# INTRODUCTION

Avocados are one of the most important fruits produced in the world. They're the go-to ingredient for guacamole dips at parties. And they're also turning up in everything from salads and wraps to smoothies and even brownies. The United States is the second largest producer of avocados after Mexico. About 90% of avocado production in the United States takes place in California by more than 5,000 growers. Avocados come in a variety of shapes, sizes, and colours. They can be shaped like a ball, a teardrop, or a football.

Avocados are native to Mexico and Central America but are now grown in many different regions across the globe. They take a significant amount of resources to cultivate, since an avocado tree takes between three and five years before it begins bearing fruit. Over 231 thousand hectares of avocado trees were planted in Mexico as of 2017/18, up from 168.11 thousand hectares in 2013/14.

Avocados are a nutrient-dense food that provide relatively many health benefits that include abundant nutrients (vitamins C, B5, B6, E, and K), potassium, folate, protein, and healthy fats. Seven of the nine grams of carbohydrates in a 100-gram serving are fibre, qualifying avocados as a low-carb food. Avocado has become a favourite ingredient in many dishes(guacamole dip, salads, sandwiches) and is now on the menu of several restaurants across the U.S. The Avocado toast, a humble breakfast that was earlier eaten mainly in California and Australia, now is a trendy menu item in many restaurants around the world. According to the USDA, U.S. demand for avocados has increased steadily over the past two decades. Per capita consumption of avocados has tripled since 2001 to 8 pounds per person in 2018. While the average consumption of avocados in the United States amounted to 2.45 billion pounds per week in 2018.

# DATA SET DESCRIPTION

The table represents weekly 2018 retail scan data for National retail volume (units) and price. Retail scan data comes directly from retailers' cash registers based on actual retail sales of Hass avocados.

Starting in 2013, the table reflects an expanded, multi-outlet retail data set. Multi-outlet reporting includes an aggregation of the following channels: grocery, mass, club, drug, dollar and military. The Average Price (of avocados) in the table reflects a per unit (per avocado) cost, even when multiple units (avocados) are sold in bags. The Product Lookup codes (PLU's) in the table are only for Hass avocados. Other varieties of avocados (e.g. green skins) are not included in this table.

Data Fields -

- Date - The date of the observation
- Average Price - the average price of a single avocado
- type - conventional or organic
- year - the year
- Region - the city or region of the observation
- Total Volume - Total number of avocados sold
- 4046 - Total number of avocados with PLU 4046 sold
- 4225 - Total number of avocados with PLU 4225 sold
- 4770 - Total number of avocados with PLU 4770 sold

# ALGORITHM

## Time series analysis -

Time series is a series of data points indexed (or listed or graphed) in time order. Therefore, the data is organized by relatively deterministic timestamps, and may, compared to random sample data, contain additional information that we can extract. Time series are one of the most common data types encountered in daily life. Financial prices, weather, home energy usage, and even weight are all examples of data that can be collected at regular intervals. Almost every data scientist will encounter time series in their daily work and learning how to model them is an important skill in the data science toolbox. One powerful yet simple method for analysing and predicting periodic data is the additive model. An additive model can show us both patterns/trends and make predictions based on these observations.

Time series is a sequence of observations recorded at regular time intervals. Depending on the frequency of observations, a time series may typically be hourly, daily, weekly, monthly, quarterly and annual. Sometimes, you might have seconds and minute-wise time series as well, like, number of clicks and user visits every minute etc. Besides, time series forecasting has enormous commercial significance because stuff that is important to a business-like demand and sales, number of visitors to a website, stock price etc are essentially time series data. Time series analysis involves understanding various aspects about the inherent nature of the series so that you are better informed to create meaningful and accurate forecasts.

The plot below shows the time series analysis on our avocado data set -
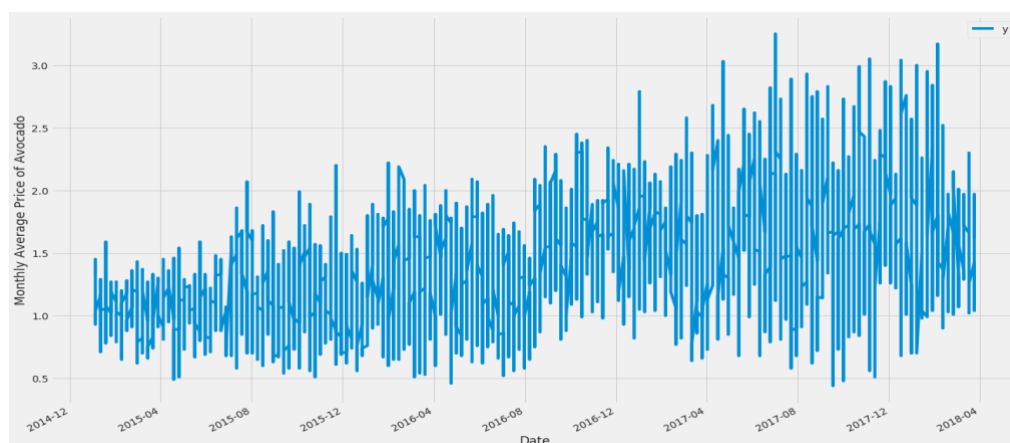
**Patterns in a time series -**

Any time series may be split into the following components: Base Level + Trend + Seasonality + Error

A trend is observed when there is an increasing or decreasing slope observed in the time series. Whereas seasonality is observed when there is a distinct repeated pattern observed between regular intervals due to seasonal factors. It could be because of the month of the year, the day of the month, weekdays or even time of the day. However, It is not mandatory that all time series must have a trend and/or seasonality. A time series may not have a distinct trend but have a seasonality. The opposite can also be true.

So, a time series may be imagined as a combination of the trend, seasonality and the error terms.

Additive and multiplicative time series

Depending on the nature of the trend and seasonality, a time series can be modelled as an additive or multiplicative, wherein, each observation in the series can be expressed as either a sum or a product of the components:
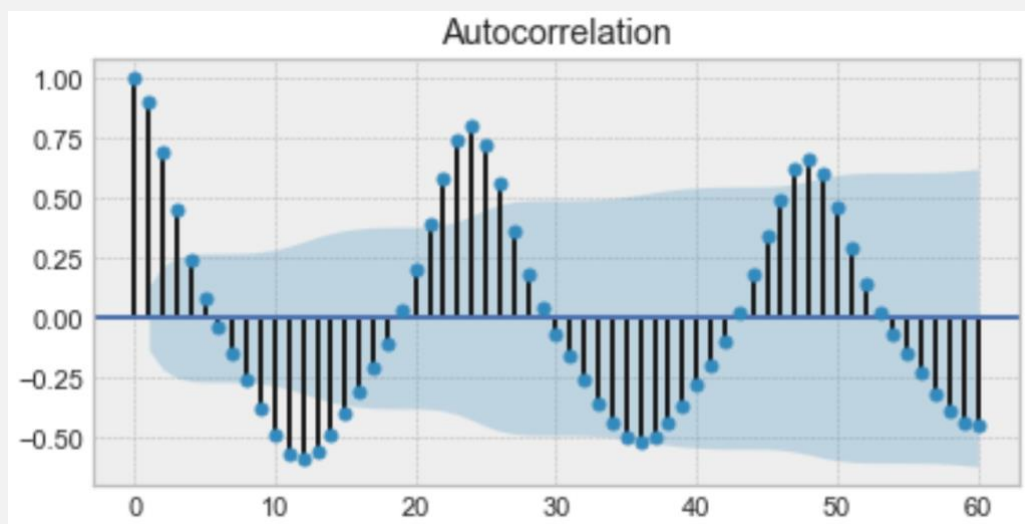
**Additive time series:**

Value = Base Level + Trend + Seasonality + Error

**Multiplicative Time Series:**

Value = Base Level x Trend x Seasonality x Error

**Autocorrelation :** Informally, autocorrelation is the similarity between observations as a function of the time lag between them.

**Seasonality :** Seasonality refers to periodic fluctuations. For example, electricity consumption is high during the day and low during night, or online sales increase during Christmas before slowing down again.
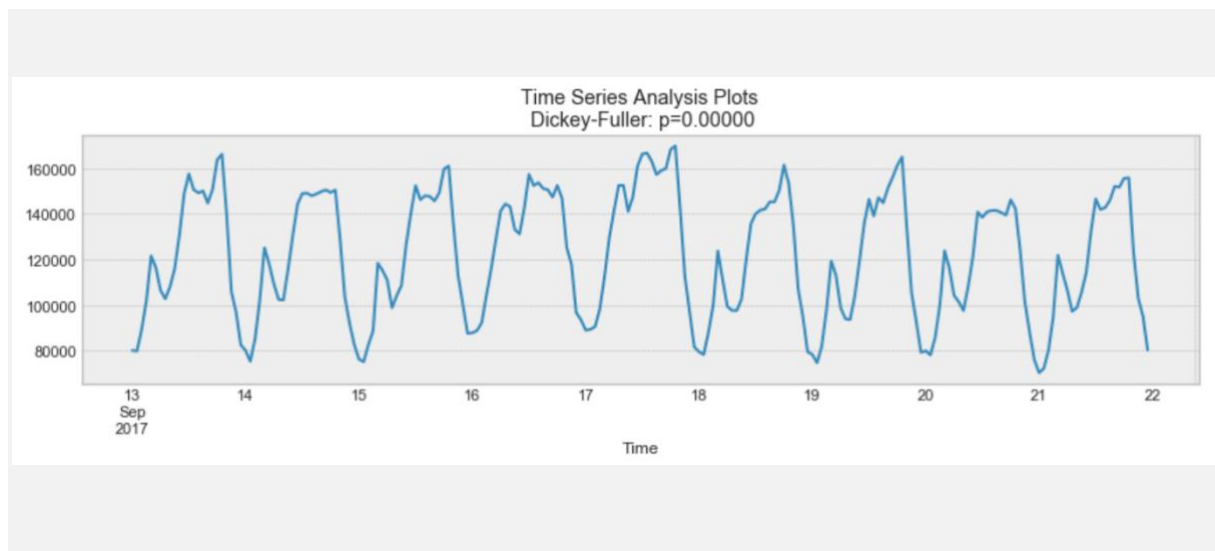


**Stationarity :** Stationarity is an important characteristic of time series. A time series is said to be stationary if its statistical properties do not change over time. In other words, it has constant mean and variance, and covariance is independent of time.
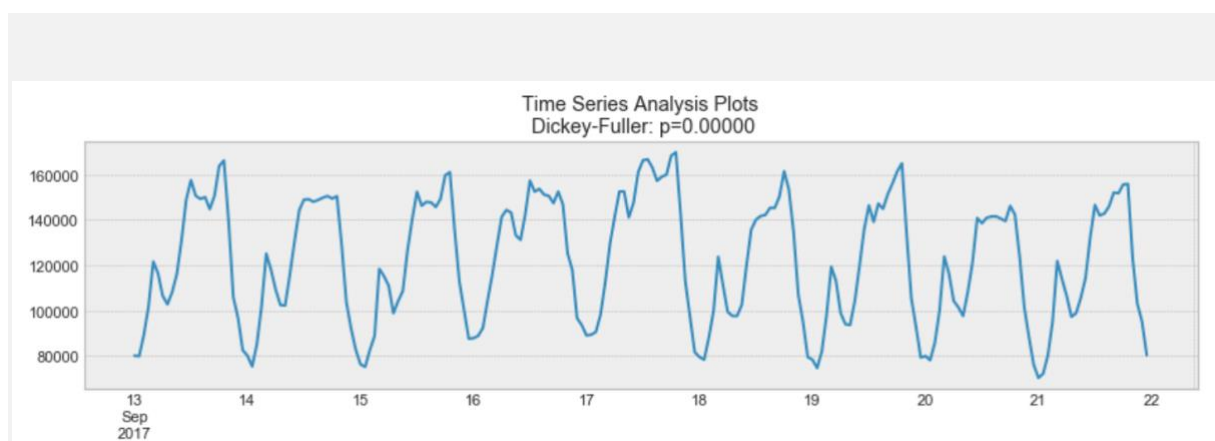
# Call prophet model -

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is a Facebook's open source time series prediction. Prophet decomposes time series into trend, seasonality and holiday. It has intuitive hyper parameters which are easy to tune.

Prophet time series = Trend + Seasonality + Holiday + error

Trend models the non-periodic changes in the value of the time series. Seasonality is the periodic changes like daily, weekly, or yearly seasonality. Holiday are the effects which occur on irregular schedules over a day or a period of days. Error terms are what is not explained by the model. The plots below show the prophet model applied on the avocado data set -
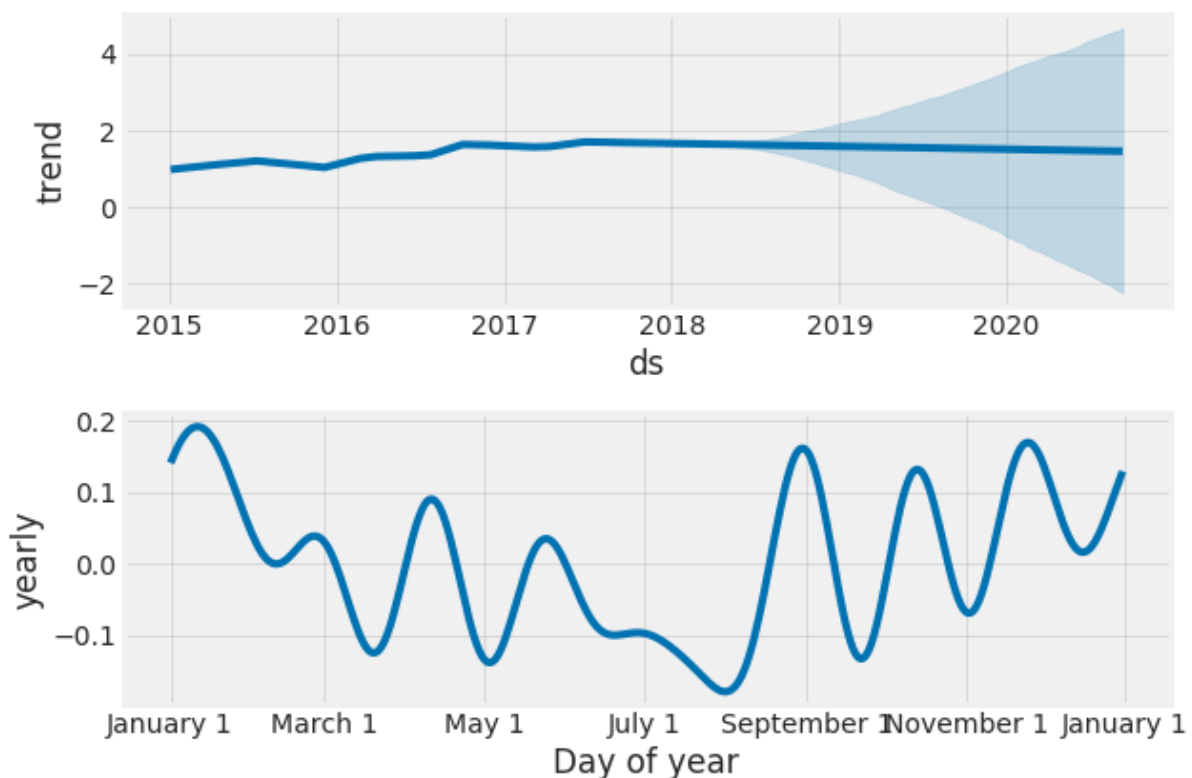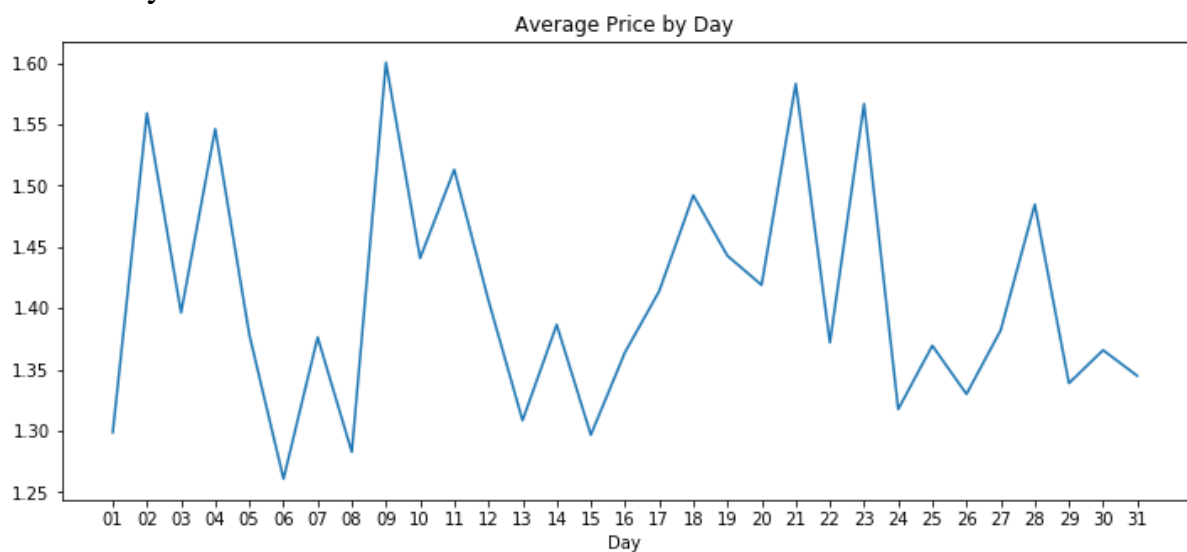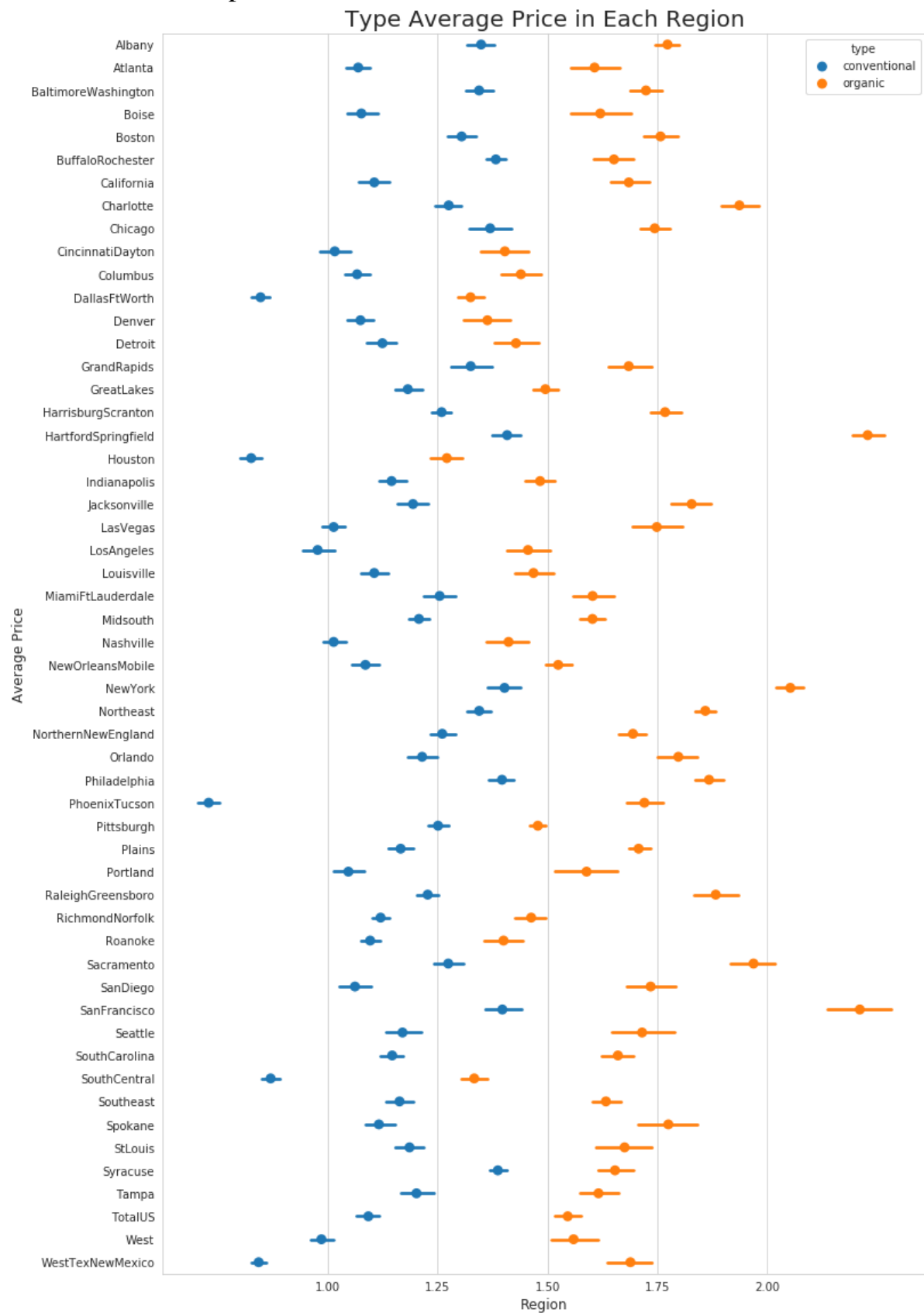
# PLOTS
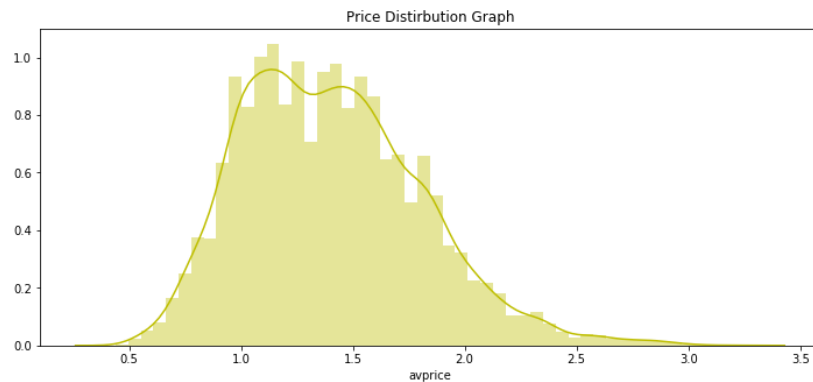


The plot below highlights that seasonal changes can affect prices.
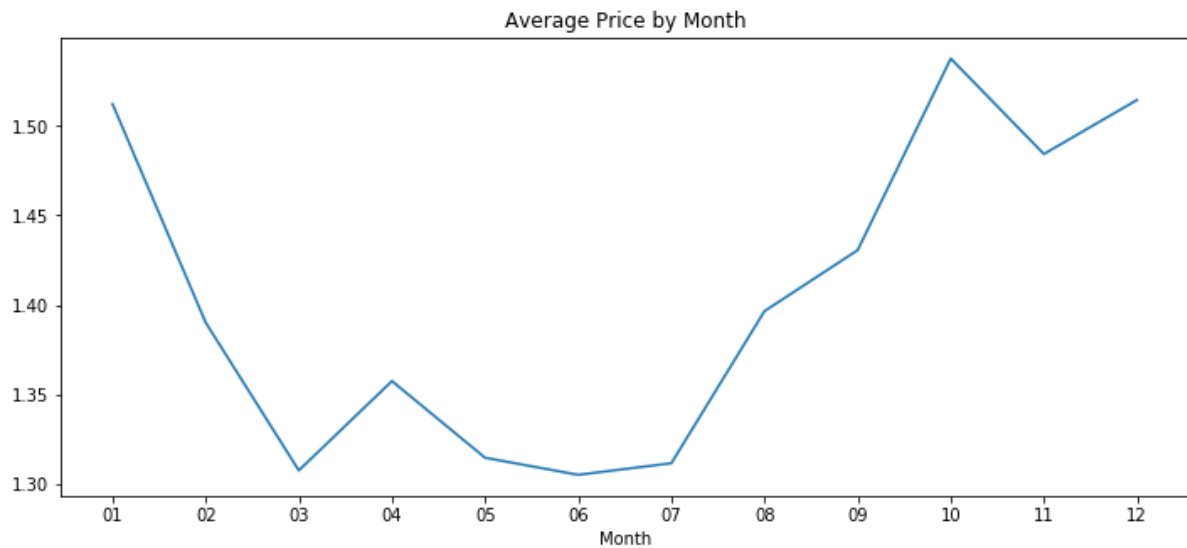


The plot below is significant because it shows us the discount prices on certain days of the week.
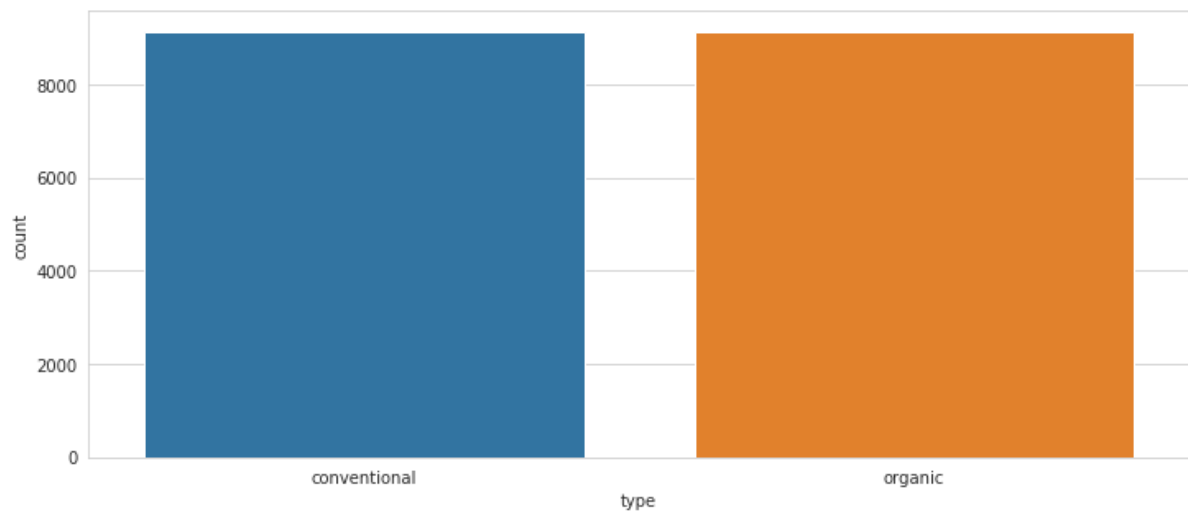
In this chart, we can see price changes according to type in cities. Organic avocado is more expensive as a standard.



Type Average Price in Each Region

The plot below shows type distribution of avocados in the data set.

# RESULT

Time series analysis is used to predict the result.

The model predicted that rising avocado prices would go down in the coming years.

```
fig, ax1 = plt.subplots(figsize=(16, 8))
ax1.plot(forecastnew.set_index('ds'), color='b')
ax1.plot(forecastedvalues.set_index('ds'), color='r')
ax1.set_ylabel('Average Prices')
ax1.set_xlabel('Date')
print("Red = Predicted Values, Blue = Base Values")
```
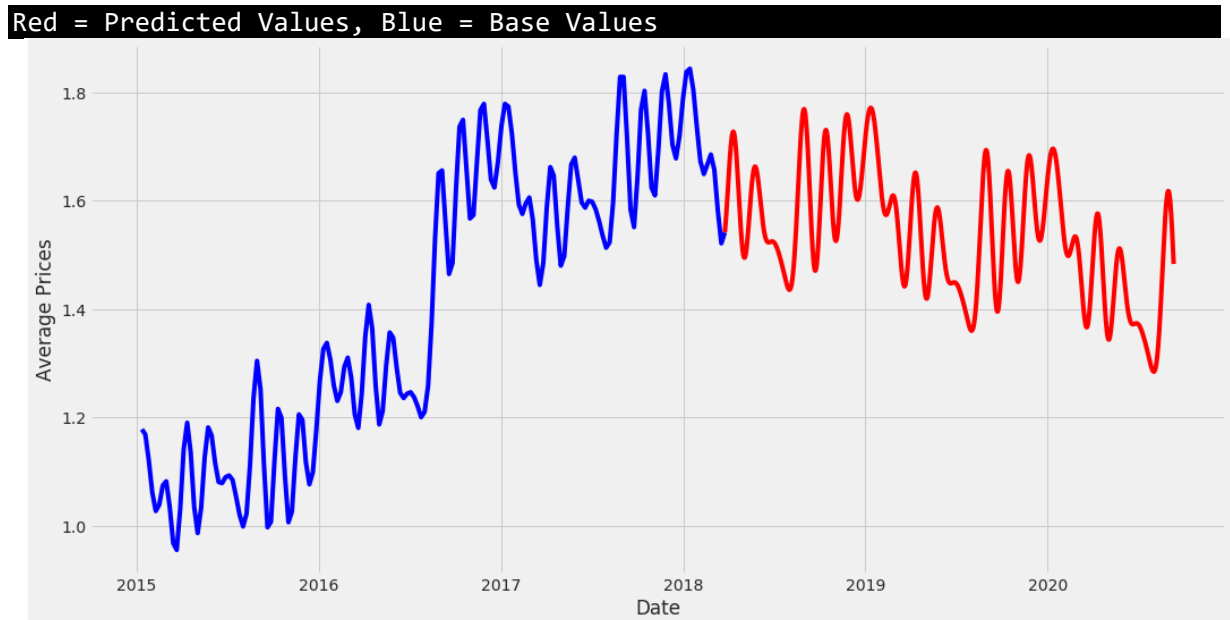
# CONCLUSION

During this project we learned a lot of skills in Data Science using Python. We learnt how to use time series analysis and call prophet model on our dataset and predicted that the prices of avocado would go down in the coming years. We chose the above models because they're more useful in our case of data amongst other algorithms.