Statistical Forecasting Project 1

Drashti Jiteshbhai Hindocha

Predictive Analysis (1498), Conestoga College

STAT 8041: Statistical Forecasting

Laura Neher

21st Feb 2024.

Data

Information is a crucial aspect in completing any project. One of the most challenging parts of this analysis is finding the right database. I searched different websites until I discovered a suitable database on Kaggle. Understanding the database is essential to completing the project successfully. For this project, the dataset contained relevant information for my individual time series analysis project. I aim to determine the relationship between a set of variables in the dataset and predict the average price of avocado based on the time series data.

Dataset

The following information was obtained from the Hass Avocado Board website in May of 2018, and put together into a single CSV file. The table shows the weekly retail scan data for 2018, regarding the national retail volume (units) and price of Hass avocados. The data is obtained directly from cash registers of retailers and represents the actual retail sales of Hass avocados.

There are two main types of avocados, each with 13 different attributes listed. The data can be explained as follows: the average price is the response variable, while the other attributes serve as predictors.

- **Date:** The date of the observation.
- **Average Price:** the average price of a single avocado.
- **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.
- **Total Bags:** Total number of bags.
- **Small Bags:** Total number of small bags.
- **Large Bags:** Total number of large bags.
- **XLarge Bags:** Extra Large Bags.
- **type:** conventional or organic.
- **year:** year of the date.
- **region:** the city or region of the observation.

The PLU (Price Look-Up) code is a 4 or 5-digit number used to identify products, typically in supermarkets or grocery stores. The International Federation for Produce Standards (IFPS) assigns the PLU codes to categorize different types of fruits and vegetables.

In the "Avocado Prices" dataset, the PLU codes 4046, 4225, and 4770 are used to categorize different types of avocados. Specifically, 4046, 4225, and 4770 are assigned to the Hass, Lamb Hass, and Green Skin varieties, respectively. The dataset provides information about the demand for each variety based on the total quantity of avocados sold for each PLU code during the given time.

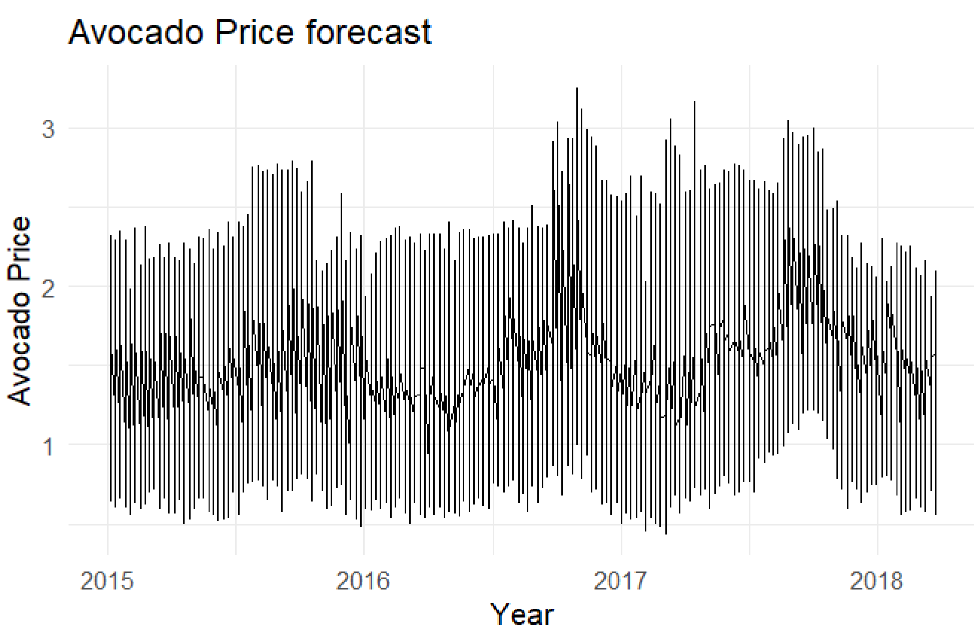
Exploratory Data Analysis

The data was presented in a time series and then divided into two parts: training and testing data. A time series has four key components:

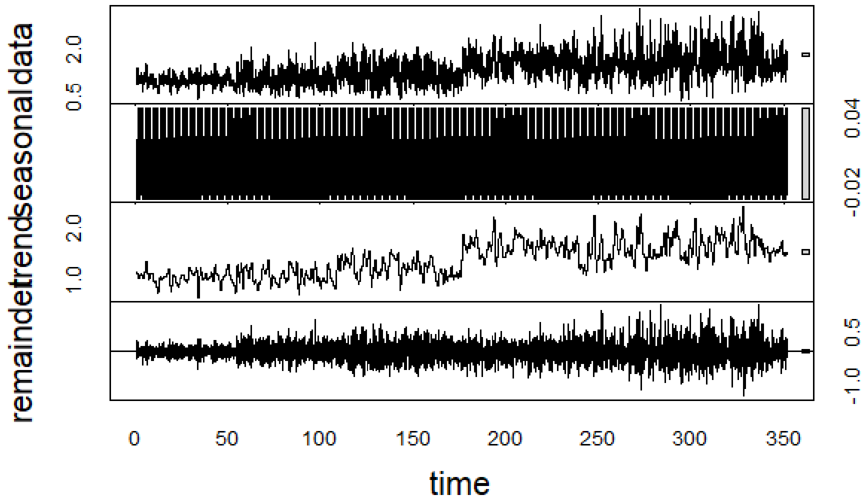
1. Trend: This shows the long-term behavior of the data.
2. Seasonality: This refers to recurring patterns in the data that are related to the time of year or weather conditions.
3. Cyclic Variations: These are recurring patterns in the data that repeat over a period of time, such as a year.
4. Irregularities: These are random and unexpected changes or fluctuations in the data.

In this phase of the project, we will concentrate mainly on one aspects of the time series forecasting analysis which consists of:

• **Seasonal Patterns**: In this section, we'll focus on reoccurring patterns in conventional and organic avocados that consistently show up month after month and year after year.

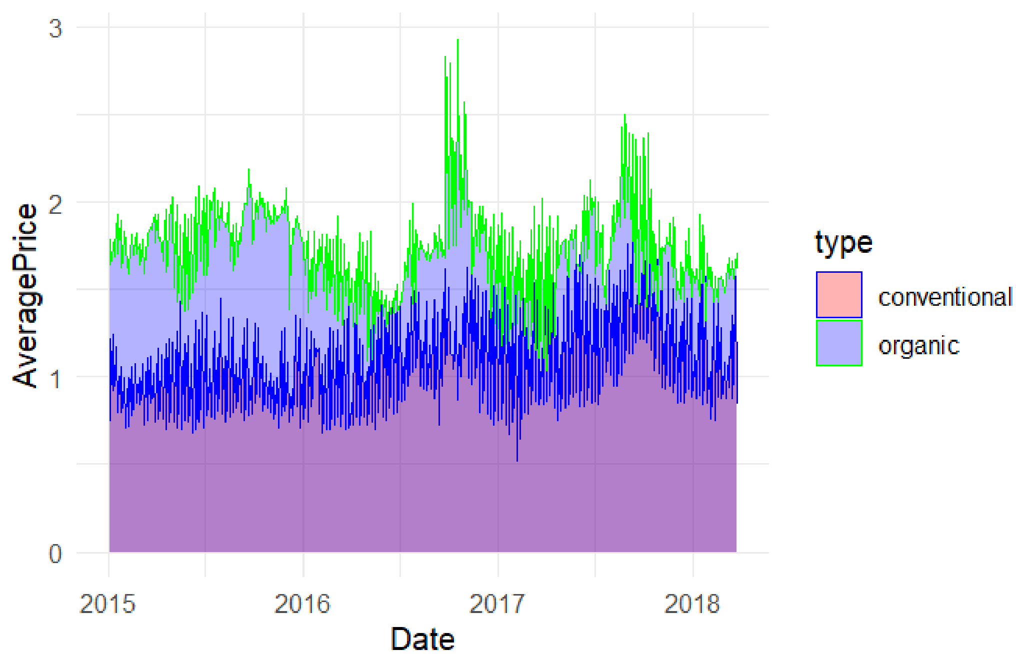


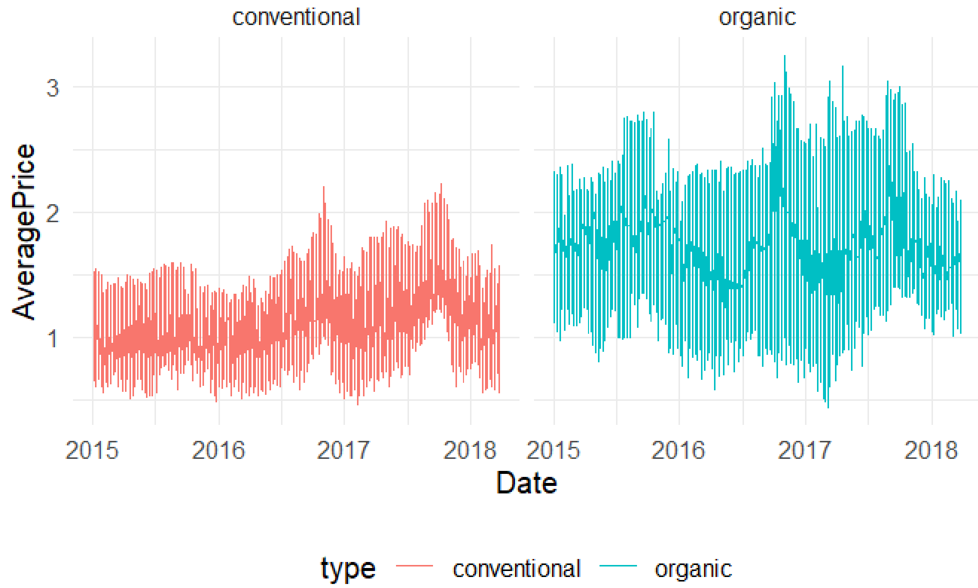
From the above graph, we can see that there is seasonality pattern that occur regularly over different interval of time. Avocado prices peak at regular intervals and then drop sharply. This cyclic behavior suggests seasonal factors influencing avocado prices.



In this part, we'll examine the various varieties of avocados included in this dataset. As mentioned earlier before, there are 2 types of avocados.
In essence, there are two varieties of avocados:

1) Organic Avocados
2) Conventional Avocados

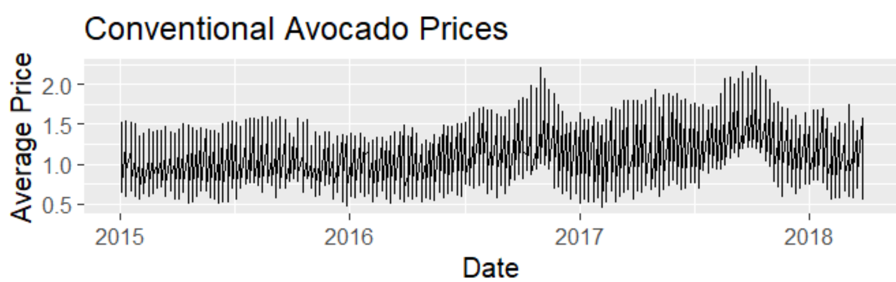


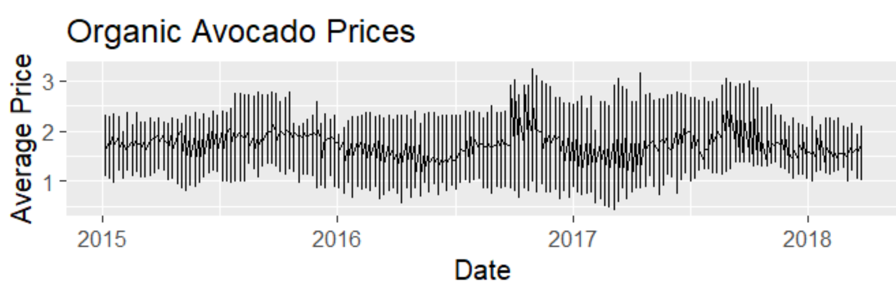


Based on the pricing variations over time, we can conclude that organic avocados have a higher price, while conventional avocados have a lower price.

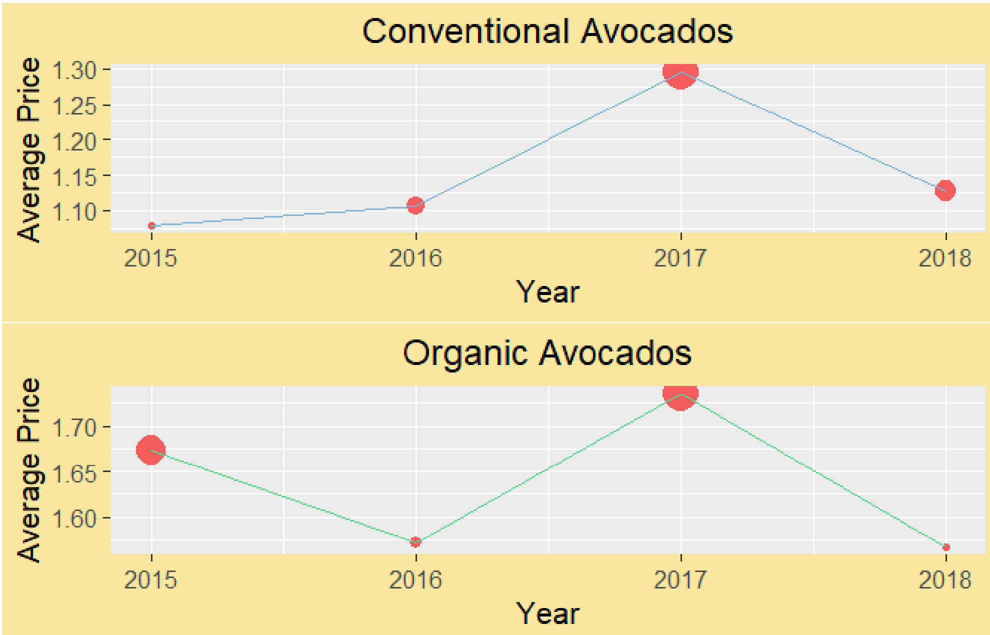
In the next section, our aim is to assess the overall price of avocados. An oversupply of avocados can have a negative impact on the market price. Therefore, we will analyse both conventional and organic avocados to see if they fit this description.

Seasonal Patterns Analysis: We will investigate if there are any notable recurring seasonal patterns. We will look for any instances where the price of avocados appears to increase regularly.

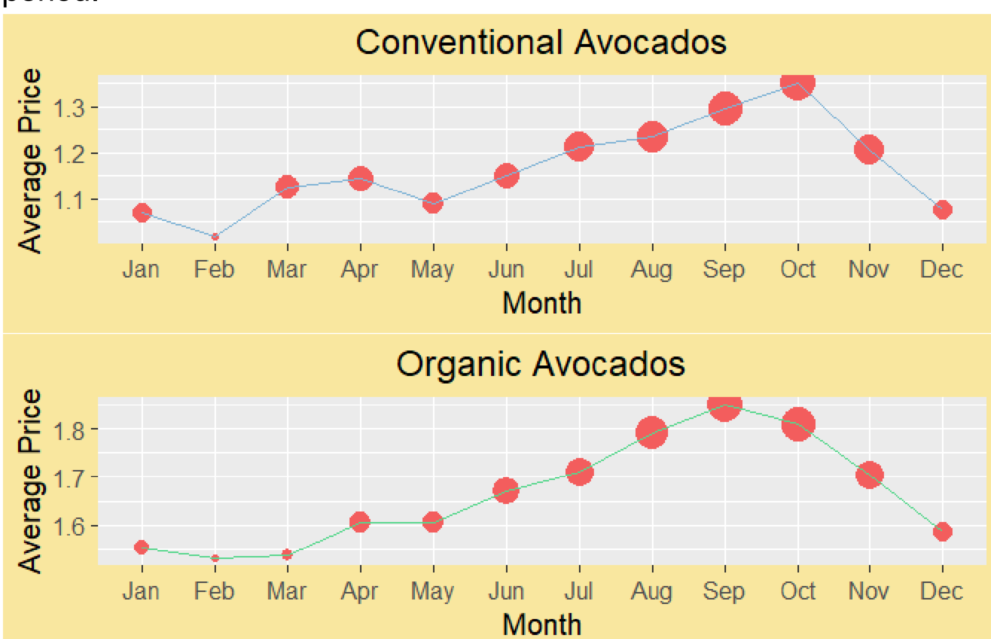




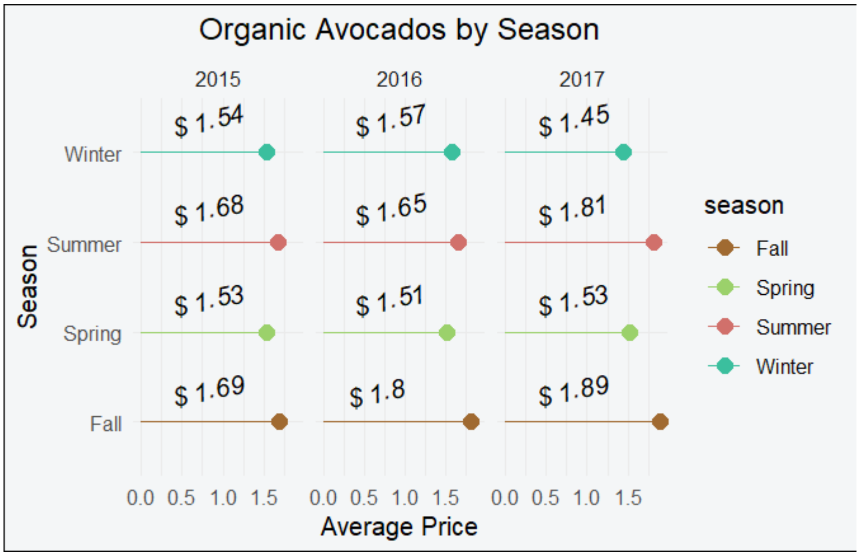
To see more precisely, we will see the graph in detail.

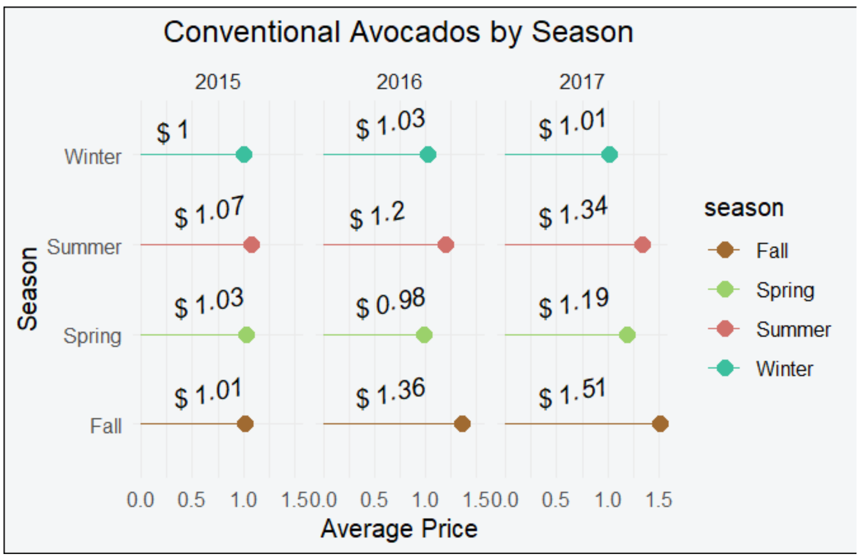


From the above graph , we can see that for conventional avocados the highest price is seen in 2017, while it’s the same for organic avocados along with the year 2015. We observe that avocado prices typically rise each year. Let’s look in mode detailed time period.



Thus, we can observe that, for a given year, avocado prices usually increase in May. Let's now examine how much the two varieties of avocados cost.





• Yearly distributions: In 2015, the majority of conventional avocado prices seemed to be around $1.00. Although in 2016 and 2017, the price density was marginally higher. The cost of organic avocados ranges from $1.50 to $1.90. They appear to be more expensive than regular avocados.

• Price peaks per month: It seems that the most expensive months for both conventional and organic avocados are September and October.

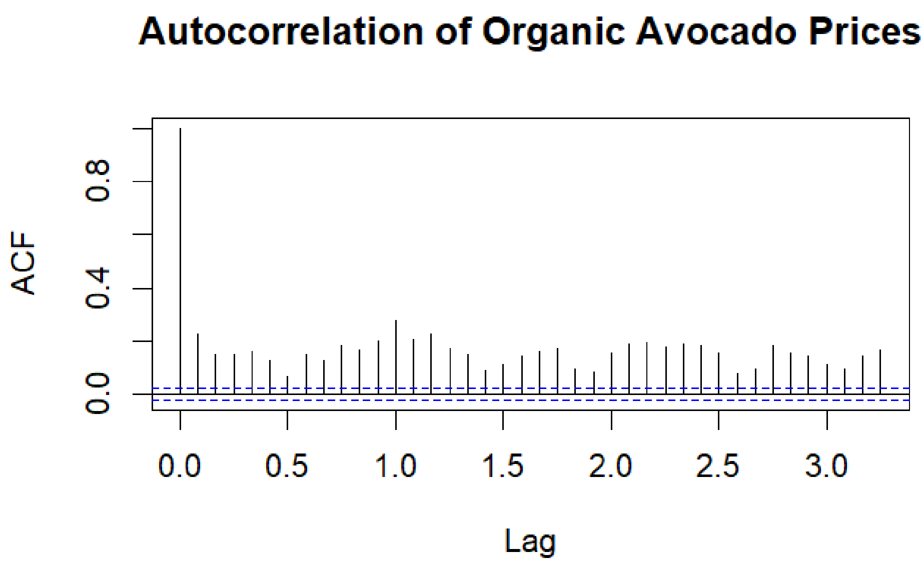
• Significant price reduction at year's end: It's interesting to observe that avocados become considerably less expensive as the year draws to a close. I'm wondering why that would be. Why is there a price reduction at the end of the year each year?

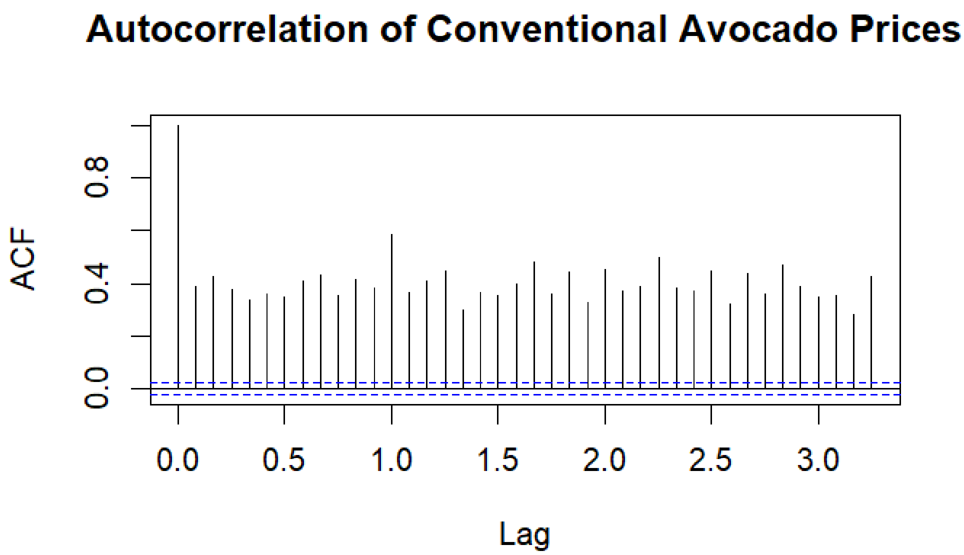
Autocorrelation

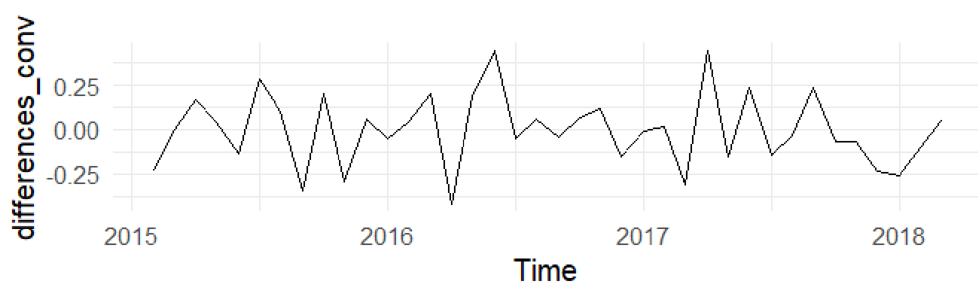
A time series' autocorrelation can be defined as the linear relationship between its lagged values. Lags can be thought of as intervals of time. As a result of its self-connection, the correlation at lag zero is always 1.

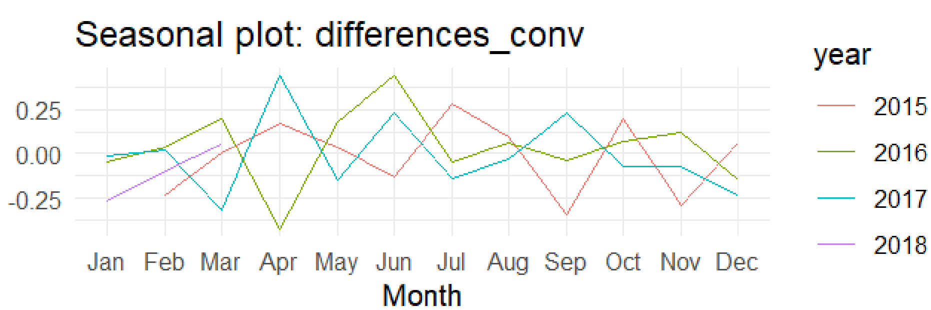
What does the term "internal correlation" mean? Essentially, our goal is to find any correlation between the price of avocados and the price of a specific date (we'll use January in the example below). In the first visualisation, lags are shown as month representations. As a month approaches January, we can see that the correlation grows, indicating that February's price will rise. There is a significant autocorrelation up until the third lag (March), suggesting that these weeks have a strong linear relationship with the first month.

Unlike the first month, there isn't a linear relationship between the subsequent months. Consequently, we can conclude that there is no consistent pattern that would suggest a linear correlation between the prices of conventional and organic avocados.




Autocorrelation of Conventional Avocado Prices
ACF
0.0
0.4
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


differences_conv
0.25
0.00
-0.25
2015
2016
2017
2018
Time

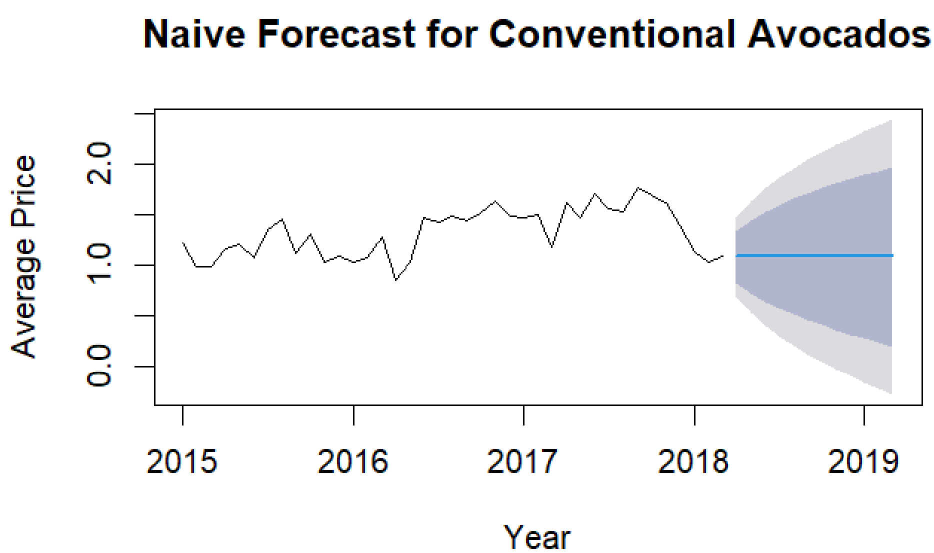

Seasonal plot: differences_conv
0.25
0.00
-0.25
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
year
2015
2016
2017
2018

Forecasting and Analysis

Method 1

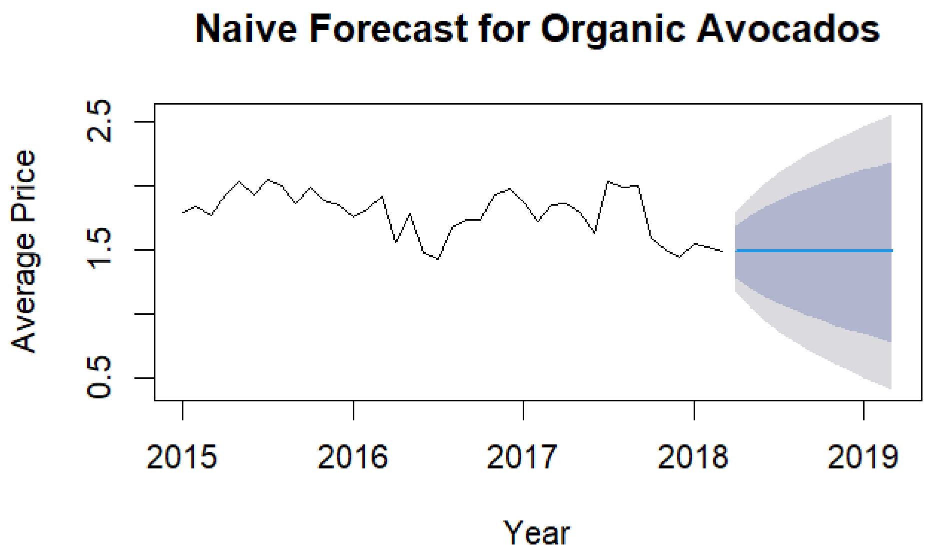
In time series analysis, the **Naive forecasting method** is a straightforward and intuitive technique for making predictions about future values based on past data. Without considering underlying patterns or factors, the method assumes that the future values will be the same as the most recent observed values. The naive method is simple to use and straightforward, but it might not be as effective at identifying complex patterns, seasonality, or trends in the time series.

According to the naive forecast, the average price of conventional avocados will remain constant at 1.08 for every month between April 2018 and March 2019, with different levels of confidence indicated by the associated prediction intervals.



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Apr 2018	1.08	0.8245511	1.335449	0.68932466	1.470675
May 2018	1.08	0.7187407	1.441259	0.52750163	1.632498
Jun 2018	1.08	0.6375496	1.522450	0.40333045	1.756670
Jul 2018	1.08	0.5691022	1.590898	0.29864931	1.861351
Aug 2018	1.08	0.5087989	1.651201	0.20642337	1.953577
Sep 2018	1.08	0.4542806	1.705719	0.12304475	2.036955
Oct 2018	1.08	0.4041458	1.755854	0.04637020	2.113630
Nov 2018	1.08	0.3574815	1.802519	-0.02499674	2.184997
Dec 2018	1.08	0.3136534	1.846347	-0.09202603	2.252026
Jan 2019	1.08	0.2721997	1.887800	-0.15542391	2.315424
Feb 2019	1.08	0.2327719	1.927228	-0.21572353	2.375724
Mar 2019	1.08	0.1950991	1.964901	-0.27333909	2.433339

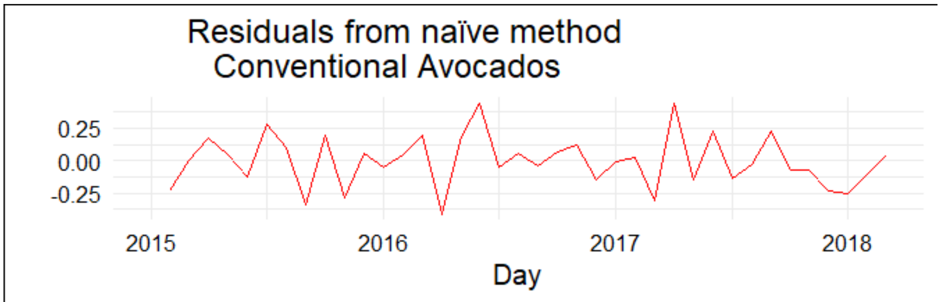
While the average price of organic avocados will remain constant at 1.48 for every month between April 2018 and March 2019, with different levels of confidence indicated by the associated prediction intervals.

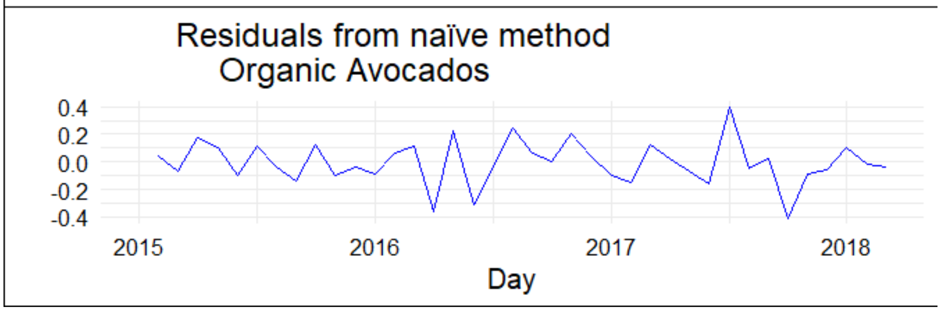


	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Apr 2018	1.48	1.2774222	1.682578	1.1701840	1.789816
May 2018	1.48	1.1935118	1.766488	1.0418541	1.918146
Jun 2018	1.48	1.1291250	1.830875	0.9433830	2.016617
Jul 2018	1.48	1.0748445	1.885156	0.8603681	2.099632
Aug 2018	1.48	1.0270224	1.932978	0.7872305	2.172770
Sep 2018	1.48	0.9837878	1.976212	0.7211090	2.238891
Oct 2018	1.48	0.9440296	2.015970	0.6603040	2.299696
Nov 2018	1.48	0.9070236	2.052976	0.6037082	2.356292
Dec 2018	1.48	0.8722667	2.087733	0.5505521	2.409448
Jan 2019	1.48	0.8393929	2.120607	0.5002759	2.459724
Feb 2019	1.48	0.8081256	2.151874	0.4524567	2.507543
Mar 2019	1.48	0.7782500	2.181750	0.4067661	2.553234

When using the naive forecasting method, **Residuals** are the differences between the observed values that are actually observed and the values that are predicted by the naive forecasting method.

To put it another way, the residual is the portion of the observed data that remains after the naive forecast has been applied, or the error. If the residual is positive, it means that the observed value exceeded the naive forecast, and if it is negative, it means that the observed value was less than the naive forecast.



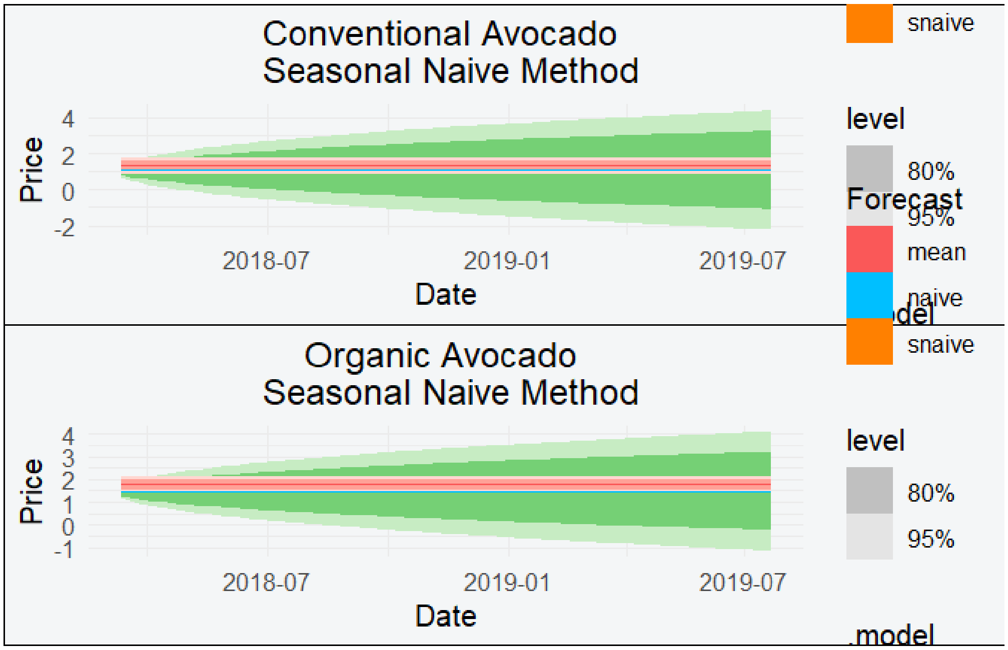


Conventional Avocados: The avocado conventional chart contains two unusual peaks. This could suggest unusual price swings, which might suggest that this was caused by some extraordinary event.

Organic Avocados: Although not to the same degree as conventional avocados, organic avocados do experience certain peaks. However, there are a few unusual peaks that require more analysis.

Method 2

An improved version of the Naive method created to handle time series data with different seasonal patterns is the **Seasonal Naive forecasting method**. By utilizing the most recent observed values from the same season in prior years, it forecasts future values. The straightforward but efficient approach is predicated on the idea that the seasonal pattern holds true over time. However, if seasonality changes or if other significant factors are at work, it might not function well. The forecasting accuracy of the method can be evaluated using assessment metrics like MAE, MSE, or RMSE.



As you can see, the blue line is level with the December price that was most recent. Now, the reason it is so low makes sense. Because avocados, both conventional and organic, were so inexpensive, this could also act as a support line. A new negative trend could begin if prices drop below this level. In the event that we use monthly data, for example, our January forecast will be equal to the January value from the previous year (2018).

2015: Residuals are denoted with NA, suggesting that either the forecasting model did not produce predictions for this period or there may be missing data. The **residuals** for every month in the years 2016–2018 are displayed for conventional avocados using seasonal naïve method.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2016	-0.19	0.08	0.28	-0.31	-0.17	0.40	0.07	0.03	0.33	0.20	0.61	0.40
2017	0.44	0.42	-0.09	0.77	0.44	0.23	0.14	0.05	0.32	0.18	-0.01	-0.09
2018	-0.34	-0.46	-0.10									

As an example: The observed value in January 2016 was 0.19 units less than the predicted value. The observed value in February 2016 was 0.08 units higher than the predicted value.

The residuals for every month in the years 2016–2018 are displayed for organic avocados using seasonal naïve method.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2016	-0.04	-0.02	0.16	-0.37	-0.25	-0.46	-0.61	-0.33	-0.13	-0.25	0.05	0.13
2017	0.12	-0.09	-0.08	0.30	0.01	0.16	0.60	0.31	0.27	-0.14	-0.43	-0.53
2018	-0.33	-0.20	-0.36									

As an example: The observed value in January 2016 was 0.04 units less than the predicted value. The observed value in February 2016 was 0.02 units less than the predicted value. The observed value in March 2016 exceeded the predicted value by 0.16 units.

Prediction Intervals

A prediction interval is a range of values that, with a certain degree of confidence, a future observation is likely to fall within. They provide an indication of the possible variability in the predicted values by providing a measure of uncertainty around a point forecast. Greater uncertainty is indicated by a wider prediction interval, whereas greater forecast confidence is indicated by a narrower interval. Confidence levels of 80%, 90%, and 95% are typical.

A prediction interval is a range that we anticipate lying within with a certain probability. This is conceptually related to confidence intervals.

For example, our prediction interval is as follows:
Prediction Interval of 95% for Conventional Avocados is: [0.64533143 1.552338]
Prediction Interval of 95% for Organic Avocados is: [0.879072602 1.626391]

Conclusion

Expensive Organic Avocados: As expected, organic avocados are significantly more expensive than regular avocados.

Similar Patterns: Although there are some differences between the two avocado varieties, most of the patterns are similar.

2017-best year: 2017 is the most notable year for avocados, probably because to the state of the economy as well as other factors that are not obvious but have an impact on market prices.

Seasonal Trend: Consider Buying Avocados Before Fall: As autumn draws near, prices for both conventional and organic avocados rise steadily. This indicates that consumers may want to think about buying avocados ahead of time.

Long-Term Declining Trend: The forecasting model predicts that the prices of both avocado varieties will decline over time. Nevertheless, despite the general downward trend over the following two years, there may be some short-term upward pressure seen.

References

https://www.kaggle.com/datasets/smokingkrils/avacado-price-prediction/data

Appendix

```
# Statistical Forecasting_Project 1_Drashti Hindocha

# Install and load packages
install.packages("tidyverse", repos = "https://cran.rstudio.com/")
install.packages("caret", repos = "https://cran.rstudio.com/")
install.packages("readxl", repos = "https://cran.rstudio.com/")
install.packages("ggplot2", repos = "https://cran.rstudio.com/")
install.packages("dplyr", repos = "https://cran.rstudio.com/")
install.packages("lubridate", repos = "https://cran.rstudio.com/")
install.packages("gridExtra", repos = "https://cran.rstudio.com/")
install.packages("tsibble", repos = "https://cran.rstudio.com/")
install.packages("forecast", repos = "https://cran.rstudio.com/")


# Load libraries
library(fpp3)
library(tidyverse)
library(forecast)
library(dplyr)
library(tsibble)
library(ggplot2)
library(readr)
library(seasonal)
library(knitr)
library(ggfortify)
library(gridExtra)
library(fpp2)
library(tibbletime)
library(cowplot)
```

```
library(zoo)
library(fable)
library(fabletools)

#Load the Avocado dataset
Avodata <- read_csv("C:/Users/admin/Documents/Conestoga college/SEM 2/Statistical Forcasting/Project 1/avocado.csv")

#Summary
summary(Avodata)

# Data Cleaning
sum(is.na(Avodata))
data <- na.omit(data)

#Since every value in the matrix is zero in this instance, your data object has no missing values (NA). This shows that none of the variables in this dataset have any empty or missing values.

# Selecting region and type of avocado to forecast
set.seed(133)

# Time series plot
ggplot(Avodata, aes(x = Date, y = AveragePrice)) +
  geom_line() +
  labs(x = "Year", y = "Avocado Price", title = "Avocado Price forecast") + theme_minimal()

# Additional plot
ggplot() +
  geom_line(data = Avodata, aes(x = Date, y = AveragePrice), color = "blue", size = 0.6) +
  geom_vline(xintercept = c(as.Date("2017-09-10"), as.Date("2018-03-25")), linetype = "dashed", size = 1) +
  annotate("rect", xmin = as.Date("2017-09-10"), xmax = as.Date("2018-03-25"), ymin = -Inf,
           ymax = Inf, alpha = 0.1, fill = "green") +
  xlab("Year") + ylab("Avocado Price($)") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "red"), legend.position = "none")


avocado_ts <- ts(Avodata$AveragePrice)

#ACF Plot
acf(avocado_ts, main = "Autocorrelation of Avocado Prices", lag.max = 52)
pacf(Avodata$AveragePrice, lag.max = 52, main = "")

# Create a time series object with weekly avocado prices
week_price <- ts(na.omit(Avodata$AveragePrice), frequency = 52)

# Decompose the time series using STL
decomp <- stl(week_price, s.window = "periodic")

# Plot the decomposition components
plot(decomp, cex = 0.6)
```

```
#Convert the "Date" Column to Date Format
Avodata$Date <- as.Date(Avodata$Date, "%Y-%m-%d")
class(Avodata$Date)
Avodata <- Avodata[order(as.Date(Avodata$Date, format="%Y-%m-%d")),]

# Create an Area Plot for Two Types of Avocado
price_trend <- Avodata %>%
  select(Date, AveragePrice, type, region) %>%
  ggplot(aes(x = Date, y = AveragePrice)) +
  geom_area(aes(color = type, fill = type), alpha = 0.3, position = position_dodge(0.8)) +
  theme_minimal() +
  scale_color_manual(values = c("blue", "green")) +
  scale_fill_manual(values = c("red", "blue"))

# Print the plot
print(price_trend)

# Create a Facet Wrap for each product
ggplot(data = Avodata, aes(x = Date, y = AveragePrice, col = type)) +
  geom_line() +
  facet_wrap(~ type) +
  theme_minimal() +
  theme(legend.position = "bottom")

# Filter by type: Organic
organic <- Avodata %>%
  select(Date, AveragePrice, type, `Total Volume`) %>%
  filter(type == "organic")

# Filter by type: Conventional
conventional <- Avodata %>%
  select(Date, AveragePrice, type, `Total Volume`) %>%
  filter(type == "conventional")

# For Organic Avocado prices:
# Convert the "Date" column to a Date type
organic$Date <- as.Date(organic$Date)

# Create a time series object
organic_ts <- ts(organic$AveragePrice, start = c(year(organic$Date[1]), month(organic$Date[1])), frequency = 12)

# Generate the ACF plot
acf(organic_ts, main = "Autocorrelation of Organic Avocado Prices")


#For Conventional Avocado Prices:
# Convert the "Date" column to a Date type
conventional$Date <- as.Date(conventional$Date)

# Create a time series object
```

```
conventional_ts <- ts(conventional$AveragePrice, start = c(year(conventional$Date[1]), month(conventional$Date[1])), frequency = 12)

# Generate the ACF plot
acf(conventional_ts, main = "Autocorrelation of Conventional Avocado Prices")


# Filter the data for conventional avocados
conventional <- subset(Avodata, type == "conventional")

# Filter the data for organic avocados
organic <- subset(Avodata, type == "organic")

# Create time series plots
conventional_plot <- ggplot(conventional, aes(x = Date, y = AveragePrice)) +
  geom_line() +
  labs(x = "Date", y = "Average Price", title = "Conventional Avocado Prices")

organic_plot <- ggplot(organic, aes(x = Date, y = AveragePrice)) +
  geom_line() +
  labs(x = "Date", y = "Average Price", title = "Organic Avocado Prices")

# Arrange the plots in a grid
grid_arrange <- gridExtra::grid.arrange(conventional_plot, organic_plot, nrow = 2)

# Display the grid of plots
print(grid_arrange)

# Detecting seasonality patterns for Conventional Avocados
conv_patterns <- Avodata %>%
  select(year, AveragePrice, type) %>%
  filter(type == "conventional") %>%
  group_by(year) %>%
  summarize(avg = mean(AveragePrice)) %>%
  ggplot(aes(x = year, y = avg)) +
  geom_point(color = "#F35D5D", aes(size = avg)) +
  geom_line(group = 1, color = "#7FB3D5") +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F9E79F")) +
  labs(title = "Conventional Avocados", x = "Year", y = "Average Price")

# Detecting seasonality patterns for Organic Avocados
org_patterns <- Avodata %>%
  select(year, AveragePrice, type) %>%
  filter(type == "organic") %>%
  group_by(year) %>%
  summarize(avg = mean(AveragePrice)) %>%
  ggplot(aes(x = year, y = avg)) +
  geom_point(color = "#F35D5D", aes(size = avg)) +
  geom_line(group = 1, color = "#58D68D") +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F9E79F")) +
  labs(title = "Organic Avocados", x = "Year", y = "Average Price")
```

```
# Arrange the plots in a grid
grid_arrange <- grid.arrange(conv_patterns, org_patterns, nrow = 2)

# Display the grid of plots
print(grid_arrange)

# Create a copy of the original dataset
seasonal_df <- Avodata
seasonal_df$month_year <- format(as.Date(Avodata$Date), "%Y-%m")
seasonal_df$month <- format(as.Date(Avodata$Date), "%m")
seasonal_df$year <- format(as.Date(Avodata$Date), "%Y")
seasonal_df$monthabb <- sapply(seasonal_df$month, function(x) month.abb[as.numeric(x)])
seasonal_df$monthabb <- factor(seasonal_df$monthabb, levels = month.abb)

# Plot for Conventional Avocados
conv_patterns <- seasonal_df %>%
  select(monthabb, AveragePrice, type) %>%
  filter(type == "conventional") %>%
  group_by(monthabb) %>%
  summarize(avg = mean(AveragePrice)) %>%
  ggplot(aes(x = monthabb, y = avg)) +
  geom_point(color = "#F35D5D", aes(size = avg)) +
  geom_line(group = 1, color = "#7FB3D5") +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F9E79F")) +
  labs(title = "Conventional Avocados", x = "Month", y = "Average Price")

# Plot for Organic Avocados
org_patterns <- seasonal_df %>%
  select(monthabb, AveragePrice, type) %>%
  filter(type == "organic") %>%
  group_by(monthabb) %>%
  summarize(avg = mean(AveragePrice)) %>%
  ggplot(aes(x = monthabb, y = avg)) +
  geom_point(color = "#F35D5D", aes(size = avg)) +
  geom_line(group = 1, color = "#58D68D") +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F9E79F")) +
  labs(title = "Organic Avocados", x = "Month", y = "Average Price")

# Display the grid of plots
grid_arrange <- grid.arrange(conv_patterns, org_patterns, nrow = 2)
print(grid_arrange)

# Set plot dimensions
options(repr.plot.width=8, repr.plot.height=6)

# Create a new column for the season
seasonal_df$season <- ifelse(seasonal_df$month %in% c("03", "04","05"), "Spring",
                  ifelse(seasonal_df$month %in% c("06","07" ,"08"), "Summer",
                     ifelse(seasonal_df$month %in% c("09","10","11"), "Fall", "Winter")))
```

```
# Plot for conventional avocados by season
seasonality.plot.conventional <- seasonal_df %>% select(season, year, AveragePrice, type) %>%
  filter(type == "conventional", year == c("2015", "2016", "2017")) %>%
  group_by(season, year) %>%
  summarize(avg=mean(AveragePrice)) %>%
  ggplot(aes(x=season, y=avg, color=season)) +
  geom_point(size=3) +
  geom_segment(aes(x=season,
                   xend=season,
                   y=0,
                   yend=avg)) +
  coord_flip() +
  facet_wrap(~as.factor(year)) +
  theme_minimal() +
  theme(plot.title=element_text(hjust=0.5), plot.background=element_rect(fill="#F4F6F7")) +
  scale_color_manual(values=c("#a06a31", "#9bd16b", "#d1706b", "#3bbf9e")) +
  labs(title="Conventional Avocados by Season", x="Season", y="Average Price") +
  geom_text(aes(x=season, y=0.01, label= paste0("$ ", round(avg,2))),
            hjust=-0.5, vjust=-0.5, size=4,
            colour="black", fontface="italic",
            angle=372)

# Plot for organic avocados by season
seasonality.plot.organic <- seasonal_df %>% select(season, year, AveragePrice, type) %>%
  filter(type == "organic", year == c("2015", "2016", "2017")) %>%
  group_by(season, year) %>%
  summarize(avg=mean(AveragePrice)) %>%
  ggplot(aes(x=season, y=avg, color=season)) +
  geom_point(size=3) +
  geom_segment(aes(x=season,
                   xend=season,
                   y=0,
                   yend=avg)) +
  coord_flip() +
  facet_wrap(~as.factor(year)) +
  theme_minimal() +
  theme(plot.title=element_text(hjust=0.5), plot.background=element_rect(fill="#F4F6F7")) +
  scale_color_manual(values=c("#a06a31", "#9bd16b", "#d1706b", "#3bbf9e")) +
  labs(title="Organic Avocados by Season", x="Season", y="Average Price") +
  geom_text(aes(x=season, y=0.01, label= paste0("$ ", round(avg,2))),
            hjust=-0.5, vjust=-0.5, size=4,
            colour="black", fontface="italic",
            angle=372)

# Display the plots
print(seasonality.plot.conventional)
print(seasonality.plot.organic)

# Selecting data
conv <- seasonal_df %>% filter(type == "conventional")
org <- seasonal_df %>% filter(type == "organic")
```

```
# Declare data as time series
conventional <- as_tbl_time(conv, index = Date)
conventional <- as_period(conventional, '1 month')
conventional$type <- NULL

organic <- as_tbl_time(org, index = Date)
organic <- as_period(organic, '1 month')
organic$type <- NULL

# Create time series objects
conv_ts <- ts(conventional$AveragePrice, start = c(2015, 1), frequency = 12)
org_ts <- ts(organic$AveragePrice, start = c(2015, 1), frequency = 12)

# Calculate the differences between consecutive months for conventional avocado prices
differences_conv <- diff(conv_ts)

# Plotting the differences and seasonality plot
main_diff <- autoplot(differences_conv) + theme_minimal()
seasonality_diff <- ggseasonplot(differences_conv) + theme_minimal()

# Arrange the plots in a grid
grid_arrange <- grid.arrange(main_diff, seasonality_diff, nrow=2)


# Forecasting

# Method 1 Naive method


# Naive Forecast for Conventional Avocados
conv_naive_forecast <- naive(conv_ts, h = 12)

# Plot the Naive forecast for Conventional Avocados
options(repr.plot.width=12, repr.plot.height=6)
plot(conv_naive_forecast, main = "Naive Forecast for Conventional Avocados", xlab = "Year", ylab = "Average Price")

# Display the Naive forecast values
print(conv_naive_forecast)

# Naive Forecast for Organic Avocados
org_naive_forecast <- naive(org_ts, h = 12)

# Plot the Naive forecast for Organic Avocados
plot(org_naive_forecast, main = "Naive Forecast for Organic Avocados", xlab = "Year", ylab = "Average Price")

# Display the Naive forecast values
print(org_naive_forecast)

# Calculate residuals for the naive method on conventional avocados
rescv_nv <- residuals(naive(conv_ts, h = 72))
```

```
p1 <- autoplot(rescv_nv, color = "red") + xlab("Day") + ylab("") +
  ggtitle("Residuals from naïve method \n Conventional Avocados") + theme_minimal() +
  theme(plot.title = element_text(hjust = 0.20, color = "black"),
        plot.background = element_rect(fill = "white"),
        axis.text.x = element_text(colour = "black"),
        axis.text.y = element_text(colour = "black"),
        axis.title = element_text(colour = "black"))

# Calculate residuals for the naive method on organic avocados
resorg_nv <- residuals(naive(org_ts, h = 72))
p2 <- autoplot(resorg_nv, color = "blue") + xlab("Day") + ylab("") +
  ggtitle("Residuals from naïve method \n Organic Avocados") + theme_minimal() +
  theme(plot.title = element_text(hjust = 0.20, color = "black"),
        plot.background = element_rect(fill = "white"),
        axis.text.x = element_text(colour = "black"),
        axis.text.y = element_text(colour = "black"),
        axis.title = element_text(colour = "black"))


grid_arrange <- grid.arrange(p1, p2, nrow=2)
sqrt(0.05354)


# Method 2 Seasonal Naive method


options(repr.plot.width=10, repr.plot.height=7)

# Convert time series data to tsibble objects
conv_ts <- as_tsibble(conventional, key = NULL, index = Date)
org_ts <- as_tsibble(organic, key = NULL, index = Date)

# Convert the data to tsibble format
conv_tsibble <- as_tsibble(conventional, key = NULL, index = Date)
org_tsibble <- as_tsibble(organic, key = NULL, index= Date )

# Fill implicit gaps in time with explicit missing values
conv_tsibble <- fill_gaps(conv_tsibble)
org_tsibble <- fill_gaps(org_tsibble)

# Specify the seasonal lag value
seasonal_lag <- 12  # Assuming monthly data

# Forecast using seasonal naive method for coventional avocados
conv_forecast_sn <- conv_tsibble %>%
  model(
    mean = MEAN(AveragePrice),
    naive = NAIVE(AveragePrice),
    snaive = SNAIVE(AveragePrice ~ RW())
  ) %>%
  forecast(h = 72) %>%
```

```
  autoplot() +
  ggplot2::labs(title = "Conventional Avocado \n Seasonal Naive Method", x = "Date", y = "Price") +
  scale_color_manual(values = c("#FA5858", "#00BFFF", "#FF8000")) +
  guides(colour = guide_legend(title = "Forecast")) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F4F6F7"))

# Forecast using seasonal naive method for organic avocados
org_forecast_sn <- org_tsibble %>%
  model(
    mean = MEAN(AveragePrice),
    naive = NAIVE(AveragePrice),
    snaive = SNAIVE(AveragePrice ~ RW())
  ) %>%
  forecast(h = 72) %>%
  autoplot() +
  ggplot2::labs(title = "Organic Avocado \n Seasonal Naive Method", x = "Date", y = "Price") +
  scale_color_manual(values = c("#FA5858", "#00BFFF", "#FF8000")) +
  guides(colour = guide_legend(title = "Forecast")) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5), plot.background = element_rect(fill = "#F4F6F7"))

# Set the plot dimensions
options(repr.plot.width = 10, repr.plot.height = 7)
grid_arrange <- grid.arrange(conv_forecast_sn, org_forecast_sn, nrow=2)

# Convert the data to a time series object
conv_ts <- ts(conventional$AveragePrice, start = c(2015, 1), frequency = 12)
org_ts <- ts(organic$AveragePrice, start = c(2015, 1), frequency = 12)
conv_ts_index <- as.Date(time(conv_ts))
org_ts_index <- as.Date(time(org_ts))


#Residual using Seasonal naive method
# Seasonal Naive Forecast for Conventional Avocados
conv_seasonal_naive_forecast <- snaive(conv_ts, h = 12)

# Extract the predicted values
conv_seasonal_naive_pred <- fitted(conv_seasonal_naive_forecast)

# Calculate residuals for Conventional Avocados
conv_residuals <- residuals(conv_seasonal_naive_forecast)

# Display the residuals for Conventional Avocados
print(conv_residuals)

# Seasonal Naive Forecast for Organic Avocados
org_seasonal_naive_forecast <- snaive(org_ts, h = 12)

# Extract the predicted values
org_seasonal_naive_pred <- fitted(org_seasonal_naive_forecast)
```

```
# Calculate residuals for Organic Avocados
org_residuals <- residuals(org_seasonal_naive_forecast)

# Display the residuals for Organic Avocados
print(org_residuals)


#Prediction Interval for Seasonal Naive method


# Calculate the standard deviation of residuals
residual_sd <- sd(conv_residuals)
residual_sd <- sd(org_residuals)

# Set the confidence level (e.g., 95%)
confidence_level <- 0.95

# Calculate the z-score for the given confidence level
z_score <- qnorm((1 + confidence_level) / 2)

# Calculate the prediction interval
conv_prediction_interval <- cbind(lower = conv_seasonal_naive_pred - z_score * residual_sd,
                         upper = conv_seasonal_naive_pred + z_score * residual_sd)
org_prediction_interval <- cbind(lower = org_seasonal_naive_pred - z_score * residual_sd,
                         upper = org_seasonal_naive_pred + z_score * residual_sd)

# Combine the prediction interval with the forecasted values
conv_with_prediction_interval <- cbind(conv_seasonal_naive_pred, conv_prediction_interval)
org_with_prediction_interval <- cbind(org_seasonal_naive_pred, org_prediction_interval)

# Print the results
print(conv_with_prediction_interval)
print(org_with_prediction_interval)
```