

Dunnhumby E-commerce SQL Analysis

We have been given **Dunnhumby** customer demographics, transaction and product tables. We have been tasked to answer the following questions to derive sales trends, customer behavior and product desirability to name a few.

Datasets link :

https://drive.google.com/drive/folders/1xU91jKUknFRtBIC9vrKUSfhSC9_kvb30

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

```
With Order_Value as(
SELECT BASKET_ID,BASKET_TOTAL,
case
    when BASKET_TOTAL >0 AND BASKET_TOTAL <= 10
    then "small"
    when BASKET_TOTAL > 10 AND BASKET_TOTAL <= 20
    then "medium"
    else "large"
end as value_bracket
FROM
(select BASKET_ID, ROUND(SUM(SALES_VALUE),2) AS BASKET_TOTAL
from `Sales_Analysis.Transaction`
GROUP BY BASKET_ID) AS BASKET_SUMMARY)

select  distinct value_bracket,
```

```
count(Order_Value.BASKET_TOTAL) over (partition by value_bracket
from Order_Value;
```

Row	value_bracket	Count
1	small	115495
2	large	68231
3	medium	49630

Question 2: Find the number of orders that are small, medium or large order value(small:0-5 dollars, medium:5-10 dollars, large:10+)

```
With Order_Value as(
SELECT BASKET_ID,BASKET_TOTAL,
case
    when BASKET_TOTAL >0 AND BASKET_TOTAL <= 5
    then "small(0-5)"
    when BASKET_TOTAL > 5 AND BASKET_TOTAL <= 10
    then "medium(5-10)"
    else "large(10+)"
end as value_bracket
FROM
(select BASKET_ID, ROUND(SUM(SALES_VALUE),2) AS BASKET_TOTAL
from `Sales_Analysis.Transaction`
GROUP BY BASKET_ID) AS BASKET_SUMMARY)

select  distinct value_bracket,
count(Order_Value.BASKET_TOTAL) over (partition by value_bracket
```

```
from Order_Value
order by Count desc;
```

Row	value_bracket	Count
1	large(10+)	117861
2	small(0-5)	69920
3	medium(5-10)	45575

Question 3: Find top 3 stores with highest foot traffic for each week (Foot traffic: number of customers transacting)

```
with store_pref as(
select  WEEK_NO, STORE_ID, FOOT_TRAFFIC,
ROW_NUMBER() OVER(PARTITION BY WEEK_NO ORDER BY FOOT_TRAFFIC DESC)
FROM
(select  WEEK_NO, STORE_ID,
COUNT(DISTINCT household_key) AS FOOT_TRAFFIC
from `Sales_Analysis.Transaction`
GROUP BY
WEEK_NO, STORE_ID) AS TRAFFIC)

select WEEK_NO, STORE_ID, FOOT_TRAFFIC
FROM store_pref
WHERE ORDERING <=3
order by WEEK_NO asc, FOOT_TRAFFIC desc;
```

Row	WEEK_NO ▼	STORE_ID ▼	FOOT_TRAFFIC ▼
1	1	32004	5
2	1	324	3
3	1	367	3
4	2	32004	7
5	2	313	6
6	2	367	5
7	3	367	10
8	3	32004	9
9	3	356	8
10	4	367	17
11	4	32004	11

Question 4: 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 d.household_key, min(t.DAY) as FIRST_VISIT, max(t.DAY) as LAST_VISIT,
count(distinct t.BASKET_ID) as VISIT_COUNT,
ROUND(SUM(t.SALES_VALUE)/count(distinct t.BASKET_ID),2) as AVG_I
round(SUM(t.SALES_VALUE),2) as TOTAL_EXP
from `Sales_Analysis.Demographic` as d
inner join `Sales_Analysis.Transaction` as t
on d.household_key = t.household_key
```

```
GROUP BY d.household_key
order by AVG_EXP_VISIT desc;
```

Row	household_key	FIRST_VISIT	LAST_VISIT	VISIT_COUNT	AVG_EXP_VISIT	TOTAL_EXP
1	973	95	710	80	85.95	6875.89
2	1899	20	705	69	83.91	5789.59
3	2479	111	706	111	62.65	6954.64
4	248	29	704	53	58.32	3090.89
5	688	70	692	27	57.74	1558.95
6	1864	103	710	148	57.68	8537.28
7	1848	105	706	97	57.34	5561.56
8	888	12	706	76	56.57	4299.36
9	392	75	709	103	54.93	5657.72
10	2134	30	711	84	54.8	4602.91
11	1662	85	702	79	54.8	4329.22

Question 5: 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 total spent)

```
with top_customer as(
select distinct d.household_key, ROUND(SUM(t.SALES_VALUE),2) as Total_Exp
from `Sales_Analysis.Demographic` as d
inner join `Sales_Analysis.Transaction` as t
on d.household_key = t.household_key
group by d.household_key
order by Total_Exp desc
limit 1)
```

```
select d.*, t.Total_Exp
from `Sales_Analysis.Demographic` as d
```

```
inner join top_customer as t
on d.household_key = t.household_key;
```

Row	AGE_DESC ▾	MARITAL_STATUS_CODE ▾	INCOME_DESC ▾	HOMEOWNER_DESC ▾	HH_COMP_DESC ▾	HOUSEHOLD_SIZE
1	45-54	A	125-149K	Homeowner	2 Adults Kids	5+

Row	HOMEOWNER_DESC ▾	HH_COMP_DESC ▾	HOUSEHOLD_SIZE_DESC ▾	KID_CATEGORY_DESC ▾	household_key ▾	Total_Exp ▾
1	Homeowner	2 Adults Kids	5+	3+	1609	13804.38

Question 6: Find products(product table : SUB_COMMODITY_DESC) which are most frequently bought together and the count of each combination bought together. do not print a combination twice (A-B / B-A)

```
with frequent_prod as(
select t1.BASKET_ID,t1.PRODUCT_ID as prod1, t2.PRODUCT_ID as prod2
from `Sales_Analysis.Transaction` as t1
join `Sales_Analysis.Transaction` as t2
on t1.BASKET_ID = t2.BASKET_ID
where t1.PRODUCT_ID > t2.PRODUCT_ID),

freq_prod_res as(
select frequent_prod.prod1,frequent_prod.prod2,count(*) as Item_Count
from frequent_prod
group by frequent_prod.prod1,frequent_prod.prod2
order by Item_Count DESC)

select fp.prod1,p1.SUB_COMMODITY_DESC,fp.prod2,p2.SUB_COMMODITY_DESC
from freq_prod_res as fp
join `Sales_Analysis.Product` as p1
```

```

on fp.prod1 = p1.PRODUCT_ID
join `Sales_Analysis.Product` as p2
on fp.prod2 = p2.PRODUCT_ID
order by fp.Item_Count desc;

```

Row	prod1	SUB_COMMODITY_DESC	prod2	SUB_COMMODITY_DESC_1	Item_Count
1	1082185	BANANAS	1029743	FLUID MILK WHITE ONLY	848
2	1082185	BANANAS	995242	FLUID MILK WHITE ONLY	728
3	1082185	BANANAS	981760	EGGS - X-LARGE	625
4	1127831	STRAWBERRIES	1082185	BANANAS	611
5	1106523	FLUID MILK WHITE ONLY	1082185	BANANAS	519
6	1082185	BANANAS	961554	CARROTS MINI PEELED	473
7	1082185	BANANAS	951590	MAINSTREAM WHITE BREAD	458
8	1082185	BANANAS	1070820	FLUID MILK WHITE ONLY	430
9	1126899	FLUID MILK WHITE ONLY	1082185	BANANAS	427
10	1098066	HOT DOG BUNS	826249	HAMBURGER BUNS	421
11	1082185	BANANAS	854852	TOMATOES HOTHOUSE ON TH...	420

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

```

with weekly_exp as(
select household_key,WEEK_NO,Weekly_Sum,
lag(Weekly_Sum) over(partition by household_key order by WEEK_NO)
from
(select d.household_key,t.WEEK_NO,
sum(t.SALES_VALUE) as Weekly_Sum
from `Sales_Analysis.Demographic` as d
inner join `Sales_Analysis.Transaction` as t
on d.household_key = t.household_key
group by d.household_key,t.WEEK_NO
order by d.household_key,t.WEEK_NO))

```

```
select household_key,WEEK_NO,weekly_exp.Weekly_Sum, Last_Week_Sur
Round((Weekly_Sum - Last_Week_Sum),2) as Revenue_Per_Acc
from weekly_exp
order by household_key,WEEK_NO;
```

Row	household_key ▼	WEEK_NO ▼	Weekly_Sum ▼	Last_Week_Sum ▼	Revenue_Per_Acc ▼
1	1	8	42.58	null	null
2	1	10	14.01	42.58	-28.57
3	1	13	14.030000000000...	14.01	0.02
4	1	14	25.71	14.030000000000...	11.68
5	1	15	10.98	25.71	-14.73
6	1	16	9.09	10.98	-1.89
7	1	17	13.98	9.09	4.89
8	1	19	47.350000000000...	13.98	33.37
9	1	20	31.77	47.350000000000...	-15.58
10	1	22	38.980000000000...	31.77	7.21
11	1	23	26.36	38.980000000000...	-12.62

Question 8 :Identify the most popular products (based on quantity sold) in each store and display the top 3 for each store.

```
with store_pref as(
select  STORE_ID,SUB_COMMODITY_DESC,Quantity,
ROW_NUMBER() OVER(PARTITION BY STORE_ID ORDER BY Quantity DESC)
FROM
(select  t.STORE_ID,p.SUB_COMMODITY_DESC,
SUM(t.QUANTITY) AS Quantity
from `Sales_Analysis.Transaction` as t
```



```

inner join `Sales_Analysis.Product` as p
on t.PRODUCT_ID = p.PRODUCT_ID
GROUP BY
    t.STORE_ID,p.SUB_COMMODITY_DESC) AS TRAFFIC)

select STORE_ID,SUB_COMMODITY_DESC,Quantity
FROM store_pref
WHERE ORDERING <=3
order by STORE_ID asc,Quantity desc;

```

Row	STORE_ID ▼	SUB_COMMODITY_DESC ▼	Quantity ▼
1	1	GASOLINE-REG UNLEADED	7249
2	1	LAXATIVES	1
3	2	FINISHED ELECTRICAL	1
4	2	PREMIUM FLOWERING PLANTS	1
5	2	BANANAS	1
6	12	BEERALEMALT LIQUORS	1
7	12	POPULAR 750ML WINES	1
8	12	SFT DRNK 2 LITER BTL CARB I...	1
9	19	GASOLINE-REG UNLEADED	1
10	19	CEREAL BARS	1
11	19	CANDY BARS MULTI PACK W/F...	1

Question 9 :Calculate the average spending per transaction across different income groups (based on the

INCOME_DESC column in the Demographic table).

```
select d.INCOME_DESC,  
       round(sum(t.SALES_VALUE)/count(distinct t.BASKET_ID),2) as Avg_  
from `Sales_Analysis.Demographic` as d  
inner join `Sales_Analysis.Transaction` as t  
on d.household_key = t.household_key  
group by d.INCOME_DESC  
order by Avg_Per_Trans desc;
```

Row	INCOME_DESC ▼	Avg_Per_Trans ▼
1	150-174K	29.6
2	200-249K	27.46
3	250K+	26.86
4	175-199K	25.1
5	125-149K	23.89
6	75-99K	22.53
7	50-74K	19.92
8	100-124K	18.68
9	35-49K	16.62
10	Under 15K	15.79
11	25-34K	15.16
...

Question 10: Determine the percentage of total

revenue contributed by each product category (**COMMODITY_DESC**)

```
SELECT p.COMMODITY_DESC, SUM(t.SALES_VALUE) as Commodity_Sales,
ROUND((SUM(t.SALES_VALUE) / (SELECT SUM(SALES_VALUE) FROM `Sales_Analysis.Transaction`)
* 100, 2) as Revenue_Cont
FROM `Sales_Analysis.Product` as p
inner join `Sales_Analysis.Transaction` as t
ON p.PRODUCT_ID = t.PRODUCT_ID
GROUP BY p.COMMODITY_DESC
ORDER BY Revenue_Cont DESC;
```

Row	COMMODITY_DESC ▼	Commodity_Sales ▼	Revenue_Cont ▼
1	COUPON/MISC ITEMS	319834.69999999...	7.94
2	SOFT DRINKS	164139.59999999...	4.07
3	BEEF	156390.77999999...	3.88
4	FLUID MILK PRODUCTS	102343.29000000...	2.54
5	CHEESE	95325.77999999...	2.37
6	FRZN MEAT/MEAT DINNERS	79587.62999999...	1.98
7	BAG SNACKS	74695.40999999...	1.85
8	BEERS/ALES	74152.49999999...	1.84
9	BAKED BREAD/BUNS/ROLLS	73137.95999999...	1.82
10	FROZEN PIZZA	73055.41999999...	1.81
11	COLD CEREAL	55622.23999999...	1.38

Insights/ Recommendations

1. Order value of **0-10 dollars** are **highest** in number, with orders valued **higher than 20 dollars** come in **second**.
2. Order value of **more than 10 dollars** are highest in number while **small sized(0-5 dollar)** orders are the next **highest**.

Recommendation : Considering the above 2 insights, items in the range of 5-20 dollars must be stocked well and marketed accordingly, as most sales are happening in the range.

3. Store **32004** and **367** is seen to be having high footfall week on week.
Recommendation : Inspect the stores closely to see the factors that work for it, so that it can be replicated in other stores as well.

4. Household number **973** has the highest spent the most on each visit per average at **85 dollars** , with **80 visits** and spent a total amount of **6875 dollars**.

5. Household number **1609** spent **13804 dollars**, with a household of **more than 5 members**. The householder is **45-54 years old** and earns **125-149K dollars** annually.

6. **Bananas-Milk, Bananas-Eggs** are the most frequently bought items. Above, Bananas-Milk is repeated several times, as **Fluid Milk White Only has several records with different Product Id**. This must be looked into, raising the question was it done because it is supplied by different Manufacturer as indicated in the Products table?

7. There is a **positive-negative** fluctuation every week , couldn't identify a clear trend.

8. **Gasoline and Coupons** are the most sold items , apart from **Groceries** for every store.

9. There is a **relationship between high spending and high income families**.
Recommendation : Suggest products at the higher end of the spectrum for families that earn more.

10. **Coupons, Soft drinks and Beef** have the largest chunk in the revenue share for the store.

