

2020DataFest – Maskman

Purchasing Behavior During COVID-19: Amazon Price/Google Trend in-depth Analysis

RShiny App: <https://whuai.shinyapps.io/2020DataFest-Maskman/>

Git Repository: <https://github.com/JunpengJiang/2020-DataFest-Maskman>

Members:

Chuqi Bian – cb102497@gmail.com

Minghan Li – kevinli980714@gmail.com

Zuxuan Huai – wendyhuai@g.ucla.edu

Baiting Zhu – baitingzbt@g.ucla.edu

Junpeng Jiang – junpengmarshall@gmail.com

Summary

As many health care products such as masks and disinfectant products are taking over the news headlines, our team has decided to place our primary focus on the shifts of the demand and prices of the consumer products. Specifically, we divided the products into four categories and selected 5 products to represent each category. With extensive data cleaning and visualizations, we have mainly examined the products' price trend, price vs. demand correlation and created an out-of-stock index across different categories.

Data Source

The dataset used to implement analysis mainly comes from Amazon. With the help of Keepa API, we web-scraped the historical price data starting from 01/01/2019. Since the raw data on Amazon is collected on a minute level, we averaged each products' prices on a daily basis. We also filtered the data such that all downloaded data are not missing since the pandemic started. Another source we used was Google Trend, we downloaded the search frequency of all the keywords we were interested in and merged them into one dataset.

Methods & Results

Before we started, we standardized the price data such that the price of each product on 01/01/2020 is 1, 1.1 then reflects there is a 10% increase in the prices. To match the range of the adjusted price data, we also rescaled google trend data. By generating heatmaps, animated line charts and scatter plot, we explored the price trend of the products. In addition, we combined all products in each category and matched their google search trend with the price on a daily basis.

From there, we constructed a scatter plot between rescaled google search trend and standardized price. In our model, we use correlation coefficient between the two variables to represent price elasticity to search rate. Besides, we constructed heatmaps and animated scatter plots to examine the google search trend and price indexes between each pair of the products. One interesting thing we found, for example, was that there is a positive correlation between Google search trend of Toilet Paper and Hand Sanitizer.

The last thing we did is to produce an out-of-stock index. By identifying each product's out-of-stock status across each day, we calculated the out-of-stock percentage and generated an animated boxplot that tracks the stock status of the products since the pandemic started. We found that consumers have shifted their buying interests from health care products (masks/hand sanitizer) to non-perishable food (canned soups), and finally to Yoga Mats, Dumbbells, and Hair Clippers for daily life and exercise nowadays.