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

1. 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;

1. 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;

1. 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;

1. 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;

1. 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;