Pizza Data Analysis

Eszter - Viki - Ozan

2020 October 23

Data Collection

Due to the pandemic situation, the data collection was made exclusively online, thereby, the time spent on research reduced significantly. Due this reason, several values such as number of tables and waiters, the size of the restaurant measured by square meter were excluded in the end. The aim of those data was to examine the exclusivity of restaurant which was solved by collecting different rating on websites. On the other hand, some data on prices were less excessive in smaller restaurants that do not delivery and are also absent on the internet which might influence the outcome of the analysis negatively. The main purpose of the analysis is to investigate the differences in cola and margarita prices on Budapest and countryside cities such as Debrecen, Nyíregyháza, Szeged, etc. The cola prices were chosen for comparison because this soft drink was served in all examined restaurants. In the following parts, the result of the full analysis will be presented shortly and a hypothesis regarding the pizza prices. Daata and further information can be found on this github repository: Github

Descriptive statistics

The below table contains a summary about the main statistic connected to the collected data about the margarita pizza and cola prices. The dataset includes a total of 69 observation from which the following was concluded: the distribution is asymmetric considering the fact it possesses a positive skewness which also suggest a right long tail. However, due to the problem of collecting unified prices on 500 ml Cola, the summary also shows a conclusion of all the cola also an estimated price of them. In other words, the estimated cola prices were assumed by calculating the price/500 ml.

variable	mean	median	std	iq_range	min	max	skew	numObs
Margarita Pizza	1733.00	1670.00	338.88	500.00	1190.00	2590.00	0.52	69
Cola	487.63	450.00	206.85	180.30	196.00	1160.00	1.42	61

Table 1: Summary statistics for the margarita and cola prices

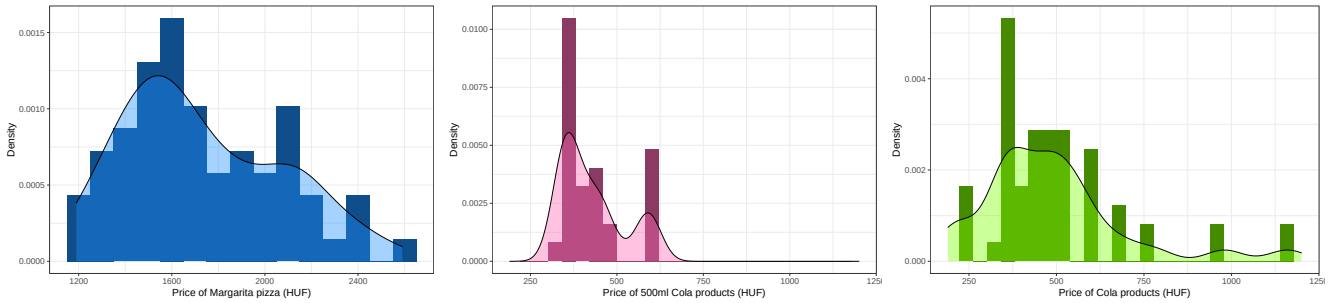
variable	mean	median	std	iq_range	min	max	skew	numObs
500ml Cola Actual	428.06	400.00	91.08	110.00	340.00	590.00	0.87	31
500ml Cola Estimated	549.19	530.30	268.75	250.38	196.00	1160.00	0.69	30

Table 2: Summary statistics for the actual 500ml cola prices and the estimated ones

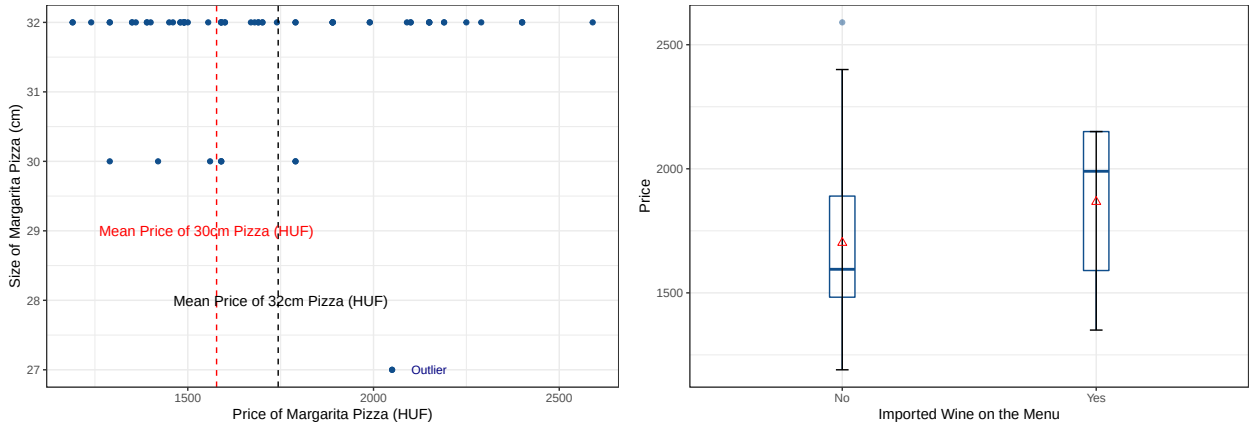
Distribution of Margarita and cola prices

As it was mentioned before, the statistics suggested a right long tail, a positive skewness which is presented in the following histograms. The main difference between the cola histograms are that the left contains the

estimated prices of all colas in 500 ml units while the right shows all of the prices even if they are 330ml, 1000ml or 1250ml bottle. The central tendency of the prices of margarita pizza are between 1400 HUF and 1600 HUF while it is around 350 HUF and 400 HUF for the estimated cola prices. On the other hand, there is an unusual increase at 600 HUF in estimated cola prices which might require further analysis in the future because the high tendency suggest a serious reason behind it.

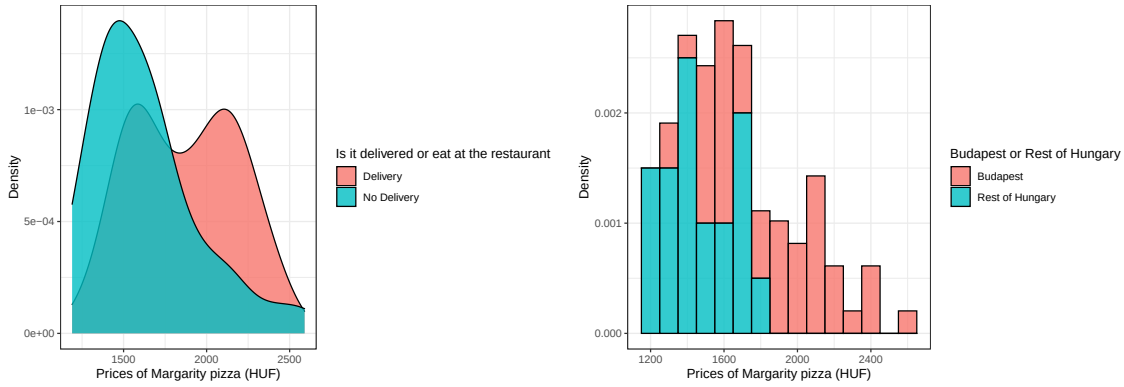


Margarita pizza prices according to their size and on restaurants with different qualities



Prices of Margarita pizza online vs offline and Budapest vs countryside

The following charts present the comparison of the online and offline price and the distribution of prices in Budapest and countryside. As it is presented, the delivery prices range higher than on-delivery prices while the Budapest prices excess the countryside prices.



Hypothesis testing if the mean price online vs offline is the same

```
##
##  Welch Two Sample t-test
##
## data:  margarita_price by delivery
## t = 3.0685, df = 63.384, p-value = 0.003163
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   82.82589 392.03565
## sample estimates:
##    mean in group Delivery mean in group No Delivery
##                  1867.200                  1629.769
```

Here we see that we reject the null hypothesis that the mean pizza price we observe online is the same as those we observe in restaurants (t = 3.0685, df = 63.384, p-value = 0.003163). The average pizza price online is 1867.200 and for those offline is 1629.769. The 95% confidence interval for the difference in prices between the two groups is (82.82589, 392.03565).

```
##
##  Welch Two Sample t-test
##
## data:  cola_price by delivery
## t = 0.51646, df = 54.09, p-value = 0.6076
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -80.49158 136.35358
## sample estimates:
##    mean in group Delivery mean in group No Delivery
##                  502.7446                  474.8136
```

For the difference between online and offline cola prices on the other hand, we cannot reject the null that the average of these two prices is not the same (t = 0.51646, df = 54.09, p-value = 0.6076). There is strong evidence suggesting that online and offline beverage prices are similar on average.