Task 2

Team A

October 25, 2020

Data collection, difficulties, problems.

Entity Resolution

To avoid duplicates and maintain consistency, all three members of the team recorded their findings (observations) in one Google Sheet. However, at some point of the process, we figured out that they are some duplicates. We decided to leave one such observation and drop the other duplicate observations as they had the same value for all variables.

Validity

Reliability

Missing Values(pizza found, Google review not found)

Extreme Values

Problem 1

Deciding which size of pizza to choose was problematic. Pizza restaurants were offering the pizzas of size 28cm, 30cm, 32 cm, 48cm and even more. Thus, we initially decided to start analyzing all the restaurants and record the price for pizza of size 28cm, 30cm, and 32 cm as we met these sizes most frequently. We had a column for price of pizza with the size 28 cm and 32 cm. As the pizza of size 30cm and 32cm were observed most frequently, we decided to drop the column for price of pizza with the size 28 cm and leave the column for the price of pizza with the size 32 cm.

Problem 2

Deciding which method to use for measuring the distance was also problematic. At the beginning, two team members used "Measure distance" functionality of the Google Map. However, 3rd member of the group used "Show directions" functionality. Even though both methods are valid (shows the distance between two points), "Show directions" functionality shows the fastest and shortest path. Therefore, it shows different distances when checked different times. It may vary based on the traffic jam, maintenance work in roads or due to other reasons. However, "Measure distance" functionality works as a "ruler on the map" manner and shows the same distance between two points no matter when the distance was checked. To maintain the comparability, 3rd member of the group also decided to change the values for variable "distance" by using "Measure distance" functionality.

Descriptive Statistics

Discuss the data collection, difficulties, problems. How you picked the beverage. How did you decide on store features to record and how did you code it? (1-2 paragraphs) [3p]

Obs	n	mean	median	min	max	sd	skew
Pizza	44	1801.18	1790	1240	2690	291.30	0.55
Beverage	44	457.41	450	300	790	101.78	1.00

Descriptive Graphs

Show two descriptive graphs of price distributions of your products in the whole data. [2p]