

Supermarket sales

Exploratory Data Analysis for Machine Learning Course Project

Caio Mescouto Terra de Souza

May 11, 2021

Brief description of the data set and a summary of its attributes

Supermarket sales data set [3] is a historical record of sales data in 3 different Branches for 3 months. Each observation has 17 features that indicates where the purchase happen (Branch, City), when (Date and Time), the Product Line, Unit Price, Quantity, Payment method, Cost of goods sold, margins, Taxes and Customer informations (Gender and Type) in addition to stratification rating on their overall shopping experience.

Branch and City represents the same information, so we will keep only one for the analysis. Tax and gross margin have no variance, the exploratory data analysis on margins or taxes are not meaningful and we will drop all related variables.

Initial plan for data exploration

According with the available attributes, we can investigate the relationship of customers and purchase, differences between branches, day of the week and time of the day and purchase, for example.

References

- [1] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The L^AT_EX Companion*. Addison-Wesley, Reading, Massachusetts, 1993.
- [2] Albert Einstein. *Zur Elektrodynamik bewegter Körper*. (German) [*On the electrodynamics of moving bodies*]. Annalen der Physik, 322(10):891–921, 1905.

- [3] Supermarket sales: Historical record of sales data in 3 different super-markets,
<https://www.kaggle.com/aungpyaeap/supermarket-sales>