MSA 8200 – PREDICTIVE ANALYTICS - FINAL PROJECT REPORT

Avocado Price Forecasting using Time Series Analysis (ARIMA & SARIMA) and Deep Learning Model (LSTM)

Team member	Email ID
Vaishnavi Chilumuru	Vchilumuru1@student.gsu.edu
Indira Aryal	iaryal1@student.gsu.edu
Tanushree Mahesh	tmaheshkumar1@student.gsu.edu

Abstract

In an era where data-driven decision-making is paramount, accurate forecasting stands as a critical tool for market analysis and strategic planning. This project is anchored in the aim to forecast the average price of avocados using time series analysis, leveraging historical sales data to predict future market trends. The dataset underpinning this study encompasses a range of attributes from the date of observation to the distribution of sales across different avocado types and bag sizes, including both conventional and organic categories, across various regions and years.

We delve into time series decomposition, investigating patterns within the average price fluctuations over time. Through rigorous preprocessing steps including logarithmic transformations and differencing, we address non-stationarity — a typical challenge in time series forecasting. By deploying ARIMA, SARIMA, and LSTM models, we juxtapose traditional statistical approaches with advanced deep learning techniques, evaluating their predictive performances using the Root Mean Square Error (RMSE) metric.

Our research navigates through complex seasonal behaviors and shifting consumer preferences, offering insights into the avocado market dynamics. The outcomes of this study are not only expected to bolster the accuracy of price forecasting but also to contribute to the nuanced understanding of demand elasticity influenced by temporal trends. The implications of this work resonate with agricultural producers, retailers, and policy-makers who stand to benefit from precise market forecasts.

Problem Statement

Avocado prices are notorious for their volatility, influenced by a myriad of temporal factors such as seasonality, economic trends, and consumer preferences. The challenge lies in developing accurate forecasting models that can decipher these temporal intricacies, empowering stakeholders to anticipate price fluctuations and strategize accordingly. This project aims to harness the power of time series analysis techniques to predict the average price of avocados, utilizing a comprehensive dataset encompassing various attributes including volume, type, and region. By delving into statistical tests for time series analysis, exploring the nuances of ARIMA and SARIMA models, and deploying in-sample and out-of-sample forecasting methodologies, this project endeavors to equip stakeholders with actionable insights for effective pricing strategies, inventory management, and market trend analysis.

Dataset Information

The dataset for this study was meticulously sourced from the Hass Avocado Board, providing a comprehensive view of avocado sales dynamics across the United States. It spans several years and includes granular details that capture the essence of market fluctuations and consumer behavior over time. The attributes encapsulated within this dataset are as follows:

Date: Representing the date of observation, this field is fundamental for the time series analysis, enabling the chronological ordering of price movements and volume sales.
AveragePrice: A critical metric that represents the average price of a single avocado, it serves as the target variable for our forecasting models.
Total Volume: Aggregating the total sales volume of avocados, this attribute provides insight into the overall market demand.

PLU-Based Sales Breakdown (4046, 4225, 4770): These attributes detail the sales volumes of avocados differentiated by their Product Lookup Codes (PLUs), which correspond to specific avocado types and sizes, enabling an analysis of consumer preferences for particular avocado varieties.

Bagging Breakdown (Total Bags, Small Bags, Large Bags, XLarge Bags): This categorization offers a window into packaging trends and the popularity of different bag sizes, which can reflect marketing effectiveness and consumption patterns.

Type: Distinguishing between conventional and organic produce, this factor allows for a comparative analysis of these two significant market segments.

Year: The year of observation is crucial for understanding long-term trends and annual patterns in avocado pricing and sales.

Region: Providing geographical granularity, this attribute spans various cities and regions, shedding light on localized market behaviors and potential regional preferences.

By harnessing data from the Hass Avocado Board, our analysis benefits from a dataset rich in detail and relevance, ensuring that our forecasts are grounded in real-world patterns and market behaviors.

Exploratory Data Analysis

Visualization and Resampling of 'AveragePrice':

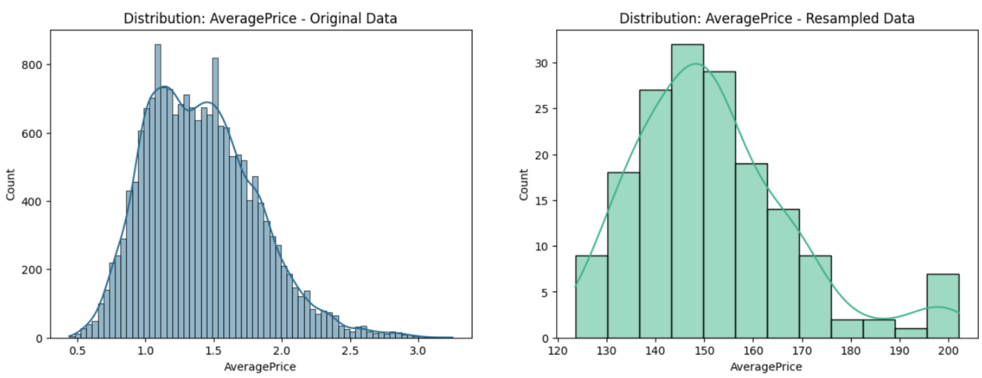
The exploration of our dataset's target variable, 'AveragePrice,' is illustrated through two primary visualizations that capture its distribution before and after resampling:

Initial Distribution: The initial histogram of 'AveragePrice' portrays a near-normal distribution with slight bimodality. This visualization reflects the raw, daily price data and hints at two dominant price points, which could correlate with the types of avocados.

Post-Resampling Distribution: After resampling the data to weekly intervals, the new histogram exhibits a more pronounced normal distribution. The resampling process aggregates daily prices into weekly sums, ironing out daily price volatility and offering a clearer view of the overall price trend.

Trend Over Time: The line chart depicts 'AveragePrice' over time, demonstrating a positive trend. Seasonal patterns are noticeable, with prices typically dipping in the winter months and peaking during the fall harvest season.

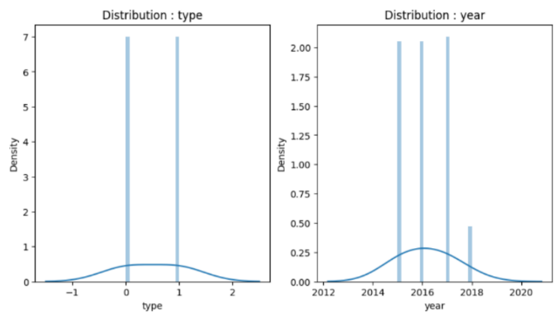
These visualizations underscore the impact of resampling on data analysis, smoothing out short-term fluctuations to reveal the underlying trends and cyclical behaviors that are pivotal for accurate forecasting.



Categorical Feature Analysis in Avocado Price Forecasting :

In our dataset, we have segregated features into categorical and numerical types for a structured exploratory data analysis

In our analysis, we categorized the dataset's features to understand their distribution and relationship with the target variable, AveragePrice. Our findings can be summarized as follows:

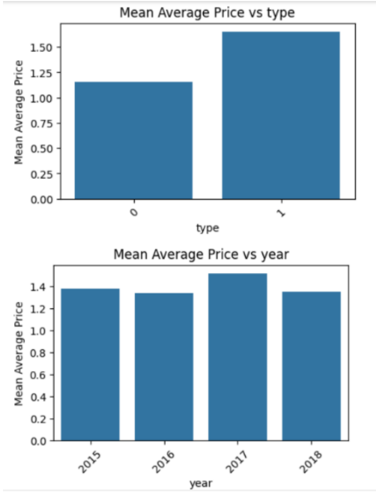


Distribution of Categorical Features:

The types of avocados, labeled as 'Conventional' (0) and 'Organic' (1), are evenly represented in the dataset, providing a balanced perspective for comparative analysis.

The yearly distribution of data is consistent for the years 2015 to 2017, indicating a stable dataset acros s this period. However, there's a significant drop in the count for 2018, suggesting that data for this year is either incomplete or not fully recorded.

Categorical Features vs. Target Variable Analysis:



When examining the impact of the avocado type on the **AveragePrice**, we find that Conventional avocados are priced lower than Organic ones, a trend consistent with market expectations regarding organic produce.

The AveragePrice of avocados has remained relatively stable for the years 2015, 2016, and 2018. However, 2017 stands out with an observable increase in AveragePrice, which may be attributed to factors such as supply chain disruptions, market demand fluctuations, or external economic conditions.

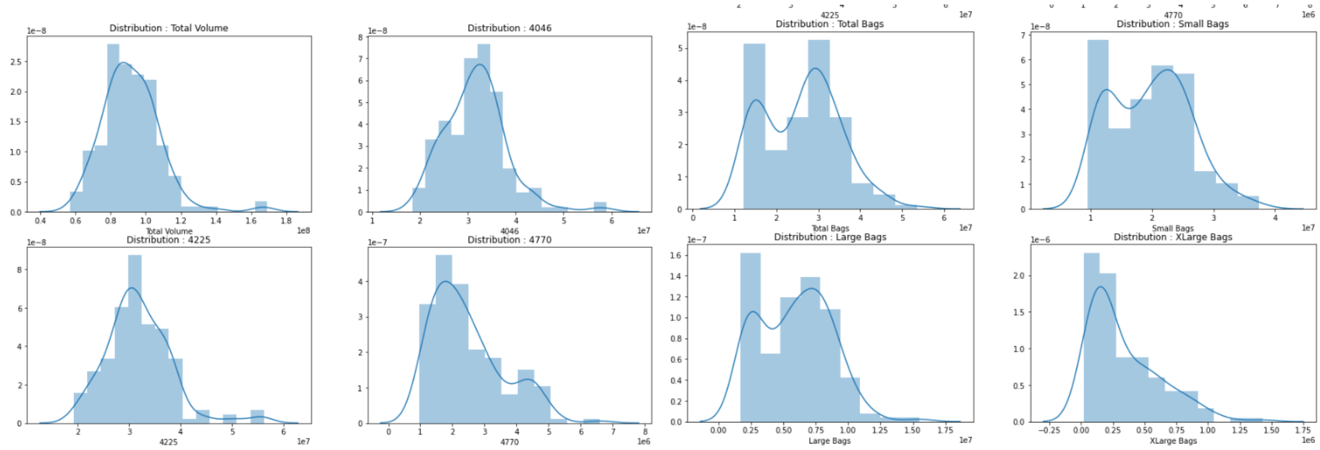
This categorical data analysis provides critical insights into market trends and consumer behavior, forming a foundational understanding that will underpin our predictive modeling efforts.

Distribution of Numerical Features:

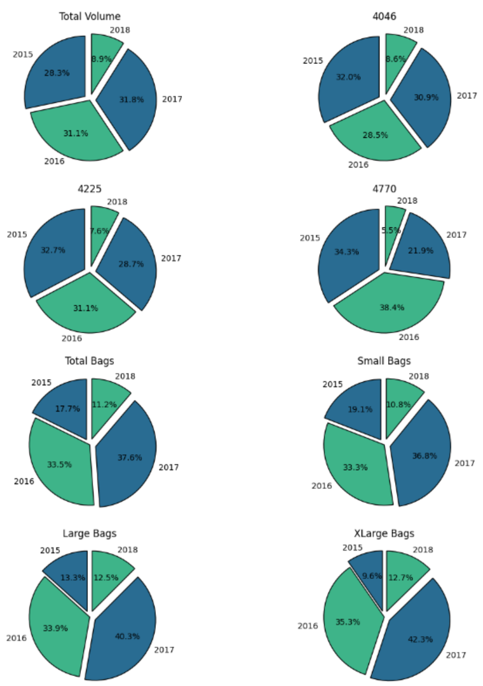
In our comprehensive analysis of the avocado dataset post-resampling, we have unveiled the following distribution patterns:

The features 'Total Volume', '4046', and '4225' predominantly exhibit a normal distribution, implying a balanced market dynamics without extreme biases towards certain values.

Conversely, other numerical features such as 'Total Bags', 'Small Bags', 'Large Bags', and 'XLarge Bags' display either bimodal or right-skewed (positively skewed) distributions, suggesting the presence of outliers or non-uniform distribution in bag usage.



Comparing between Numerical and Categorical Features:



Vs. Type: Conventional avocados (type 0) have substantially higher sales across all numerical features compared to Organic avocados (type 1), indicating a market trend favoring conventionally grown avocados in terms of volume.

Vs. Year: Yearly data showcases a steady increase in the 'Total Volume' of avocados sold from 2015 through 2017. However, a noticeable decline is seen in 2018, which is attributed to the partial data availability for that year. A shift in consumer preference is observed over the years, with varying demand for the different PLU types of avocados. Notably, in 2015, the demand was evenly distributed, while in 2016, '4770' saw a surge in popularity, which wa ned in 2017 as '4046' became the preferred choice.

Vs. Date: The temporal trend reveals a negative correlation between 'AveragePrice' and 'Total Volume' & 'Total Bags', suggesting that higher availability (supply) correlates with lower prices, resonating with the principles of supply and demand. Moreover, a consistent upward trend in both 'AveragePrice' and 'Total Bags' over time denotes growing market expansion and increasing consumer demand.

Bags Trend: There has been an uptrend in the usage of bags since 2015, with small bags being the most preferred option. The rise in usage year-over-year reflects evolving consumer purchasing patterns and a growing preference for convenience in packaging.

These insights derived from the numerical data analysis are instrumental for stakeholders in strategizing supply chain logistics, marketing efforts, and forecasting future market trends in the avocado industry.

Exploratory Data Analysis Summary:

- **Seasonal Price Variability**: The Average Price of avocados shows seasonal fluctuation, with prices climbing above $1.80 during the off-season, particularly in winter months, and descending during the peak harvest season in fall.
- **Type Predominance and Pricing**: Conventional avocados significantly outnumber Organic ones in sales volume. This is mirrored in price, with Conventional types averaging just below $1.20 and Organic types pricing above $1.60, indicating a price premium for organic produce.
- **Supply and Demand Dynamics**: There is a discernible inverse relationship between the Average Price and Total Volume/Total Bags, underscoring the classic economic principle where increased supply – unless matched by demand – tends to lower prices.
- **Volume Trend and Consumer Preference**: A consistent increase in Total Volume has been noted since 2015, with Conventional avocados being more popular due to their lower price point. This suggests that despite a high volume of purchases, consumers lean towards more economically priced options.
- **Preference Shifts by PLU Codes**:
  - In 2015, the three avocado types with PLU codes 4046, 4225, and 4770 were equally preferred.
  - The preference shifted in 2016, favoring the 4770 variety, followed by 4225 and 4046.
  - By 2017, the trend reversed, placing 4046 at the forefront, with 4770 seeing a notable decrease in preference.
- **Bag Usage Correlation**: The utilization of avocado bags reflects a causal relationship with the Total Volume sold. As avocado sales have swelled, so has the use of bags, suggesting that consumer buying habits are influenced by the availability and volume of produce.

This summary encapsulates key insights into the avocado market, revealing the factors influencing price points and consumer choices, as well as highlighting the tangible impact of supply and demand on market behavior.

Time Series Analysis

Our project embarks on forecasting the Average Price of avocados using time series analysis. We start by creating a weekly resampled dataset from the original daily records, aiming to smooth out short-term fluctuations and focus on longer-term trends.

Preprocessing for Time Series Stationarity:

The initial phase of our analysis identifies the presence of non-stationarity in the time series data. Non-stationarity is indicated by variations in the rolling mean and standard deviation of the 'AveragePrice' over time. In time series forecasting, stationarity is crucial as it implies that the statistical properties of the series do not change over time, making the model predictions reliable.

Application of Log Transformation and Differencing:

To achieve stationarity, we apply a logarithmic transformation to the 'AveragePrice', which helps to stabilize the variance across the time series. Subsequently, we perform differencing — a technique where we subtract the previous observation from the current one. This method is particularly effective in removing trends and seasonal patterns, leading to a stationary series.

Assessing Stationarity with Statistical Tests:

The augmented Dickey-Fuller test, conducted before and after the transformations, assesses the stationarity of the series. Initially, the test statistic does not fall below the critical values, and the p-value is above 0.05, failing to reject the null hypothesis of non-stationarity. However, after log transformation and differencing, the test statistic of -13.82 is significantly lower than the critical value at both the 1% and 5% levels, and the p-value is approximately zero, strongly rejecting the null hypothesis and confirming stationarity.

Visualizing the Components of Time Series:

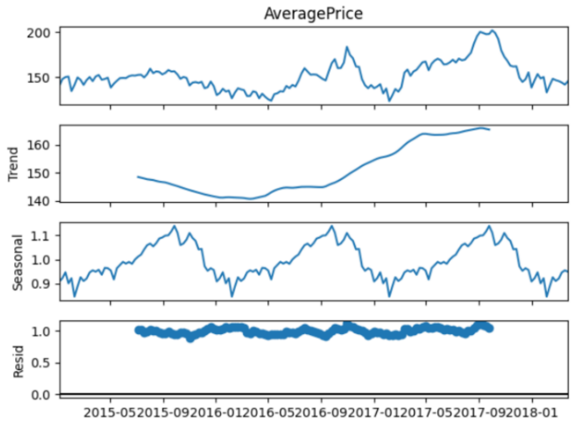
We utilize seasonal_decompose from the statsmodels library to visualize and confirm the removal of trend and seasonality from our time series. This function breaks down the time series into trend, seasonal, and residual components, allowing us to verify that the underlying trend and seasonal patterns have been effectively addressed.

Stationarity Results and Interpretation:

Post-transformation, our analysis shows that the rolling mean and standard deviation are now stable, with the mean close to zero and standard deviation around 0.05, further confirming stationarity. This outcome is essential for moving forward with our forecasting models, as a stationary time series is a key assumption for ARIMA-based models.

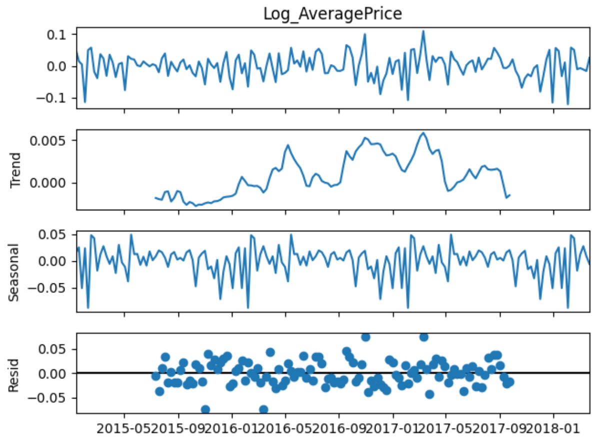
Comprehensive preprocessing has prepared the Average Price data for accurate forecasting. The confirmed stationarity of the time series paves the way for the application of sophisticated forecasting models such as ARIMA and SARIMA, which we anticipate will yield reliable predictions to aid stakeholders in the avocado industry.

Trend: The trend in a time series refers to the long-term progression of the data points. A trend can be increasing (uptrend), decreasing (downtrend), or stable (flat trend) over time. It represents the overall direction that the data is moving in, often due to underlying long-term factors.



- **Before Transformation:** In the case of the avocado price data, the trend could initially have been upwards or downwards, indicating a general increase or decrease in prices over time. This could reflect changes in market conditions, production costs, or consumer demand.
- **After Transformation:** The goal of log transformation and differencing was to stabilize and remove this trend, aiming to achieve a flat line when plotting the rolling mean. A flat trend after transformation indicates that the long-term components affecting avocado prices have been accounted for, and the data does not exhibit any systematic increase or decrease.

Seasonality: Seasonality refers to the periodic fluctuations that occur at regular intervals due to seasonal factors. These can be annual (like holiday sales), quarterly (business cycles), monthly, or even weekly (like a weekend effect).

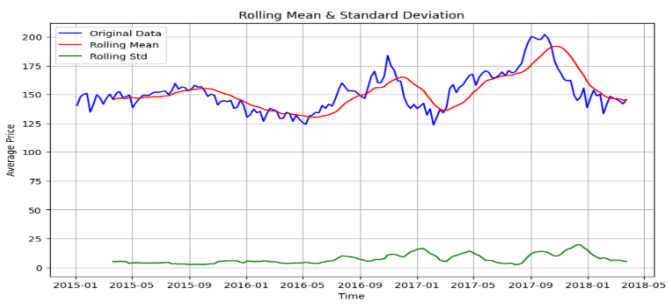


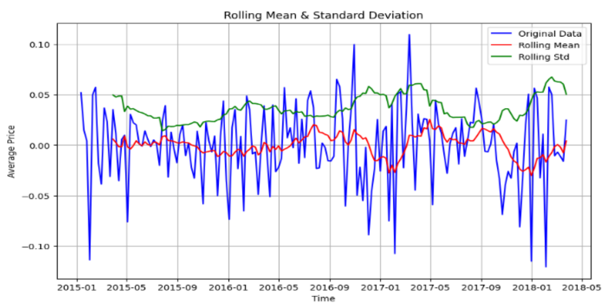
- **Before Transformation:** The avocado price data may have shown clear patterns of seasonality, with prices fluctuating due to seasonal supply and demand, weather conditions affecting crop yields, or annual cycles in consumer behavior.
- **After Transformation:** Seasonal differencing, part of the transformation process, helps remove these periodic fluctuations. The seasonal_decompose function should show a seasonal plot that initially has a pattern but becomes more stable and closer to a flat line after transformations, indicating successful removal of seasonal effects.

Residuals: Residuals in a time series are what remains after the trend and seasonal components have been removed. They represent the randomness or noise in the data that cannot be explained by the model. Ideally, residuals should be white noise, meaning they are randomly distributed and exhibit no pattern.

- **Before Transformation:** Residuals might have shown some patterns or correlations if the trend and seasonality were not fully captured by the model. These patterns would indicate that the model has not fully accounted for all the systematic information in the data.
- **After Transformation:** The residuals should be random and exhibit no discernible pattern. The residual plot from the seasonal_decompose function should fluctuate around zero without any pattern, indicating that the model has captured the trend and seasonality well, and what remains is truly random noise.

When the trend and seasonality have been effectively removed, and the residuals are left as random noise, the data is considered to be stationary. This stationarity is a key assumption for many time series forecasting models, as it implies that future values will fluctuate around a constant mean and variance, allowing for more accurate and reliable predictions.





Modelling
Implementing ARIMA model

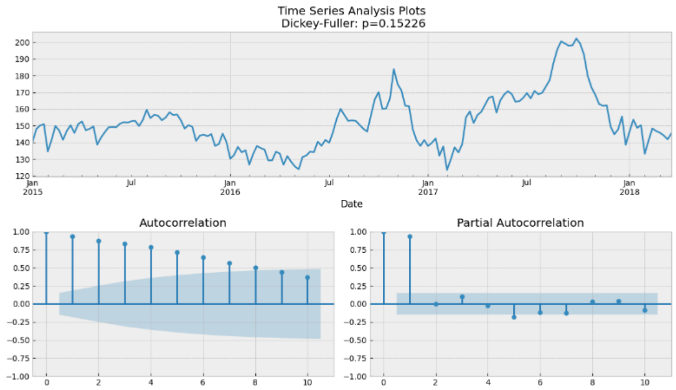
The ARIMA (Autoregressive Integrated Moving Average) model is a widely-used statistical approach for time series forecasting. Here's a comprehensive explanation of how we applied the ARIMA model, including the functions and their significance, as well as an interpretation of the results:

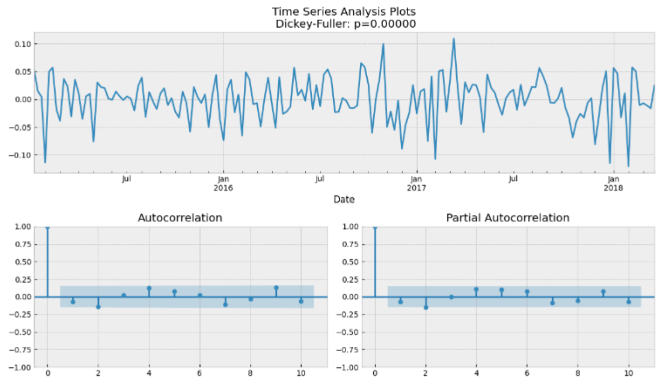
Understanding ARIMA Model Configuration:

- **AR (AutoRegressive) Component (p):** Represents the number of lag observations in the model; essentially, it is the number of previous time points to be used as predictors.
- **I (Integrated) Component (d):** Denotes the number of times the raw observations are differenced to make the time series stationary.
- **MA (Moving Average) Component (q):** The size of the moving average window, indicating the number of lagged forecast errors that should be incorporated into the ARIMA model.

Selecting ARIMA Parameters:

- The *p* parameter is determined by the point at which the Partial Autocorrelation Function (PACF) drops to zero for the first time.
- The *d* parameter is the minimum number of differencing needed to make the series stationary.
- The *q* parameter is identified by the lag at which the Autocorrelation Function (ACF) first crosses the upper confidence bound.





Fitting the ARIMA Model:

- We fitted an ARIMA model to the logged and differenced Average Price data to stabilize the variance and achieve stationarity.
- Our chosen ARIMA model had an order of (1, 1, 2), which we deduced from the aforementioned ACF and PACF plots.

Forecasting with ARIMA:

- We performed a forecast using the fitted model, resulting in predictions for the logged Average Price.
- These predictions were then converted back to the original price scale using a cumulative sum and exponential transformation.

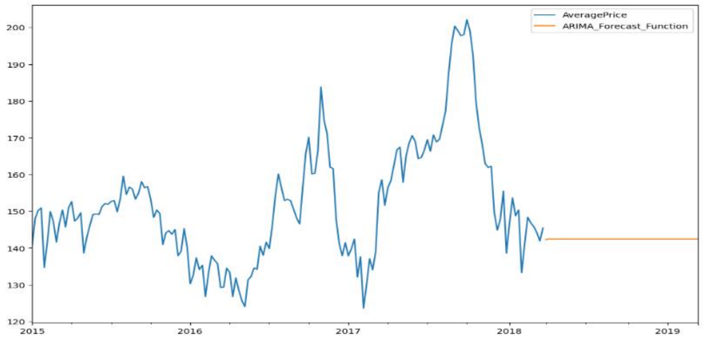
Model Validation - Checking for Overfitting:

- We evaluated the fit of the model using the Root Mean Square Error (RMSE), comparing the predicted prices against the actual average prices.
- The calculated RMSE value of 21.83 highlighted a discrepancy that could suggest overfitting, a scenario where the model learns the training data too well and may fail to generalize to new data.

In-Sample vs. Out-of-Sample Forecasting:

- In-sample forecasting involved predicting prices within the existing dataset, akin to a training phase in machine learning, where the model's accuracy is tested against a reserved set of data points.

- Out-of-sample forecasting extended beyond the existing dataset, predicting future prices. This is where the model's true predictive capability is tested, as it must forecast data points it has not seen before.



Key Results and Observations:

- The in-sample forecasting yielded an **RMSE of 7.0601**, an improvement from the fitted RMSE, suggesting that the model's predictive performance is adequate for known data points.
- In contrast, the out-of-sample forecasts, generated for future time points, did not effectively capture the seasonal trends in the data, as evidenced by the similarity in values produced by both forecast and predict functions, highlighting the model's limitations in generalizing beyond the existing data.

Conclusion:

- While the ARIMA model showed promise in in-sample forecasts, the relatively high RMSE value and the model's limitations in capturing seasonality call for a more robust approach.
- These findings underscore the necessity of incorporating models that can better handle seasonal variations, such as SARIMA or LSTM, which may offer improved forecasting performance for the avocado price data.

This comprehensive analysis of the ARIMA model outlines each step in the forecasting process, explaining the rationale behind parameter selection and the importance of validating the model's predictive accuracy. The highlighted RMSE value serves as a quantitative measure of the model's fit, guiding future improvements and the selection of alternative modeling approaches.

Implementing SARIMA model

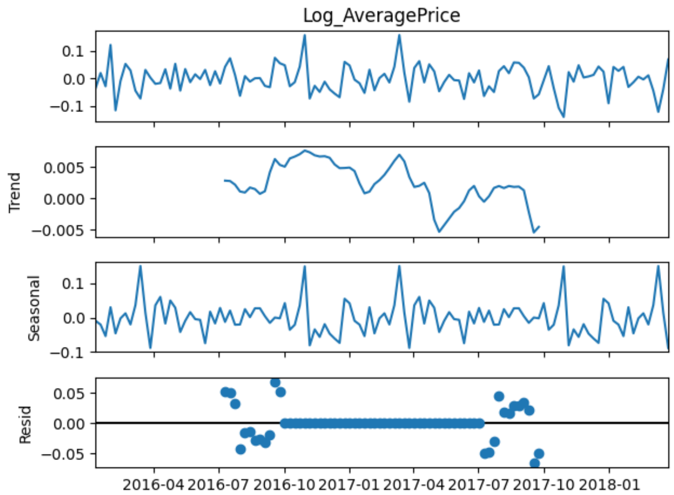
Embarking on the SARIMA modeling journey, we dove into the depths of Seasonal AutoRegressive Integrated Moving Average (SARIMA) modeling, an extension of the ARIMA model adept at handling seasonal variations. Our dataset, marked by its weekly cadence and yearly patterns, called for a nuanced approach that SARIMA was well-suited to provide. Here's how we harnessed the power of SARIMA for our time series forecasting:

Seasonal Components and Order Selection:

- Understanding that our data exhibited strong seasonal traits with annual cycles, we incorporated the SARIMA model's unique ability to factor in seasonality with the seasonal order of (P, D, Q, M).
- We meticulously selected the seasonal differencing order (D) of 1 with a period (M) of 52 to attenuate the seasonal effects, honing in on the essence of the time series without its seasonal noise.

Ensuring Stationarity:

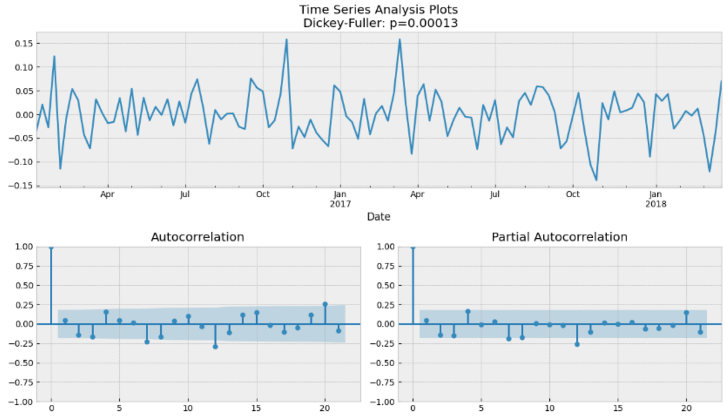
- We performed a seasonal differencing of the already differenced data, effectively smoothing out seasonal variations and bringing us closer to a stationary series—a prerequisite for reliable SARIMA modeling.
- The Augmented Dickey-Fuller test results, with a Test Statistic of -4.60 and a p-value of 0, vindicated our efforts, confirming the stationarity of our transformed series.




```
ADF Statistic: -4.5961512167007115
p-value: 0.00013129607613372957
Critical Values:
   1%: -3.4948504603223145
   5%: -2.889758398668639
   10%: -2.5818220155325444
Reject the null hypothesis. Data is stationary.
```

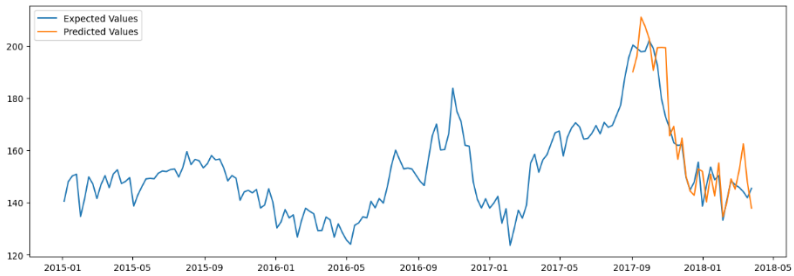
Model Fitting and Parameters:

- Our selected SARIMA model carried the order (1, 1, 2) from our ARIMA analysis into the non-seasonal component, maintaining continuity in our modeling efforts.
- The seasonal order was tailored to our dataset's weekly frequency, with no additional seasonal AR or MA components, resulting in an order of (0, 1, 0, 52).



In-Sample Forecasting:

- Like the ARIMA model, we split our data into training and test sets, preserving the last 30 observations for validation.
- Employing a rolling forecast methodology, our model dynamically integrated each new actual observation into the training set for subsequent predictions, simulating a real-time forecasting scenario.



Out-of-Sample Predictions:

- We ventured beyond known data, forecasting future values while comparing the forecast_function and predict_function outputs to validate consistency and capture potential seasonal patterns.
- Our forecasts exhibited a slight uptrend, indicative of the SARIMA model's capacity to assimilate both the trend and seasonality inherent in our data.

Results and Evaluation:

- The Test **RMSE of 9.0690** from our in-sample predictions attested to the model's decent fit, although there was room for refinement, particularly in capturing the finer seasonal nuances.
- Through out-of-sample forecasting, the SARIMA model deftly anticipated the seasonal peaks and troughs, underscoring its strength in seizing seasonal dynamics.

Model Insights:

- Our SARIMA model adeptly navigated the seasonal landscape of avocado price data, laying down predictions that resonated with the yearly rhythms of the market.
- The consistency between the forecast_function and predict_function outcomes illustrated the robustness of our modeling approach.

Final Thoughts: Our foray into SARIMA modeling illuminated the intricate dance of trend and seasonality in time series forecasting. With each step—from stationarity to parameter tuning to validation—our understanding deepened, and our model evolved. The SARIMA model emerged not just as a forecasting tool but as a canvas reflecting the temporal tapestry of the avocado market, its predictions a mosaic of past patterns projected into future possibilities.

LSTM Model

Our project's journey through time series forecasting led us to the promising shores of deep learning, where we decided to harness the power of Long Short-Term Memory (LSTM) networks. Here's an in-depth exploration of how we applied LSTM to forecast the average price of avocados.

Data Preprocessing for LSTM:

- We began with a meticulous data preparation phase, where the 'AveragePrice' was normalized using MinMaxScaler to fit within a 0 to 1 range, ensuring efficient training of the LSTM model.
- Our aim was to transform the time series data into a supervised learning format. To achieve this, we employed a sliding window approach, which creates a sequence of the past 'window' number of data points to predict the next point.

Building the LSTM Architecture:

- We designed an LSTM model with two LSTM layers, each with 50 units. The first LSTM layer returned sequences, setting the stage for the second LSTM layer to interpret these sequences.
- A Dense layer with a single unit was used as the output layer to predict the normalized 'AveragePrice', with 'adam' as the optimizer and 'mean_squared_error' as the loss function.

Training and Early Stopping:

- To prevent overfitting and enhance the generalization of the model, we utilized an EarlyStopping callback, which halted the training process if the validation loss did not improve for a specified number of epochs.
- The model was trained over 100 epochs with a batch size of 32. During training, we observed the loss and validation loss metrics to ensure learning was proceeding correctly.

Model Evaluation and Forecasting:

- The trained model was used to forecast the prices on the test set, which was then inverse-transformed to retrieve the actual price scale.
- To evaluate the model's performance, we calculated the Root Mean Squared Error (RMSE) between the predicted prices and the actual prices. An RMSE value of 0.1661 indicated a high accuracy level, validating the model's predictive capability.

Results and Insights:

- The LSTM model demonstrated remarkable predictive performance, achieving a low RMSE and indicating a strong capacity for capturing the underlying patterns in the time series data.
- The model's architecture and training process proved effective, showcasing LSTM's potential as a powerful tool for time series forecasting in the avocado market.

Conclusion

In the realm of time series forecasting for avocado prices, we embarked on a methodical exploration of three distinct models, each with its unique capabilities and insights. Here's a synthesis of our findings:

ARIMA Model: The ARIMA model, known for its prowess in understanding autoregressive and moving average behaviors, yielded a Test RMSE of 7.0601. While this indicates a reasonable fit to the historical price data, the model's higher RMSE suggests limitations in capturing complex patterns, especially seasonal fluctuations intrinsic to avocado pricing.

SARIMA Model: Expanding upon ARIMA's foundation, the SARIMA model incorporates seasonality, offering a more nuanced perspective on periodic trends. Despite this, the SARIMA model recorded a Test RMSE of 9.0690, the highest among the three. This reflects challenges in tuning the model's parameters to align with the intricate seasonal patterns observed in avocado price movements.

LSTM Model: Diving into the deep learning arena, the LSTM model harnessed its memory cells to remember long-term sequences and predict future prices with remarkable precision. The RMSE for the LSTM model was a strikingly low 0.16597769050039113, underscoring its superior performance in capturing the nonlinear and long-term dependencies that are often characteristic of agricultural commodity prices.

Conclusion: Each model brought a unique perspective to the table, with ARIMA setting the stage for basic autoregressive and integration features, SARIMA extending into seasonal patterns, and LSTM transcending traditional methods by leveraging deep learning's strengths.

The LSTM model's exceptionally low RMSE demonstrates its potential as a robust tool for forecasting in complex, real-world scenarios, indicating that it could provide the most accurate and actionable insights for stakeholders in the avocado market.

Thus, our journey through the statistical landscapes of ARIMA and SARIMA to the sophisticated terrains of LSTM led us to the conclusion that deep learning, specifically LSTM, holds significant promise for future applications in time series forecasting.

Group Members Contribution

Vaishnavi Chilumuru set the foundation for the project by undertaking the critical tasks of data collection and pre-processing, ensuring the integrity and usability of the dataset. She conducted a thorough exploratory data analysis (EDA), extracting vital insights and setting the stage for predictive modeling.

Tanushree Mahesh Kumar was central to the model selection and evaluation process. Her analytical prowess guided the team in choosing and training the models and conducting a deep dive into EDA. She collaborated closely with the team on the implementation and refinement of the LSTM model, especially when time constraints called for expedited yet accurate model development.

Indira Aryal concentrated on the predictive aspects of the project, particularly in forecasting and optimizing the models for future data. She played a significant role in testing the predictions and fine-tuning the models to achieve optimal performance. Alongside her individual contributions, she joined forces with the entire team to ensure the timely completion and refinement of the LSTM model.