Time Series Analysis Final Project

Fashion Trend Forecasting

Team 12 Member:	Jiaying (Grace) Zheng
	Hazel Huang
	Shuhan (Judy) Wang
	Yu-Ning Yu
Class:	Applied Time Series Analysis for Forecasting
Section:	MW 11:00 am

Executive Summary

This paper investigates the cyclical nature of fashion trends through an analysis aimed at predicting the resurgence of a specific trend, the leather jacket, using Google Trends information as time series data. Grounded in the idea that past trends may resurface cyclically in the fashion world, this study seeks to investigate whether fashion trends follow predictable cycles or arise randomly. Incorporating diverse predictors such as temperature, Consumer Confidence Index (CCI), unemployment rate, and precipitation, the analysis aims to forecast future fashion cycles.

The study's methodology involves model fitting and assessment, exploring various forecasting models such as Seasonal Naive Forecasting, Holt-Winter's Exponential Smoothing, ARIMA, and SARIMA. Following a thorough evaluation, the SARIMA(1, 0, 1)x(1, 1, 1)12 model emerged as the most promising performer, exhibiting the least Mean Absolute Percentage Error (MAPE) at 11.42% for training and 10.68% for testing. With consideration of external factors, we have also exclusively utilized temperature and precipitation as independent variables to forecast the future value of 2023 Nov - 2024 Oct.

Leveraging this SARIMA model, the study presents forecasts for the anticipated popularity trend of leather jackets in 2024 as valuable insights to fashion enthusiasts and industry brands. The demand for these jackets is forecasted to experience a significant surge during the winter months, particularly from November 2023 to January 2024. Furthermore, strategic initiatives are outlined for companies and fashion influencers in the fashion/retail sector, aligning their strategies with forecasted trends to capitalize on emerging opportunities and maximize revenue.

Introduction

In the world of fashion, there is a phrase "What goes around comes around". Styles that were popular in the 1990s might resurface in the 2020s, such as bike shorts. Leveraging Google Trends data, we aim to explore whether fashion is a cyclical phenomenon in the public's perception or merely a random occurrence. While fashion is a broad concept, this project focuses on specific fashion trends to yield meaningful results. We aim to predict the return of a particular fashion trend, leather jacket using time series data, benefiting fashion enthusiasts and brands. Additionally, we intend to incorporate pertinent data from various predictors, such as temperature, CCI (Consumer Confidence Index), unemployment rate, and precipitation, to authenticate their contributions to forecasting the trend of leather jackets.

Business Problem Statement: This analysis investigates the cyclical nature of specific fashion trends and aims to forecast their future cycles to benefit fashion enthusiasts and brands.

Objective of the analysis: By predicting future fashion cycles, our analysis intends to provide value to stakeholders in the fashion industry and fashion followers. For fashion brands and buyers seeking to maintain their position in the industry or improve their bottom line, our analysis offers insights into the potential resurgence of specific fashion trends. Armed with this information, they can proactively prepare in various business aspects to outpace their competition when they understand their customers' desires. Design, supply chain, marketing, and event planning teams can all have ample time to prepare products before the new trend reemerges. Additionally, major e-commerce platforms like Amazon, which heavily relies on data for personalization and predicting future customer interests, can use this information to prepare in advance and introduce the latest trends to their users. Lastly, fashion enthusiasts and influencers looking to stay ahead of their peers can gain insights into upcoming trends, allowing them to showcase these trends before they become widely recognized in the fashion world.

Data Description

Data Source: We believe search volume at Google is a good proxy to evaluate popularity. To capture fashion trends, we extracted data from the Google Trend website.

Data Description and Frequency: The main variable in this study is the search interest of the leather jacket on the Google website in the Los Angeles area. The definition of search interest is the popularity of the search term. A value of 100 is the peak popularity of the leather jacket. A value of 50 means half of the popularity.

We chose monthly data from 2004 to 2023 Oct. From the plot (Exhibit 1), we observed there was level, seasonality, upward trend, and noise.

Methodology

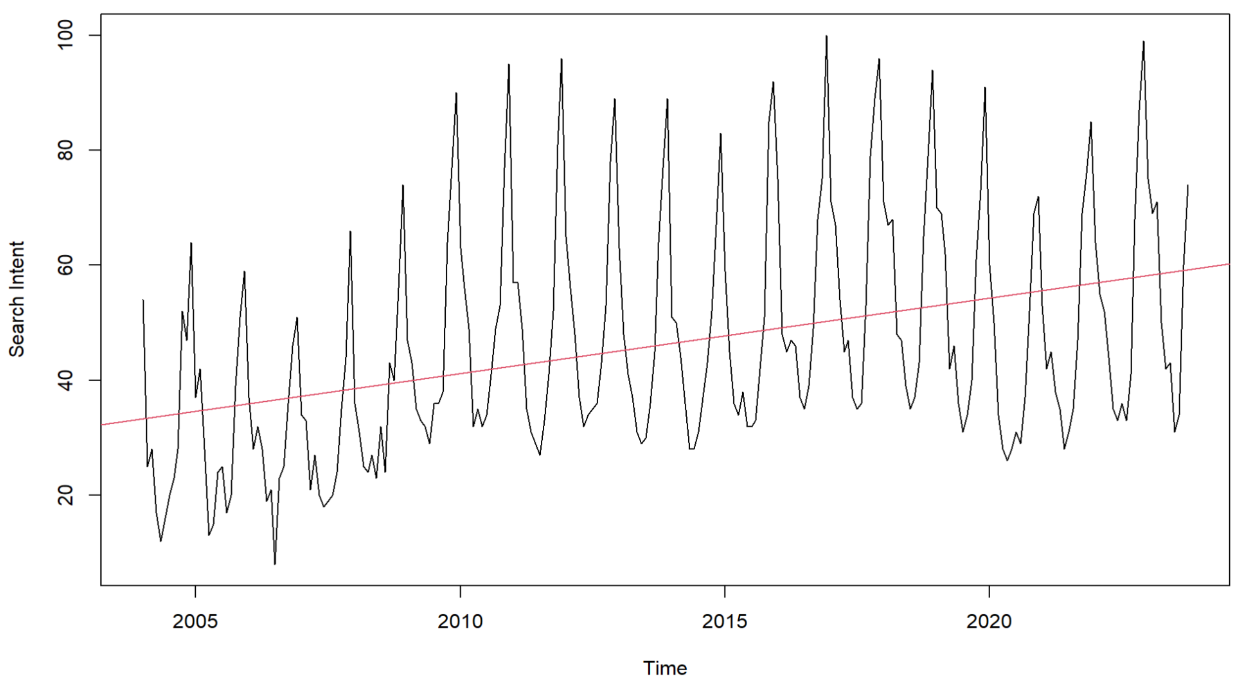
Data Preprocessing: After collecting the historical data, we performed classical decomposition to understand the components in the time series data. From the Linear Regression model and Seasonal Linear regression model, we observed an upward trend and seasonality.
Model Fitting: We partitioned our data into training and validation sets by selecting 2004 to 2021 as the training period, and 2021-2023 October as the testing period. Since data shows trend and seasonality, we explored models, including Seasonal Naive Forecasting, Holt Winter's Exponential Smoothing, ARIMA, and SARIMA. After that, we fit models with validation sets and chose the models with the least residuals.
Forecasting: Considering the external factors may influence our dependent variables, we incorporated the SARIMA model with external variables, such as temperature and precipitation, to forecast the future value of 2023 November - 2024 October.

Model Results

Classical Decomposition

I. Linear Regression Model

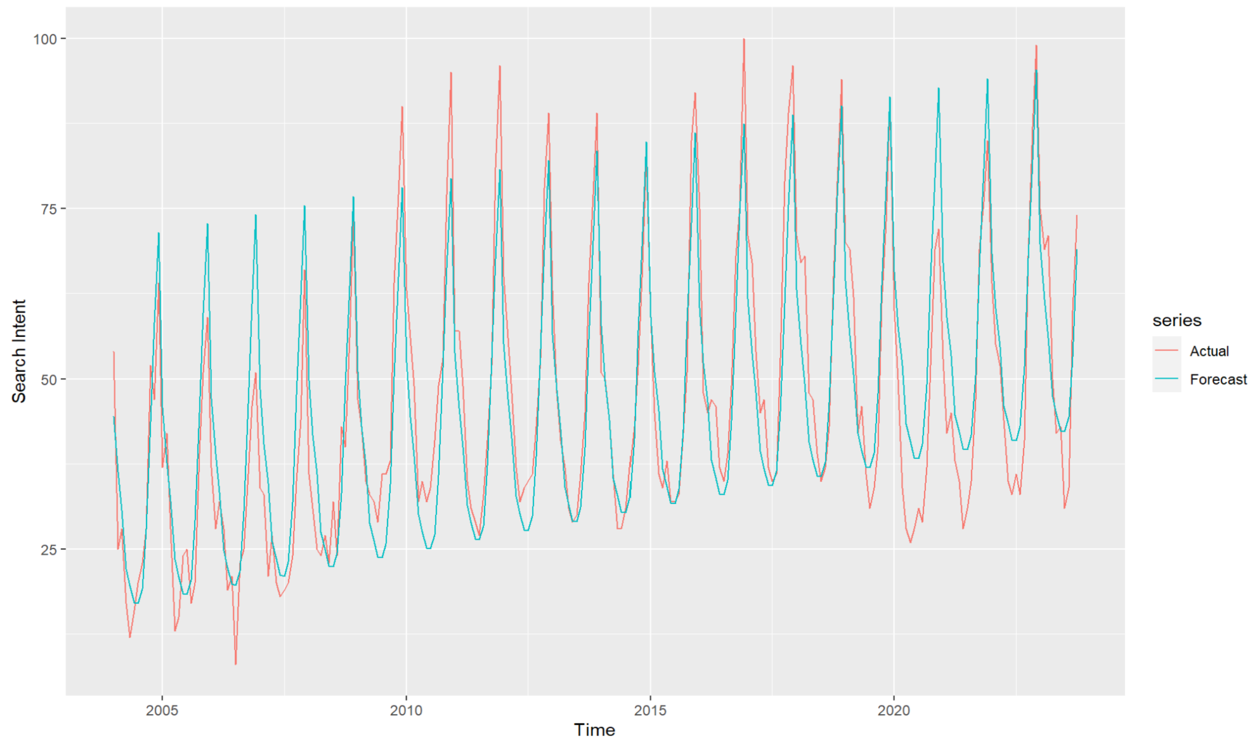
We apply linear regression to the data, which shows an upward trend.



Linear Regression Model

II. Seasonal Linear Regression Model

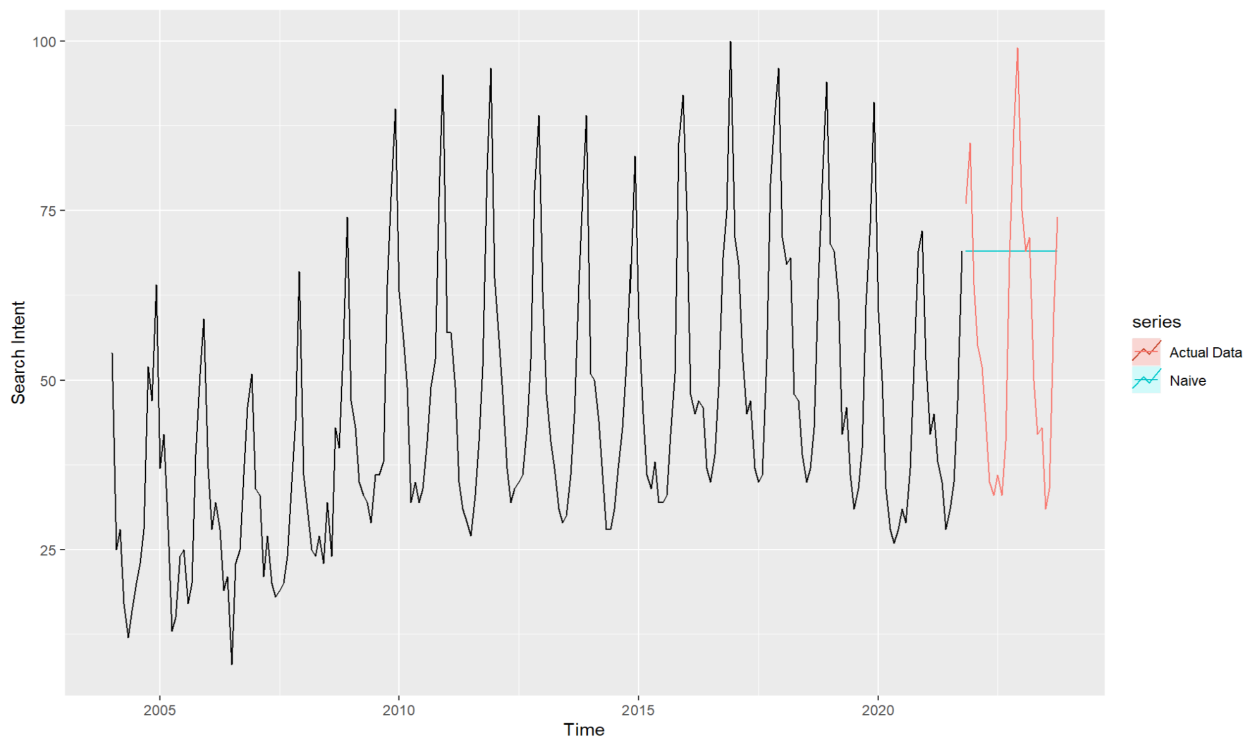
Since there is an obvious seasonality presented in the data, we employ seasonal classical decomposition to capture the pattern in the data. The adjusted R-squared is 0.8224.



Seasonal Linear Regression Model

Naive Forecasting

For Naive Forecasting, there are patterns existing in the residual plot. The RMSE for the training set is 13.01, and 23.15 for testing. Meanwhile, the MAPE is 22.85 for training and 45.81 for testing.

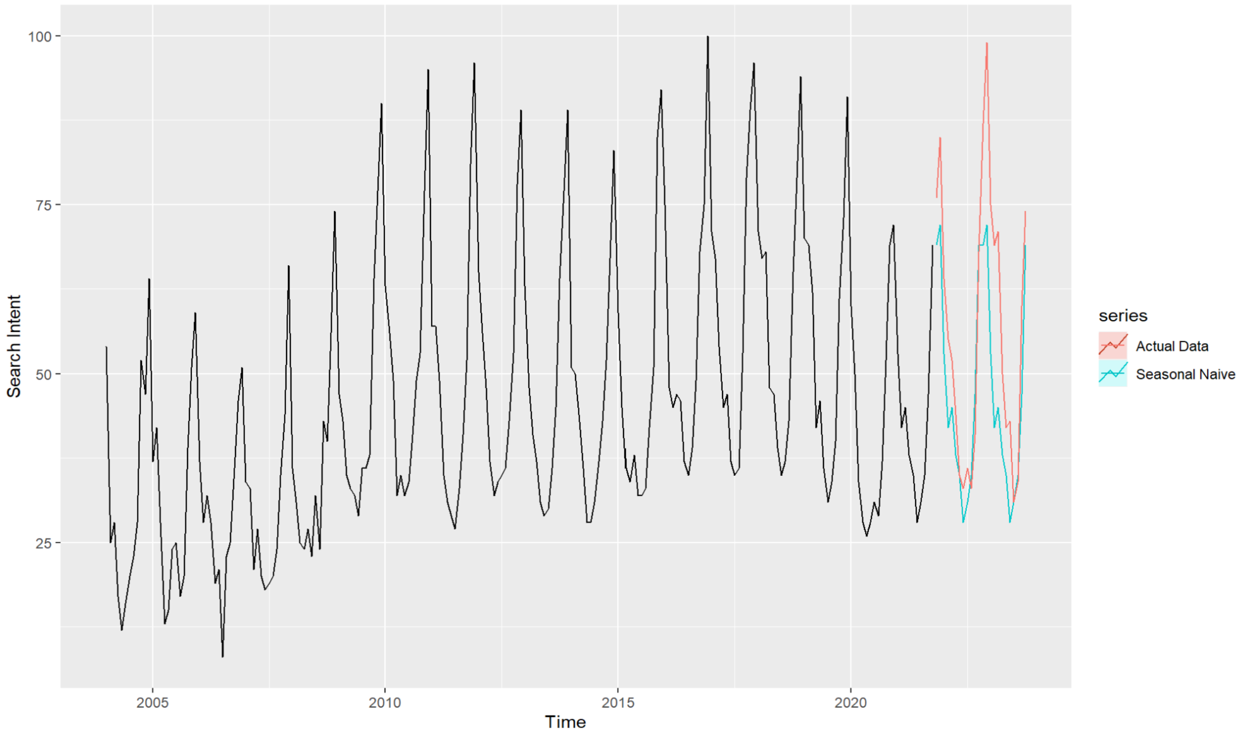


Naive Forecasting Model

Seasonal Naive Forecasting

For Seasonal Naive Forecasting, the RMSE for the training set is 8.38, and for testing, it is 13.29. Meanwhile, the MAPE is 15.97 for training and 16.82 for testing. When we check the residuals of this model, there are still some patterns that exist in the ACF plot.

In the graph below, it struggles to capture the spikes during winter.

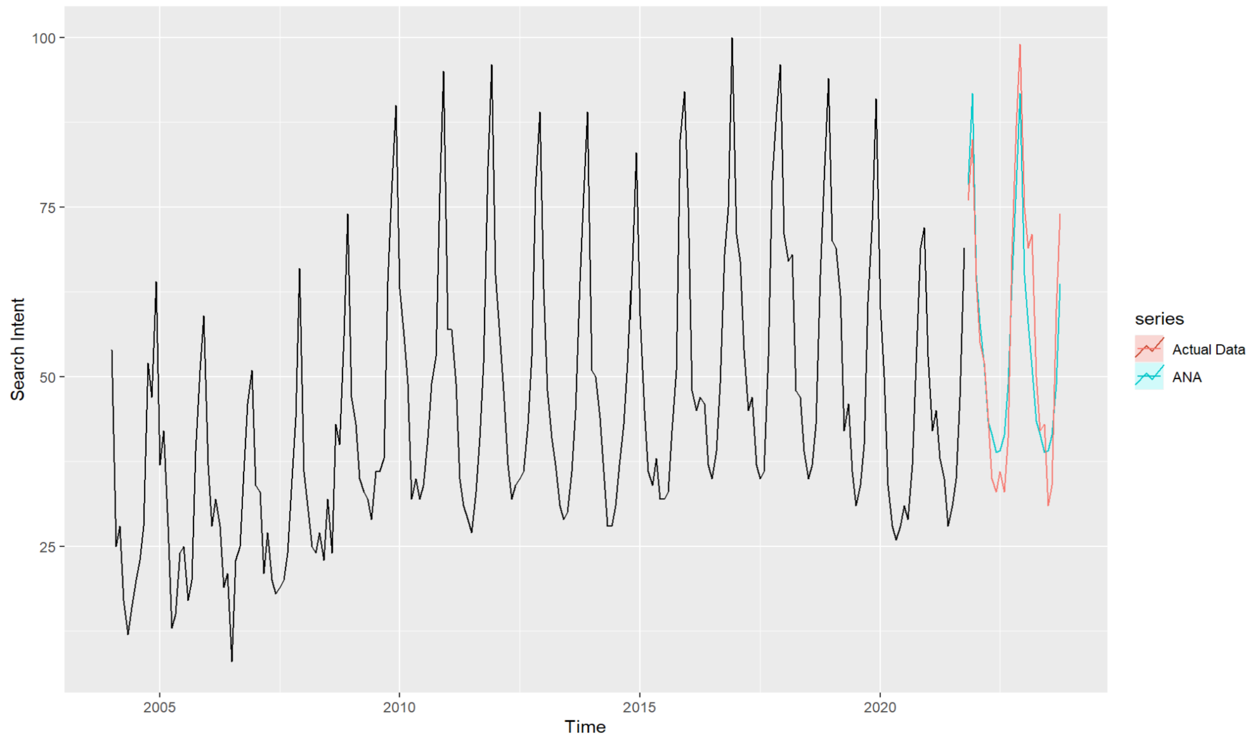


Seasonal Naive Forecasting Model

Holt-Winter's Exponential Smoothing Method

For Holt-Winter's Exponential Smoothing Method, we allow the algorithm to choose the optimal model for our dataset, which turns out to be the ANA model. This model has an RMSE of 5.78 for training and 7.75 for testing. The MAPE values are 12.22 for training and 12.28 for testing. As we check the residuals, it is similar to the normal distribution, with most lags falling within the threshold in ACF plot, representing no significant deviation from zero.

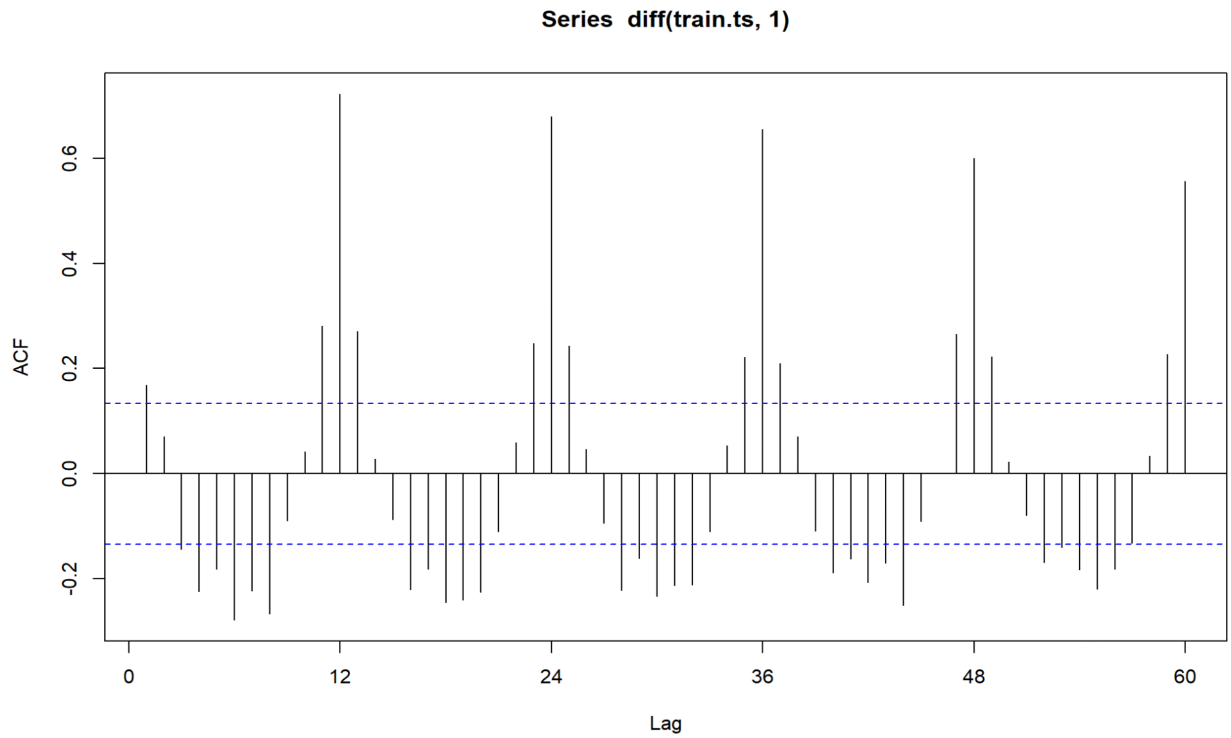
We use the training set to build the model and use the validation set to test the model's performance. In this graph, the red line represents actual data and the blue line represents forecast data.



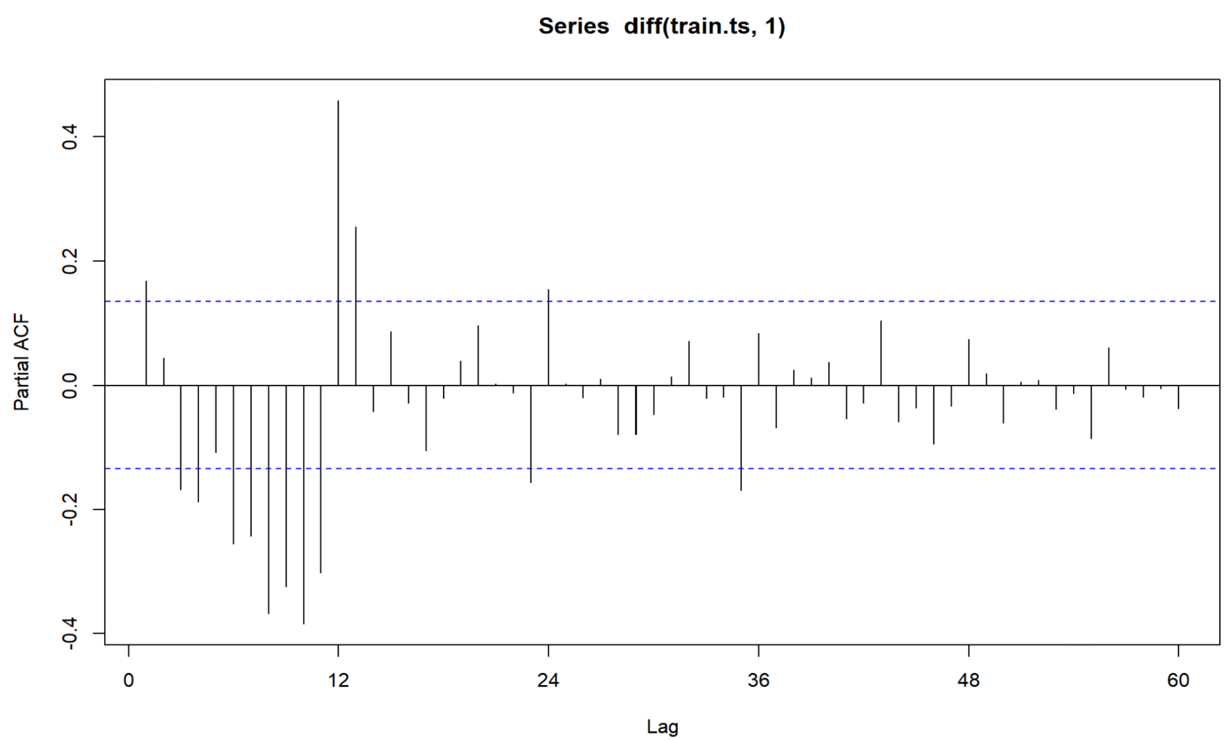
ANA Model

ARIMA Model

For ARIMA model, we first difference the trend and look at ACF and PACF to decide which model we should use. Since the graphs don't have clear signs showing its cutting off or tailing off for trend, we fit the data into 3 different models by changing ARIMA models for trend and keeping SARIMA at (1, 0, 0). SARIMA(1, 1, 1)x(1, 0, 0)12 performs the best among these 3 models, with an RMSE of 7.19 for training and 11.10 for testing. The MAPE values are 14.50 for training and 14.33 for testing.

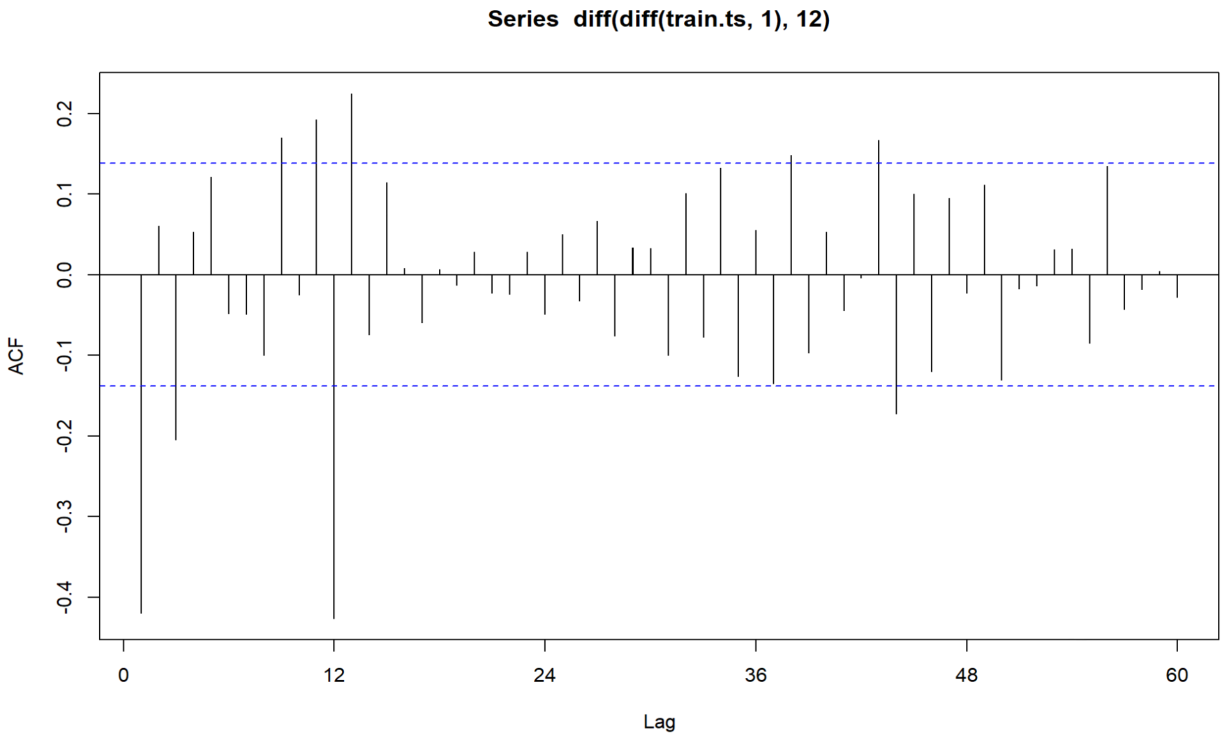


ACF (Differencing Trend)

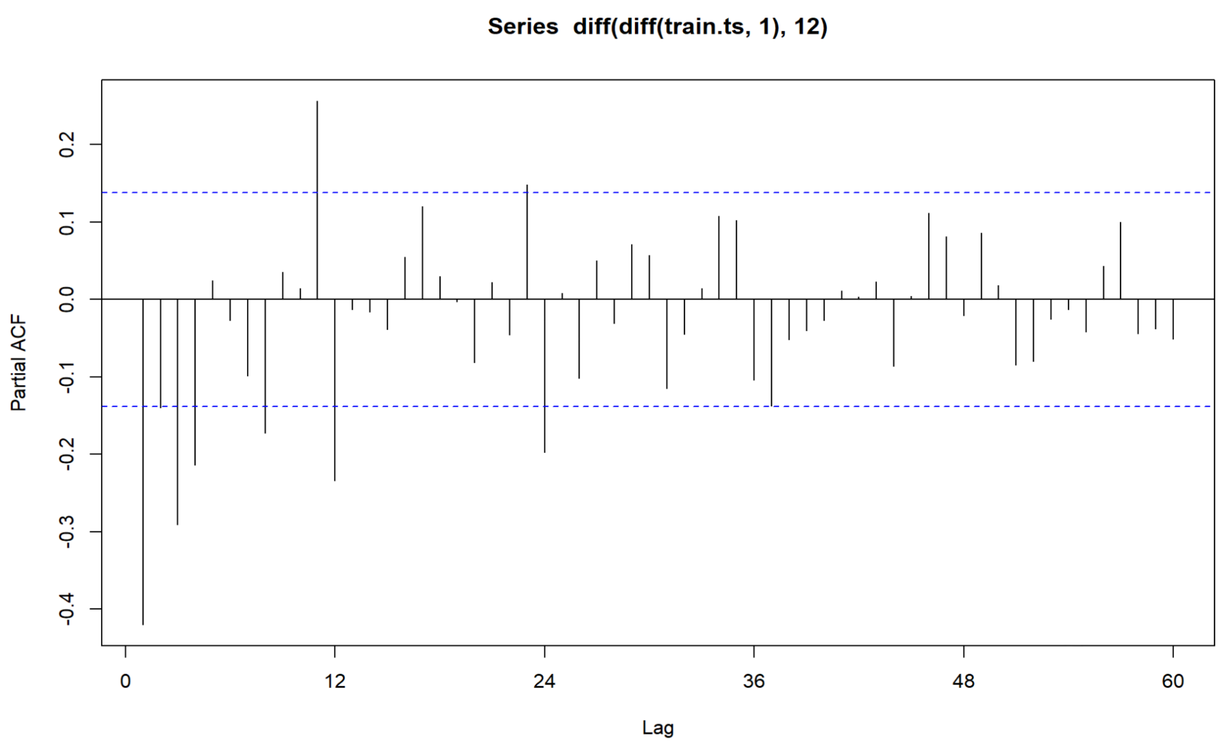


PACF (Differencing Trend)

We then difference both trend and seasonality to try out different models.



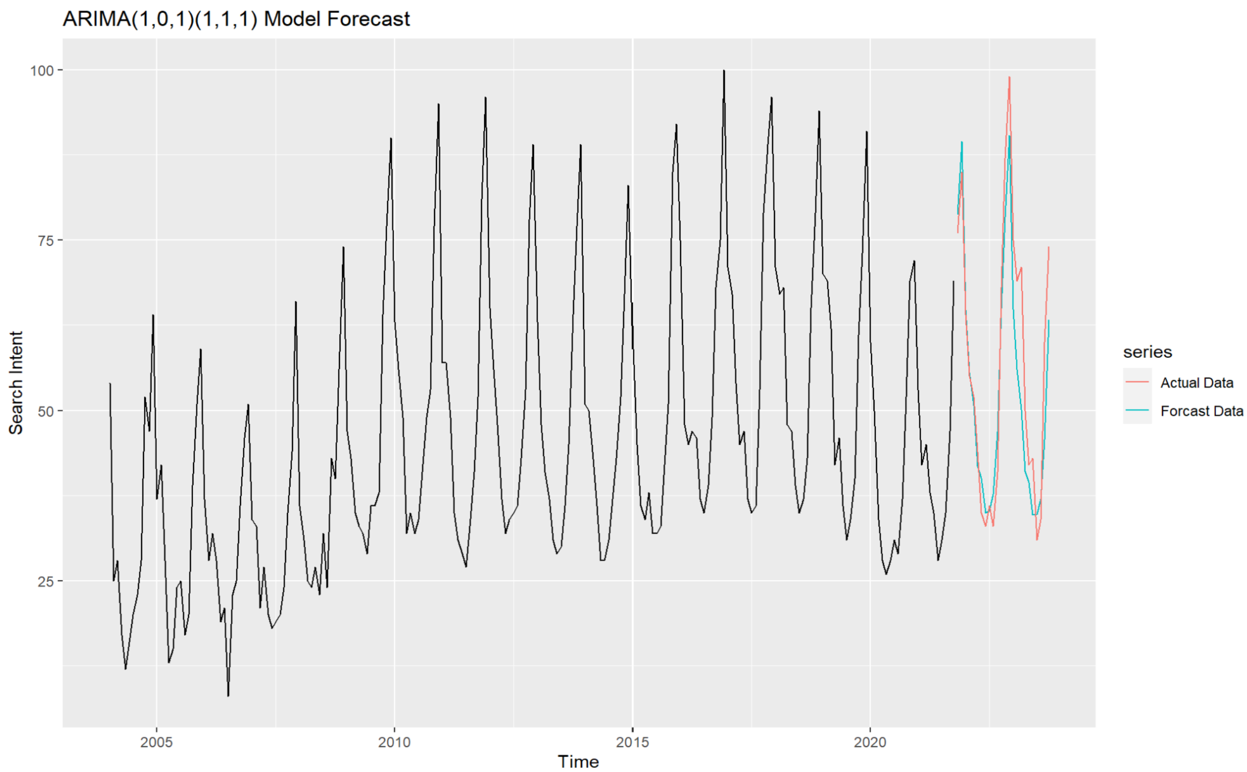
ACF (Differencing Trend and Seasonality)



PACF (Differencing Trend and Seasonality)

SARIMA(1, 0, 1)x(1, 1, 1)12 performs the best among all models, with an RMSE of 5.91 for training and 7.87 for testing. The MAPE values are 11.42 for training and 10.68 for testing. We check the residuals to see if the assumptions are met for this model. The residuals displayed a distribution close to normal, with lags mostly within the acceptable range.

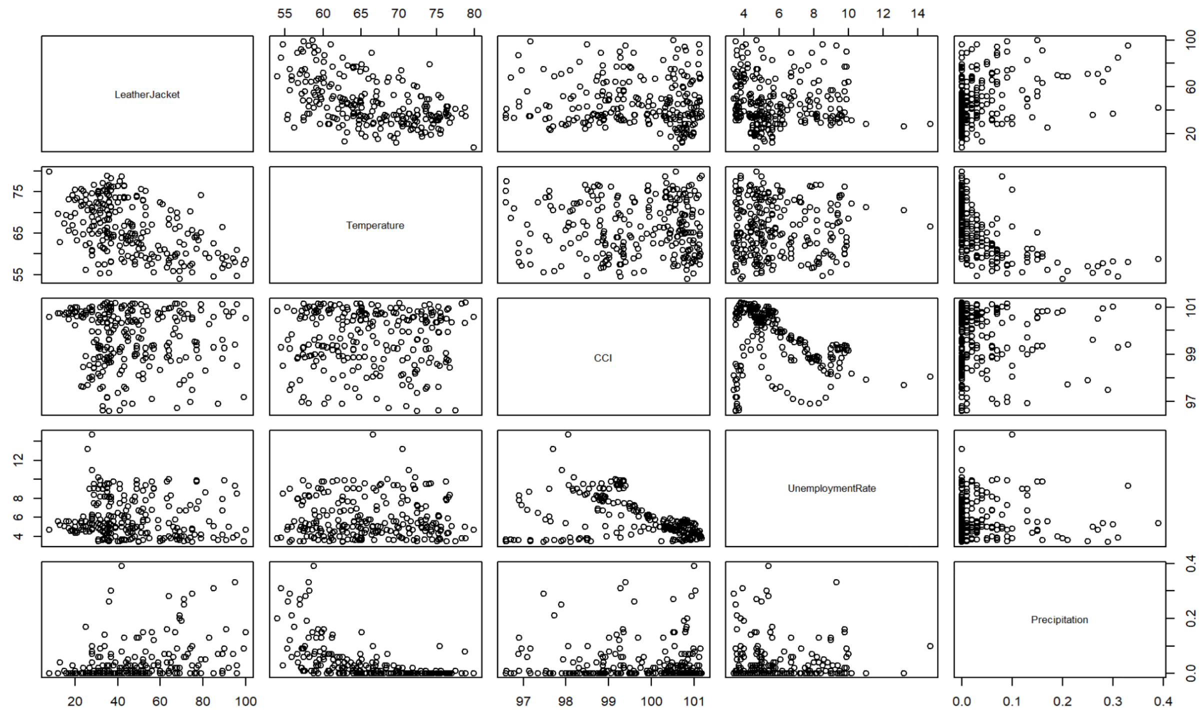
We use the training set to build the model and use the validation set to test the model performance. In this graph, the red line represents actual data and the blue line represents forecast data.



SARIMA(1, 0, 1)(1, 1, 1) Model

Multiple Linear Regression with External Variables

We also consider other datasets (temperature, CCI, unemployment rate, and precipitation) to see the relationship between fashion trends and these variables.



Scatter Plot between Variables

Correlation Coefficient:

	Temperature	CCI	Unemployment Rate	Precipitation
Leather Jacket Trend	-0.5060187	-0.06781938	-0.06845179	0.3634446

ADF Test:

	Temperature	CCI	Unemployment Rate	Precipitation
P-Value	< 0.01	0.9128	0.2253	< 0.01

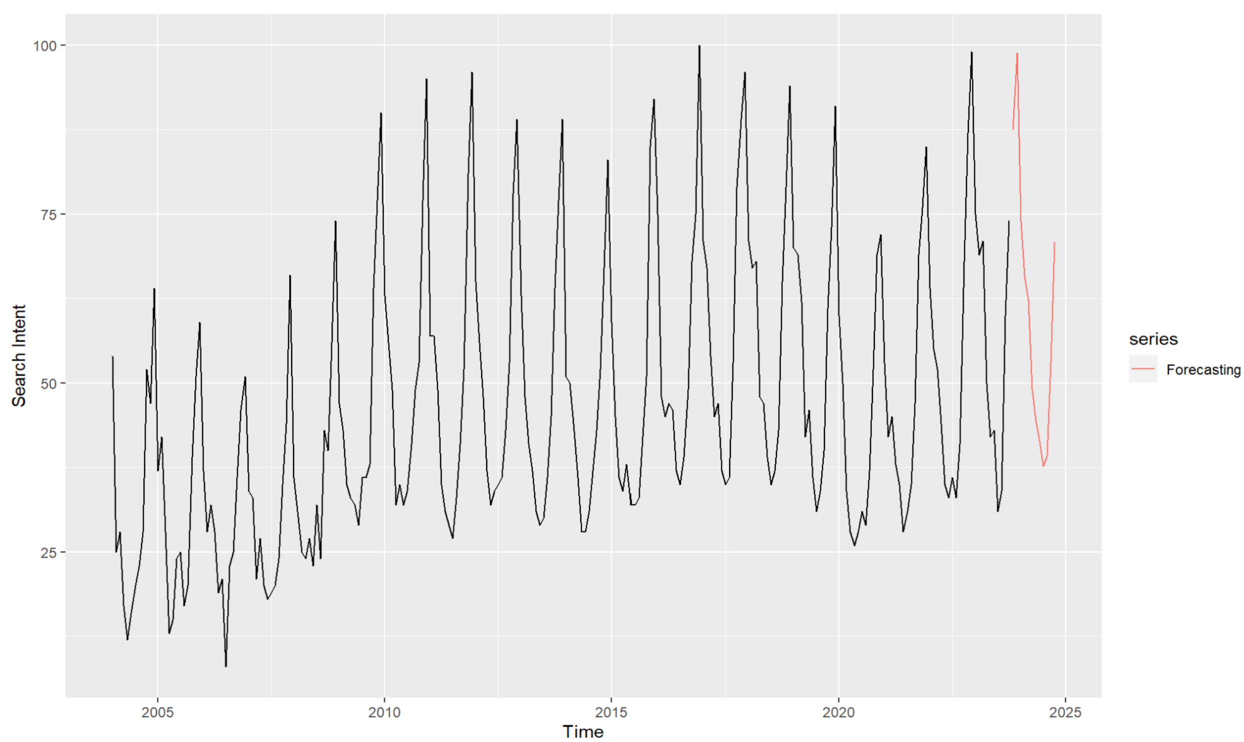
Due to the nearly zero correlation coefficient between CCI and unemployment in relation to the leather jacket trend, and the non-stationary nature of both CCI and unemployment rate data, we have chosen to exclusively utilize temperature and precipitation as independent variables.

We test three linear regression models. In the first model, only temperature is included as an independent variable. In the second model, only precipitation is added as an independent variable. The third model incorporates both temperature and precipitation as independent variables.

	Temperature	Precipitation	Temperature + Precipitation
Adjusted R-Squared	0.8331467	0.8241694	0.832786
RMSE	7.882811	8.092093	7.873693
MAPE	15.6054	16.45719	15.56268

Forecasting
SARIMA(1, 0, 1)x(1, 1, 1)12 emerged as the top performer in terms of MAPE, recording 11.42 for training and 10.68 for testing, and the assumptions are met for this model. Therefore, we selected SARIMA(1, 0, 1)x(1, 1, 1)12 as our final choice to forecast the future leather jacket trend.

To anticipate the upcoming trend, we utilized the entire dataset to train the SARIMA model, providing us with a robust framework for predicting the leather jacket trend over the next year, as illustrated in the graph below.



Forecast Trend for Next Year

Implications and Recommendations

Drawing on our chosen final model, SARIMA(1, 0, 1)x(1, 1, 1)12, and the anticipated popularity trend of leather jackets in 2024, we present strategic initiatives for companies and fashion influencers within the fashion/retail sector. To maximize revenue and enhance their reputation, it is important to grasp the underlying seasonality embedded in our predictions. Our recommendations leveraged the latest available data and served as relevant assumptions that support our forecasts. By aligning strategies with the forecasted trends, stakeholders can position themselves advantageously in the market and capitalize on emerging opportunities.

In our analysis forecasting the popularity of leather jackets in the year 2024, the outcomes generated by SARIMA(1, 0, 1)x(1, 1, 1)12 model reveal a pattern in consumer interest. Specifically, the demand for these jackets experiences a significant surge during the winter months, particularly from November to January (the beginning and end of the year). Conversely, consumer interest is observed to be at its lowest during the summer season, from June to August. This season-specific trend suggests that fashion/retail companies should be well prepared by outlining promotional efforts and inventory management of the leather jacket in the months leading up to the peak seasons in which the demands are significantly higher.

Implications

I. **Seasonal Inventory Management:** Given the peak interest in leather jackets during the winter months (November - January), fashion and retail companies should focus on optimizing their inventory management strategies. They should increase stock levels and ensure a variety of styles and sizes are available in the month leading up to the winter season to meet the heightened demand. Conversely, companies should consider implementing lean inventory practices during the summer to avoid excess stock and potential markdowns.

II. **Targeted Marketing Campaigns:** For fashion/retail companies, another strategy is to tailor marketing efforts to align with the seasonal fluctuations in leather jacket popularity. During the peak months in winter, companies can launch targeted advertising campaigns, promotions, and social media initiatives to capture consumer attention. Particularly, marketers can highlight the warmth, style, and versatility of leather jackets, emphasizing their suitability for cold weather to enhance the visibility of the brand during heightened demands. During the summer, the marketing strategies should focus on other seasonal clothing items, but consider running clearance sales to get rid of excess remaining leather jacket inventory.

III. **Content Creation by Fashion Influencers:** While the first two strategies are tailored for companies in the fashion and retail industries, the third strategy focuses on content creation specifically directed at our stakeholders, who are influential figures in the fashion industry. Fashion influencers should collaborate with other influencers and well-known leather jacket brands to create engaging content that highlights the appeal and versatility of leather jackets. During the winter season with high demands, they can focus on winter styling and transitioning to lighter styles for summer.

Limitations

I. **Limited Predictive Power:** While Google Trend data provides insights into the search interest for leather jackets, it solely serves as a proxy for purchasing behavior. It is imperative to acknowledge that individuals searching for specific items may not necessarily translate into actual purchases. Therefore, this analysis should be approached cautiously and complemented with additional purchasing data to make informed business decisions.

II. **Temporal Misalignment:** There might exist a time lag between searches and consumer behavior might differ. For instance, increased searches may not immediately translate to increased purchases. Additional relevant variables related to consumer behavior should be incorporated into the analysis for business decisions.

III. **Data quality:** Google Trends data has its limitations. The trending interest does not provide absolute search volumes but only relative interest percentage to its highest search record, and its accuracy might vary.

IV. **Geographical differences:** Consumer behavior might differ regionally. This model only captures the trend of Los Angeles areas and would require adjustments if applied to global or national trends.

V. **External Factors:** There are other variables influencing leather jacket purchases that could not be captured by our models. For example, fashion innovation, unforeseen events like pandemics, or promotions, are not accounted for in this analysis, but they are still important factors heavily influencing leather jacket interests.

Appendix

Exhibit 1. Time Series data from 2004 to 2023 October

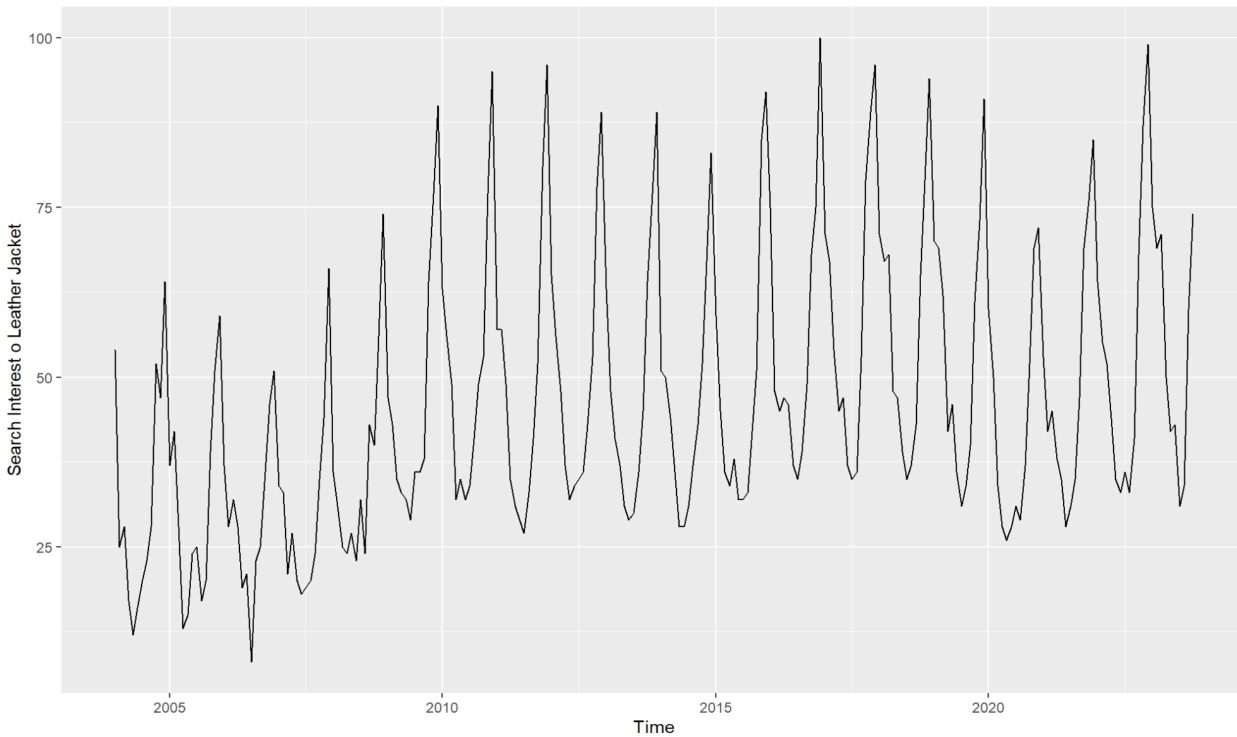


Exhibit 2. Season Plot

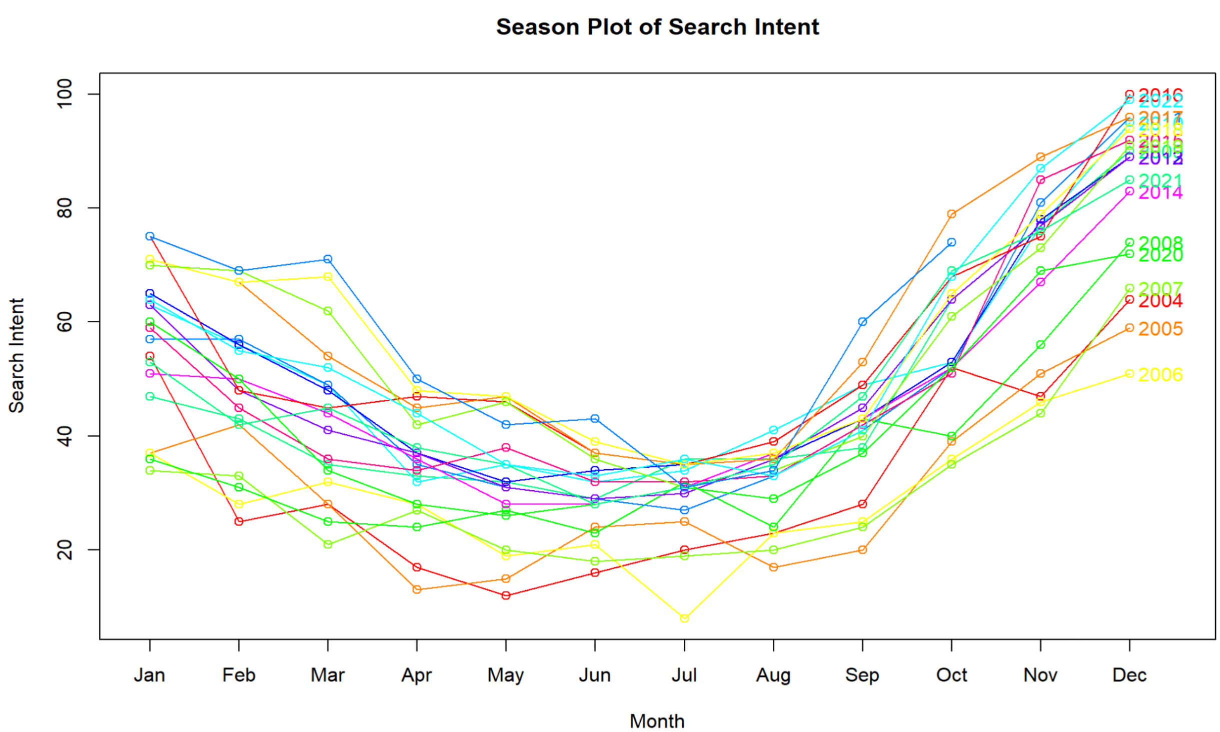


Exhibit 3. Seasonal Linear Regression Model Residuals

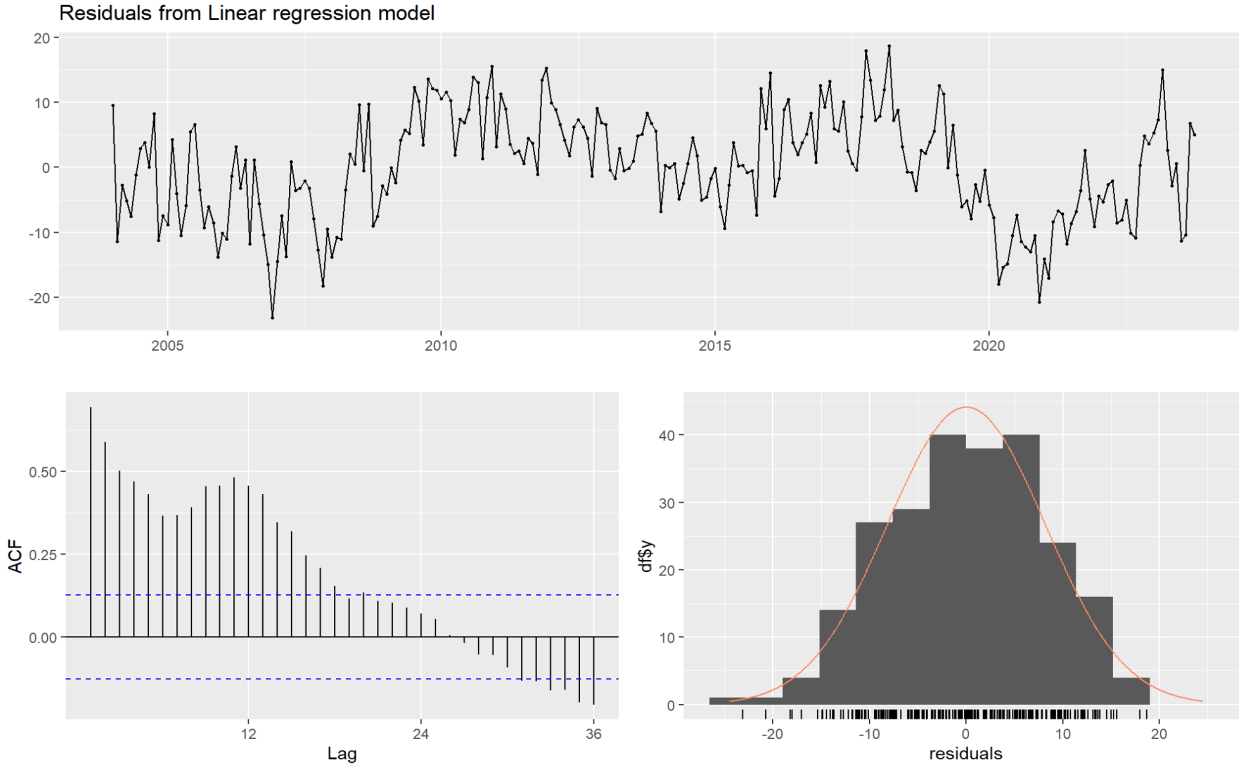


Exhibit 4. Naive Forecasting Model Residuals

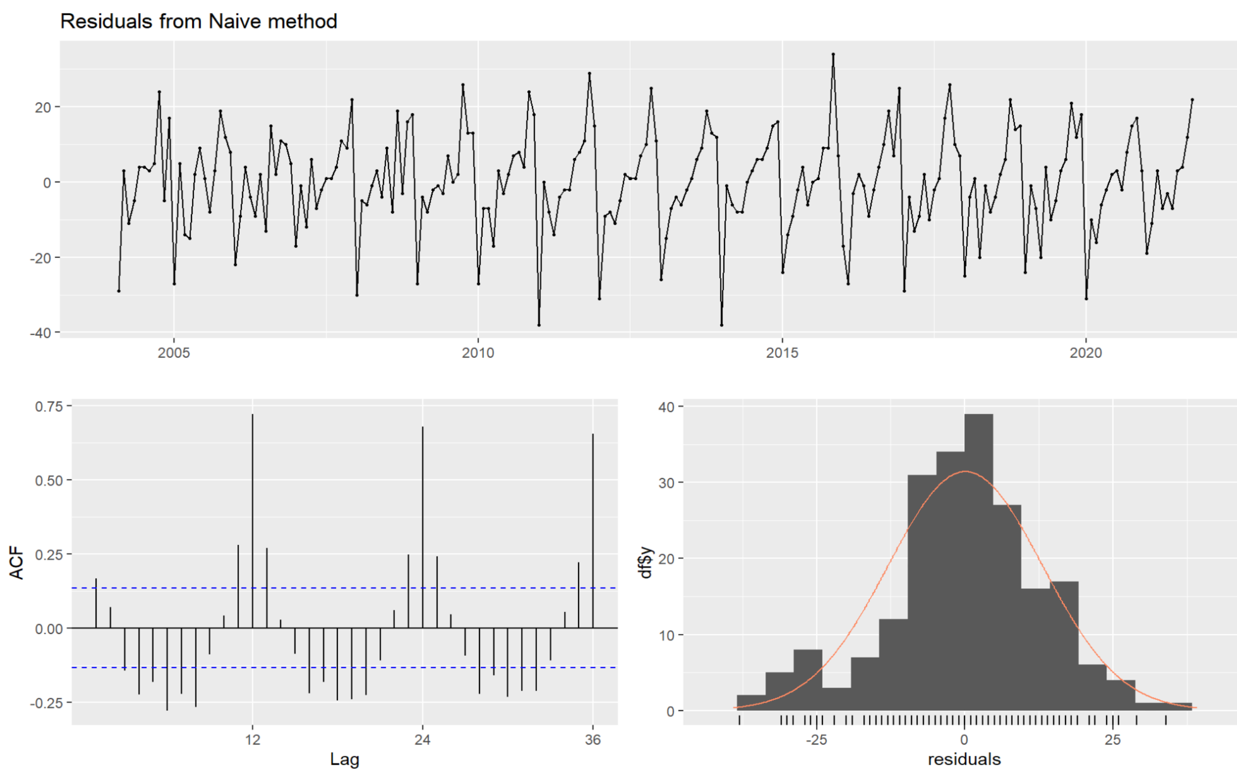


Exhibit 5. Seasonal Naive Forecasting Model Residuals

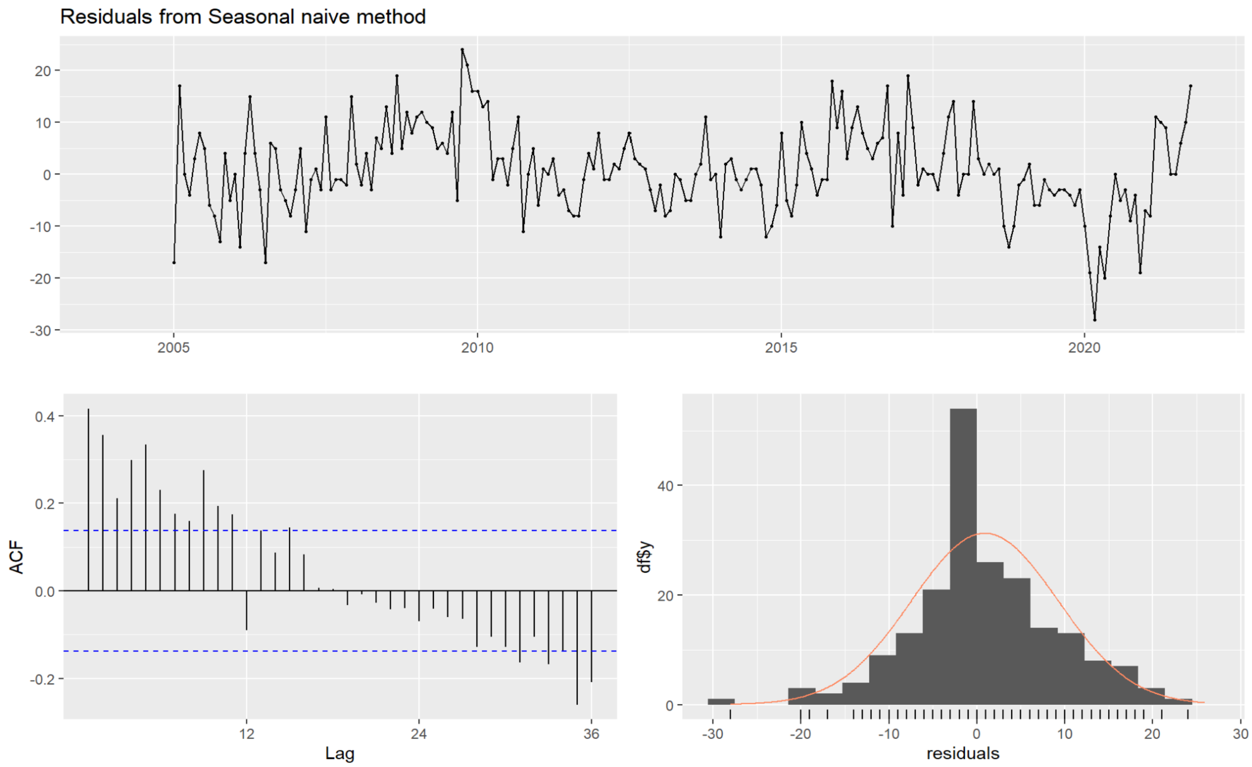


Exhibit 6. ANA Model Residuals

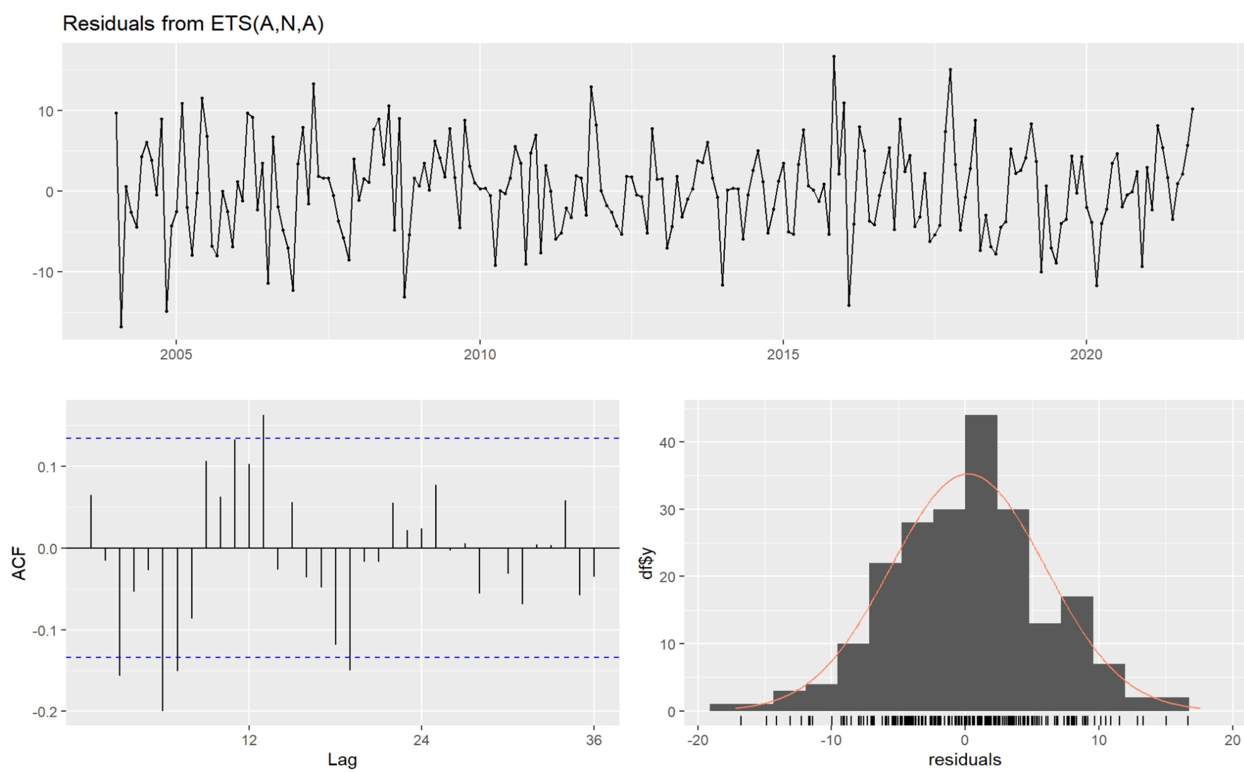


Exhibit 7. SARIMA(1, 0, 1)(1, 1, 1) Residuals

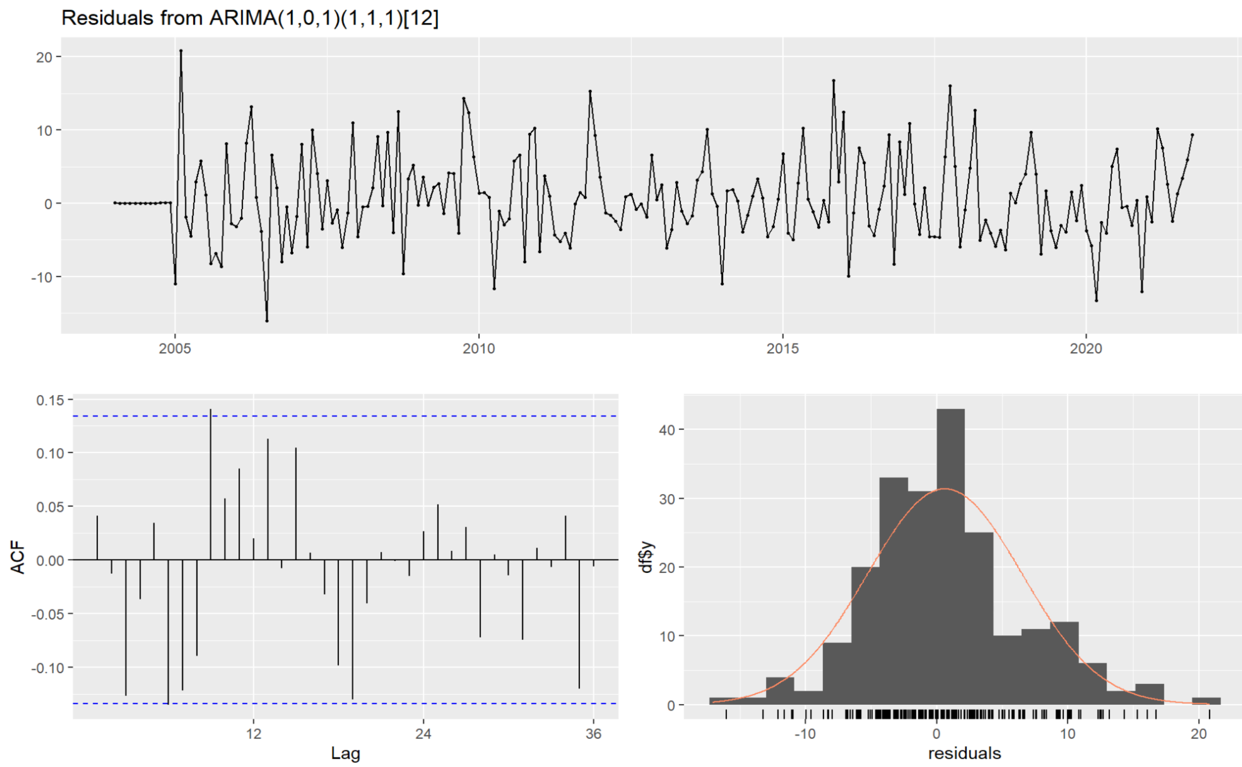


Exhibit 8. Moving Average Method

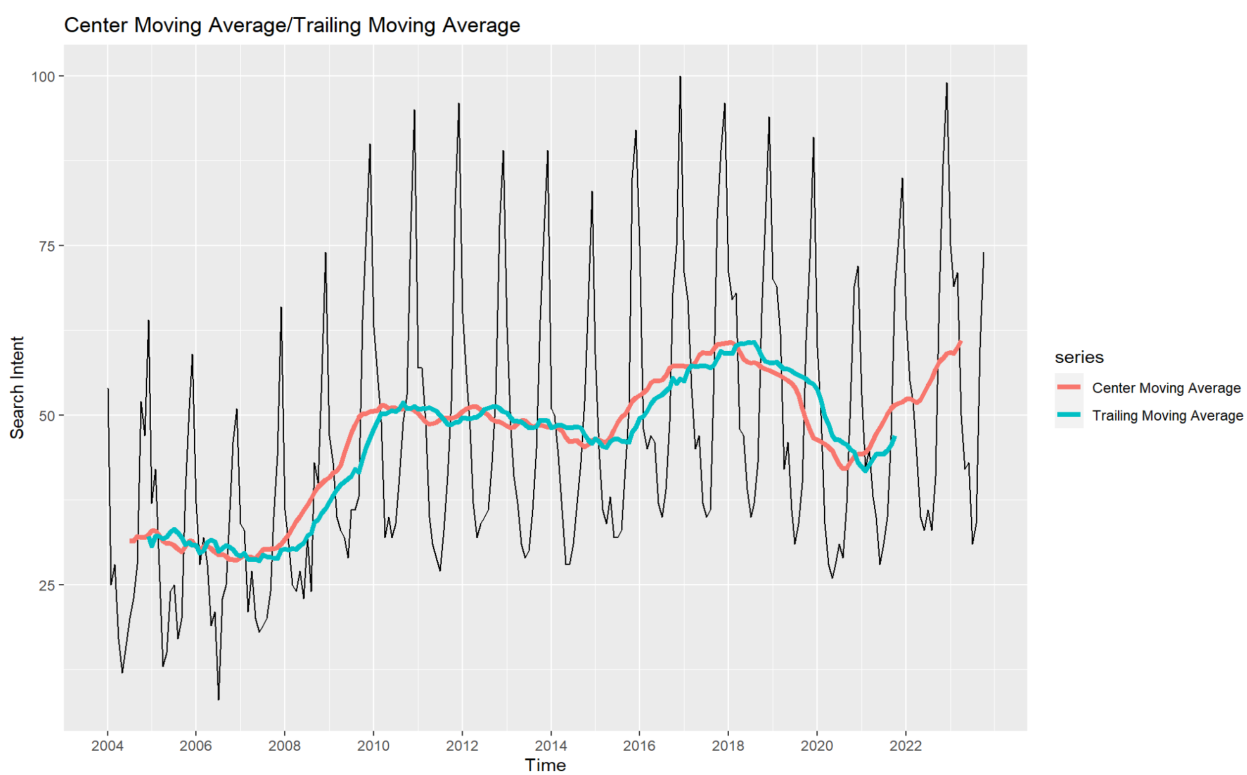


Exhibit 9. Moving Average Prediction