TITLE:
REPORT ON TIME SERIES ANALYSIS ON MONTHLY SALES USING DATA FROM THE ADVANCE MONTHLY RETAIL TRADE SURVEY (MARTS) IN THE UNITED STATES OF AMERICA.

PREPARED BY: JOEBRIGHT BOAKYE-DANKYI

TABLE CONTENT.

i. INTRODUCTION

The Advance Monthly Retail Trade Survey (MARTS) provides early indicators of retail and food service sales in the United States. It covers businesses in the North American Industry Classification System (NAICS) that offer merchandise and related services to end consumers. MARTS gathers data on sales figures, reporting timelines, and the count of retail locations. MARTS estimates monthly sales in food service establishments and drinking venues.

In this study, the steps for conducting a detailed time series analysis using data from the Advance Monthly Retail Trade Survey (MARTS) were outlined. The data was examined to generate questions for exploration. Trends and seasonal patterns in retail and food service sales were identified. Following this, the remaining differences were analysed to assess the model fit. A suitable time series model, such as ARIMA was applied, and the fit was evaluated again.
Future sales were forecasted. The findings were summarised, conclusions drawn, and relevant sources referenced to aid decision-making and planning in the retail industry.
All analysis was done in R programming (Using libraries magrittr and ggplot2).

ii. ANALYSIS OUTLINE

a. EXPLORATORY DATA ANALYSIS

The sourced data from the Advanced Monthly Retail Trade Survey (MARTS) was obtained, and the analysis focused on monthly retail sales spanning from January 1992 to February 2019 (Thus 320 months). The original sales values were recorded in millions of dollars; however, for ease of manipulation in our analysis, these values were converted to billions of dollars equivalent. The following table (*Table 1.0*) presents initial summary statistics derived from the data.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
146.4	244.4	328.5	326.6	397.6	558.0

Table 1.0

RESEARCH QUESTIONS

The following research questions were developed based on the initial explorations done on the data and the report will assist in answering all these questions.

1. How have trends in retail and food service sales evolved according to the MARTS dataset, and what are the implications for future consumer demand?
2. Are there any distinct seasonal patterns observed in retail and food service sales data collected from MARTS, and how do these patterns influence sales strategies and inventory management?
3. Can a robust time series forecasting model be developed using MARTS data to predict future sales accurately, and what factors should be considered to improve the accuracy of these forecasts?

b. TREND ANALYSIS

To find out the trend analysis for our data, regression models were used. After considering different regression models, including linear, quadratic, and cubic. The linear model was determined to be the best-fitting model based on the lowest Akaike Information Criterion (AIC) value among all evaluated models.

Table 1.1 below shows all the AIC values obtained for each model.

Model Name	AIC Value Obtained
Linear	3046.326
Quadratic	3048.295
Cubic	3048.295

Table 1.1

The best-fitted model was concluded to be the linear Regression as displayed in *Fig. 1.0*

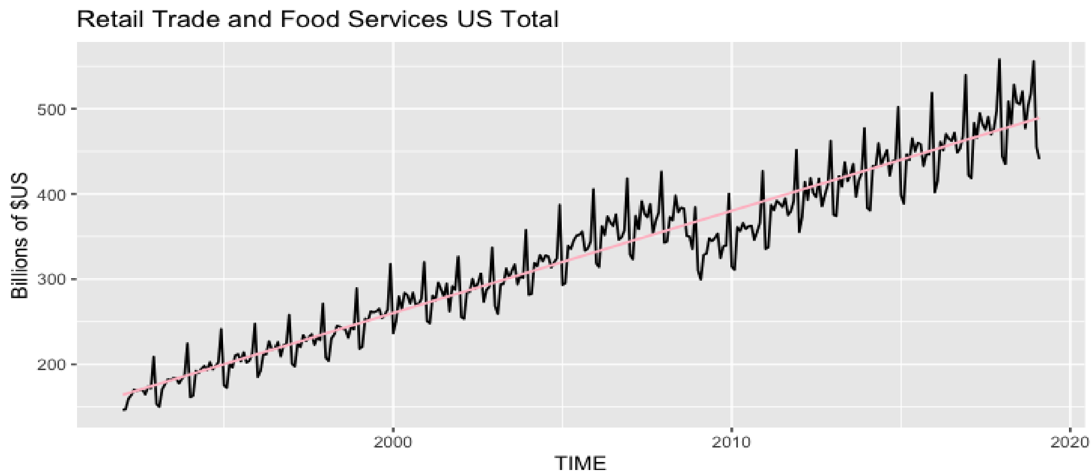

Fig. 1.0

The trend analysis on this linear model *(Fig. 1.0*) shows a steady rise in retail trade and food services sales in the United States over the observed period. The pink linear regression line illustrates this trend, indicating a consistent increase in sales value over time. Despite minor fluctuations, the overall pattern suggests continuous growth in the retail and food services sector.

c. <u>SEASONALITY ANALYSIS</u>

The general seasonality of the Retail Trade and Food Services appears to exhibit a cyclical pattern with fluctuations recurring at regular intervals over the observed period. The plot (*Fig 1.1*) shows periodic peaks and troughs, indicating periods of increased and decreased activity, respectively. This suggests that there are distinct seasonal trends influencing retail and food services sales throughout the time frame depicted. Despite some irregularities, the overall pattern suggests a recurring cycle in sales activity, which could be attributed to factors such as seasonal holidays, economic cycles, and consumer behaviour trends.

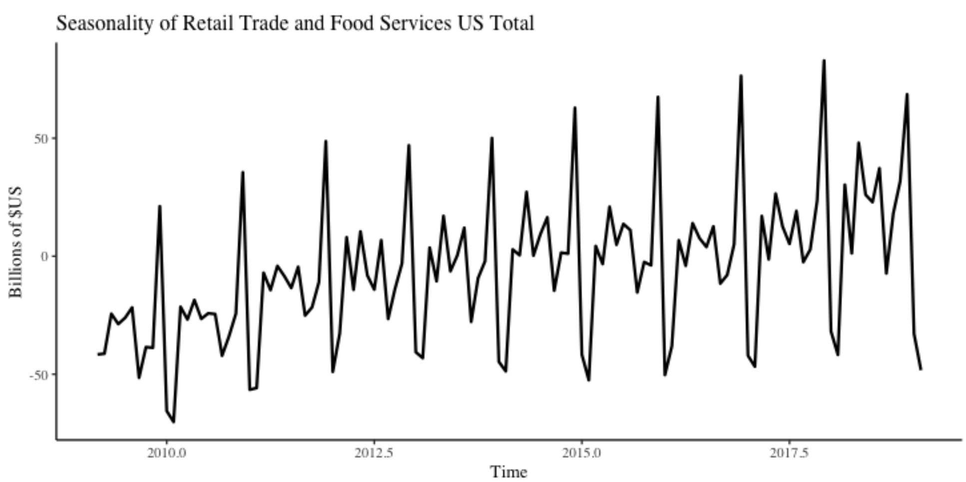

Fig. 1.1

To determine the best-fitted seasonality of the data, the COSINE-SINE model was used to fit the seasonality of the data. Values ranging from 1 to 6 were used with corresponding AIC computed, the least value of AIC was chosen as the best-fitted seasonality model.

Table 1.2 below shows all the AIC values obtained for the 6 models.

Model value	AIC Values for Seasonality
1	3033.56
2	2973.87
3	2878.096
4	2812.381
5	2705.518
6	2684.944

Table 1.2

The best-fitted seasonality model was with a value 6 as displayed in *Fig.1.2* below.

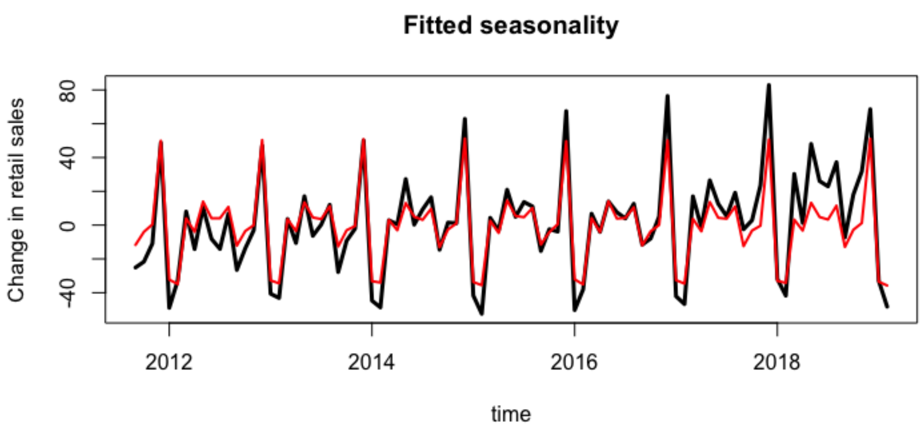

Fig. 1.2

The fitted seasonality plot (*Fig.1.2*) depicts the seasonal patterns in changes in retail sales over time. The red line represents the fitted seasonal component, showing the expected fluctuations in sales over the observed period. The peaks and troughs in the red line indicate periods of increased and decreased sales activity, respectively. The black line represents the observed changes in retail sales, which align closely with the fitted seasonal pattern. This suggests that the model has effectively captured the seasonal variations in retail sales, allowing for a better understanding and forecasting of sales trends over time.

d. RESIDUAL ANALYSIS

In Autoregressive Integrated Moving Average (ARIMA) analysis, assessing residuals is crucial for evaluating the model's fit. Residuals represent the differences between observed and predicted values. Analysing residuals helps detect any patterns or errors not accounted for by the model, ensuring it captures data patterns effectively. This process involves checking if residuals are randomly distributed around zero, indicating a good fit. Diagnostic plots like autocorrelation and partial autocorrelation plots are also used to identify any remaining correlations.

ACF AND PACF PLOTS BEFORE ARIMA

A close observation of the ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) plots (*Fig. 1.3 and Fig. 1.4 below*) of residuals show significant autocorrelation at any lag. In the ACF plot, all autocorrelation values fall above the dashed lines, indicating that they are statistically different from zero. In the PACF plot, most partial autocorrelation values also fall outside the blue dashed lines, indicating significant partial autocorrelation at some lags.

These observations suggest that there are remaining patterns or correlations in the residuals that need to be addressed, further confirming why we need to apply the ARIMA model in capturing the underlying data patterns.

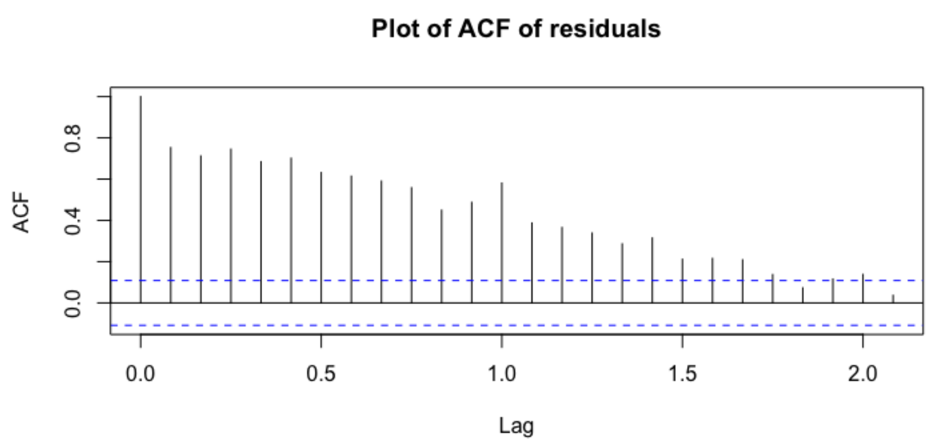

Fig. 1.3

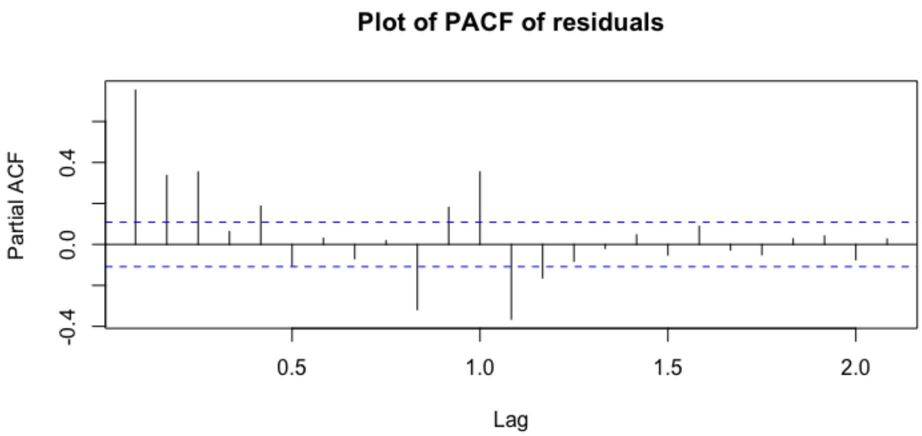

Fig. 1.4

USING ARIMA

Considering the AIC values obtained in *Table 1.3* below, the appropriate model corresponding to the least AIC is ARIMA (2,0,3).

p,q	0	1	2	3
0	2662.885	2504.820	2479.496	2407.930
1	2390.513	2317.939	2319.590	2304.489
2	2353.099	2319.769	2307.359	**2300.247**
3	2313.075	2317.630	2302.070	2300.300

Table 1.3

From *Fig. 1.5* below. The plot also shows the differences between the observed values and the values predicted by the ARIMA model over time. The residuals appear to fluctuate randomly around zero, indicating that the model adequately captures the variability in the data. There are no discernible patterns or trends in the residuals, suggesting that the ARIMA model effectively accounts for the underlying data patterns. Overall, the plot demonstrates that the residuals meet the assumption of randomness, homoscedasticity and constant variation validating the goodness-of-fit of the ARIMA (2,0,3) model to the data.

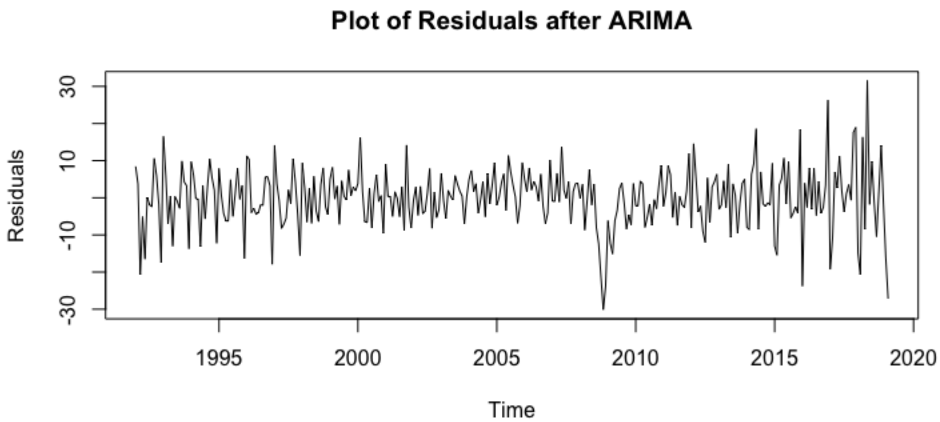

Fig. 1.5

PACF AND ACF PLOTS AFTER ARIMA

Both the PACF (Partial Auto Correlation Function) and ACF (Auto Correlation Function) plots (*Fig. 1.6* and *Fig.1.7* below) after using the best ARIMA model (ARIMA (2,0,3)) indicate no significant autocorrelation at any lag in the series. In the PACF plot, all partial autocorrelation values fall within the blue dashed lines, suggesting no significant partial autocorrelation at any lag. In the ACF plot, all autocorrelation values also fall within the blue dashed lines, indicating no significant autocorrelation at any lag. These observations confirm that the best ARIMA model adequately captures the underlying data patterns, as there are no remaining correlations or patterns in the residuals confirmed in *Fig.1.5* above.

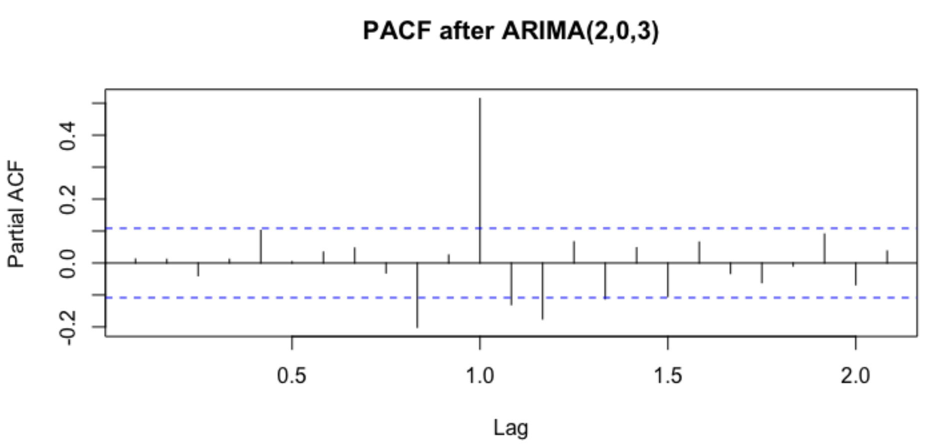

Fig. 1.6

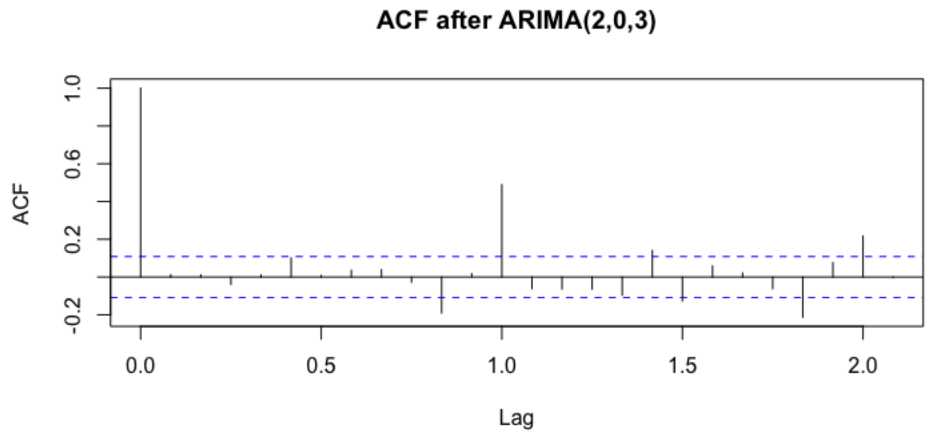

Fig. 1.7

iii. FORECASTING

Forecasting in time series analysis is about predicting future data points using past information. It includes recognizing patterns, trends, and seasonal changes in the data to make educated guesses about what will happen next. Forecasting methods usually involve using statistics, machine learning, or both to estimate future values based on what has already occurred. The aim is to offer useful insights for decision-making, helping businesses and organizations plan, manage resources, and reduce risks.

FORECAST – FUTURE VALUES

The forecast in *Fig.1.8* below shows the predicted values (in red) for retail trade and food services in the US Total, considering a well-fitted linear trend, properly fitted seasonality, and accurately applied ARIMA to the residuals. The black line represents the actual values of retail sales over time. The forecasted values closely follow the observed trend and seasonal patterns, capturing the general fluctuations and peaks in sales activity. This suggests that the model effectively incorporates both the underlying trend and seasonality, along with addressing any remaining patterns in the residuals, to generate accurate predictions of future sales. Overall, the forecast provides valuable insights for decision-making in the retail and food services sector, enabling businesses to plan, allocate resources, and strategize effectively.

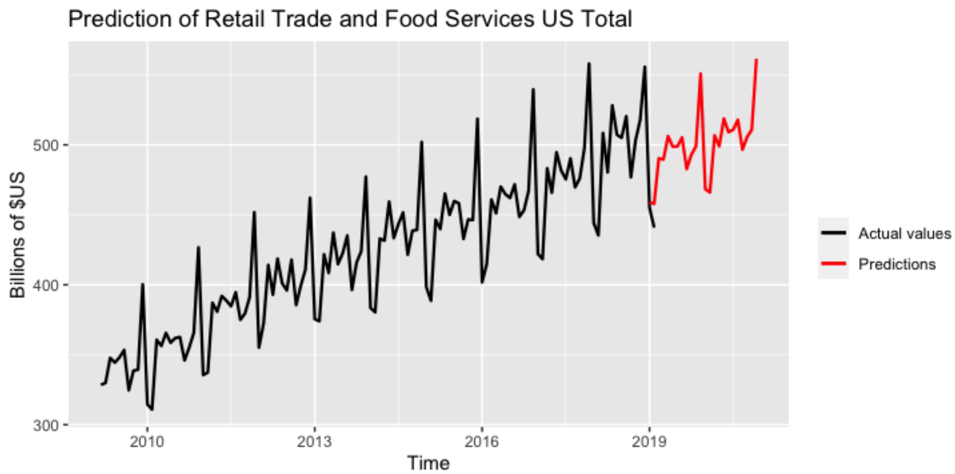

Fig. 1.8

<u>FORECAST FOR EXISTING MONTHLY VALUES FOR THE LAST TWO YEARS OF DATA.</u>

The prediction depicted in red (*Fig.1.9* below), closely follows the observed monthly values of existing retail trade and food services in the United States. It captured the overall trend and fluctuations in the data, indicating a reasonable level of predictive accuracy. There are instances where the predicted values deviate slightly from the observed ones, suggesting areas where the model may need refinement.

Also, *Fig.1.9* shows good alignment with the observed monthly values of the last 2 years of the data (in black). To evaluate its accuracy quantitatively, a comparison of the figures was done using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The RMSE value was 422.038, MSE was 418.395, and MAPE was 483.159. These metrics provided insights into the predictive performance of the model, with lower values indicating better accuracy.

The prediction provided a useful estimate of previous values based on historical trends and patterns in the data.

Fig. 1.9

iv. CONCLUSION

In conclusion, the comprehensive analysis of retail and food service sales data from the Advance Monthly Retail Trade Survey (MARTS) has provided crucial insights into the retail industry. Through thorough exploratory data analysis, trend analysis, seasonality analysis, and ARIMA modelling, significant patterns and trends were uncovered within the dataset.

The linear trend model emerged as the most suitable representation of the underlying sales trend, illustrating a consistent upward trajectory over the observed period. This model enables stakeholders to grasp the long-term market dynamics, aiding in strategic decision-making.

Also, the identification of a six-seasonality model through seasonality analysis offers a nuanced understanding of recurring patterns in retail and food service sales. This model effectively captures cyclicality, including fluctuations to be driven by seasonal factors like holidays and economic cycles. By integrating seasonality into the analysis, businesses can tailor their strategies to capitalize on peak sales periods and navigate challenges during slower seasons.

Again, the application of the ARIMA (2,0,3) model to address residual variations has proved highly effective in accurately forecasting future sales. The ARIMA model's ability to account for residual patterns ensures that forecasted values closely align with observed trends and seasonal fluctuations, providing stakeholders with reliable insights for strategic planning and risk management. While minor deviations between forecasted and observed values indicate areas for potential model refinement, continuous monitoring and refinement will be essential to adapt to evolving market conditions and enhance predictive accuracy over time. One can confidently conclude that all the research questions raised have been curatively answered by our analysis.

v. REFERENCES

The Advance Monthly Retail Trade Survey (MARTS) data available at: https://www.census.gov/econ_export/?format=xls&mode=report&default=&errormode=Dep&charttype=&chartmode=&chartadjn=&submit=GET+DATA&program=MARTS&startYear=1992&endYear=2024&categories%5B0%5D=44X72&dataType=SM&geoLevel=US&adjusted=false&notAdjusted=true&errorData=false&vert=1 (accessed: 22 March 2024).

https://www.census.gov/econ/currentdata/?programCode=MARTS&startYear=1992&endYear=2024&categories[]=44X72&dataType=SM&geoLevel=US&adjusted=1&notAdjusted=1&errorData=0#table-results (accessed: 22 March 2024).