Customer Segmentation

By Google BigQuery
Srisuda Wongvoraruj

K-Means Clustering Model

- CHOOSE CLUSTERS = 3

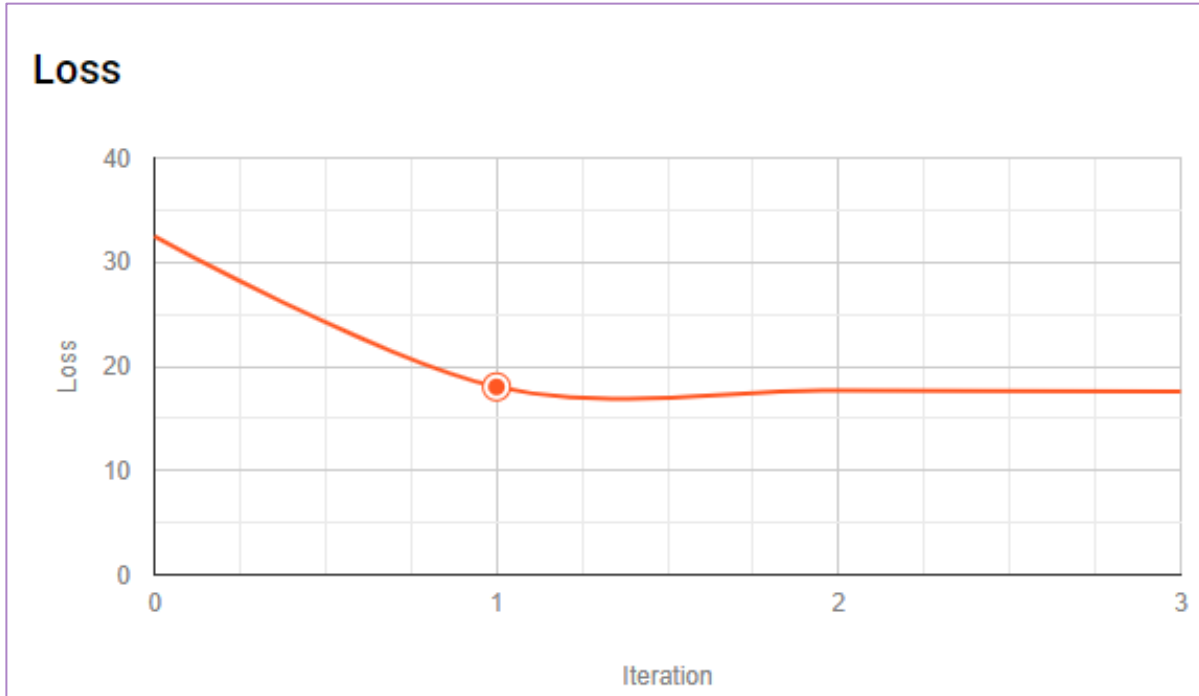
Query editor

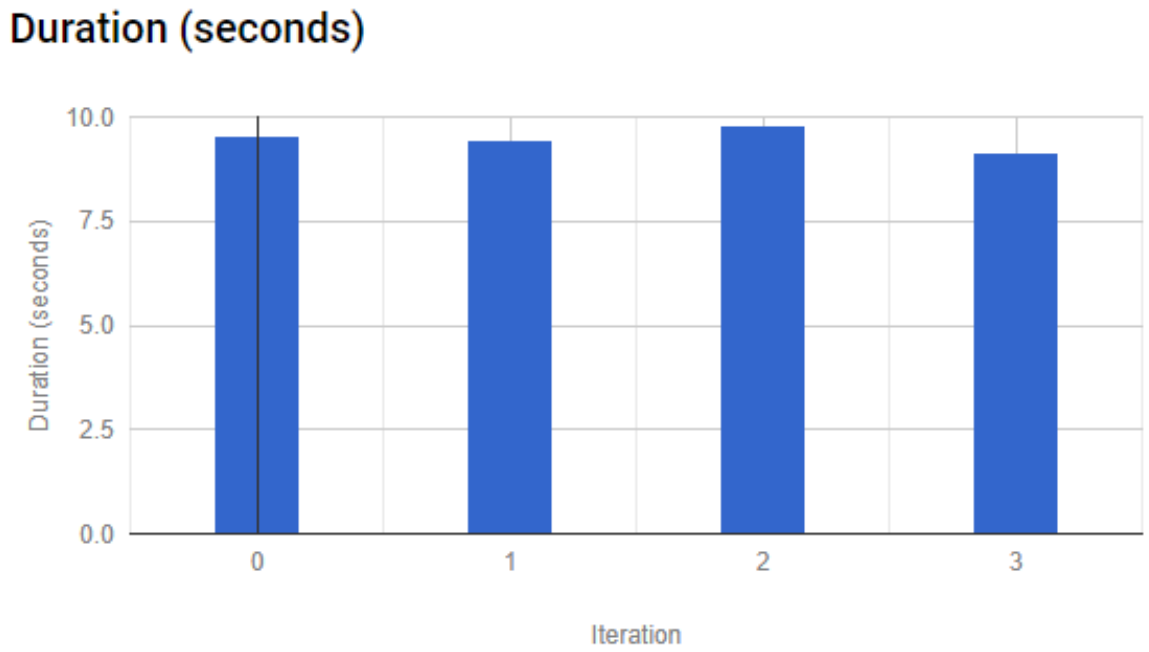
```
1 CREATE MODEL`bads7105-318514.Supermarket.Cluster`
2 OPTIONS( MODEL_TYPE='KMEANS',NUM_CLUSTERS=3,KMEANS_INIT_METHOD='RANDOM')
3 AS SELECT*FROM`bads7105-318514.Supermarket.DATA`
```

- Result

Iteration	Training data loss	Duration (seconds)	Cluster centroid ID	Cluster radius	Cluster size
3	17.5912	9.14	1	4.18406836	363259
			2	4.12616775	238189
			3	4.24942298	355126
2	17.7058	9.78	1	4.23296348	389656
			2	4.14968398	219556
			3	4.21604721	347362
1	18.0301	9.45	1	4.29138006	438394
			2	4.18940282	200071
			3	4.21896085	318109
0	32.4737	9.54	1	5.62991389	499183
			2	5.79740319	139243
			3	5.76167585	318148

Results


Loss
40
30
20
10
0
Loss
0
1
2
3
Iteration


Duration (seconds)
10.0
7.5
5.0
2.5
0.0
Duration (seconds)
0
1
2
3
Iteration

Results

Metrics

Davies–Bouldin index	4.1469
Mean squared distance	17.5912

Numeric features

This table shows the centroid value for each feature. Use the select menu to view more numeric features.

Selected features
QUANTITY, SHOP_HOUR, SHOP_WEEKDAY, SPEND

Centroid ID	Count	QUANTITY	SHOP_HOUR	SHOP_WEEKDAY	SPEND
1	363,259	1.4751	15.4575	3.5984	1.9144
2	238,189	1.5320	13.5177	5.6405	1.6631
3	355,126	1.5479	15.2878	3.4026	1.9556

Results

Categorical features

Each chart below shows the category value distribution for a particular feature. Use the select menu to view more categorical features.

Selected features
10 options selected