

Pizza Data Analysis

Eszter - Viki - Ozan

2020 October 23

Data Collection

Due to the pandemic situation, we decided to collect the data for our research exclusively from online sources. This had the advantage, that we did not have to visit several restaurants, which would have been time consuming but also caused some difficulties and also some changes in our initial set of variables. We were not able to record the area of the places, the number of waiters or the number of tables, with which we wanted to measure the exclusivity of the places, instead we decided to collect rating on Wolt, NetPincér, Tripadvisor, Google and also Facebook, as well as the number of reviews for reliability. . . .

Descriptive statistics

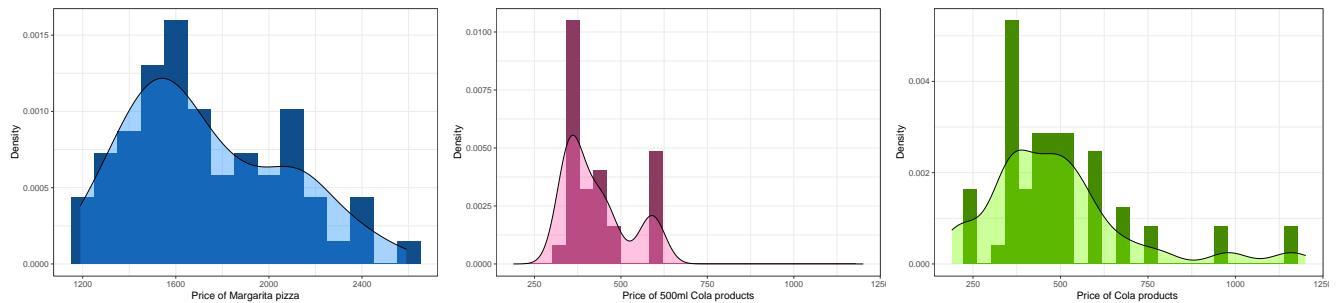
variable	mean	median	std	iq_range	min	max	skew	numObs
Margarita Pizza	1733.00	1670.00	338.88	500.00	1190.00	2590.00	0.52	69
Cola	487.63	450.00	206.85	180.30	196.00	1160.00	1.42	61

Table 1: Summary statistics for the margarita and cola prices

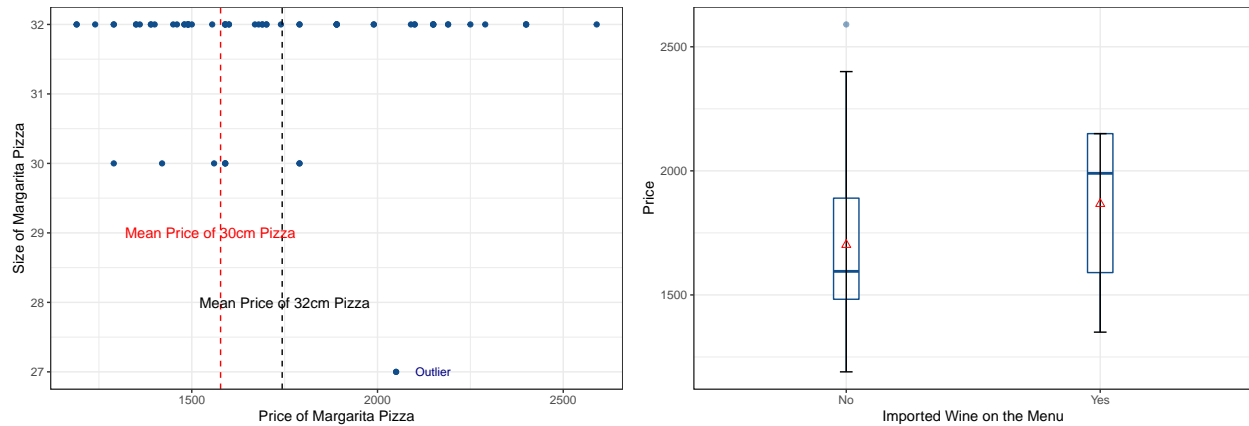
variable	mean	median	std	iq_range	min	max	skew	numObs
500ml Cola Actual	428.06	400.00	91.08	110.00	340.00	590.00	0.87	31
500ml Cola Estimated	549.19	530.30	268.75	250.38	196.00	1160.00	0.69	30

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

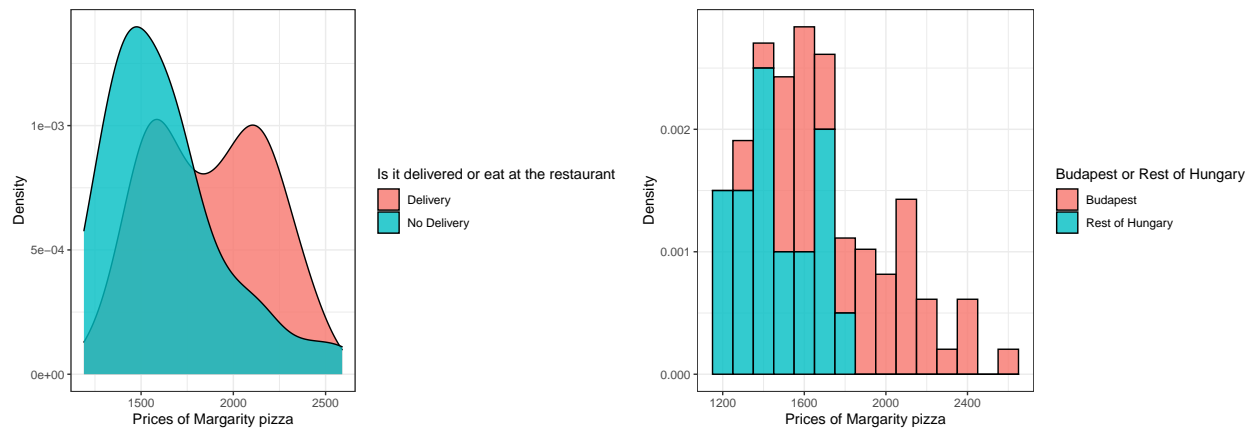
Distribution of Margarita and cola prices



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



Prices of Margarita pizza online vs offline



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 ($t = 3.0685$, $df = 63.384$, $p\text{-value} = 0.003163$). 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 ($t = 0.51646$, $df = 54.09$, $p\text{-value} = 0.6076$). There is strong evidence suggesting that online and offline beverage prices are similar on average.

Hypothesis testing if the mean price in Budapest or in Rest of Hungary is the same

```
##
## Welch Two Sample t-test
##
## data: margarita_price by is_budapest
## t = 6.0714, df = 60.651, p-value = 9.022e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 252.6088 500.7585
## sample estimates:
## mean in group Budapest mean in group Rest of Hungary
## 1842.184 1465.500
```

Here we easily reject the null hypothesis that the mean pizza price we observe in Budapest is the same as those we observe in the rest of Hungary ($t = 6.0714$, $df = 60.651$, $p\text{-value} = 9.022e-08$). The average pizza price in Budapest is most likely to be higher than the rest of Hungary.

```
##
## Welch Two Sample t-test
##
## data: cola_price by is_budapest
## t = 2.8381, df = 54.756, p-value = 0.006354
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 31.11565 180.69479
## sample estimates:
## mean in group Budapest mean in group Rest of Hungary
## 515.4128 409.5076
```

For cola prices, we again reject the null of same average prices in Budapest and the rest of Hungary but the evidence is not as strong this time ($t = 2.8381$, $df = 54.756$, $p\text{-value} = 0.006354$).