Retail Customer Analysis

TABLE OF CONTENTS

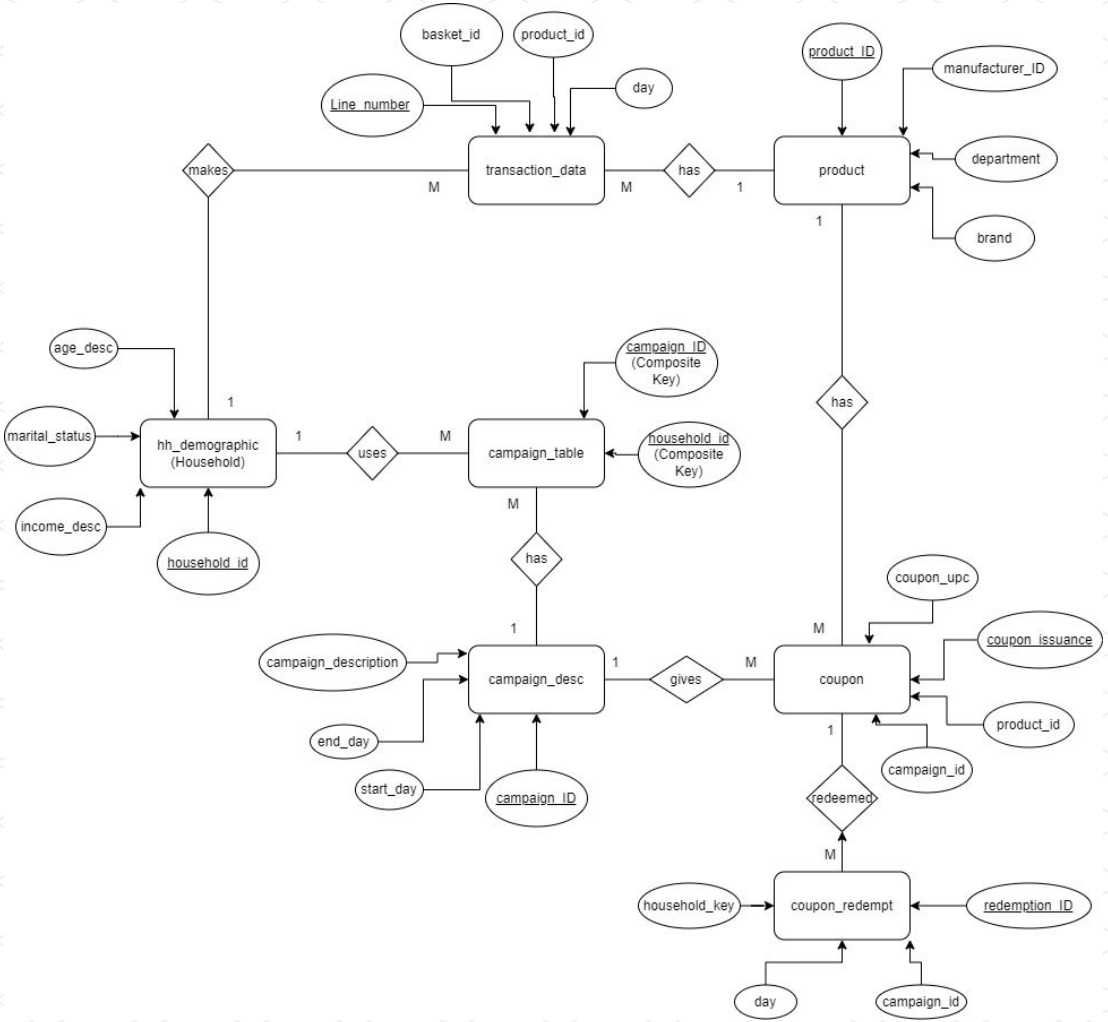
Overview

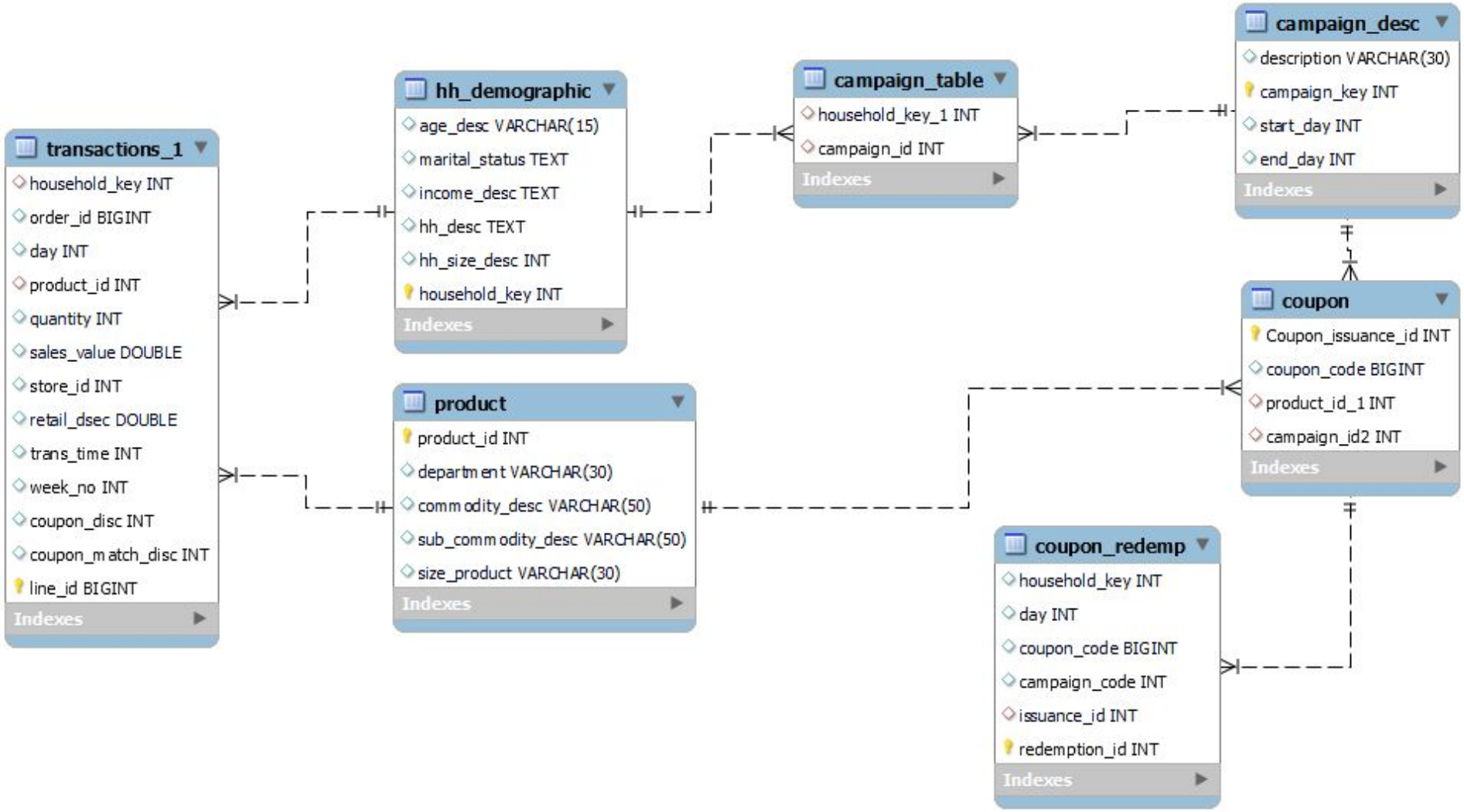
Our dataset contains transaction level details of a retail brand with demographic information of 2500 households where customers are targeted through marketing campaigns which promote their products through different offers and coupons.

Link - https://www.kaggle.com/datasets/frtgnn/dunnhumby-the-complete-journey

Data Overview Details

Table Name	Columns	Rows	Size
Transactions	13	2,595,732	138 MB
Household	6	735	30 KB
Product	5	92,353	5.1 MB
Campaign Table	2	7,208	94 KB
Campaign Desc	4	30	1 KB
Coupon	4	124,548	3.6 MB
Coupon Redemption	6	2318	79 KB


basket_id
product_id
day
Line_number
transaction_data
makes
has
product
product_ID
manufacturer_ID
department
brand
M
M
1
1
age_desc
marital_status
income_desc
household_id
hh_demographic (Household)
uses
campaign_table
campaign_ID (Composite Key)
household_id (Composite Key)
1
M
M
has
has
campaign_description
end_day
start_day
campaign_ID
campaign_desc
gives
coupon
coupon_upc
coupon_issuance
product_id
campaign_id
redeemed
household_key
coupon_redempt
redemption_ID
day
campaign_id


transactions_1
household_key INT
order_id BIGINT
day INT
product_id INT
quantity INT
sales_value DOUBLE
store_id INT
retail_dsec DOUBLE
trans_time INT
week_no INT
coupon_disc INT
coupon_match_disc INT
line_id BIGINT
Indexes
hh_demographic
age_desc VARCHAR(15)
marital_status TEXT
income_desc TEXT
hh_desc TEXT
hh_size_desc INT
household_key INT
Indexes
campaign_table
household_key_1 INT
campaign_id INT
Indexes
campaign_desc
description VARCHAR(30)
campaign_key INT
start_day INT
end_day INT
Indexes
product
product_id INT
department VARCHAR(30)
commodity_desc VARCHAR(50)
sub_commodity_desc VARCHAR(50)
size_product VARCHAR(30)
Indexes
coupon
Coupon_issuance_id INT
coupon_code BIGINT
product_id_1 INT
campaign_id2 INT
Indexes
coupon_redemp
household_key INT
day INT
coupon_code BIGINT
campaign_code INT
issuance_id INT
redemption_id INT
Indexes

Data Loading

Method 1

- For 3 Tables which are huge we created the skeleton of the table by uploading a sample file with 10 entries
- Then we truncated the table
- Then using local_infile method we loaded the data

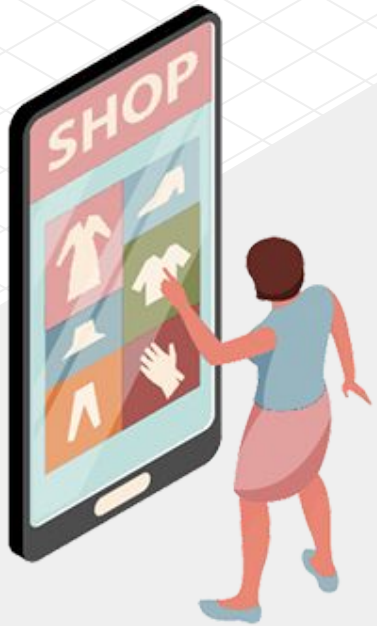
```
truncate table retail.coupon;
set global local_infile=on;
use retail;
LOAD DATA LOCAL INFILE 'C:\\project\\coupon.csv'
INTO TABLE coupon
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

Method 2

For the rest of the 4 tables we simply imported the entire data using the import wizard, we didn't face any major problem while completing this phase.

Insights

With Tableau Dashboard and Graphs



Overview of Data

Transactions
276,484

Coupons
46,805

Households
2,500

Days
711

Sales
8,507,563 $

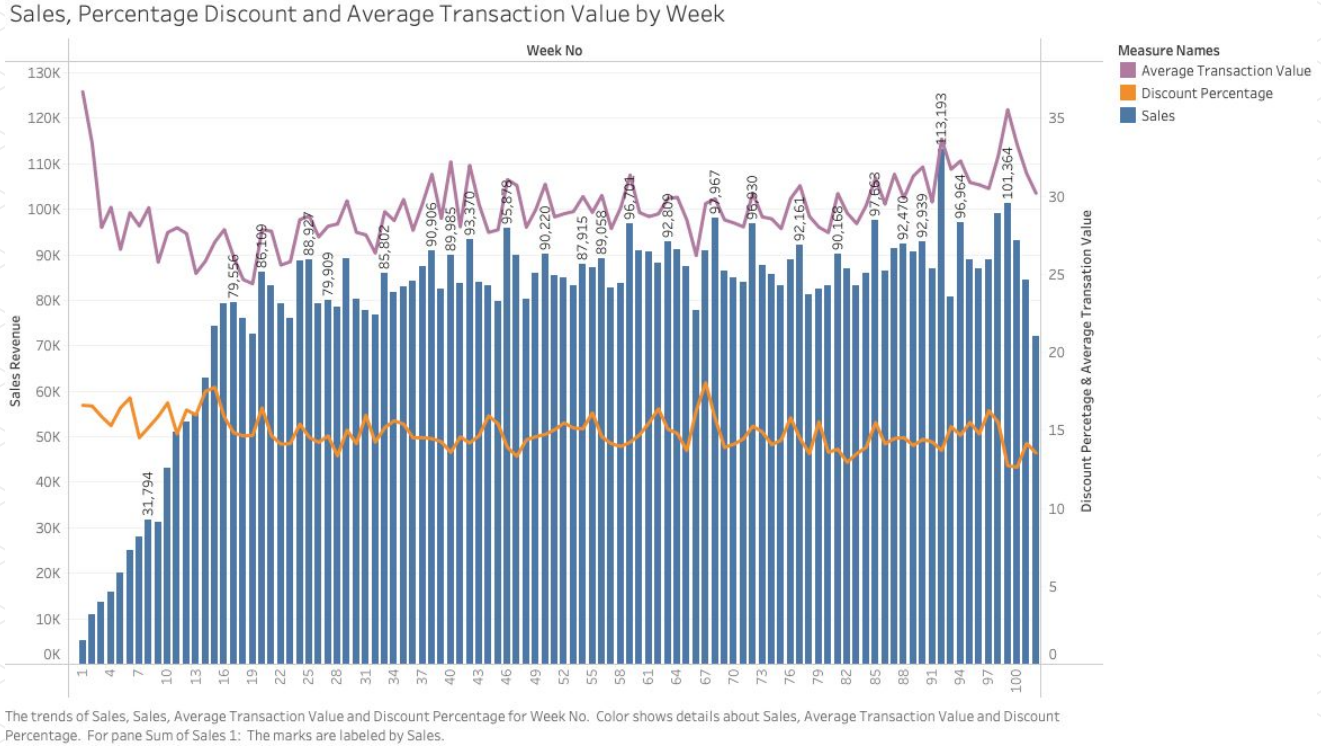
Weeks
102

Discount
1,398,335 $

% Discount
17%

Transactions have nearly 17% discount attached to them through the coupons.

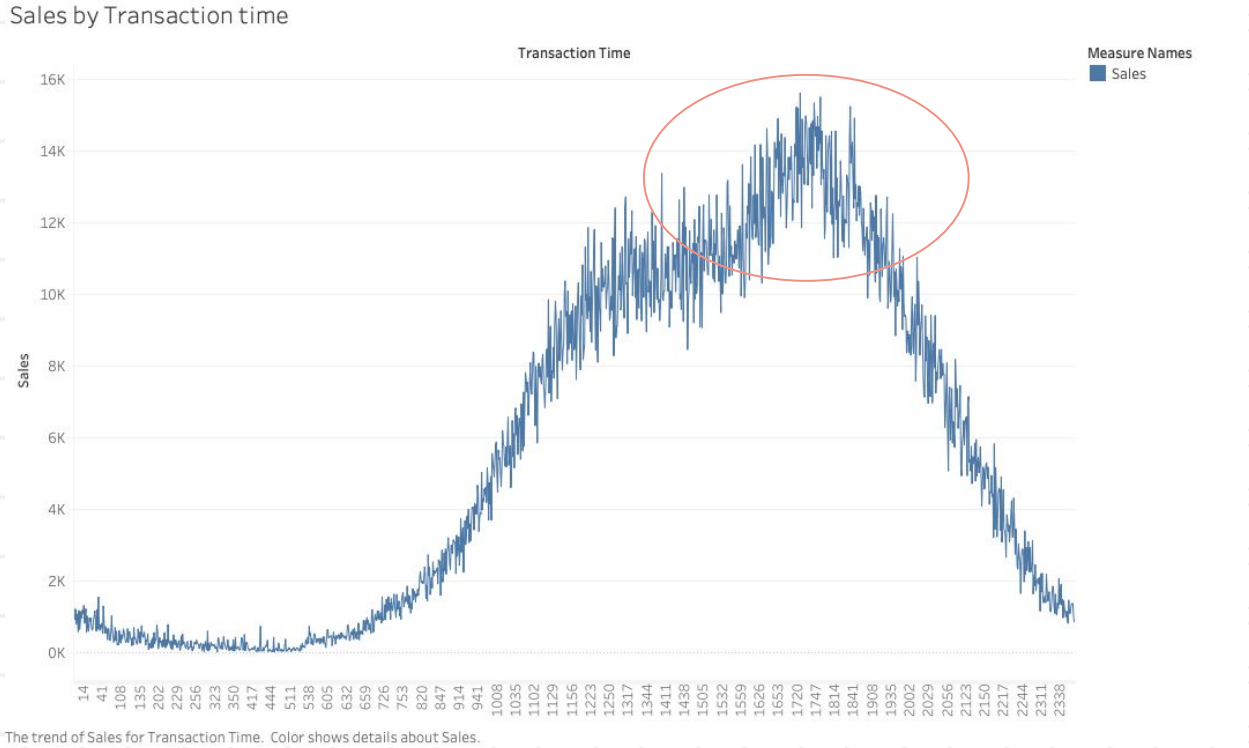
sales increase every 4-5th week which indicates payweek



The trends of Sales, Sales, Average Transaction Value and Discount Percentage for Week No. Color shows details about Sales, Average Transaction Value and Discount Percentage. For pane Sum of Sales 1: The marks are labeled by Sales.

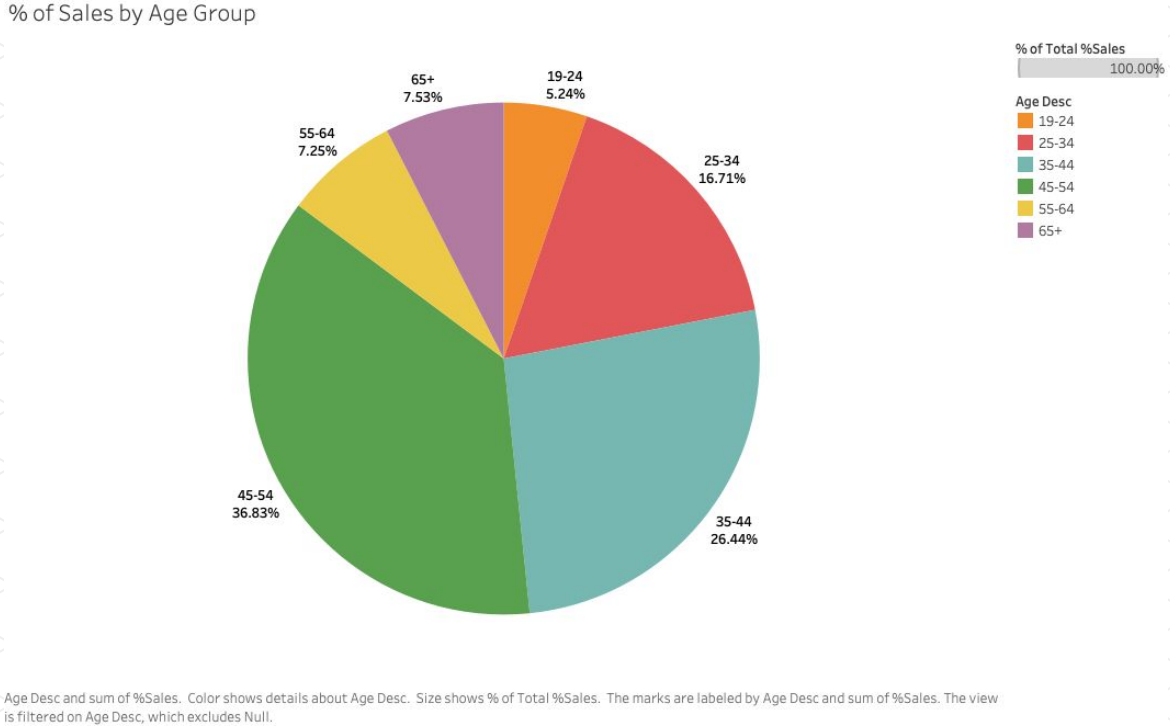
Sales during the non-pay weeks can be increased by catering lucrative offers and campaigns

The highest sales are captured during the time duration of 5pm-7pm



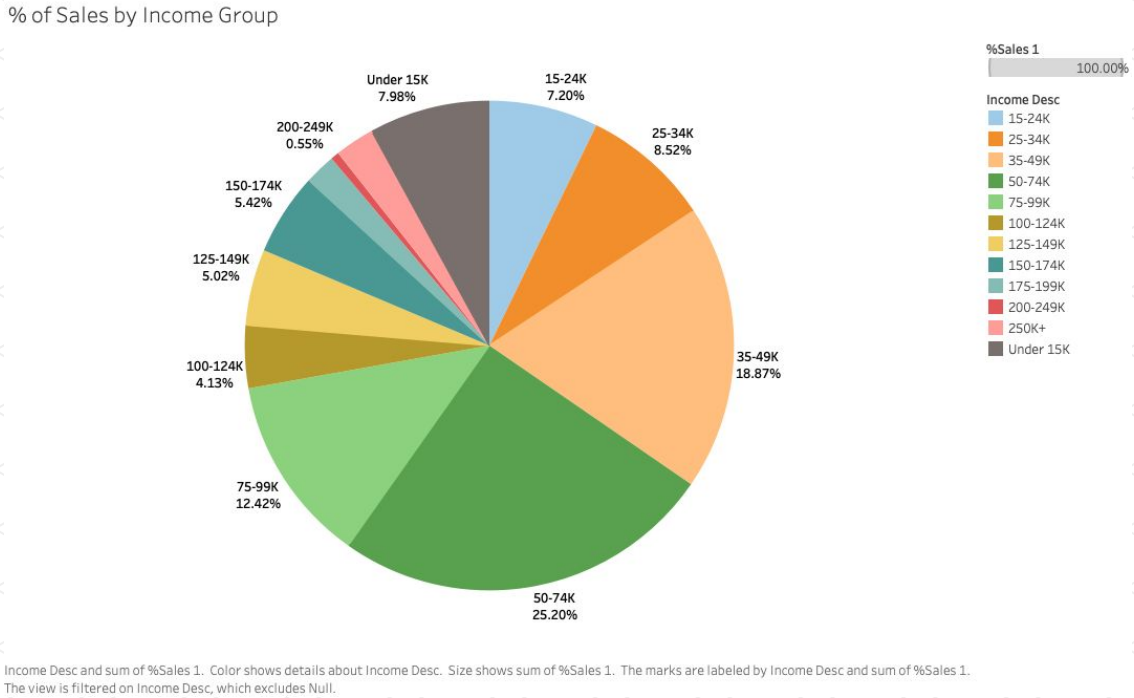
Happy hours, Rush hour sales can boost the business during the non peak time

age-group 45-54 has maximum contribution towards the sales - 37%



The business should consider catering attractive offers to the age groups which are low contributors of the sales by offering customized and targeted campaigns, coupons or discounts to them

Income Slab 50-74K has maximum contribution towards the sales - 25%



The business should focus on the income slabs next to the highest sales contributors (75-99k and 25-34k) as they can be converted into better contributors considering they might have similar needs and desires as the highest contributing slabs.

Coupon transactions make 154% more sale as compared to no coupon Txn

Detail	Transactions	Avg Sale	Avg Coupon Disc
Coupon Bills	15,672	68.2	-3.0
No Coupon Bills	260,812	26.8	0.0

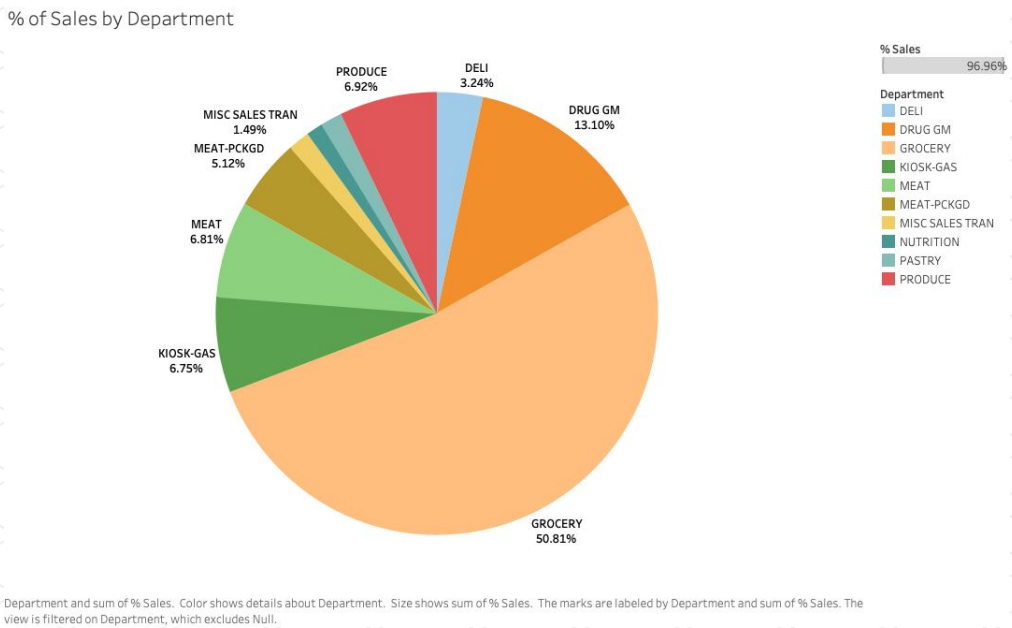
As it’s established that coupons aid in accelerating the sales so the business should keep up on that track.

Lower Ticket Sales Transactions have higher discount as compared to high ticket size customers

Ntile	Sales Bucket	Avg Sale	Avg Coupon Disc	% Discount
1	0-6.96	3.43	-1.44	-42%
2	6.96-17.07	11.52	-1.97	-17%
3	17.07-36.28	25.34	-2.27	-9%
4	36.28-961.49	76.28	-3.50	-5%

As we witnessed that customers who fall in the highest sales bucket are barely impacted by the discount and regardless of coupons they may still proceed with the purchase so the business should be more conservative in offering this slab more discount.

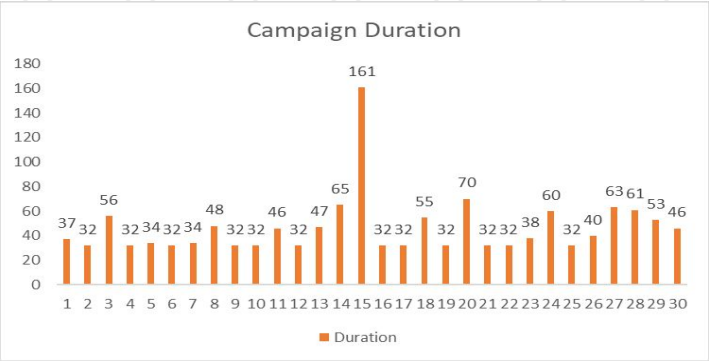
Grocery contributes to 50% of the sales followed by drugs which is at 13%

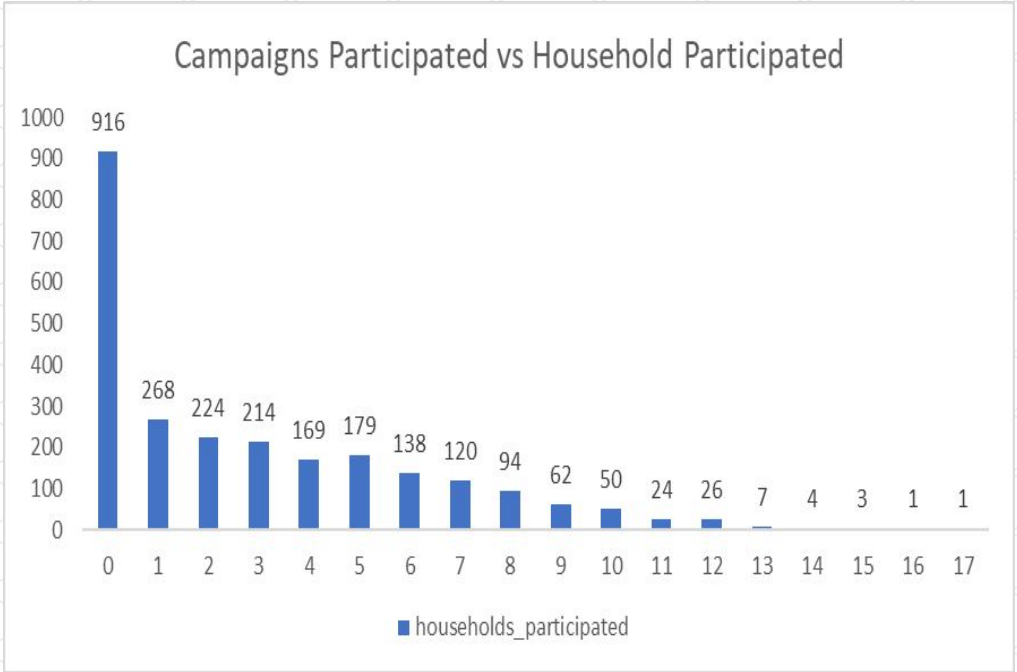


Business should come up with combo offers where they can club products of other departments with grocery which will help in introducing products to customers which can be sold independently later on.

3 Types of campaigns are being executed - Type A, Type B and Type C

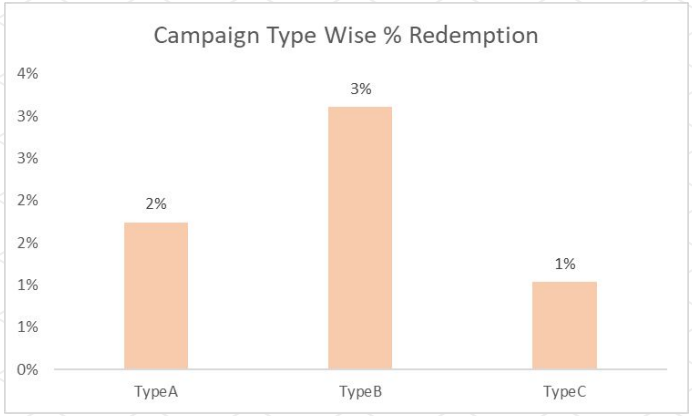
Description	Campaigns
TypeA	5
TypeB	19
TypeC	6

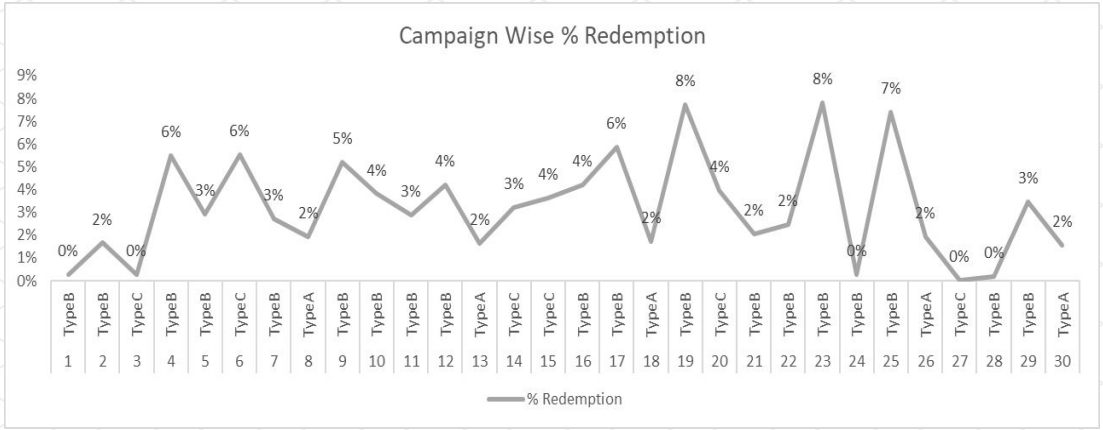




After running our analysis we discovered that the average campaign duration is 46.6 days. The longest campaign is 161 days long.

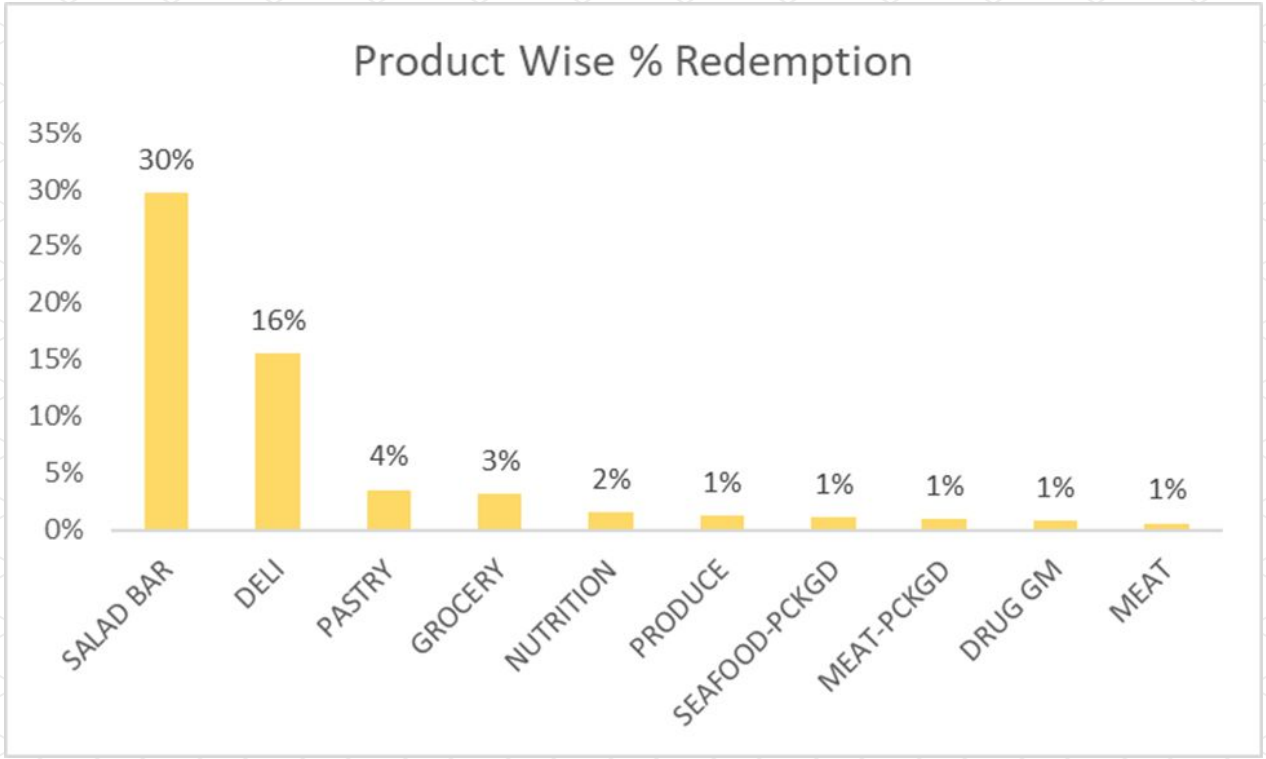
Type B Campaigns have a high success rate, Campaign 19 and Campaign 23 performed best where 8% coupons were redeemed.





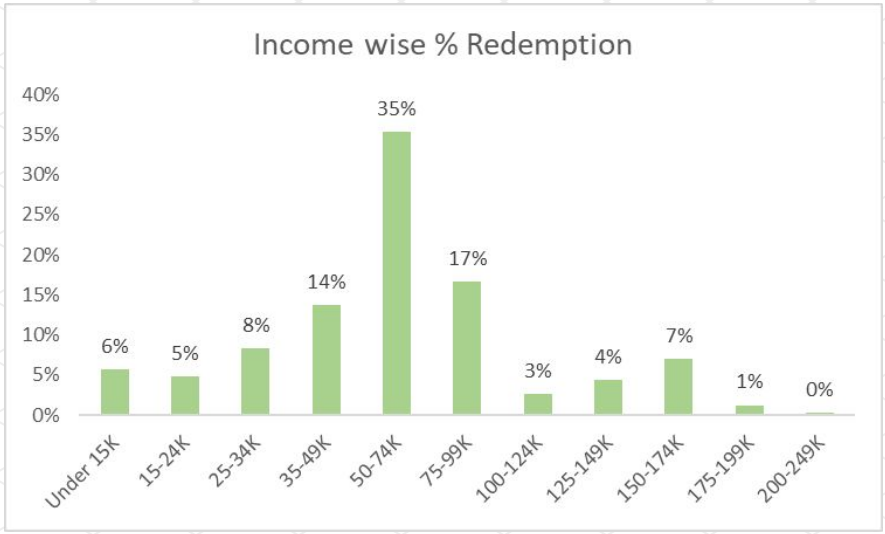
As established and can be seen in the above chart Type B campaigns have performed extraordinarily so business should do more of those to fetch better sales numbers for the store.

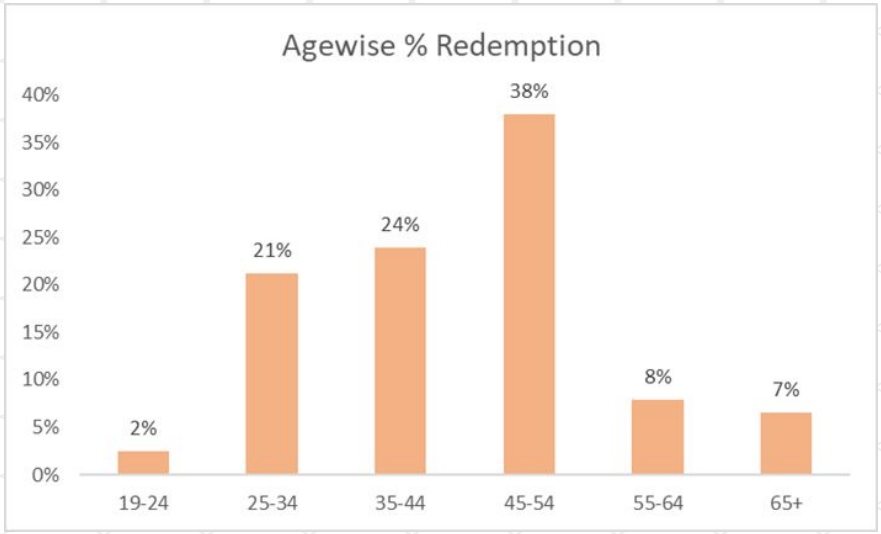
Salad bar(30%) from the food section is driving the maximum number of coupon redemption followed by Deli (16%)



Business can consider clubbing up highest grossing products which products having lesser sales value for example, salad bar can be clubbed with produce etc.

50-74K(35%) income bracket, age bracket of 45-54(38%)are redeeming more coupons





Business should be more concerned about expanding its customer base and should promote itself amongst different age groups and customers from different income brackets through promotion, marketing, coupons and campaigns.

THANK YOU

Complex SQL Queries

Creating Sales Bracket Using Ntile Function and CTE

```
-------------------------------------------------------
 ------ sale brackets ntile
WITH baskets
     AS (SELECT *,
                Ntile(4)
                  OVER(
                    ORDER BY sales) AS ntile_1
         FROM   (SELECT order_id,
                        household_key,
                        Sum(sales_value)       AS sales,
                        Sum(coupon_disc)       AS coupon_disc,
                        Sum(coupon_match_disc) AS COUPON_MATCH_DISC
                 FROM   transactions_1
                 GROUP  BY 1,
                           2) AS a)
SELECT ntile_1,
       Max(sales)                     AS max_sales,
       Min(sales)                     AS min_sales,
       Count(DISTINCT CASE
                        WHEN coupon_disc != 0 THEN order_id
                      END)            AS coupon_bills,
       Count(DISTINCT order_id)       AS transactions,
       Count(DISTINCT household_key)  AS households,
       Avg(sales)                     AS sales_mean,
       Avg(CASE
             WHEN coupon_disc < 0 THEN coupon_disc
           END)                       AS coupon_disc
FROM   baskets
GROUP  BY 1;
-------------------------------------------------------------------
```

Creating Product,Department Overview using Window Function and CTE

```
-------Product overview
SELECT department,
       commodity_desc,
       Sum(transactions)              AS transactions,
       Sum(households)                AS households,
       Sum(sales)                     AS sales,
       Sum(sales)
         OVER (
           partition BY department)   AS department_sales,
       Sum(discount)                  AS discount,
       Sum(coupons)                   AS coupons,
       Sum(quantity)                  AS quantity
FROM   (SELECT department,
               commodity_desc,
               Count(DISTINCT order_id)       AS transactions,
               Count(DISTINCT household_key)  AS households,
               Sum(sales_value)               AS sales,
               Sum(retail_dsec)               AS discount,
               Sum(coupon_disc)               AS coupons,
               Sum(quantity)                  AS quantity
        FROM   transactions_1 h
               JOIN product p
                 ON h.product_id = p.product_id
        GROUP  BY 1,
                  2) h
GROUP  BY 1,
          2
ORDER  BY 1,
          2;
```