The number of clients who ordered products around Brazil

Goal:

To effectively develop e-commerce, it is essential to assess market penetration in various regions to identify areas of high and low activity. This will aid in identifying expansion opportunities in regions adjacent to these areas. Data on Brazil's e-commerce market was obtained for analysis for 2018.

Insights:

These areas have the highest number of clients placing orders. In these areas, the population has “trust” in e-commerce, which is a very important factor for order volume (“trust” is one of the important factors in marketing). High e-commerce activity is concentrated in urban centers such as São Paulo and Rio de Janeiro (11200 and 3924, respectively), as well as regions such as coastal areas and inland cities such as Porto Alegre and Belo Horizonte (1100 and 2982, respectively), while other regions show significant disparities in digital commerce participation. Historically, the high population density in coastal areas has been associated with a greater proportion of trade.

Data Abstraction:

For optimal e-commerce enhancement, analyze market penetration across different regions to pinpoint areas of strong and weak activity, and explore growth prospects in areas neighboring these high-density regions. I have analyzed only 2018 year because in this year the activity of costumers in commerce increased dramatically (information given in the second report which analyzed customers review).

Dataset Type: The dataset is a table resulting from the merging of three tables (olist_geolocation_datasett.csv ,olist_orders_datasett.csv and olist_customers_datasett.csv).

olist_geolocation table consists of 5 columns and given explanation only used columns:

geolocation_zip_code_prefix: An integer representing the zip code prefix for the geolocation.

geolocation_lat: A float representing the latitude coordinate of the geolocation.

geolocation_lng: A float representing the longitude coordinate of the geolocation.

olist_orders table contains 8 columns and given explanation only used columns:

order_id: A unique identifier for each order.

order_delivered_customer_date: The date the order was delivered to the customer.

customer_id: A unique identifier for each customer.

olist_customers contains 5 columns and given explanation only used columns:

customer_id: A unique identifier for each customer.

customer_zip_code_prefix: The zip code prefix for the customer's address.

customer_state: The state of the customer.

geolocation_zip_code_prefix , **customer_zip_code_prefix** and **customer_id** were used to join three tables together (olist_geolocation, olist_customers and olist_orders).

Items:

Each bar in the graph represents an item, including the order quantity and geolocation of orders.

Data Types of Attributes:

- Order quantity (Quantitative, Sequential).

Task Abstraction:

For optimal e-commerce enhancement, analyze market penetration across different regions to pinpoint areas of strong and weak activity, and explore growth prospects in areas neighbouring these high-density regions.

***Marks*:**

Data points: number of clients in different cities

***Channel*:**

Position: Placement of the marks on the map corresponds to the geographic location of the cities.

Size: The size of the marks indicates the relative number of clients in each city.

Color: The marks' color seems to be within a gradient that may correspond to the number of clients, with a separate color legend indicating that the gradient goes from 0 to 30,000 in order quantity.

Action and Target:

Level	Action	Target	
High-Level	**Identify**	**Outliers**	Discover areas outliers with the highest or lowest e-commerce activity to understand regional market dynamics and consumer engagement in online shopping
Mid-Level	**Lookup**	**One attribute**	Consumers lookup since the location and target of data are known.
Low-Level	**Compare**	**Features**	Stakeholders such as the marketing manager and data analysts compare the dynamics of market neighboring areas and make future financial plans for increasing activity of consumers.

Additional data source:

- In this project, **Python** was used for analyzing data.
- **URL** for data: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce