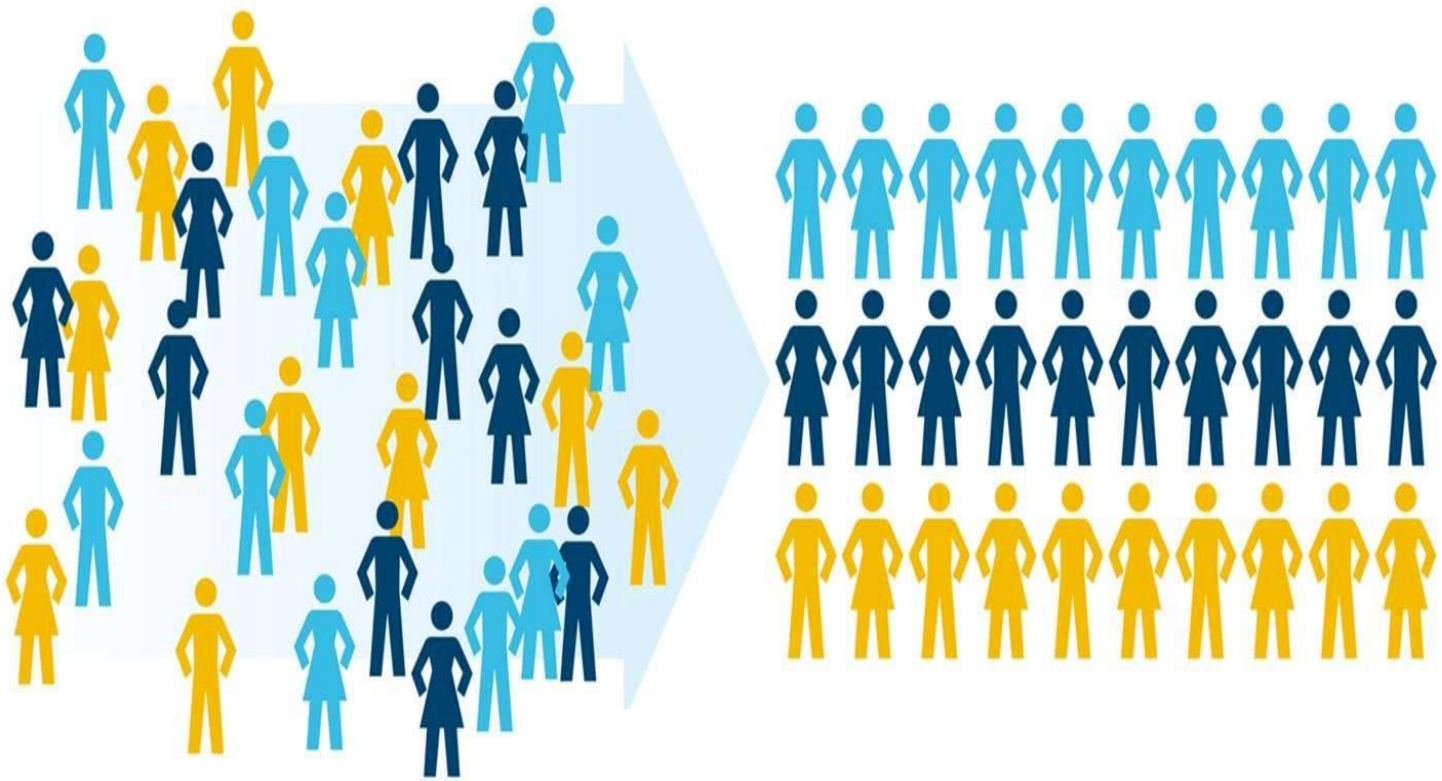
Customer Segmentation “Supermarket”



CUSTOMER SEGMENTATION

Google Big Query (SQL)

Query editor

```
1  Create or replace model
2  `crm-7105.Supermarket.RU_Cluster`
3  options (model_type='kmeans',num_clusters=7)
4  as(
5  SELECT
6  count(distinct basket_id) as total_visit,
7  sum(spend) as total_spend
8  FROM `crm-7105.Supermarket.Segment_RU`
9  where cust_code is not null
10 group by cust_code
11
```

Processing location: US

Run | Save query | Save view | Schedule query | More

RU_Cluster

Centroid Id	Count	total_visit	total_spend
1	896	4.7656	113.1891
2	81	12.8519	4,579.5444
3	564	8.0798	213.8332
4	347	11.6888	388.9621
5	264	12.0947	1,137.2407
6	176	12.5227	2,319.9650
7	3772	1.5215	17.0615

Segment by K-mean (K=7)

Google Big Query (SQL)

Query editor

```sql
SELECT
* EXCEPT(nearest_centroids_distance)
FROM
ML.PREDICT(MODEL `crm-7105.Supermarket.RU_Cluster`,
(
select
cust_code,
count(distinct basket_id) as total_visit,
sum(spend) as total_spend
from `crm-7105.Supermarket.Segment_RU`
where cust_code is not null
group by cust_code


))
```

Processing location: US

Run | Save query | Save view | Schedule query | More

Query results | SAVE RESULTS | EXPLORE DATA

Query complete (14.8 sec elapsed, 129.1 MB processed)

Job information | Results | JSON | Execution details

Row	CENTROID_ID	cust_code	total_visit	total_spend
1	5	CUST0000190355	12	1281.1099999999963
2	7	CUST0000830651	2	2.1

USE Python To Plot graph Separate by group

PYTHON (PANDAS & Matplotlib)

In [ ]:
```
import pandas as pd
import numpy as np
my_data = pd.read_csv('centriod result.csv')
df = pd.DataFrame(my_data)
df['CENTROID_ID'] = df['CENTROID_ID'].replace(1, 'r')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(2, 'g')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(3, 'b')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(4, 'c')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(5, 'm')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(6, 'y')
df['CENTROID_ID'] = df['CENTROID_ID'].replace(7, 'k')
new_df1 = pd.DataFrame({
        'cust_code' : df['cust_code'],
        'total_spend' : df['total_spend'],
        'total_visit' : df['total_visit'],
        'Colors' : df['CENTROID_ID']
    })
new_df
```

In [ ]:
```
import matplotlib.pyplot as plt
%matplotlib qt
plt.title( "total_visit VS total_spend",fontsize=20)
plt.xlabel( "total_visit",fontsize=17)
plt.ylabel( 'total_spend',fontsize=17 )
plt.scatter(new_df1['total_visit'], new_df1['total_spend'], color=new_df1['Colors'], alpha=0.5, edgecolor='k')
plt.show()
```