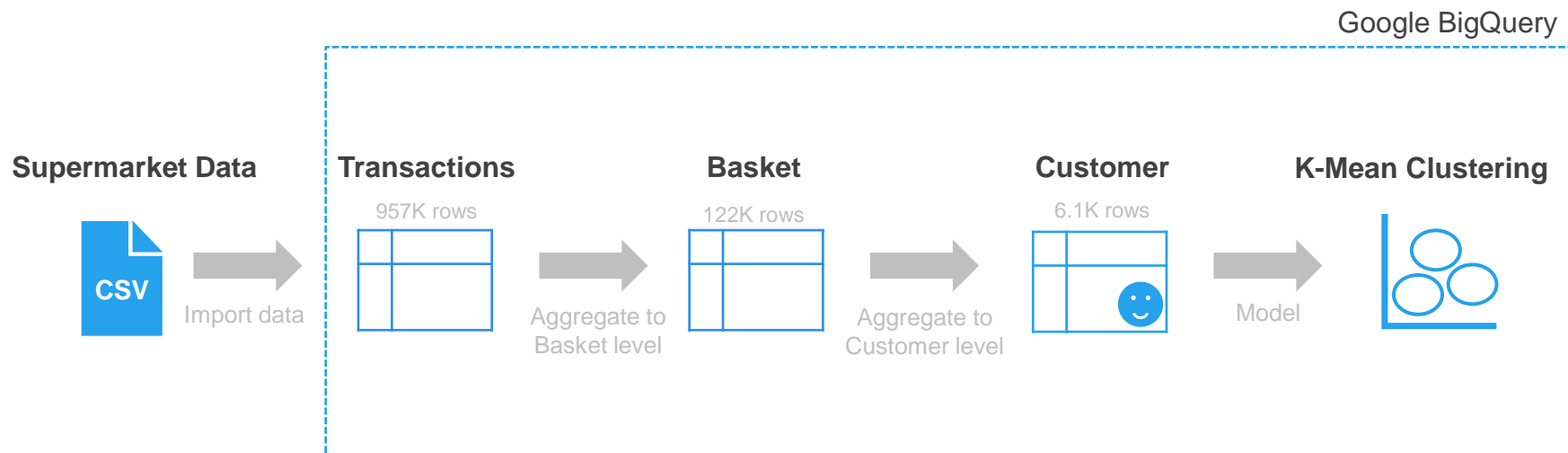




CUSTOMER SEGMENTATION

DUNNHUMBY DATABASE

PROJECT FLOW



FEATURE SELECTION

Customer

6.1K rows



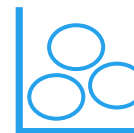
Feature Selection

Features	Definition
AGE	No of month between first and last visit
LAST_VISIT	No of month between last visit to this month
PROB_AFTERNOON	Probability to visit on afternoon
PROB_EVENING	Probability to visit on evening
PROB_BAS_M	Probability to purchase basket size "M"
PROB_BAS_S	Probability to purchase basket size "S"
PROB_WEEKEND	Probability to visit on weekend
PROB_WEEKDAY	Probability to visit on weekday



Model

K-Mean Clustering



BIGQUERY RESULT

Metrics

Davies-Bouldin index	1.6476
Mean squared distance	4.0413

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_M, PROB_BAS_S, PROB_EVEN...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_M	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	1,778	<div><div></div></div> 19.5832	<div><div></div></div> 2.4454	<div><div></div></div> 0.6297	<div><div></div></div> 0.4068	<div><div></div></div> 0.2554	<div><div></div></div> 0.1423	<div><div></div></div> 0.7292	<div><div></div></div> 0.2708
2	1,361	<div><div></div></div> 19.2572	<div><div></div></div> 2.3784	<div><div></div></div> 0.2497	<div><div></div></div> 0.3161	<div><div></div></div> 0.4160	<div><div></div></div> 0.5773	<div><div></div></div> 0.6900	<div><div></div></div> 0.3100
3	795	<div><div></div></div> 1.2528	<div><div></div></div> 14.8201	<div><div></div></div> 0.4882	<div><div></div></div> 0.3274	<div><div></div></div> 0.4794	<div><div></div></div> 0.3038	<div><div></div></div> 0.0574	<div><div></div></div> 0.9426
4	1,122	<div><div></div></div> 2.4688	<div><div></div></div> 13.6738	<div><div></div></div> 0.4551	<div><div></div></div> 0.0171	<div><div></div></div> 0.9608	<div><div></div></div> 0.2869	<div><div></div></div> 0.9421	<div><div></div></div> 0.0579
5	1,044	<div><div></div></div> 1.4531	<div><div></div></div> 15.2443	<div><div></div></div> 0.4024	<div><div></div></div> 0.6389	<div><div></div></div> 0.0560	<div><div></div></div> 0.3994	<div><div></div></div> 0.9617	<div><div></div></div> 0.0383

AVG.
by cluster group

CENTROID_ID	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_EVENING	PROB_BAS_M	PROB_BAS_S	PROB_WEEKEND	PROB_WEEKDAY
1	20.24	2.26	0.62	0.15	0.4	0.26	0.28	0.72
2	18.75	2.52	0.24	0.59	0.32	0.42	0.31	0.69
3	1.27	14.81	0.49	0.3	0.33	0.48	0.94	0.06
4	2.45	13.68	0.45	0.29	0.02	0.96	0.06	0.94
5	1.52	15.0	0.43	0.37	0.64	0.05	0.04	0.96

INTREPRETING RESULTS

Loyalty housewife	1.7K of members (29% of base) 66% of revenue contribution Long period membership Medium to high basket size Low Recency Weekday afternoon visit
Offer coupon discount for maintain frequency	

Loyalty salaryman	1.3K of members (22% of base) 32% of revenue contribution Long period membership Small to medium basket size Low Recency Weekday evening visit
Offer promotion to increase basket size and maintain frequency	

High value Win-back	1.1K of members (18% of base) 1.5% of revenue contribution Short period membership Medium to high basket size High Recency Weekday afternoon visit
Offer attractive promotion to win-back	

Weekend Win-back	0.8K of members (13% of base) 0.7% of revenue contribution Short period membership Small to medium basket size High Recency Weekend afternoon visit
Offer promotion to win-back	

Weekday Win-back	1.1K of members (18% of base) 0.4% of revenue contribution Short period membership Small to medium basket size High Recency Weekday afternoon visit
Offer promotion to win-back	

APPENDIX

BigQuery : Transaction to Basket

```
CREATE TABLE `bads7105-308204.supermarket.basket` AS
```

```
SELECT *, DATE_DIFF(DATE'2008-07-06',SHOP_DATE,MONTH) AS LAST_VISIT  
FROM (SELECT
```

```
  CAST(BASKET_ID AS STRING) AS BASKET_ID,  
  CUST_CODE,  
  CUST_PRICE_SENSITIVITY,  
  CUST_LIFESTAGE,  
  PARSE_DATE("%Y%m%d",  
    CAST (SHOP_DATE AS STRING)) AS SHOP_DATE,  
  SHOP_WEEKDAY,  
  (CASE WHEN shop_weekday IN (1, 7) THEN 1 ELSE 0 END) AS is_weekend,  
  SHOP_HOUR,  
  (CASE WHEN shop_hour <= 11 THEN 'morning'  
    WHEN shop_hour BETWEEN 12 AND 16 THEN 'afternoon'  
    ELSE 'evening' END ) AS shop_daytime,  
  BASKET_SIZE,  
  BASKET_TYPE,  
  BASKET_PRICE_SENSITIVITY,  
  BASKET_DOMINANT_MISSION,  
  SUM(SPENDING) AS TOTAL_SPENDING
```

```
FROM `bads7105-308204.supermarket.supermarket`
```

```
WHERE CUST_CODE IS NOT NULL
```

```
GROUP BY BASKET_ID, CUST_CODE, CUST_PRICE_SENSITIVITY, CUST_LIFESTAGE, SHOP_DATE,  
SHOP_WEEKDAY, SHOP_HOUR, BASKET_SIZE, BASKET_TYPE, BASKET_PRICE_SENSITIVITY, BASKET_DOMINANT_MISSION )
```

BigQuery : Basket to Customer

CUSTOMER SINGLE VIEW

```
CREATE TABLE `bads7105-308204.supermarket.customer` AS
```

```
SELECT *
```

```
# CUSTOMER PRICE SENSIVITY ("LA"=Less Affluent, "MM"=Mid Market, "UM"=Up Market, "XX"=Unclass)
```

```
(CASE WHEN CUST_PRICE_SENSITIVITY = 'LA' THEN 1 ELSE 0 END) AS CUS_PRICE_LA,
```

```
(CASE WHEN CUST_PRICE_SENSITIVITY = 'MM' THEN 1 ELSE 0 END) AS CUS_PRICE_MM,
```

```
(CASE WHEN CUST_PRICE_SENSITIVITY = 'UM' THEN 1 ELSE 0 END) AS CUS_PRICE_UM,
```

```
#CUSTOMER LIFESTAGE ("OA"=Older Adult, "OF"=Older Families, "OT"=Other, "PE"=Pensioners, "YA"=Young Adults, "YF"=Young Families)
```

```
(CASE WHEN CUST_LIFESTAGE = 'OA' THEN 1 ELSE 0 END) AS CUS_LIFE_OA,
```

```
(CASE WHEN CUST_LIFESTAGE = 'OF' THEN 1 ELSE 0 END) AS CUS_LIFE_OF,
```

```
(CASE WHEN CUST_LIFESTAGE = 'OT' THEN 1 ELSE 0 END) AS CUS_LIFE_OT,
```

```
(CASE WHEN CUST_LIFESTAGE = 'PE' THEN 1 ELSE 0 END) AS CUS_LIFE_PE,
```

```
(CASE WHEN CUST_LIFESTAGE = 'YA' THEN 1 ELSE 0 END) AS CUS_LIFE_YA,
```

```
(CASE WHEN CUST_LIFESTAGE = 'YF' THEN 1 ELSE 0 END) AS CUS_LIFE_YF,
```

```
FROM (
```

```
SELECT CUST_CODE, CUST_PRICE_SENSITIVITY, CUST_LIFESTAGE,
```

```
COUNT(DISTINCT BASKET_ID) AS TOTAL_VISIT,
```

```
SUM(TOTAL_SPEND) AS TOTAL_SPEND,
```

```
ROUND(AVG(TOTAL_SPEND),2) AS AVG_SPEND,
```

```
MIN(LAST_VISIT) AS LAST_VISIT,
```

```
COALESCE(SAFE_DIVIDE(COUNT(DISTINCT BASKET_ID),DATE_DIFF(MAX(SHOP_DATE),MIN(SHOP_DATE),MONTH)),0) AS AVG_VISIT_PER_MONTH,
```

```
# NO OF MONTH BETWEEN FIRST VISIT AND LAST VISIT
```

```
DATE_DIFF(MAX(SHOP_DATE),MIN(SHOP_DATE),MONTH) AS AGE,
```

```
# PROBABILITY TO VISIT ON WEEKDAY & WEEKEND
```

```
COALESCE(SAFE_DIVIDE(SUM(is_weekend),COUNT(DISTINCT BASKET_ID)),0) AS PROB_WEEKEND,
```

```
COALESCE(SAFE_DIVIDE(SUM(IF(is_weekend=0,1,0)),COUNT(DISTINCT BASKET_ID)),0) AS PROB_WEEKDAY,
```


BigQuery : Basket to Customer

PROBABILITY TO VISIT ON TIME

```
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN shop_daytime = 'morning' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_MORNING,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN shop_daytime = 'afternoon' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_AFTERNOON,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN shop_daytime = 'evening' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_EVENING,
```

PROBABILITY OF BASKET SIZE TO TOTAL VISIT

```
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_SIZE='L' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_L,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_SIZE='M' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_M,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_SIZE='S' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_S,
```

PROBABILITY OF BASKET TYPE TO TOTAL VISIT

```
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_TYPE='Full Shop' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_TYPE_FULL,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_TYPE='Small Shop' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_TYPE_SMALL,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_TYPE='Top Up' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_TYPE_TOPUP,
```

PROBABILITY OF BASKET PRICE SENSITIVITY TO TOTAL VISIT

```
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_PRICE_SENSITIVITY='LA' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_PRICE_LA,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_PRICE_SENSITIVITY='MM' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_PRICE_MM,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_PRICE_SENSITIVITY='UM' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_PRICE_UM,
```

PROBABILITY OF BASKET DOMINANT TO TOTAL VISIT

```
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_DOMINANT_MISSION='Fresh' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_DOMI_FRESH,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_DOMINANT_MISSION='Grocery' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_DOMI_GROCE  
RY,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_DOMINANT_MISSION='Mixed' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_DOMI_MIXED,  
COALESCE(SAFE_DIVIDE(SUM(CASE WHEN BASKET_DOMINANT_MISSION='Nonfood' THEN 1 ELSE 0 END),COUNT(DISTINCT BASKET_ID)),0) AS PROB_BAS_DOMI_NONFO  
OD
```

```
FROM `bads7105-308204.supermarket.basket`
```

```
GROUP BY CUST_CODE,CUST_PRICE_SENSITIVITY,CUST_LIFESTAGE)
```

BigQuery : K-Mean Clustering

Modelling

```
CREATE MODEL
`bads7105-308204.supermarket.mymodel`

OPTIONS
( MODEL_TYPE='KMEANS',
  NUM_CLUSTERS=7,
  KMEANS_INIT_METHOD='RANDOM') AS

SELECT
AGE,
LAST_VISIT,
PROB_AFTERNOON,
PROB_EVENING,
PROB_BAS_M,
PROB_BAS_S,
PROB_WEEKEND,
PROB_WEEKDAY
FROM `bads7105-308204.supermarket.customer`
WHERE CUST_CODE IS NOT NULL
```

Model Interpreting

```
CREATE TABLE `bads7105-308204.supermarket.result7_3` AS
SELECT
CENTROID_ID,
ROUND(AVG(AGE),2) AS AGE,
ROUND(AVG(LAST_VISIT),2) AS LAST_VISIT,
ROUND(AVG(PROB_AFTERNOON),2) AS PROB_AFTERNOON,
ROUND(AVG(PROB_EVENING),2) AS PROB_EVENING,
ROUND(AVG(PROB_BAS_M),2) AS PROB_BAS_M ,
ROUND(AVG(PROB_BAS_S),2) AS PROB_BAS_S,
ROUND(AVG(PROB_WEEKEND),2) AS PROB_WEEKEND,
ROUND(AVG(PROB_WEEKDAY),2) AS PROB_WEEKDAY

from
(select
* except(nearest_centroids_distance)
from
ML.PREDICT(MODEL `bads7105-308204.supermarket.mymodel7_3`,
(

SELECT AGE, LAST_VISIT,
PROB_AFTERNOON, PROB_EVENING,
PROB_BAS_M, PROB_BAS_S,
PROB_WEEKEND, PROB_WEEKDAY
FROM `bads7105-308204.supermarket.customer`
WHERE CUST_CODE IS NOT NULL
)))
GROUP BY CENTROID_ID
```

BigQuery : K-Mean Clustering

Metrics

Davies–Bouldin index	3.0893
Mean squared distance	24.4388

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_DOMI_FRESH, PROB_BAS_M, ...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_DOMI_FRESH	PROB_BAS_M	PROB_BAS_PRICE_MM	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	805	<div><div></div></div> 2.3739	<div><div></div></div> 14.1801	<div><div></div></div> 0.7169	<div><div></div></div> 0.7041	<div><div></div></div> 0.4829	<div><div></div></div> 0.3434	<div><div></div></div> 0.4985	<div><div></div></div> 0.2330	<div><div></div></div> 0.7988	<div><div></div></div> 0.2012
2	334	<div><div></div></div> 26.0539	<div><div></div></div> 0.3174	<div><div></div></div> 0.4628	<div><div></div></div> 0.5327	<div><div></div></div> 0.3630	<div><div></div></div> 0.5042	<div><div></div></div> 0.2263	<div><div></div></div> 0.3027	<div><div></div></div> 0.7161	<div><div></div></div> 0.2839
3	1,042	<div><div></div></div> 17.0413	<div><div></div></div> 4.4885	<div><div></div></div> 0.4728	<div><div></div></div> 0.4046	<div><div></div></div> 0.3145	<div><div></div></div> 0.4231	<div><div></div></div> 0.1352	<div><div></div></div> 0.3538	<div><div></div></div> 0.7152	<div><div></div></div> 0.2848
4	1,453	<div><div></div></div> 17.6105	<div><div></div></div> 3.5836	<div><div></div></div> 0.4707	<div><div></div></div> 0.5398	<div><div></div></div> 0.5041	<div><div></div></div> 0.6058	<div><div></div></div> 0.4029	<div><div></div></div> 0.3540	<div><div></div></div> 0.6636	<div><div></div></div> 0.3364
5	656	<div><div></div></div> 5.0107	<div><div></div></div> 11.3430	<div><div></div></div> 0.0682	<div><div></div></div> 0.3286	<div><div></div></div> 0.2947	<div><div></div></div> 0.7612	<div><div></div></div> 0.6941	<div><div></div></div> 0.0253	<div><div></div></div> 0.7226	<div><div></div></div> 0.2774
6	765	<div><div></div></div> 6.3673	<div><div></div></div> 11.3111	<div><div></div></div> 0.4778	<div><div></div></div> 0.4856	<div><div></div></div> 0.1998	<div><div></div></div> 0.5855	<div><div></div></div> 0.0375	<div><div></div></div> 0.3809	<div><div></div></div> 0.7264	<div><div></div></div> 0.2736
7	1,045	<div><div></div></div> 4.0144	<div><div></div></div> 12.5598	<div><div></div></div> 0.4398	<div><div></div></div> 0.1881	<div><div></div></div> 0.1757	<div><div></div></div> 0.9141	<div><div></div></div> 0.8141	<div><div></div></div> 0.5161	<div><div></div></div> 0.6898	<div><div></div></div> 0.3102

1st Model Result

- From all features, We found top10 features that nearest centroid.

BigQuery : K-Mean Clustering

Metrics

Davies-Bouldin index	1.911
Mean squared distance	5.2666

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_DOMI_FRESH, PROB_BAS_M, ...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_DOMI_FRESH	PROB_BAS_M	PROB_BAS_PRICE_MM	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	557	<div><div></div></div> 1.3698	<div><div></div></div> 14.9892	<div><div></div></div> 0.6985	<div><div></div></div> 0.4089	<div><div></div></div> 0.3223	<div><div></div></div> 0.6501	<div><div></div></div> 0.4872	<div><div></div></div> 0.0238	<div><div></div></div> 0.0650	<div><div></div></div> 0.9350
2	1,746	<div><div></div></div> 21.8368	<div><div></div></div> 1.7520	<div><div></div></div> 0.5707	<div><div></div></div> 0.5076	<div><div></div></div> 0.3845	<div><div></div></div> 0.5250	<div><div></div></div> 0.2809	<div><div></div></div> 0.1778	<div><div></div></div> 0.7073	<div><div></div></div> 0.2927
3	1,070	<div><div></div></div> 4.1187	<div><div></div></div> 11.9364	<div><div></div></div> 0.5264	<div><div></div></div> 0.3351	<div><div></div></div> 0.0272	<div><div></div></div> 0.8190	<div><div></div></div> 0.9383	<div><div></div></div> 0.1527	<div><div></div></div> 0.9118	<div><div></div></div> 0.0882
4	634	<div><div></div></div> 1.5047	<div><div></div></div> 15.5978	<div><div></div></div> 0.0396	<div><div></div></div> 0.4912	<div><div></div></div> 0.3713	<div><div></div></div> 0.5044	<div><div></div></div> 0.3320	<div><div></div></div> 0.8908	<div><div></div></div> 0.9709	<div><div></div></div> 0.0291
5	1,026	<div><div></div></div> 18.9152	<div><div></div></div> 2.3587	<div><div></div></div> 0.2407	<div><div></div></div> 0.4498	<div><div></div></div> 0.3657	<div><div></div></div> 0.5652	<div><div></div></div> 0.3582	<div><div></div></div> 0.6221	<div><div></div></div> 0.7108	<div><div></div></div> 0.2892
6	760	<div><div></div></div> 2.5974	<div><div></div></div> 13.3553	<div><div></div></div> 0.7246	<div><div></div></div> 0.4600	<div><div></div></div> 0.6774	<div><div></div></div> 0.5391	<div><div></div></div> 0.0488	<div><div></div></div> 0.0493	<div><div></div></div> 0.9332	<div><div></div></div> 0.0668
7	307	<div><div></div></div> 2.2117	<div><div></div></div> 12.9805	<div><div></div></div> 0.0214	<div><div></div></div> 0.4530	<div><div></div></div> 0.3079	<div><div></div></div> 0.6642	<div><div></div></div> 0.5112	<div><div></div></div> 0.9494	<div><div></div></div> 0.1348	<div><div></div></div> 0.8652

2nd Model Result

- Delete features “PROB_BAS_PRICE” and “PROB_BAS_DOMI” because it don’t have different mean between cluster that use for segmentation

AVG.
by cluster group

CENTROID_ID	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_EVENING	PROB_BAS_S	PROB_BAS_M	PROB_WEEKEND	PROB_WEEKDAY	PROB_BAS_PRICE_MM	PROB_BAS_DOMI_FRESH
1	1.38	15.0	0.7	0.03	0.49	0.32	0.93	0.07	0.65	0.41
2	21.78	1.78	0.57	0.18	0.28	0.39	0.29	0.71	0.53	0.5
3	4.32	11.97	0.57	0.09	0.93	0.03	0.09	0.91	0.8	0.36
4	1.39	15.18	0.03	0.92	0.41	0.33	0.02	0.98	0.57	0.44
5	18.88	2.37	0.24	0.62	0.37	0.36	0.29	0.71	0.57	0.45
6	2.44	13.53	0.72	0.05	0.05	0.68	0.06	0.94	0.53	0.47
7	2.06	13.12	0.02	0.95	0.51	0.31	0.87	0.13	0.66	0.46

BigQuery : K-Mean Clustering

Metrics

Davies-Bouldin index	1.5206
Mean squared distance	3.4565

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_M, PROB_BAS_S, PROB_EVEN...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_M	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	760	<div><div></div></div> 14.2632	<div><div></div></div> 3.9776	<div><div></div></div> 0.5655	<div><div></div></div> 0.1484	<div><div></div></div> 0.7128	<div><div></div></div> 0.1345	<div><div></div></div> 0.7037	<div><div></div></div> 0.2963
2	819	<div><div></div></div> 17.5421	<div><div></div></div> 2.7656	<div><div></div></div> 0.1917	<div><div></div></div> 0.3316	<div><div></div></div> 0.3775	<div><div></div></div> 0.7116	<div><div></div></div> 0.6905	<div><div></div></div> 0.3095
3	763	<div><div></div></div> 1.2202	<div><div></div></div> 14.8375	<div><div></div></div> 0.4808	<div><div></div></div> 0.3307	<div><div></div></div> 0.4776	<div><div></div></div> 0.3117	<div><div></div></div> 0.0406	<div><div></div></div> 0.9594
4	941	<div><div></div></div> 1.2019	<div><div></div></div> 15.0659	<div><div></div></div> 0.4360	<div><div></div></div> 0.0111	<div><div></div></div> 0.9717	<div><div></div></div> 0.3143	<div><div></div></div> 0.9727	<div><div></div></div> 0.0273
5	657	<div><div></div></div> 1.9132	<div><div></div></div> 14.6195	<div><div></div></div> 0.0704	<div><div></div></div> 0.6165	<div><div></div></div> 0.0738	<div><div></div></div> 0.6606	<div><div></div></div> 0.9494	<div><div></div></div> 0.0506
6	595	<div><div></div></div> 2.7328	<div><div></div></div> 13.4151	<div><div></div></div> 0.9065	<div><div></div></div> 0.6005	<div><div></div></div> 0.0606	<div><div></div></div> 0.0222	<div><div></div></div> 0.9320	<div><div></div></div> 0.0680
7	1,565	<div><div></div></div> 23.1048	<div><div></div></div> 1.4383	<div><div></div></div> 0.5295	<div><div></div></div> 0.4411	<div><div></div></div> 0.2068	<div><div></div></div> 0.2257	<div><div></div></div> 0.7140	<div><div></div></div> 0.2860

AVG.
by cluster group

CENTROID_ID	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_EVENING	PROB_BAS_M	PROB_BAS_S	PROB_WEEKEND	PROB_WEEKDAY
3	1.2	14.83	0.48	0.31	0.33	0.48	0.96	0.04
2	17.25	2.84	0.19	0.72	0.33	0.37	0.31	0.69
7	22.73	1.52	0.53	0.22	0.44	0.21	0.28	0.72
4	1.07	15.22	0.44	0.32	0.01	0.97	0.02	0.98
1	14.26	4.14	0.55	0.14	0.13	0.77	0.3	0.7
5	1.89	14.63	0.07	0.66	0.62	0.07	0.05	0.95
6	2.7	13.44	0.91	0.02	0.59	0.06	0.07	0.93

3rd Model Result

- DBI and MSD index decreased from 2nd model
- Try to decreased no. of K in next model

BigQuery : K-Mean Clustering

Metrics

Davies–Bouldin index	1.5321
Mean squared distance	3.6578

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_M, PROB_BAS_S, PROB_EVEN...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_M	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	1,874	<div><div></div></div> 21.8431	<div><div></div></div> 1.7006	<div><div></div></div> 0.5588	<div><div></div></div> 0.3723	<div><div></div></div> 0.3109	<div><div></div></div> 0.1855	<div><div></div></div> 0.7153	<div><div></div></div> 0.2847
2	1,037	<div><div></div></div> 17.7541	<div><div></div></div> 2.7319	<div><div></div></div> 0.2171	<div><div></div></div> 0.3192	<div><div></div></div> 0.4143	<div><div></div></div> 0.6423	<div><div></div></div> 0.6841	<div><div></div></div> 0.3159
3	773	<div><div></div></div> 1.2704	<div><div></div></div> 14.7477	<div><div></div></div> 0.4830	<div><div></div></div> 0.3273	<div><div></div></div> 0.4830	<div><div></div></div> 0.3074	<div><div></div></div> 0.0452	<div><div></div></div> 0.9548
4	1,092	<div><div></div></div> 2.2601	<div><div></div></div> 13.7949	<div><div></div></div> 0.4558	<div><div></div></div> 0.0141	<div><div></div></div> 0.9649	<div><div></div></div> 0.2836	<div><div></div></div> 0.9440	<div><div></div></div> 0.0560
5	656	<div><div></div></div> 1.9756	<div><div></div></div> 14.6265	<div><div></div></div> 0.0642	<div><div></div></div> 0.6101	<div><div></div></div> 0.0765	<div><div></div></div> 0.6624	<div><div></div></div> 0.9550	<div><div></div></div> 0.0450
6	668	<div><div></div></div> 3.3234	<div><div></div></div> 12.7919	<div><div></div></div> 0.8872	<div><div></div></div> 0.6025	<div><div></div></div> 0.0725	<div><div></div></div> 0.0353	<div><div></div></div> 0.8993	<div><div></div></div> 0.1017

AVG.
by cluster group

CENTROID_ID	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_EVENING	PROB_BAS_M	PROB_BAS_S	PROB_WEEKEND	PROB_WEEKDAY
1	21.73	1.74	0.56	0.19	0.38	0.31	0.29	0.71
2	17.68	2.77	0.22	0.65	0.31	0.42	0.32	0.68
3	1.26	14.76	0.48	0.31	0.33	0.48	0.95	0.05
4	2.29	13.81	0.45	0.28	0.01	0.97	0.06	0.94
5	2.18	14.44	0.07	0.66	0.62	0.08	0.05	0.95
6	3.11	12.94	0.9	0.03	0.6	0.07	0.08	0.92

4th Model Result

- DBI and MSD index slightly increased from 3rd model but acceptable.
- Try to decreased no. of K in next model

BigQuery : K-Mean Clustering

Metrics

Davies-Bouldin index	1.6476
Mean squared distance	4.0413

Numeric features

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

Selected Features

AGE, LAST_VISIT, PROB_AFTERNOON, PROB_BAS_M, PROB_BAS_S, PROB_EVEN...

Centroid Id	Count	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_BAS_M	PROB_BAS_S	PROB_EVENING	PROB_WEEKDAY	PROB_WEEKEND
1	1,778	<div><div></div></div> 19.5832	<div><div></div></div> 2.4454	<div><div></div></div> 0.6297	<div><div></div></div> 0.4068	<div><div></div></div> 0.2554	<div><div></div></div> 0.1423	<div><div></div></div> 0.7292	<div><div></div></div> 0.2708
2	1,361	<div><div></div></div> 19.2572	<div><div></div></div> 2.3784	<div><div></div></div> 0.2497	<div><div></div></div> 0.3161	<div><div></div></div> 0.4160	<div><div></div></div> 0.5773	<div><div></div></div> 0.6900	<div><div></div></div> 0.3100
3	795	<div><div></div></div> 1.2528	<div><div></div></div> 14.8201	<div><div></div></div> 0.4882	<div><div></div></div> 0.3274	<div><div></div></div> 0.4794	<div><div></div></div> 0.3038	<div><div></div></div> 0.0574	<div><div></div></div> 0.9426
4	1,122	<div><div></div></div> 2.4688	<div><div></div></div> 13.6738	<div><div></div></div> 0.4551	<div><div></div></div> 0.0171	<div><div></div></div> 0.9608	<div><div></div></div> 0.2869	<div><div></div></div> 0.9421	<div><div></div></div> 0.0579
5	1,044	<div><div></div></div> 1.4531	<div><div></div></div> 15.2443	<div><div></div></div> 0.4024	<div><div></div></div> 0.6389	<div><div></div></div> 0.0560	<div><div></div></div> 0.3994	<div><div></div></div> 0.9617	<div><div></div></div> 0.0383

AVG.
by cluster group

CENTROID_ID	AGE	LAST_VISIT	PROB_AFTERNOON	PROB_EVENING	PROB_BAS_M	PROB_BAS_S	PROB_WEEKEND	PROB_WEEKDAY
1	20.24	2.26	0.62	0.15	0.4	0.26	0.28	0.72
2	18.75	2.52	0.24	0.59	0.32	0.42	0.31	0.69
3	1.27	14.81	0.49	0.3	0.33	0.48	0.94	0.06
4	2.45	13.68	0.45	0.29	0.02	0.96	0.06	0.94
5	1.52	15.0	0.43	0.37	0.64	0.05	0.04	0.96

5th Model Result

- DBI and MSD index slightly increased from 4th model but acceptable.