

Bogota - Data Analysis 1 Team Assignment

Bogota team members: Natalia Iriarte, Péter Kaiser, Rauhan Nazir, Sára Vargha, Xibei Chen

Data Collection



Our Tasks

- Pick two districts, one in “inner city”, the other in “outer city”;
- Choose another product in addition to Coca Cola 0.5l plastic bottle;
- Pick some features about the store as variables;
- Visit 10-15 stores in each district to collect data about the price of the two products and other features about the store.

Our Experience

- **How we selected the locations**
 1. We picked districts different from the ones already picked by others and that were closer to where we live;
 2. We tried to cover both the inner and outer part within each district to mitigate the selection bias;
 3. We tried to select a wide variety of shops.
- **How we chose another product and other features**
 1. We thought Nutella is quite popular and 400g is the regular volume;
 2. We thought about collecting data about the size of the store, but we found it hard to measure in person or find the information online, so we dropped these features as variables;
 3. Instead, we picked some other features to help us understand what nationality the store is, how busy the store is, and if the products are among the products that the store intends to sell the most.
- **Problems we encountered**
 1. There are stores where we could not find the exact product that we want, for example they only sell zero sugar Coca Cola or Nutella 700g.
 2. We only realized some stores on the border of the 9th district actually belong to the 8th, when our local teammates checked the postcode in the end. So we changed the district for those observations to 8, but we kept the data in the table. And we still reached the min. number of stores required for the 9th district.
- **Choices we had to make during the process**
 1. When we created the data record table, we tried to use the tidy data approach with long format to make sure that each row is only for one observation, e.g. product as variables instead of product 1 and product 2. However, in this way we had to sacrifice by typing in the information about the store twice for each product.
 2. If the product was out of stock, but there is price tag, we still recorded the data;
 3. We did not use our individual criteria for the store type, instead we used the type shown on Google Maps;
 4. We counted two people going shopping together in queue as one;
 5. We typed in the position of the product located around the height of our line of sight as middle instead of top even though it was not exactly in the middle;
 6. We saw red label instead of white label in SPAR as discounted, but there was no "Akció" on the label, and translation was not available there, so we were not sure if it was discounted;
 7. We were not sure about the exact nationality of some Asian stores, so we just labeled them Asian.