Customer shopping behaviors exploration

a.

- The number of tore shopping trips recorded in the database

```
Select count(hh_id)
from trips;
```

- The number of households appearing in the database

```
Select Count(Distinct(hh_id))
from households;
```

- The number of stores of different retailers in the database

```
SELECT Count(distinct(TC_retailer_code))
FROM db_consumer_panel2.trips;
```

- The number of Different products that are recorded
  - i. Products per category and products per module

```
Select group_at_prod_id,Count(Prod_id)
from products group by
group_at_prod_id;

Select module_at_prod_id,Count(Prod_id)
from products
group by module_at_prod_id;
```

  - ii. Plot the distribution of products and modules per department

```
Select department_at_prod_id,group_at_prod_id,Count(Prod_id)
from products
group by department_at_prod_id,group_at_prod_id;

Select department_at_prod_id,module_at_prod_id,Count(Prod_id)
from products
group by department_at_prod_id,module_at_prod_id;
```

b. Aggregate the data at the household-monthly level

- households do not shop at least once on a 3 month periods

```
drop table if exists t1;
create table t1 SELECT hh_id, TC_date , ROW_NUMBER() OVER (ORDER BY TC_date) as ID FROM trips
where year(TC_date)=2004 order by hh_id, TC_date;

drop table if exists t2;
create table t2
```

```
select *, ID+1 as ID_2 from t1 order by hh_id, TC_date;

drop table if exists t3;
create table t3
select
A.hh_id as hh_id_0 ,
A.TC_date as TC_date_0 ,
B.hh_id as hh_id_1 ,
B.TC_date as TC_date_1 ,
datediff(B.TC_date ,A.TC_date )/30 as TIME_WINDOW
from t2 as A inner join t1 as B on A.ID_2 = B.ID;
select sum(TIME_WINDOW)
from (select count(TIME_WINDOW) as TIME_WINDOW from t3 where TIME_WINDOW>3
union
select count(distinct(hh_id)) from t2) a;
```

- Among the households who shop at least once a month, found what percent of them concentrate at least 80% of their grocery expenditure (on average) on single retailer and also among 2 retailers.

```
# hh_id's with shopping every month
CREATE TEMPORARY TABLE HH_once_month as (SELECT *
FROM
(Select hh_id,Count(hh_id) AS Months from
(Select hh_id,month(TC_date) from trips
group by hh_id,month(TC_date)) A group by hh_id) B
WHERE months>11);

#Left joining trips data for those who shop once a month Create temporary table
HH_once_month_all as
(select A.* from trips A
left join HH_once_month B on A.hh_id=B.hh_id
where B.hh_id is not null);

#Calculating avg spent per household DROP TABLE HH_AVG_SPENT;
CREATE TABLE HH_AVG_SPENT AS
select hh_id,sum(tc_total_spent)/12 AS AVG_SPENT from HH_once_month_all
group by hh_id;
#having hh_id='9001556';

#Calculating average spent per retailer DROP TABLE HH_AVG_SPENT_RETAILER;
CREATE TABLE HH_AVG_SPENT_RETAILER AS
```

```
select hh_id,tc_retailer_code,sum(tc_total_spent)/12 AS AVG_SPENT_PER_RETAILER from
HH_once_month_all
group by hh_id,tc_retailer_code; #having hh_id='9001556';

#retailers percentage share
create table HH_PERCENTAGE_OF_SPENT AS SELECT
A.hh_id,A.tc_retailer_code,A.AVG_SPENT_PER_RETAILER,B.AVG_SPENT,(A.AVG_SPENT_
PER_RETAILER/B.AVG_SPENT)*100 AS 'PERCENTAGE_OF_SPENT'
FROM HH_AVG_SPENT_RETAILER A LEFT JOIN HH_AVG_SPENT B
ON A.HH_ID=B.HH_ID;

#one retailers with more than 80% share select *
from HH_PERCENTAGE_OF_SPENT
where 'PERCENTAGE_of_spent'>=80;

#two retailers with more than 80% share CREATE TABLE HH_2_RETAILERS_80 AS
select A.HH_ID,A.TC_RETAILER_CODE,A.AVG_SPENT_PER_RETAILER AS
RETAILER_1,B.AVG_SPENT_PER_RETAILER AS
RETAILER_2,(A.AVG_SPENT_PER_RETAILER+B.AVG_SPENT_PER_RETAILER),
((A.AVG_SPENT_PER_RETAILER+B.AVG_SPENT_PER_RETAILER)/A.AVG_SPENT)*100 AS
PERCENTAGE_SPENT
from HH_PERCENTAGE_OF_SPENT A LEFT JOIN HH_PERCENTAGE_OF_SPENT B ON A.HH_ID=B.HH_ID
AND A.TC_RETAILER_CODE<>B.TC_RETAILER_CODE WHERE
((A.AVG_SPENT_PER_RETAILER+B.AVG_SPENT_PER_RETAILER)/A.AVG_SPENT)*100>80;
```

i. Are their demographics remarkably different? Are these people richer? Poorer?

```
SELECT A.HH_ID,B.HH_INCOME
FROM HH_2_RETAILERS_80 A
LEFT JOIN HOUSEHOLDS B ON
A.HH_ID=B.HH_ID;
```

ii. What is the retailer that has more loyalists?

```
SELECT TC_RETAILER_CODE,COUNT(HH_ID) FROM TRIPS
GROUP BY TC_RETAILER_CODE
ORDER BY COUNT(HH_ID) DESC
LIMIT 1;
```

iii. Where do they live? Plot the distribution by state.

```
SELECT TC_RETAILER_CODE,COUNT(HH_ID) FROM TRIPS
  GROUP  BY  TC_RETAILER_CODE
  ORDER  BY  COUNT(HH_ID)  DESC
  LIMIT 1;
```