Report on Snickers and Coca-Cola Prices

Lagos Team

7 November 2021

Introduction

Can we encounter lower prices in the outer districts then in the inner districts? What if we go farther away from the city center? In order to examine these questions, product prices were collected from two districts of Budapest, the 13^{th} and the 10^{th}, representing the inner and outer parts of the city. Our objective with the data collection was to compare retail prices of two chosen products, Coca-Cola 500ml plastic bottle and Snickers 50gr chocolate bar in the two areas. In order to examine the differences in the price distributions, product tags were gathered from shops of different sizes including supermarkets, gas stations, as well as small local groceries. Our data set and artifacts can be found on our Github repo

Data Quality and Cleaning

The focus of our analysis is on the prices of the two products. In order to increase the comparability between the registered prices, we recorded if the products were offered at a discounted price. Items with discounted price were later excluded from our analysis (2 records). Additional data wrangling was done in order to prepare our data for analysis: missing values were encoded with the conventional NA, product names were shortened and all the variables in the dataset were assigned proper data type. Moreover, we created two groups from the shop types, where we aggregated the shops into big supermarket chains and small local groceries categories.

Descriptive Statistics

To further understand and analyze the price differences between the the two selected districts, a summary statistics table was created.

Table 1: Descriptive statistics of product prices by district

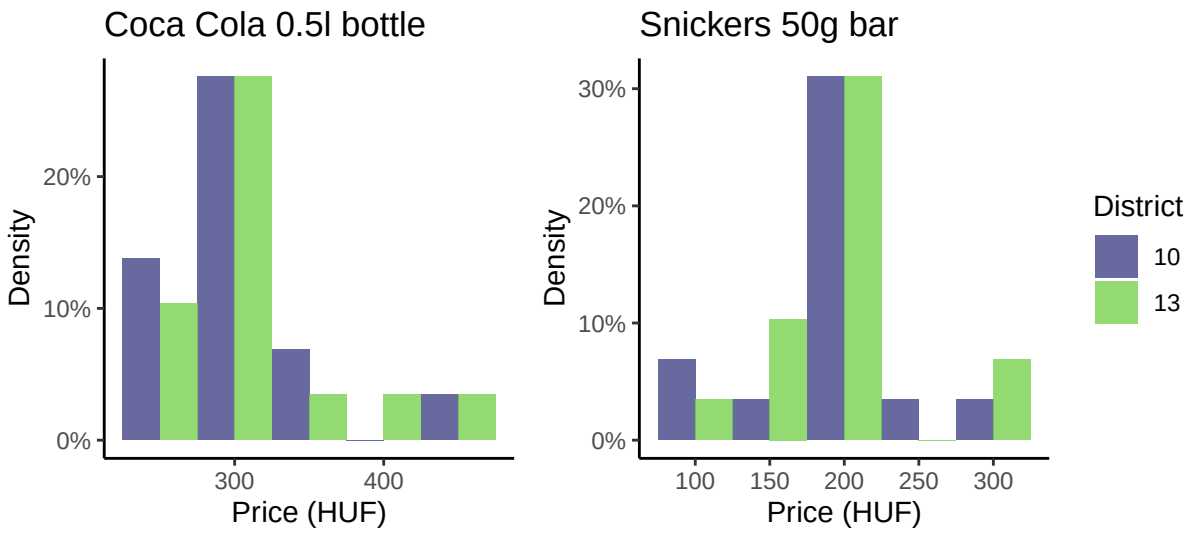
District		N	Missing	Mean	Median	SD	Min	Max	Range	P05	P95
10	Coke 0.5l	15	0	304.5	299	52.43	230	449	219	244	380
	Snickers 50g	14	1	193.8	189	49.50	119	319	200	120	274
13	Coke 0.5l	14	1	318.4	304	52.63	259	439	180	259	426
	Snickers 50g	15	0	192.6	180	46.96	119	299	180	147	299

As per the summary statistics, the mean price of Coke is slightly higher in the inner 13^{th} district (HUF 318.4) compared to the 10^{th} district (HUF 304.5). As for Snickers, we cannot see considerable difference between average prices (HUF 1.2). Regarding the spread of prices, there's less variation in Snickers' prices compared to Coke's. The range of prices for both the products were in the interval of 180 to 219, while 95% of the Coke prices were lower than 380 and 426 for the 10th and 13th district, respectively.

Price Distribution based on District

We chose histograms to show the distribution of prices for both the products in each district. We set the bin size to HUF 50 and the positioning of bars to 'dodged', as it is easier to interpret a HUF 50 step in prices, while it did not impose additional clutter.

Overall, Coke prices have distribution with long right tails in both districts, while in case of Snickers, the distribution is centered around HUF 200.The majority of the shops offer Coke in the price interval between HUF 275 and 325 in both district while in case of Snickers the mode is at the 175-225 price bin.

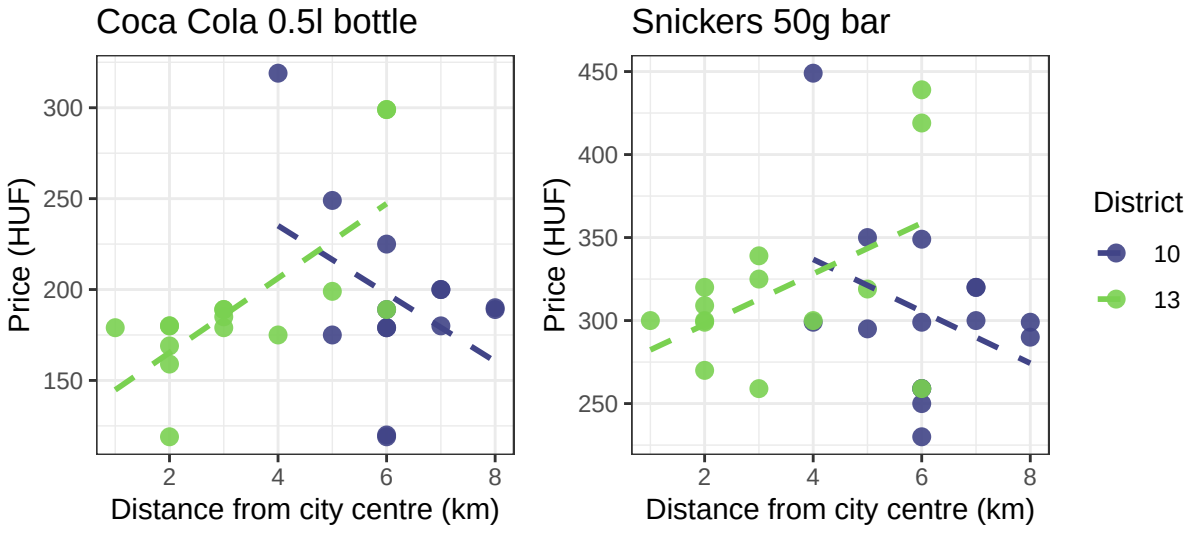


Observations with extreme price values at the tail of the distribution are driven by the prices of gas stations.

Price Distribution based on Distance

We used scatterplots to display our findings on the relationship between the distance from city center and the product prices.

As shown in the graph below, for both products there is a positive association between the distance and the price in the 13^{th} district. In contrary to that, 10^{th} district prices tend to decrease with distance.



Conclusion

To sum up, our analysis indicates that prices in the inner district are generally higher than the outer districts of Budapest. After examining the relationship between distance and price, we can conclude that there exists an upward tendency in the 13^{th} district, as opposed to the 10^{th} district where the relationship seems negative.

NOTE: There are two further analysis (box whisker plot & t-test) made as a part of this exercise in R Markdown but included in the report.