

Store Business Performance Analysis



1. Project Background

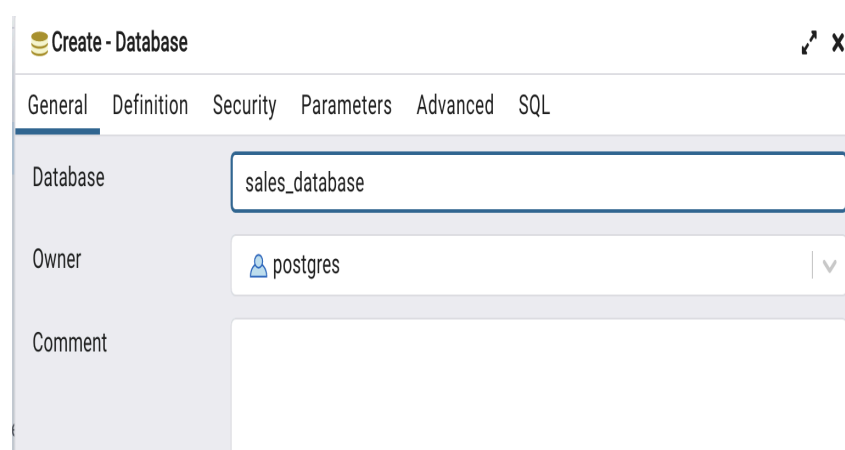
US-Superstore is a US-based superstore that has opened stores since 2015. The store sells 3 product categories including technology, office Supply and furniture.

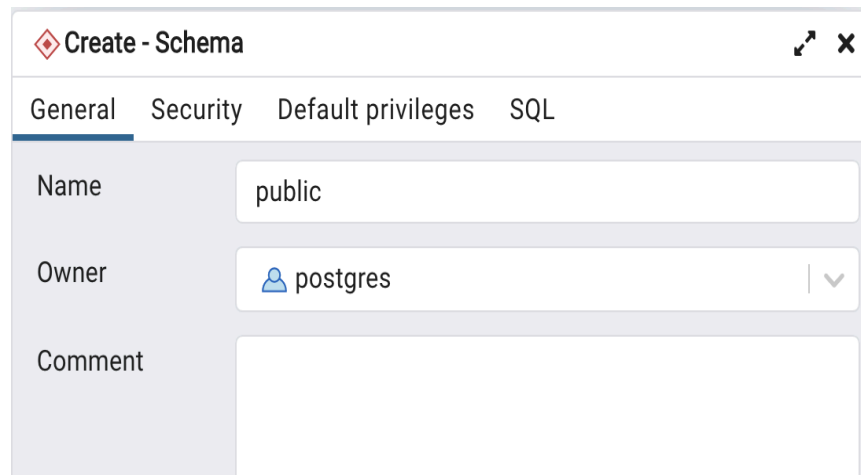
The owner has captured some very basic data from last 4 years of operation but have no idea how to use the data to run the business. So he needs your help to use their data store to quantify business performance and tell the story of how you have been able to generate that growth.

2. Data Preparation

2.1 Create database and schema

For this case study, I used PostgreSQL then I created the database and schema by using menu options after right clicking on database as shown in following image:



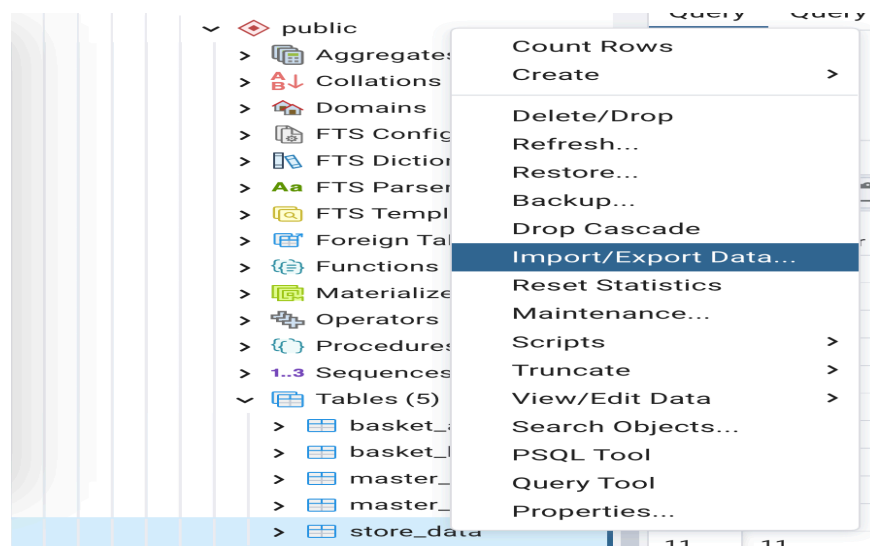


2.2 Tables

Once database and schema are created, the next step is to create tables. Here the dataset that I use is the dummy data that I got from the boot camp platform: MySkill.

Therefore we can take advantage of the data import feature which can be done by preparing data from Excel and saving it in the .csv extension then importing it into a table in the PostgreSQL database.

Next, we can login to the pgAdmin application and open the sales_database database that I created earlier. On the name of the table that has been created, right click and select Import/Export...



a. Store_data

Records consist of customer's orders with customer id, date when the order is ordered, product id of item ordered, order_id, date when the order is shipped, total amount sales and discount value. There are 9800 rows in this table.

	row_id	order_id	order_date	ship_date	ship_mode	customer_id	product_id	sales	discount
	character varying (10)	character varying (50)	timestamp without time zone	timestamp without time zone	character varying (10)	character varying (10)	character varying (50)	double precision	double precision
1	1	CA-2017-152156	2017-11-08 00:00:00	2017-11-11 00:00:00	Second Class	CG/12520	FUR-BO-10001798	3929400	0.02
2	2	CA-2017-152156	2017-11-08 00:00:00	2017-11-11 00:00:00	Second Class	CG/12520	FUR-CH-10000454	10979100	0.01
3	3	CA-2017-138688	2017-06-12 00:00:00	2017-06-16 00:00:00	Second Class	DV/13045	OFF-LA-10000240	219300	0.01
4	4	US-2016-108966	2016-10-11 00:00:00	2016-10-18 00:00:00	Standard Class	SO/20335	FUR-TA-10000577	14363662.5	0.02
5	5	US-2016-108966	2016-10-11 00:00:00	2016-10-18 00:00:00	Standard Class	SO/20335	OFF-ST-10000760	335520	0.03

b. Master_customer

Records consist of customer's orders with customer id, name of the customer, segmentation of customer, age and information region of customers. There are 793 rows in this table.

	customer_id	customer_name	segment	country	city	state	postal_code	region	age
	character varying (10)	character varying (50)	character varying (10)	character varying (10)	character varying (10)	character varying (10)	character varying (10)	character varying (10)	integer
1	CG/12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	42
2	DV/13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	47
3	SO/20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	19
4	BH/11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	39
5	AA/10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	31

c. Master_product

Records consist of product_id, name of the product, category and sub category each product. There are 1861 rows in this table.

	product_id	category	sub_category	product_name
	character varying (50)	character varying (100)	character varying (100)	text
1	FUR-BO-10001798	Furniture	Bookcases	Bush Somerse...
2	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fa...
3	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive ...
4	FUR-TA-10000577	Furniture	Tables	Bretford CR45...
5	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N ...

3. Project Goal

The store needs your help to gain insight into business performance

3.1 Objectives

- Store business performance

3.2 Problems Question

1. How many customers has store ever had?
2. What is the average monthly active users per year?
3. What is the number of new customers per year?
4. What is the number of customers who placed a repeat order (retained)?
5. What is the average order per year?
6. What has the trend been for sales versus revenue from 2015 to 2018?
7. Which state contributed the most to sales and revenue?
8. Which category contributed the most to sales and revenue for a selected state?
9. What is the largest total amount each customer spent at the store?
10. Which market/segment is the largest and smallest?
11. What was the first item purchased by all customer?
12. What is the most purchased item and how many times was it purchased by all customers?
13. Which sub category was the most popular for each customer?

4. Analysis and Visualization

Q1. How many customers has store ever had?

```
3 select count(distinct customer_id) as total_customer
4 FROM master_customer;
```

Result :

	total_customer bigint 
1	793

Q2. What is the average monthly active users per year?

```

9  with yearly_number AS(
10     select extract(year from order_date)as years,
11            extract(month from order_date)as months,
12            count(distinct customer_id)as number_customer
13     from store_data
14     group by 1,2
15     order by 1
16 )
17 SELECT years,
18        round(avg(number_customer),2)as avg_active_cust
19 from yearly_number
20 group by 1;

```

Result :

	years numeric	avg_active_cust numeric
1	2015	75.08
2	2016	79.08
3	2017	99.08
4	2018	125.33

Q3. What is the number of new customers per year?

```

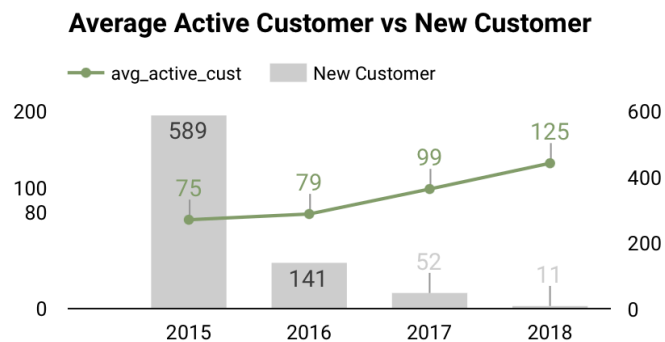
23 select extract(year from first_date)as years,
24        count(distinct customer_id)as num_customer
25 from (
26     select customer_id,
27            min(order_date)as first_date
28     from store_data
29     group by 1
30     order by 2
31 )as table_annual
32 group by 1
33 order by 1;

```

Result :

	years numeric	num_customer bigint
1	2015	589
2	2016	141
3	2017	52
4	2018	11

Insight :



As we can see there was a significant decrease in the number of new customers up to 60% in 2016 and this trend continued to decline until 2018.

Meanwhile, the average trend of active customers showed otherwise. The average number of active customers increased gradually every year, reaching 125 customers in 2018.

Q4. What is the number of customers who placed a repeat order (retained)?

```

36 with repeat_order AS(
37     select extract(year from order_date) as years,
38           t1.customer_id,
39           count(distinct t1.customer_id) as number_customer
40 from store_data t1
41 left join master_customer t2 ON t1.customer_id = t2.customer_id
42 group by 1,2
43 having count(order_id) > 1
44 )
45 SELECT years,
46        count(number_customer) as total_customer_repeat
47 from repeat_order
48 group by 1;

```

Result :

	years numeric 🔒	total_customer_repeat bigint 🔒
1	2015	426
2	2016	437
3	2017	493
4	2018	593

Q5. What is the average order per year?

```

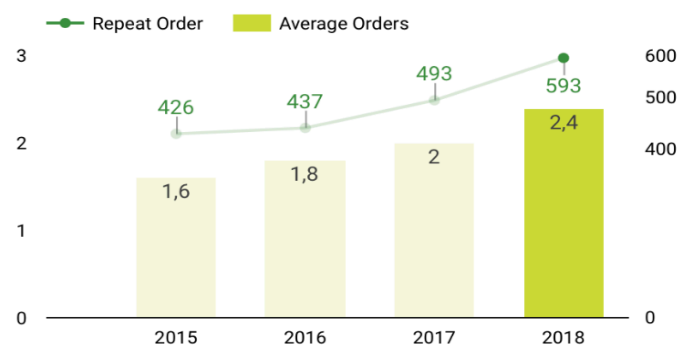
51 with orders as(
52     select extract(year from order_date) as years,
53            customer_id,
54            count(distinct order_id) as number_order
55     from store_data
56     group by 1,2
57     order by 1
58 )
59 select years,
60        round(avg(number_order),2) as avg_order
61     from orders
62     group by 1;

```

Result :

	years numeric 🔒	avg_order numeric 🔒
1	2015	1.61
2	2016	1.8
3	2017	2.04
4	2018	2.41

Insight :

Average Orders vs Repeat Orders

There was an increasing trend from 2015 to 2018 for the number of customers who made purchases more than once.

Likewise, the average number of orders placed by customers is also slowly increasing every year. although it doesn't change much each year, the average customer placed an order more than twice in 2018. This shows the probability that most of the customers were repeat customers

Q6. What has the trend been for sales versus revenue from 2015 to 2018?

```

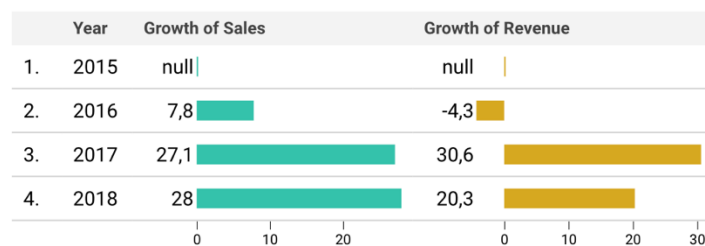
68 WITH first_number AS (
69   SELECT extract(year from order_date) as year,
70          COUNT(DISTINCT order_id) as total_order,
71          SUM(sales) as revenue
72   FROM store_data
73   WHERE sales IS NOT NULL
74   GROUP BY 1
75   ORDER BY 1
76 ),
77 -- add column LAG and LEAD
78 before_after AS(
79   SELECT year,
80          total_order,
81          LAG(total_order) OVER (order by year) as total_order_before,
82          LEAD(total_order) OVER (ORDER BY year) as total_order_after,
83          revenue,
84          LAG(revenue) OVER (order by year) as revenue_before,
85          LEAD(revenue) OVER (ORDER BY year) as revenue_after
86   FROM first_number
87   ORDER BY 1
88 )
89   SELECT year, total_order, revenue,
90          SUM(total_order - total_order_before)*100/SUM(total_order_before) as growth_order,
91          SUM(revenue - revenue_before)*100/SUM(revenue_before) as growth_revenue
92   FROM before_after
93   GROUP BY 1,2,3
94   ORDER BY 1;

```

Result :

	year numeric	total_order bigint	revenue double precision	growth_order numeric	growth_revenue double precision
1	2015	944	7197675061.5	[null]	[null]
2	2016	1018	6891453411	7.838983050847458	-4.254452276374133
3	2017	1294	9002798430	27.111984282907663	30.63715145530714
4	2018	1656	10830622458	27.975270479134466	20.30284296835023

Insight :

Sales and Revenue Growth from 2015 to 2018

There was a significant decrease in revenue causing outlets to experience a negative trend of -4.3% in 2016.

Meanwhile sales shows an increasing trend every year with a peak trend in 2017 of 27.1%.

Q7. Which state contributed the most to sales and revenue?

```

97  SELECT state,
98         COUNT(DISTINCT order_id)as total_order,
99         row_number() over(order by COUNT(DISTINCT order_id) desc)as rank_sales,
100        SUM(sales)as revenue,
101        row_number() over(order by SUM(sales) desc)as rank_revenue
102  FROM store_data t1
103 left join master_customer t2 ON t1.customer_id = t2.customer_id
104 WHERE sales IS NOT NULL
105 GROUP BY 1
106 ORDER BY 2 DESC, 4 DESC;

```

Result :

	state character varying (50)	total_order bigint	rank_sales bigint	revenue double precision	rank_revenue bigint
1	California	978	1	6545658700.5	1
2	New York	541	2	4159658812.5	2
3	Texas	457	3	2875922773.5	3
4	Pennsylvania	306	4	2128042605	4
5	Illinois	297	5	1684604460	7

The best city with the highest number of consumers placing orders with 978 transactions or sales and spending a lot of money of 6.5 billion was dominated by California.

It shows the locations of clients who make the most purchases overall allowing the sales team to drive sales and marketing efforts in this area to retain existing customers and attract new customers.

Q8. Which category contributed the most to sales and revenue?

```

109  SELECT category,
110         COUNT(DISTINCT order_id)as total_order,
111         row_number() over(order by COUNT(DISTINCT order_id) desc)as rank_sales,
112        SUM(sales)as revenue,
113        row_number() over(order by SUM(sales) desc)as rank_revenue
114  FROM store_data t1
115 left join master_product t2 ON t1.product_id = t2.product_id
116 WHERE sales IS NOT NULL
117 GROUP BY 1
118 ORDER BY 2 DESC, 4 DESC;

```

Result :

	category character varying (100)	total_order bigint	rank_sales bigint	revenue double precision	rank_revenue bigint
1	Office Supplies	3665	1	10580847480	3
2	Furniture	1727	2	10929878635.5	2
3	Technology	1518	3	12411823245	1

Although included in the most orders, Office Supplies was not the best product category of total sales. Technology generated the most sales and this can be attributed to the unit price of the product.

Q9. What is the largest total amount each customer spent at the store?

```

121 SELECT t1.customer_id,
122         SUM(sales)as revenue,
123         row_number() over(order by SUM(sales) desc)as rank_revenue
124 FROM store_data t1
125 left join master_customer t2 ON t1.customer_id = t2.customer_id
126 WHERE sales IS NOT NULL
127 GROUP BY 1
128 ORDER BY 2 DESC
129 limit 1;

```

Result :

	customer_id character varying (50)	revenue double precision	rank_revenue bigint
1	SM/20320	375645750	1

SM/20320 was the best customer by contributing to total revenue of 375,6 millions.

Q10. Which market/segment is the largest and smallest?

```

132 SELECT segment,
133         count(distinct t1.customer_id)as number_customer,
134         count(distinct order_id)as number_orders
135 from store_data t1
136 join master_customer t2 on t1.customer_id = t2.customer_id
137 group by 1;

```

Result :

	segment character varying (50)	number_customer bigint	number_orders bigint
1	Consumer	409	2537
2	Corporate	236	1491
3	Home Office	148	894

The consumer segment dominates both the number of customers and the number of orders.

Q11. What was the first item purchased by all customer?

```

148 with index_rank as(
149     select order_date,
150            category,
151            product_name,
152            row_number() over(order by order_date) as rnk
153 from store_data t1
154     left join master_product t2 on t1.product_id = t2.product_id
155 )
156
157 select order_date,
158        category,
159        product_name
160 from index_rank
161 where rnk = 1
162 order by 1;
```

Result :

	order_date timestamp without time zone	category character varying (100)	product_name text
1	2015-01-03 00:00:00	Office Supplies	Message Boo...

The Message Book from Office Supplies Category is the first item purchased

Q12. What is the most purchased item and how many times was it purchased by all customers?

```

165 select t1.product_id,
166        product_name,
167        count(t1.product_id) as number_purchase,
168        row_number() over(partition by t1.product_id
169                           order by count(t1.product_id) desc) as rank_order
170 from store_data t1
171     left join master_product t2 on t1.product_id = t2.product_id
172 group by 1,2
173 order by 3 desc
174 limit 1;
```

Result :

	product_id character varying (50)	product_name text	number_purchase bigint	rank_order bigint
1	OFF-PA-10001970	Xerox 1881	19	1

The most popular item that was bought the most was the Xerox 1881

Q13. Which sub category was the most popular for all customer?

```

177 with index_rank as(
178     select sub_category,
179            count(sub_category)as number_product,
180            row_number() over(order by count(sub_category) desc)as rnk
181     from store_data t1
182     left join master_product t2 on t1.product_id = t2.product_id
183     group by 1
184     order by 2 desc
185 )
186
187 select sub_category,
188        number_product
189 from index_rank
190 where rnk = 1
191 order by 1;

```

Result :

	sub_category character varying (100)	number_product bigint
1	Binders	1492

The most popular sub-category that was bought the most was Binders

5. Insight and Recommendation

5.1 Insight :

- Over the past 4 years, the store has a total of 793 customers, with the average active customer increasing gradually each year, reaching 125 customers in 2018.
- Even though a significant decrease in the number of new customers, reaching 60% in 2016. It is likely that most of these customers are regular customers. This can be seen from the increasing trend from 2015 to 2018 in the number of customers who make purchases more than once and the average number of orders made by customers.

- In 2016 the store experienced a negative revenue growth of -4.3% even though sales growth showed an increasing trend with a peak trend in 2017 of 27.1%.
- California is the best city with the highest number of consumers who order and spend a lot of money. It shows where the clients are making the most purchases overall allowing the sales team to drive sales and marketing efforts in this area to retain existing customers and attract new customers.
- The consumer segment dominates both the number of customers and the number of orders.
- SM/20320 was the best customer by contributing to total revenue of 375,6 millions.
- Despite being a popular category, Office Supplies is not the best product category from total sales. Technology generates the most sales and this can be attributed to the unit price of the product.

5.2 Recommendation :

- Explore marketing strategies that will help target and bring in new clients.
- The store should utilize customer and product information for marketing strategies that will help in retaining top customers.
- Find out what makes California as customers favourite state and apply the same strategy in other states.
- The Office Supplies category seems to be the community's favourite activity in the company's sales area, which is indicated by orders and the technology category being the largest total sales.
- The store must optimize sales by providing more products in both technology and office supplies categories and adapt their marketing campaigns to this niche market.