Retail data analysis & network location planning

1. Problem description

Online retailers sell a vast number of products to a disperse set of customers. Often only few central warehouses are operated to store the products intermediately. The Olist data set contains data of Brazilian online retailer. A description of the data set can be found here https://www.kaggle.com/olistbr/brazilian-ecommerce. Please download the data directly from kaggle. An illustration of the retailer homepage and an overview of the data set is shown here:

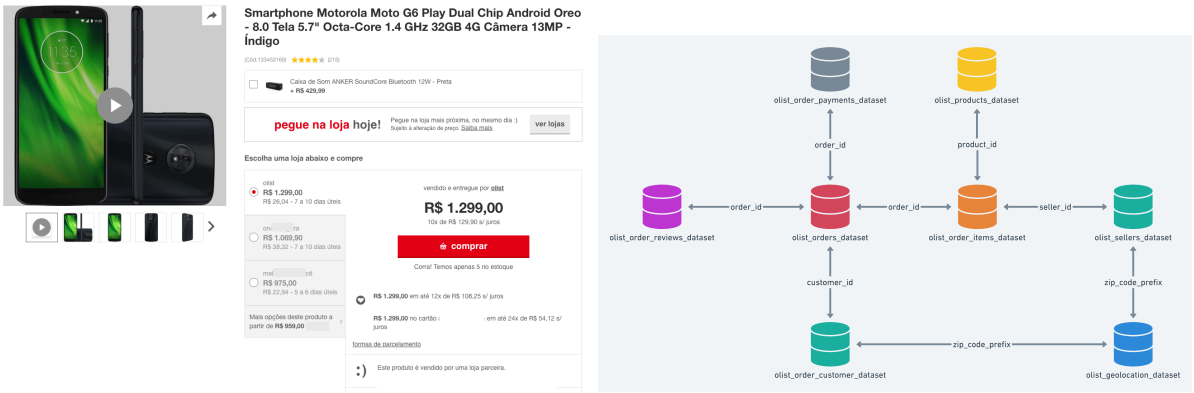

Figure 1: Sampled product on retail homepage and structure of the data set

The data set consists of different data bases containing customer data, order data, product data, geolocation data and some more. Basically, an order consists of the product information, a customer, and the delivery information.

2. Task

The task is to analyze the order data. Data about sellers, payments, and reviews can be omitted. Compose a report that displays the temporal and spatial distribution of customer orders w.r.t. to economic criteria (e.g., order value, delivery frequency etc.). Identify most relevant products and customer areas. Based on the geoinformation of customer locations, find potential (geo)locations of 5 warehouses in Brazil and assign customer locations to these warehouses by minimizing distance.