The project

For this project we examine data of S&P 500 and Trends of interest in several words. We are particularly interested in the correlation between stock prices and interest in particular goods during peak of COVID-19(between March 2020 and September 2020). This blog is structured as follows. First, we go into the S&P 500 stock data. We structure it into dimension we want and add some additional variables that will come in handy at a later point. Next, we show the data from PyTrends. Thereafter a number of visualisations are show to get more insight into the analyzis.We finish with the showing correlation between stock value and interest in products.

Data

In order to get more insight into market values, there has been made use of historical data of S&P 500 stock data. The data was retrieved from Yahoo and can be found via [this](https://finance.yahoo.com/quote/%5EGSPC/history/?guccounter=1) link. The data set hold information of more than 100 days of historical data. The variables that are especially interesting in our research are Date, High, Low, and Volume. Below there is an overview of the first 5 days in the data set to get a feeling what we are talking about.

Table

Description automatically generated

Figure 1. Overview of the variables in the dataset

Data Cleaning

The data set looks clean at first sight but there are some adjustments we would like to make. First of all, we would like to make our analysis on weekly basis rather than daily. Second, we find out that the type of the Date variable is an object and we convert it to an integer. Next, we add an extra variable to our dataframe, namely a date\_average. This additional variable will be helpful in the clarity later in the analyzis.The date\_average was obtained by finding mean of High and Low columns. Furthermore, there is one wine with a very high price as shown in the figure below.

Table

Description automatically generated

Figure 2. Overview of the variables in the data set after changes

PyTrends

After our research, we understood that because of COVID-19, people tended to stay home the number of people ordering food and beauty items has increased. Thus, we selected Cosmetics and Delivery. As the number of orders has increased, restaurants need more Tupperware. Finally, provided that many individuals were forced to stay home, the number of depressed and lonely people decreased. A study in the late 90th proved that people with those symptoms tend to consume more sweets; hence we selected Candies.

**Data**

In order to get more insight into people’s interests, there has been made use PyTrends. The data was retrieved from google searches of each good by people in the Netherlands in aforementioned period. Figure 3 illustrates an overview of the first 5 days in the data set to get a feeling what we are talking about. Table

Description automatically generated

Figure 3. Overview of variables in the PyTrends data set

Date Cleaning

The method of daily summation wad used.In similar fashion to Yahoo data, PyTrends data was transformed into weekly data set, as shown in figure 4.Thereafter, both data sets were integrated into one. This was done in order to carry out analyzis discussed later.The integrated table is shown in Figure 5.Furthermore, interest per good was analyzed and presented in Figure 6

Table

Description automatically generated

Figure 4. The data set after transformation.

Table, Excel

Description automatically generated

Figure 5. Integrated data set

Chart, bar chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Figure 6. Interest per product

Data visualisation

To get more insight in the data we explore which Dutch provinces have the most interest in our goods .Below we see an overview of this. In addition it shows that although some countries may not be the most popular wine countries, it does not seem that they are rated lower. As illustrated in Figure 7, Flevoland was leading in Cosmetics, Friesland in Snoep, North Holland in Delivery, Zeeland in Tupperware

Table

Description automatically generated

Figure 7. The interest in each province to our products

For further analysis, we view correlation between date\_average and interest in products selected. In the scatter plot below, we can see an overview correlation for each good. Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Figure 8. Correlation per product with date\_average

Results

In conclusion, it can be stated that pandemic caused an increased interest in the number of products in the Netherlands and across the globe. Even though stable increase in the number of searches as prices of stock decrease can be observed, there were enough cases that show that price changes were not the only reasons for the given trend. In order to increase the accuracy of given research, increasing the time frame to a much larger period and comparing the data with pre-corona records and observing the difference can be applied. Meanwhile, the project brought some impressive results; more work needs to be done.