

# Budapest Pizza Report

KFC (Kata, Fanni, Cosmin) Team

October 25th, 2020

The purpose of the report is to explore data about pizza prices in Budapest. The following repository contains the dataset, walkthrough explanation of the variables and the R codes used to create this descriptive analysis: [GitHub](#)

## Data collection, difficulties & problems

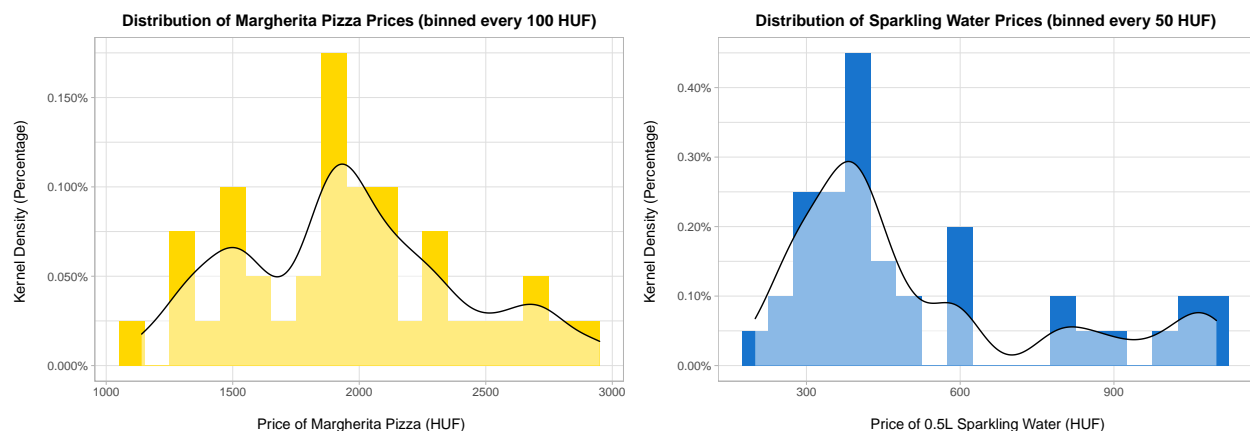
The KFC team chose to limit the data collection to the Budapest area, with on-site prices being gathered only within the inner city districts (i.e. 2, 5, 6, 7, 8, 9, 11, 13), while delivery prices were documented irrespective of inner city district limits. The Google Maps API was used in order to compile the list (custom Google Maps shared lists) of on-site locations, gather customer reviews and compute the travel time between the restaurant and CEU. The choice of 0.5L sparkling water was convened with the drawback that the distribution of prices, brands and mark-up would be more heterogeneous than with the margherita pizza.

Variables like district, address and restaurant were added for reference purposes and to be able to have a unique identifier for every pizza place in case the dataset were to ever be expanded. The binary indicator variable 'delivery' which differentiates between restaurants' menu and online merchants' prices. The creation of this variable leads into the 'delivery\_site' column which designates the online vendor. Lastly, since distance from the CEU Nador 15 building is hard to quantify with regards to driving a car (as there are numerous one-way streets and other detours which might add unnecessary distance units), the shortest walking distance was chosen and the metric of choice is minutes.

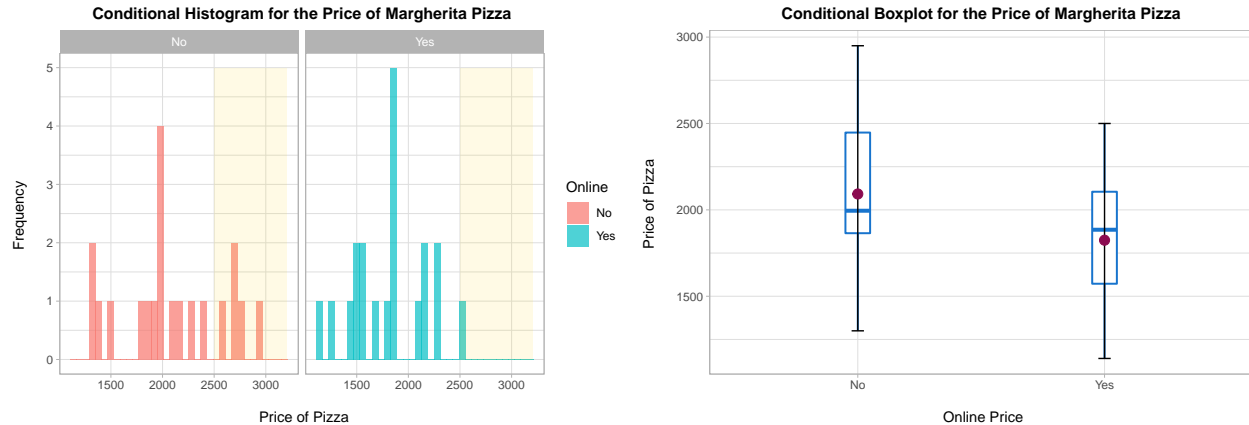
## Descriptive statistics for prices of margherita pizza and sparkling water

### Histograms of product prices (kernel density showcases skewness)

The following plots describe the distribution of margherita pizza prices and sparkling water prices.



## Visualizations of the price of margherita pizza conditioning on online vs offline price



## Test of whether the online and offline prices of margherita pizza are the same

### Summary of findings regarding unconditional and conditional price distributions

Based on the histograms of product prices it can be clearly seen that even though the price distribution for margherita pizza tends to have a few smaller modes, it closely resembles a normal distribution. With the help of binning, the distribution can be brought to look much closer like a normal distribution. The same cannot be said about the distribution of sparkling water prices, which showcases a heavy right skew, having a mean much higher than the median. This is perhaps caused by the large differences in sparkling water brands (from local Hungarian to expensive imported Italian water). As such, the range of products is much higher than with respect to the pizza. For further analysis, it would be worthwhile to inspect distributions of local sparkling water prices and imported water prices in separate graphs.

As for the conditional distribution of the price of margherita pizza it can be concluded that both the online and offline price distributions approximate a normal distribution. Furthermore, when comparing the two we can see that offline prices are higher on average than online prices since the mean of the former is higher than the latter. Also, on the histogram it is visible that the maximum online price is 2500 HUF while the maximum offline price is almost 3000 HUF. Lastly, the boxplot shows that the spread of offline prices is greater than that of the online prices.