# 

# **Dunnhumby - The Complete Journey**

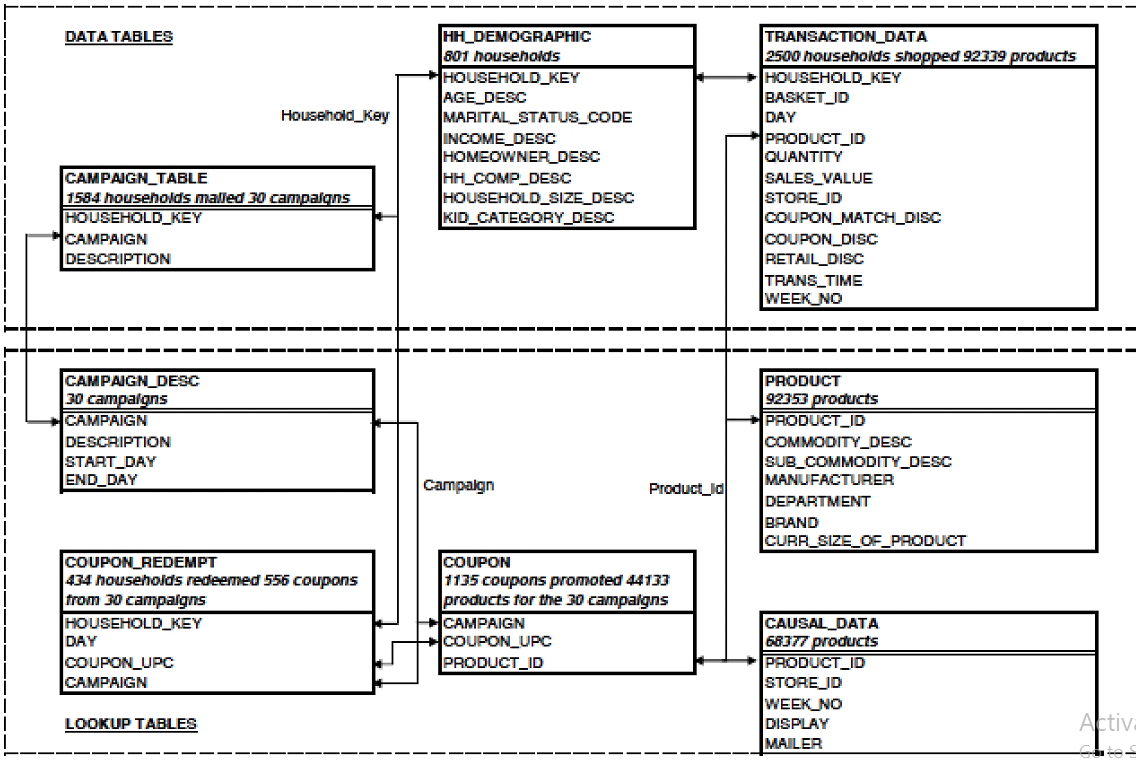
**Context**

Global leader in Customer data science and analytics, dunnhumby has experts in working with brands, grocery retail, retail pharmacy, and retailer financial services.

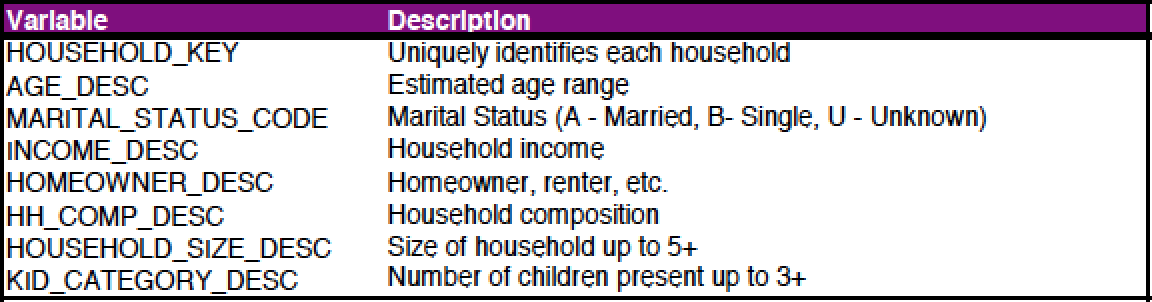
With deep heritage and expertise in retail — one of the world's most competitive markets, with a deluge of multi-dimensional data — dunnhumby today enables businesses all over the world, across industries, to be Customer First.

This business case has

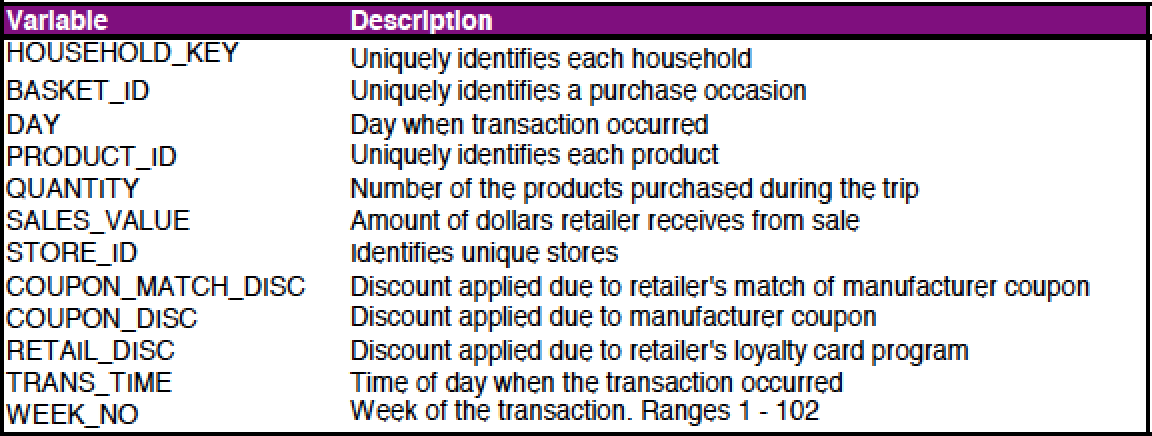
* Household level transactions over two years from a group of 2,500 households who are frequent shoppers at a retailer
* All of a household’s purchases within the store, not just those from a limited number of categories
* Demographics and direct marketing contact history for select households

****

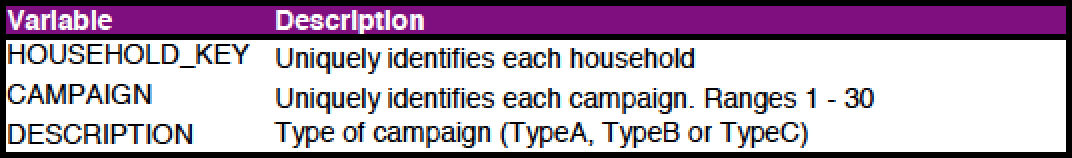
* **hh\_demographic**

****

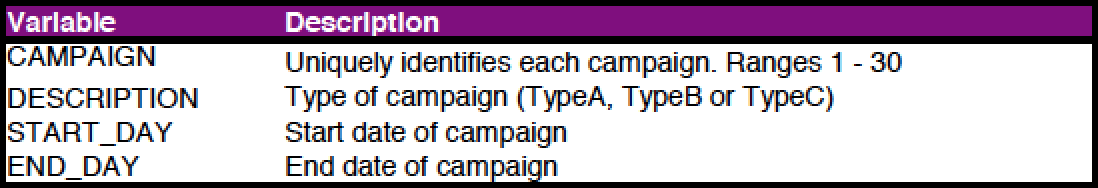
* **transaction\_data**

****

* **campaign\_table**

****

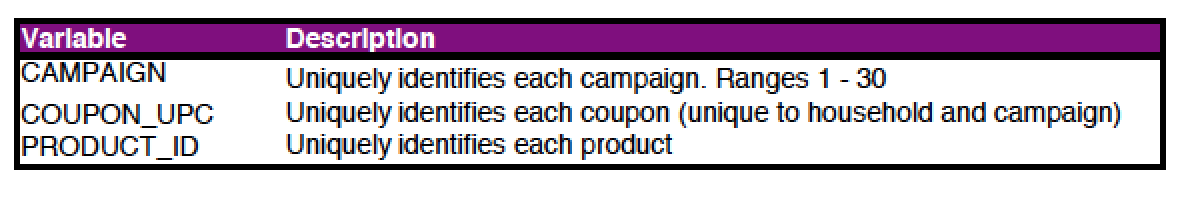
* **capaign\_desc**

****

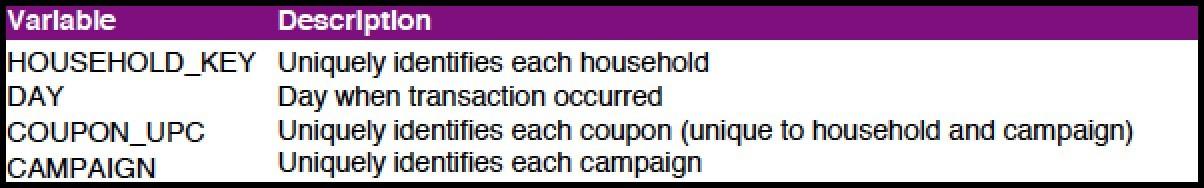
* **product**

****

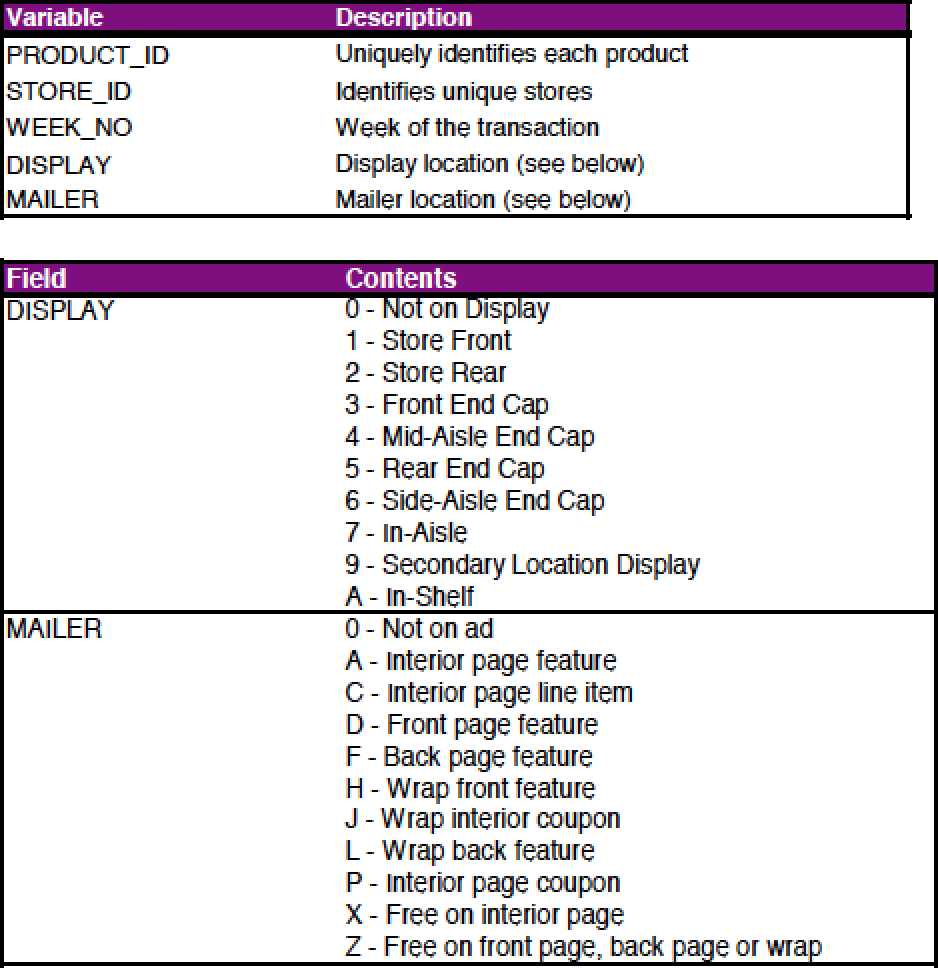
* **coupon**

****

* **coupon\_redempt**

****

* **Causal\_data - event info and some other details**

****

**Top questions:**

1. Find the number of orders that are small, medium or large order value(small:0-5$, medium:5-10$, large:10+)
2. Find top 3 stores with highest foot traffic for each week (Foot traffic: number of customers transacting )
3. Create a basic customer profiling with first, last visit, number of visits, average money spent per visit and total money spent order by highest avg money
4. Do a single customer analysis selecting most spending customer for whom we have demographic information(because not all customers in transaction data are present in demographic table)(show the demographic as well as profiling data)
5. Find products(product table :SUB\_COMMODITY\_DESC) which are most frequently bought together
6. **Find out on which weeks does each household shop and find their cumulative spending over time(sum of all previous) (uses sum over partition)**
7. **Find the weekly change in Revenue Per Account (RPA) (spending by each customer compared to last week)(use lag function)**
8. **Find number of returning customers and percent of returning customers for all week**
9. **Quarterly analysis: sales comparison: total sale amount (create a new quarter column using case where,12 weeks(3 months)=1 quarter)**

**(Use cte tables)**

1. **How are the sales for individual stores changing over the quarters**
2. **Customer churn analysis for each quarter (churned customers : that never shop after that particular quarter)**
3. **Find the retained customers for each quarter(retained :Households who were there in previous quarters and are there in the current quarter)**
4. Calculate Customer lifetime value(CLV) for different age group

Average purchase value — the value of all customer purchases over a particular time frame , divided by the number of purchases in that period

Average purchase frequency — divide the number of purchases in that same time period by the number of individual customers who made a transaction over the same period

Customer value — the average purchase frequency multiplied by the average purchase value

Average customer lifespan — the average length of time a customer continues buying from you

CLV = customer value X average customer lifespan

**Answers:**

1. **Find the number of orders that have small, medium or large order value (small:0-10$, medium:10-20$, large:20+)**

**select basket\_size , count(\*) as num\_orders**

**from (**

**select case**

**when sales\_value between 0 and 10 then 'small'**

**when sales\_value between 10 and 20 then 'medium'**

**when sales\_value >20 then 'large'**

**end as basket\_size**

**from(**

**select sum(SALES\_VALUE) as sales\_value from `dunnhumbysql.complete.transaction\_data`**

**group by BASKET\_ID))**

**group by basket\_size**

****

**2. Find week over week top 3 stores with highest foot traffic (Foot traffic: number of households transacting )**

**select \* from(**

**select STORE\_ID, WEEK\_NO, count(household\_key) as foot\_traffic,**

**RANK() OVER(PARTITION BY WEEK\_NO order by count(household\_key) desc ) as rnk**

**from `dunnhumbysql.complete.transaction\_data`**

**group by STORE\_ID,WEEK\_NO)**

**where rnk<4**

**order by WEEK\_NO, foot\_traffic desc**

* ****

**Variation:**

**Find week over week top 3 stores with most number of distinct households transacting**

**with base as (select**

**STORE\_ID, WEEK\_NO, count(distinct household\_key) as footfall**

**from `dunnhumby.transaction\_data`**

**group by 1,2**

**),**

**base\_2 as (**

**select \*, dense\_rank() over (partition by base.week\_no order by footfall desc) as ranker**

**from base**

**order by week\_no asc**

**)**

**select week\_no, store\_id, footfall, ranker from base\_2**

**where ranker <=3**

**order by week\_no, ranker asc**

****

**3. Create a basic customer profiling with first, last visit, number of visits, average money spent per visit and total money spent order by highest avg money**

**select household\_key, min(WEEK\_NO) as first\_visit,**

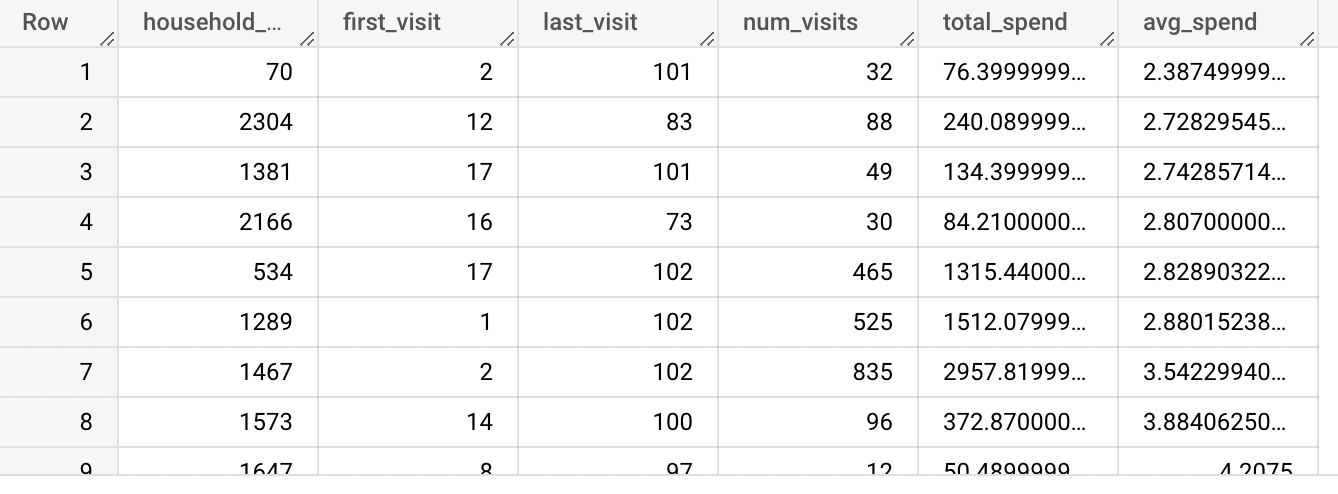
**max(WEEK\_NO) last\_visit, count(distinct(BASKET\_ID)) as num\_visits,**

**sum(SALES\_VALUE) as total\_spend, (sum(SALES\_VALUE)/count(distinct(BASKET\_ID))) as avg\_spend**

**from `dunnhumbysql.complete.transaction\_data`**

**group by household\_key**

**order by avg\_spend**

****

**4. Do a customer analysis for the most spending customer for whom we have demographic information**

**with cte as(**

**select t.household\_key, sum(SALES\_VALUE) as total\_spendfrom `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on d.household\_key=t.household\_key**

**group by t.household\_key**

**order by total\_spend desc**

**limit 1**

**)**

**select cte.\*, d.\* from cte**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on cte.household\_key=d.household\_key**

****

**5. Find products(product table:SUB\_COMMODITY\_DESC) which are most frequently bought together**

**with cte as (**

**SELECT \***

**FROM `dunnhumbysql.complete.product` p**

**join `dunnhumbysql.complete.transaction\_data` t**

**on p.PRODUCT\_ID=t.PRODUCT\_ID**

**)**

**select t.SUB\_COMMODITY\_DESC as item\_1, t2.SUB\_COMMODITY\_DESC as item\_2,**

**count(distinct t.BASKET\_ID) as num\_orders**

**from cte t**

**inner join cte t2**

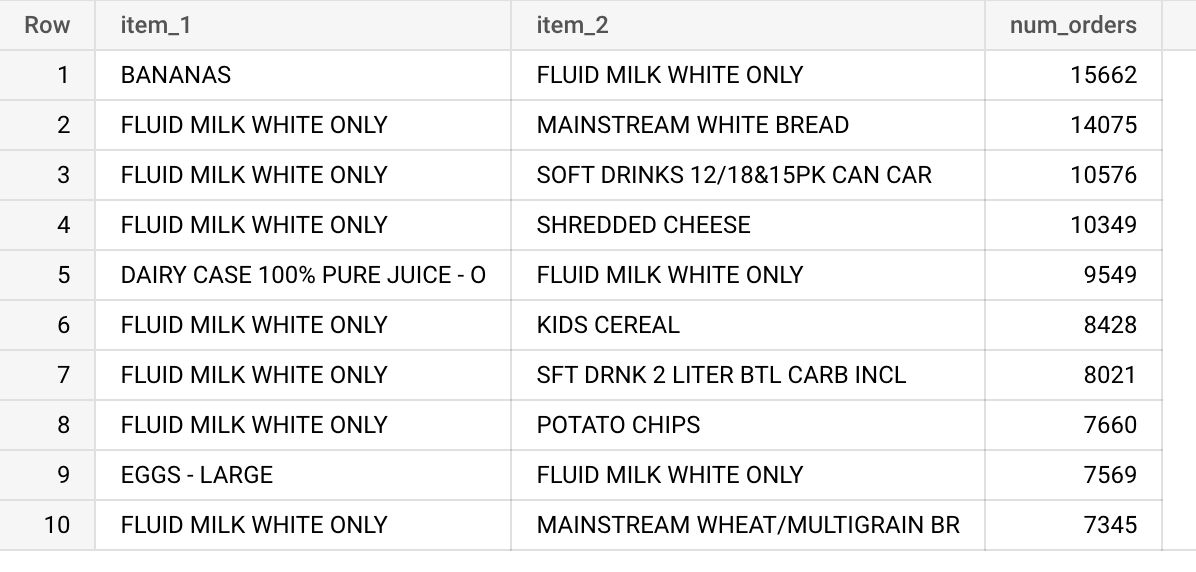
**on t.BASKET\_ID=t2.BASKET\_ID**

**and t.SUB\_COMMODITY\_DESC<t2.SUB\_COMMODITY\_DESC**

**group by t.SUB\_COMMODITY\_DESC, t2.SUB\_COMMODITY\_DESC**

**order by num\_orders desc**

**limit 10**

****

**6. Find out on which weeks does each household shop and find their cumulative spending over time**

**with cte as(**

**select WEEK\_NO , household\_key , sum(SALES\_VALUE) as sales**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by WEEK\_NO, household\_key**

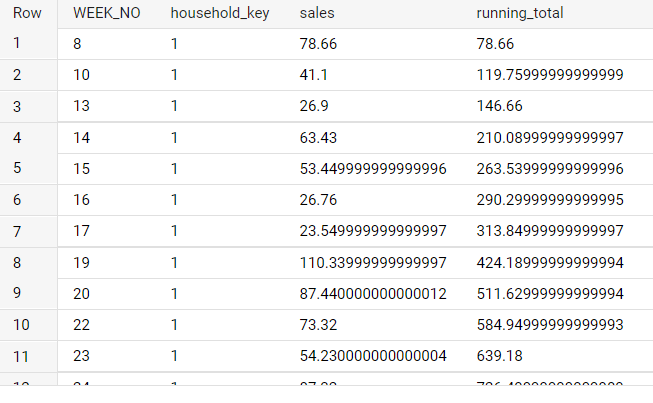
**)**

**SELECT**

**\*,**

**SUM(sales) OVER (PARTITION BY household\_key ORDER BY week\_no) AS running\_total**

**from cte**

****

**7. Find the weekly change in Revenue Per Account (RPA) (spending by each customer compared to last week)(use lag function)**

**with cte as(**

**select WEEK\_NO , household\_key , sum(SALES\_VALUE) as sales**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by WEEK\_NO, household\_key**

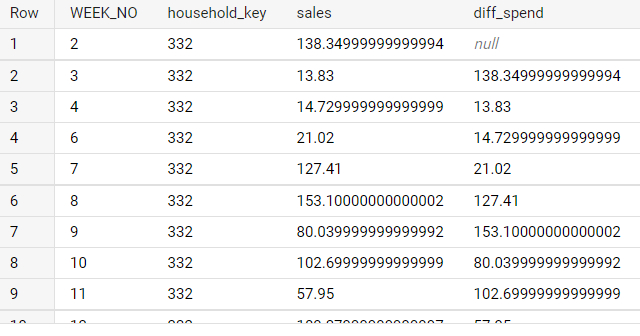
**)**

**SELECT**

**\*,**

**lag(sales)OVER (PARTITION BY household\_key ORDER BY week\_no) as diff\_spend,**

**from cte**

****

**8. Find number of returning customers and percent of returning customers for all week**

**with cte as(**

**select b.week\_no, a.household\_key,**

**CASE when min(a.week\_no)<b.week\_no then 1 else 0**

**end as decider**

**from `dunnhumbysql.complete.transaction\_data` a**

**left join `dunnhumbysql.complete.transaction\_data` b**

**on a.household\_key=b.household\_key**

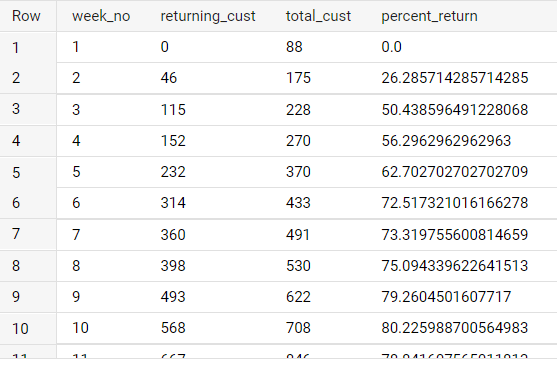
**group by b.week\_no, a.household\_key**

**) select week\_no,sum(decider) as returning\_cust, count(decider) as total\_cust, (sum(decider)/count(decider))\*100 as percent\_return**

**from cte**

**group by 1**

**order by 1**

****

**9. Quarterly analysis: sales comparison over quarters**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

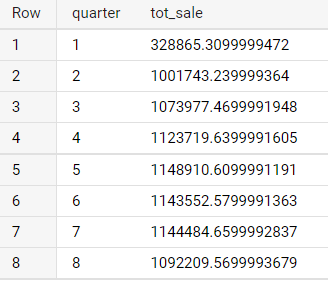
**end as quarter**

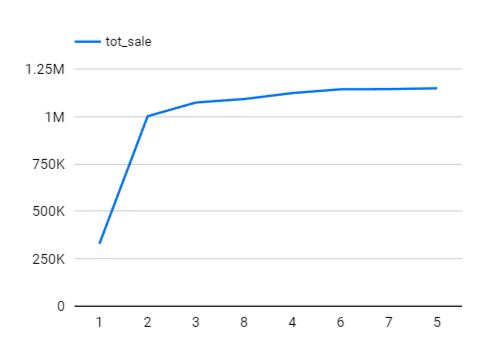
**from `dunnhumbysql.complete.transaction\_data`)**

**select quarter, sum(sales\_value) as tot\_sale**

**from cte**

**group by quarter**

****

****

**10. How are the sales for individual stores changing over the quarters**

**select STORE\_ID ,sum(sales) as sales, case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

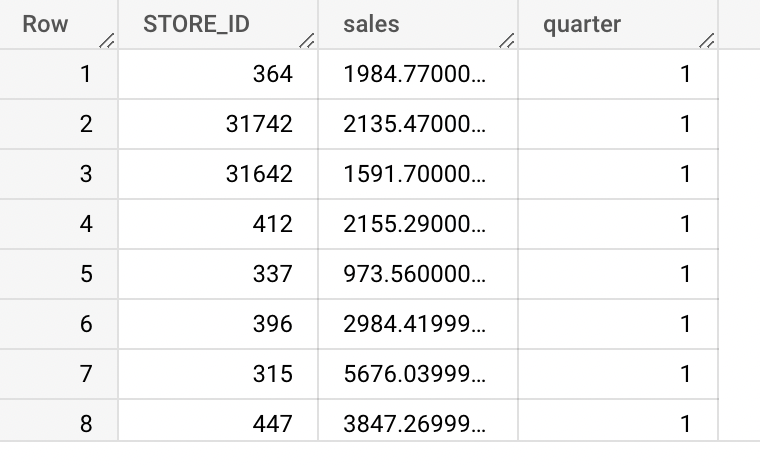
**when week\_no between 91 and 102 then 8**

**end as quarter**

**from( select week\_no, STORE\_ID, sum(SALES\_VALUE) as sales,**

**from `dunnhumbysql.complete.transaction\_data` group by STORE\_ID, WEEK\_NO)**

**group by quarter, store\_id**

****

**11. Customer churn analysis for each quarter**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

**end as quarter**

**from `dunnhumbysql.complete.transaction\_data`)**

**select a.quarter, count(distinct(a.household\_key)) as chrned**

**from cte a**

**FULL OUTER JOIN cte b**

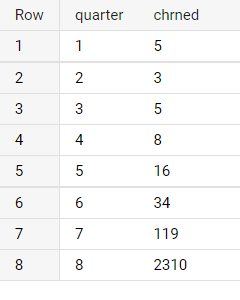
**on a.household\_key=b.household\_key**

**and a.quarter<b.quarter**

**where b.household\_key is NUll**

**group by a.quarter**

**order by quarter**

****

**12. Find the retained customers for each quarter(Households who were there last quarters and are there in the current quarter**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

**end as quarter**

**from `dunnhumbysql.complete.transaction\_data`)**

**select a.quarter, count(distinct(a.household\_key)) as retained**

**from cte a**

**left join cte b**

**on a.household\_key=b.household\_key and a.quarter>b.quarter**

**group by a.quarter**

****

**13. Calculate Customer lifetime value(CLV) for different age group**

* + **Average purchase value —** the value of all customer purchases over a particular time frame , divided by the number of purchases in that period
  + **Average purchase frequency —** divide the number of purchases in that same time period by the number of individual customers who made a transaction over the same period
  + **Customer value —** the average purchase frequency multiplied by the average purchase value
  + **Average customer lifespan —** the average length of time a customer continues buying from you
  + **CLV = customer value X average customer lifespan**

**select AGE\_DESC, (avg\_purch\_val\*avg\_purch\_freq\*avg\_cust\_lifespan) as clv**

**from(**

**with cte as (**

**select household\_key, (max(WEEK\_NO)- min (WEEK\_NO)) as cust\_duration**

**from `dunnhumbysql.complete.transaction\_data`**

**group by household\_key**

**)**

**select AGE\_DESC, sum(SALES\_VALUE)/count(distinct(BASKET\_ID)) as avg\_purch\_val,**

**count(distinct(BASKET\_ID))/count(distinct(d.household\_key)) as avg\_purch\_freq,**

**(sum(cte.cust\_duration)/count(1)) as avg\_cust\_lifespan,**

**from `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on t.household\_key=d.household\_key**

**join cte**

**on cte.household\_key=d.household\_key**

**group by AGE\_DESC**

**)**

****

**Questions**

1. **Exploratory queries**
   1. Find out which age group is the most active shopper( join hh\_demographic and transaction\_data)
   2. Which week had the best sales
   3. What is the average basket size for shoppers (Divide it in small, medium, large)
   4. Find foot traffic for each store per week. (Foot traffic: number of customers transacting)
   5. Top5 spending customers (households) with sales value in integer
2. **Customer profiling**
   1. Create a basic customer profiling with first last visit and total money spent for all customers
   2. Do a single customer analysis selecting most spending customer for whom we have demographic information(because not all customers in transaction data are present in demographic table)
   3. What products does the customer buy most
   4. Which promotional campaigns were they a part of(campaign\_table)
3. **Product analysis**
   1. Find the most selling product
   2. When did the product sell the most and where
   3. Where was the product placed in store and featured in ad for that particular store and week
   4. Was it a part of some campaigns
   5. How many household did actually redeem coupons for this product in each campaign
   6. Which products were the best seller(top 3) for each week and what quantity did they sell
4. **Advance analysis and queries**
   1. Find out on which weeks does each household shop and find their cumulative spending over time(sum of all previous) (uses sum over partition)
   2. Find the trend in spending for each customer (spending compared to last purchase)(use lag function)
   3. Find number of returning customers and percent of returning customers for all week
   4. Quarterly analysis: sales comparison(create a new quarter column using case when,12 weeks(3 months)=1 quarter)

(Use cte tables)

* 1. Are the customers spending more or less over time ( group in 25 week segments)
  2. Customer churn analysis for each quarter
  3. Find the retained customers for each quarter

**Question and queries**

1. **Exploratory queries**

* **1. A. Find out which age group is the most active shopper**

**SELECT distinct(AGE\_DESC), count(1) as num\_cust\_trans**

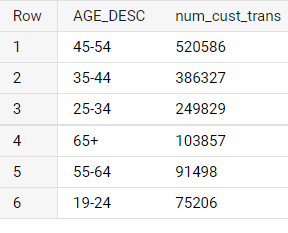
**FROM `dunnhumbysql.complete.hh\_demographic`h**

**join `dunnhumbysql.complete.transaction\_data`t**

**on t.household\_key=h.household\_key**

**group by AGE\_DESC**

**order by num\_cust\_trans desc**

****

* **1.b. Find out which income group shops the most**

**SELECT distinct(INCOME\_DESC), count(1) as num\_cust\_trans**

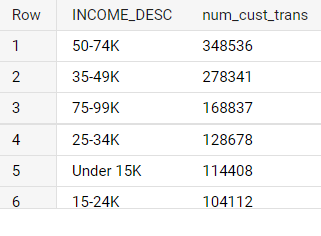
**FROM `dunnhumbysql.complete.hh\_demographic`h**

**join `dunnhumbysql.complete.transaction\_data`t**

**on t.household\_key=h.household\_key**

**group by INCOME\_DESC**

**order by num\_cust\_trans desc**

****

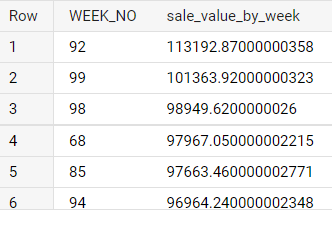
* **1.c. Which week had the best sales**

**SELECT WEEK\_NO, sum(SALES\_VALUE) as sale\_value\_by\_week**

**from `dunnhumbysql.complete.transaction\_data`**

**group by WEEK\_NO**

**order by sale\_value\_by\_week desc**

****

* **1.f. Top5 spending households with sales value in integer**

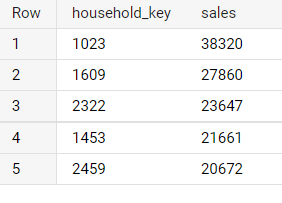
**select household\_key,cast( sum(SALES\_VALUE) as int) as sales**

**from `dunnhumbysql.complete.transaction\_data`**

**group by household\_key**

**order by sales desc**

**Limit 5**

****

1. **Customer analysis**

* **2.a Create a basic customer profiling with first, last visit, number of visits, average money spent per visit and total money spent**

**select household\_key, min(WEEK\_NO) as first\_visit,**

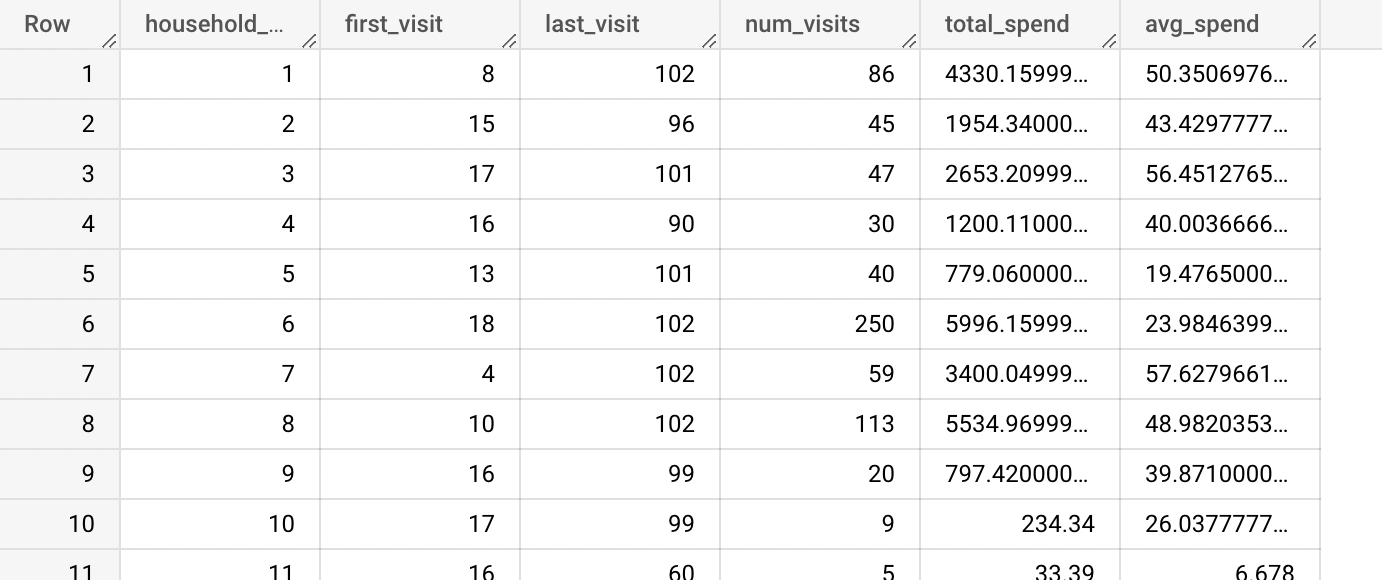
**max(WEEK\_NO) last\_visit, count(distinct(BASKET\_ID)) as num\_visits,**

**sum(SALES\_VALUE) as total\_spend, (sum(SALES\_VALUE)/count(distinct(BASKET\_ID))) as avg\_spend**

**from `dunnhumbysql.complete.transaction\_data`**

**group by household\_key**

**order by household\_key**

****

* **2.b. Do a customer analysis for the most spending customer for whom we have demographic information**

**with cte as(**

**select t.household\_key, sum(SALES\_VALUE) as total\_spend,**

**min(WEEK\_NO) as first\_visit,max(WEEK\_NO) last\_visit**

**from `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on d.household\_key=t.household\_key**

**group by t.household\_key**

**order by total\_spend desc**

**limit 1**

**)**

**select cte.\*, d.\* from cte**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on cte.household\_key=d.household\_key**

****

* **2.c. Get the demographic information for that**

**select\***

**from `dunnhumbysql.complete.hh\_demographic`**

**where household\_key =(**

**select d.household\_key**

**from `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

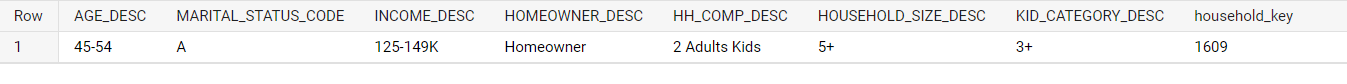
**on d.household\_key=t.household\_key**

**group by household\_key**

**order by sum(SALES\_VALUE) desc**

**limit 1**

**)**

****

* **2.d. What products does the top spender buys the most?**

**select household\_key, PRODUCT\_ID, count(QUANTITY) as quant**

**from `dunnhumbysql.complete.transaction\_data`**

**where household\_key in(**

**select d.household\_key,**

**from `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on t.household\_key=d.household\_key**

**group by d.household\_key**

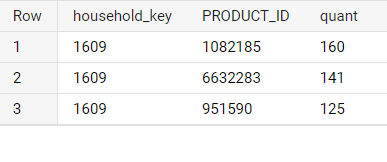
**order by sum(SALES\_VALUE) desc**

**limit 1)**

**group by household\_key, PRODUCT\_ID**

**order by quant desc**

**limit 3**

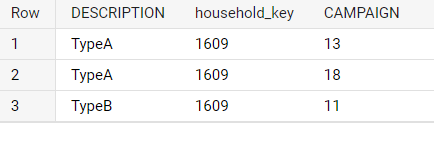
****

* **2.e. How many campaigns were they a part of?**

**select\***

**from `dunnhumbysql.complete.campaign\_table`**

**where household\_key=1609**

****

1. **Product analysis**

* **3.a. For this we’ll choose the best selling product from the transaction data ie :product\_id:1082185**

**select \* from `dunnhumbysql.complete.product`**

**where PRODUCT\_ID = (**

**SELECT PRODUCT\_ID,**

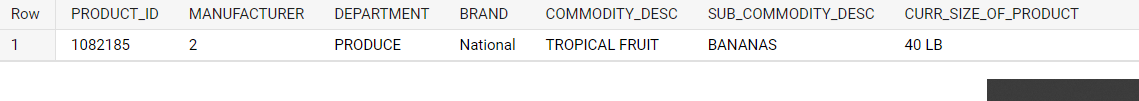
**FROM `dunnhumbysql.complete.transaction\_data`**

**group by PRODUCT\_ID**

**order by count(1)**

**limit 1**

**)**

****

* **3.b. When did the product sell the most and where (top 3)**

**select PRODUCT\_ID, count(QUANTITY) as qn, WEEK\_NO, STORE\_ID**

**from `dunnhumbysql.complete.transaction\_data`**

**where PRODUCT\_ID=(**

**SELECT PRODUCT\_ID,**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by PRODUCT\_ID**

**order by count(1) desc**

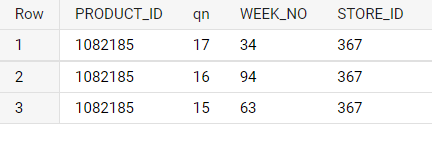
**limit 1**

**)**

**group by PRODUCT\_ID, WEEK\_NO, STORE\_ID**

**order by qn desc**

**limit 3**

****

* **3.c. Where was the product placed and featured for that particular store and week**

**with cte as(**

**select PRODUCT\_ID, WEEK\_NO, STORE\_ID**

**from `dunnhumbysql.complete.transaction\_data`**

**where PRODUCT\_ID=(**

**SELECT PRODUCT\_ID,**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by PRODUCT\_ID**

**order by count(1) desc**

**limit 1**

**)**

**group by PRODUCT\_ID,STORE\_ID, WEEK\_NO**

**order by count(1) desc**

**limit 3**

**)**

**select c.\***

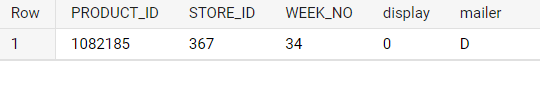
**from `dunnhumbysql.complete.causal\_data` c**

**right join cte t**

**on c.PRODUCT\_ID=t.PRODUCT\_ID**

**where c.WEEK\_NO in (t.WEEK\_NO )**

**and c.STORE\_ID in (t.STORE\_ID)**

****

* **3.f. Which products were the best seller(top 3) for each week and what quantity did they sell**

**SELECT \***

**FROM (**

**SELECT WEEK\_NO, PRODUCT\_ID,COUNT(PRODUCT\_ID)AS NUM\_SALES, ROW\_NUMBER() OVER (PARTITION BY WEEK\_NO ORDER BY COUNT(PRODUCT\_ID) DESC) AS n**

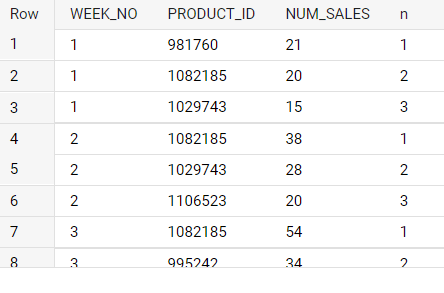
**FROM `dunnhumbysql.complete.transaction\_data`**

**GROUP BY WEEK\_NO , PRODUCT\_ID**

**ORDER BY WEEK\_NO**

**) AS x**

**WHERE n <= 3**

****

1. **Advance analysis and queries**

* **4.a. Find out on which weeks does each household shop and find their cumulative spending over time**

**with cte as(**

**select WEEK\_NO , household\_key , sum(SALES\_VALUE) as sales**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by WEEK\_NO, household\_key**

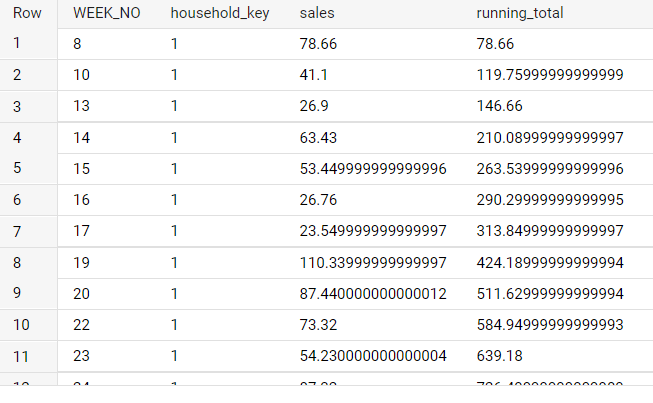
**)**

**SELECT**

**\*,**

**SUM(sales) OVER (PARTITION BY household\_key ORDER BY week\_no) AS running\_total**

**from cte**

****

* **4.b. Find the trend in spending for each customer**

**with cte as(**

**select WEEK\_NO , household\_key , sum(SALES\_VALUE) as sales**

**FROM `dunnhumbysql.complete.transaction\_data`**

**group by WEEK\_NO, household\_key**

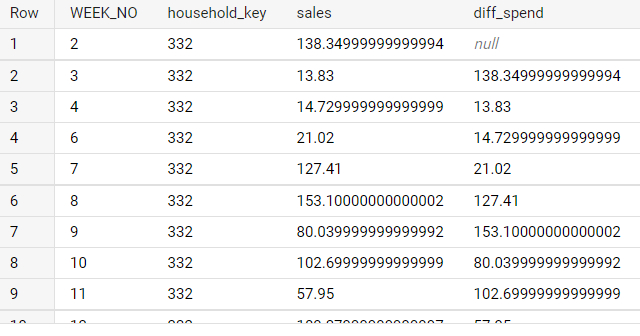
**)**

**SELECT**

**\*,**

**lag(sales)OVER (PARTITION BY household\_key ORDER BY week\_no) as diff\_spend,**

**from cte**

****

* **4.c. Find number of returning customers and percent of returning customers for all week**

**with cte as(**

**select b.week\_no, a.household\_key, CASE**

**when min(a.week\_no)<b.week\_no then 1 else 0**

**end as decider**

**from `dunnhumbysql.complete.transaction\_data` a**

**left join `dunnhumbysql.complete.transaction\_data` b**

**on a.household\_key=b.household\_key**

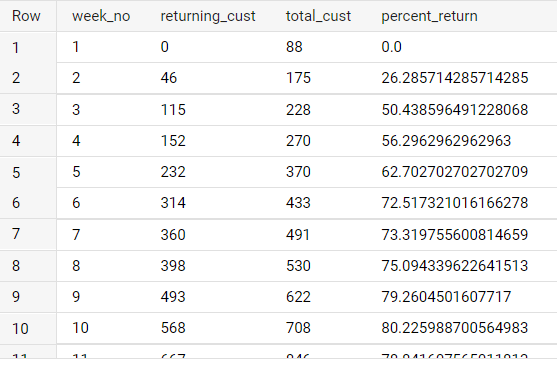
**group by b.week\_no, a.household\_key**

**) select week\_no,sum(decider) as returning\_cust, count(decider) as total\_cust, (sum(decider)/count(decider))\*100 as percent\_return**

**from cte**

**group by 1**

**order by 1**

****

* **4.d. Quarterly analysis: sales comparison over quarters**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

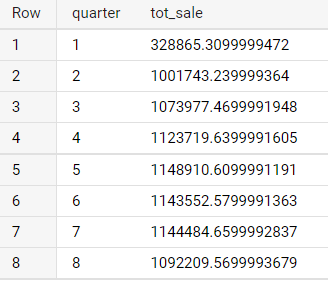
**end as quarter**

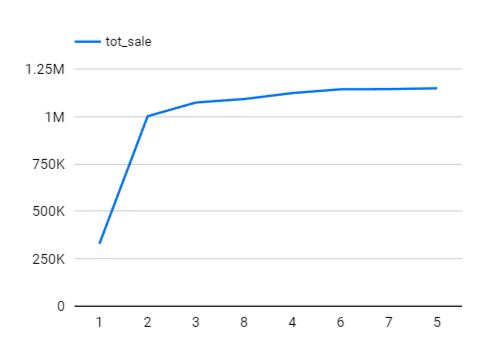
**from `dunnhumbysql.complete.transaction\_data`)**

**select quarter, sum(sales\_value) as tot\_sale**

**from cte**

**group by quarter**

****

****

* **4.e. Are the customers spending more or less over time**

select household\_key,sum(sales), case

when week\_no between 0 and 12 then 1

when week\_no between 13 and 25 then 2

when week\_no between 26 and 38 then 3

when week\_no between 39 and 51 then 4

when week\_no between 52 and 64 then 5

when week\_no between 65 and 77 then 6

when week\_no between 78 and 90 then 7

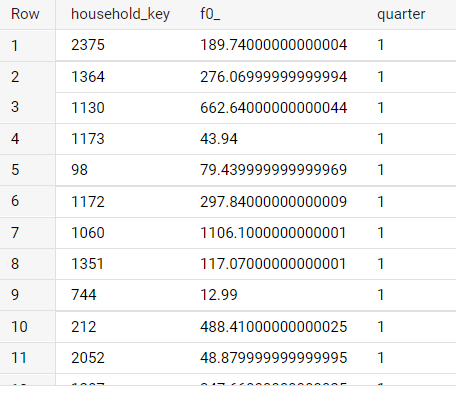
when week\_no between 91 and 102 then 8

end as quarter

from( select week\_no, household\_key, sum(SALES\_VALUE) as sales,

from `dunnhumbysql.complete.transaction\_data` group by household\_key, WEEK\_NO)

group by household\_key, quarter

****

**For better understanding selecting a single customer**

**select household\_key,sum(sales), case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

**end as quarter**

**from( select week\_no, household\_key, sum(SALES\_VALUE) as sales,**

**from `dunnhumbysql.complete.transaction\_data` group by household\_key, WEEK\_NO)**

**where household\_key=2375**

**group by household\_key, quarter**

****

* **4.f. Customer churn analysis for each quarter**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

**end as quarter**

**from `dunnhumbysql.complete.transaction\_data`)**

**select a.quarter, count(distinct(a.household\_key)) as chrned**

**from cte a**

**FULL OUTER JOIN cte b**

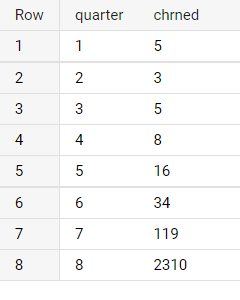
**on a.household\_key=b.household\_key**

**and a.quarter<b.quarter**

**where b.household\_key is NUll**

**group by a.quarter**

**order by quarter**

****

* **4.g. Find the retained customers for each quarter**

**with cte as(**

**select \*,**

**case**

**when week\_no between 0 and 12 then 1**

**when week\_no between 13 and 25 then 2**

**when week\_no between 26 and 38 then 3**

**when week\_no between 39 and 51 then 4**

**when week\_no between 52 and 64 then 5**

**when week\_no between 65 and 77 then 6**

**when week\_no between 78 and 90 then 7**

**when week\_no between 91 and 102 then 8**

**end as quarter**

**from `dunnhumbysql.complete.transaction\_data`)**

**select a.quarter, count(distinct(a.household\_key)) as retained**

**from cte a**

**left join cte b**

**on a.household\_key=b.household\_key and a.quarter>b.quarter**

**group by a.quarter**

****

* **4.h. Calculate Customer lifetime value(CLV) for different age group**
  + **Average purchase value —** the value of all customer purchases over a particular time frame , divided by the number of purchases in that period
  + **Average purchase frequency —** divide the number of purchases in that same time period by the number of individual customers who made a transaction over the same period
  + **Customer value —** the average purchase frequency multiplied by the average purchase value
  + **Average customer lifespan —** the average length of time a customer continues buying from you
  + **CLV = customer value X average customer lifespan**

**select AGE\_DESC, (avg\_purch\_val\*avg\_purch\_freq\*avg\_cust\_lifespan) as clv**

**from(**

**with cte as (**

**select household\_key, (max(WEEK\_NO)- min (WEEK\_NO)) as cust\_duration**

**from `dunnhumbysql.complete.transaction\_data`**

**group by household\_key**

**)**

**select AGE\_DESC, sum(SALES\_VALUE)/count(distinct(BASKET\_ID)) as avg\_purch\_val,**

**count(distinct(BASKET\_ID))/count(distinct(d.household\_key)) as avg\_purch\_freq,**

**(sum(cte.cust\_duration)/count(1)) as avg\_cust\_lifespan,**

**from `dunnhumbysql.complete.transaction\_data` t**

**inner join `dunnhumbysql.complete.hh\_demographic` d**

**on t.household\_key=d.household\_key**

**join cte**

**on cte.household\_key=d.household\_key**

**group by AGE\_DESC**

**)**

****

* **4.i. Find products(product:SUB\_COMMODITY\_DESC) which are most frequently bought together**

**with cte as (**

**SELECT \***

**FROM `dunnhumbysql.complete.product` p**

**join `dunnhumbysql.complete.transaction\_data` t**

**on p.PRODUCT\_ID=t.PRODUCT\_ID**

**)**

**select t.SUB\_COMMODITY\_DESC as item\_1, t2.SUB\_COMMODITY\_DESC as item\_2,**

**count(distinct t.BASKET\_ID) as num\_orders**

**from cte t**

**inner join cte t2**

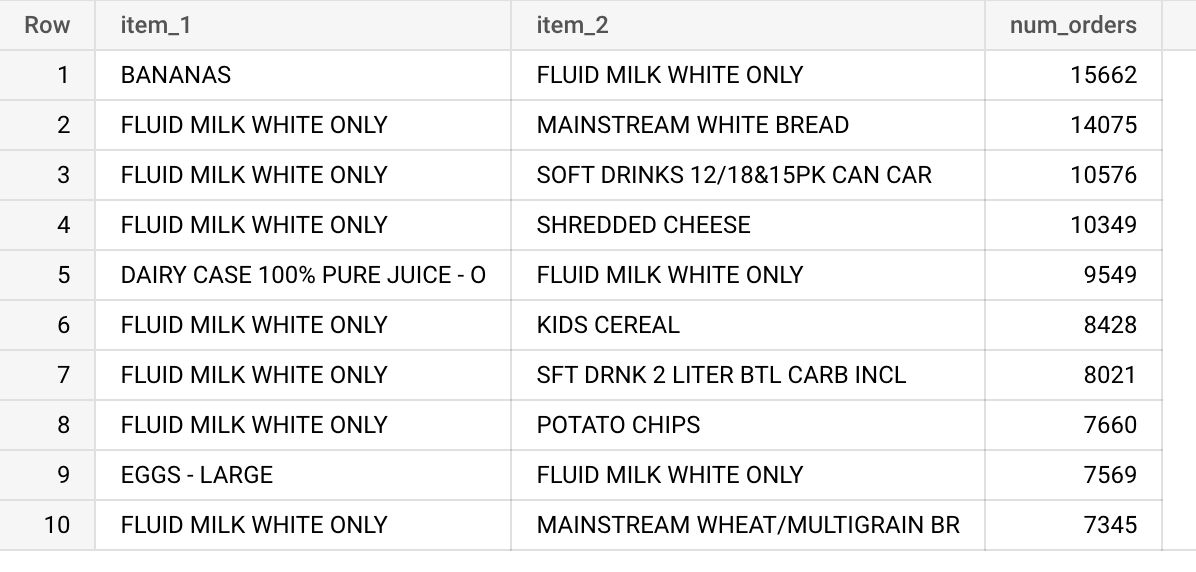
**on t.BASKET\_ID=t2.BASKET\_ID**

**and t.SUB\_COMMODITY\_DESC<t2.SUB\_COMMODITY\_DESC**

**group by t.SUB\_COMMODITY\_DESC, t2.SUB\_COMMODITY\_DESC**

**order by num\_orders desc**

**limit 10**

****