**Introduction**

Avocado, a bright green fruit with a large pit and dark leathery skin, has become a fashion among millennials in America. Avocados featured in its unique flavor, rich texture, and high percentage of healthy fat, which is an ideal food for the healthy diet. However, the average price for avocados is not cheap, which made those millennials unable to achieve their “Avocado Freedom” dream. In addition, as an applicant in American master and a fan of avocado, it is of significant attraction for me to conduct the explorative data analysis of avocados, hoping to find out some hidden patterns, like seasonal patterns or cyclical patterns. Furthermore, related factors were also evaluated to indicate the influence on the avocados’ price, which could help these millennials to find the cheaper avocados in a scientific way.

**Methods**

**Collection and preprocessing of the avocado dataset**

The dataset was downloaded from Kaggle (*Avocado Prices (2020) | Kaggle*, 2020), which contains the data from 4 January 2015 to 17 May 2020. Specifically, the dataset directly scanned from retailers’ cash registers based on actual retail sales of Hass avocados. Avocados in this dataset could be seperated into to two categories, either conventional or organic; the corresponding average price and volume of avocados sold differs on types and geographical locations. In this experiment, the dataset does not contain any missing dataset and I mainly focused on six parameters including date, average price, total volume, year, type and geography, aiming to find the best strategy in buying avocados.

**Explorative data analysis**

**Comparison between different avocado types**

In this dataset, there are two types of avocados, conventional and organic. I would like to compare the difference in the average price and the volume of these avocados sold. Hence, I plotted a density plot to demonstrate the distribution of average price among two types of avocados. Figure 1 indicates that the average price for organic avocados distributed more evenly and the average price could be more expensive than conventional type. Conventional avocados’ average price centered around 1 dollar while organic centered around 1.5.

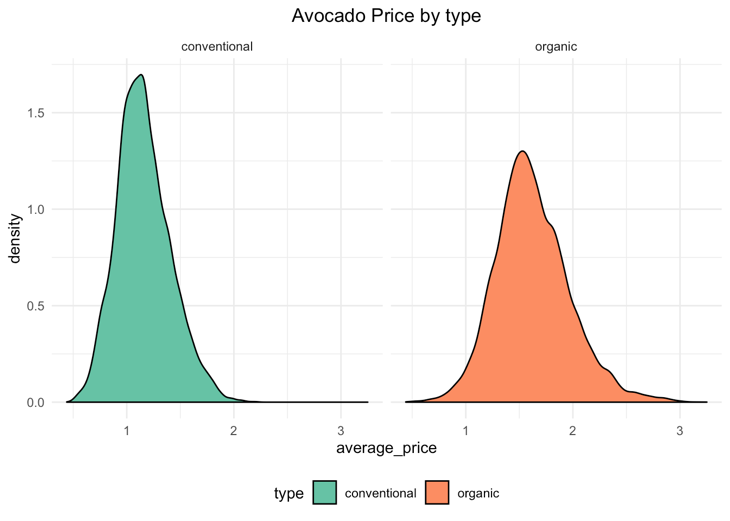


Figure 1: Density plot of avocado price by type

Table 1 was created for better illustration of the correlation among these two types. As can be seen from the table, conventional avocado could account for approximately 97% in the markets and the average price would be 28% lower than organic type. Furthermore, the trends of average price among the two types could be visualized in Figure 2. It is clear that the average price of conventional avocados is continuously lower than organic one since green stands for organic type while red stands for conventional type.

**Table 1: Average volume and price comparison among two types**

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Average volume | Average price | Volume percent |
| Conventional | 1,818,206 (1.8 M) | 1.16 $ | 96.8% |
| Organic | 60,127 (0.06 M) | 1.62 $ | 3.2% |

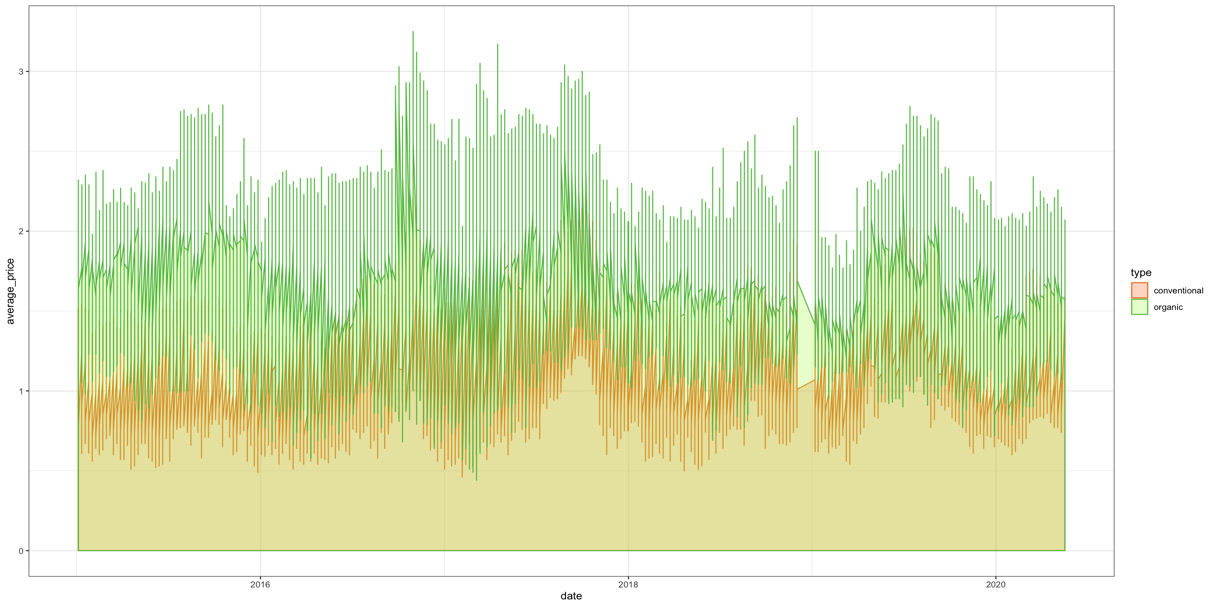


Figure 2: Avocado price with the date in two types of avocados

**Correlation between avocados’ price and volume**

Additionally, it is a common view that the volume and the price could be correlated since the price would increase as the customers’ demand increase. I would like to see whether two types of avocados follow these rules. In this case, Figure 3 illustrates the average price and volume among organic and conventional avocados from 2015 to 2020 in each month. As can be seen from the figure, the overall trend for conventional avocados volume increases steadily while the organic avocados volume starts to decrease after 2019. Furthermore, we could notice that the volume undergoes some periodic change. In other words, the avocado price and volume may have some seasonal patterns.

图形用户界面, 图表

描述已自动生成

Figure 3: Average price and volumes trends of conventional and organic avocados from 2015 to 2020

**Seasonal patterns of price and volume analysis per year**

In order to elucidate the seasonal patterns of avocados, I categorized the samples based on seasons: Spring (Mar-May), Summer (Jun-Aug), Autumn (Sep-Nov), Winter (Dec-Feb), and calculated the average price and mean volume respectively from 2015 to 2020 (Figure 4). From the perspective of output volume, Spring would be the highest in either conventional or organic avocados sold, indicating American people really love to eat avocado in Spring. From the perspective of average price, however, the patterns are not consistent, either winters or spring could be the season with lowest price, which stimulates me to explore the data monthly.

表格

描述已自动生成

Figure 4: Average price and volumes sold of two types avocados by season

**Monthly patterns analysis of price and volumes per year**

图表, 折线图

描述已自动生成

表格

描述已自动生成**Geographical analysis of the cheapest avocados**

**Future average price of avocados trend prediction**

图表

描述已自动生成

图表, 折线图

描述已自动生成

Limitations & improvements

图表, 折线图

描述已自动生成

图表

描述已自动生成