Webscrapping of Supermarkets Products Information Related

Group members:

André Oliveira 20222156

Rafael Chamusca 20222174

Vasco 20222155

Diogo 20222152

Israel Lucena 20211720

Abstract

The problem being explored is one lives everyday when thinks where is the best place to buy groceries. Often, one does not know what is the best place (the cheapest) to buy food. We will focus on the data cleaning, exploration, and visualization of supermarkets focusing mainly on groceries. Namely, Pingo Doce, Continente and Auchan.

Since the main goal of the project is to get, clean and explore the data, our main focus will be to retrieve and clean the data. In visualization, we can compare the different prices of the supermarkets focusing our analysis on the well-know "basic food basket" (or the recently 0% "IVA" products).

One problem we foresee is the different categorization each supermarket does of products. In that sense, we expect as our main difficulty to be the cleaning and arranging the data. From a quick research about webscrapping limitations we found that these are very short. So, the data will be retrieved from the Supermarket's websites. For example, from here to Pingo Doce, here to Continente, and here to Auchan.