PORTFOLIO PROJECT REPORT

Confectionery sales Data

Prepared by: Saima Naz
Palwasha khalid
Tayyaba
Approved by:
Advisor: Hadiqa

Business overview:

This is a dynamic enterprise **operating in the heart of New York State,** where it has successfully established **8 retail outlets, a well-equipped warehouse, and a centralized headquarters.** These strategic locations facilitate a seamless supply chain and operational efficiency.

Dedicated to serving the local community, each retail outlet have its own outlet manager, responsible for overseeing day-to-day operations and ensuring customer satisfaction. **A team of 55 individuals,** comprising skilled managers and dedicated field workers, collaborates across all outlets.

This business is really good at offering lots of different things to customers. They organize their **product into five special groups.** Inside each group, there are categories, and in each category, there are different types of things to buy. And inside those types, there are individual products.

In a recent effort to enhance the joy of tasting, business introduced **two fresh treats: "Happy Holidays Hot Chocolate,"** a delightful seasonal indulgence found within the beverage collection (product category: drinking chocolate; product type: seasonal drink), and **"Rio Night"** a superb creation highlighted in the beverage group (product category: coffee; product type: specialty coffee).

In the spirit of continuous improvement, they **sets monthly sales targets for each product group within every outlet.** These targets not only drive the business forward but also maintain the brand's dedication to delivering quality and innovation, except in the case of Add-ons.

About DATA:

We possess a data-set related to a confectionery bakery. Within this data-set, there are nine tables that contain information exclusively from the month of April 2019. Across these tables, the combined total of columns reaches 79. Furthermore, our data-set exclusively covers data from three specific outlets: Outlet 3, Outlet 5, and Outlet 8.

Scope of project:

Our project's focus is to explore the performance of three specific sales outlets: Outlet 3, Outlet 5, and Outlet 8, during the month of April 2019. We aim to understand how these outlets performed in terms of sales. Additionally, our goal is to analyze how clients were distributed among generation categories.

Scope Exclusions:

The following aspects are not within the scope of this project:

- Detailed analysis of individual product categories or items.
- Investigation of factors beyond the month of April 2019.
- Analysis of the performance of outlets other than Outlet 3, Outlet 5, and Outlet 8.

Project Deliverables:

- Detailed report summarizing the sales performance and client generation distribution analysis for outlets 3, 5, and 8 in April 2019.
- Visualizations and charts depicting trends and insights.
- Actionable recommendations for each outlet.

Data Sources

- Data sourced provide by the guidance of the project lead.

Purpose of project:

- The data is intended solely for research purposes, without any claims or assertions.

Stackholder of project:

- We imagin the stakeholders of this project to be the sales and marketing teams.

Problem statment:

The goal of our research is to determine from sales outlet 3, 5 and 8 which have successfully achieved or not achieved their sales goals. Additionally, we aim to uncover the reasons behind both success and failure. Furthermore, we seek to analyze the distribution of client generations among our clients, providing valuable insights for future sales achievement.

Objective:

Sales target:

Question 1: Which sales outlets successfully met their total quantity sold goals in the month of April?

Question 2: How well did each sales outlet perform in achieving their monthly sales goals across various product categories (beans, beverages, food, merchandise)?

Product potential:

To assess product potential, we can examine consumer buying behavior. One approach is to determine whether we have repeat customers, as customers tend to return only if they have a favorable impression of the product. If a customer doesn't make a return visit, it might indicate a lack of product appeal.

Question 3: Which top 10 products are consistently purchased by repeat customers?

Question 4: Which products are consistently purchased by repeat customers, And to which product groups do these products belong?

Question 5: compare the products that are frequently purchased by repeat clients within each product group, along with the total number of products in each group?

Client Generation Distribution:

Question 6: What is the distribution of clients across different generation categories in april-2019? Which generation category has the highest representation among clients in april-2019?

Question 7: How many years does each generation category span, and what is the average number of customers per year within each category?

Question 8: What is the count of avg repeat customers within each generation category?

Problem in our data which we identify during data cleaning process

Controlable factor: we change them

☑ Abbreviation of
- HQ
- WH

☑ **Yn**_stand for

☑ **Manager staff id** is missing in sales outlet table ID 2

☑ There are four measurement types in the **unit of measurement**.

☑ The **line item amount** is a derived column calculated as the unit price multiplied by the quantity. While **99.5% of the values are correct, 0.5%** of the values do not display the correct values.

☑ The **unit prices** have some errors. Please cross-check them with the unit prices listed in the product table. For example, there is a case where one product has two records."

☑ In the "Sales Target" table, the column named **"Merchandise Goal"** contains an additional space between the words "Merchandise" and "Goal." We need to correct this spacing discrepancy in the column name.

☑ We are adjusting the **data type** to align with our data dictionary's specifications.

☑ Establish a **primary key** and define **foreign key** relationships.

Uncontrolable factor:

we don't making any adjustments or modifications to factors that is beyond our control.

☐ In the sales table, the **sales outlet column shows only IDs 3, 5, and 8.** However, the staff members listed for other outlets seem to contradict this information.

☐ In the sales table, the **staff IDs 11, 21, 22, 23, 24,** and **31 to 40** never appear. Unfortunately, staff IDs **31 to 40 belong to outlets 8 and 9.** This implies that we lack data for outlets 8 and 9, which contradicts the statement above.

☐ **50% of customer IDs** have a value of 0.(also data missing in customer table)

☐ The **'instore_yn'** does not make sense.

☐ The **demo data in the pastry_table table**, for example, shows that the quantity we sold for product 69 never matches the sales quantity sold on the specific date. This inconsistency also leads to incorrect waste and percentage values. Both of these columns are derived columns.

☐ Three tables have **no primary keys** and do not require them.

☐ In the product table, please verify the **splitting of product IDs 69 and 75** by cross-referencing the pastry inventory.

☐ The **promotional items** differ in two columns.

☐ The date of **April 30th** is missing.

☐ **Line_item_id** column have no relation with any table

☐ Abbreviation of **"FL"**

Data Dictionary

Sales table

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraints
transaction_id	The identifier provided for the transaction.	INT	
transaction_date	The date when the transaction occurred.	DATE	
transaction_time	The time at which the transaction was recorded.	TIME	
sales_outlet_id	Identifier for the sales outlet where the transaction took place.	INT	FK
staff_id	Identifier for the staff member processing the transaction.	INT	FK
customer_id	Identifier for the customer involved in the transaction.	INT	FK
instore_yn	Indicates whether the transaction was made in-store (Yes/No)	ENUM ('n','y')	
orders	Identifier for the overall order associated with the transaction.	INT	
product_id	An identifier for the product sold in the transaction.	INT	FK
Quantity	The quantity of the product sold in the transaction.	INT	
line_item_amount	Total amount of the transaction, obtained by multiplying the unit price by the quantity of items purchased.	DECIMAL (10, 2)	
unit_price	The price per unit of the item.	DECIMAL (10, 2)	
promo_item_yn	This indicates whether an item is a promotional item or not.	ENUM ('n','y')	

Sales outlet

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraints
sales_outlet_id	A unique identifier for each sales outlet.	INT	PK
sales_outlet_type	The categorization or type of the sales outlet (e.g., retail store, headquarter, warehouse etc.).	VARCHAR (250)	
store_square_feet	The size of the sales outlet in square feet.	INT	
store_address	The street address of the sales outlet.	VARCHAR (250)	
store_city	The city where the sales outlet is located.	VARCHAR (250)	
store_state_province	The state or province where the sales outlet is located.	VARCHAR (250)	
store_telephone	The telephone number of the sales outlet.	VARCHAR (12)	
store_postal_code	The postal code or ZIP code of the sales outlet's location.	VARCHAR (10)	
store_longitude	The longitude coordinate of the sales outlet's location.	DECIMAL (10, 2)	
store_latitude	The latitude coordinate of the sales outlet's location.	DECIMAL (10, 2)	
manager	The manager unique staff id who is responsible for the sales outlet.	INT	FK
Neighborhood	The neighborhood city where the sales outlet is situated.	VARCHAR (250)	

Pastry inventory

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraints
sales_outlet_id	A unique identifier for each sales outlet.	INT	FK
transaction_date	The date when the transaction occurred.	DATE	
product_id	A unique identifier for each product.	INT	FK
start_of_day	The quantity available when the sales outlet opened for the day.	INT	
quantity_sold	The quantity of the product sold during the day.	INT	
waste	The quantity of the product that went to waste	INT	
% waste	The percentage of waste calculated based on the wast and total quantity available.	DECIMAL (10, 2)	

Customer table

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraints
Customer id	A unique identifier for each customer	INT	PK
Home store	Identifier of the sales outlet where the customer made purchases	IN	FK
customer_first-name	The first name of the customer	VARCHAR (250)	
customer_email	The email address associated with the customer's account.	VARCHAR (50)	
customer_since	The date when the customer's account was created	DATE	
loyalty_card_number	The unique identifier of the customer's loyalty card.	VARCHAR (12)	
birthdate	The date of birth of the customer.	DATE	
gender	The gender of the customer.	ENUM	
Birth year	The birth year of the customer.	DATE	FK

Dates

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or freeform text)	Constraints
transaction_date	The date when the transaction occurred.	DATE	
Date_ID	A unique identifier for each date.	INT	PK
Week_ID	A unique identifier for each week in a calendar year.	INT	
Week_Desc	A descriptive label for the week, often containing the week number	VARCHAR (250)	
Month_ID	A unique identifier for each month in a calendar year.	INT	
Month_Name	The name of the month.	VARCHAR (250)	
Quarter_ID	A unique identifier for each quarter in a calendar year.	INT	
Quarter_Name	The name of the quarter.	VARCHAR (250)	
Year_ID	A unique identifier for each calendar year.	INT	

Sales target table

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraint
sales_outlet_id	A unique identifier for each sales outlet.	INT	FK
year_month	The specific year and month to which the sales goals apply.	DATE	
beans_goal	The sales target for the "beans" group for the specified month.	INT	
beverage_goal	The sales target for the "beverage" group for the specified month.	INT	
food_goal	The sales target for the "foods" group for the specified month.	INT	
merchandise _goal	The sales target for the "merchandise" group for the specified month.	INT	
total_goal	The overall sales target for the sales outlet for the specified month, summing up all category goals.	INT	

staff

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or free form text)	Constraints
Staff_id	A unique identifier for each staff member.	INT	PK
First name	The first name of the staff member.	VARCHAR (250)	
Last name	The last name of the staff member.	VARCHAR (250)	
position	The job title or position held by the staff member.	VARCHAR (250)	
Start date	The date when the staff member began employment.	DATE	
location	The unique identifier of the sales outlet where the staff member is work.	INT	FK

Generation

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or freeform text)	Constraints
Birth year	The year in which an individual was born.	DATE	PK
generation	A label representing the generation to which an individual belongs (e.g., Baby Boomer, Generation X, Millennial, etc.).	VARCHAR (250)	

Product

Field name	Description	Data Type/ Format (e.g., numeric, date, currency, string or freeform text)	Constraint
product_id	A unique identifier for each product.	INT	PK
product_group	A broader classification that groups related products together.	VARCHAR (250)	
product_category	A more specific classification that categorizes products within a group.	VARCHAR (250)	
product_type	The type or variant of the product.	VARCHAR (250)	
product	The name or title of the specific product.	VARCHAR (250)	
product_description	A brief description of the product's features or characteristics.	VARCHAR (250)	
unit_of_measure	The unit used to measure the product's quantity (e.g single, pump etc.).	VARCHAR (250)	
measurement	This field represents the measurement of liquid in ounce	DECIMAL (10, 2)	
current_wholesale_price	The current whole sale price	DECIMAL (10, 2)	
current_retail_price	The current price at which the product is sold to retail customers.	DECIMAL (10.2)	
tax_exempt_yn	Indicates whether the product is tax-exempt (Yes) or not (No).	ENUM	
promo_yn	Indicates whether the product is a promotional item (Yes) or not (No).	ENUM	
new_product_yn	Indicates whether the product is a new addition (Yes) or not (No).	ENUM	

ERD EXPLANATION:

One-to-Many Relationship: each sales outlet can be associated with multiple sales, but each sale is associated with only one sales outlet.

Many-to-One Relationship: each sale have one and only customer, and each customer can have multiple sales.

One_to_many Relationship: each customer can be associated with only one sales outlet, but each sales outlet can serve multiple customers.

one-to-many relationship: each generation is associated with multiple customers generation can have multiple customers associated with it based on their birth year.

one-to-one Relationship where each manager is associated with exactly one sales outlet, and each sales outlet is managed by exactly one manager.

Many-to-One Relationship: multiple employees can work at the same sales outlet, but each employee works at only one sales outlet.

Many-to-One Relationship: multiple sales recipes can be linked to a single staff member (employee), but each sales recipe corresponds to only one staff member.

Many-to-One Relationship: multiple sales recipes can be associated with one product, but each sales recipe corresponds to only one specific product.

One-to-Many Relationship: each product can have multiple inventory entries, but each inventory entry corresponds to only one specific product.

One-to-Many Relationship: each sales outlet can have multiple sales targets, but each sales target corresponds to only one specific sales outlet.

One-to-Many Relationship: each sales outlet can have multiple entries in the pastry inventory, but each inventory entry corresponds to only one specific sales outlet.
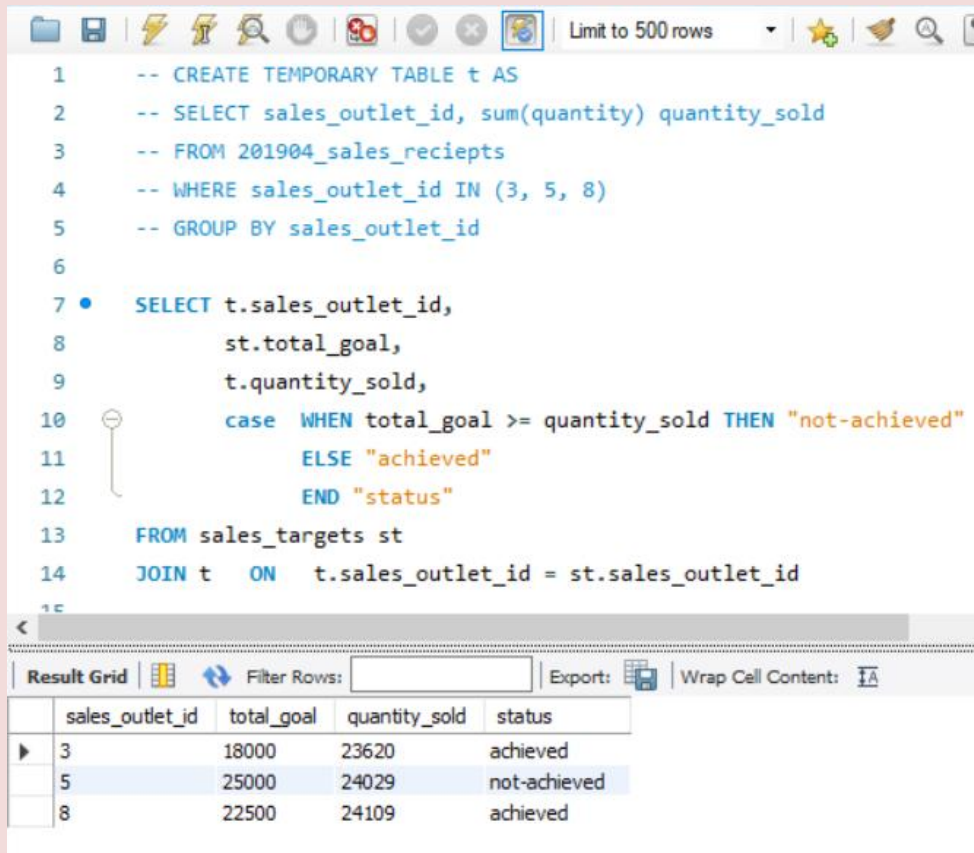
ERD:

This diagram presents the Entity-Relationship structure of my database, highlighting the interconnections between various tables that encompass business-related information. It visually communicates how these tables are linked and interact, offering a comprehensive view of the relationships that define the business data within the system.
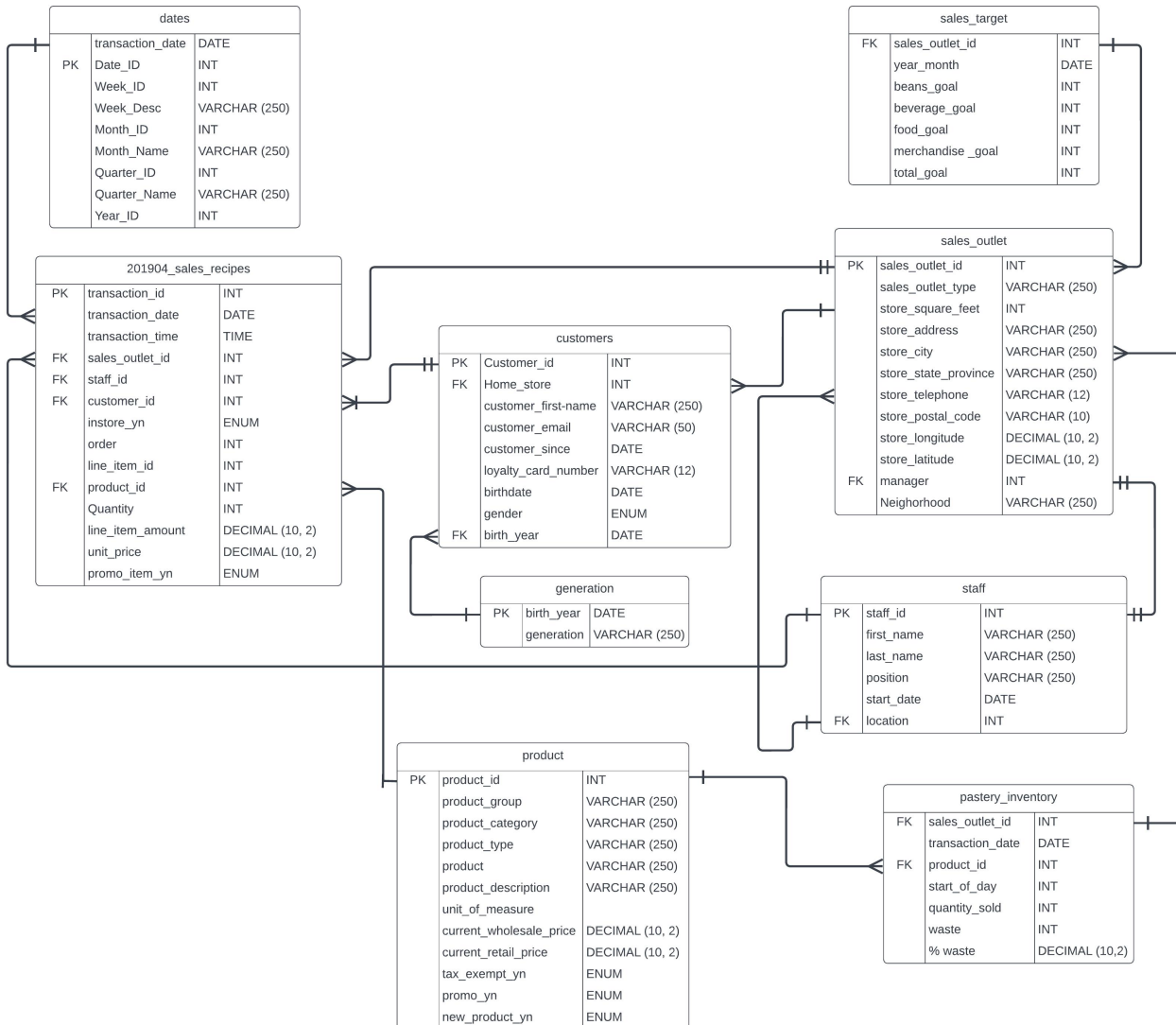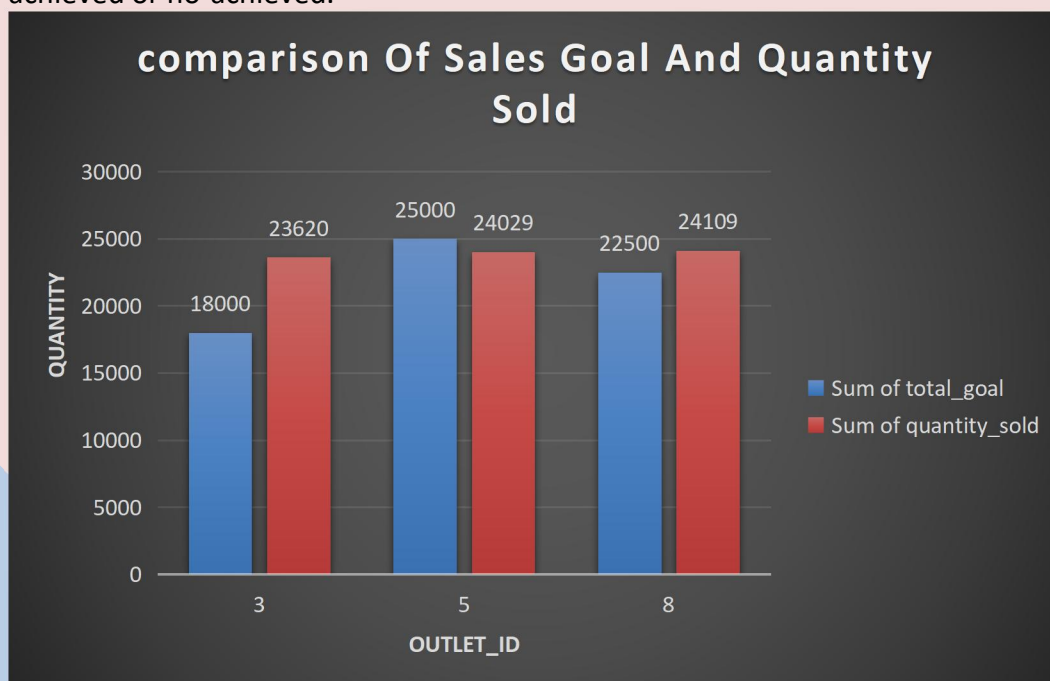
Queries:

Question 1: Which sales outlets successfully met their total quantity sold goals in the month of April?

```sql
1    -- CREATE TEMPORARY TABLE t AS
2    -- SELECT sales_outlet_id, sum(quantity) quantity_sold
3    -- FROM 201904_sales_reciepts
4    -- WHERE sales_outlet_id IN (3, 5, 8)
5    -- GROUP BY sales_outlet_id
6
7  • SELECT t.sales_outlet_id,
8         st.total_goal,
9         t.quantity_sold,
10    case  WHEN total_goal >= quantity_sold THEN "not-achieved"
11              ELSE "achieved"
12              END "status"
13    FROM sales_targets st
14    JOIN t   ON   t.sales_outlet_id = st.sales_outlet_id
```

sales_outlet_id	total_goal	quantity_sold	status
3	18000	23620	achieved
5	25000	24029	not-achieved
8	22500	24109	achieved

First I create a temporary **table 't'** to find total quantities sold for **sales outlets 3, 5, and 8** from the **'201904_sales_reciepts'** table. The main query joins 't' with 'sales_targets', calculating goal achievement status. The result shows outlet IDs, total goals, quantities sold, and whether goals are achieved or no-achieved.

Insight: it's notable that sales outlets 3 and 8 have successfully met their sales goals, whereas outlet 5 has fallen short. To gain deeper insights, we will conduct an analysis based on product groups. This analysis aims to identify the specific product groups that have contributed to the achievements of outlets 3 and 8, as well as to ascertain the factors that might have contributed to the under performance of outlet 5.

Question 2: How well did each sales outlet perform in achieving their monthly sales goals across various product group (beans, beverages, food, merchandise)?

Initially, the sales target data is structured in a pivoted table format. To process this data, we begin by transforming it from its pivoted state into a more conventional tabular format. This transformation is carried out within a temporary table named "t1."

I'm unpivoting the data because I need to create a specific set of sales data for outlet IDs 3, 5, and 8, grouped by each product category. I'm using a "union" operation because SQL doesn't provide a built-in way to directly convert columns into rows.

```
1 •    SELECT *   FROM t1
2      -- CREATE temporary TABLE t1 as
3      -- SELECT sales_outlet_id, product_group , goal
4      -- FROM (
5      --        SELECT   sales_outlet_id,      'beans_goal' AS product_group ,       beans_goal  AS goal         FROM sales_targets
6      --        UNION ALL
7      --        SELECT   sales_outlet_id,      'beverage_goal' AS product_group ,    beverage_goal  AS goal      FROM sales_targets
8      --        UNION ALL
9      --        SELECT   sales_outlet_id,      'food_goal' AS product_group ,        food_goal  AS goal          FROM sales_targets
10     --        UNION ALL
11     --        SELECT   sales_outlet_id,      'merchandise_goal' AS product_group,  merchandise_goal AS goal    FROM sales_targets
12     --        ) sub
13     -- WHERE sales_outlet_id  IN ( 3, 5, 8)
14     -- ORDER BY  1, 2
15
```

sales_outlet_id	product_group	goal
3	beans_goal	720
3	beverage_goal	13500
3	food_goal	3420
3	merchandise_goal	360
5	beans_goal	1000
5	beverage_goal	18750
5	food_goal	4750

Next, I'm calculating the total quantity of items sold by sales outlets 3, 5, and 8 within each product group.

```
24 •    SELECT * FROM t2
25
26      -- UPDATE t2
27      -- SET product_group =  CASE
28      --                WHEN   product_group = 'Beverages'        THEN   'beverage_goal'
29      --                WHEN   product_group = 'Food'             THEN   'food_goal'
30      --                WHEN   product_group = 'Merchandise'      THEN   'merchandise_goal'
31      --                WHEN   product_group = 'Whole Bean/Teas'  THEN   'beans_goal'
32      --                  ELSE 0
33      --                  END;
```

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

sales_outlet_id	product_group	quantity_sold
3	beverage_goal	20001
3	food_goal	2489
3	merchandise_goal	94
3	beans_goal	313
5	beverage_goal	19282
5	food_goal	2694
5	merchandise_goal	125
5	beans_goal	362
8	beverage_goal	19822
8	food_goal	2581
8	merchandise_goal	36
8	beans_goal	493

Now we've gathered data about the sales goals and the quantity of items sold for business outlets with IDs 3, 5, and 8 in each product group. We've stored this information in two separate temporary tables. Now, I'm planning to combine these tables by matching the sales outlet IDs and product groups. This way, we can compare the data and determine whether we've achieved our goals or not.

However, while preparing for this comparison, we identified a problem with the product group column. It appears that the product group names are not consistent across both tables. They're displayed with varying names and different orders. This inconsistency could potentially affect the accuracy of our merging process. To address this issue, we plan to update the temporary tables "t2" and correct the inconsistencies in the product group column before proceeding with the data merging and comparison

```
35    -- CREATE TEMPORARY TABLE t3
36    -- SELECT *
37    -- FROM t2
38    -- ORDER BY  sales_outlet_id, product_group
39 •     SELECT * FROM t3
```

Result Grid | Filter Rows: | Export: | Wrap

sales_outlet_id	product_group	quantity_sold
3	beans_goal	313
3	beverage_goal	20001
3	food_goal	2489
3	merchandise_goal	94
5	beans_goal	362
5	beverage_goal	19282
5	food_goal	2694
5	merchandise_goal	125
8	beans_goal	493
8	beverage_goal	19822
8	food_goal	2581
8	merchandise_goal	36

FINAL RESULT TABLE IS BELOW, ABOVE ALL QUERRY ARE JUST PROCESS

```
41
42 •  SELECT t1.sales_outlet_id ,
43         t1.product_group,
44         t1.goal,
45         t3.quantity_sold,
46         CASE  WHEN t1.goal >= t3.quantity_sold THEN "not-achieved"     ELSE "achieved" END AS "status"
47    FROM t1
48    JOIN t3    ON   t1.sales_outlet_id = t3.sales_outlet_id   AND  t1.product_group = t3.product_group
49
50
51
```
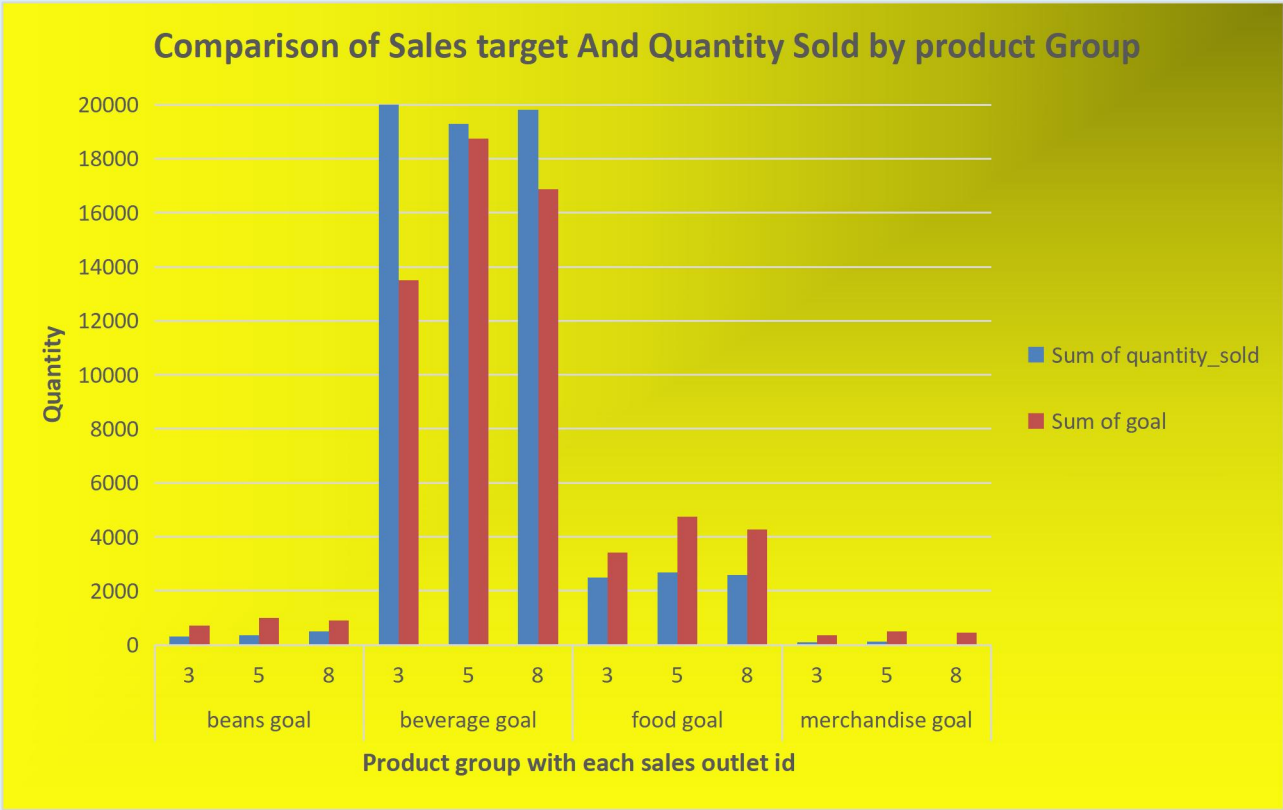
Result Grid | Filter Rows: | Export: | Wrap Cell Content: IA

sales_outlet_id	product_group	goal	quantity_sold	status
3	beans_goal	720	313	not-achieved
3	beverage_goal	13500	20001	achieved
3	food_goal	3420	2489	not-achieved
3	merchandise_goal	360	94	not-achieved
5	beans_goal	1000	362	not-achieved
5	beverage_goal	18750	19282	achieved
5	food_goal	4750	2694	not-achieved
5	merchandise_goal	500	125	not-achieved
8	beans_goal	900	493	not-achieved
8	beverage_goal	16875	19822	achieved
8	food_goal	4275	2581	not-achieved
8	merchandise_goal	450	36	not-achieved

Insights: It is evident from our analysis that the **beverage product group is the sole achiever** of its sales goals **across all sales outlets.** In contrast, other product categories have fallen short of their targets.

This points to the fact that the achievements of outlets 3 and 8 can be primarily attributed to the exceptional performance of the beverage group. This success has managed to offset the shortcomings of the other product groups. While we have attained our monthly sales targets for outlets 3 and 5, we must not overlook the complete lack of success in the remaining product groups.

In the case of outlet 5, the beverage group has demonstrated commendable performance; however, its success in this area alone cannot compensate for the overall under performance of the other product groups. This discrepancy is the key reason behind our inability to attain the total sales goal. To enhance the sales performance of these other product groups.
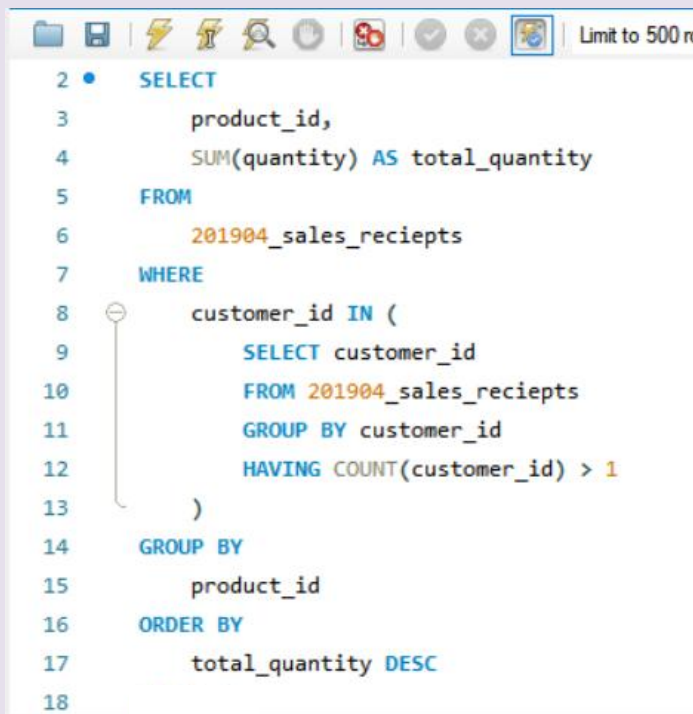

Comparison of Sales target And Quantity Sold by product Group

Question 3:Which products are consistently purchased by repeat customers and their quantity?

```sql
2 •   SELECT
3           product_id,
4           SUM(quantity) AS total_quantity
5     FROM
6           201904_sales_reciepts
7     WHERE
8       ⊖   customer_id IN (
9               SELECT customer_id
10              FROM 201904_sales_reciepts
11              GROUP BY customer_id
12              HAVING COUNT(customer_id) > 1
13          )
14    GROUP BY
15          product_id
16    ORDER BY
17          total_quantity DESC
18
```

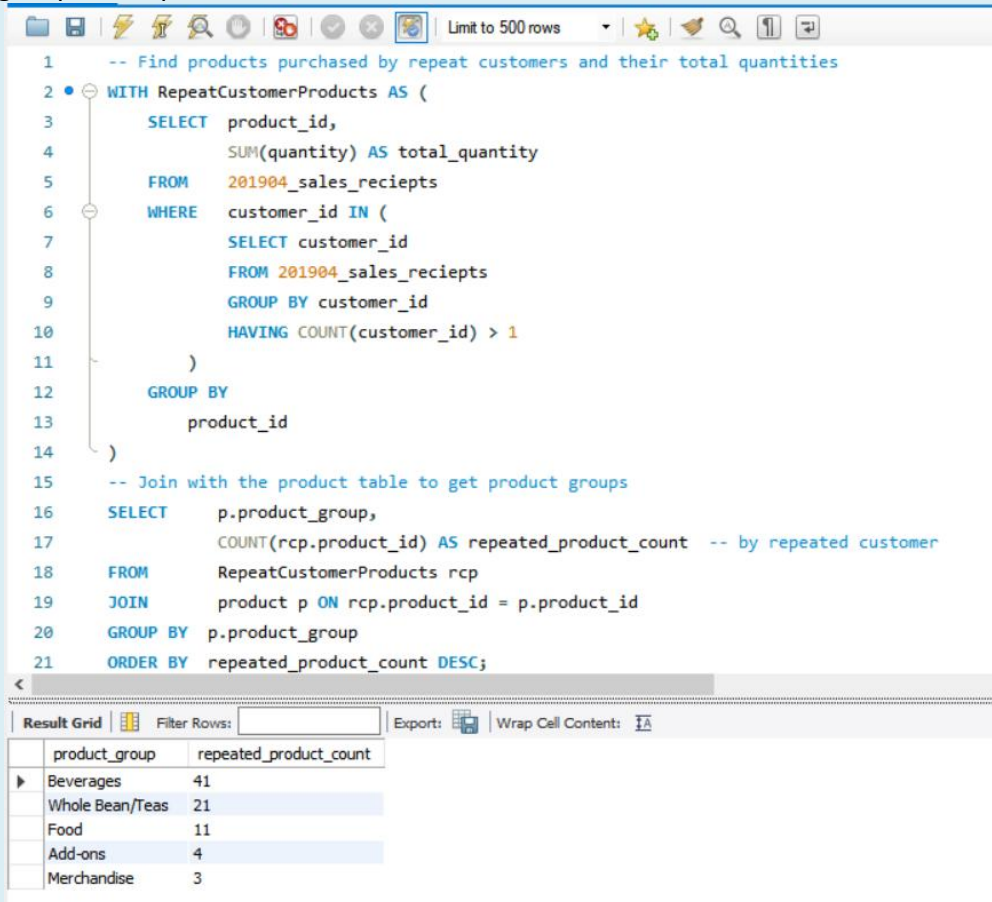| product_id	total_quantity
54	799
38	795
52	793
50	791
29	784
22	781
25	779
42	775

Initially, we identify products purchased by repeat customers. We then analyze which product groups these purchases encompass. This groundwork will enable us to support our point in the upcoming question.

NOTE: without limit 80 row comes that means 80 product buy by repeat customer and total no of product are 88.
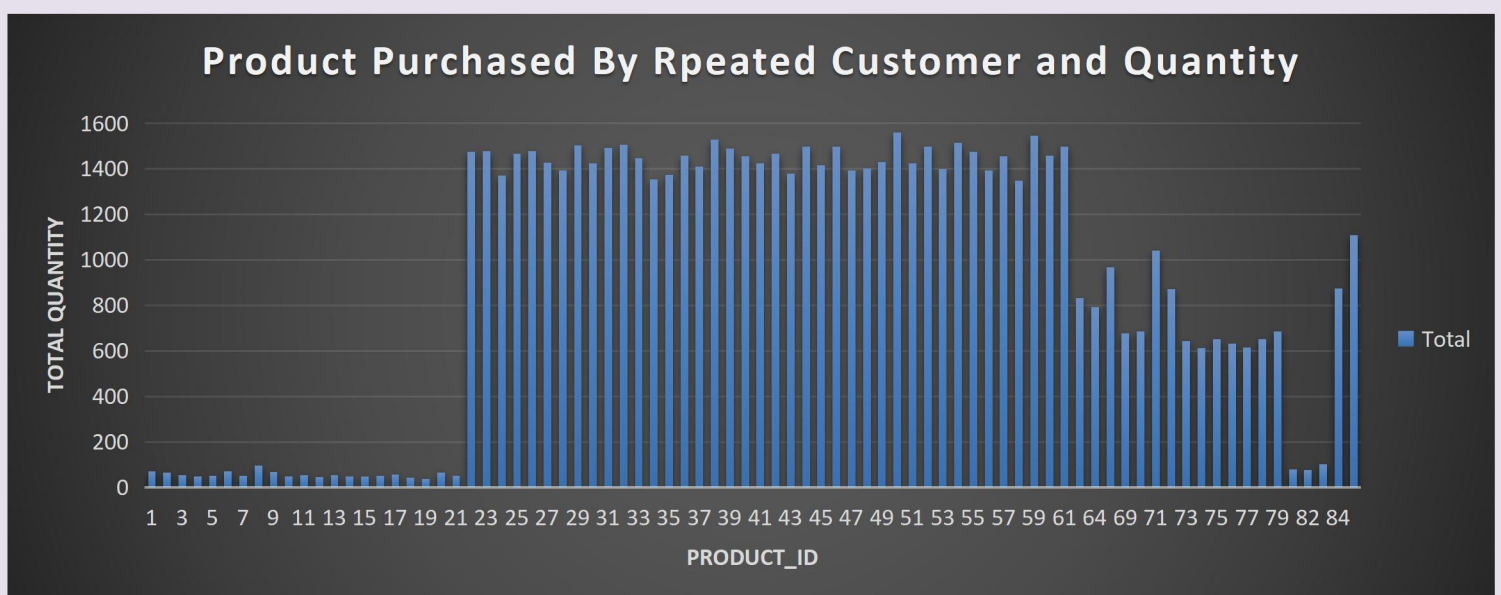
Product Purchased By Rpeated Customer and Quantity

Question 4: Which products are consistently purchased by repeat customers, And to which product groups do these products belong?
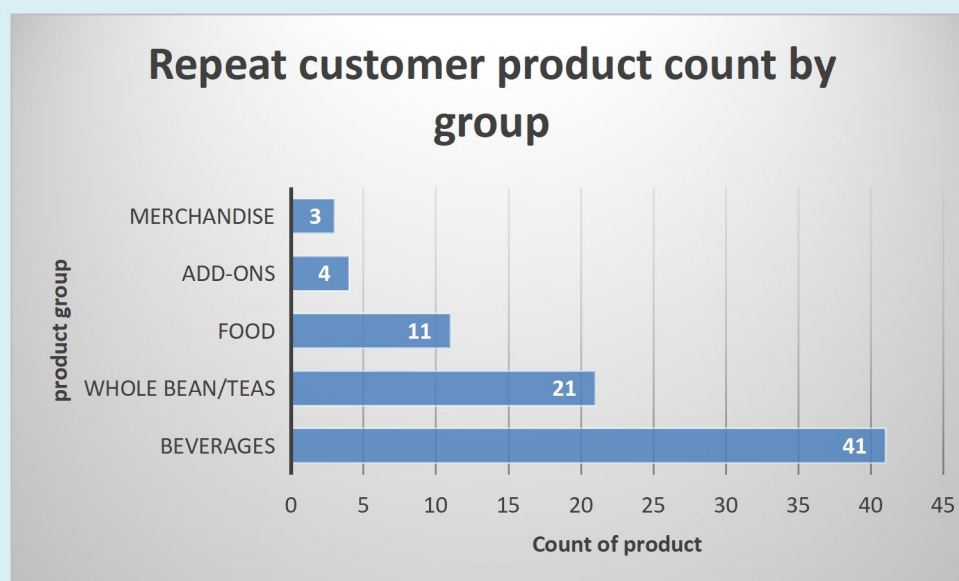
aims to identify specific products that are frequently bought by the same customers more than once. It then seeks to understand which product groups these products are categorized under. This analysis helps in identifying products with a strong customer base that returns for repeat purchases and allows you to group these products based on their characteristics.

```sql
-- Find products purchased by repeat customers and their total quantities
WITH RepeatCustomerProducts AS (
    SELECT  product_id,
            SUM(quantity) AS total_quantity
    FROM    201904_sales_reciepts
    WHERE   customer_id IN (
            SELECT customer_id
            FROM 201904_sales_reciepts
            GROUP BY customer_id
            HAVING COUNT(customer_id) > 1
        )
    GROUP BY
        product_id
)
-- Join with the product table to get product groups
SELECT      p.product_group,
            COUNT(rcp.product_id) AS repeated_product_count   -- by repeated customer
FROM        RepeatCustomerProducts rcp
JOIN        product p ON rcp.product_id = p.product_id
GROUP BY    p.product_group
ORDER BY    repeated_product_count DESC;
```

product_group	repeated_product_count
Beverages	41
Whole Bean/Teas	21
Food	11
Add-ons	4
Merchandise	3

Secondly, we are associating the count of product IDs of items purchased by repeat customers with their corresponding product groups.

Repeat customer product count by group

product group	Count of product
MERCHANDISE	3
ADD-ONS	4
FOOD	11
WHOLE BEAN/TEAS	21
BEVERAGES	41

Question 5: compare the products that are frequently purchased by repeat clients within each product group, along with the total number of products in each group?

```
31      -- CREATE TEMPORARY TABLE t AS
32      -- SELECT
33      --     product_group,
34      --         COUNT(product_id) as total_product_in_group
35      -- FROM product
36      -- GROUP BY product_group
37
38  •   SELECT  t2.product_group,
39              t.total_product_in_group,
40              t2.repeated_product_count
41      FROM t2
42      JOIN t              ON t2.product_group = t.product_group
```
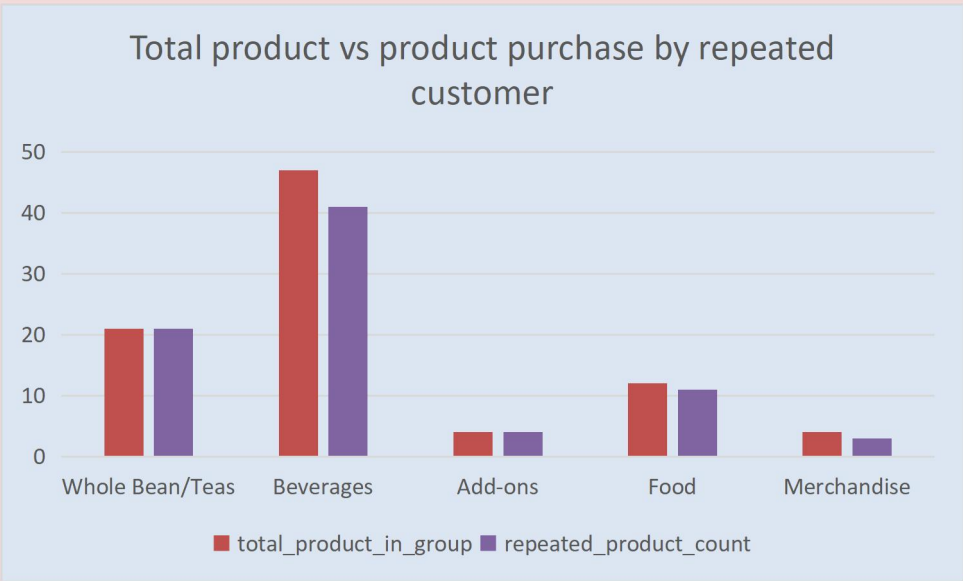
product_group	total_product_in_group	repeated_product_count
Whole Bean/Teas	21	21
Beverages	47	41
Add-ons	4	4
Food	12	11
Merchandise	4	3

Here the t2 is a temporary table which I make from the result of question 7.

Insight:

Lastly, we compare the total number of products within each product group to those products purchased by repeat clients. Notably, we observe that nearly all of the products have been bought by repeat clients. This trend signifies the potential of our products, as the presence of repeat clients indicates a positive product experience and suggests that our products do not possess any major issues.

Question 6: What is the distribution of clients across different generation categories in april-2019? Which generation category has the highest representation among clients in april-2019?

```sql
1   SELECT
2       CONCAT(MIN(g.birth_year), '-', MAX(g.birth_year)) AS year_range,
3       g.generation,
4       COUNT(sr.customer_id) AS no_of_customer
5   FROM
6       201904_sales_reciepts sr
7   JOIN
8       customer c ON c.customer_id = sr.customer_id
9   JOIN
10      generations g ON g.birth_year = c.birth_year
11  GROUP BY
12      g.generation
13  ORDER BY
14      FIELD(g.generation, 'Baby Boomer', 'Gen X', 'Older Millennials', 'Younger Millennials', 'Gen Z');
15
```
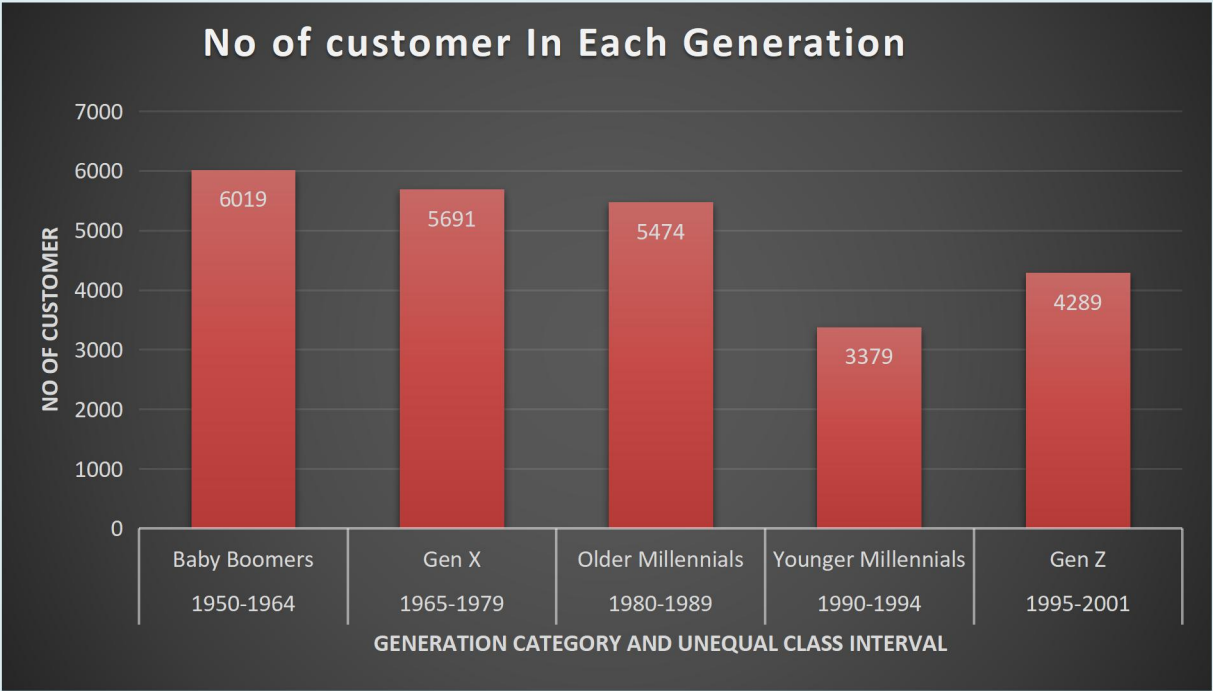
year_range	generation	no_of_customer
1950-1964	Baby Boomers	6019
1965-1979	Gen X	5691
1980-1989	Older Millennials	5474
1990-1994	Younger Millennials	3379
1995-2001	Gen Z	4289

Result 4 ×

Insight: we can observe that 'Younger Millennial's have the fewest number of customers, and this observation holds true. However, it overlooks the fact that **'Younger Millennial's have the narrowest range of years, only four years,** in comparison to the 'Baby Boomers' who have a range of fourteen years. Let's explore our data further and attempt to mitigate these limitations.

No of customer In Each Generation

Generation	Year Range	No of Customer
Baby Boomers	1950-1964	6019
Gen X	1965-1979	5691
Older Millennials	1980-1989	5474
Younger Millennials	1990-1994	3379
Gen Z	1995-2001	4289

GENERATION CATEGORY AND UNEQUAL CLASS INTERVAL

Question 7: How many years does each generation category span, and what is the average number of customers per year within each category?
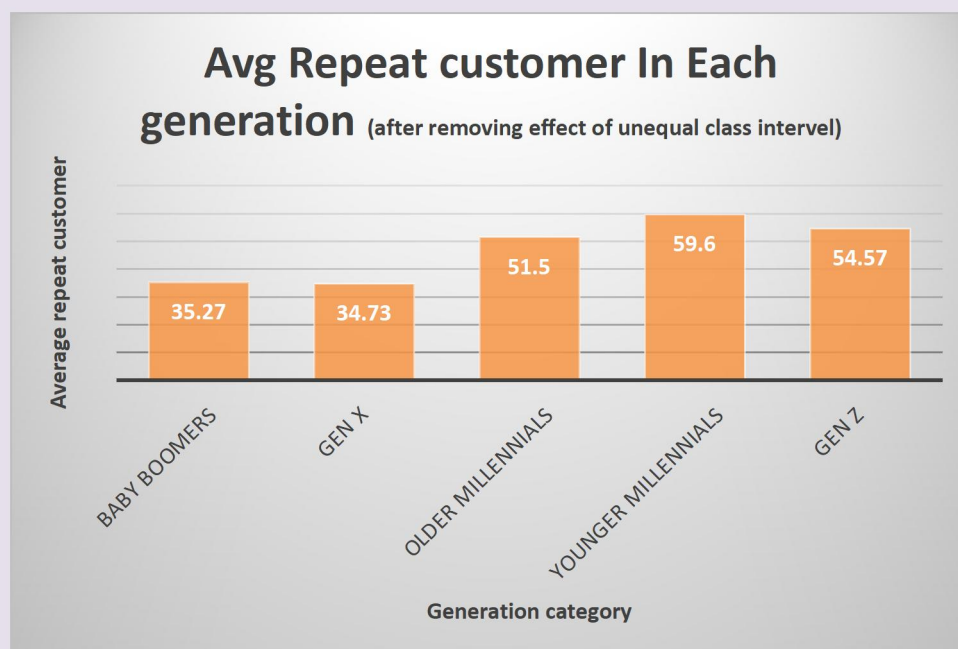
```sql
SELECT
    CONCAT(MIN(g.birth_year), '-', MAX(g.birth_year))                    AS year_range,
    (MAX(g.birth_year) - MIN(g.birth_year) + 1)                         AS span_years,
    g.generation,
    COUNT(sr.customer_id)                                               AS no_of_customer,
    ROUND((COUNT(sr.customer_id)/ (MAX(g.birth_year) - MIN(g.birth_year) + 1)),2)  AS average_customer
FROM
    201904_sales_reciepts sr
JOIN
    customer c ON c.customer_id = sr.customer_id
JOIN
    generations g ON g.birth_year = c.birth_year
GROUP BY
    g.generation
ORDER BY
    FIELD(g.generation, 'Baby Boomer', 'Gen X', 'Older Millennials', 'Younger Millennials', 'Gen Z');
```

year_range	span_years	generation	no_of_customer	average_customer
1950-1964	15	Baby Boomers	6019	401.27
1965-1979	15	Gen X	5691	379.40
1980-1989	10	Older Millennials	5474	547.40
1990-1994	5	Younger Millennials	3379	675.80
1995-2001	7	Gen Z	4289	612.71

INSIGHT: Now, the distinction becomes evident when looking at the average figures. Among the categories, **'Younger Millennial' stand out with the highest average number of customers. And Gen Z in 2nd number.** Also, it's noticeable that **'Baby Boomers' have the 2nd smallest average** number of customers.

Avg Repeat customer In Each generation (after removing effect of unequal class intervel)

| Generation category	Average repeat customer
BABY BOOMERS	35.27
GEN X	34.73
OLDER MILLENNIALS	51.5
YOUNGER MILLENNIALS	59.6
GEN Z	54.57

Question 8: What is the count of avg repeat customers within each generation category?


```sql
1 •   SELECT
2         CONCAT(MIN(g.birth_year), '-', MAX(g.birth_year))            AS year_range,
3         (MAX(g.birth_year) - MIN(g.birth_year) + 1)                  AS span_years,
4         g.generation,
5         COUNT( DISTINCT sr.customer_id )    AS repeat_customer ,
6         ROUND((COUNT(DISTINCT sr.customer_id)/ (MAX(g.birth_year) - MIN(g.birth_year) + 1)),2)  AS avg_repeat_customer
7     FROM
8         201904_sales_reciepts sr
9     JOIN
10        customer c ON c.customer_id = sr.customer_id
11    JOIN
12        generations g ON g.birth_year = c.birth_year
13    GROUP BY
14        g.generation
15    ORDER BY
16        FIELD(g.generation, 'Baby Boomer', 'Gen X', 'Older Millennials', 'Younger Millennials', 'Gen Z');
17
```

year_range	span_years	generation	repeat_customer	avg_repeat_customer
1950-1964	15	Baby Boomers	529	35.27
1965-1979	15	Gen X	521	34.73
1980-1989	10	Older Millennials	515	51.50
1990-1994	5	Younger Millennials	298	59.60
1995-2001	7	Gen Z	382	54.57

REASONE: for finding the avg_repeat_customer we can divide the repeat_customer by span year because our class interval is unequal so here we use the relative frequency concept.

Use Relative Frequencies: Instead of using the actual frequencies in each interval, calculate the relative frequencies (proportions) by dividing the frequency in each interval by the total number of observations. This way, you can compare the proportions across intervals regardless of their size.
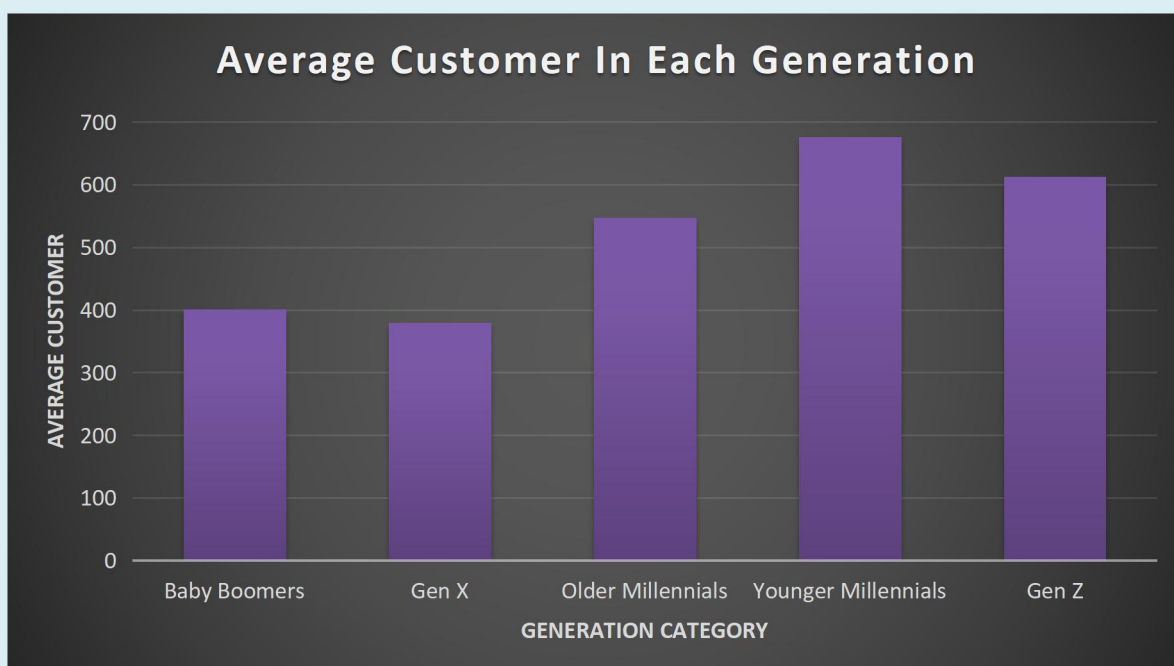
Analysis And Recommendation:

it's notable that sales outlets **3 and 8 have successfully met their sales goals**, whereas outlet 5 has fallen short.

And also It is evident from our analysis that the **beverage product group is the solo achiever** of its sales goals **across all sales outlets.** In contrast, other product categories have fallen short of their targets.

This points to the fact that the **achievements of outlets 3 and 8** can be primarily attributed to the **exceptional performance of the beverage group.** This success has managed to offset the shortcomings of the other product groups. While we have attained our monthly sales targets for outlets 3 and 5, we must not overlook the complete lack of success in the remaining product groups.

In the case of **outlet 5,** the beverage group has demonstrated commendable performance; however, its success in this area alone cannot compensate for the overall under performance of the other product groups. This discrepancy is the key reason behind our inability to attain the total sales goal. To enhance the sales performance of these other product groups.

To **improve the performance of the other product groups,** we need to identify areas of concern and weakness. It's essential to determine whether the **products lack sales potential,** or there are challenges in **effectively reaching customers,** or if **pricing in the market is comparatively high**, Or you are **target the wrong audience.**

To determine if our products have potential, we **tested a hypothesis** by investigating the performance of product groups. The results **revealed a positive outcome: a significant number of our products have a consistent customer base. This recurring customer trend indicates the potential of our products, as customers are returning to make repeat purchases.**

This difference indicates that a significant portion of our customers belong to the 'Baby Boomer' category due to its large market size. Despite the relatively short market size of the 'Younger Millennial's category, our products are notably popular among them.

we can observe that 'Younger Millennial's have the fewest number of customers, and this observation holds true. However, it overlooks the fact that **'Younger Millennial's have the narrowest range of years, only four years,** in comparison to the 'Baby Boomers' who have a range of fourteen years.

Moreover, we're also gaining popularity among the 'Gen Z' category. The decision to focus on either market size or future growth depends entirely on higher management.

My recommendation leans towards future growth, particularly targeting 'Gen Z.' Currently, they are young and have limited resources, but their market size and purchasing power are expected to grow in the future. This growth could substantially boost our sales.

It's important to foster brand loyalty through effective promotion among 'Gen Z.' Offering student-friendly deals and creating a mission to ensure every 'Gen Z' member tries our product can further enhance our success. It's worth considering that the 'Baby Boomer' market may shrink in the future due to factors like diabetes. Ultimately, the key lies in making strategic decisions that align with your company's goals.

- The "avg_repeat_customer" indicates the average number of repeat customers for each generation.
- Younger Millennial's (1990-1994) have the highest average repeat customer count (59.60), suggesting strong customer loyalty within this group.

Older Millennial's (1980-1989) and Gen Z (1995-2001) also exhibit high average repeat customer counts, indicating their engagement with brands and products.

This suggests that a significant portion of customers in each generation category is engaging with your products or services repeatedly.

Having an average count of repeat customers high side is a positive sign as it indicates customer loyalty and ongoing engagement. It show our product have a potential and number of customers are finding value in your offerings and are returning for additional transactions. This can be beneficial for business growth and revenue stability.

It's important to continue analyzing and understanding customer behavior to identify patterns, preferences, and opportunities for further engagement and growth.