

XI'AN JIAOTONG-LIVERPOOL UNIVERSITY

西 交 利 物 浦 大 学

INT303 FINAL PROJECT REPORT COVER

Name	Yuxuan	Wu
ID Number	<i>1716309</i>	
Programme	<i>Big Data Analysis</i>	
Assignment Title	<i>Final project report No 2 Final Report</i>	
Submission Deadline	<i>Time of submission: 12.27</i>	
Group	<i>TA name: Linlin Du</i>	

I **certify** that:

- I have read and understood the University's definitions of COLLUSION and PLAGIARISM (available in the Student Handbook of Xi'an Jiaotong-Liverpool University).

With reference to these definitions, I **certify** that:

- I have not colluded with any other student in the preparation and production of this work;
- this document has been written solely by me and in my own words except where I have clearly indicated and acknowledged that I have quoted or used figures from published or unpublished sources (including the web);
- where appropriate, I have provided an honest statement of the contributions made to my work by other people including technical and other support staff.

I understand that unauthorised collusion and the incorporation of material from other works without acknowledgement (plagiarism) are serious disciplinary offences.

Signature吴宇轩.....

Date2020.12.20.....

Your marks to your team mates:

This is an individual project and all the works were finished by my own

Deadline: **27th of December 2020, 23:59(Beijing Time)**

For Academic Office use:	Date Received	Days Late	Penalty

Introduction

Avocado, bright green fruit with a large pit and dark leathery skin, has become a fashion among millennials in America. Avocados are featured in their unique flavor, rich texture, and the 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 degree and a fan of avocados, it is of significant interest for me to conduct the explorative data analysis of avocados, hoping to figure out some hidden patterns, like seasonal patterns or cyclical patterns. Furthermore, related factors were also evaluated to suggest the influence on the avocados’ price, which could help these millennials to find the cheaper avocados in a scientific way.

Methods and Results

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 is directly scanned from retailers’ cash registers based on actual retail sales of Hass avocados. Avocados in this dataset could be separated into two categories, either conventional or organic; the corresponding average price and volume of avocados sold differ on types and geographical locations. In this experiment, the dataset does not contain any missing value 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 both the average price and the volume of these avocados sold. Hence, I firstly 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 is distributed more evenly and the average price could be more expensive than the conventional type. Conventional avocados’ average price centered around 1 dollar while organic centered around 1.5 dollars.

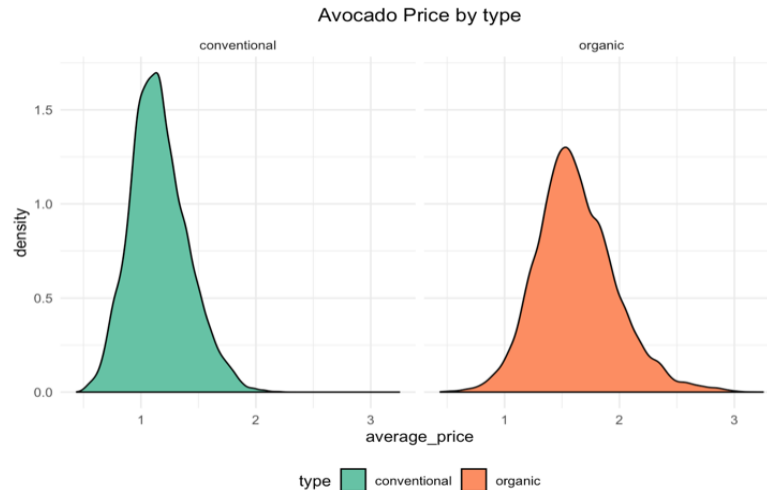


Figure 1: Density plot of avocado price by type

Table 1 was created for a better illustration of the correlation between 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 the 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 ones 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%

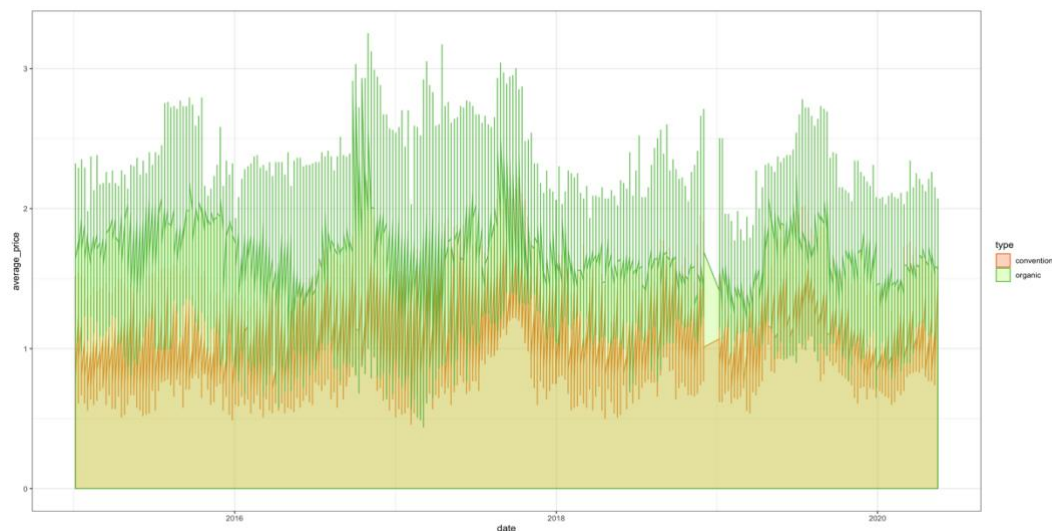


Figure 2: Avocado price with the date in two types of

Correlation between avocados' price and volume

Additionally, it is a common view that the volume and the price could be negatively correlated since the customers' demand would increase as the avocado price decrease. I would like to see whether the 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.

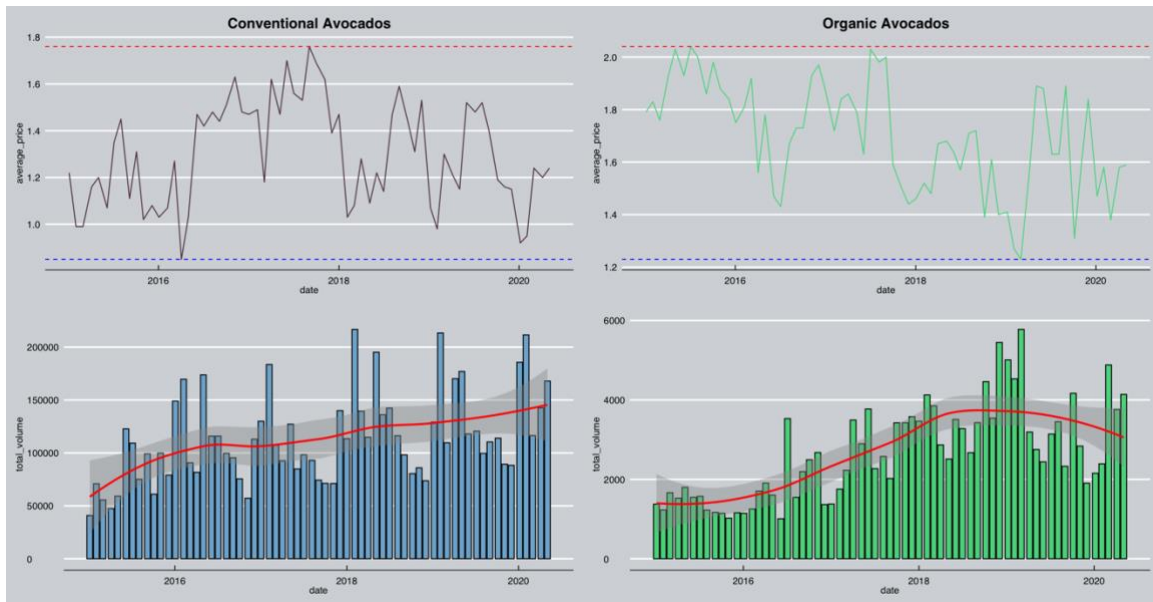


Figure 3: Average price and volumes trends of conventional and organic avocados from 2015 to 2020

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 volume respectively from 2015 to 2020 (Figure 4). From the perspective of avocado volume sold, Spring would be the highest in either conventional or organic avocados sold, indicating that American people love to buy avocados in Spring. From the perspective of average price, however, the patterns are not consistent, either winter or spring could be the season with the lowest price, which stimulates me to explore the data monthly.

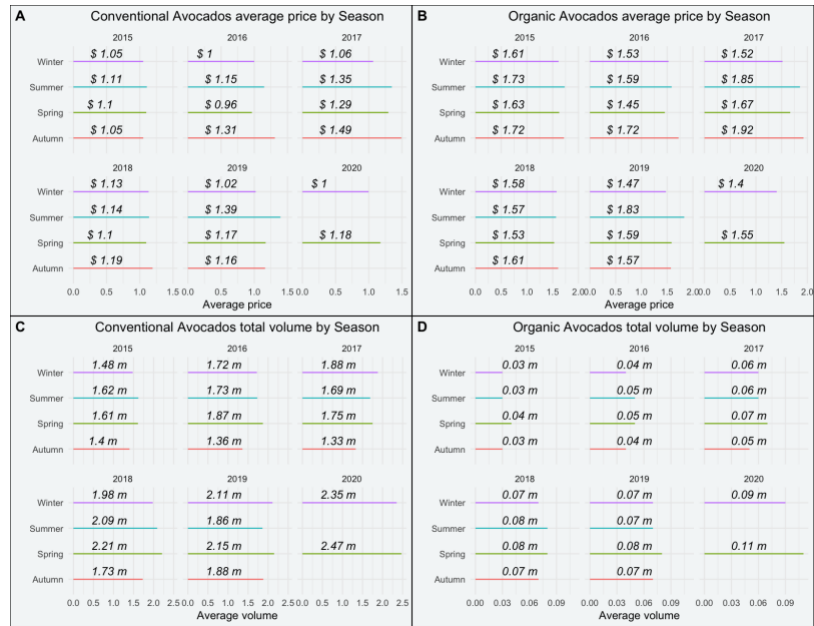


Figure 4: Average price and volumes sold of two types avocados by season

Monthly patterns analysis of price and volumes per year

Figure 5 compares the average price and volume per month from 2015 to 2020. As can be seen from the figure that the average price would reach the peak during September or October and reach the bottom in next year's February. Regarding the average volume sold, American people prefer to buy avocados in next year's February or May and the willingness achieved lowest in November.

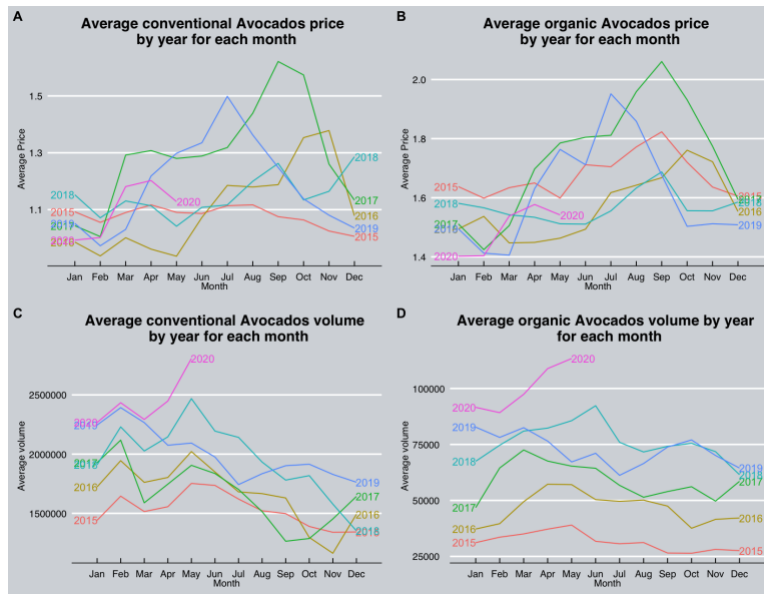


Figure 5: Average price and volumes sold of two types avocados per month

The geographical analysis of the cheapest avocados

To find the cities or regions in the United States with the cheapest avocados, I thus returned the top six places and visualized them with the flipped bar chart (Figure 6). Phoenix/Tucson, with the average price of 0.78 dollars, would be the cheapest region in the conventional avocados. Meanwhile, organic avocado could find its lowest average (1.34\$) price in Dallas/Ft. Worth.

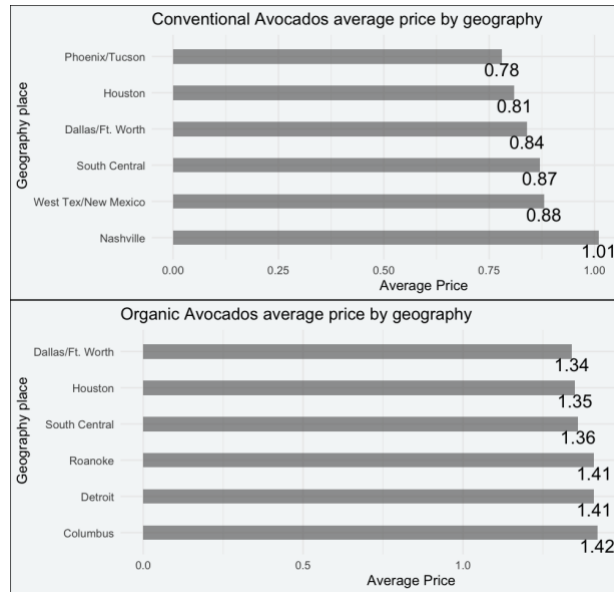


Figure 6: Six cities or regions with the lowest average price in the United States

The future average price of avocados trend prediction

In addition to the explorative data analysis (EDA), it is of great significance to predict the future price. Therefore, I utilized the R package “prophet” to conduct the predictive data analysis. (Taylor and Letham, 2018). Figure 7 is the combination of both the history avocado price and the predicted price indicated by the dashed plot. The predicted price of avocado would continue to grow but the increased rate would slow. Figure 8 includes the trend of avocado average price from daily, weekly, monthly and yearly. The positive trend value indicates the price increase while the negative value indicates the decrease. From subplot A, the average price of avocados continues to grow and reaches a peak in 2017, with a speed higher than 1.3. Regarding future price prediction, the range is about [1.20-1.34]. Furthermore, the subplot B reconfirms the previous research that starting from May, the average price continues to grow until October. Subplot C suggests the average price at weekends was relatively higher than on weekdays. The subplot D indicates the time per day, indicating the price change occurred at midnight.

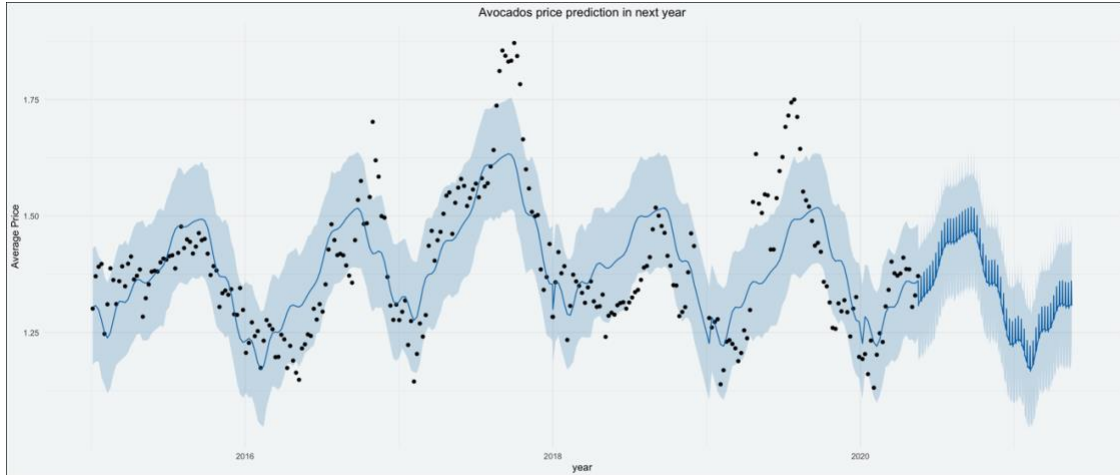


Figure 7: Average avocado price trend in next year

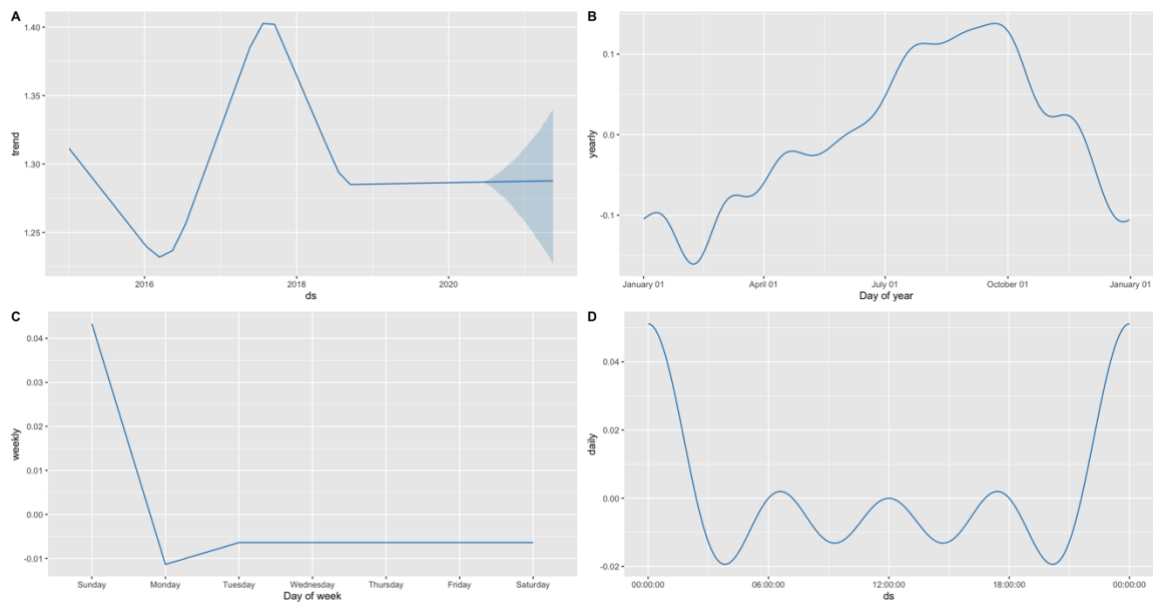


Figure 8: Average avocado price trend per day/month/week/year

Discussion and Conclusion

The explorative-based and prediction-based data analysis indicates the avocado market situation in the United States. There are two major types of avocados, either organic or conventional. Conventional avocados with an average 1.16\$ (28% lower than organic type) account for almost 97% of the market, suggesting most American people are sensitive to price, they would choose conventional avocado compared with organic type under the same condition. Moreover, the average price and the volume of avocado sold indicate some

negative correlation, which obeys the common view that customers' demand would decrease as the price increase. Besides, avocados' price and volume indicate some seasonal patterns which are normally reflected in the fruits. The highest price of the whole year would during September or October while the highest volume sold was in February or May. The research indicated that avocado would be ripe in approximately September per year (*Is that avocado ready for harvest? Here's what to look for in your garden | Home/Garden | nola.com*, 2017). With the necessary conservation and transportation, avocados would appear on the market during October. At that time, the previous year's remaining avocados would reach the minimum and this situation would trigger the price increase. As a result, with the consistent supply of avocados, the price starts to decrease until February. Simultaneously, since the price reached the button, the avocado volume sold would reach the peak. In addition, the top six regions with the cheapest avocado price were returned from historical data. Except for Nashville, the top five cheapest regions of conventional avocado are all in the middle west of America, near to Mexico, where avocados originated. Based on the research, these places are the centers of avocado planting which therefore achieving such a low price compared with others (*What Types of Fruit Trees Grow in Florida? | Hunker*, no date). Apart from the explorative data analysis, the future price trend on avocados was predicted to conduct the predictive data analysis. The avocado price in the following year would continue to grow, but with a lower increasing rate, ranging from [1.20-1.34]. Furthermore, with the strike of the Covid 19 pandemic, people's consumption of avocados would decrease concomitantly. In this case, the following year's avocado price increasing rate would be about 1.20.

It is unavoidable that the research could contain some limitations and need to be improved in future analysis, specifically, the prediction model. In this experiment, I utilized the existing R package developed by Facebook to conduct the avocado price prediction, however, the internal mechanisms of this package are difficult to elucidate because of its intrinsic complexity. In other words, the price prediction model was a "black box problem". Therefore, in the future, I would build the time-series prediction model individually and emphasize the interpretability of the model, for instance, the Smoothing Moving Average (SMA), Seasonal Naive Method and Drift Method.

Reference

Avocado Prices (2020) | Kaggle (no date). Available at:
<https://www.kaggle.com/timmate/avocado-prices-2020> (Accessed: 4 December 2020).

Is that avocado ready for harvest? Here's what to look for in your garden | Home/Garden | nola.com (no date). Available at:
https://www.nola.com/entertainment_life/home_garden/article_c44ca588-7f9c-5402-8338-a12f9c531fcd.html (Accessed: 16 December 2020).

Taylor, S. J. and Letham, B. (2018) 'Forecasting at Scale', *American Statistician*, 72(1), pp. 37–45. doi: 10.1080/00031305.2017.1380080.

What Types of Fruit Trees Grow in Florida? / *Hunker* (no date). Available at: <https://www.hunker.com/12312954/what-types-of-fruit-trees-grow-in-florida> (Accessed: 16 December 2020).

Appendix

The whole script could be accessed from following links:

Github: <https://github.com/yuxuanwu17/INT303-Avocado-prediction>

Kaggle: <https://www.kaggle.com/yuxuanwu17/eda-of-avocado-price-from-2015-to-2020>