# Brazilian E-commerce Dataset

This is a dataset of Brazilian e-commerce transactions, containing information about orders, customers, reviews, and products. The dataset was obtained from Kaggle and includes data from September 2016 to October 2018.

# Contents

The dataset includes the following files:

*olist\_customers\_dataset.csv:* information about customers, including their unique IDs, customer city, and customer state.

*olist\_geolocation\_dataset.csv:* geographic information about customers and sellers, including latitude and longitude coordinates.

*olist\_order\_items\_dataset.csv:* information about the items included in each order, including product ID, seller ID, price, and freight value.

*olist\_order\_payments\_dataset.csv:* information about payments made for each order, including payment type and value.

*olist\_order\_reviews\_dataset.csv:* information about reviews written for each order, including review score and comments.

*olist\_orders\_dataset.csv:* information about orders, including order ID, customer ID, order status, and order purchase date.

*olist\_products\_dataset.csv:* information about products, including product ID, product category, and product price.

olist\_sellers\_dataset.csv: information about sellers, including seller ID, seller city, and seller state.

# Data Analysis

This dataset can be used to analyse various aspects of Brazilian e-commerce, including:

* Customer behaviour and demographics
* Popular products and categories
* Order fulfillment and shipping times
* Payment types and fraud prevention
* Seller performance and reputation

# SQL and Power BI Analysis

To analyse this dataset, we can use SQL to query the data and Power BI to create visualizations. Here are some examples of analyses that can be performed:

## SQL Extraction and finding patterns

To make it easier for you to access and analyses the data, we have loaded the dataset into the target dataset. All the SQL queries that were performed on this dataset can be found in the "Brazilian E-commerce.sql" file. This file contains a series of SQL commands that were used to extract and analyse various pieces of information from the dataset, allowing you to gain valuable insights into the behaviour of customers and the performance of the e-commerce platform. By reviewing the queries in this file, you can learn more about the structure and content of the dataset and start to form your own hypotheses about the factors that drive e-commerce success.

## Power BI

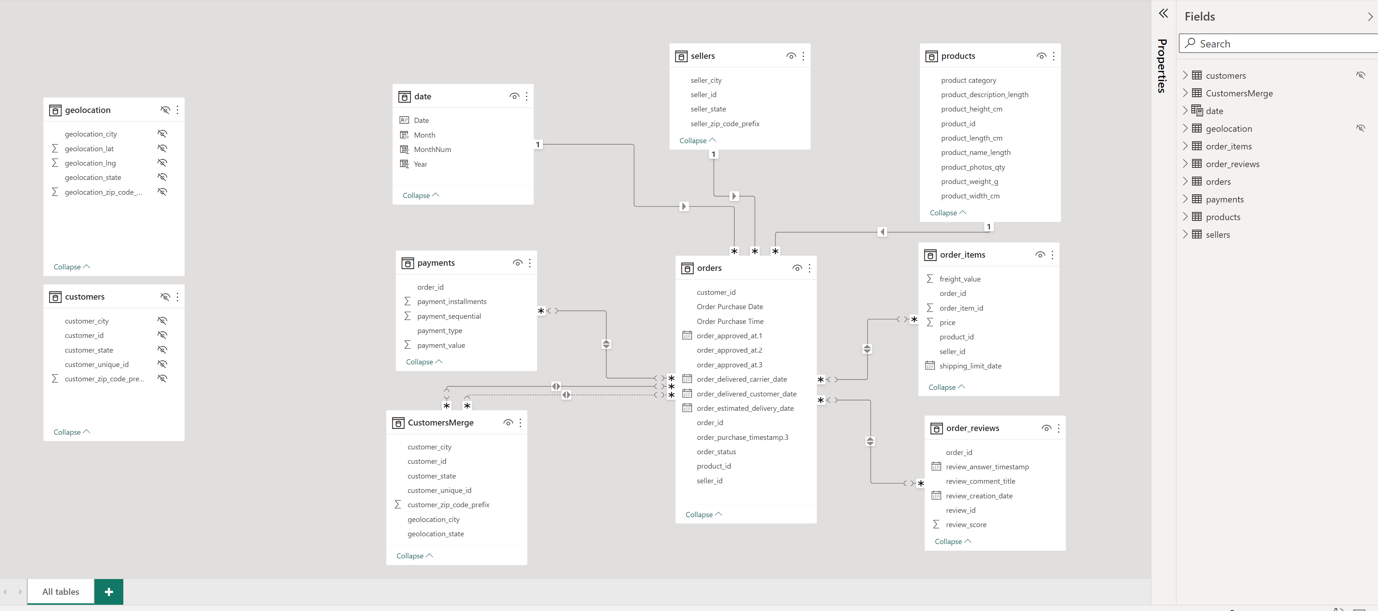
## Data Model:

A star schema is a type of data modelling technique that is commonly used in data warehousing. In a star schema, the data is organized into a central fact table and a set of dimension tables. The fact table contains the measures or metrics, such as sales, revenue, or any other quantitative data that is being analysed. The dimension tables contain the attributes or characteristics that describe the measures in the fact table, such as time, product, location, customer, etc.

The star schema is so named because the diagram of the schema resembles a star, with the fact table at the centre and the dimension tables radiating out from it like the points of a star. This structure is intended to make querying and analysis of the data as fast and efficient as possible, since it minimizes the number of tables that need to be joined and allows for easier aggregation of the data.

Star schemas are often used in business intelligence and data analytics applications, as they provide a simplified and optimized structure for querying and analysing large datasets. Additionally, star schemas are easily scalable since new dimensions can be added without affecting the structure of the fact table.

Star schema for the dataset.



# License

This dataset is licensed under the Open Database License. You are free to use, share, and adapt the data, but you must give attribution to the original source and make any changes publicly available under the same license.