AMAZON BRAZIL’S DATA ANALYSIS

To identify trends, customer behaviours, and preferences that could be leveraged in the Indian market.

SWATI SHARMA
Sharma.swati.515@gmail.com

Project Context:

Amazon, a global leader in e-commerce, has achieved significant success in almost all parts of world. To identify trends, customer behaviours, and preferences that could be leveraged in the Indian market, Amazon India is analysing customer and sales data from Amazon Brazil, since the two have similarities - such as large populations and diverse consumer bases. The main goal of this analysis is to identify customer behaviours, product preferences, and payment patterns to enhance customer experience and discovering new opportunities in Indian market.

Tables used in analysis: Customers, Orders, Order Items, Product, Sellers, and Payments. Through SQL queries, various business-critical questions and providing actionable insights are shared.

Question 1.1

Problem Statement:

To simplify its financial reports, Amazon India needs to standardise payment values. Round the average payment values to integer (no decimal) for each payment type and display the results sorted in ascending order.

Analysis approach:

1. Identifying relevant tables and columns:

- Table: amazon_brazil.payments
- Columns: payment_type, payment_value

2. Average payment value calculation:

- AVG () aggregate function is used on payment value to get average payment value.
- GROUP BY function is used to get the average payment value for each payment type.
- ROUND () function is used to round the average payment values to the nearest integer.
- WHERE function is used to give condition to eliminate payment type which has zero payment values.

3. Sorting the Results:

- Used ORDER BY function to order the final results in ascending order of rounded average payment value.

SQL query:

```
select payment_type,
round(avg(payment_value)) as rounded_avg_payment
from amazon_brazil.payments
where payment_type <> 'not_defined'
group by payment_type
order by rounded_avg_payment;
```

Output:

payment_type	rounded_avg_payment
Voucher	66
debit_card	143
Boleto	145
credit_card	163

Recommendations:

1. **Prioritize best performing payment methods:**
   - Credit cards have the highest average payment (163 BLR). To further capitalize on this, continue offering incentives and exclusive credit card offers.
   - For debit card and boleto offer incentives like cashback or loyalty points to encourage more frequent use.
2. **Improvising least performing payment method:**
   - Vouchers have the lowest average payment value. Focus on increasing voucher usage by offering promotions or gift cards.

Question 1.2

Problem statement:

To refine its payment strategy, Amazon India wants to know the distribution of orders by payment type. Calculate the percentage of total orders for each payment type, rounded to one decimal place, and display them in descending order.

Analysis approach:

1. Identifying relevant tables and columns:

- Table: amazon_brazil.payments
- Columns: payment_type, order_id

2. Calculating percentage of total orders:

- Used COUNT () function on order id to get the number of orders across each payment type.

- Used count (*) to get the total number of orders.
- Divided the count of orders per payment type by the total number of orders, and multiplied by 100.
- Rounded the percentage values to one decimal place using ROUND () function.

5. Sorting Results:

- Used ORDER BY () DESC function to order the percentage values in descending orders.

SQL query:

```
select payment_type,
round(count(order_id) * 100.0 / (select count (*) from amazon_brazil.payments),1)
as percentage_orders
from amazon_brazil.payments
group by payment_type
order by percentage_orders desc;
```

Output:

payment_type	percentage_orders
credit_card	73.9
Boleto	19
Voucher	5.6
debit_card	1.5

Recommendations:

1. **Optimize credit card payments (73.9%):**
   - ❖ Focus on ensuring a seamless experience for credit card users, who form the majority.
   - ❖ Consider incentives like cashback, rewards, or no-cost EMI to maintain and grow this base.
2. **Encourage boleto users (19%):**
   - ❖ Introduce promotions or discounts for boleto users to either retain them or shift them to more efficient methods like credit or debit cards.
3. **Promote and expand voucher (5.6%) and debit card (1.5%) usage:**
   - ❖ Vouchers and debit card have lower usage. Expand partnerships, offer gift vouchers, or run targeted campaigns to boost voucher transactions and make it a more appealing option.

Question 1.3

Problem statement:

Amazon India seeks to create targeted promotions for products within specific price ranges. Identify all products priced between 100 and 500 BRL that contain the word 'Smart' in their name. Display these products, sorted by price in descending order.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.order_items, amazon_brazil.product
- Columns: product_id, price

2. Joined tables:

- Used JOIN function to combine two tables order items and product ON product id.

3. Filtering desired values:

- Used WHERE function to find product id with price between 100 and 500 BLR and the product category name having word 'Smart'.

4. Ordered the result:

- Used ORDER BY price DESC to order the price in descending order.

SQL query:

```
select pr.product_id, o.price as price
from amazon_brazil.order_items o
join amazon_brazil.product pr
on o.product_id = pr.product_id
where o.price between 100 and 500
and product_category_name like lower('%Smart%')
order by price desc;
```

Output: Total number of rows in output are 33, here showing top 10 of that.

product_id	Price
1df1a2df8ad2b9d3aa49fd851e3145ad	439.99
7debe59b10825e89c1cbcc8b190c85e2	349.99
ca86b9fe16e12de698c955aedff0aea2	349
ca86b9fe16e12de698c955aedff0aea2	349
0e52955ca8143bd179b311cc454a6caa	335
7aeaa8f3e592e380c420e8910a717255	329.9
7aeaa8f3e592e380c420e8910a717255	329.9

7aeaa8f3e592e380c420e8910a717255	329.9
7aeaa8f3e592e380c420e8910a717255	329.9
…..	…..

Recommendations:

1. **Focus on high-value "Smart" products:**
   - Products with highest value (439.99 BRL) or in the price range between 350 – 500 BLR offer a higher revenue opportunity. Promote these products with attractive offers or deals to drive more sales.
2. **Promotion of mid-range and low range products:**
   - Products priced between 200–350 BRL are popular and could attract customers. Targeted discounts or limited-time offers can make these products more attracting.
   - Items priced between 100–200 BRL can be promoted to price-sensitive customers. Emphasize affordability without sacrificing quality to attract large number of customers.

Question 1.4

Problem statement:

To identify seasonal sales patterns, Amazon India needs to focus on the most successful months. Determine the top 3 months with the highest total sales value, rounded to the nearest integer.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders, amazon_brazil.order_items
- Columns: order_purchase_timestamp, price, order_id

2. Combined two tables:

- Used JOIN function to combine orders and order items table on order id.

3. Calculated total sales per month:

- Used EXTRACT (MONTH FROM) function to get the months of order.
- Used SUM () of price to get the total sale value for each month.
- ROUNDED the total sales to nearest integer for convenience.

4. Grouping and Ordering the output:

- Used GROUP BY month to get month wise total sales.
- Used ORDER BY DESC to get total sales in descending order.
- Used LIMIT BY 3 to get top 3 sale values.

SQL query:

```
select extract (month from o.order_purchase_timestamp) as month,
round(sum(oi.price)) as total_sales
from amazon_brazil.orders o
join amazon_brazil.order_items oi
on o.order_id = oi.order_id
group by month
order by total_sales desc
limit 3;
```

Output:

month	total_sales
5	1502589
8	1428658
7	1393539

Recommendations:

1. Focus on month of highest sales:

- May generates the highest total sales. Launching major seasonal offers, and marketing campaigns during this month can help to further boost revenue.
- July and August are also strong sales months. Implementing targeted sales strategies such as end-of-season sales, and new product launch can help to further strengthen the sales figures.

2. Inventory and logistics management:

- Ensure adequate inventory, efficient supply chain management, and smooth logistics operations to handle the increased demand and minimize stockouts or delivery delays.

Question 1.5

Problem statement:

Amazon India is interested in product categories with significant price variations. Find categories where the difference between the maximum and minimum product prices is greater than 500 BRL.

Analysis approach:

1. Identified relevant tables and columns:

- Tables: amazon_brazil.product, amazon_brazil.order_items

- Columns: product_category_name, price, product_id

2. Joined tables:

- Used JOIN function to combine product and order items table on product id.

3. Calculating price difference:

- Used MAX () function to find the maximum price of product across categories.
- Used MIN () function to find the minimum price of product across categories.
- Used MAX () – MIN () to get the price variations within categories.

4. Grouping data:

- GROUPED the data BY product category name to get category wise price variation.

5. **Filtering categories:**

- Used HAVING function to give condition for filtering categories having price variation more than 500 BLR.

SQL query:

```
select p.product_category_name, max(oi.price) - min(oi.price)
as price_difference
from amazon_brazil.product p
join amazon_brazil.order_items oi
on p.product_id = oi.product_id
group by p.product_category_name
having max(oi.price) - min(oi.price) > 500;
```

Output: Total number of rows in output are 57, here showing first 10 out of that.

product_category_name	price_difference
Climatizacao	1588.1
livros_importados	730.01
NULL	3977
ferramentas_jardim	3923.65
dvds_blu_ray	1411.1
cine_foto	867.19
beleza_saude	3122.8
livros_interesse_geral	893.9
tablets_impressao_imagem	875.09
Papelaria	1690.71
…..	…..

Recommendations:

1. Focus on high variation categories:

- Categories like utilidades domesticas, pcs, artes, eletroportateis, instrumentos musicais, consoles games etc shows high variation (above 4000 BLR) and offers opportunities to target a wide range of consumers, from budget-conscious to premium shoppers.

2. Explore Niche Opportunities:

- Categories like perfumaria, moveis quarto, bebidas, audio, artigos de festas present substantial price gaps.
- These categories may benefit from promotions or premium vs. budget positioning.

Question 1.6

Problem statement:

To enhance the customer experience, Amazon India wants to find which payment types have the most consistent transaction amounts. Identify the payment types with the least variance in transaction amounts, sorting by the smallest standard deviation first.

Analysis approach:

1. Identified relevant tables and columns:

- Tables: amazon_brazil.payments
- Columns: payment_type, payment_value

2. Calculating standard deviation of payment value:

- Used STDDEV () function to get the variance in transaction amount for each payment type.
- ROUNDED the standard deviation for better readability.

3. Grouping the output:

- GROUPED the data BY payment type to get values across each payment type.

SQL query:

```
select payment_type, round(stddev(payment_value)) as std_deviation
from amazon_brazil.payments
where payment_type <> 'not_defined'
group by payment_type
order by std_deviation;
```

Output:

payment_type	std_deviation
voucher	116
Boleto	214
credit_card	222
debit_card	246

Recommendations:

1. Promote voucher payments:

- Voucher payments have the least standard deviation (116). Encouraging customers to use vouchers can help streamline payment processes and predict transaction patterns more accurately.

2. Improving boleto and credit card support:

- Both boleto and credit card payments have moderate deviation (214 and 222, respectively). Ensure robust support for these payment methods, as they provide predictable transaction amounts.

3. Monitor debit card transactions:

- Debit card payments have the highest standard deviation (246), indicating greater variability in transaction amounts. Amazon could monitor these transactions more closely.

Question 1.7

Problem statement:

Amazon India wants to identify products that may have incomplete name in order to fix it from their end. Retrieve the list of products where the product category name is missing or contains only a single character.

Analysis approach:

1. Identified relevant tables and columns:

- Tables: amazon_brazil.product
- Columns: product_id, product_category_name

2. Identifying product category:

- Used WHERE function to get the product id whose product category name is null.
- Used LEN () function to get the product id whose product category name contains single character.

SQL query:

```
select product_id, product_category_name
from amazon_brazil.product
where product_category_name is null
or length(product_category_name) = 1;
```

Output: Total number of rows in output are 614, here showing first 10 of that.

product_id	**product_category_name**
a41e356c76fab66334f36de622ecbd3a	NULL
d8dee61c2034d6d075997acef1870e9b	NULL
56139431d72cd51f19eb9f7dae4d1617	NULL
46b48281eb6d663ced748f324108c733	NULL
5fb61f482620cb672f5e586bb132eae9	NULL
e10758160da97891c2fdcbc35f0f031d	NULL
39e3b9b12cd0bf8ee681bbc1c130feb5	NULL
794de06c32a626a5692ff50e4985d36f	NULL
7af3e2da474486a3519b0cba9dea8ad9	NULL
…..	…..

Recommendations:

1. Identifying and correcting the names of category:

- There are 614 product category names which are either null or having just one letter. Replace these names with the accurate names.

Question 2.1

Problem statement:

Amazon India wants to understand which payment types are most popular across different order value segments (e.g., low, medium, high). Segment order values into three ranges: orders less than 200 BRL, between 200 and 1000 BRL, and over 1000 BRL. Calculate the count of each payment type within these ranges and display the results in descending order of count.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.payments

- Columns: payment_type, payment_value

2. Order value segmentation:

- Used CASE WHEN function to categorise order value as 'low', 'high', and 'medium'.

3. Counting each payment type within these segmentations:

- Used COUNT () function to count the payment type for each segmentation.
- GROUPED the data BY order value segmentation and payment type to get count of each payment type across each order value segment.

4. Ordering the result:

- ORDERED the result in descending order of count.

SQL query:

```
select
case
when payment_value > 1000 then 'high'
when payment_value between 200 and 1000 then 'medium'
when payment_value < 200 then 'low'
else 'NA'
end as order_value_segment, payment_type,
count(payment_type)
from amazon_brazil.payments
group by order_value_segment, payment_type
order by count desc;
```

Output:

order_value_segment	payment_type	count
Low	credit_card	60548
Low	Boleto	16444
Medium	credit_card	15303
Low	Voucher	5476
Medium	Boleto	3162
Low	debit_card	1287
High	credit_card	944
Medium	Voucher	286
Medium	debit_card	227
High	Boleto	178
High	debit_card	15
High	Voucher	13
Low	not_defined	3

Recommendations:

1. Prioritize popular payment methods:

- ❖ Credit cards are the most popular payment method across all order value segments, particularly in both low and medium-value orders. Offering rewards or discounts, could drive more sales and enhance user experience for credit card transactions.
- ❖ Boleto is the second most popular payment method. Promoting boleto payment options for higher-order values may encourage broader adoption.

3. Promote unpopular payment methods:

- ❖ Voucher usage is more common in low order value segments. Amazon could run promotions offering vouchers to increase sales in the medium and high segments.
- ❖ Debit card payments are less common. Amazon might consider offering additional incentives, such as cashback or instant rewards, to promote debit card usage across all order value segments.
- ❖ Address the small number of "not_defined" payments in low-value segments by ensuring all payment methods are clearly categorized.

Question 2.2

Problem statement:

Amazon India wants to analyse the price range and average price for each product category. Calculate the minimum, maximum, and average price for each category, and list them in descending order by the average price.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.product, amazon_brazil.order_items
- Columns: product_category_name, price

2. Calculated maximum, minimum and average price:

- Used MAX () function to get the maximum price.
- Used MIN () function to get minimum price.
- Used AVG () function to get the average price and used ROUND () function to round the average.

3. Grouping and ordering the data:

- GROUP BY product category name to get price across each category.
- ORDER BY average price in descending order to get category with maximum price first.

SQL query:

```
select p.product_category_name,
min(o.price) as min_price,
max(o.price) as max_price,
round(avg(o.price)) as avg_price
from amazon_brazil.product p
join amazon_brazil.order_items o
on p.product_id = o.product_id
group by product_category_name
order by avg_price desc;
```

Output: Total number of rows in output are 79, here showing first 10 of that.

product_category_name	**min_price**	**max_price**	**avg_price**
Pcs	34.5	6729	1098
portateis_casa_forno_e_cafe	10.19	2899	624
eletrodomesticos_2	13.9	2350	476
agro_industria_e_comercio	12.99	2990	342
instrumentos_musicais	4.9	4399.87	282
Eletroportateis	6.5	4799	281
portateis_cozinha_e_preparadores_de_alimentos	17.42	1099	265
telefonia_fixa	6	1790	226
construcao_ferramentas_seguranca	8.9	3099.9	209
relogios_presentes	8.99	3999.9	201
.....	.....	.....	.....

Recommendations:

1. Focus on high and mid value categories:

- Categories such as pcs, portateis casa forno e cafe and eletrodomesticos 2 have the highest average prices. Amazon should focus on premium positioning andexclusive deals to attract high-spending customers.
- Categories like agro industria e comercio, instrumentos musicais and portateis cozinha e preparadores de alimentos show average prices in the mid-range. Amazon could explore bundling and cross-selling opportunities to increase sales.

2. Promoting low value categories:

- ❖ Categories such as artes, brinquedos, and papelaria have relatively lower average prices. Focusing on volume sales through discounts or bulk purchase offers would be effective for these segments.

Question 2.3

Problem statement:

Amazon India wants to identify the customers who have placed multiple orders over time. Find all customers with more than one order, and display their customer unique IDs along with the total number of orders they have placed.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.customers, amazon_brazil.orders
- Columns: customer_unique_id, order_id

2. Finding customers with multiple orders:

- Used COUNT () function to get number of orders.
- GROUPED BY customer unique id to get total orders across each customer.
- Used HAVING function to filter customer with more than one order.

SQL query:

```
select c.customer_unique_id,
count(o.order_id) as total_orders
from amazon_brazil.customers c
join amazon_brazil.orders o
on c.customer_id = o.customer_id
group by c.customer_unique_id
having count(o.order_id) > 1;
```

Output: Total number of rows in output are 3140, here showing first 10 of that.

customer_unique_id	total_orders
00172711b30d52eea8b313a7f2cced02	2
004288347e5e88a27ded2bb23747066c	2
004b45ec5c64187465168251cd1c9c2f	2
0058f300f57d7b93c477a131a59b36c3	2
00a39521eb40f7012db50455bf083460	2
00cc12a6d8b578b8ebd21ea4e2ae8b27	2
011575986092c30523ecb71ff10cb473	2

011b4adcd54683b480c4d841250a987f	2
012452d40dafae4df401bced74cdb490	2
012a218df8995d3ec3bb221828360c86	2
…..	…..

Recommendations:

1. Emphasis on customers with multiple orders:

- There is total 3140 customers who have placed multiple orders from 2 to 16 each. Offering cashbacks and free gifts can help to retain them and encourage them to shop more.

2. Encouraging customers with single order:

- Offering coupon discounts, festive sale and first purchase offer to encourage those customers who shop less or never shops.

Question 2.4

Problem statement:

Amazon India wants to categorize customers into different types ('New' – order qty. = 1; 'Returning' – order qty. 2 to 4; 'Loyal' – order qty. >4) based on their purchase history. Use a temporary table to define these categories and join it with the customers table to update and display the customer types.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.customers, amazon_brazil.orders
- Columns: customer_id, order_id

2. Creating temporary table:

- Created temp table as customer categories to categorise customers.
- Used CASE WHEN function to identify 'New', 'Returning' and 'Loyal'.
- GROUP BY customer id to get category across each customer

3. Joined tables:

- Used JOIN function to join customer_categories table with customers table to get category across each customer.

SQL query:

```
create temporary table customer_categories as
```

```
select customer_id,
case
when count(order_id) = 1 then 'New'
when count(order_id) between 2 and 4 then 'Returning'
when count(order_id) > 4 then 'Loyal'
else 'NA'
end as customer_type
from amazon_brazil.orders
group by customer_id;
select c.customer_id, cc.customer_type
from amazon_brazil.customers c
join customer_categories cc
on c.customer_id = cc.customer_id;
```

Output: Total number of rows in output are 98348, here showing first 10 of that.

customer_id	customer_type
00012a2ce6f8dcda20d059ce98491703	New
000161a058600d5901f007fab4c27140	New
0001fd6190edaaf884bcaf3d49edf079	New
0002414f95344307404f0ace7a26f1d5	New
000379cdec625522490c315e70c7a9fb	New
0004164d20a9e969af783496f3408652	New
000419c5494106c306a97b5635748086	New
00046a560d407e99b969756e0b10f282	New
00050bf6e01e69d5c0fd612f1bcfb69c	New
000598caf2ef4117407665ac33275130	New
…..	…..

Recommendations:

1. Retaining Loyal and Returning customers:

❖ There are 98 Loyal (more than 4 orders) and 117 Returning (2 - 4 orders). To retain them Amazon could offer cashback point or free delivery on every purchase.

2. Focus on New customers:

❖ Maximum number (98133) of customers are 'New' (1 order). To encourage them to shop more amazon could provide best deals and offers.

Question 2.5

Problem statement:

Amazon India wants to know which product categories generate the most revenue. Use joins between the tables to calculate the total revenue for each product category. Display the top 5 categories.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.product, amazon_brazil.order_items
- Columns: product_category_name, price

2. Calculating total revenue:

- Used SUM () function on price to get the total revenue.
- ROUND () the total revenue for convenience.

3. Grouping and ordering data:

- Used GROUP BY product category name to get total revenue across each product category.
- Used ORDER BY total revenue in descending order and LIMIT BY 5 to get top 5 categories with maximum revenue.

SQL query:

```
select p.product_category_name, round(sum(o.price)) as total_revenue
from amazon_brazil.product p
join amazon_brazil.order_items o
on p.product_id = o.product_id
group by product_category_name
order by total_revenue desc
limit 5;
```

Output:

product_category_name	total_revenue
beleza_saude	1257865
relogios_presentes	1203060
cama_mesa_banho	1032269
esporte_lazer	985881
informatica_acessorios	910605

Recommendations:

1. Prioritize high-revenue categories:

- ❖ Categories like Beleza saude, relogios presentes, and cama mesa banho etc should be focused, as they generate the highest revenue. Consider targeted campaigns and loyalty programs to retain and expand customer bases in these areas.
- ❖ Ensure a strong supply chain for these high-revenue categories. Maintaining adequate stock levels and reducing delivery times could help increase customer satisfaction and sales.

2. Cross-selling and upselling of products:

- ❖ Promote complementary products within these top categories to further boost revenue.
- ❖ Identify products that have high demand but lower visibility. Highlight these through personalized marketing, offering exclusive deals to maximize revenue.

Question 3.1

Problem statement:

The marketing team wants to compare the total sales between different seasons. Use a subquery to calculate total sales for each season (Spring, Summer, Autumn, Winter) based on order purchase dates, and display the results. Spring is in the months of March, April and May. Summer is from June to August and Autumn is between September and November and rest months are Winter.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders, amazon_brazil.order_items
- Columns: price, order_purchase_timestamp

2. Subquery to identify season:

- Used EXTRACT MONTH FROM function to extract months from timestamp.
- Used CASE WHEN to categorise season as 'Spring', 'Summer', 'Winter'.

3. Calculating total sales:

- Used SUM () function on price to get total sales and ROUNDED the SUM.
- Used JOIN function to combine orders table with order items on order id.

4. Calculating total sales for each season:

- Used GROUP BY season to get total sales across each season.

SQL query:

```
select season, round(sum(oi.price)) as total_sales
from
(
select o.order_id,
case
when extract (month from order_purchase_timestamp) in (3, 4, 5) then 'Spring'
when extract (month from order_purchase_timestamp) in (6, 7, 8) then 'Summer'
when extract (month from order_purchase_timestamp) in (9, 10, 11) then 'Autumn'
else 'Winter'
end as season
from amazon_brazil.orders o
)
as order_season
join amazon_brazil.order_items oi
on order_season.order_id = oi.order_id
group by season;
```

Output:

season	total_sales
Winter	2905750
Autumn	2348813
Spring	4216722
Summer	4120360

Recommendations:

1. Focus on high sale seasons:

- ❖ Spring and summer generate the highest sales. Prioritize major promotions, new product launches, and discounts during these seasons to increase consumer spending.

2. Boost low sale seasons:

- ❖ Autumn shows comparatively lower sales. Targeted promotions, seasonal deals and festival sales can help to boost demand in this season.
- ❖ Winter sales are relatively strong. Introducing end-of-year sales, holiday promotions, and winter-specific product lines can further drive revenue during this season.

Question 3.2

Problem statement:

The inventory team is interested in identifying products that have sales volumes above the overall average. Write a query that uses a subquery to filter products with a total quantity sold above the average quantity.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.order_items
- Columns: product_id, order_id

2. Calculating total quantity sold:

- In SUBQUERY used COUNT () of order id and GROUP BY product id to get total quantity sold of each product.

3. Filtering products:

- Used WHERE function to get products with a total quantity sold above the average quantity.

SQL query:

```
select product_id, total_quantity_sold
from
(
select product_id, count (distinct order_id) as total_quantity_sold
from amazon_brazil.order_items
group by product_id
) as product_totals
where total_quantity_sold >
(select avg(total_quantity_sold)
from
(select product_id, count (distinct order_id) as total_quantity_sold
from amazon_brazil.order_items
group by product_id
) as avg_totals
)
order by total_quantity_sold desc;
```

Output: Total number of rows in output are 5538, here showing first 10 of that.

product_id	total_quantity_sold
99a4788cb24856965c36a24e339b6058	467

aca2eb7d00ea1a7b8ebd4e68314663af	431
422879e10f46682990de24d770e7f83d	352
d1c427060a0f73f6b889a5c7c61f2ac4	323
389d119b48cf3043d311335e499d9c6b	311
53b36df67ebb7c41585e8d54d6772e08	306
368c6c730842d78016ad823897a372db	291
53759a2ecddad2bb87a079a1f1519f73	287
154e7e31ebfa092203795c972e5804a6	269
2b4609f8948be18874494203496bc318	259
…..	…..

Recommendations:

1. Prioritize high - demand products:

- Products with higher sales volumes should be prioritized for inventory restocking. This will avoid stockouts and meet ongoing customer demand.
- Analyse historical trends for high-demand products to forecast future demand. This can help optimize procurement and logistics.
- Highlight best-selling products in marketing campaigns to leverage their popularity. Also bundling these products with related items can further increase sales.

2. Monitor low - demand products:

- Reducing stock or offer discounts to clear slower -moving inventory.

Question 3.3

Problem statement:

To understand seasonal sales patterns, the finance team is analysing the monthly revenue trends over the past year (year 2018). Run a query to calculate total revenue generated each month and identify periods of peak and low sales. Export the data to Excel and create a graph to visually represent revenue changes across the months.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.order_items, amazon_brazil.orders
- Columns: order_purchase_timestamp, price

2. Calculating total revenue for each month in 2018:

- Used EXTRACT MONTH FROM function to get months from timestamp.

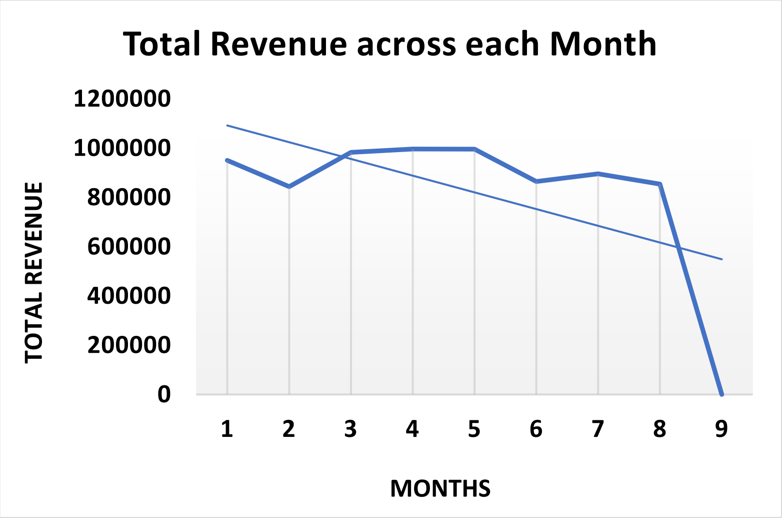
- Used SUM () function on price to get total revenue and ROUNDED the SUM.
- Used WHERE function to EXTRACT YEAR = 2018.
- Used GROUP BY month to get revenue across each month of 2018.

SQL query:

```
select extract (month from order_purchase_timestamp) as month,
round(sum(oi.price)) as total_revenue
from amazon_brazil.orders o
join amazon_brazil.order_items oi
on o.order_id = oi.order_id
where extract (year from (order_purchase_timestamp)) = 2018
group by month;
```

Output:

Month	total_revenue
1	950030
2	844179
3	983213
4	996648
5	996518
6	865124
7	895507
8	854686
9	145



Recommendations:

1. Target peak sales months:

- ❖ March, April, and May show the highest revenue, each exceeding 950,000 BRL. These months should be targeted for major promotional campaigns to capitalize on high demand.

2. Improvise low sales months:

- ❖ February and August exhibit comparatively low sales. Introduce strategic promotions or limited-time offers during these periods to stimulate demand.

3. Catering anomalies in data:

- September shows unusually low revenue (145 BRL), which may indicate a data entry error or an outlier. Investigate the cause of this drop and ensure accurate reporting for future analysis.

Question 3.4

Problem statement:

A loyalty program is being designed for Amazon India. Create a segmentation based on purchase frequency: ‘Occasional’ for customers with 1-2 orders, ‘Regular’ for 3-5 orders, and ‘Loyal’ for more than 5 orders. Use a CTE to classify customers and their count and generate a chart in Excel to show the proportion of each segment.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders
- Columns: customer_id, order_id

2. Classifying customer segments:

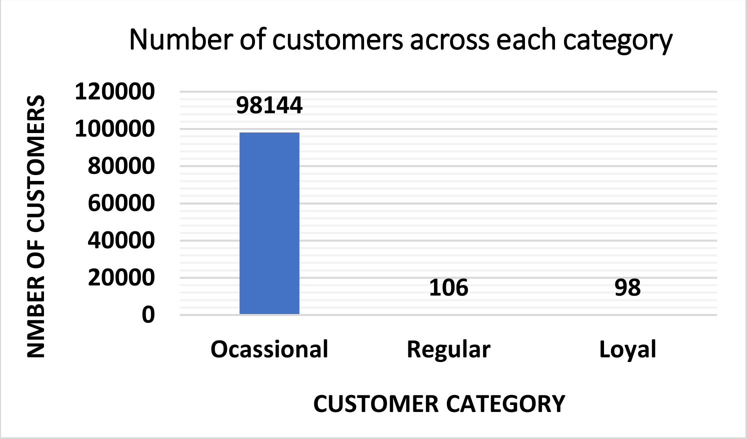
- Created CTE customer_segmentation to identify customer type.
- Used CASE WHEN function to get customer segmentation.
- Used COUNT () function to get the number of customers across each segment.
- GROUPED BY customer type.

SQL query:

```
with customer_segmentation as
(
select customer_id, count(order_id) as count,
case
when count(order_id) > 5 then 'Loyal'
when count(order_id) between 3 and 5 then 'Regular'
when count(order_id) between 1 and 2 then 'Ocassional'
else 'NA'
end as customer_type
from amazon_brazil.orders
group by customer_id
)
select customer_type, count(customer_type) as count from customer_segmentation
group by customer_type;
```

Output:

customer_type	Count
Ocassional	98144
Regular	106
Loyal	98



Recommendations:

1. Focus on Occasional Customers:

- ❖ Majority of customers (98,144) fall into the Occasional segment. Target this group with personalized offers, discounts, and incentives to encourage repeat purchases and converting them into Regular segment.
- ❖ Re-engage occasional customers through abandoned cart emails

2. Grow the Regular Segment:

- ❖ Regular segment (106 customers) can be further nurtured into becoming Loyal. Rewards can encourage them to increase their purchase frequency, such as discounts after a certain number of orders or additional benefits for higher spending.

3. Reward Loyal Customers:

- ❖ Loyal customer segment is relatively small (98 customers), they represent the most engaged users. Design exclusive loyalty rewards, early access to sales, and premium services to retain their loyalty and drive long-term value.

Question 3.5

Problem statement:

Amazon wants to identify high-value customers to target for an exclusive rewards program. You are required to rank customers based on their average order value (avg_order_value) to find the top 20 customers.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders, amazon_brazil.order_items
- Columns: customer_id, price

2. Calculating average order value across each customer.

- Used AVG () function on price to calculate average of order value.
- Used RANK function to arrange customers in rank based on their order value in descending order.
- Used LIMIT BY 20 to get top 20 high value customers.

SQL query:

```
select o.customer_id, avg(oi.price) as avg_order_value,
rank () over (order by avg(oi.price) desc) as customer_rank
from amazon_brazil.orders o
join amazon_brazil.order_items oi
on o.order_id = oi.order_id
group by o.customer_id
order by avg_order_value desc
limit 20;
```

Output: Total number of rows in output are 20, here showing first 10 of that.

customer_id	avg_order_value	customer_rank
c6e2731c5b391845f6800c97401a43a9	6735	1
f48d464a0baaea338cb25f816991ab1f	6729	2
3fd6777bbce08a352fddd04e4a7cc8f6	6499	3
df55c14d1476a9a3467f131269c2477f	4799	4
24bbf5fd2f2e1b359ee7de94defc4a15	4690	5
3d979689f636322c62418b6346b1c6d2	4590	6
1afc82cd60e303ef09b4ef9837c9505c	4399.87	7
35a413c7ca3c69756cb75867d6311c0d	4099.99	8
e9b0d0eb3015ef1c9ce6cf5b9dcbee9f	4059	9
c6695e3b1e48680db36b487419fb0398	3999.9	10
…..	…..	…..

Recommendations:

1. Prioritise high order value customers:

- ❖ Exclusive reward program should be there for the identified top 20 customers. Offer tailored benefits such as premium customer service, early access to sales, or personalized product recommendations.

- Track the purchasing behaviour of these customers to identify changes in buying patterns. If their spending drops, re-engage them with targeted campaigns or customer service follow-ups to prevent churn.
- Loyalty reinforcement can be done by providing personalized offers, loyalty points or limited-edition product launches to keep them connected to the brand.
- Increase retention through incentives like discounts or cashback on future purchases.

Question 3.6

Problem statement:

Amazon wants to analyse sales growth trends for its key products over their lifecycle. Calculate monthly cumulative sales for each product from the date of its first sale. Use a recursive CTE to compute the cumulative sales (total_sales) for each product month by month.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders, amazon_brazil.order_items
- Columns: product_id, order_purchase_timestamp, price

2. Computing cumulative sales for each product using CTE:

- Used EXTRACT MONTH FROM to get sale month.
- Used SUM () function on price to get monthly sale of each product.
- Used JOIN function to join orders table with customers.
- Used PARTION BY product id and ORDER BY sale month to get total sale for each product month by month.

SQL query:

```
with monthly_sales as
(
select oi.product_id,
extract (month from o.order_purchase_timestamp) as sale_month,
sum(oi.price) as monthly_sale
from amazon_brazil.orders o
join amazon_brazil.order_items oi
on o.order_id = oi.order_id
group by product_id, sale_month
)
select product_id, sale_month,
sum(monthly_sale) over (partition by product_id order by sale_month) as total_sales
from monthly_sales;
```

Output: Total number of rows in output are 60796, here showing first 10 of that.

product_id	sale_month	total_sales
00066f42aeeb9f3007548bb9d3f33c38	5	101.65
00088930e925c41fd95ebfe695fd2655	12	129.9
0009406fd7479715e4bef61dd91f2462	12	229
000b8f95fcb9e0096488278317764d19	8	117.8
000d9be29b5207b54e86aa1b1ac54872	4	199
0011c512eb256aa0dbbb544d8dffcf6e	12	52
00126f27c813603687e6ce486d909d01	9	498
001795ec6f1b187d37335e1c4704762e	10	38.9
001795ec6f1b187d37335e1c4704762e	11	116.7
001795ec6f1b187d37335e1c4704762e	12	350.1
…..	…..	……

Recommendations:

1. Focus on product with high sales:

- Through promotions and discounts retain the high sales of these products.
- List these products on top in search to enhance chances of confirm orders.
- Offer large discounts on bulk purchase to increase sale volume and revenue.

2. Identify and target product with low sales:

- Offer free shipping, clearance sale discounts, clubbing with high sale products etc can help enhance the sales of these products.

Question 3.7

Problem statement:

To understand how different payment methods affect monthly sales growth, Amazon wants to compute the total sales for each payment method and calculate the month-over-month growth rate for the past year (year 2018). Write query to first calculate total monthly sales for each payment method, then compute the percentage change from the previous month.

Analysis approach:

1. Identifying relevant tables and columns:

- Tables: amazon_brazil.orders, amazon_brazil.order_items, amazon_brazil.payments
- Columns: payment_type, order_purchase_timestamp, price

2. Calculating total sale for each payment method:

- Used EXTRACT MONTH FROM function to get months.
- Used SUM () function on price to get total sale for each payment method.
- Used JOIN function to combine table orders, order_items and payments.
- Used WHERE condition to specify year as 2018.

3. Computing percentage change from previous month.

- Used LAG () function to get the percentage change month over month.

SQL query:

```
with total_sale as
(
select p.payment_type,
extract (month from o.order_purchase_timestamp) as sale_month,
round(sum(oi.price)) as monthly_total
from amazon_brazil.orders o
join amazon_brazil.order_items oi
on o.order_id = oi.order_id
join amazon_brazil.payments p
on o.order_id = p.order_id
where extract (year from o.order_purchase_timestamp) = 2018
group by p.payment_type, sale_month
)
select payment_type, sale_month, monthly_total,
case
when lag(monthly_total) over () = 0 then null
else
round ((monthly_total - lag(monthly_total) over())/lag(monthly_total) over() * 100)
end as monthly_change
from total_sale;
```

Output: Total number of rows in output are 33, here showing first 10 of that.

payment_type	sale_month	monthly_total	monthly_change
boleto	1	170651	NULL
boleto	2	153166	-10

boleto	3	157807	3
boleto	4	162941	3
boleto	5	166572	2
boleto	6	126380	-24
boleto	7	162938	29
boleto	8	118214	-27
credit_card	1	760253	543
credit_card	2	680199	-11
…..	…..	…..	…..

Recommendations:

1. Focus on credit card sales:

- ❖ Credit cards consistently generate the highest monthly sales, with some months showing growth (e.g., a 20% increase in March). Amazon should prioritize credit card promotions by providing more incentives for customers using this method.

2. Boleto payment volatility:

- ❖ Boleto payment method sales show fluctuating growth rates, with significant drops in June (-24%) and August (-27%). Amazon may need to investigate reasons for this inconsistent purchasing behaviour and consider offering special incentives to stabilize sales growth for this payment method.

3. Growth opportunity for debit cards:

- ❖ Debit cards show impressive growth from May onward (256% growth in June and consistent increases afterward). Amazon should explore opportunities to promote debit card payments, such as by offering targeted discounts or cashback offers to increase adoption.

4. Reducing Voucher decline:

- ❖ Voucher-based payments show declining sales in multiple months, with a particularly steep drop in September (-100%). Amazon should evaluate whether the voucher program needs adjustments or better promotion to attract more users.

THANK YOU