<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data>

Handling duplicates:

The database has no duplicated value and if there are similar Data rows, it depends on the definition conditions in the original dataset. (go to the database documentation)

Customer segmentation

1. Based on location:

See code block 6 – CSV is stored: 3

1. Based on payment method/payment method + location code block 7 and 8
2. Based on location and product category code block 9
3. Code block 10 gives you changes in revenue and cost labeled by season
4. Code block 11 defines the total value of revenue based on date and location
5. Code block 12 Give you total of revenue and cost based on year and season

Sale Performance

1. code block 13: total Sales and freight per category
2. code block 14: total Sales and revenue per Seller
3. code block 15: most sold products
4. code block 16: most gained revenue from each product
5. code block 17: revenue per capita for each product
6. code block 18: (based on price - freight value - warehousing cost considered by the size of product)

bigger products have more warehousing costs, filtering values if a product had a high number of sales and high price and low weight and low size, is a super beneficial product

Product analysis

1. code block 19: of canceled products
2. code block 20: infographic of orders
3. code block 21: infographic of canceled orders based on location
4. code block 22: Infographic of canceled products based on location
5. code block 23: Infographic of canceled products based on product category

seller analysis

1. code block 24: a rolling total for getting a rolling total of sellers in each city.
2. Code block 25: average seller revenue per city
3. Code block 26: geolocation performance of seller >> to see each seller has customers from how many locations.