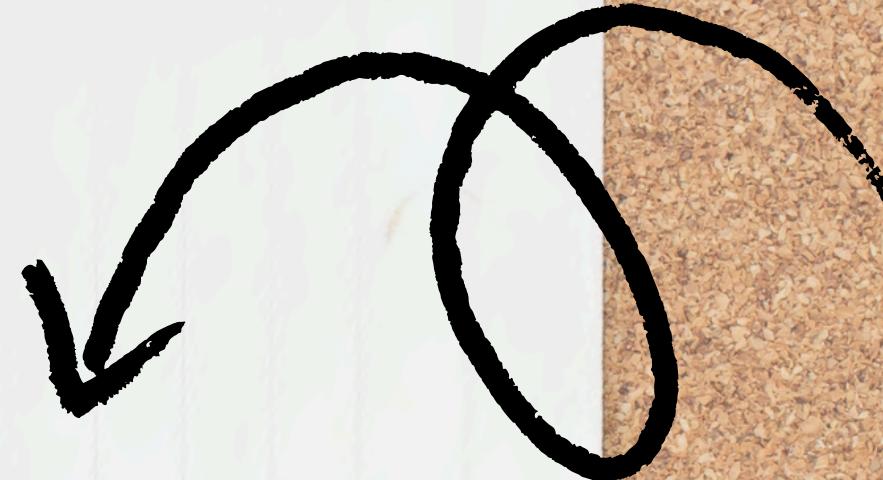


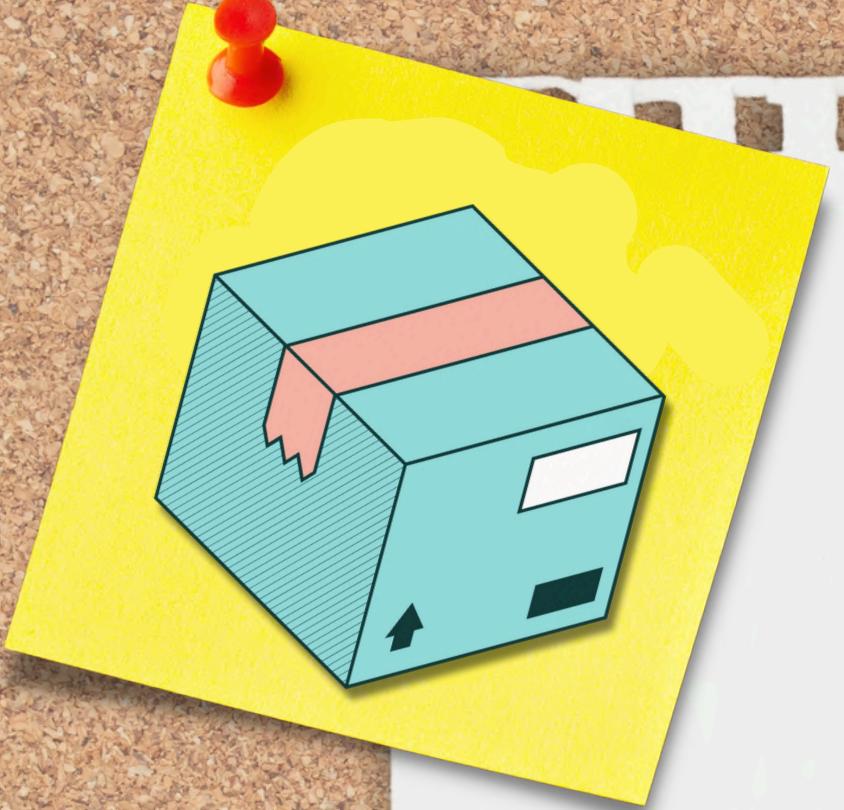
Data Mart ANALYSIS

Presented by



Simala Om Prakash [in](#)
Data Analyst





Introduction

Data Mart Sales Performance Analysis

- Welcome to the presentation on analyzing the sales and performance of Data Mart.
- We'll delve into the impact of sustainable packaging methods implemented in June 2020.
- Let's explore the trends and insights derived from the data.



SCHEMA USED: WEEKLY_SALES TABLE

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

	week_date	region	platform	segment	customer_type	transactions	sales
▶	2020-08-31	ASIA	Retail	C3	New	120631	3656163
	2020-08-31	ASIA	Retail	F1	New	31574	996575
	2020-08-31	USA	Retail	null	Guest	529151	16509610
	2020-08-31	EU	EUROPE	Retail	C1	New	4517
	2020-08-31	AFRICA	Retail	C2	New	58046	1758388
	2020-08-31	CANADA	Shopify	F2	Existing	1336	243878
	2020-08-31	AFRICA	Shopify	F3	Existing	2514	519502
	2020-08-31	ASIA	Shopify	F1	Existing	2158	371417



Data Cleansing Steps

In a single query, perform the following operations and generate a new table in the data_mart schema named clean_weekly_sales:

1. Add a week_number as the second column for each week_date value, for example any value from the 1st of January to 7th of January will be 1, 8th to 14th will be 2, etc.
2. Add a month_number with the calendar month for each week_date value as the 3rd column
3. Add a calendar_year column as the 4th column containing either 2018, 2019 or 2020 values
4. Add a new column called age_band after the original segment column using the following mapping on the number inside the segment value
5. Add a new demographic column using the following mapping for the first letter in the segment values:
segment | demographic |
C | Couples |
F | Families |
6. Ensure all null string values with an "unknown" string value in the original segment column as well as the new age_band and demographic columns
7. Generate a new avg_transaction column as the sales value divided by transactions rounded to 2 decimal places for each record

segment	age_band
1	Young Adults
2	Middle Aged
3 or 4	Retirees



PROJECT QUESTIONS

1. Which week numbers are missing from the dataset?
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?
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?
6. What is the percentage of sales by demographic for each year in the dataset?
7. Which age_band and demographic values contribute the most to Retail sales?



Data Cleansing Steps

- In this first step, I cleaned the data to ensure accuracy and consistency.
- Operations included adding week and month numbers, categorizing age bands and demographics, and handling null values.
- A new table named 'clean_weekly_sales' was created in the data_mart schema.

```
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 is null then 'Unknown'
else segment
end as segment,
case
when right(segment,1)='1' then 'Young Age'
when right(segment,1)='2' then 'Middle Age'
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_sales'
from weekly_sales
```

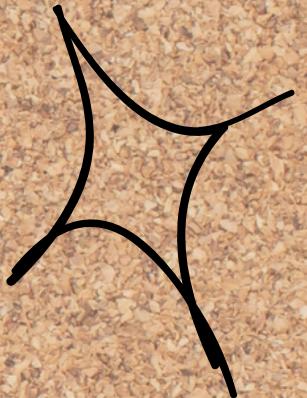
New Table

	week_date	week_number	month_number	calendar_year	region	platform	segment	age_band	demographic	customer_type	transactions	sales	avg_sales
▶	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 Age	Families	New	31574	996575	31.56
	2020-08-31	35	8	2020	USA	Retail	null	Unknown	Unknown	Guest	529151	16509610	31.20
	2020-08-31	35	8	2020	EUROPE	Retail	C1	Young Age	Couples	New	4517	141942	31.42
	2020-08-31	35	8	2020	AFRICA	Retail	C2	Middle Age	Couples	New	58046	1758388	30.29
	2020-08-31	35	8	2020	CANADA	Shopify	F2	Middle Age	Families	Existing	1336	243878	182.54
	2020-08-31	35	8	2020	AFRICA	Shopify	F3	Retirees	Families	Existing	2514	519502	206.64
	2020-08-31	35	8	2020	ASIA	Shopify	F1	Young Age	Families	Existing	2158	371417	172.11
	2020-08-31	35	8	2020	AFRICA	Shopify	F2	Middle Age	Families	New	318	49557	155.84
	2020-08-31	35	8	2020	AFRICA	Retail	C3	Retirees	Couples	New	111032	3888162	35.02
	2020-08-31	35	8	2020	USA	Shopify	F1	Young Age	Families	Existing	1398	260773	186.53
	2020-08-31	35	8	2020	OCEANIA	Shopify	C2	Middle Age	Couples	Existing	4661	882690	189.38
	2020-08-31	35	8	2020	SOUTH ...	Retail	C2	Middle Age	Couples	Existing	1029	38762	37.67
	2020-08-31	35	8	2020	SOUTH ...	Shopify	C4	Retirees	Couples	New	6	917	152.83
	2020-08-31	35	8	2020	EUROPE	Shopify	F3	Retirees	Families	Existing	115	35215	306.22
	2020-08-31	35	8	2020	OCEANIA	Retail	F3	Retirees	Families	Existing	551905	30371770	55.03
	2020-08-31	35	8	2020	ASIA	Shopify	C3	Retirees	Couples	Existing	1969	374327	190.11
	2020-08-31	35	8	2020	AFRICA	Retail	F1	Young Age	Families	Existing	97604	5185233	53.13
	2020-08-31	35	8	2020	OCEANIA	Retail	C2	Middle Age	Couples	New	111219	2980673	26.80
	2020-08-31	35	8	2020	USA	Retail	F1	Young Age	Families	New	11820	463738	39.23
	2020-08-31	35	8	2020	SOUTH ...	Retail	F3	Retirees	Families	Existing	1363	65730	48.22
	2020-08-31	35	8	2020	AFRICA	Retail	C3	Retirees	Couples	Existing	284971	14430196	50.64
	2020-08-31	35	8	2020	ASIA	Retail	F2	Middle Age	Families	New	70496	2176980	30.88

Q1. Missing Week Numbers

```
create table seq100
(x int not null auto_increment primary key);
insert into seq100 values (),(),(),(),(),(),(),();
insert into seq100 select x + 50 from seq100;
select * 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);
```

week_day
1
2
3
4
5
6
7
8
9
10
11
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52



Q2. Total Transactions by Year

```
SELECT  
    calendar_year,  
    SUM(transactions) AS total_transactions  
FROM clean_weekly_sales group by calendar_year;
```

	calendar_year	total_transactions
▶	2020	375813651
	2019	365639285
	2018	346406460

- Analyzing the total transactions per year revealed trends in sales volume over time.
- This insight aids in understanding overall business performance and growth.

Q3. Total Sales by Region and Month

```
select region,month_number,sum(sales) as total_sales  
from clean_weekly_sales  
group by region,month_number
```

- Total sales were analyzed by region and month.
- This breakdown provides valuable insights into regional performance and seasonal variations.

	region	month_number	total_sales
▶	ASIA	8	1663320609
	USA	8	712002790
	EUROPE	8	122102995
	AFRICA	8	1809596890
	CANADA	8	447073019
	OCEANIA	8	2432313652
	SOUTH AMERICA	8	221166052
	AFRICA	7	1960219710
	CANADA	7	477134947
	USA	7	760331754
	EUROPE	7	136757466
	OCEANIA	7	2563459400
	SOUTH AMERICA	7	235582776
	ASIA	7	1768844756
	OCEANIA	6	2371884744
	USA	6	703878990
	SOUTH AMERICA	6	218247455
	EUROPE	6	122813826
	ASIA	6	1619482889
	CANADA	6	443846698
	AFRICA	6	1767559760
	EUROPE	5	109338389
	USA	5	655967121
	AFRICA	9	276320987
	CANADA	9	69067959

Q4. Transactions by Platform

```
select platform,sum(transactions) as total_transactions  
from clean_weekly_sales  
group by platform
```

	platform	total_transactions
▶	Retail	1081934227
	Shopify	5925169

- We examined the total count of transactions for each platform.
- This analysis helps in understanding customer preferences and platform effectiveness.

Q5. Retail vs. Shopify Sales

```
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
)
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

- Percentage analysis of sales for Retail vs. Shopify was conducted for each month.
- This comparison highlights the distribution of sales across different platforms.

Q6. Sales by Demographic

```
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

- Sales percentage by demographic for each year was analyzed.
- Understanding demographic preferences can inform marketing strategies and product targeting.

Q7. Top Contributors to Retail Sales

```
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 Age	Families	4354091554
	Young Age	Couples	2602922797
	Middle Age	Couples	1854160330
	Young Age	Families	1770889293

- Identified top age bands and demographics contributing to Retail sales.
- This insight helps in optimizing marketing efforts and tailoring product offerings.

Conclusion

- The analysis provided valuable insights into Data Mart's sales performance.
- Recommendations based on findings can guide future strategies and initiatives.
- Continuous monitoring and analysis are crucial for maintaining competitive advantage.



Thanks!

@Simala Om Prakash

