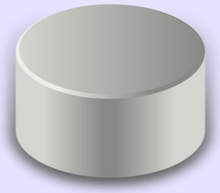
MySQL

CASE STUDY DATA MART

Presented By : Nikhil Kumar Roy

Data Dart is my latest venture and I want your help to analyze the sales and performance of my venture. In June 2020 - large-scale supply changes were made at Data Mart. All Data Mart products now use sustainable packaging methods in every single step from the farm all the way to the customer.



SCHEMA

Column name	Data type
week_date	date
region	varchar(20)
platform	varchar(20)
segment	varchar(10)
customer	varchar(20)
transactions	int
sales	int

```sql
CREATE TABLE clean_weekly_sales AS SELECT week_date,
    WEEK(week_date) AS week_number, MONTH(week_date) AS month_number, YEAR(week_date) AS calendar_year,
    region,platform,
    CASE
        WHEN segment = 'null' THEN 'Unknown'
        ELSE segment
    END AS segment,
    CASE
        WHEN RIGHT(segment, 1) = '1' THEN 'Young Adults'
        WHEN RIGHT(segment, 1) = '2' THEN 'Middle Aged'
        WHEN RIGHT(segment, 1) IN ('3' , '4') THEN 'Retirees'
        ELSE 'Unknown'
    END AS age_band,
    CASE
        WHEN LEFT(segment, 1) = 'C' THEN 'Couples'
        WHEN LEFT(segment, 1) = 'F' THEN 'Families'
        ELSE 'Unknown'
    END AS demographic,
    customer_type, transactions, sales,
    ROUND(sales / transactions, 2) AS avg_transaction FROM
    weekly_sales;
```

DATA CLEANSING

week_date	week_number	month_number	calendar_year	region	platform	segment	age_band	demographic	customer_type	transactions	sales	avg_transaction
2020-08-31	35	8	2020	ASIA	Retail	C3	Retirees	Couples	New	120631	3656163	30.31
2020-08-31	35	8	2020	ASIA	Retail	F1	Young Adults	Families	New	31574	996575	31.56
2020-08-31	35	8	2020	USA	Retail	Unknown	Unknown	Unknown	Guest	529151	16509610	31.20
2020-08-31	35	8	2020	EUROPE	Retail	C1	Young Adults	Couples	New	4517	141942	31.42

1. Which week numbers are missing from the dataset?

```sql
CREATE TABLE seq100 (
    x INT NOT NULL AUTO_INCREMENT PRIMARY KEY
);
insert into seq100 values (),(),(),(),(),(),(),(),(),();
insert into seq100 values (),(),(),(),(),(),(),(),(),();
insert into seq100 values (),(),(),(),(),(),(),(),(),();
insert into seq100 values (),(),(),(),(),(),(),(),(),();
insert into seq100 values (),(),(),(),(),(),(),(),(),();
insert into seq100 select x + 50 from seq100;

CREATE TABLE seq52 AS (SELECT x FROM seq100
LIMIT 52);
SELECT DISTINCT x AS week_day
FROM seq52
WHERE
    x NOT IN (SELECT DISTINCT week_number
        FROM clean_weekly_sales);

select distinct week_number from clean_weekly_sales;
```

Result Grid

week_number
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14

```sql
SELECT
    calendar_year, SUM(transactions) AS total_transactions
FROM
    clean_weekly_sales
GROUP BY calendar_year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

calendar_year	total_transactions
2020	375813651
2019	365639285
2018	346406460

2.How many total transactions were there for each year in the dataset?

3.What are the total sales for each region for each month?

```sql
SELECT
    month_number, region, SUM(sales) AS total_sales
FROM
    clean_weekly_sales
GROUP BY month_number , region
ORDER BY month_number , region;
```

month_number	region	total_sales
4	CANADA	484552594
4	EUROPE	127334255
4	OCEANIA	2599767620
4	SOUTH AMERICA	238451531
4	USA	759786323
5	AFRICA	1647244738
5	ASIA	1526285399
5	CANADA	412378365
5	EUROPE	109338389
5	OCEANIA	2215657304
5	SOUTH AMERICA	201391809
5	USA	655967121
6	AFRICA	1767559760
6	ASIA	1619482889
6	CANADA	443846698
6	EUROPE	122813826
6	OCEANIA	2371884744
6	SOUTH AMERICA	218247455
6	USA	703878990
7	AFRICA	1960219710
7	ASIA	1768844756
7	CANADA	477134947

```sql
SELECT
    platform, SUM(transactions) AS total_transactions
FROM
    clean_weekly_sales
GROUP BY platform;
```

	platform	total_transactions
▶	Retail	1081934227
	Shopify	5925169

4.What is the total count of transactions for each platform

5. What is the percentage of sales for Retail vs Shopify for each month?

```sql
WITH cte_monthly_platform_sales AS (
  SELECT
    month_number,calendar_year,
    platform,
    SUM(sales) AS monthly_sales
  FROM clean_weekly_sales
  GROUP BY month_number,calendar_year, platform)
```

```sql
SELECT
  month_number,calendar_year,
  ROUND(
    100 * MAX(CASE WHEN platform = 'Retail' THEN monthly_sales ELSE NULL END) /
      SUM(monthly_sales),
    2
  ) AS retail_percentage,
  ROUND(
    100 * MAX(CASE WHEN platform = 'Shopify' THEN monthly_sales ELSE NULL END) /
      SUM(monthly_sales),
    2
  ) AS shopify_percentage
FROM cte_monthly_platform_sales
GROUP BY month_number,calendar_year
ORDER BY month_number,calendar_year;
```

month_number	calendar_year	retail_percentage	shopify_percentage
3	2018	97.92	2.08
3	2019	97.71	2.29
3	2020	97.30	2.70
4	2018	97.93	2.07
4	2019	97.80	2.20
4	2020	96.96	3.04
5	2018	97.73	2.27
5	2019	97.52	2.48
5	2020	96.71	3.29
6	2018	97.76	2.24
6	2019	97.42	2.58
6	2020	96.80	3.20
7	2018	97.75	2.25
7	2019	97.35	2.65
7	2020	96.67	3.33
8	2018	97.71	2.29
8	2019	97.21	2.79
8	2020	96.51	3.49
9	2018	97.68	2.32
9	2019	97.09	2.91

6.What is the percentage of sales by demographic for each year in the dataset?

```sql
SELECT
  calendar_year,
  demographic,
  SUM(SALES) AS yearly_sales,
  ROUND(
    (
      100 * SUM(sales)/
        SUM(SUM(SALES)) OVER (PARTITION BY demographic)
    ),
    2
  ) AS percentage
FROM clean_weekly_sales
GROUP BY
  calendar_year,
  demographic
ORDER BY
  calendar_year,
  demographic;
```

calendar_year	demographic	yearly_sales	percentage
2018	Couples	3402388688	30.38
2018	Families	4125558033	31.25
2018	Unknown	5369434106	32.86
2019	Couples	3749251935	33.47
2019	Families	4463918344	33.81
2019	Unknown	5532862221	33.86
2020	Couples	4049566928	36.15
2020	Families	4614338065	34.95
2020	Unknown	5436315907	33.27

7.Which age_band and demographic values contribute the most to Retail sales?

```sql
SELECT
    age_band, demographic, SUM(sales) AS total_sales
FROM
    clean_weekly_sales
WHERE
    platform = 'Retail'
GROUP BY age_band , demographic
ORDER BY total_sales DESC;
```

age_band	demographic	total_sales
Unknown	Unknown	16067285533
Retirees	Families	6634686916
Retirees	Couples	6370580014
Middle Aged	Families	4354091554
Young Adults	Couples	2602922797
Middle Aged	Couples	1854160330
Young Adults	Families	1770889293

THANK YOU

https://github.com/NikhilRoyDA

https://www.linkedin.com/in/nikhil-kumar-roy/

nikhilroy744@gmail.com