1. Load dataset to Big Query

Supermarket Data

SCHEMA | DETAILS

Field name	Type	Mode	Policy Tags	Description
SHOP_WEEK	INTEGER	NULLABLE		
SHOP_DATE	INTEGER	NULLABLE		
SHOP_WEEKDAY	INTEGER	NULLABLE		
SHOP_HOUR	INTEGER	NULLABLE		
QUANTITY	INTEGER	NULLABLE		
SPEND	FLOAT	NULLABLE		
PROD_CODE	STRING	NULLABLE		
PROD_CODE_10	STRING	NULLABLE		
PROD_CODE_20	STRING	NULLABLE		
PROD_CODE_30	STRING	NULLABLE		
PROD_CODE_40	STRING	NULLABLE		
CUST_CODE	STRING	NULLABLE		
CUST_PRICE_SENSITIVITY	STRING	NULLABLE		
CUST_LIFESTAGE	STRING	NULLABLE		
BASKET_ID	FLOAT	NULLABLE		
BASKET_SIZE	STRING	NULLABLE		
BASKET_PRICE_SENSITIVITY	STRING	NULLABLE		
BASKET_TYPE	STRING	NULLABLE		
BASKET_DOMINANT_MISSION	STRING	NULLABLE		
STORE_CODE	STRING	NULLABLE		
STORE_FORMAT	STRING	NULLABLE		
STORE_REGION	STRING	NULLABLE		

2. Pre-process the dataset to group by SHOP MONTH and CUST_CODE and Created the new SUMMARY Table.

```
SELECT CONCAT(CAST(a.SHOP_MONTH AS STRING), '01'), a.CUST_CODE, a.SHOW,
    (SELECT COUNT(b.CUST_CODE) FROM `rugged-reality-335113.sample_dataset12345.Supermarket Data` b
        WHERE a.CUST_CODE = b.CUST_CODE AND
            DATE_DIFF(
                PARSE_DATE('%Y%m%d', CONCAT(a.SHOP_MONTH,'01')),
                PARSE_DATE('%Y%m%d', CAST(b.SHOP_DATE AS STRING)), DAY) > 0
    ) AS PAST,
    (SELECT COUNT(c.CUST_CODE) FROM `rugged-reality-335113.sample_dataset12345.Supermarket Data` c
        WHERE a.CUST_CODE = c.CUST_CODE AND
        DATE_DIFF(
            DATE_SUB(PARSE_DATE('%Y%m%d', CONCAT(a.SHOP_MONTH, '01')), INTERVAL 1 MONTH), PARSE_DATE("%Y%m%d", CAST(c.SHOP_DATE AS STRING)),
            DAY) > 0
    ) AS PAST_EXCLUDE,
    FROM (SELECT SUBSTRING(CAST(SHOP_DATE AS STRING), 0, 6) AS SHOP_MONTH, CUST_CODE, COUNT(DISTINCT(CUST_CODE)) AS SHOW FROM `rugged-reality-335113.
    GROUP BY SHOP_MONTH, CUST_CODE HAVING CUST_CODE IS NOT NULL)
    a
```

Result :

Row	f0_	CUST_CODE	SHOW	PAST	PAST_EXCLUDE
1	20060401	CUST0000293743	1	0	0
2	20060401	CUST0000246027	1	0	0
3	20060401	CUST0000167145	1	0	0
4	20060401	CUST0000136043	1	0	0
5	20060401	CUST0000140993	1	0	0
6	20060401	CUST0000268287	1	0	0

3. Classify the customer types into New customer, Repeated customer, Reactivated customer and Churned customers and quantify the numbers of each type of customers with SUMMARY TABLE.

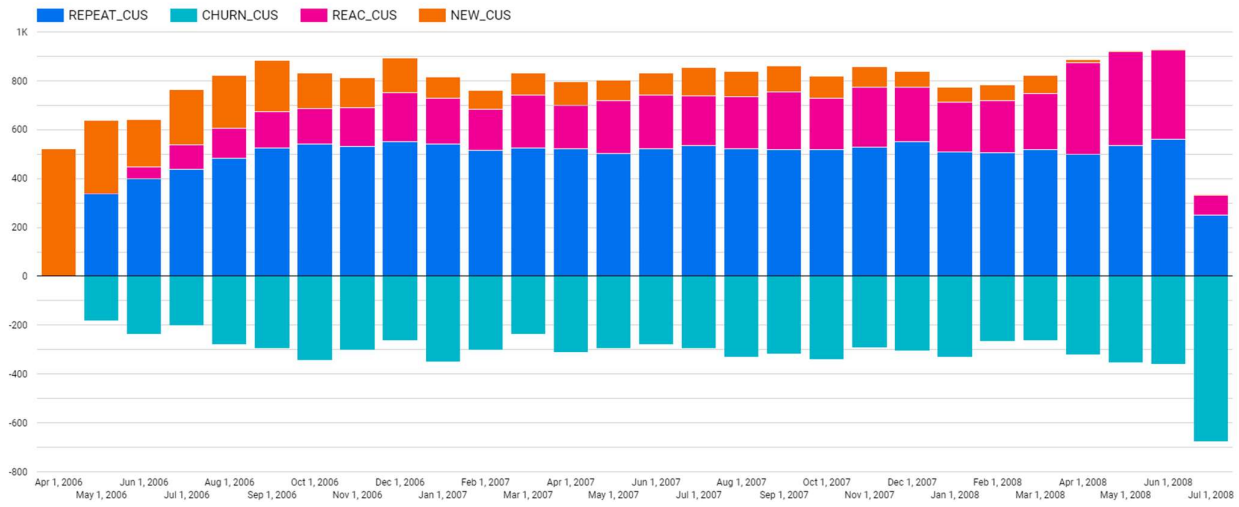
```
SELECT a.SHOP_MONTH, a.NEW_CUS, a.REPEAT_CUS, a.REAC_CUS, - ((b.NEW_CUS + b.REPEAT_CUS + b.REAC_CUS) - a.REPEAT_CUS) as CHURN_CUS
FROM (SELECT f0_ AS SHOP_MONTH,
    (SUM(IF(PAST > 0,0,1))) AS NEW_CUS,
    (SUM(IF(PAST > PAST_EXCLUDE,1,0))) AS REPEAT_CUS,
    (SUM(IF((PAST != 0 AND PAST = PAST_EXCLUDE),1,0))) AS REAC_CUS FROM `rugged-reality-335113.sample_dataset12345.SUMMARY` GROUP BY SHOP_MONTH) a
    LEFT JOIN (SELECT FORMAT_DATE('%Y%m%d', DATE_ADD(PARSE_DATE('%Y%m%d', f0_), INTERVAL 1 MONTH)) AS SHOP_MONTH,
        (SUM(IF(PAST > 0,0,1))) AS NEW_CUS,
        (SUM(IF(PAST > PAST_EXCLUDE,1,0))) AS REPEAT_CUS,
        (SUM(IF((PAST != 0 AND PAST = PAST_EXCLUDE),1,0))) AS REAC_CUS FROM `rugged-reality-335113.sample_dataset12345.SUMMARY` GROUP BY SHOP_MONTH) b
        ON b.SHOP_MONTH = a.SHOP_MONTH

```

Result :

Row	SHOP_MONTH	NEW_CUS	REPEAT_CUS	REAC_CUS	CHURN_CUS
1	20060401	522	0	0	*null*
2	20060501	299	338	0	-184
3	20060601	191	398	51	-239
4	20060701	225	437	102	-203
5	20060801	217	483	122	-281
6	20060901	213	526	146	-296

4. Visualize the data through Google Data Studio as following.

Result :



end