# **Amazon Brazil Data Analysis Project**

## **Executive Summary**

**This project analyzes Amazon Brazil's e-commerce data to identify trends, customer behaviors, and preferences that could be leveraged in the Indian market. By examining various aspects like payment preferences, product pricing, customer segmentation, and seasonal patterns, this analysis aims to provide actionable insights for Amazon India's strategic decision-making.**

## **Project Background**

**Amazon, as a global e-commerce leader, has achieved remarkable success across markets including the U.S., Europe, and Asia. In Brazil, Amazon has established itself as a key player connecting small and medium businesses with millions of customers. Given the similarities between Brazil and India—both featuring large populations and diverse consumer bases—there's a significant opportunity to apply successful strategies from the Brazilian market to India.**

## **Objectives**

1. **Analyze customer demographics and behavior patterns**
2. **Identify regional trends and customer density**
3. **Track order lifecycles and understand product preferences**
4. **Evaluate seller performance and payment preferences**
5. **Develop actionable recommendations based on comprehensive data analysis**

## **Methodology**

**The analysis employs PostgreSQL queries of varying complexity, from basic aggregations to advanced concepts including:**

* **Joins across multiple tables**
* **Common Table Expressions (CTEs)**
* **Window functions**
* **Subqueries and recursive queries**
* **Statistical calculations and categorization**

## **Data Overview**

**The analysis utilizes seven interconnected tables:**

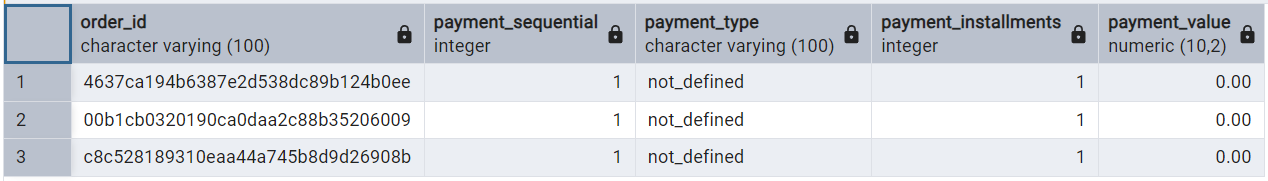
* **Customers: Customer information and demographics**
* **Orders: Order details including status and timestamps**
* **Order Items: Individual items with price and shipping information**
* **Product: Product specifications and categories**
* **Seller: Seller information and location**
* **Payments: Transaction details and payment methods**

**Through systematic analysis of these datasets, this project aims to uncover valuable insights that can inform Amazon India's market strategies and enhance customer experience.**

**DATA CLEANING**

SELECT \* FROM payments

WHERE payment\_type = 'not\_defined' ;

****

During the data cleaning process, a set of transactions was identified with the payment\_type labeled as 'not\_defined' and payment\_value = 0.00. These records are likely the result of placeholder values, data entry errors, or missing payment information.

**Key Observations**:

* All such records had a payment value of zero, which does not contribute to revenue.
* The payment\_type was undefined, offering no insight into customer payment preferences or behavior.
* These records accounted for a very small fraction of the dataset.

**Action Taken:**

DELETE FROM payments

WHERE payment\_type = 'not\_defined' AND payment\_value = 0.00;

**Impact:**

* Improved the accuracy of payment-type distribution analysis.
* Prevented misleading results when calculating average payment values or order proportions by payment method.
* Ensured that all records included in the analysis reflect valid transactions.

**Analysis - I**

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

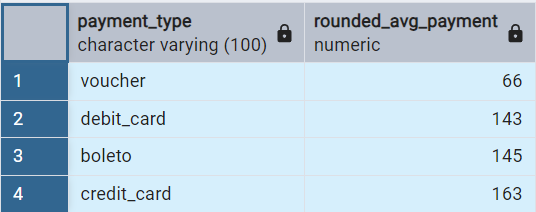
SELECT payment\_type,

ROUND(AVG(payment\_value)) AS rounded\_avg\_payment

FROM payments

GROUP BY payment\_type

ORDER BY rounded\_avg\_payment;



### **Insights:**

* **Voucher** has the lowest average payment (66), indicating it’s used for smaller transactions.
* **Debit card** (143) and **boleto** (145) are similar in value, suggesting they may be used for mid-range transactions.
* **Credit card** has the highest average payment (163), likely indicating it’s preferred for larger purchases.

### **Recommendations:**

* **Voucher**: Promote more vouchers for smaller purchases, offering incentives to increase usage.
* **Debit card & Boleto**: Consider offering targeted promotions or discounts to encourage customers to spend more with these methods.
* **Credit card**: Focus on loyalty programs for higher-value transactions to encourage repeat spending.
* **Payment Diversification**: Analyze usage patterns and introduce new options if certain payment types are underutilized.

2.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

SELECT payment\_type,

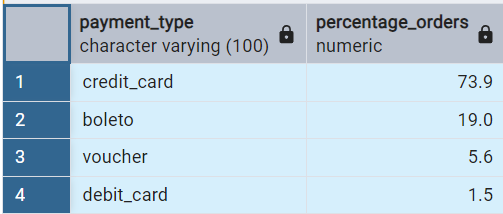
ROUND(COUNT(order\_id) \* 100.0 / total\_orders, 1) AS percentage\_orders

FROM payments,

(SELECT COUNT(order\_id) AS total\_orders FROM payments) AS t

GROUP BY payment\_type, t.total\_orders

ORDER BY percentage\_orders DESC;



### **Insights:**

* **Credit Card** dominates with 73.9% of total orders, suggesting it's the preferred payment method.
* **Boleto** follows with 19.0%, indicating it's also a significant choice.
* **Voucher** and **Debit Card** have smaller shares, 5.6% and 1.5% respectively, showing limited use compared to other methods.

### **Recommendations:**

* **Credit Card**: Amazon India could explore ways to incentivize and maintain high usage of credit card payments.
* **Boleto**: It could be beneficial to offer promotions or loyalty programs to increase boleto usage.
* **Voucher & Debit Card**: Focus on strategies to boost these payment methods, such as offering special deals or discounts for users preferring them.
* **Payment Method Optimization**: Reassess payment options to ensure that customers have diverse and convenient payment choices.

3.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.

Output: product\_id, price

SELECT DISTINCT p.product\_id,oi.price

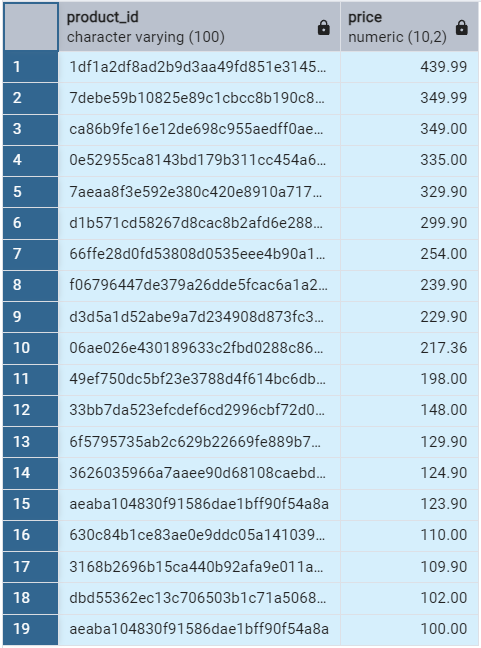
FROM product p

JOIN order\_items oi ON p.product\_id = oi.product\_id

WHERE oi.price BETWEEN 100 AND 500

AND p. product\_category\_name ILIKE '%smart%'

ORDER BY oi.price DESC;



### **Insights:**

1. **Product Demand**: The "Smart" category is popular within the **100 to 500 BRL range**, indicating strong interest in smart technology products at a mid-range price point.
2. **Price Distribution**: Higher-priced products (closer to 500 BRL) are prominent in the results, suggesting that customers are willing to spend more on smart devices within this range.
3. **Market Focus**: The data suggests that targeting the **"Smart"** category could be highly effective in promotions, as these products are consistently in the middle price range.

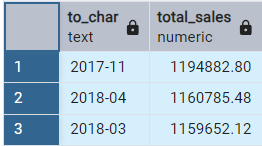
### **Recommendations:**

1. **Focus on Premium Products**: Highlight **premium smart devices** priced near 500 BRL for high-end targeted promotions.
2. **Segmented Promotions**: Run **promotions and discounts** for products priced around 300-500 BRL, as they appear to have the most demand.
3. **Targeted Campaigns**: Use the **smart technology trend** to create **seasonal or flash sale campaigns** to boost sales in this category.
4. **Explore Loyalty Programs**: For the lower-priced products, consider introducing **loyalty rewards or bundle offers** to increase repeat purchases.

4.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.

Output: month, total\_sales

SELECT TO\_CHAR(order\_purchase\_timestamp, 'YYYY-MM') AS order\_month,  
SUM(payment\_value) AS total\_sales  
FROM orders o  
JOIN payments p ON o.order\_id = p.order\_id  
GROUP BY TO\_CHAR(order\_purchase\_timestamp, 'YYYY-MM')  
ORDER BY total\_sales DESC  
LIMIT 3;



### **Insights:**

1. **Top Sales Months**: The highest sales occurred in **November 2017**, followed by **April 2018** and **March 2018**, indicating these months have been the most successful.
2. **Seasonal Peaks**: **November 2017** saw the **highest total sales** (1,194,882.80 BRL), likely benefiting from holiday shopping or promotions.
3. **Consistency Across Months**: **March** and **April 2018** also show significant sales, suggesting **early spring** could be another high-sales period.

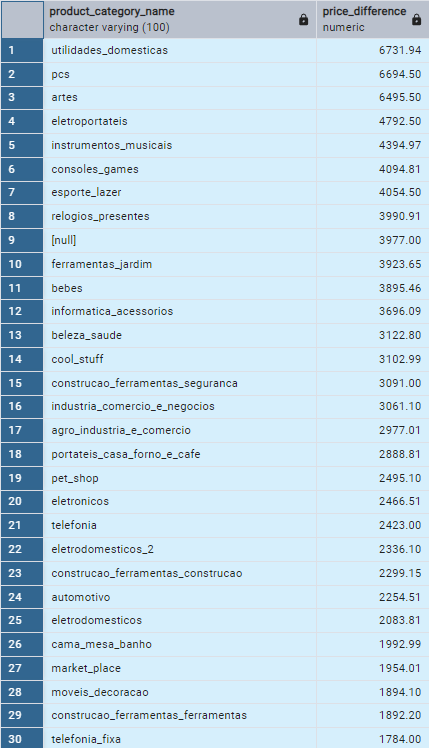
### **Recommendations:**

1. **Focus on Strong Months**: Run **special promotions** in **November, March, and April** to maximize sales during historically high periods.
2. **Holiday and Seasonal Campaigns**: Take advantage of **November sales** with targeted **holiday campaigns** or early-bird promotions.
3. **Understand Sales Drivers**: Investigate specific events in **March and April** to identify what drove sales, helping to replicate this success in future months.

5.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.

Output: product\_category\_name, price\_difference

SELECT product\_category\_name,  
MAX(price) - MIN(price) AS price\_difference  
FROM order\_items oi  
JOIN product p ON p.product\_id = oi.product\_id  
GROUP BY product\_category\_name  
HAVING MAX(price) - MIN(price) > 500;


### **Insights:**

1. **Extremely High Price Variation**: Categories like **"artes" (6495.50 BRL)**, **"pcs" (6694.50 BRL)**, and **"utilidades\_domesticas" (6731.94 BRL)** show substantial price differences. These likely include both basic and premium products, possibly with significant technological or brand-based differences.
2. **Wide Category Spread**: Over **50 product categories** show price variation >500 BRL, suggesting **broad product diversity** across the catalog.
3. **Opportunity to Refine Pricing Strategy**: Some consumer segments, like **"perfumaria" (684.91)** or **"smart" (1444.50)**, show moderate differences, indicating potential for **tiered branding or bundling strategies**.

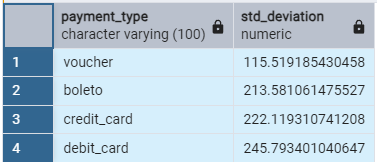
### **Recommendations:**

1. **Detailed Price Band Analysis**: Break down price ranges within top-variation categories to design **segment-specific offers** (e.g., basic, mid-tier, premium).
2. **Targeted Category Campaigns**: Use insights to launch **category-specific promotions**, highlighting **value products** and **premium alternatives** in the same space.
3. **UX Enhancements**: Provide filters and comparison tools in these categories to help users navigate wide price spreads and improve **decision-making**.

6.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.

Output: payment\_type, std\_deviation

SELECT payment\_type,   
 STDDEV(payment\_value) AS std\_deviation  
FROM payments  
GROUP BY payment\_type  
ORDER BY std\_deviation ASC;



**Insights:**

* Voucher payments show the lowest standard deviation (115.52), making them the most consistent in terms of transaction values.
* Boleto (213.58) and Credit Card (222.12) payments exhibit moderate consistency, with only slight variability in transaction amounts.
* Debit Card payments have the highest standard deviation (245.79), indicating greater inconsistency and wider variation in payment values compared to other methods.

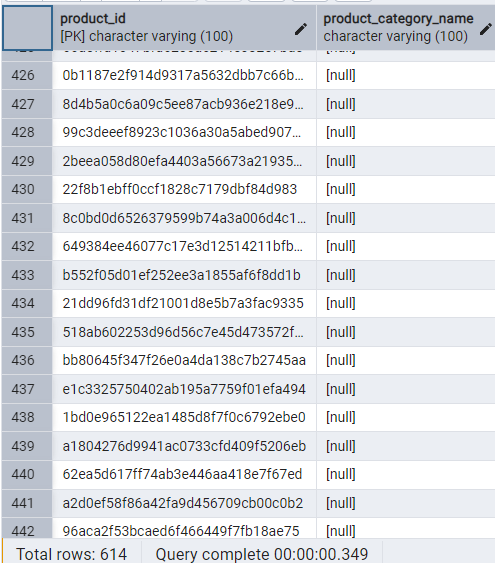
**Recommendations:**

* Promote voucher payments for categories or campaigns that require stable and predictable pricing.
* Monitor debit card usage to identify customer segments or purchase patterns causing the high variability.
* Highlight consistent payment methods (voucher, boleto) at checkout to improve user confidence and enhance the purchasing experience.

7.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.

Output: product\_id, product\_category\_name

SELECT product\_id, product\_category\_name  
FROM product  
WHERE product\_category\_name IS NULL   
OR CHAR\_LENGTH(product\_category\_name) = 1;



**Insights**:

* 614 products have missing or very short category names (length = 1), indicating potential data quality issues.

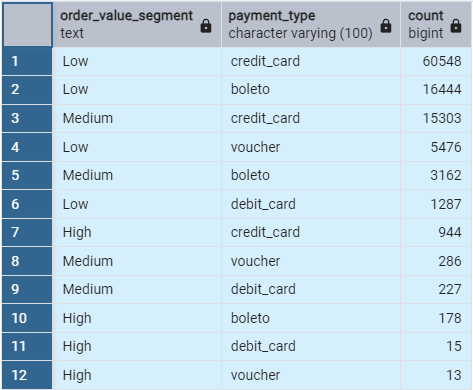
**Recommendations**:

1. Clean data by identifying and correcting these entries.
2. Label products with missing categories as "Needs Review."
3. Investigate the root cause and improve data entry validation.
4. Assess the business impact and prioritize high-impact fixes.

**Analysis - II**

1. 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

SELECT   
 CASE   
 WHEN payment\_value < 200 THEN 'Low'  
 WHEN payment\_value BETWEEN 200 AND 1000 THEN 'Medium'  
 WHEN payment\_value > 1000 THEN 'High'  
 END AS order\_value\_segment,  
 payment\_type,   
 COUNT(\*) AS count  
FROM payments  
GROUP BY payment\_type,order\_value\_segment  
ORDER BY count DESC;



**Insights**:

1. The **credit card** payment type is the most popular across all segments, with a significant dominance in the "Low" and "Medium" order value segments.
2. **Boleto** is the second most popular payment method, particularly in the "Low" and "Medium" segments.
3. **Voucher** and **debit card** are less popular across all segments, with very few transactions in the "High" order value segment.
4. As order values increase, **credit card** remains the most frequent payment method, while **boleto** and **voucher** see a decrease in usage.

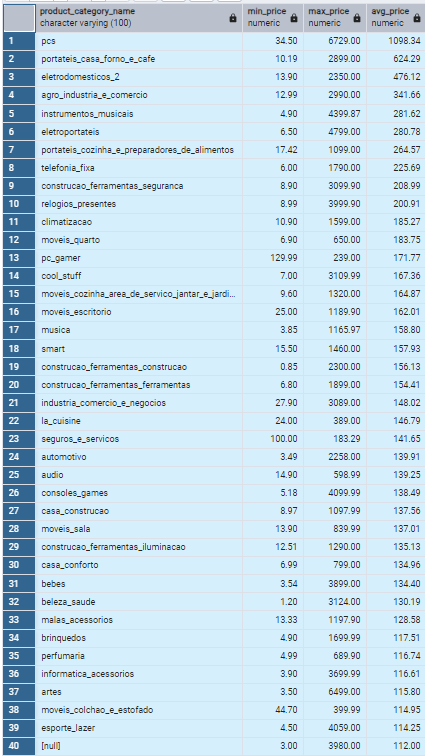
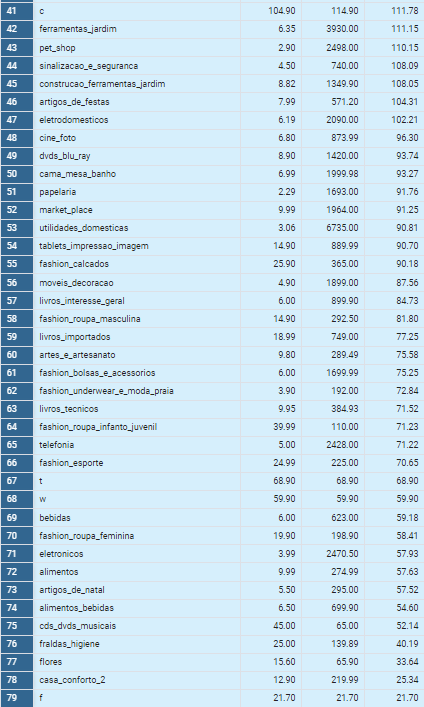
**Recommendations**:

1. Focus marketing and promotions around **credit card** and **boleto** payment options, particularly for low- to medium-value orders.
2. Consider offering incentives to encourage **voucher** and **debit card** usage, especially in higher-value orders.
3. Analyze the reasons behind the lower adoption of **voucher** and **debit card** to target potential barriers to usage in higher-value segments.

2.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.

Output: product\_category\_name, min\_price, max\_price, avg\_price

SELECT product\_category\_name,   
 MIN(price) AS min\_price,   
 MAX(price) AS max\_price,   
 ROUND(AVG(price), 2) AS avg\_price  
FROM product  
JOIN order\_items ON product.product\_id = order\_items.product\_id  
GROUP BY product\_category\_name  
ORDER BY avg\_price DESC;

**Insights**:

1. The **"pcs"** category has the highest average price (1098.34 BRL), with a wide price range between 34.50 BRL (minimum) and 6729.00 BRL (maximum).
2. **"portateis\_casa\_forno\_e\_cafe"** and **"eletrodomesticos\_2"** are other categories with relatively high average prices, at 624.29 BRL and 476.12 BRL respectively.
3. Categories like **"f"**, **"c"**, and **"t"** have the lowest average prices (21.70 BRL, 59.90 BRL, and 68.90 BRL), with narrow price ranges.
4. **"bebes"** and **"beleza\_saude"** also have low average prices, at 134.40 BRL and 130.19 BRL, respectively, but with broader price ranges compared to categories with very low average prices.

**Recommendations**:

1. Focus on premium products within categories like **"pcs"** and **"portateis\_casa\_forno\_e\_cafe"** to target high-value customers.
2. Explore potential marketing strategies for **"f"**, **"c"**, and **"t"** to increase customer engagement despite their low average prices.
3. Consider bundling products in categories with low average prices, like **"bebes"**, to increase perceived value.

3.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.

Output: customer\_unique\_id, total\_orders

Select \* from Orders;

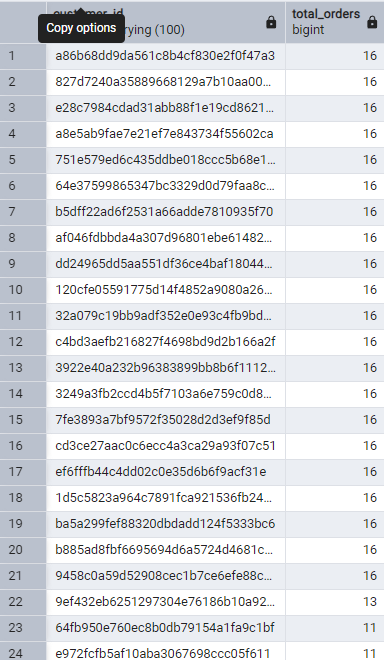
Select customer\_id, count(\*) as total\_orders

from orders

group by customer\_id

having count(\*) > 1

order by total\_orders desc;


### **Insights:**

1. **Frequent Orders**: A few customers placed 16 orders, showing high loyalty. Others have fewer orders, indicating potential for increased engagement.
2. **Retention Opportunities**: Customers with fewer orders (4-6) suggest room for improvement in retention.
3. **Top Customers**: Focus on those with the most orders for loyalty programs.

### **Recommendations:**

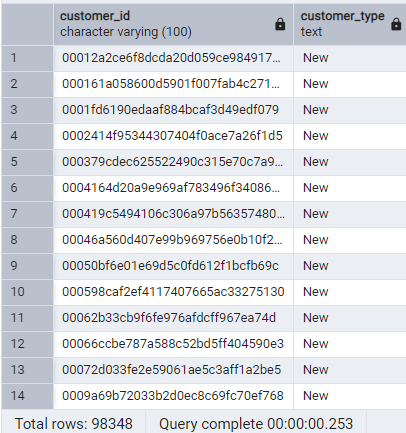
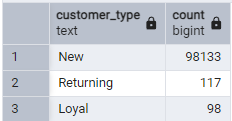
1. **Loyalty Programs**: Reward high-frequency customers with exclusive deals.
2. **Retention Campaigns**: Engage customers with 4-6 orders through personalized offers and reminders.
3. **Customer Segmentation**: Tailor marketing strategies for high, moderate, and low-frequency buyers.
4. **Targeted Ads**: Focus on products bought by frequent customers for more precise marketing.
5. **Referral Programs**: Encourage loyal customers to refer others for rewards.

4.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.

Output: customer\_id, customer\_type

CREATE TEMPORARY TABLE customer\_types AS  
SELECT o.customer\_id,  
CASE   
 WHEN COUNT(o.order\_id) = 1 THEN 'New'  
 WHEN COUNT(o.order\_id) BETWEEN 2 AND 4 THEN 'Returning'  
 WHEN COUNT(o.order\_id) > 4 THEN 'Loyal'  
 END AS customer\_type  
FROM orders o  
GROUP BY o.customer\_id;

SELECT customer\_id,customer\_type   
FROM customer\_types ;

  
  
SELECT CUSTOMER\_TYPE,COUNT(\*) AS COUNT  
FROM CUSTOMER\_TYPES  
GROUP BY CUSTOMER\_TYPE ;  


### **Insights:**

1. **Majority Are New**: Most customers (98,133) have made only one purchase, indicating a high drop-off after the first order.
2. **Loyal Customers Are Few**: Only 98 customers fall into the 'Loyal' category, showing a significant opportunity to boost retention.
3. **Moderate Engagement**: 117 customers are 'Returning' (2–4 orders), suggesting a small but growing group that may convert to loyal with the right push.

### **Recommendations:**

1. **Improve First Experience**: Enhance post-purchase engagement to convert ‘New’ customers to ‘Returning’.
2. **Re-engagement Campaigns**: Target ‘New’ customers with timely offers or reminders.
3. **Upsell to ‘Returning’**: Encourage this group with loyalty points or exclusive discounts to promote more frequent purchases.
4. **Reward Loyalty**: Strengthen relationships with ‘Loyal’ customers through premium rewards, early access, or recognition programs.

5.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.

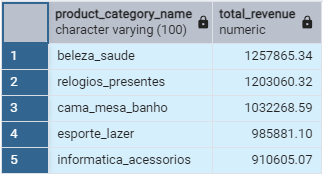
Output: product\_category\_name, total\_revenue

SELECT p.product\_category\_name,  
 ROUND(SUM(oi.price), 2) AS total\_revenue  
FROM order\_items oi  
JOIN product p ON oi.product\_id = p.product\_id  
GROUP BY p.product\_category\_name  
ORDER BY total\_revenue DESC  
LIMIT 5;  
  
**Insights:**

1. **Top Category – Beauty & Health**: 'beleza\_saude' leads in revenue, indicating high demand in personal care.
2. **Watches & Gifts Popular**: 'relogios\_presentes' ranks second, showing strong performance in lifestyle/gifting.
3. **Home Essentials Thrive**: 'cama\_mesa\_banho' suggests consistent consumer investment in home products.
4. **Active Lifestyle Products Sell Well**: 'esporte\_lazer' indicates interest in fitness and recreation.
5. **Tech Accessories Are Lucrative**: 'informatica\_acessorios' reflects growing tech usage and gadget reliance.

### **Recommendations:**

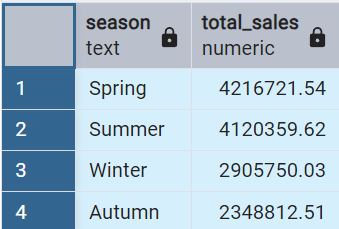
1. **Promote Top Sellers**: Increase visibility and ad spend on the top 3 categories.
2. **Bundle Products**: Create combo offers within high-performing categories to boost order value.
3. **Stock Strategically**: Ensure inventory is aligned with demand in these high-revenue areas.
4. **Cross-Sell Opportunities**: Suggest related items from top categories during checkout.



**Analysis - III**

1.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.

Output: season, total\_sales

SELECT   
CASE   
 WHEN EXTRACT(MONTH FROM o.order\_purchase\_timestamp) IN (3, 4, 5) THEN 'Spring'  
 WHEN EXTRACT(MONTH FROM o.order\_purchase\_timestamp) IN (6, 7, 8) THEN 'Summer'  
 WHEN EXTRACT(MONTH FROM o.order\_purchase\_timestamp) IN (9, 10, 11) THEN 'Autumn'  
 ELSE 'Winter'  
 END AS season, SUM(oi.price) AS total\_sales  
FROM orders o  
JOIN order\_items oi ON o.order\_id = oi.order\_id  
GROUP BY season;  


### **Insights:**

1. **Spring Leads in Sales**: With over **4.2M** in revenue, Spring is the most profitable season.
2. **Summer is Close Behind**: Summer sales are also strong, suggesting good consumer activity during mid-year.
3. **Winter and Autumn Lag**: Lower sales in Winter and Autumn may indicate reduced shopping activity or fewer promotions.

### **Recommendations:**

1. **Boost Autumn/Winter Sales**: Launch seasonal promotions or campaigns to increase traction during slower months.
2. **Leverage Spring Momentum**: Plan major product launches or marketing pushes during Spring.
3. **Analyze Seasonal Preferences**: Tailor offerings by season to match customer behavior and needs.

2.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.

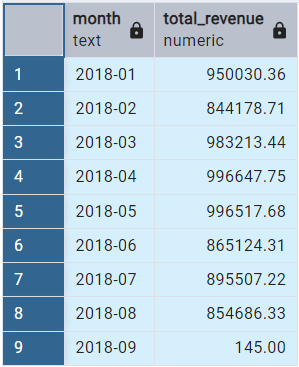
Output: product\_id, total\_quantity\_sold

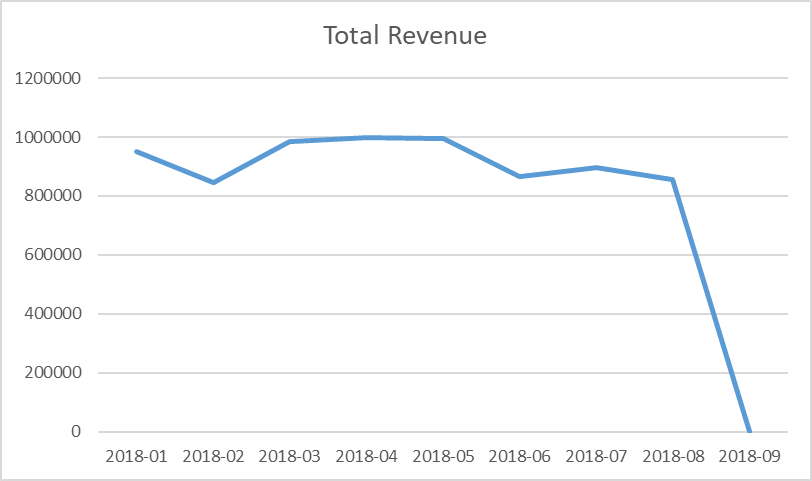
SELECT product\_id,   
 COUNT(\*) AS total\_quantity\_sold  
FROM order\_items  
GROUP BY product\_id  
HAVING COUNT(\*) > (  
 SELECT AVG(product\_sales)   
 FROM (  
 SELECT COUNT(\*) AS product\_sales  
 FROM order\_items  
 GROUP BY product\_id   
 ) AS avg\_subquery

);

3.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.

Output: month, total\_revenue

SELECT   
 TO\_CHAR(o.order\_purchase\_timestamp, 'YYYY-MM') AS month,  
 SUM(oi.price) AS total\_revenue  
FROM orders o  
JOIN order\_items oi ON o.order\_id = oi.order\_id  
WHERE EXTRACT(YEAR FROM o.order\_purchase\_timestamp) = 2018  
GROUP BY TO\_CHAR(o.order\_purchase\_timestamp, 'YYYY-MM')  
ORDER BY month;  
  




**Insights:**

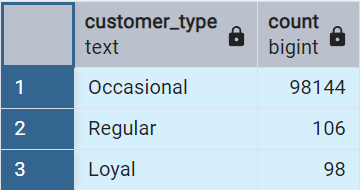
1. **Peak Sales**: The months with the highest sales were **April**, **May**, and **March**, with total revenues close to 1 million.
2. **Low Sales**: **February** and **September** had lower sales, with February being significantly lower. September, in particular, saw a sharp decline in revenue to **145.00**, which may indicate a data issue or an unexpected event.
3. **Seasonal Trends**: The months of **March through May** seem to show the highest performance, suggesting possible seasonal demand peaks in spring.

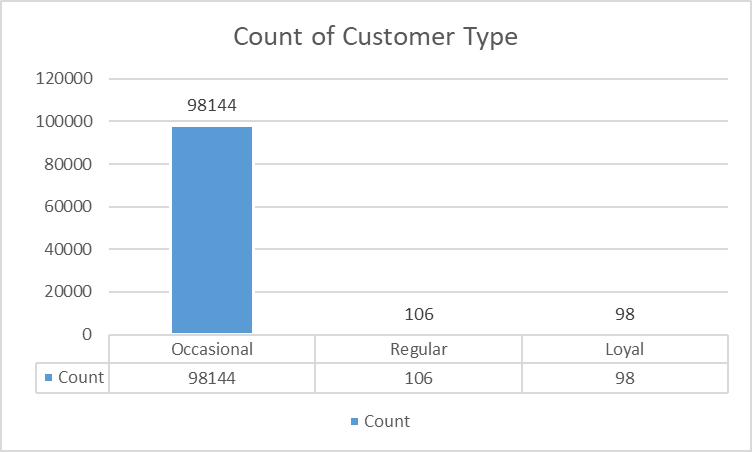
**Recommendations:**

1. Investigate the **September data anomaly** (145.00) to ensure accuracy.
2. Focus marketing and promotional efforts around the **spring months (March-May)** to capitalize on peak sales trends.
3. Consider **targeted campaigns** in **February** and **September**, given their relatively lower sales.
4. Explore whether external factors (e.g., holidays or events) contributed to the peaks and troughs.

4.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.

Output: customer\_type, count

WITH order\_counts AS (  
 SELECT customer\_id,COUNT(order\_id) AS order\_count  
 FROM orders  
 GROUP BY customer\_id  
),  
customer\_segments AS (  
 SELECT   
 Customer\_id,  
 CASE   
 WHEN order\_count BETWEEN 1 AND 2 THEN 'Occasional'  
 WHEN order\_count BETWEEN 3 AND 5 THEN 'Regular'  
 ELSE 'Loyal'  
 END AS customer\_type  
 FROM order\_counts  
)  
SELECT Customer\_type, COUNT(\*) AS count  
FROM customer\_segments  
GROUP BY customer\_type;  


  
**Insights:**

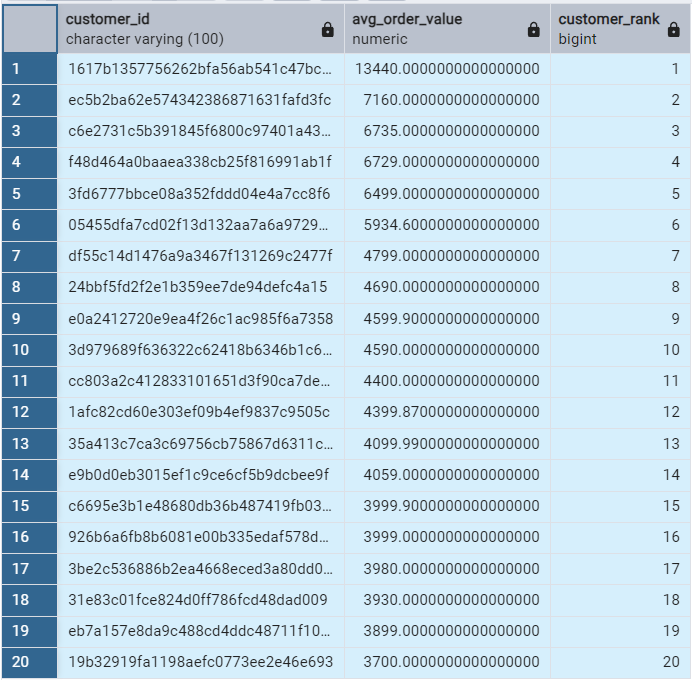
1. **Occasional Customers**: The majority of customers fall into the "Occasional" category, with **98,144** customers making 1-2 orders.
2. **Regular Customers**: A relatively small group, **106** customers, make 3-5 orders.
3. **Loyal Customers**: An even smaller group, **98** customers, have placed more than 5 orders.

**Recommendations:**

1. **Focus on "Occasional" Segment**: Since most customers are occasional buyers, offer incentives (e.g., discounts, targeted promotions) to encourage them to become "Regular" or "Loyal" customers.
2. **Loyalty Program Tailoring**: Create exclusive rewards for "Loyal" customers to retain them and increase customer lifetime value.
3. **Promote Regular Purchases**: For the "Regular" segment, offer personalized deals or notifications to encourage them to become "Loyal."
4. **Proportion Chart**: In Excel, create a pie chart to visualize the distribution of each segment:
   * "Occasional" should dominate the chart.
   * "Regular" and "Loyal" segments will form smaller portions.

5.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.

Output: customer\_id, avg\_order\_value, and customer\_rank

WITH order\_totals AS (  
 SELECT   
 O.customer\_id,  
 O.order\_id,  
 SUM(oi.price) AS order\_value  
 FROM orders o  
 JOIN order\_items oi ON o.order\_id = oi.order\_id  
 GROUP BY o.customer\_id, o.order\_id  
),  
customer\_avg\_order AS (  
 SELECT   
 Customer\_id,  
 AVG(order\_value) AS avg\_order\_value  
 FROM order\_totals  
 GROUP BY customer\_id  
)  
SELECT   
 Customer\_id,  
 avg\_order\_value,  
 RANK() OVER (ORDER BY avg\_order\_value DESC) AS customer\_rank  
FROM customer\_avg\_order  
LIMIT 20;  
  
**Insights:**

* The top customer spends **13,440** on average, with the top 5 ranging from **13,440** to **6,499**.
* The 20th customer has an average of **3,700**.

**Recommendations:**

* Target the top 20 customers for an exclusive rewards program.
* Analyze their spending patterns for personalized marketing.
* Create a bar chart in Excel to visualize their ranks and order values.

6.Amazon wants to analyze 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.

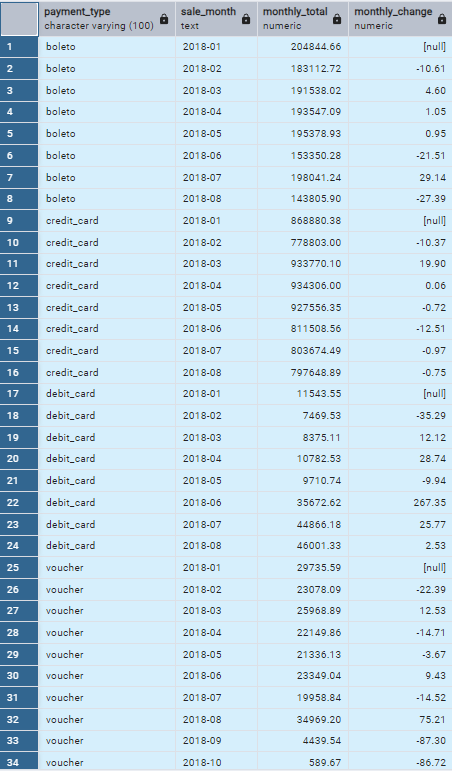
Output: product\_id, sale\_month, and total\_sales

WITH recursive product\_sales AS (  
 SELECT   
 oi.product\_id,  
 DATE\_TRUNC('month', o.order\_purchase\_timestamp) AS sale\_month, SUM(oi.price) AS monthly\_sales  
 FROM order\_items oi  
 JOIN orders o ON oi.order\_id = o.order\_id  
 GROUP BY oi.product\_id, DATE\_TRUNC('month', o.order\_purchase\_timestamp)  
),  
recursive\_sales AS (  
 SELECT ps.product\_id, ps.sale\_month,ps.monthly\_sales, ps.monthly\_sales AS total\_sales  
 FROM product\_sales ps  
 WHERE NOT EXISTS (  
 SELECT 1 FROM product\_sales ps2   
 WHERE ps2.product\_id = ps.product\_id   
 AND ps2.sale\_month < ps.sale\_month  
 )  
UNION ALL  
 SELECT ps.product\_id,ps.sale\_month,ps.monthly\_sales,  
 rs.total\_sales + ps.monthly\_sales AS total\_sales  
 FROM product\_sales ps  
 JOIN recursive\_sales rs   
 ON ps.product\_id = rs.product\_id   
 AND ps.sale\_month = rs.sale\_month + INTERVAL '1 month'  
)  
SELECT product\_id, sale\_month, total\_sales  
FROM recursive\_sales  
ORDER BY product\_id, sale\_month;

7.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.

Output: payment\_type, sale\_month, monthly\_total, monthly\_change.

WITH monthly\_sales AS (  
 SELECT payment\_type,  
 TO\_CHAR(o.order\_purchase\_timestamp, 'YYYY-MM') AS sale\_month, SUM(p.payment\_value) AS monthly\_total  
 FROM payments p  
 JOIN orders o ON p.order\_id = o.order\_id  
 WHERE EXTRACT(YEAR FROM o.order\_purchase\_timestamp) = 2018  
 GROUP BY payment\_type, TO\_CHAR(o.order\_purchase\_timestamp, 'YYYY-MM')  
),  
sales\_with\_lag AS (  
 SELECT   
 payment\_type,  
 sale\_month,  
 monthly\_total,  
 LAG(monthly\_total) OVER (PARTITION BY payment\_type ORDER BY sale\_month) AS prev\_month\_total  
 FROM monthly\_sales  
)  
SELECT payment\_type,  
 sale\_month,  
 monthly\_total,  
 ROUND(100.0 \* (monthly\_total - COALESCE(prev\_month\_total, 0)) / NULLIF(prev\_month\_total, 0), 2 ) AS monthly\_change  
FROM sales\_with\_lag  
ORDER BY payment\_type, sale\_month;



**Insights:**

* **Boleto** and **Credit Card** show fluctuating sales, with Boleto peaking in July (+29.14%) and Credit Card declining from March onwards.
* **Debit Card** has strong growth, especially in June (+267.35%).
* **Voucher** experiences sharp declines in Sept (-87.30%) and Oct (-86.72%).

**Recommendations:**

* Boost promotions for **Boleto** and **Voucher** during low growth months.
* Focus on growing **Debit Card** sales.
* Investigate reasons behind **Credit Card** decline.
* Visualize data in Excel for clearer trend analysis.