# Assignment

Team C

## Data Collection Process

The data collection process took place in a primarily online setting. Main sources included searches through Google and NetPincer websites. The restaurants were selected in random fashion. Restaurants outside of Budapest, were selected from cities with the population ranging from 60 thousand to 200 thousand, in an attempt to increase coverage. The number of observations was limited to three to avoid bias. The cola beverage was chosen as a variable in order to avoid missing values. Due to high popularity, the assumption was that, Cola is available almost anywhere in some shape or form. In addition, no distinction was made between nominal brands of cola (e.g. Pepsi-Cola vs. Coca-Cola). In order to manage the geographic restrictions and the limited opening hours, a binary variable was selected. The variable measures if the observed restaurant serves pizza only or has additional items on the menu (e.g. pasta, soup, non-Italian cuisine, etc.). In case of further research, the variable can be used as a conditioning variable, to explore correlation between it and other variables. The main cahllenges faced during the collection of the data are listed below

- Some of the restaurants don't have an online presence, which makes it harder to record observations and thus increases the chance of incomplete coverage.
- Most of the places don't have a uniform menu features, meaning some of the places don't indicate the pizza size while others don't have cola on the menu. This resulted in missing values for some observations.
- Due to the pandemic and geographic restrictions, the available variable pool was decreased. For example, there was no opportunity to visit all of the restaurants and actually try the pizza, thus qualitative variables such as taste, crust texture, etc. were not available for analysis.

## Descriptive Statistics for Margherita and Cola Prices in Hungary

Table 1: Summmary Statistics for Margherita (1) and Cola (2)

| Mean | Median | Std.dev. | Min | Max | IQ range | Skewness | Observations |
|---|---|---|---|---|---|---|---|
| 1651.6304 | 1690 | 365.4156 | 790 | 2300 | 572.5 | -0.1165506 | 46 |
| 492.8261 | 450 | 218.7582 | 170 | 1160 | 240.0 | 0.1957690 | 46 |

## Price Distributions

Figure 1.1 Cola Price Distribution

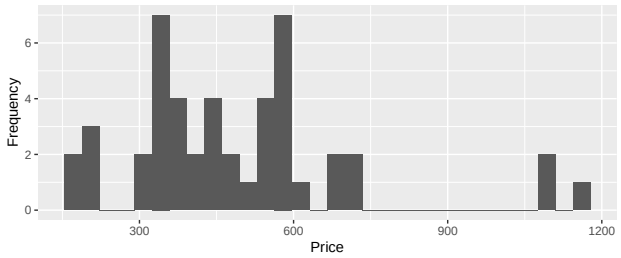
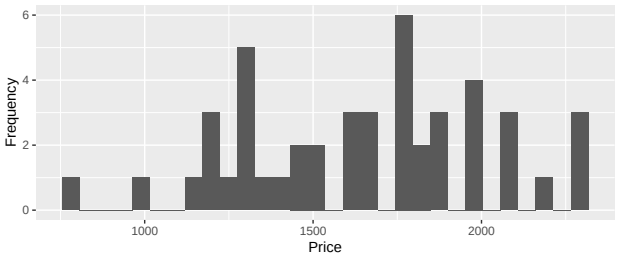
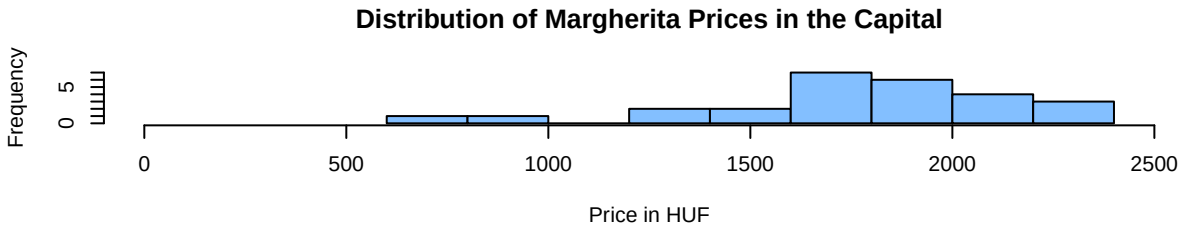
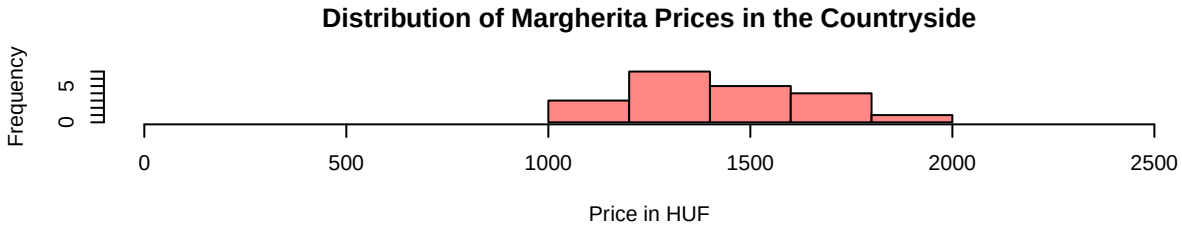



Figure 1.2 Margherita Pizza Price Distribution




## Figure 2.1 Price Distributions Conditioned on Region




## Statistical Test Results

Table 2: Test Results

| Region | mean_margherita_price | se_margherita_price | num_obs | t_stat |
|---|---|---|---|---|
| Capital | 1797.462 | 74.84814 | 26 | 24.01478 |
| Countryside | 1462.050 | 53.69448 | 20 | 27.22906 |

We could conclude based on the t-stats that the average price of margherita pizza in the capital and in the countryside is not the same. The t-stats are over 2, therefore we could reject that the prices are the same.

## Summary

The Capital distribution of prices covers more of the higher values and it is more spread out than the Countryside distribution of prices. The pizza prices in the Capital tend to be higher than in the Countryside.