WORKFORCE PLANNING THROUGH VOLUME FORECASTING ON VIDEO-SHARING SERVICE INDUSTRY: A YOUTUBE CASE STUDY

Carmona, Jollie Ann
Fortuno, Vernideth
Jumangit, Ryan

Executive Summary

The explosive growth of user-generated video content on platforms like YouTube, with an estimated 3.5 million videos uploaded daily, necessitates proactive strategies for managing user inquiries and complaints. This paper proposes a forecasting model to predict the volume of incoming contacts, enabling video streaming companies to optimize resource allocation and workforce planning for efficient complaint resolution. By aligning staffing with anticipated demand, companies can enhance customer satisfaction and reduce operational costs.

This paper will use a 3-year horizon in identifying the trends and seasonality, if any, in the YouTube video streaming company data. Naive, ETS (Error Trend Seasonal) method, ARIMA (Autoregressive Integrated Moving Average) model and ARIMAX will be considered in the forecasting and selection of the final forecast model. An ARIMAX model as a candidate model was considered in order to account launches and events that may affect forecasts.

Comparison of forecasting accuracy for the three methods will be the basis for the final model. For the data, ARIMAX model shows comparatively better performance compared to the other methods with the lowest Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Akaike Information Criteria (AIC).

Given the output of the final model, this paper will provide the company a better outlook on the demand needed for their workforce. With this, the company can supplement the needs thru hiring additional manpower or upskilling their current workforce to serve their customers.

Table of Contents

1. Introduction

1.1 Background

Social media is defined as "a collection of internet websites, services and practices that facilitate collaboration, community building, participation, and sharing" (Reynol et al., 2011). People frequently use social media engines such as Facebook fan pages, Twitter accounts and YouTube channels to express their views and thoughts, disseminate information about product launches and events and create articles and video content that may be of interest to users. Despite the benefits of social media platforms, it also brought a surge of harmful content such as fake news, rumors, hate speech, aggression, and cyberbullying that can negatively impact mental health and lead to irreversible losses.

YouTube is a video-sharing and social media platform that allows users to upload, watch, and share videos. It offers a stage for content creators and a targeted audience for advertisers. The company offers free and premium membership tiers. Free members enjoy the basic features like sharing and uploading, while premium members enjoy ad-free streaming and offline video downloads. To ensure a safe and respectful environment for all, social media platforms have systems for reporting, reviewing, and responding to user feedback. Users can submit their concerns and feedback to the platform through contact channels like chat or emails.

In this paper, we will investigate how YouTube capitalize the workforce given the contact volume of their platform. Forecast of volumes and its approach to staffing of company's workforce to which talent allocation, job realignment, attrition, allocation of workload and even its cost savings in the business operations would be one viable impact.

1.2 Business Problem

One of the core goals of businesses is to optimize the workforce through strategic planning. In order to develop an effective cost reduction method, one of the best practices in workforce planning is by using forecasting techniques in the workforce.

In its broadest sense, it aims to provide answers to these two fundamental business questions:
1) How much volume of chat and email do we expect in the next 52 weeks?
2) How many people will we staff in the next 3 months and how can we account for events and launches in these forecasts?

1.3 Analytics Problem

In a forecasting outlook, we generate volume forecasts using different methods. For this study, we would be using Naive, ETS, ARIMA and ARIMAX.

Analytics questions that our research would like to identify are the following:

1. Best forecast model that will provide prediction in the next 52 weeks on the volume of email and chat. Basis for the best model will be the lowest accuracy metrics, such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Akaike Information Criteria (AIC).
2. Given additional variables, such as campaigns and launches, can we use ARIMAX in the forecast to consider these variables?

1.4 General and Specific Objectives

The objective of the study is to provide guidance on the company which forecasting model will help identify the volume of chat and emails from customers in the next 52 weeks. This is aligned with the core goal of the company to optimize their current workforce and ensure that workforce planning is followed.

With the workforce planning, YouTube may be able to identify when to supplement the support needed by the existing team and when training can be done in order to upskill the workforce.

To meet this objective, the study aims to satisfy the following:

1. Identify the current contact volume (per channel) of the company. For this study, we can segregate chat and email for the means for obtaining user complaints.
2. Determine trend, seasonality and error that may influence the contact volume of emails and chat for YouTube.
3. Consider a different candidate forecasting models, such as Naive, ETS, ARIMA and ARIMAX, for i. Chat and ii. Emails. Researchers would like to segregate the forecast for chat and emails given the difference on handling and turnaround time of ticket resolution.
4. Compare the different forecasting models and identify the final model through the use of lowest accuracy metric, such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Akaike Information Criteria (AIC).
5. Incorporate launches and events in the forecast by using ARIMAX and investigate the performance versus ARIMA, ETS and Naive forecast models.

1.5 Scope and Limitations

This analysis is going to be limited for platform subscribers of YouTube from June 2020-October 2023 specifically on those determined in the English market or language. Contact volume stream would include chat and email tickets received through specific media platforms across different English languages where data tables will be pulled through company specific tools and engines. A ticket refers to a unique record or identifier assigned to a specific inquiry or problem raised by YouTube platform users. In this ticketing system, each subscriber issue or question is treated as a separate "ticket," even if a single user has multiple issues. This ensures a clear and organized approach with each problem addressed individually. The types of tickets received can vary between technical issues and content-related problems to which it generates a distinct ticket for each submission.

An RStudio script will be set up to run results for forecasting models and accuracy metrics. Best model will also be determined throughout the analysis. Coverage of post-mortem analysis on the effect of forecast in the current business model will be discussed.

2. Review of Related Literature

Workforce planning, as discussed in “Seven Steps of Effective Workforce Planning”, is a strategic alignment of the company’s human capital versus the demand of the business. This is a process of identifying and optimizing the current workforce by identifying the current workforce, current demands and also take into account the future demands. (Cotten, 2007). In doing so, company will be able to benefit with an efficient system on allocating resources and eliminate scenarios of shortages or surpluses in the workforce. (Girotra, 2022). Additionally, continual improvement to its employees by having more time for training and less stressful environment will be achieved with proper workforce planning.

Multiple researches have happened in terms of workforce planning. One of which is related to the medical industry in the UK. Research was regarding predicting the demands for healthcare workers in the UK using different Time Series techniques. Forecast was also compared with the growth of the economy, using GDP as indicator, and the population growth. Girotra considered using ARIMA models in this research. However, upon validating the initial results, he noticed the limitations encountered while using the model. Thus, incorporating and settling for the Seasonal ARIMA models. Comparison of the best value for model selection using Bayesian Information Criteria (BIC), Corrected Akaike Information Criteria (AICc) and Akaike Information Criteria (AIC) was used. Lastly, Granger causality null hypothesis test was used to discover if there was a relationship between the population data and the health worker data (Girotra, 2022).

Another study also looks at the number of expected calls for a call center team led by Ibrahim and Ye. This is a literature survey of different models used by different companies in order to forecast the number of calls. They highlighted the complexity of developing a forecast model and at the same time the importance of this model and the benefits that it will create to a more efficient operational decisions (Ibrahim and Ye, 2016). Approached ranged from modeling using standard forecasting techniques such as ARIMA models,

models over several days and models over a single day, wherein Poisson process is highlighted and used.

Forecasting Analytics is also used in the electricity demand of a country. In the study of Rabbi etal, they attempted to forecast the yearly electricity demand of Bangladesh. This aims to ensure that the study will be useful in predicting and ensuring ample supply of electricity to every household (Rabbi etal, 2020). Researchers used the ARIMAX model with exogenous variables identified as population and GDP per capita. In their methodology, researchers pointed out that before starting with the forecasting, they analyzed the relation of the two exogenous variables versus the response variable. They found strong linear correlation for the electricity demand and population and also electricity demand and GDP per capita.

Meanwhile, Anggregani, Vinarti and Kurniawati studied demand forecasting for clothing demand particularly on possible effect of the Eid holidays. Purpose of the study aims to compare the results of ARIMA and ARIMAX models using data from the clothing industry and Eid holidays as a factor for increase in sales during a specific period. From the study, researchers used dummy variables in order to tag and incorporate the variation effects before Eid holidays. Some of the considerations they took are month before the holiday, during the holiday and month after the holiday. Results showed that ARIMAX performed better, this means having smaller value compared to ARIMA, when they compared the Akaike Information Criterion (AIC), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Additionally, results of plot of training and testing data shows a better forecast when they used ARIMAX compared to ARIMA.

Lastly, researchers also looked at different forecasting accuracy measures to identify the best forecasting models. One of many forecast accuracy measures are coming from the scale-dependent measures and measures based on percentage errors. To be exact, Root Mean Square Error (RMSE), from scale-dependent measures, has the measuring ability to specify its mean but has the advantage of being expressed on the same scale series while being sensitive to outliers when its results are aggregated across multiple time series (Koutsandreas etal, 2021). Mentioned also is Mean Absolute Percentage Error (MAPE), which is considered being the most popular choice in accuracy metrics (Fildes & Goodwin, 2007). MAPE enables evaluation of forecasting accuracy across multiple time series of different scales and most especially easy to communicate to business and organizations (Kolassa and Martin, 2011).

3. Data and Methods

3.1 Data Sources

The dataset used in the analysis revolves around tickets generated only by paid subscribers on the platform of YouTube extracted from the internal database and dashboard. Out of 26 languages in the pipeline, the focus of forecasting will be English chat and email volumes, regardless of the subscriber's country of origin. Separate forecasting will be done for email and another for chat since we are considering the turnaround time needed for these types of interaction. This targeted approach allows YouTube to leverage a team of skilled agents who specialize in English language support. These agents are trained to handle a wide range of issues ensuring efficient and effective resolution for English-language concerns.

In instances where tickets are submitted in languages other than English, such as Japanese, Spanish and Hindi, these will be routed and assigned to specialized agents fluent in those languages. This approach ensures that each subscriber's concern is addressed by an agent with proficiency in the language of the submission, providing a personalized and tailored support experience. Since a large volume of the tickets comes from English speakers, the dataset used focused on this language of support. In terms of resolving tickets, chats were resolved real time as live agents were accommodating the request while emails were resolved within the day. In case that the ticket is too complicated and needs escalation, another specialized team will handle the request.
The analysis is crucial for resource planning and management within YouTube's contact centers, providing insights into staffing needs to efficiently handle user interactions and support requests from English-speaking subscribers.

3.2 Data Preprocessing

Data Gathering and Data Cleaning

- Data will be pulled using the company's database. To ensure accuracy of data, this will be cross validated to existing reporting dashboards.
- The English language will be the primary filter for the data. This is aligned with the scope and limitation of the study.
- Data cleaning includes removal of volumes from incomplete weeks and/or missing weeks.

Forecasting Process

1. Analyze available historical data in more detail

i. Create subseries plots

Ii. Decompose the data to determine how various components (trend, seasonality, error) impact the time series

iii. Create autocorrelation (ACF) and partial autocorrelation (PACF) plots to further the analysis.

iv. Values of p and q are determined based on the ACF and PACF plots and value of d depends on level of stationarity in the data. In PACF plot, the number of spikes indicate the order of autoregression (value of p in ARIMAX(p,d,q)) while number of spikes in ACF indicate moving average (value of q in ARIMAX(p,d,q))

2. Decide how to validate models

i. Applying the train and test split methodology, the data will be split into 80% training dataset and 20% test set. In the data, the training set includes 156 weeks which covers the weeks from January 2020 to December 2022 while the test set includes 44 weeks and covers January 2022 to October 2023.

3. Create preliminary forecasting models

i. Depending on data patterns and decomposition analysis, this should include all potential models in the initial model building phases. Candidate models are Naive, ETS, ARIMA, and ARIMAX. Determine if data transformation and/or differencing is necessary. If needed, use BoxCox transformation and/or differencing to stabilize data and make it stationary (i.e. remove heteroscedasticity and/or trend)

4. Use test sets and/or Cross Validation to calculate accuracy metrics for the determining potential models

i. Good candidates for accuracy analysis include: RMSE (root mean square error), MAE (mean absolute error), MASE (mean absolute scaled error), and MAPE (mean absolute percentage error). Other metrics that could be used are AIC (Akaike's Information Criteria) and/or BIC (Bayesian Information Criteria). Select models with lowest RMSE, MAE and/or MAPE and highest MASE and/or with lowest AIC/BIC to determine the "best models".

ii. Among the accuracy metrics, MAPE (Mean Absolute Percentage Error) is the most suitable choice for evaluating models like ARIMAX as it effectively penalizes large errors. It simply calculates the average percentage difference between predicted and actual values. (Koutsandreas, D. etal, 2021).

3.3 Analysis

Adding of Launches and Events

Initially, project managers will provide all known events and/or developments that may potentially impact volumes (new launches, policy changes, process changes, etc.), as well as the estimated impact of those events and/or developments. For example, in the next three months, there will be an anticipated NBA league, NFL, Black Friday Sale, Phone product release, YouTube product and service and etc that will impact the volume of the forecast. An estimated volume of impact per month will be provided and will be broken down into weeks through percentage representation. For example, an additional

200 volumes for the next three months are computed as 0.05% per week for both email and chat. To account for the launches and events, ARIMAX (AutoRegressive Integrated Moving Average with exogenous variables) and time series regression with input variables was used. ARIMAX allows the integration of external factors (exogenous variables) into the time series model to capture their impact on the data of customer support tickets. Time series regression with input variables involves using additional factors alongside the time series data to enhance the model's predictive capabilities. Through ARIMAX, these product launches and events will be treated as an exogenous variable or as an additional regressor to the forecasting equation.

Forecast Breakdown and Presentation

The forecast results will be assessed to break down projected ticket volumes from a weekly to a daily perspective. This detailed breakdown aims to assist YouTube's contact centers in making informed staffing decisions. By understanding daily variations in ticket submissions, the contact centers will be able to optimize headcount to ensure staffing levels to efficiently address user interactions each day.

3.4 Methods

Naive, ETS, ARIMA, and ARIMAX will be used in the analysis. To ensure model suitability, the data will undergo thorough cleaning and preprocessing. This also includes essential steps such as training and testing set creation, STL decomposition, and examination of autocorrelation functions (ACF) and partial autocorrelation functions (PACF).

Accuracy measures, including Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), and Akaike Information Criterion (AIC) will be considered for comprehensive model evaluation.

Training and Testing Set Creation

To strengthen the robustness and accuracy of the forecasting models, a series of steps will be done in the data cleaning and preprocessing phase. The first step is to pull all necessary data from the dashboard and remove those weeks with incomplete data volume. Next is to create separate training and testing sets so that it can assess model performance. A split of 80% to 20% will be used in this analysis to cover all weeks in a month. Majority of the data will be training and the remaining portion will be testing.

Seasonal-Trend decomposition using LOESS (STL)

To know the important elements of the time series, we will do the Seasonal-Trend decomposition using LOESS (STL). Through this process, we will be able to see and identify remaining irregularities in the data as well as trends and seasonality. This is an important step because it will help us better understand the dataset and its underlying structure by looking into the basic patterns and behavior.

Analyzing Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF)

In this process, it will help us choose the right parameters for Autoregressive Integrated Moving Average (ARIMA) models. This is very vital since we are looking into ARIMAX where exogenous variables play a major role in the forecasting behavior. Moreover, these plots will help identify the suitable lag values, guiding the selection of appropriate Autoregressive (AR) and Moving Average (MA) components.

Forecast Accuracy

Akaike Information Criterion (AIC) is one of the vital measures to look into as it directs the selection of the best model by balancing the trade-off between goodness of fit and model complexity. Ideally, the model with lower AIC indicates balance and optimum goodness of it. Lastly, the precision of the models will be evaluated through Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). It is important to note that MAPE evaluates the percentage difference between the expected and actual values. Whereas RMSE measures the square root of the average squared differences. Just like AIC, lower values or results in MAPE is considered the most accurate model as these measures the percentage of error.

52 weeks forecast

For the 52-week forecast, the time series data will be initially transformed and split into training (80%) and testing sets (20%). Naive, ETS, ARIMA and ARIMAX models will be used to generate the 52 weeks ahead volume. By doing the previous steps, it will ensure a robust 52-week forecast, considering data splitting, model fitting, and thorough accuracy evaluation.

3.5 Why the model

Data Modelling through Naive, ETS, ARIMA and ARIMAX

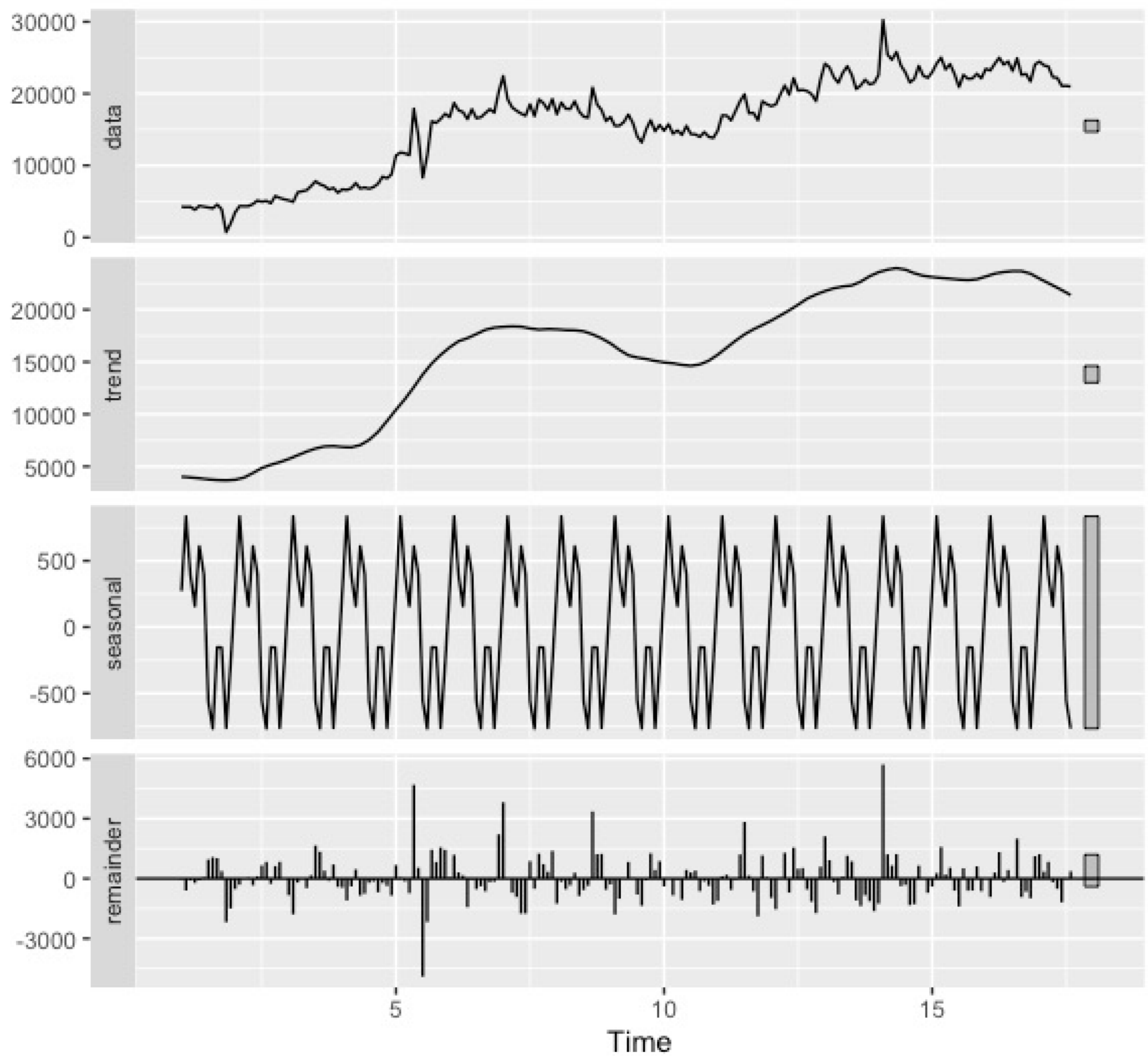
The Naive model is a simple baseline that is useful for data with little trend or seasonality because it assumes that future values will reflect the most recent observation. On the other hand, we use the ETS model because it effectively captures and models the error, trend, and seasonality in time series data, making it a reliable choice for accurate predictions in various applications. Using stepwise and disabling approximation, ARIMA model will be used thoroughly as it is intended for time series with trend and seasonality. It performs best when there is a trend or seasonality in the data. By including exogenous variables, such as launches and events, the ARIMAX model will be highly considered as it improves accuracy over the ARIMA model. By enabling seasonality, ARIMAX will capture time series patterns influenced by outside factors. In the case of our data, these were the launches and events.

4. Results

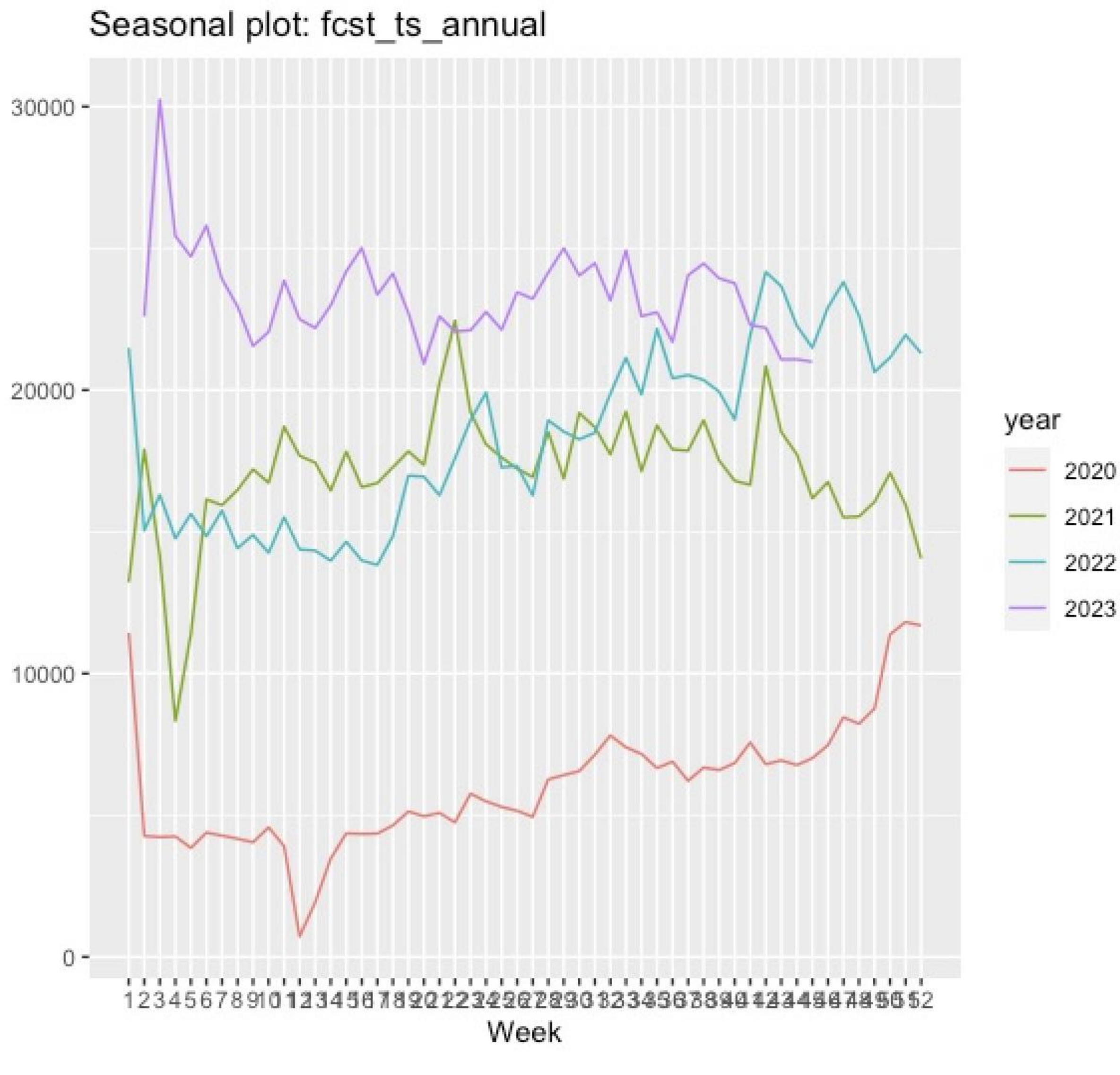
4.1 Results - English Chat

This section covers the analysis for English chat volume. This includes the time series decomposition, training and testing, and the evaluation of the models.

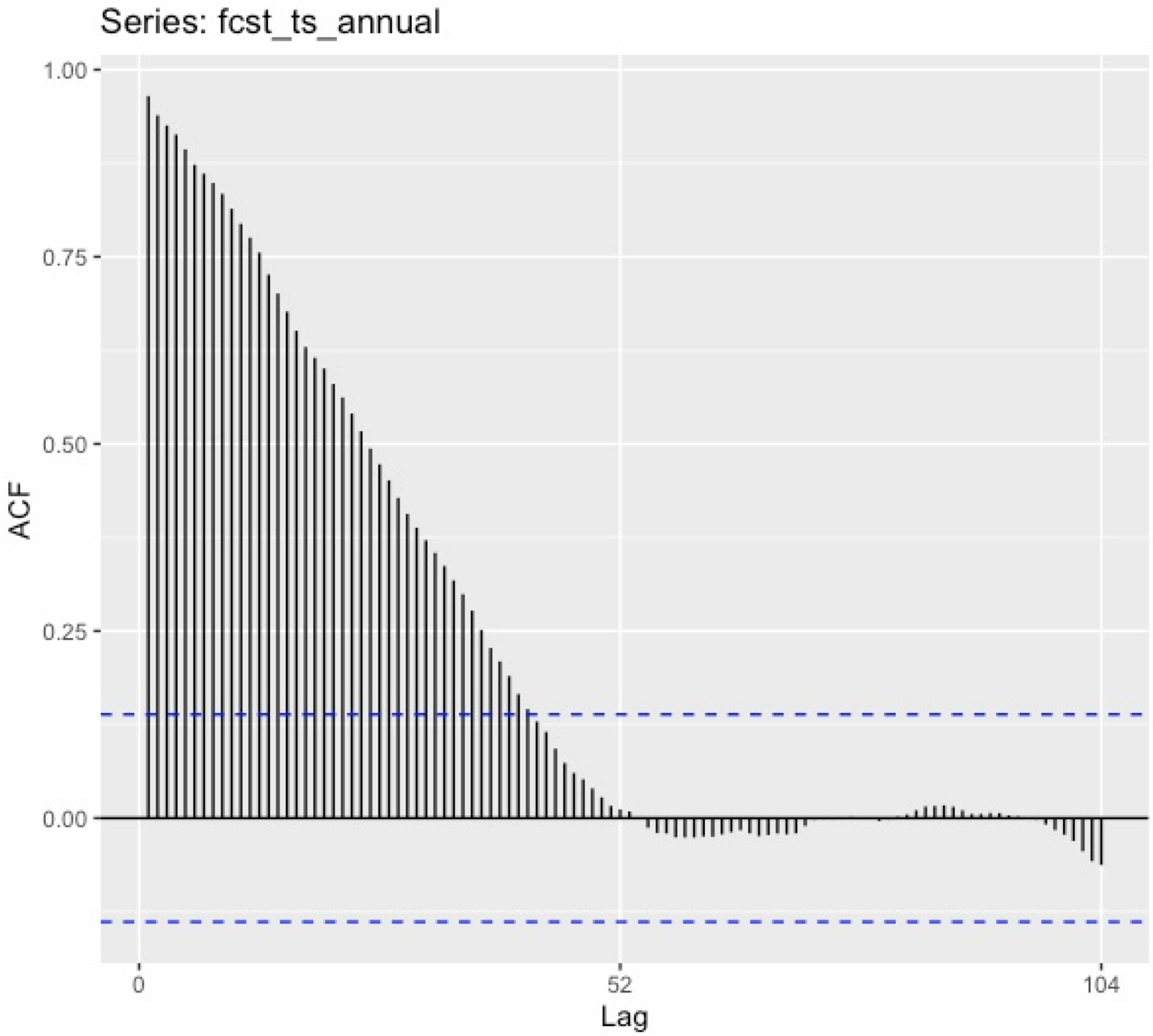
4.1.1 Exploratory Data Analysis



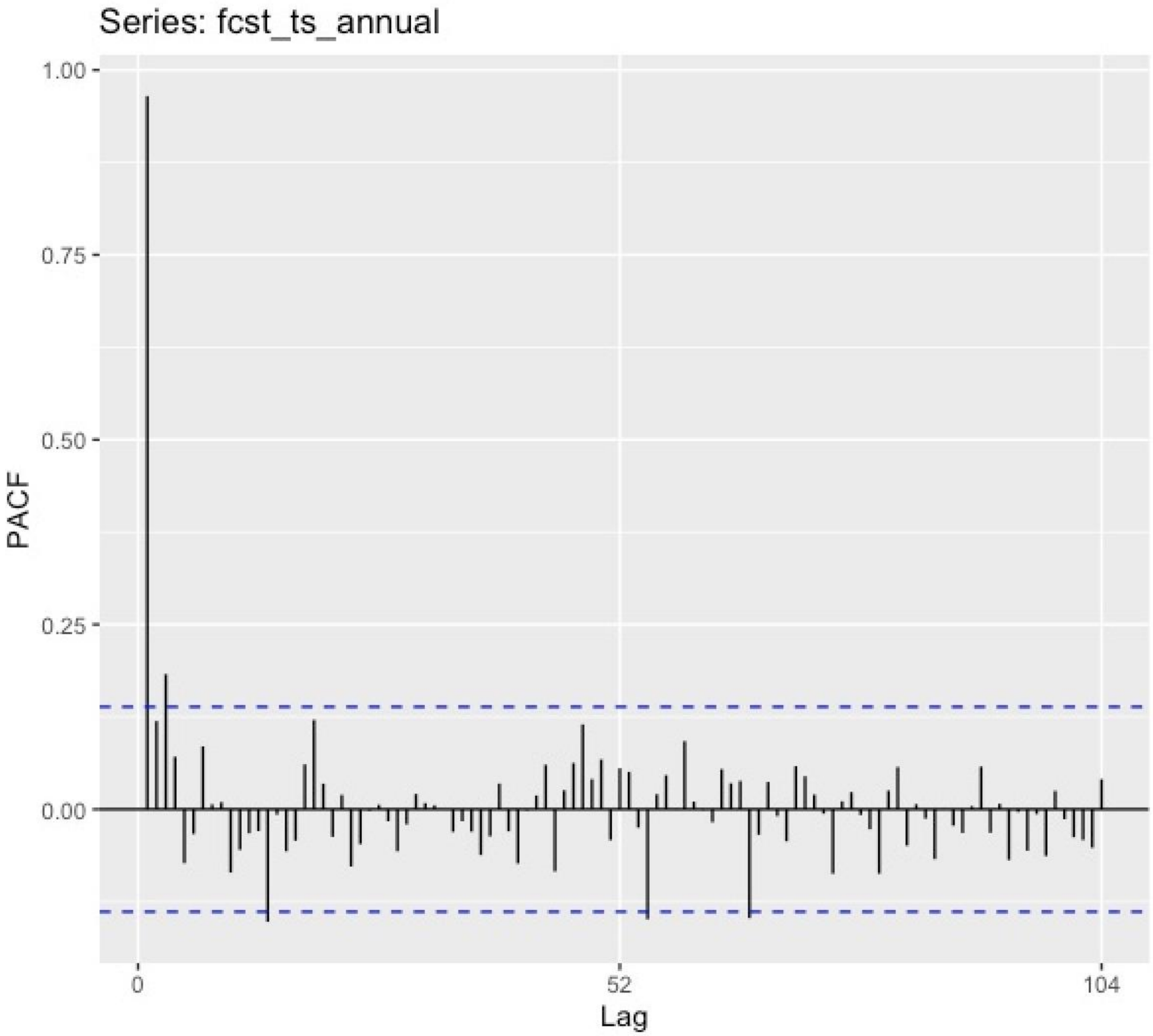
The graph shows time series decomposition into three key components: seasonal, trend, and remainder. The seasonal component shows fluctuations, indicating varying seasonal patterns throughout the year. For instance, in January, the seasonal effect is positive at 271.7584, while in July, it turns significantly negative at -558.3458 (check results from R Script chat run). On the other hand, the trend component reveals an overall upward trajectory, showing a consistent increase over the months. The remainder or outlier captures unexplained or random fluctuations in the data, highlighting periods with unexpected variations.



The seasonal plot shows volume fluctuations from January 5, 2020, to October 29, 2023. Notable patterns hint at a yearly cycle with regular ups and downs. Observing the weeks like March 15, 2020, and December 6, 2020, evidently shows unusual spikes. Overall, there's a consistent upward trend, especially post-mid-2020, indicating a general increase in volume over time.

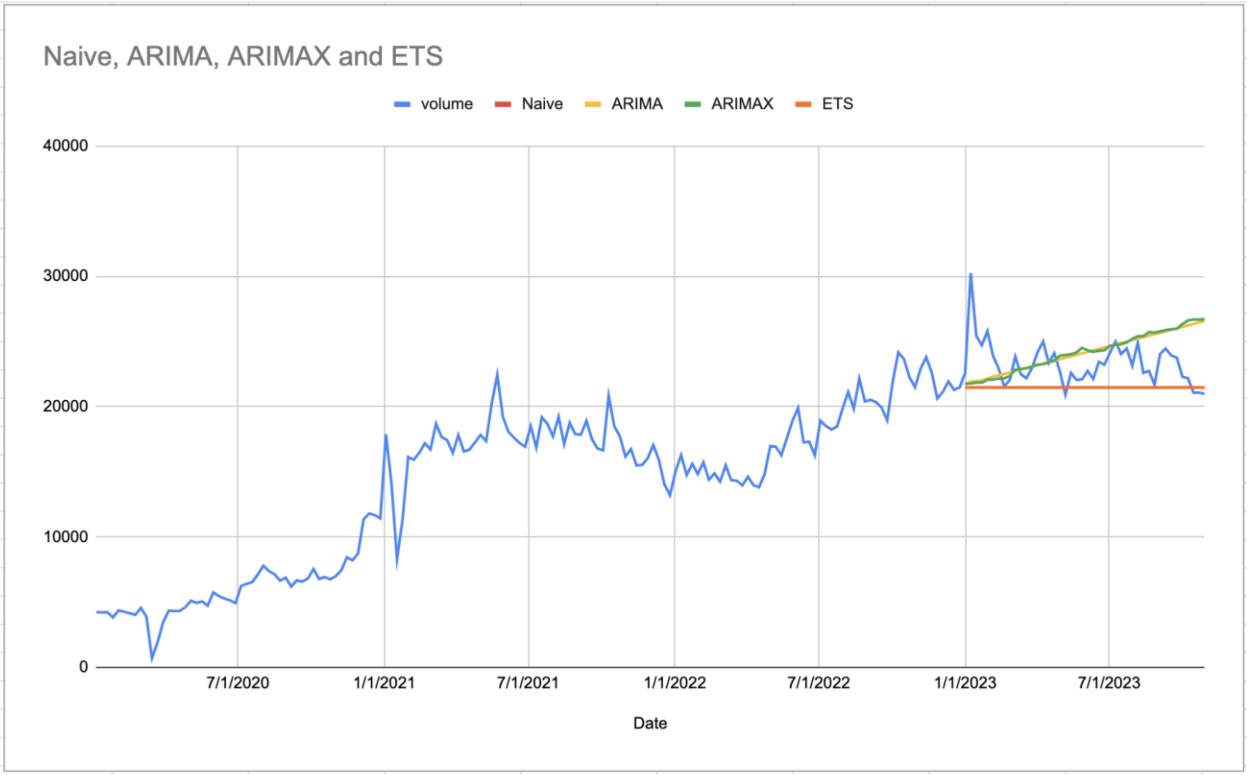


The autocorrelation function (ACF) shows a strong positive correlation at lag 0 (1.000), indicating a perfect correlation with itself. As the lag increases, the autocorrelation gradually decreases, suggesting a positive correlation between observations and their previous values. The values remain positive throughout, indicating a consistent positive relationship between the time series and its lagged values at different time points.



The partial autocorrelation function (PACF) for the 'fcst_ts_annual' time series indicates significant correlations at the initial lags, especially at the first lag (0.965) and a few subsequent lags. The PACF values then become less significant as the lag increases, with occasional fluctuations. This suggests a potential autoregressive (AR) component in the time series, particularly at the first lag.

4.1.2 Main Analysis



MODEL	DATA SPLIT	MAPE	RMSE	AIC
Naive	Training set	10.606627	1464.891	Since the Naive method doesn't involve estimating parameters or selecting a model, it doesn't have an AIC.
	Testing	7.865943	2450.915	
ETS	Training set	10.237256	1419.979	3058.398
	Testing	7.915841	2461.794	
ARIMA	Training set	10.094787	1322.626	2679.358
	Testing	8.630959	2645.74	
ARIMAX	Training set	10.2086	1321.669	2681.131
	Testing	7.825297	2478.214	

Naive, ETS, ARIMA, and ARIMAX models were evaluated based on their performance metrics, including Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), on both training and testing sets. The Naive model, which relies on the last observed value for forecasting, serves as a simplistic benchmark. In the training set, it demonstrated a MAPE of 10.61% and an RMSE of 1464.89, while AIC was not applicable, as the Naive method doesn't involve parameter estimation since it just simply uses the most recent observation as a forecast. In the testing set, the Naive model achieved a MAPE of 7.87% and an RMSE of 2450.92.

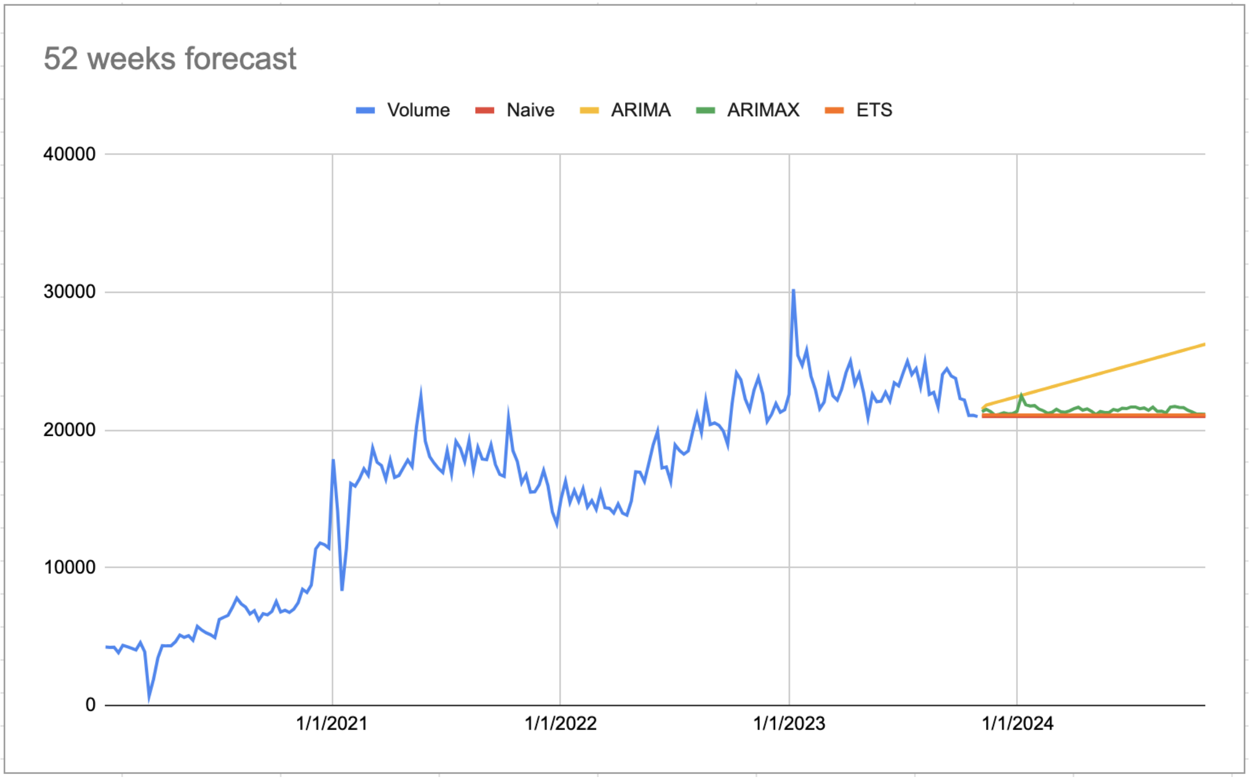
The ETS (Error, Trend, and Seasonality) model shows competitive performance in both the training and testing sets. In the training set, ETS achieves a MAPE of 10.24%, an RMSE of 1419.98, and an AIC of 3058.40. These metrics indicate its ability to capture and model the underlying patterns in the historical data. During testing, the ETS model shows a MAPE of 7.92% and an RMSE of 2461.79. These findings emphasize the effectiveness of the ETS model in predicting future data, demonstrating its usefulness in forecasting time series patterns which is valuable for its thorough handling of error, trend, and seasonality.

The ARIMA model, which captures autocorrelation and seasonality in the data, outperformed the Naive model. In the training set, it achieved a MAPE of 10.09% and RMSE of 1322.63. In the testing set, the ARIMA model exhibited a MAPE of 8.63% and an RMSE of 2645.74 while AIC evaluation is 2679.36.

The ARIMAX model, an extension of ARIMA that incorporates exogenous variables, shows competitive accuracy. In the training set, it achieved a MAPE of 10.21% and an RMSE of 1321.67. For the testing set, the ARIMAX model achieved a MAPE of 7.83% and an RMSE of 2478.21 while AIC evaluation is 2681.131.

Overall, both ARIMA and ARIMAX outperformed Naïve in MAPE metric, with ARIMAX showing a slight advantage in testing accuracy. It shows that the exogenous variable that represents launches and events, greatly affects the behavior of the model. As a general interpretation, the lower the values, the more accurate the model is.

4.1.3 Results based on the Objectives



In the plot above, forecasting analysis was conducted using the Naive, ETS ARIMA, and ARIMAX models to predict trends/forecast over the next 52 weeks. The Naive model, serving as a baseline, simply relies on the last observation thus not a good sole basis. In contrast, the ARIMA model demonstrates an upward forecast, raising concerns about potential over forecasting in the long run. This could have implications for staffing plans, especially in a contact center setting where excessive forecasts might lead to challenges in managing agent headcounts.

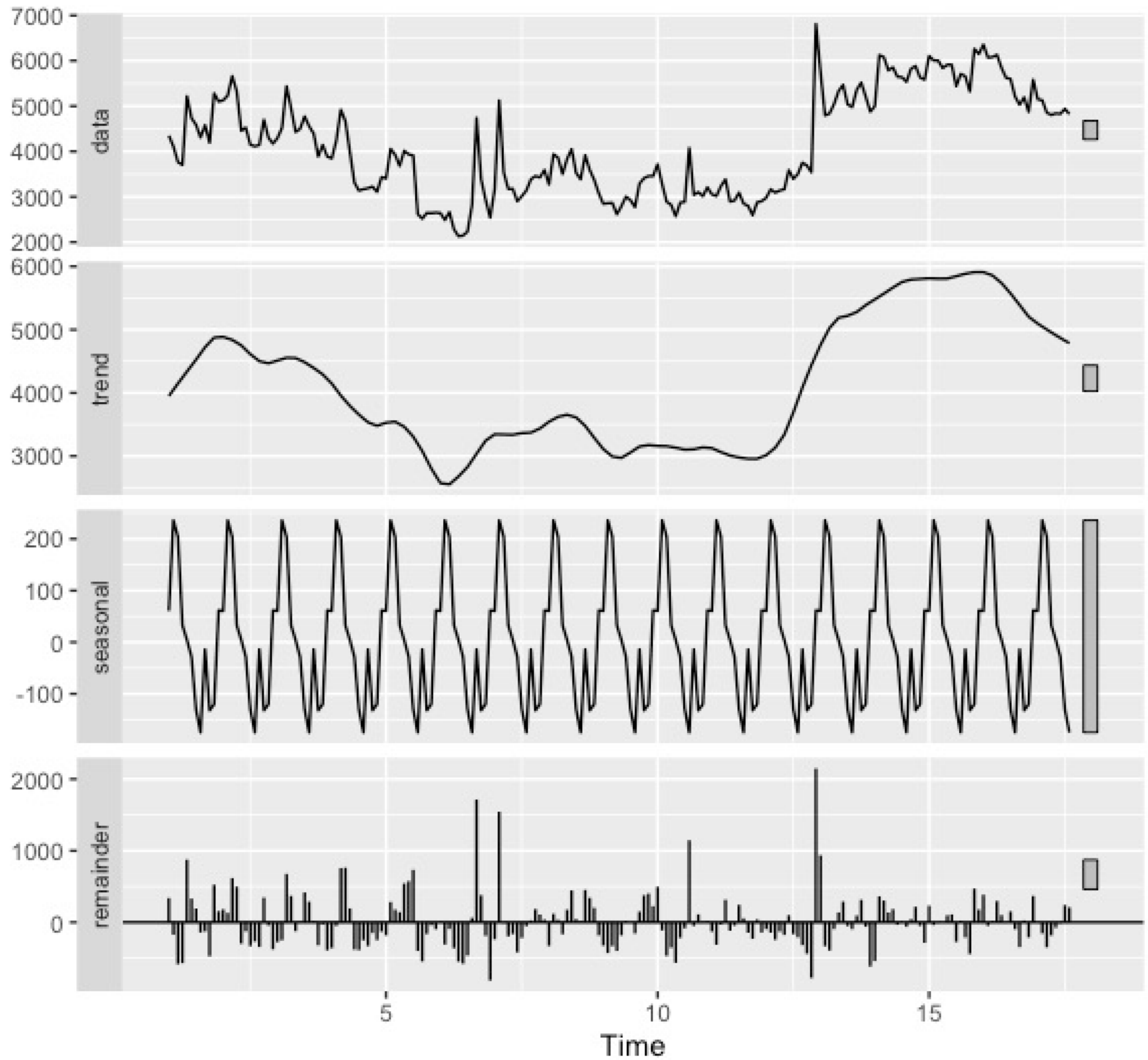
In the context of headcount calculations in workforce planning, maintaining an overly high forecast could result in a rapid increase in agent numbers and might oversee other languages of support. While this may seem initially advantageous, it indicates risks during potential ramp-down periods for YouTube vertical and campaigns. The ARIMAX model, incorporating exogenous variables such as launches and events, emerges as a more favorable choice. By considering external factors, this model facilitates a more gradual approach to ramping up and down agent headcounts. Connecting to the current events right now, tech companies such as Google, Microsoft and Amazon pronounced consecutive layoffs due to headcount reduction and cost cutting.

In the end, adopting the ARIMAX model aligns with optimal workforce planning for English chat support. It ensures a more accurate designation of agents and supports a well-structured, week-by-week headcount plan based on the anticipated volumes. This approach enhances adaptability to fluctuations in demand and enables a more balanced and efficient staffing strategy, addressing the challenges associated with both high and low forecasting situations.

4.2 Results - English Email

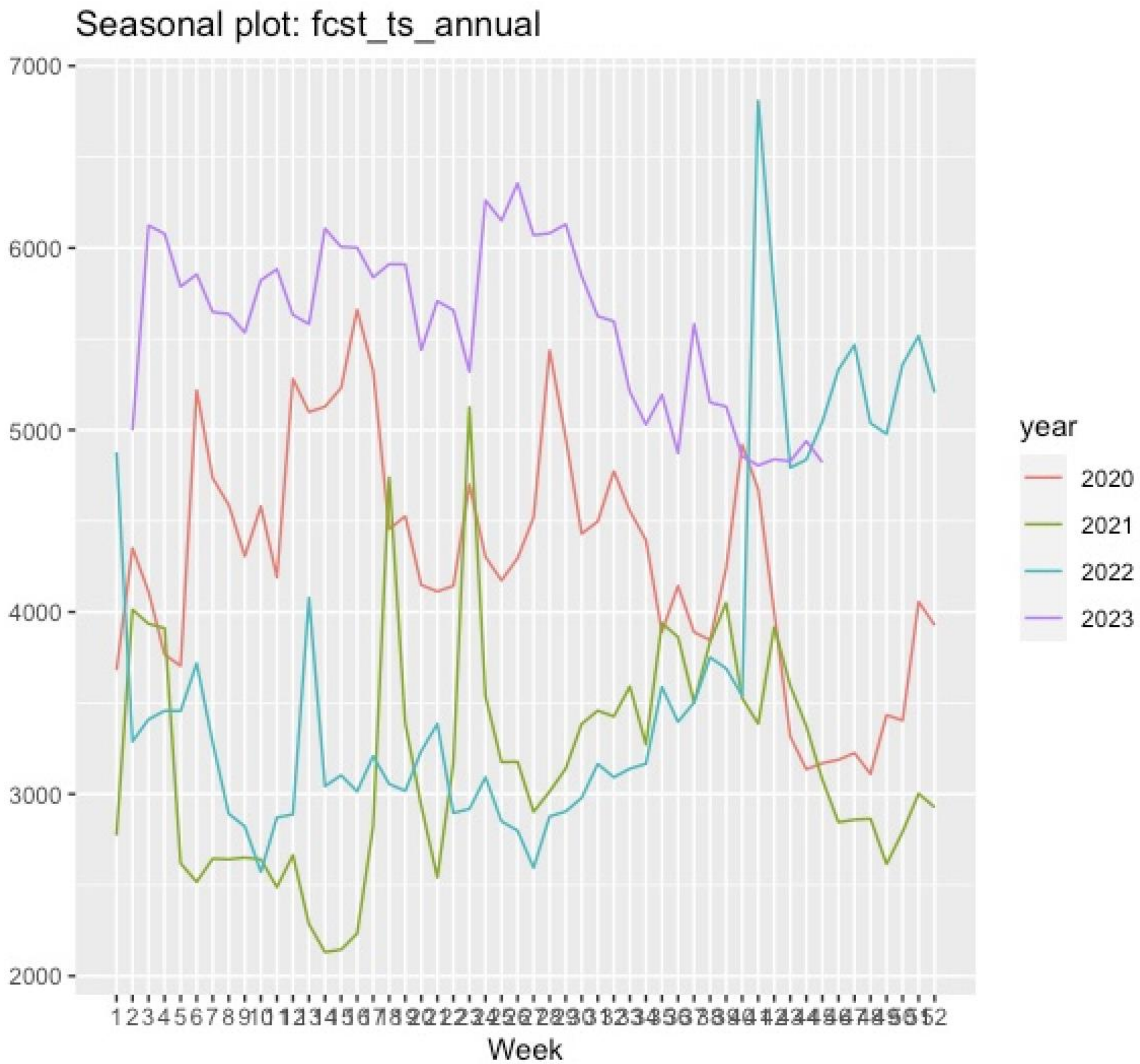
This section covers the analysis for English email volume. This includes the time series decomposition, training and testing , and the evaluation of the models.

4.2.1 Exploratory Data Analysis

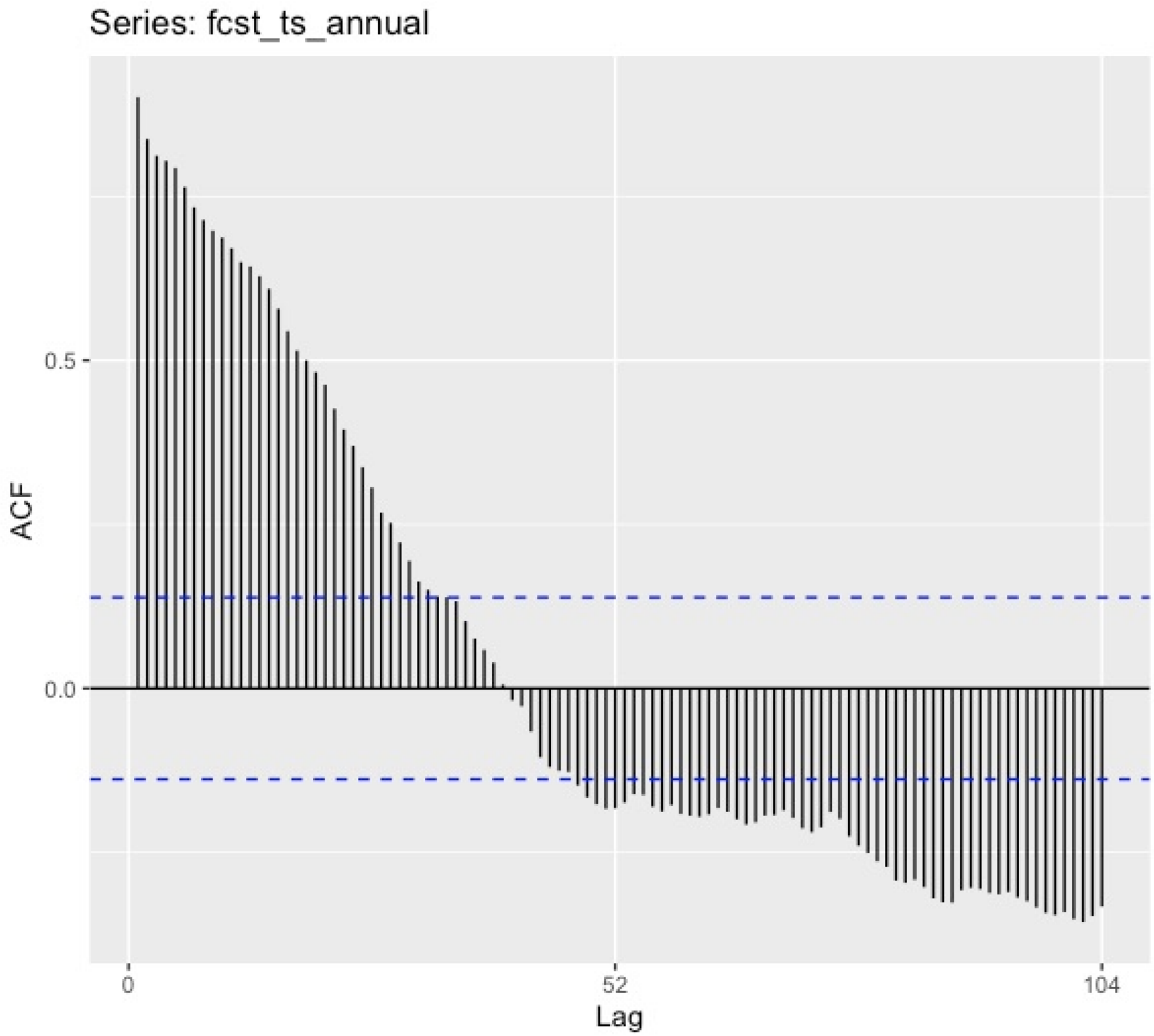


The graph above shows fluctuations in the data volume with varying seasonal patterns. Interestingly, July consistently shows a negative seasonal effect, reaching -174.87, indicating a regular decline, while January consistently shows a positive seasonal effect, reaching 60.45, indicating a regular increase. The trend component reveals the overall trajectory of the data shows a rising trend. For instance, trend values increase from 4147.614 to 5909.545 between March of year 1 and December of year 15, indicating a consistent rise in the data over the observed period of time.

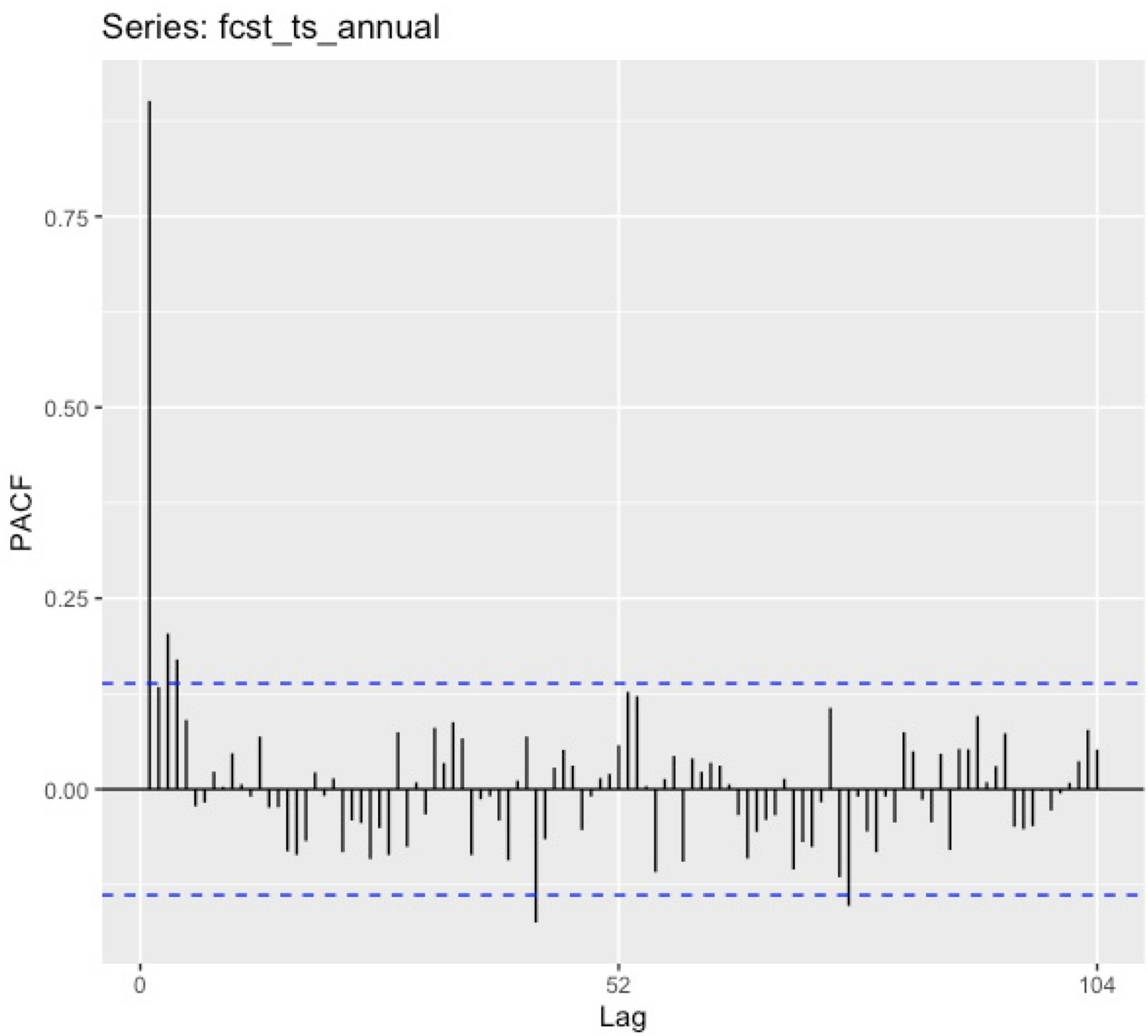
On the other hand, the remainder, or outlier component, captures unexplained fluctuations in the data. Positive or negative remainder values highlight periods with unexpected variations not explained by the seasonal and trend components. Also, May of year 5 shows a notably high remainder at 543.996, signifying an unexpected increase, while August of year 8 has a remainder of -399.572, indicating an unexpected decrease.



The seasonal plot indicates recurring patterns in weekly volumes, with consistent peaks and troughs (dip and spikes). End-of-year peaks suggest a seasonal surge.
For example, the weekly volume spike in the week of October 2, 2022, at 6809, stands out as a notable datapoint, reflecting a significant deviation from the usual trend in the data. The overall trend shows a gradual increase, with occasional deviations while outliers may be influenced by external factors too. As observed, periods of changing seasonal behavior are notable particularly in late 2022 and early 2023.

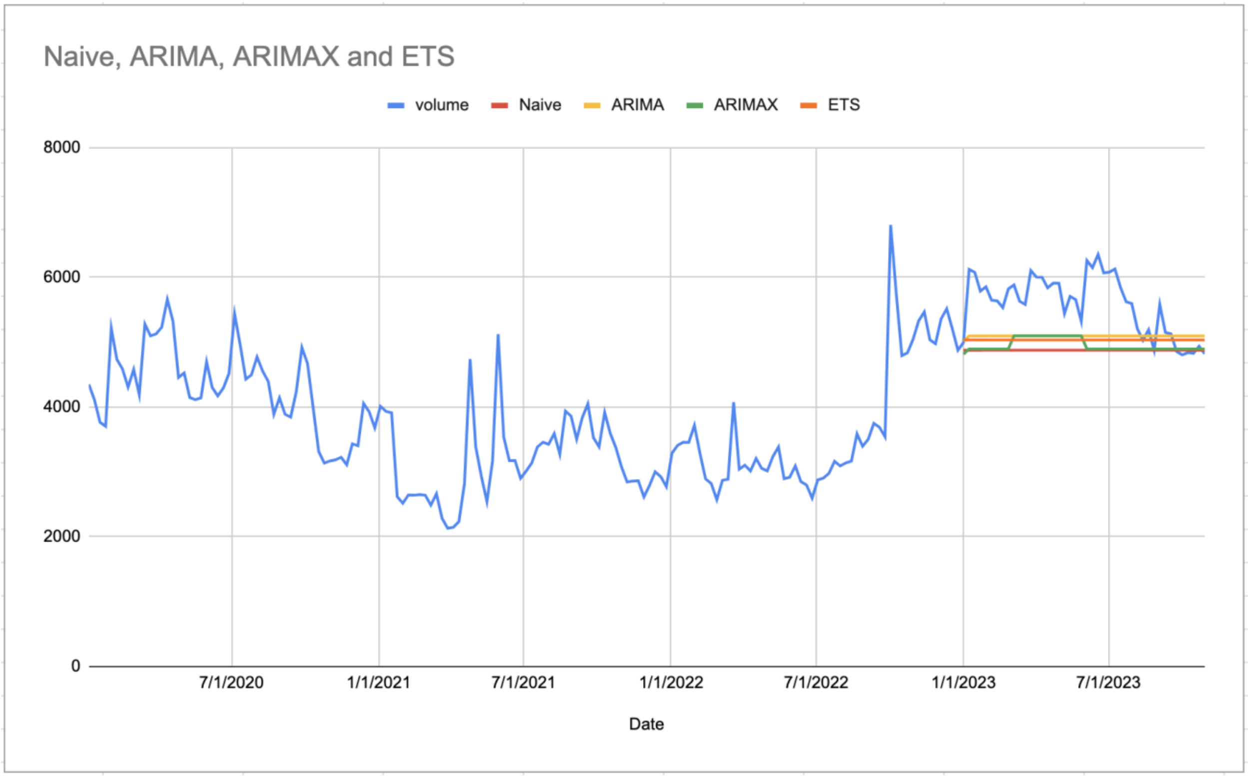


Same from the behavior of the chat data, the autocorrelation function (ACF) shows a strong positive correlation at lag 0 (1.000), indicating a perfect correlation with itself. It also shows significant autocorrelation peaks at lags of 52, indicating a strong annual seasonality in the weekly dataset. This means there's a recurring pattern every 52 weeks. No significant autocorrelation is observed at other lags, emphasizing the pronounced yearly cycle. For example, a peak at lag 52 implies correlation between values at the same week in different years.



The partial autocorrelation function (PACF) for the 'fcst_ts_annual' time series displays a significant spike at lag 1, indicating a strong partial autocorrelation at this lag. This suggests that the current week's observation is highly dependent on the observation from the previous week. Other lags do not show significant partial autocorrelation, emphasizing the immediate week's influence on the data.

4.2.2 Main Analysis



MODEL	DATA SPLIT	MAPE	RMSE	AIC
Naive	Training set	8.513672	539.2571	Since the Naive method doesn't involve estimating parameters or selecting a model, it doesn't have an AIC.
	Testing	12.3363	844.3162	
ETS	Training set	8.621981	520.1485	2745.061
	Testing	10.508928	844.3162	
ARIMA	Training set	8.334973	502.3315	2375.145
	Testing	9.868369	669.414	
ARIMAX	Training set	8.615413	499.6411	2375.048
	Testing	11.322857	779.8292	

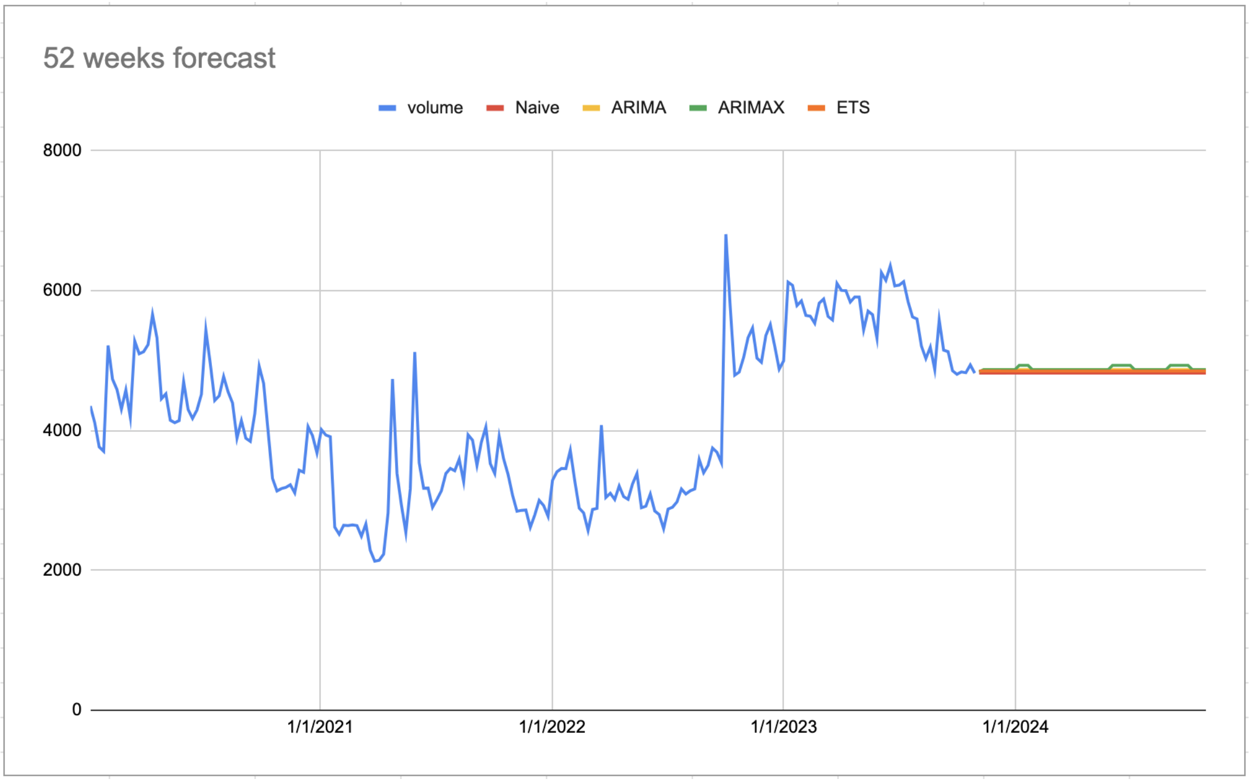
The evaluation results reveal the forecasting performance of four models: Naive, ETS ARIMA, and ARIMAX. In the training set, the Naive method yields a Mean Absolute Percentage Error (MAPE) of 8.51% and a Root Mean Squared Error (RMSE) of 539.26, while in the testing set, the MAPE is 12.34% and the RMSE is 844.32. Since the Naive method doesn't involve estimating parameters or selecting a model, it doesn't have an AIC.

The ETS (Error, Trend, and Seasonality) model, which captures the components of error, trend, and seasonality, which is particularly effective in modeling historical data, shows competitive performance. In the training set, ETS achieves a MAPE of 8.62%, and RMSE of 520.15, and an AIC of 2745.06. In the testing set, ETS maintains its competitive edge with a MAPE of 10.51% and an RMSE of 844.32.

The ARIMA model, applied without external variables, achieves a training set MAPE of 8.33%, RMSE of 502.33, its testing set shows a MAPE of 9.87% and RMSE of 669.41 while Akaike Information Criterion (AIC) of 2375.15.

The ARIMAX model, incorporating exogenous variables, shows a training set MAPE of 8.61%, RMSE of 499.64, its testing set demonstrates a MAPE of 11.32% and RMSE of 799.83 while the AIC is 2375.05. Comparatively, both ARIMA and ARIMAX outperform the Naive and ETS method on the testing set in terms of RMSE metric. As an evaluation, the lower AIC for ARIMAX suggests a better model fit compared to ARIMA. Overall, ARIMAX emerges as the more accurate forecasting model, especially when considering external factors.

4.2.3 Results based on the Objectives



The results of the analysis shows that email behavior across four models indicates slight variations in their patterns. However, an assessment done leads to the conclusion that the ARIMAX model still outperforms the other three as shown on the accuracy metrics in the model evaluation. There are minimal spikes and dips, inherited from the training data, and its forecasted values closely aligns with the slightly lower previous data points. It also shows that email data behavior is quite stable and implies a calm and manageable nature, with no significant adjustments required. The observation also highlights lower volumes in email compared to chat, which aligns with the reasonable assumption that the resolution turnaround time for email tickets is slower than chat. This also indicates that there is a stable staffing for agents handling emails and that can help them allocate time with other training that is necessary for upskilling. Moreover, stable volumes ease staffing concerns.

5. Summary and Conclusion

This paper presents time series models that enhance accuracy and efficiency of workforce planning by leveraging structured contextual data like events and launches, removing the need for subjective judgment.

In this study, we have demonstrated time series models for forecasting volumes of complaints for YouTube submitted through emails and chat. The proposed model and methodology were applied to prepare forecasts for chat and email volume ticket data of 200 weeks, from Jan 2020-October 2023. Data includes YouTube subscribers with English as the main language.

Our analysis revealed a significant disparity in ticket volume, with email exhibiting notably lower levels compared to chat. This finding suggests a potential association between communication channel and user behavior. With the slower expected turnaround time associated with email potentially discouraging user preference, leading to a preference for the faster and real-time nature of chat interactions.

The models considered in this paper are Naïve, ETS, ARIMA and ARIMAX. While some variations in behavior patterns were observed across the four models employed, a comprehensive evaluation based on accuracy metrics revealed that the ARIMAX model consistently outperformed the others. We find that adding exogenous variables like launches and events improves the predictability of the model.

Analysis reveals a high degree of stability in email data behavior, suggesting a predictable and manageable forecasting environment, potentially requiring minimal adjustments to established models. Stable email ticket volume presents an opportunity for efficient resource allocation, enabling agents to dedicate time to essential upskilling and training initiatives, ultimately enhancing their capabilities.

6. Recommendation

For future research direction, extending to other forecasting models can be recommended and comparing this to the aforementioned models in this study. Neural Networks, ETSX (Error, Trend and Seasonal with Exogenous variables) method and Long Short-Term Memory (LSTM) are good candidates for future study.

Enriching the dataset with further data points, such as complaint type, problem category, resolver group, timestamps, and even resolution times, could significantly enhance the granularity and comprehensiveness of the analysis.

Another research issue may be to investigate the effects of the forecasting model holistically in the company. This may be correlated with developments of other solutions for the customers, such as chatbots and sentiment analysis on chats / emails for better handling of complaint cases.

Lastly, there is also a potential for identifying and understanding the customer satisfaction after the deployment of the model. This may be helpful in scoping the full impact of the forecast models as this will incorporate the view of customers, in terms of waiting time or issue resolution, rather than the view of the company and employees only. Given that the objective of workforce planning is also to provide better services / resolutions to customers / users given the adequate manpower in handling these ticket issues.

7. References

[1] Cotten, A. (2007). Seven Steps of Effective Workforce Planning, Human Capital Management Series, pp. 9 – 23

[2] Girotra, R. (2022). Workforce Planning using Time Series [Master of Science in Computer Science - Data Science dissertation, Trinity College Dublin]. https://publications.scss.tcd.ie/theses/diss/2022/TCD-SCSS-DISSERTATION-2022-070.pdf

[3] Ibrahim et al (2016). Modeling and Forecasting Call Center Arrivals: A Literature Survey. https://www.iro.umontreal.ca/~lecuyer/myftp/papers/arrivals-survey15.pdf

[4] Anggraeni,W. Vinarti, R., Kurniawati, Y. (2015). Performance Comparisons Between ARIMA and ARIMAX Method in Moslem Kids Clothes Demand Forecasting: Case Study. Procedia Computer Science. Volume 72, 2015, pp. 630-637

[5] Rabbi, F., Tareq, S., Islam, M., Chowdhury, A., Kashem, M. (2020). A Multivariate Time Series Approach for Forecasting of Electricity Demand in Bangladesh Using ARIMAX Model. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI).

[6] Koutsandreas, D., Petropoulous, F., Spiliotis, E., Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. Journal of the Operational Research Society.

[7] Fildes, R., & Goodwin, P. (2007). Against your better judgment? how organizations can improve their use of management judgment in forecasting. Interfaces, 37(6), 570–576.

[8] Kolassa, S., & Martin, R. (2011). Percentage errors can ruin your day (and rolling the dice shows how). Foresight: The International Journal of Applied Forecasting, 23, 21–27.