Shop Nest Store Capstone

Data Cleaning:

Here's a summary of the datasets and the presence of null values and duplicates:

1. **Orders Dataset**:
   - Null Values: Some columns like order_delivered_carrier_date and order_delivered_customer_date have
   - Duplicates: None.
2. **Products Dataset**:
   - Null Values: Multiple columns such as product_category_name, product_name_lenght, product_description_lenght, etc., have nulls
   - Duplicates: None.
3. **Sellers Dataset**:
   - Null Values: No null values.
   - Duplicates: None.
4. **Product Categories Dataset**:
   - Null Values: No null values.
   - Duplicates: None.
5. **Geolocation Dataset**:
   - Null Values: No null values in most columns, except for geolocation_city and geolocation_state (but these have minimal nulls).
   - Duplicates: 261,836 duplicate rows.
6. **Customers Dataset**:
   - Null Values: No null values.
   - Duplicates: None.
7. **Order Items Dataset**:
   - Null Values: No null values.
   - Duplicates: None.
8. **Order Payments Dataset**:
   - Null Values: No null values.
   - Duplicates: None.
9. **Order Reviews Dataset**:
   - Null Values: Columns review_comment_title and review_comment_message have significant nulls
   - Duplicates: None.

Set Up Data Model

- **Create relationships**: Once all datasets are imported, create relationships between them in the **Model** view. For example:
  - Link orders_dataset with order_items_dataset via order_id.
  - Link order_items_dataset with products_dataset via product_id.
  - Link order_payments_dataset with orders_dataset via order_id.
  - Link order_reviews_dataset with orders_dataset via order_id.
  - Link geolocation_dataset with orders_dataset via customer_id or any other relevant relationship.

Dashboard Report

1. Total Orders by State

- **Description**: This section represents the total number of orders placed from each state. The metric Total order by State provides an overview of the geographical distribution of the orders, showing that there are **9,358 total orders** across all states. This visualization can be used to identify which state has the highest number of orders, helping to assess regional sales performance.

2. Average Rating of Orders

- **Description**: The **Average Rating** for products is calculated to be **4**, which indicates a generally positive customer experience. This value is derived from the review_score in the order_reviews_dataset. The bar chart visualizing the Order_id by review_score reveals that most products have received a rating of 5, which represents the best customer feedback.
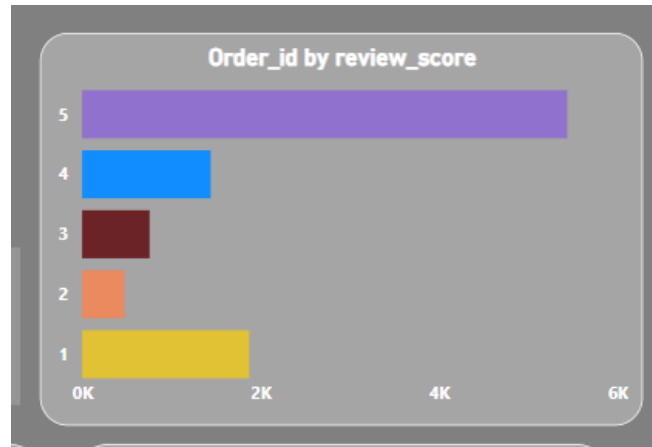
3. Order Status by Delayed Orders

- **Description**: The **Order Status** analysis provides insight into the percentage of delayed orders. According to the dashboard, **5% of the orders were delayed**, while the remaining **95% of the orders were delivered on time**. The pie chart represents this distribution, with the orange section representing delayed orders and the grey section representing on-time orders. This information is valuable for improving delivery timelines and operational efficiency.
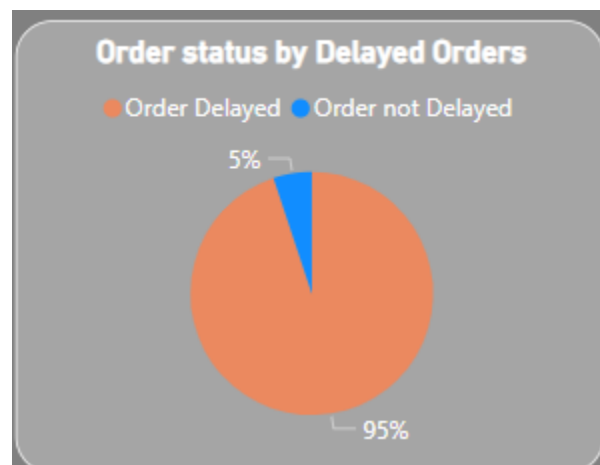

4. Count of Order IDs by Delayed Orders

- **Description**: This chart shows the **Count of Order IDs** by their delay status. The bar chart reveals that there is a significant difference in the number of on-time versus delayed orders. This comparison helps in understanding how the business is performing in terms of timely deliveries and identifying areas for improvement in the supply chain.

Count of order_id by Delayed Orders
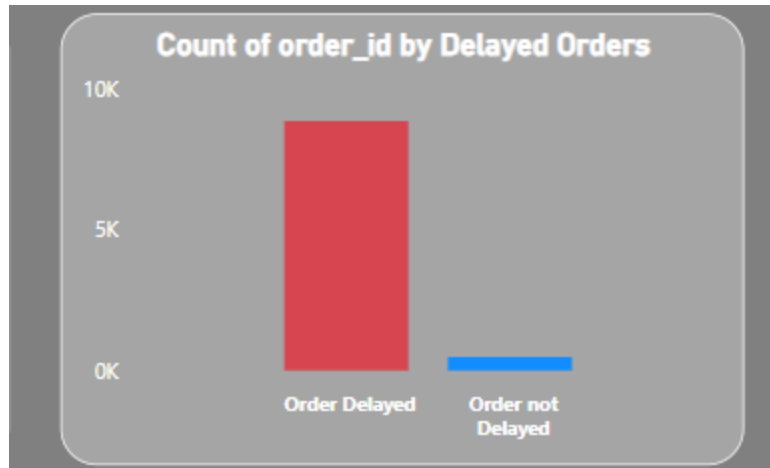
5. Payment Method Breakdown

- **Description**: The **Sum of Payment Sequential by Payment Type** pie chart shows the breakdown of payment methods used by customers. The data reveals that **68% of the payments** were made via **credit card**, followed by **13% using boleto** and **17% using debit cards**. This analysis provides insights into customers' preferred payment methods, allowing for targeted marketing strategies and payment method optimization.

Sum of payment_sequential by payment_type
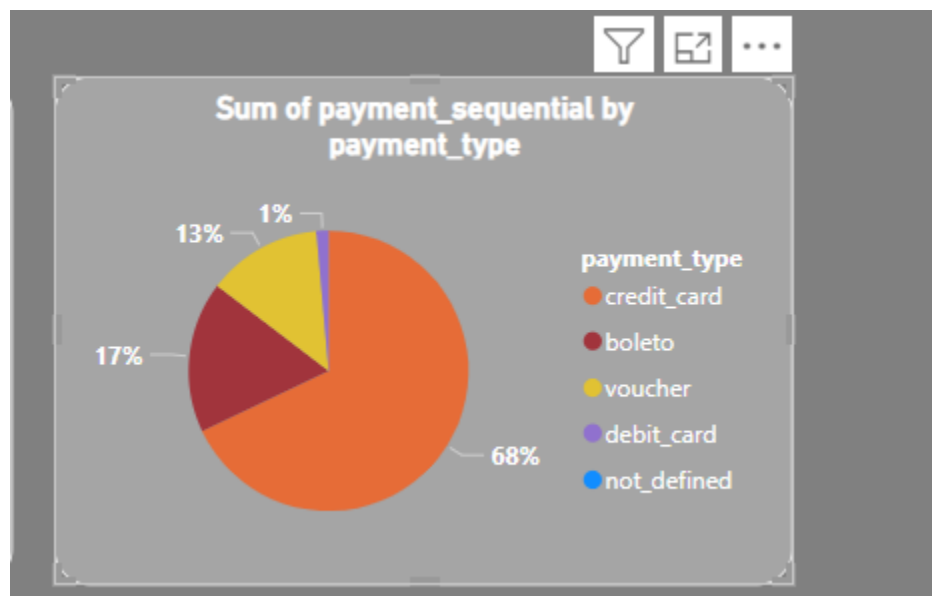
6. Total Sales by Product Category

- **Description**: The **Sum of Total Price by Product Category** bar chart visualizes the total sales across different product categories. The top category, **"cama_mesa_banho"**, leads with the highest sales. The chart helps identify the most and least profitable product categories, providing valuable information for inventory and sales strategy planning.