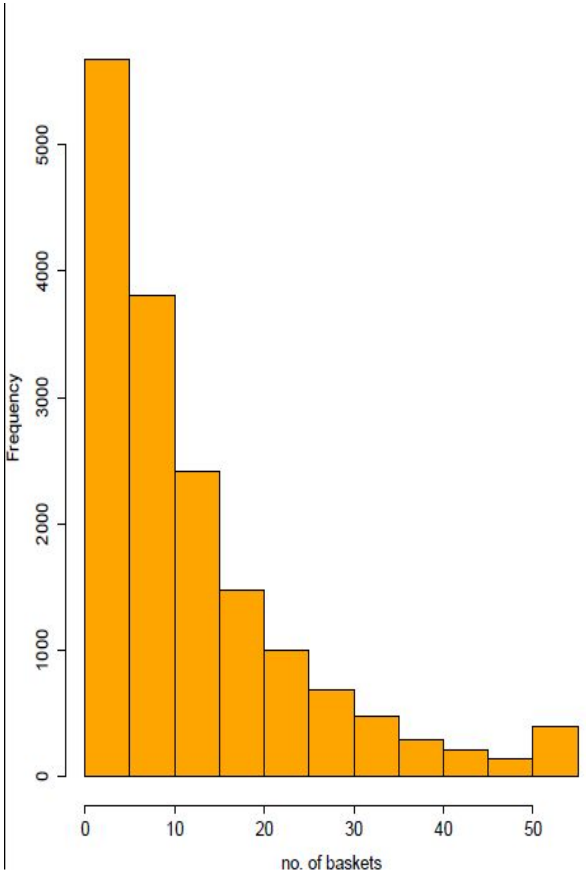


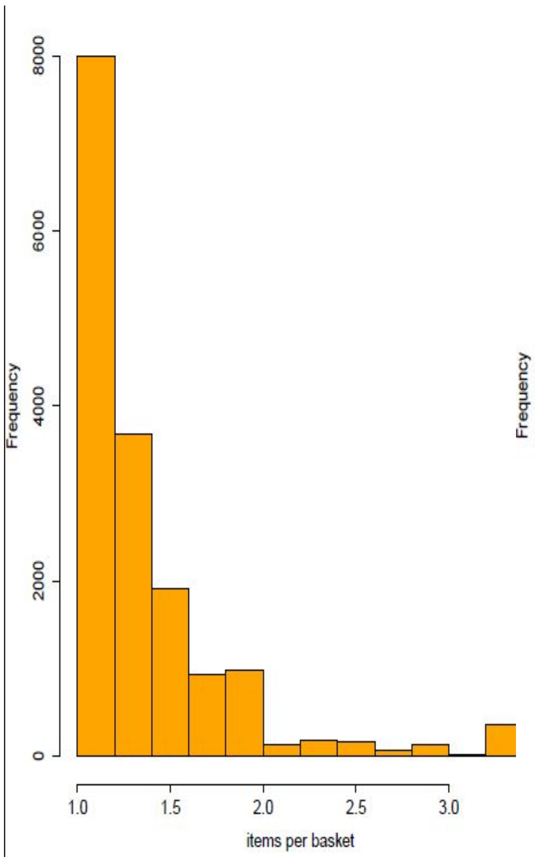
Feature Creation

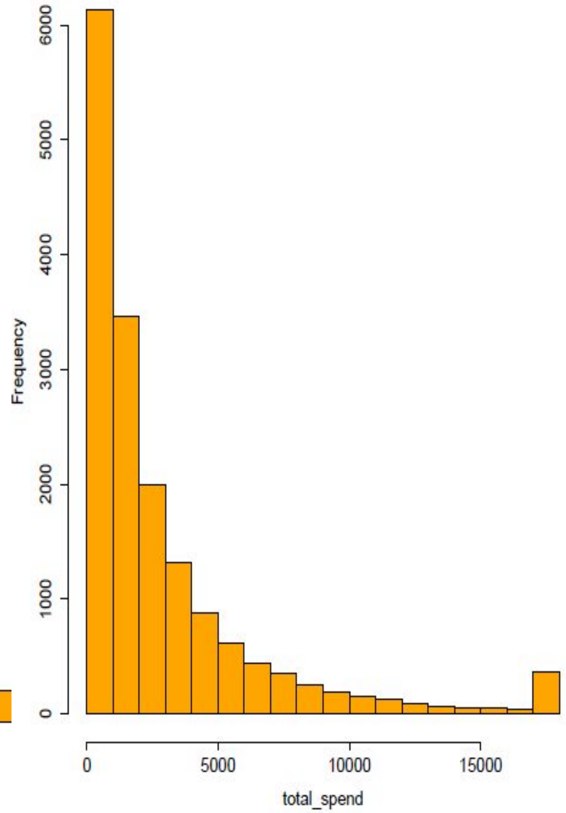
	cust_id	nbask	nitem	spend	cost	mean_item_cost	items_per_basket	total_profit	profit_per_basket
1	141833	25	31	2038	1676	54.06452	1.240000	362	14.480000
2	1376753	8	12	1001	1165	97.08333	1.500000	-164	-20.500000
3	1603071	13	13	618	509	39.15385	1.000000	109	8.384615
4	1738667	42	45	3517	2844	63.20000	1.071429	673	16.023810
5	2141497	11	11	1291	979	89.00000	1.000000	312	28.363636
6	1868685	4	4	580	431	107.75000	1.000000	149	37.250000
7	1101270	20	22	5191	3974	180.63636	1.100000	1217	60.850000
8	1754698	8	9	2385	1944	216.00000	1.125000	441	55.125000
9	1027365	9	12	1485	1208	100.66667	1.333333	277	30.777778

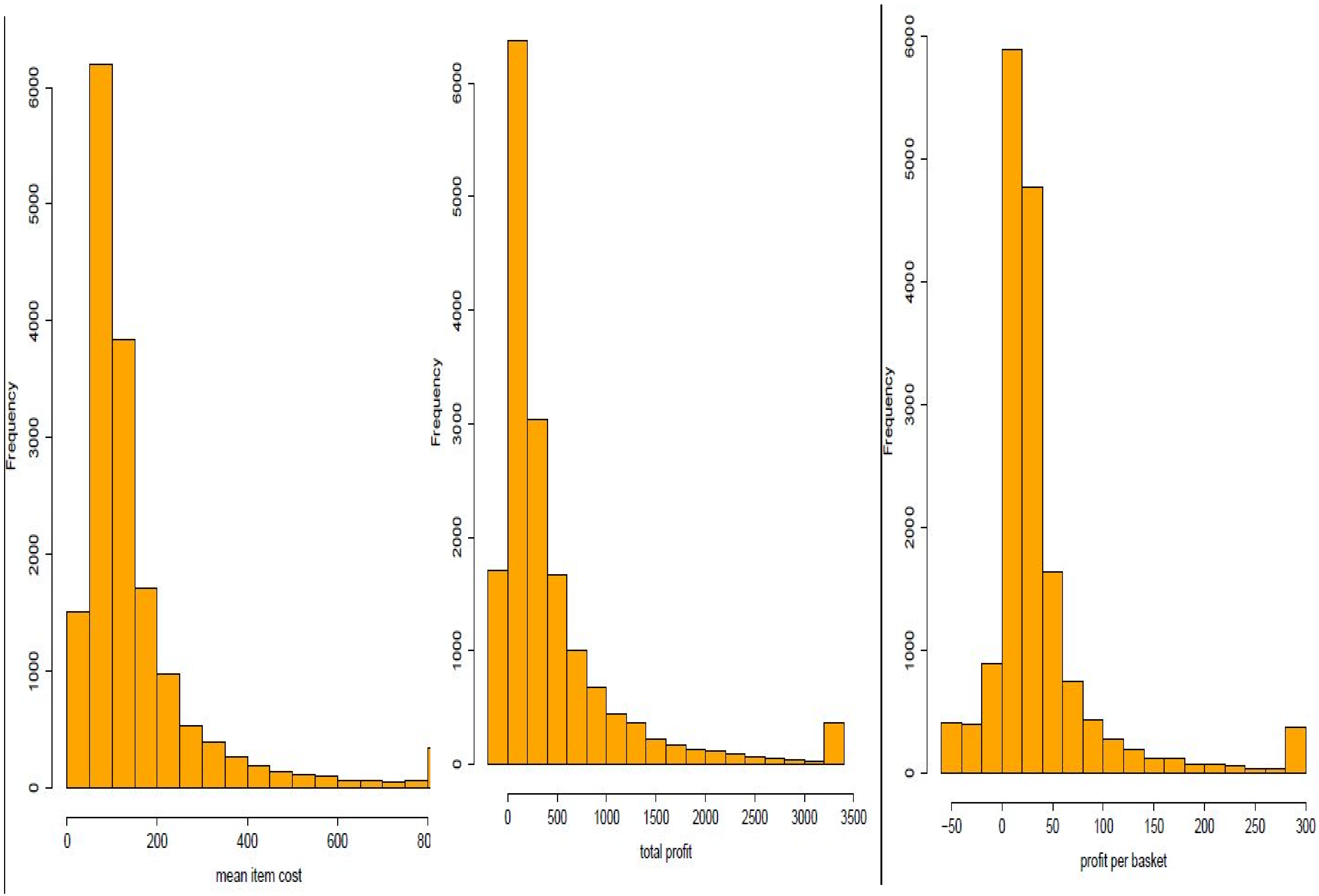
Data Cleaning and Exploratory Analysis

- Detecting outlier and deleting the corresponding observations.
- Some Plots








Frequency
0
1000
2000
3000
4000
5000
6000
0
200
400
600
80
mean item cost
Frequency
0
1000
2000
3000
4000
5000
6000
0
500
1000
1500
2000
2500
3000
3500
total profit
Frequency
0
1000
2000
3000
4000
5000
6000
-50
0
50
100
150
200
250
300
profit per basket

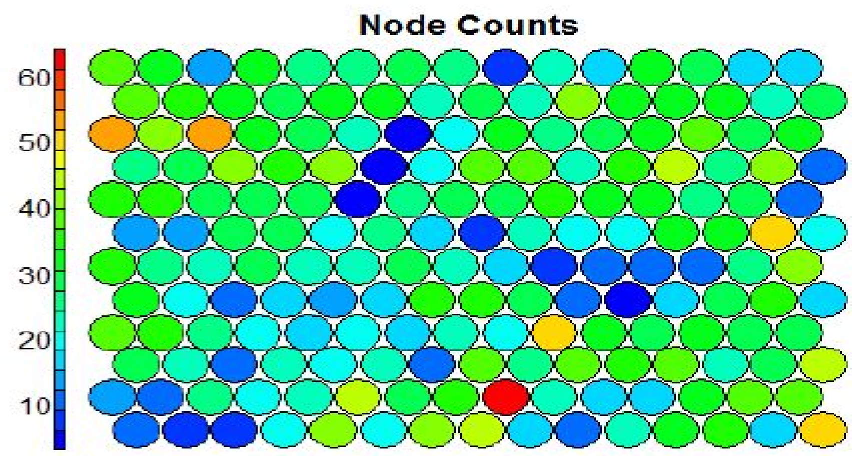
Feature Selection

mean_item_cost	items_per_basket	total_profit	profit_per_basket
54.06452	1.240000	362	14.480000
97.08333	1.500000	-164	-20.500000
39.15385	1.000000	109	8.384615
63.20000	1.071429	673	16.023810
89.00000	1.000000	312	28.363636
107.75000	1.000000	149	37.250000
180.63636	1.100000	1217	60.850000
216.00000	1.125000	441	55.125000
100.66667	1.333333	277	30.777778
100.40000	1.136364	703	31.954545
220.00000	2.000000	100	100.000000
[illegible]	[illegible]	[illegible]	[illegible]

Model Creation

Self Organizing Maps (SOM):

- A Self-Organizing Map (SOM) is a form of unsupervised neural network that produces a low (typically two) dimensional representation of the input space of the set of training samples.
- The SOM visualisation is made up of several nodes.
- Input samples are mapped to the most similar nodes on the SOM. All attributes in input data are used to determine similarity.
- Each node has a weight vector of same size as the input space.
- There is no variable/meaning to the x and y axes



All nodes have:

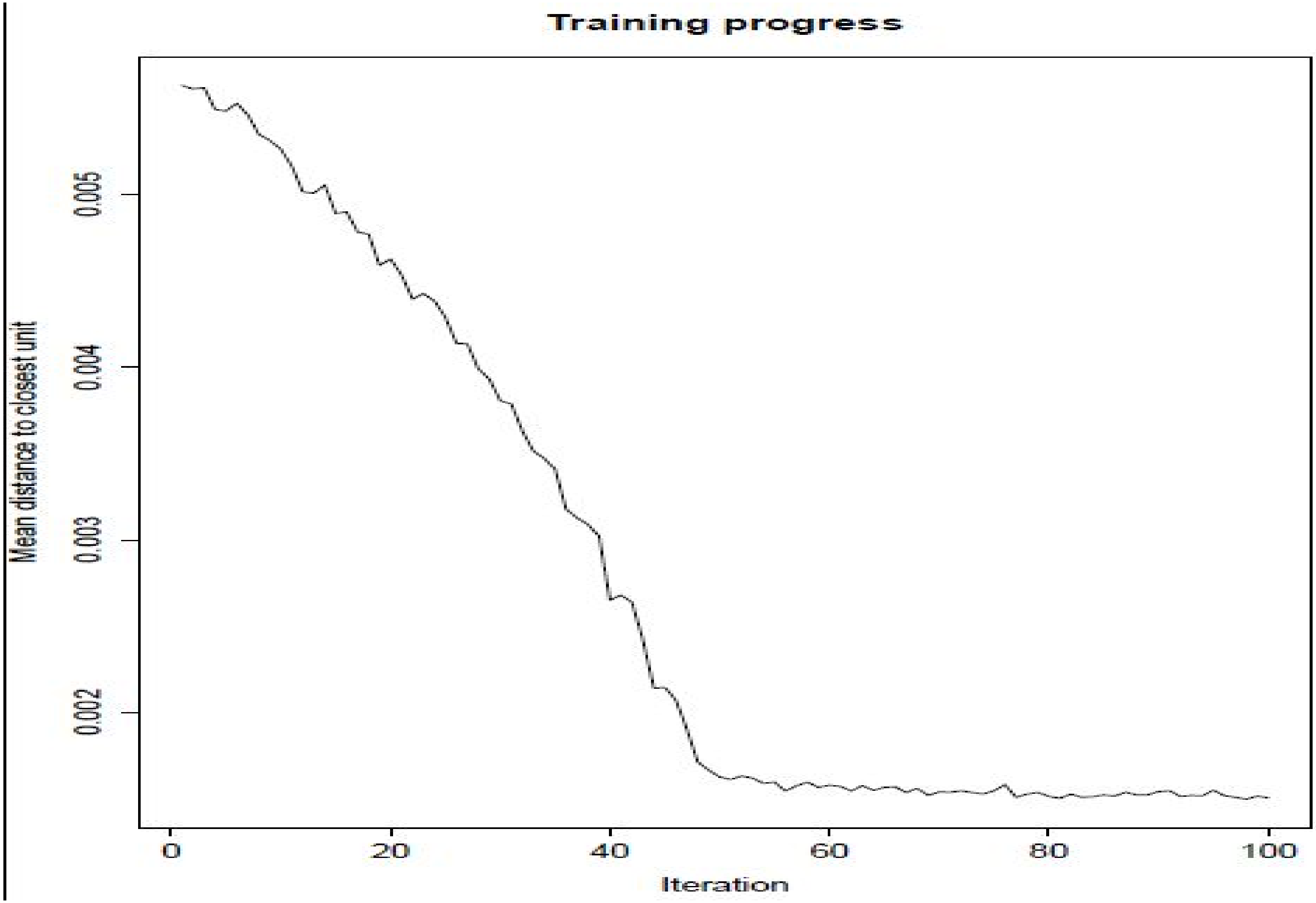
- Position
- Weight Vector
- Associated input samples

SOM: code snapshot

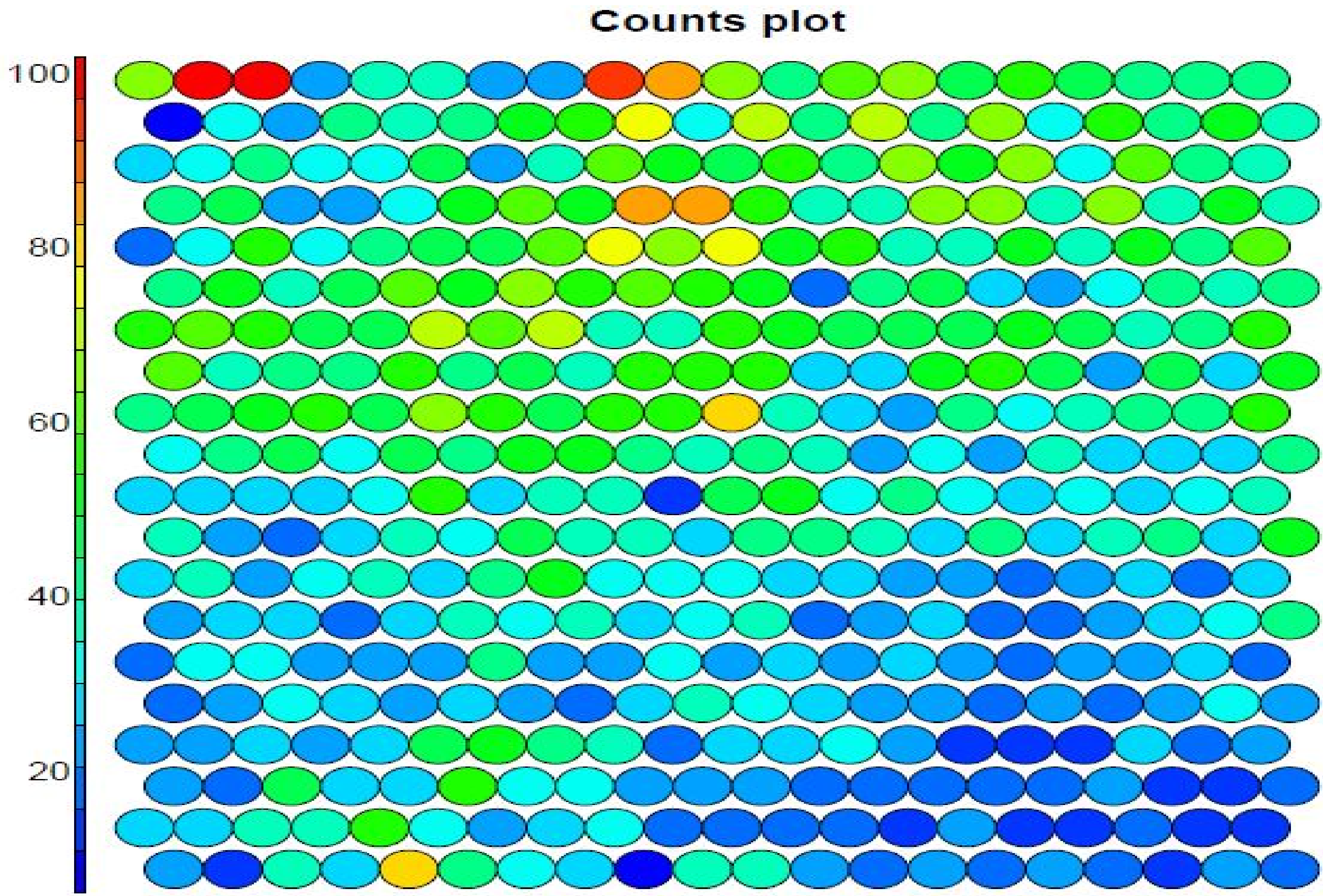
```
require(kohonen)
data_train <- counts3[,c(6,7,8,9)]
data_train_matrix <- as.matrix(scale(data_train))
som_grid <- som_grid <- somgrid(xdim = 20, ydim=20, topo="hexagonal")
som_model <- som(data_train_matrix,
                 grid=som_grid,
                 rlen=100,
                 alpha=c(0.05,0.01),
                 keep.data = TRUE,
                 n.hood="circular")
summary(som_model)
plot(som_model, type = "changes")

coolBlueHotRed <- function(n, alpha = 1)
  {
  rainbow(n, end=4/6, alpha=alpha)[n:1]
}
plot(som_model, type = "counts",palette.name = coolBlueHotRed)
plot(som_model, type = "dist.neighbours")
plot(som_model, type = "codes")
```

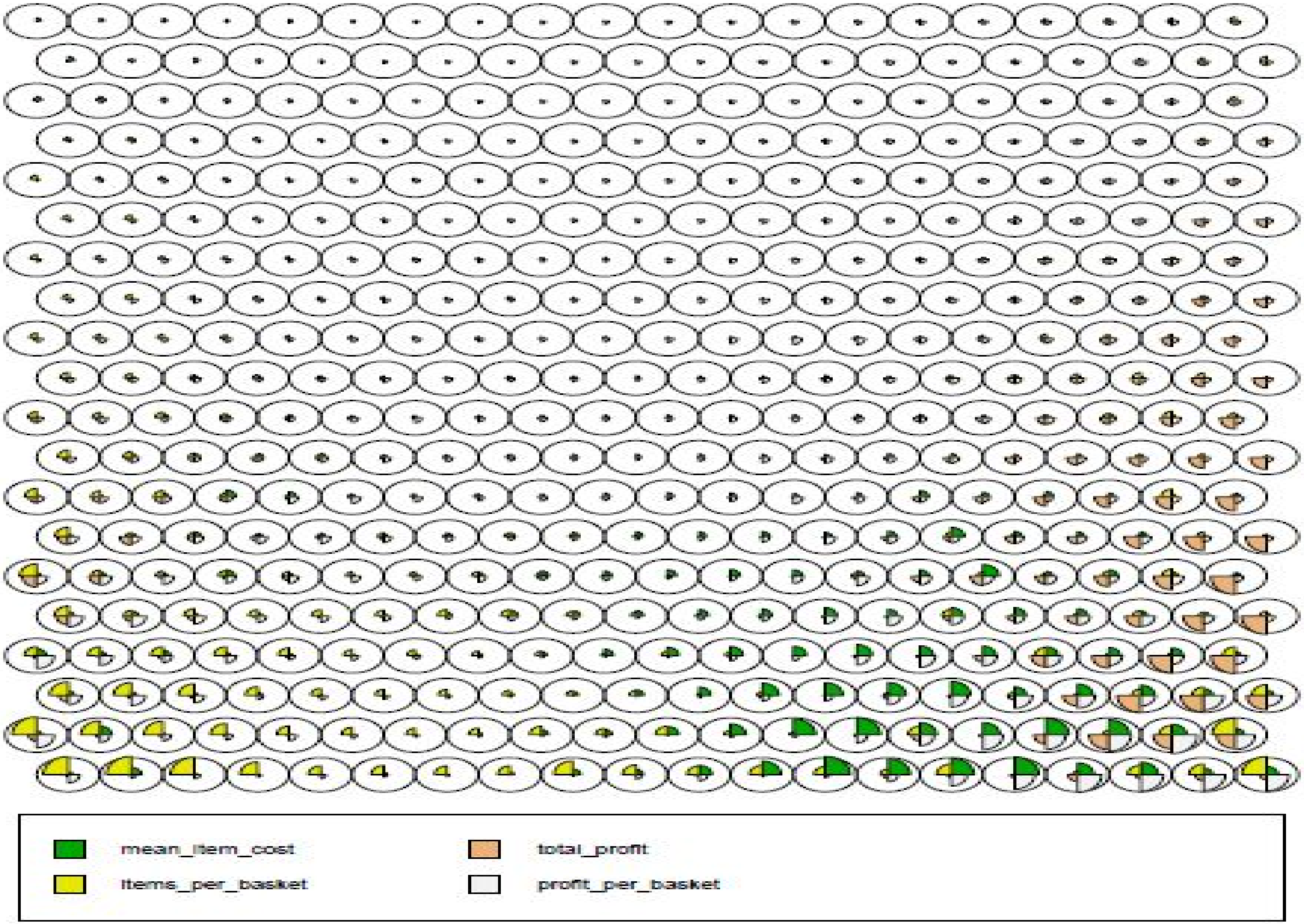
Results and plots



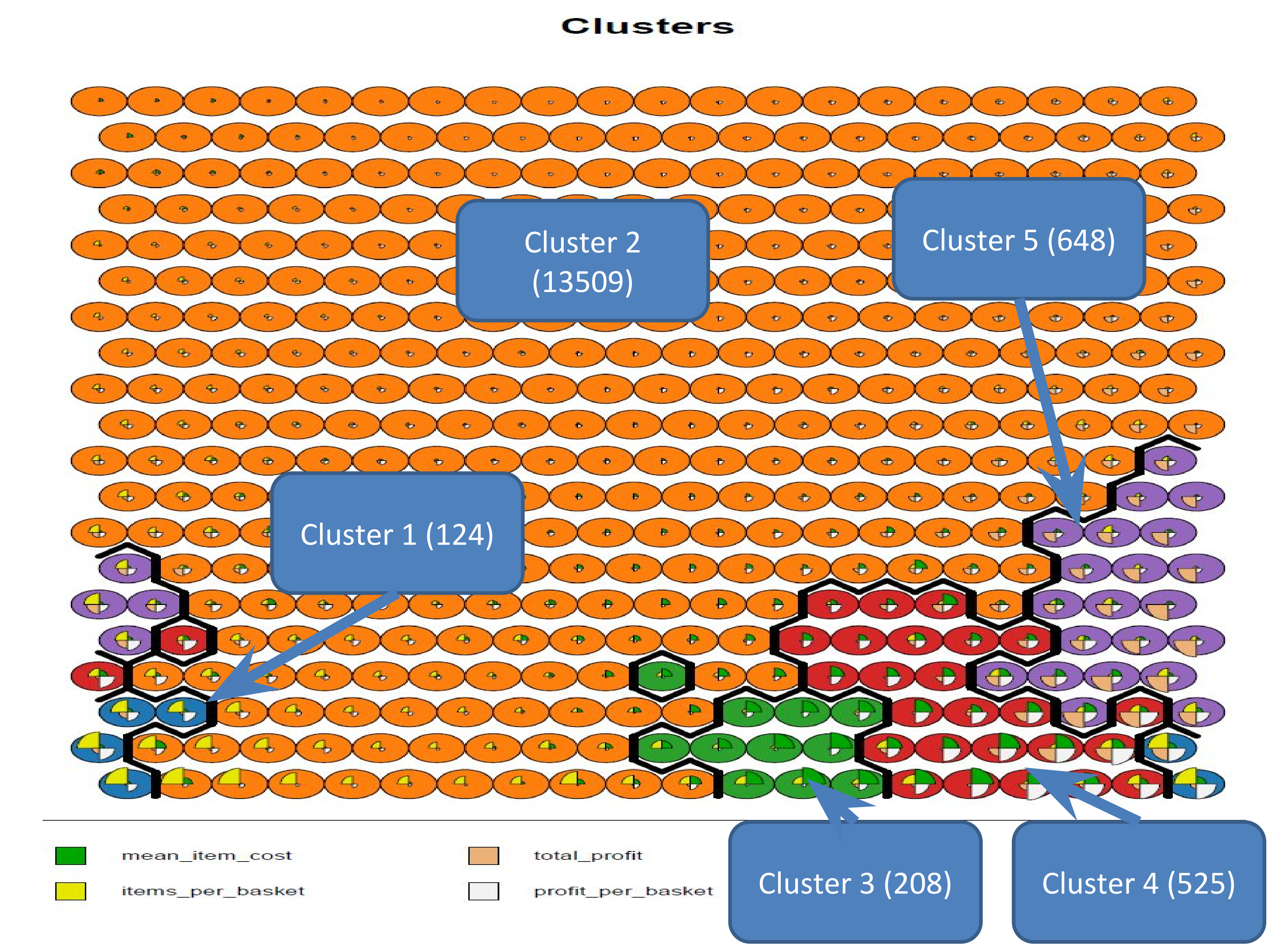
SOM Heat Map



Fan Diagram



mean_item_cost
items_per_basket
total_profit
profit_per_basket



Clusters
Cluster 2 (13509)
Cluster 5 (648)
Cluster 1 (124)
Cluster 3 (208)
Cluster 4 (525)
mean_item_cost
items_per_basket
total_profit
profit_per_basket

Interpretations of Clusters

S.No.	Clusters	Membership	Characteristics
1	1	124	High items per basket, profit per basket
2	2	13509	Average values of features
3	3	208	High mean item cost
4	4	525	High mean item cost, profit per basket
5	5	648	High total profit, profit per basklet

Recommendations

Cluster1: High items per basket, profit per basket

- Customers buy in bulks and so profit per basket is more.
- Total profit is not much.
- These customers probably use to buy fast moving and replenishing products and can be daily consumption low value products.
- Can give promotions and try to drive them towards buying high value products.

Cluster2: Average values of features.

- Since the customers belonging to this cluster do not show specific inclinations to features
- Need not to focus much on them.

Recommendation continued….

Cluster3: High mean item cost

- Customers buy costly products.
- Less no. of baskets.
- Profit margin is not much.
- Can give promotions and try give them discounts on fast moving products.

Cluster4: High mean item cost, profit per basket

- Value segment of customers.
- Purchase high value items
- Bulk purchasing as profit per basket is high
- Can give promotions and keep motivating them so that these customers should not churn.

Recommendation continued….

Cluster5: High total profit, profit per basket

- Customers buy high margin products (like cold drinks..etc)
- Bulk purchasing
- Can give promotions and try drive them towards high value products and FMCG.

Snapshots of customers belongingness.

	id	cluster
1	1069	2
2	1113	2
3	1823	2
4	2189	5
5	4282	2
6	4978	2
7	5241	5
8	5357	2
9	6668	2
10	7795	5
11	8198	2

Thanks