CONTACT CENTER EFFICIENCY

A Research Project Report
Presented to
The Department of Mathematics and Physics

by

SIBUSISO MBATHA
STUDENT NO. 221775846

As an Assessment for the Module
MATHEMATICAL SCIENCES PROJECT 3 (MSP360S)
Within the Qualification
DIPLOMA IN MATHEMATICAL SCIENCES

ACADEMIC SUPERVISOR: Ms. T Wannenberg
COMPANY MENTOR(S): Mr. Y Samsodien & Mr. E Brits

Cape Peninsula University of Technology
November 2023

DECLARATION

I, Sibusiso Mbatha, declare that the contents of this research report represent my own work. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own. Each contribution to, and quotation in, this report from the work(s) of other people has been attributed, has been cited, and referenced.

I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

Click on the box to confirm your agreement: ☒

Date of Declaration: 03 November 2023

ACKNOWLEDGEMENTS

I want to acknowledge Ms. T Wannenberg for the most contribution in this project. RCS group provides me the data to work with that why I am acknowledging RCS for the great contribution in this project. I would like to thank my best manager and my mentors for their great contribution to this project to be successful. I would like to thank my family and friends for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ARIMA= Autoregressive integrated moving average.

ADF= Augmented Dickey-Fuller.

ACF = Auto Correlation Function.

PACF= Partial Autocorrelation function.

AIC = Akaike Information Criterion.

GLOSSARY OF TERMS

ARIMA= AN AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL IS A STATISTICAL ANALYSIS MODEL THAT LEVERAGES TIME SERIES DATA TO FORECAST FUTURE TRENDS.

ADF =In statistics, an augmented Dickey–Fuller test (ADF) evaluates the null hypothesis that a unit root is present in a time series sample.

AIC = is the mathematical method for assessing how well a model fits the data from which it was generated.

ABSTRACT

This research aims at modeling and forecasting the number of transfer calls to the RCS Group Company. from 31 January 2021 to 31 October 2023. Data used in this research show a decreasing variance coupled with trend and seasonal variations. According to the dataset we have python software package that the best ARIMA model is (5,1,5) (0,0,0)12. The seasonal shows that in the trend there is a decreasing pattern. The chosen model passed the major diagnostic statistical tests and showed high accuracy performance in modelling the data. The forecasts were made at various confidence levels (95 per cent). According to the predicted values RCS must have more agents during the week which can maintain the calls/ transfer calls so that the customer does not wait a long time to be assisted.

1. INTRODUCTION

1.1 Background

Contact centers, originally known as call centers, began to gain prominence in the mid-20th century as businesses recognized the need for centralized communication hubs to manage customer inquiries and support. Early call centers relied heavily on manual processes. Agents took customer calls, wrote down information on paper, and manually routed calls to the appropriate departments. The late 1990s and early 2000s saw the integration of technologies like Interactive Voice Response (IVR) systems, Customer Relationship Management (CRM) software, and Automatic Call Distribution (ACD) systems. These technologies aimed to streamline operations and improve efficiency. As customer preferences diversified, contact centers evolved to provide multichannel and omnichannel support, allowing customers to interact via phone, email, chat, social media, and more. The integration of these channels posed new challenges and opportunities for efficiency. The COVID-19 pandemic accelerated the adoption of remote work and cloud-based contact center solutions. These changes have made it essential to ensure efficiency in a distributed work environment.

1.2 Research Problem

The aim of this research problem is the Credit Card CS skill which is experiencing a high volume of calls that should not actually go to their department. This is evident in the transfer rate being above 50%. The actual aim of forecasting is to predict how many agents are needed every single day.

1.3 Research Questions

- Forecast how many agents are needed per day to maintain the transfer calls.
- Which month/ days experience the most transfer calls?

1.4 Research Objective

Develop Arima model which can help us to forecast the future transfer calls and develop the schedule of Agents.

1.5 Research Hypotheses

The hypothesis of an ARIMA model is that the observed time series data can be effectively modeled and forecasted by considering a combination of its own past values, differing to achieve stationarity, and modeling the error terms.

1.6 Significance of Research

By using ARIMA models to forecast transfer calls, contact centers can better allocate resources, optimize staff schedules, and reduce the operational costs associated with handling transfers. This leads to increased efficiency and cost saving.

2. LITERATURE REVIEW

Recent improvements in customer service call centers have improved service quality, reduced operating costs, and increased sales income (Brett H, 2019). The innovations include modern technology, new data sources and new predictive analytics methodologies. By conducting three empirical investigations of caller behavior under call center improvements, we investigate how business might utilize some of these advances in this dissertation (Brett H, 2019). It starts back in the main offices, the happiness of employees has incredibly good impact on productivity of the company (Clement S and George W, 2021). The combination of contact center design and environment factors to offer a quasi-experimental examination of how pleasure affects output. We discover evidence of a beneficial effect on sales performance, which is driven by changes in labor productivity primary through workers converting more calls into sales and to a lesser extent by making more calls per hour and keeping a more rigid schedule (Clement S and Goerge W, 2021). Call center agents often struggle with the performance paradox of being courteous, efficient, and effective. To gain insight into how agents address this performance paradox, we carried out a hybrid-methods study of call center agents at a major financial services organization using a discourse analysis of calls. They analyzed agents' moves, they identified to discourse moves that explain two new ways to approach customer service in the contact center: solidarity building, and conversation control (Colin M, Ulrike M, and Priscilla S, 2016). Contact center operations should not run at occupancy levels that are extremely close to 100%. Particularly, such practices may cause severe service degradation (for instance, because of inaccurate demand estimates), agent "burnout" and a rise in client attrition rates. To assess the economics of a trade-off between staffing expenses and customer service, we have offered a rule of thumb study (John C and Moshe B, 2001). To efficiently run the bank's workforce call center, the authors (D.K. Barrow) investigated time series forecasting techniques. The seasonal moving average (SMA) approach and artificial neural networks (ANNs) are utilized for systematic assessments and intraday call arrival forecasting, respectively.

3. METHODOLOGY

3.1 Data Source

The data is collected at RCS Group and is collected as a primary data source. The data is collected at the contact center department where the transfer call and the customer service take place.

3.2 Data Description

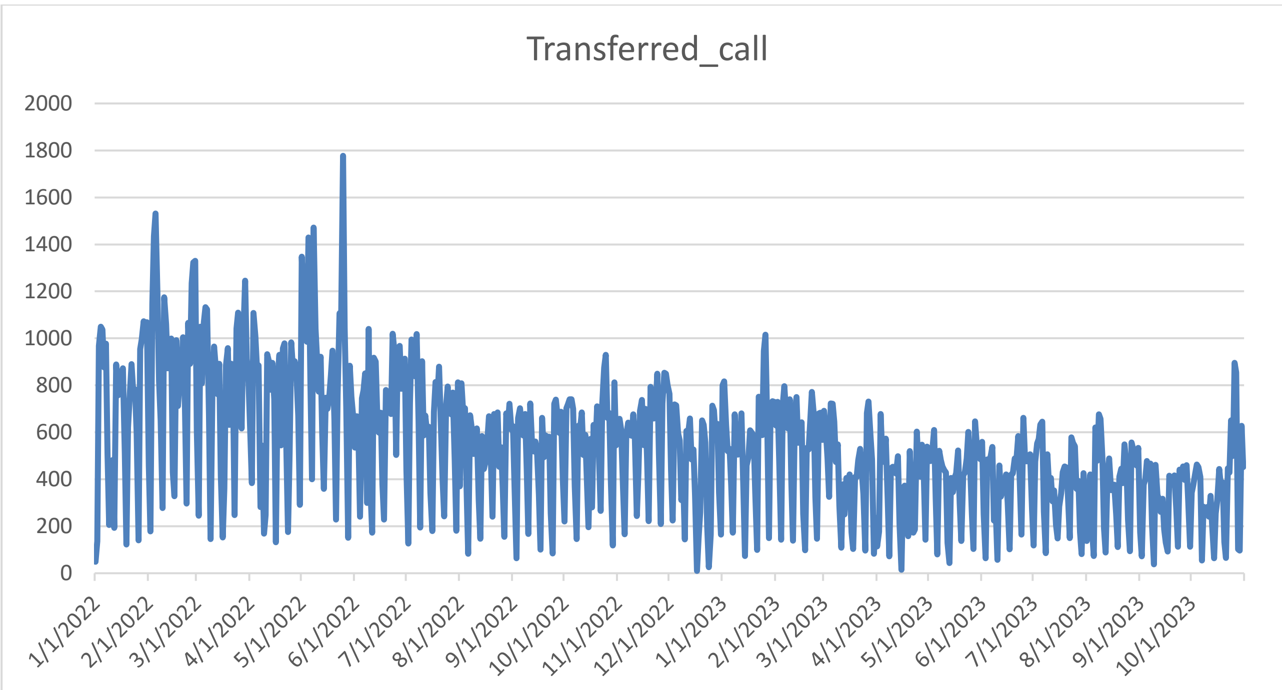

Figure 1: Data collected from RCS.

This graph is the sample data that was taken from 01January 2022 to 31 October 2023. The graph shows that the pattern is decreasing as time goes by.

3.3 Data Preparation

The RCS database was used to clean the data using SAS Enterprise Guide. We filter the data by selecting the calls that were transferred from the first call resolution agent to the other agent. We used the year from 2021 to2023. As Figure 1shown in the above graph.

3.4 Data Analysis

Daily Transferred calls data was obtained from the RCS group. The dataset is extracted from 31 January 2021 to 31 October 2023. The Daily data from 31January 2021 to 30 June 2022 were used as train dataset and used in developing a forecasting model and the data from 01 July 2023 to 31 October 2023 was then used as testing dataset for comparing the sample forecasted values with the actual values to check for the goodness of the model in fitting a given dataset. We used ARIMA model to the whole dataset, and the model was then used for out of sample forecasts. The Python and SAS software packages were used in the data analysis.

3.5 ARIMA Model Modeling Process

Autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and differential autoregressive moving average (ARIMA) model are famous time series forecasting methods proposed by Box and Jenkins in the early 1970s.

Time series model modeling is usually divided into the following steps:

(1) Stationarity test

(2) Stationary series are obtained by differential processing of non-stationary time series

(3) Determining model order

(4) Testing model significance

(5) Predicting future values of time series

(1) Stationary test

When modeling time series data, it is necessary to require that the series be stationary. In this research we used ADF test and P-value to test if our data is stationary or non-stationary.

(2) Establishing ARIMA model

To select the optimal model from several initially identified models, AIC criterion is used to determine the order, and the minimum model is the best model. In this paper, using Python software to calculate, it is found that when p=5, q=1, d=5, AIC value is the smallest, so ARIMA (5, 1, 5) model is established for the original sequence.

(3) Model test

It verifies the rationality of the model by using a white noise test on the residuals and provides a basis for determining whether the model is reasonable or not. If the model passes the white noise test, it is proved that the model is reasonable. If the model does not pass the white noise test, it is proved that the model is unreasonable and other models need to be searched.

4. RESULTS AND DISCUSSION

The descriptive statistics identified by the ARIMA model are shown in table 1 below. The minimum value of transferred calls in RCS per is 10 and the maximum value is 1777. The range of this dataset is 1767 which indicates that the values are not far apart. This will help when forecasting the future.

Count	Mean	Std	Min	25%	50%	75%	Max	Sum
669	541.83	283.57	10	121.14	528	1077.74	1777	362484

Table 1: Descriptive statistics

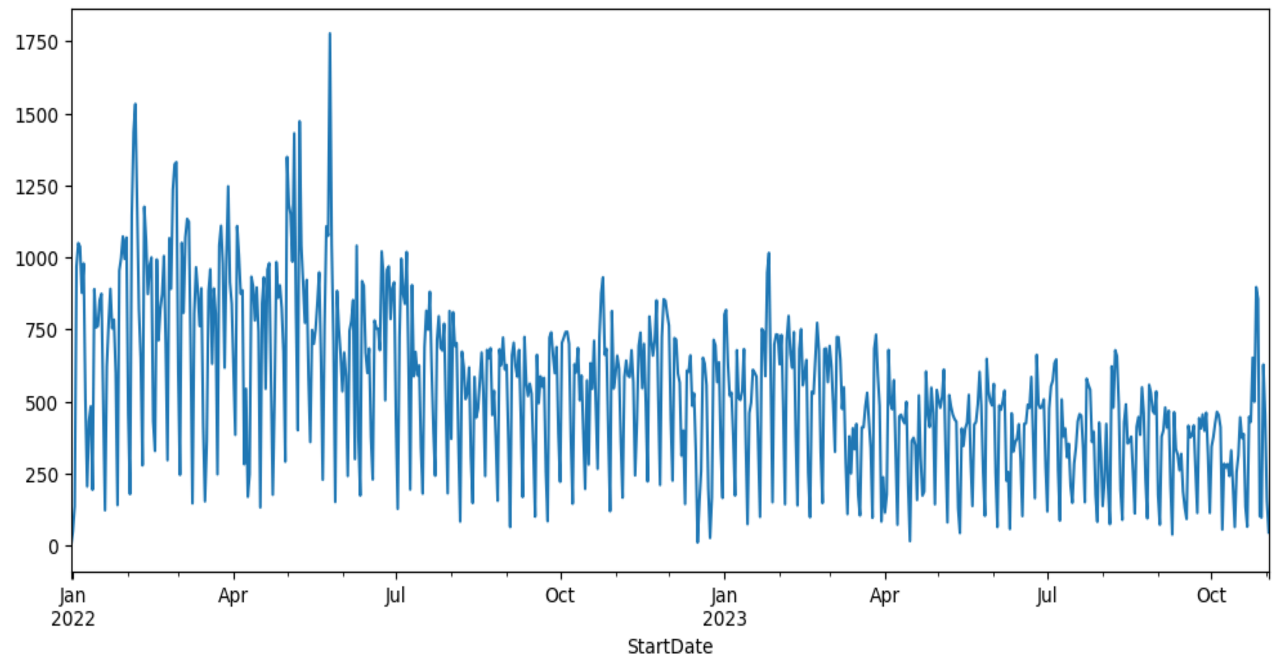

Figure 2: Sample Time Series Plot

The time plot of the Transfer calls in RCS from 31 December 2021 to 31 October 2023 in figure. This plot indicates that the data follows a decreasing pattern with decreasing variance and seasonality. The data was evaluated for stationarity using the Augmented Dickey-Fuller (ADF) test and the p-value of 0.633 which confirms that the data is not stationary.

Variable	ADF Test	1% Critical value	5% Critical value	10% Critical value	P-value	Results

Original sequence	-1.292	-3.44	-2.87	-2.57	0.633	Non-Stationary

Table 2 : Sample Data stationarity test results

The sequence's unit root is tested using the ADF test. It is shown that there is a unit root in the original sequence if the P value is larger than the confidence level alpha. Differentiating the sequence is therefore necessary to maintain its stability.

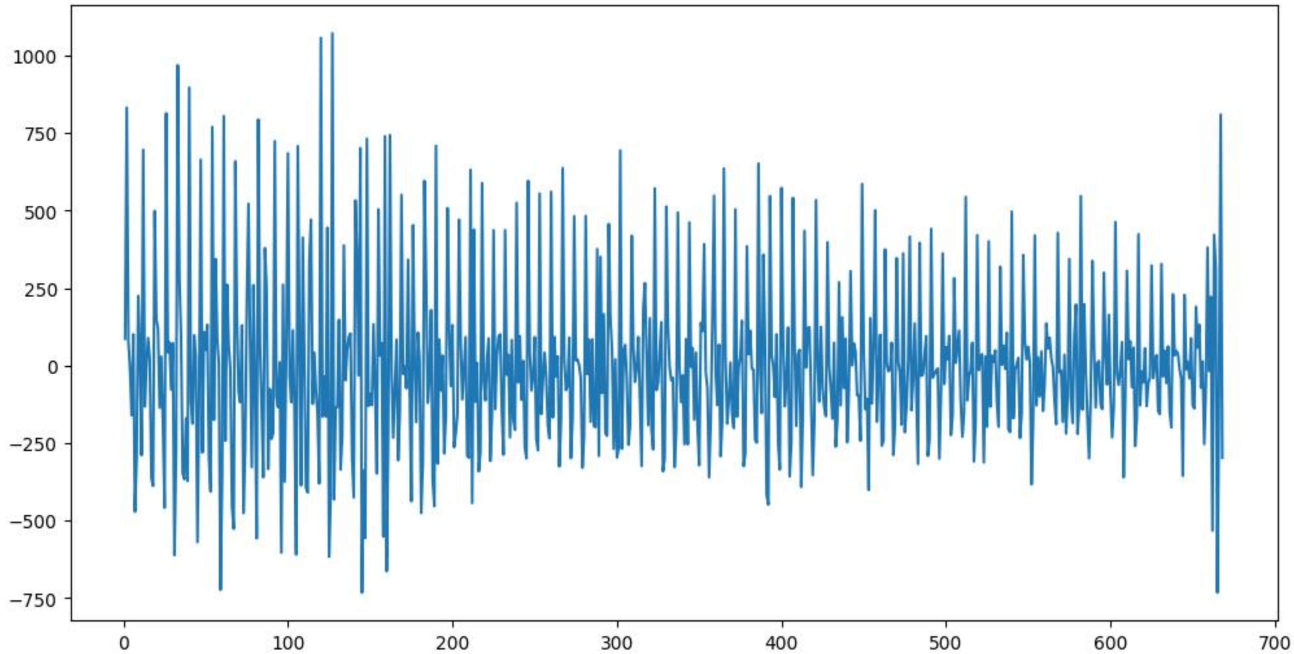

Figure 3: Time series after first-order difference

The above figure shows that the time series is now stationary. Looking at the graph it shows that there is no pattern or trend in the graph. The time series techniques, such as those used for forecasting, are based on the idea that the underlying patterns and relationships in the data are stable over time.

Variable	ADF Test statistics	1% Critical Value	5% Critical value	10 % Critical value	P-value	Results
Sequence (-1)	-10.671	-3.44	-2.87	-2.57	0.04165	Stationary

Table 3: First-order difference stationary test

The original sequence is non-stationary, but after first-order difference, it passes the stationarity test where the p-value is less than the confidence level alpha. At the same time, it can be seen from Figure 3 that the sequence value after first-order difference fluctuates up and down at 0, so the value of D in the ARMA (p, d, q) model is 1.

MODEL	**AIC**
ARIMA (0,1,1) (0,0,0)	9281.113
ARIMA (1,1,1) (0,0,0)	9222.881
ARIMA (2,1,1) (0,0,0)	9196.236
ARIMA (3,1,0) (0,0,0)	9352.797
ARIMA (2,1,2) (0,0,0)	9199.655
ARIMA (4,1,1) (0,0,0)	9183.457
ARIMA (5,1,5) (0,0,0)	**8946.970**

Table 4: Model fitting and adequacy checking.

The ARIMA (5,1,5) (0,0,0) was chosen as the best model due to its low value of AIC shown in Table 4. Consequently, to select the best ARIMA model for forecasting from the above suggested models, the lowest values in errors (AIC, BIC and /or MAE) considered.

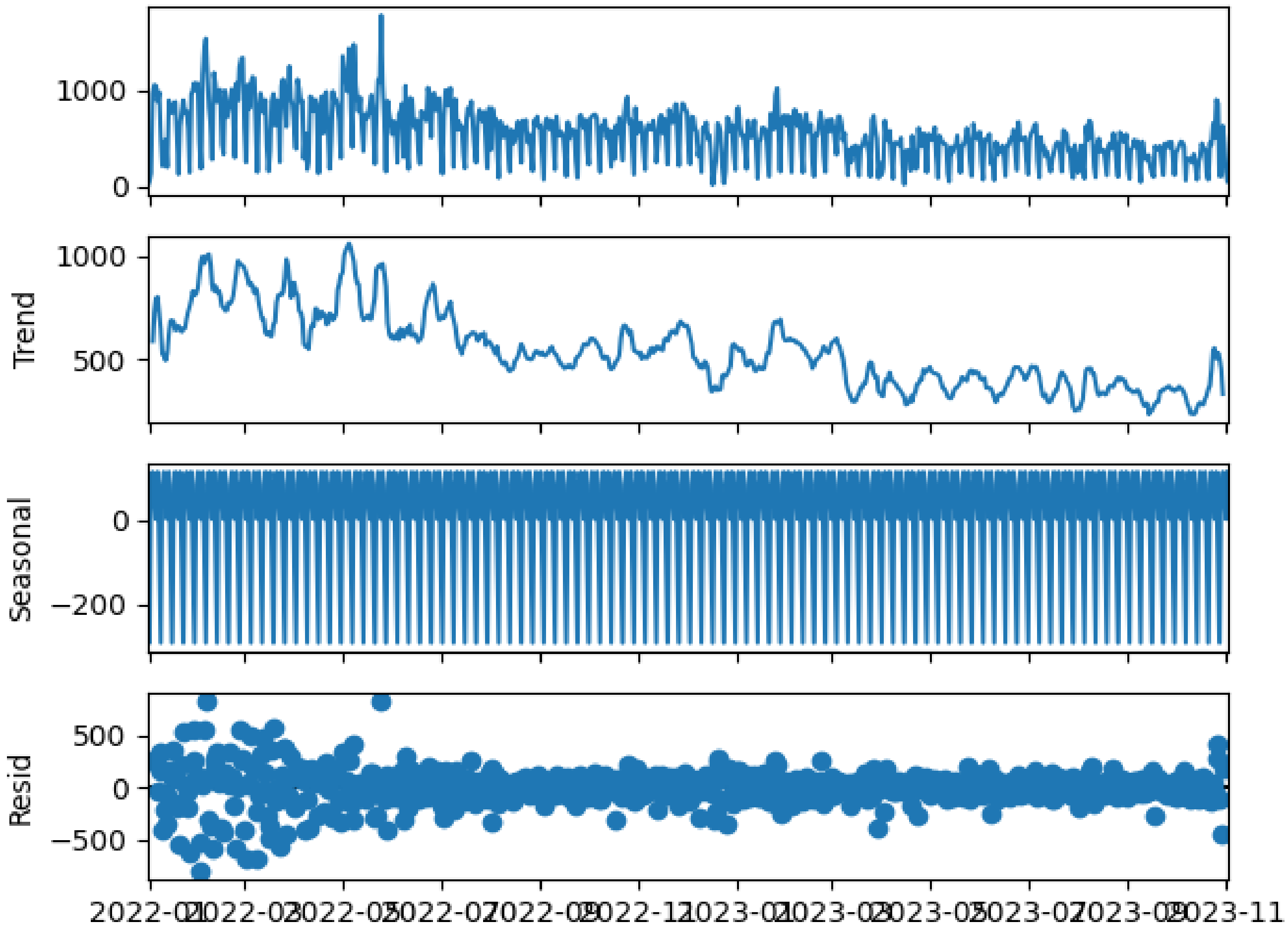

Figure 4: Decomposition of additive time series

The four-graph shown in the above figure shows the decomposition of the additive time series. The first (Top) plot Figure 4 is the original observed plot, second from top is the estimated trend in the transferred calls in RCS, third plot from the top is estimated seasonal factors and the bottom plot is the estimated irregular component in the series. In the trend it shows that the is a decreasing pattern in transfer calls.

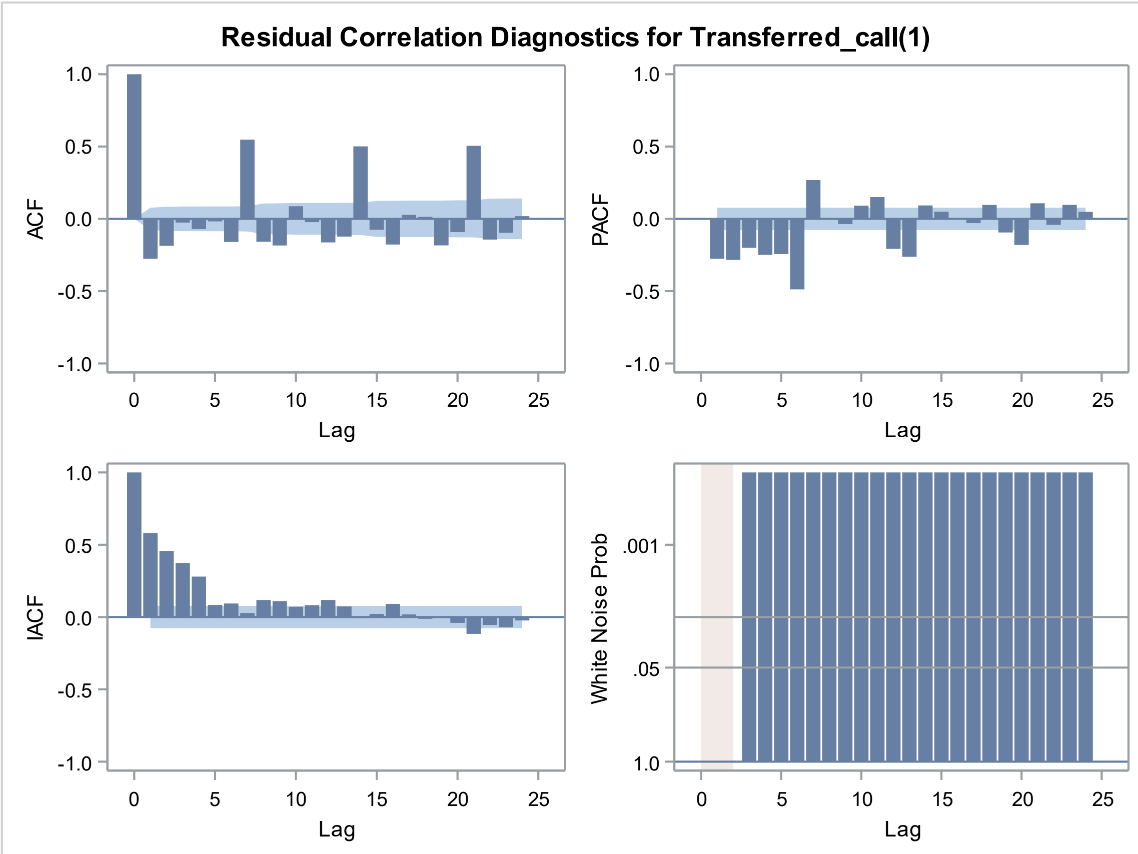

Figure 5: ACF and PACF of Transferred Calls

The ACF and PACF plot of the residuals indicate a white noise process despite the presence of many lags outside the shaded. The appearance of white noise, that is, the series is independent and identically distributed with mean zero and constant variance. As shown that ACF (Autocorrelation) is to be close to zero. The model is a good fit.

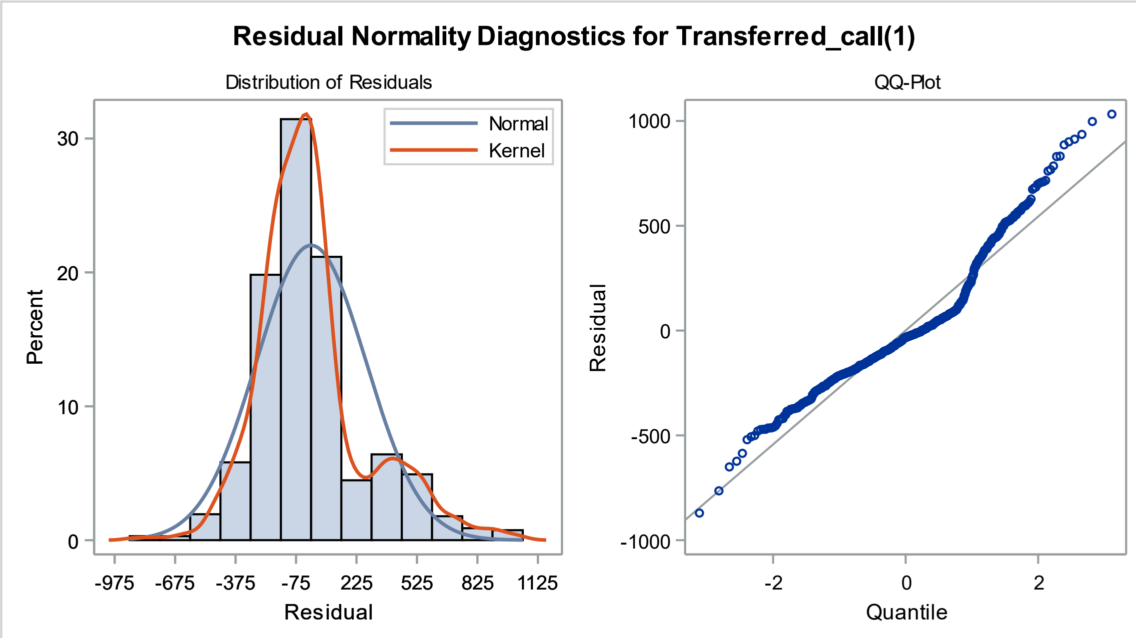

Figure 6: QQ and Histogram plot

In Histogram the Kernel Density Estimation (KDE) nearly overlaps with the Normal (blue curve). The results indicate that the residual has a normal distribution, with a mean of 0 and a standard deviation of 1. And the QQ Plot the residuals sequence of the model satisfies the normal distribution, and the D-W (Durbin Watson) test result value obtained by Python language is 2.21. when the D-W test is between 1.5 and 2.5 it shows that there is no correlation, which indicates the model is good.

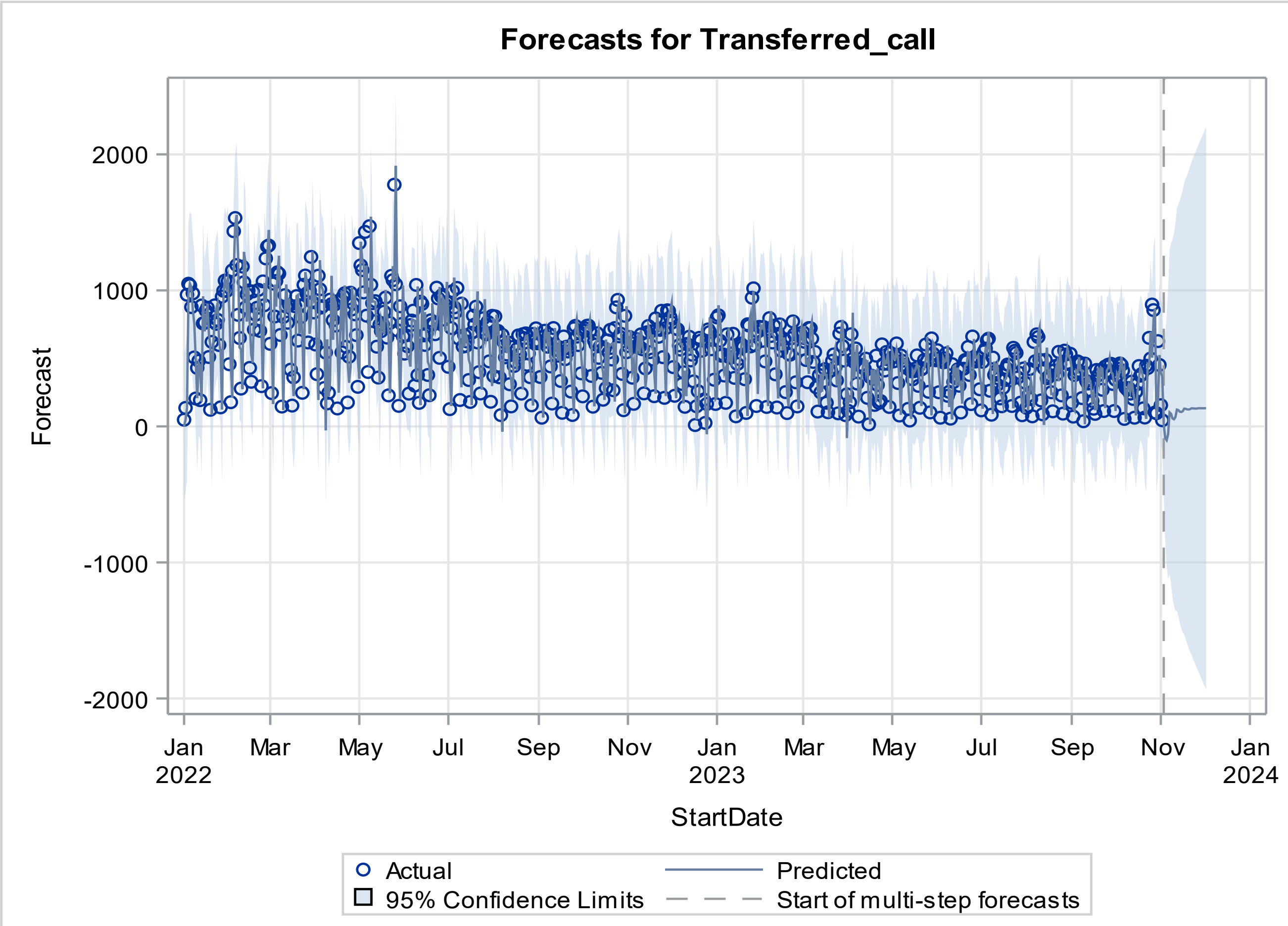

Figure 7: Plot of the Forecasted values

The plot of the predicted values in figure 7 indicates that the forecasted values can be used for decision making and formulation. The forecasted values from the model are behaving well.

Dates	Transfer calls	Forecast	STD	L95	U95	Residual
25OCT2023	499	694.7	272.14	161.33	1228.12	-195.72
26OCT2023	896	506	272.14	-27.58	1039.21	390.19
27OCT2023	855	786	272.14	252.52	1319.31	69.08
280CT2023	102	867	272.14	333.56	1400.35	-764.96
29OCT2023	97	52	272.14	-481.18	585.61	44..79
30OCT2023	627	165	272.14	-368.09	698.69	461.7
31OCT2023	452	523	272.14	-10.43	1056.37	-70.96

Table 5: Forecasted values against actual

L95 is the lower 95% confidence limit and U95is the upper 95% confidence limit. The forecasted values and actual values are quite close to each other as shown in the table above. The ARIMA model (5,1,5) is suggested as a good fit for this dataset. The test dataset was used to compare the forecasted values and actual values to see the fitness of the model in the sample dataset.

Forecasts for variable Transferred_call				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
670	395.5343	272.1455	-137.8612	928.9297
671	608.1381	384.8719	-146.1969	1362.4732
672	578.9075	471.3699	-344.9604	1502.7755
673	476.4049	544.2910	-590.3859	1543.1958
674	460.1845	608.5359	-732.5239	1652.8929
675	440.8800	639.9258	-813.3516	1695.1116
676	513.5650	669.8464	-799.3099	1826.4398
677	503.5716	698.4865	-865.4368	1872.5800
678	468.5280	725.9976	-954.4012	1891.4572
679	462.9826	752.5036	-1011.8974	1937.8626
680	456.3827	772.0445	-1056.7967	1969.5622
681	481.2323	791.1029	-1069.3009	2031.7655
682	477.8158	809.7128	-1109.1922	2064.8237
683	465.8351	827.9045	-1156.8280	2088.4981
684	463.9392	845.7050	-1193.6122	2121.4905
685	461.6828	861.4465	-1226.7212	2150.0869
686	470.1784	876.9054	-1248.5247	2188.8815
687	469.0104	892.0966	-1279.4668	2217.4875

Forecasts for variable Transferred_call				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
688	464.9144	907.0333	-1312.8382	2242.6670
689	464.2662	921.7280	-1342.2874	2270.8199
690	463.4948	935.6775	-1370.3995	2297.3891
691	466.3993	949.4222	-1394.4340	2327.2326
692	466.0000	962.9707	-1421.3878	2353.3878
693	464.5996	976.3311	-1448.9742	2378.1735
694	464.3780	989.5112	-1475.0284	2403.7845
695	464.1143	1002.3559	-1500.4671	2428.6957
696	465.1073	1015.0380	-1524.3305	2454.5451
697	464.9708	1027.5635	-1549.0168	2478.9583
698	464.4920	1039.9383	-1573.7495	2502.7336
699	464.4163	1052.1675	-1597.7941	2526.6266

Table 6: Future forecasted values of 30 days

The data in table above show that the forecasted values are remarkably close to the actual values, suggesting that ARIMA (5,1,5) $(0,0,0)_{12}$ as the good fit for this dataset. The model is fitted to the whole dataset and forecasts are now made outside of the sampling period. The predicted values are which are taken for 30 days from 01 Nov 2023 to 30 Nov 2023.

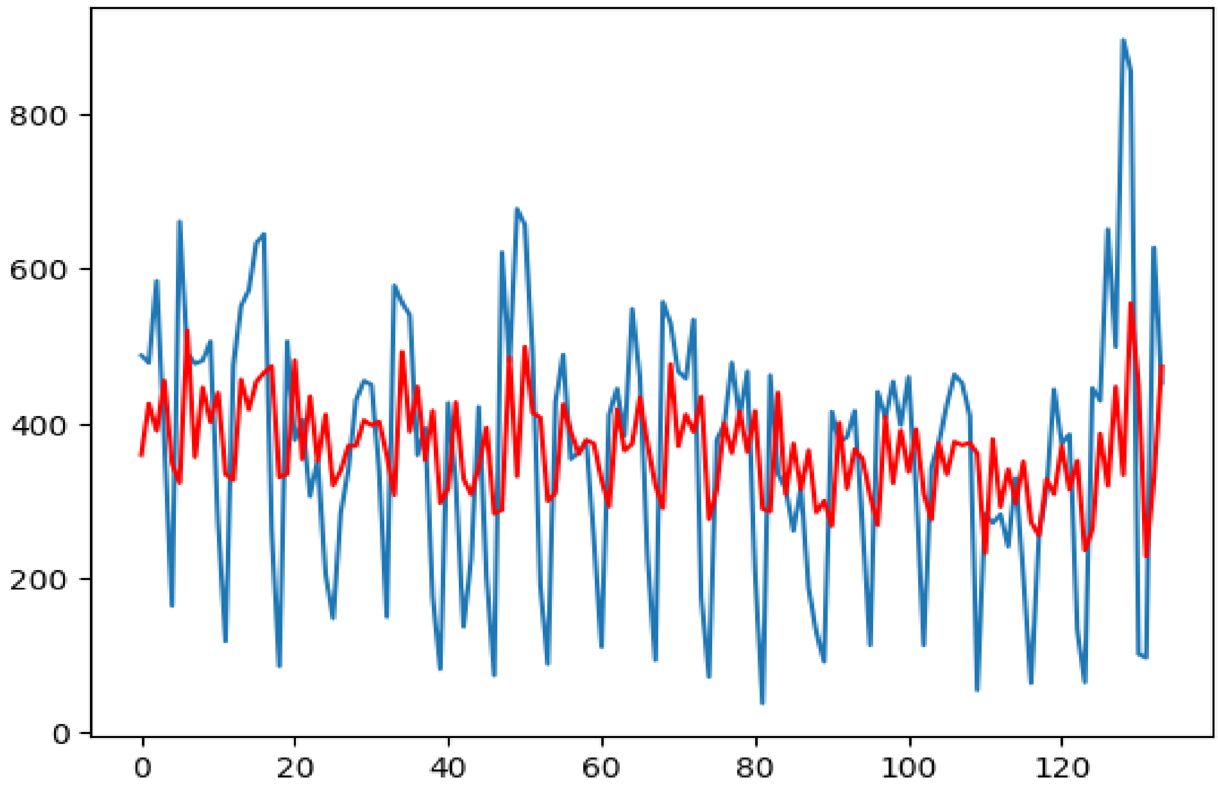

Figure 8: Forecasts against actual outcomes

The blue line in the graph shows the actual values in the test dataset we used, and the red line shows the forecasted values with the test dataset. This graph indicates that RCS group company is receiving less transfer calls as time goes by. On my suggestion RCS received many calls and transfer calls during COVID-19 where people were having to use calls to communicate with RCS regarding their account and loans.

5. CONCLUSION AND RECOMMENDATIONS

This research aims at modeling and forecasting the number of Transferred calls to RCS Group Company from 31 January 2021 to 31 October 2023. Data used in this research show a decreasing variance coupled with trend and seasonal variations. According to the dataset we have python software package that the best ARIMA model is (5,1,5) $(0,0,0)_{12.}$ The seasonal shows that in the trend there is a decreasing pattern. The chosen model passed the major diagnostic statistical tests and showed high accuracy performance in modelling the data. The forecasts were made at various confidence levels (95 per cent). According to the predicted values RCS must have more agents during the week which can maintain the calls/ transfer calls so that the customer does not wait long time to be assisted. We advise RCS Group to employ this methodology to schedule and staff at the proper levels to maximize resource efficiency and boost customer satisfaction. Additionally, it can assist future researchers in developing a model using a vast quantity of data to improve the model's accuracy and additional research on multivariate time series transfer call is needed to improve the forecasting of call kinds.

APPENDIX A. PYTHON CODE TO GET THROUGH ARIMA MODEL AND FORECASTING

```
pip install pmdarima
import pandas as pd
import numpy as np
from google.colab import files
uploaded= files.upload()

df=pd.read_excel('Daily transfer calls.xlsx',
index_col='StartDate',parse_dates=True)
df=df.dropna()
print('Shape of data',df.shape)
df.head()

df['Transferred_call'].plot(figsize=(12,5))
```

Check For Stationarity

```
from statsmodels.tsa.stattools import adfuller

def ad_test(dataset):
  dftest =adfuller(dataset,autolag='AIC')
  print("1. ADF:", dftest[0])
  print("2. P-Value :", dftest[1])
  print("Critical Values")
  for key, val in dftest[4].items():
    print("\t",key," : ",val)


stepwise_fit= auto_arima(df['Transferred_call'], trace=True,
suppress_warnings=True)
stepwise_fit.summary()

decomposition= sm.tsa.seasonal_decompose(df,model='additive')
fig= decomposition.plot()
plt.show()
```

REFERENCES

1. Brett H, 2019, Data- driven studies of caller behavior under call center innovations, University of North Carolina, Chapel Hill.
2. Clement S and George W, 2021, 'Management science', *Does employees happiness have an import on productivity?* Volume 3, Issue 1, 11 May 2023, https://doi.org/10.1287/mnsc.2023.4766.
3. Colin M, Ulrike M, and Priscilla S, 2016, 'Academy of management discoveries', *The call center agent's performance paradox: a mixed-methods study of discourse strategies and paradox resolution,* vol 5, no 2, 4 June 2019, https://doi.org/10.5464/amd.2016.0024.
4. John C and Moshe b, 2001, 'European Journal of Operational research', *Towards "zero abandonments" in call center performance,* Volume 135, Issue 1, 16 November 2001, pp 50-56, https://doi.org/10.1016/50377-2217(00)00289-7.
5. D. K. Barrow, "Forecasting intraday call arrivals using the seasonal moving average method," J. Bus. Res., vol. 69, no. 12, pp. 6088–6096, 2016, doi: 10.1016/j.jbusres.2016.06.016.