AN INNOVATIVE TIME SERIES BASED METHOD OF FORECASTING MONTHLY SALES OF CHAMPAGNE

Aseen Saxena, Siddharth Nanda
School of Engineering, Ajeenkya D.Y. Patil University Pune, India -412105

Journal	Samvakti Journal of Research in Information Technology https://www.sjrit.samvaktijournals.com Volume 1 Issue 1 (2020); Page No. 77 - 88
Discipline	Artificial intelligence and Machine Learning

Date Received	: July 18, 2020	**ID**	: 202007.50.281
Publication Date	: October 29, 2020	**doi**	: 10.46402/202007.50.281
Paper Type	: Journal	**doi URL**	: https://dx.doi.org/10.46402/202007.50.281

ABSTRACT

Time series analysis is a process which deals with data to identify trends and forecast future happening. The data used in this analysis is known time series data which is in a time interval format. The data was used is based on the data which is taken in a different interval of time, known as time-series data. Balancing of demand and supply is based on the accurate prediction of sales in future, if there is a lack of efficient forecasting than it can a challenge to run the business and make a good profit. Time-series forecasting is one of the most commonly used approaches in day-to-day in many organizations. Marketing research also uses time-series data to find their future predictions. Its main strength is to study the change of various event in time so that it can be a strong tool for marketers. Approaches of time series analysis are autoregressive integrated moving average(ARIMA), seasonal autoregressive integrated moving average(SARIMA), autoregressive moving average(ARMA), moving average(MA) and autoregression(AR). This paper consists of predicting the monthly sales of champagne by using time series and will also predict the future monthly sales of champagne. ARIMA and SARIMAX model was used for forecasting and predicted the sales of champagne for 10 years. Champagne dataset was used in this research and using time series, predicting model was prepared. It is predicting good and giving good results.

Keyword: time series analysis, forecasting, sales prediction, future prediction

INTRODUCTION

The main purpose of this paper is to find the future forecasting of the monthly sales of champagne. ARIMA and SARIMA model were used to make a model and which would be useful to predict the monthly sales of French champagne. We will find a plot which shows the prediction for The monthly sales of the champagne dataset is a time series dataset, which can be used to find the future prediction based on the past data points of the dataset. The dataset consists of months and total sales from 01/1964 to 09/1972. There are 104 entries, and the sales are counted in millions[8].

Forecasting is the process of determining the future happening. Forecasting can be used in many fields such as what is the GDP rate after five years, profits or loss after one year, decrease or increase in shares of a company, demand of a product after six months, climate forecasting. Forecasting is required to run an organization, so if there is an unfortunate happening in the future so it can be handled easily. The data which is recorded based on time is called time-series data. Examples are the annual income of an organization, the demand for a product, day-to-day temperature, and the number of passengers travelling. The data which is calculated per second, per minute, hourly, daily, weekly, monthly, yearly are all comes under time-series data. It can be used to do better planning and can suggest what should be done to increase profit and productivity. Taking a real-world example, if a shopkeeper knows the future sells of their products, so it is easy to calculate the amount of product to keep in stock. Owner of a website knows the number of users visit the website so it can be useful to handle website traffic efficiently. The model of time series predicts future values according to previous values. On the other hand, multivariate analysis is an analysis which involves finding one or more statistical outcome at a time. Time series are used in various fields such as in signal processing, econometrics, finance, astronomy, meteorology, communication, earthquake prediction and others.

ARIMA models are the types of models which is used to forecast future prediction using time series data. The main purpose of this model is to describe and find the autocorrelations in the time series data. For example, in a company how many average products sold in a year Figure 1, we can see that the black line shows the actual product sold in years and the blue lines indicates the data which is predicted by the time series model, by this we can understand that we are using past data to predict the future sales.

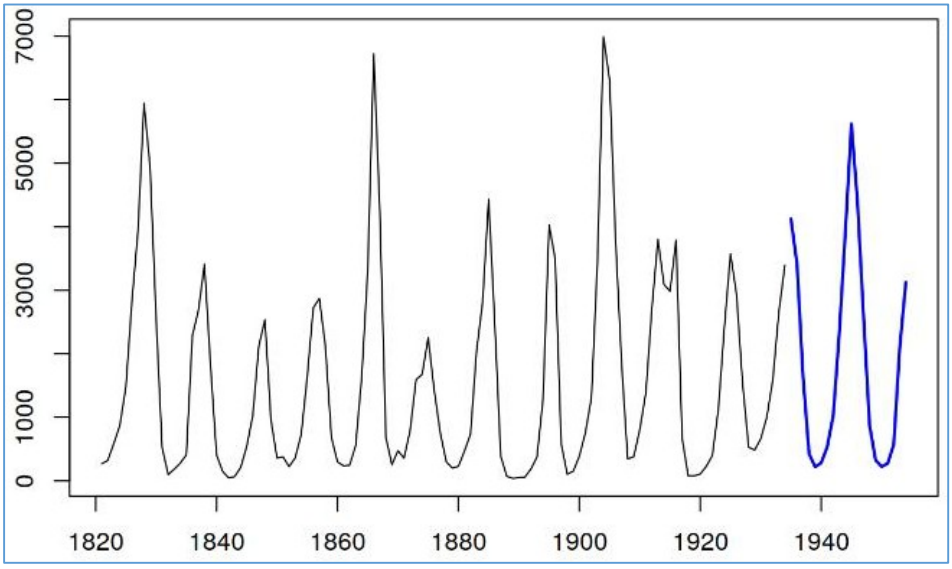

Figure 1 *: Average sales of product of a company*

LITERATURE SURVEY

Lars Dannecker et al[1]. has mentioned that their approach is to forecast the hierarchy of organization of energy market and also introduced a framework for forecasting, which can easily forecast efficiently and forecast the maintenance of time series model that derives through the continuous stream of the model. Their framework consists of model adaption and evaluation techniques that build up the model maintenance. Their approach forecasts accurately and can be predicted with the same time budget. In future, they will work on a high-level strategy.

M.S. Roulston et al[2]. explained that there are different techniques used in ECMWF to predict how much wind energy generators are checked. All of the forecasts based on ECMWF forecast products outperform climatology takes about five days. In their results, the ECMWF forecast can be useful for wind energy producers. It provides the best forecast over climatology. Their predictions lead to twice or thrice days income increased by 75%.

E.J. Ekpenyong and C.O. Omekara[3] mentioned that their research was based on periodic time series. Periodic time series model was made. Fourier series analysis method was used to make their model and model was used to predict future data values of such data. Ordinary least square method in multiple regression was used to estimate the parameter of Fourier series model. P-P plot was used to test the overall goodness of fit and forecast the values and can predict the future temperature.

Kalyan Joshi et al.[4] explained that there are three stages of the dataset. Multiple time series is used to identify a subgroup of time series. Whether the subset of time series includes time period with inactivity can be determined. Whether the time series consists of magnitude spark increase with a data above a preset magnitude can be predicted. When time series (i) lacks the time period with inactivity, (ii) exhibits the repeating characteristic, and (iii) increase in magnitude with the data above the preset magnitude.

M. Das et al.[5] explained that a model was made, which uses digital images and used for adaptive predictive coding. Multiplicative time series, one-dimensional and recursion used for their model. The performance was compared to the conventional one-dimensional model technique because of the correlation of all images taken into account. The time stability of the model was also excellent.

Lisa Farrell et al.[6] explained that their research based on the addiction of lotto players. The physical addiction of things such as cigarettes has tended to do this research. Their data suggested to them that there is some addiction and rollovers induce it since sales are increasing than average. In their results, lotto is less addictive when compared to goods such as cigarettes.

S.G. Kapoor et al.[7] mentioned that finding accurate and efficient forecast of sales in marketing strategy. The method used in this mainly is regression and correlation analysis. These techniques do not have dynamics or memory of the procedure involved in treating the data, which are in the form of time series. Two series of data were used for analysis. On comparison of minimum mean squared error forecast, which obtained from their model with observations about three period shows that actual data were good with 95% probability limits of the forecasts.

EXISTING SYSTEM APPROACH

ARIMA model includes a part of a statistical model for forecasting and predicting time series data. ARIMA stands for autoregressive integrated moving average. It mainly works with time series data and provides a powerful method that creates the forecast very efficient and accurate. It inspired by the autoregressive moving average(ARMA).

ARIMA consists of :

1. Autoregression(AR) - This method predicts the next data point in a linear function sequence in accordance with previous time steps. AR uses a dependent relationship between lagged observation and a actual observation.
2. Integrated(I) - it uses the difference of raw observation to obtain a statistic stationary.
3. Moving average(MA) - It uses the dependency between a residual error and an observation. It predicts the next data based on the linear function of the residual error.

PROPOSED SYSTEM APPROACH

1. Importing packages and libraries - In figure 1, imported the required packages and libraries.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import mean_squared_error
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
plt.style.use('ggplot')
import statsmodels.api as sm
```

Figure 2 : Importing packages and libraries

2. Importing dataset - In figure 2, imported Champagne dataset, which is a time-series dataset and which consist of two columns, i.e. months and sales and there are 104 rows. Dataset is taken from https://www.kaggle.com/anupamshah/perrin-freres-monthly-champagne-sales. Accessed on June 2020.

```
dataset = pd.read_excel('champagne.xlsx', skiprows = 12)
dataset.head()
```

	Month	Perrin Freres monthly champagne sales millions '64-'73
0	1964-01-01	2815
1	1964-02-01	2672
2	1964-03-01	2755
3	1964-04-01	2721
4	1964-05-01	2946

```
dataset.tail()
```

	Month	Perrin Freres monthly champagne sales millions '64-'73
100	1972-05-01	4618
101	1972-06-01	5312
102	1972-07-01	4298
103	1972-08-01	1413
104	1972-09-01	5877

Figure 3 : Importing Champagne dataset

3. Checking parameters - In figure 3 and 4, checked dataset to understand all parameters and dependency and edited dataset for convince. Also, displayed plot of the dataset, so, it can easy to understand and find pattern. In graph, we can see that sales is increasing as the year changes.

```
dataset['Month']=pd.to_datetime(dataset['Month'])
dataset.set_index('Month',inplace=True)
dataset.head()
```

Month	Sales
1964-01-01	2815
1964-02-01	2672
1964-03-01	2755
1964-04-01	2721
1964-05-01	2946

Figure 4*: Head of the dataset*

```
dataset.describe()
```

	Sales
count	105.000000
mean	4761.152381
std	2553.502601
min	1413.000000
25%	3113.000000
50%	4217.000000
75%	5221.000000
max	13916.000000

```
dataset.plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1efe705dec8>
```

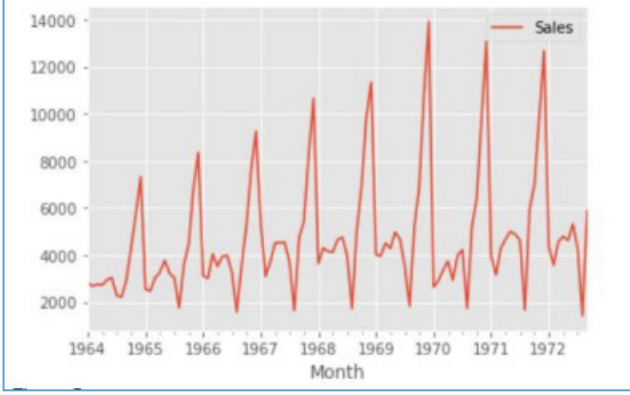

Figure 5 : Plot of the dataset

4. Seasonal plot - In figure 5, there is seasonal plot of the dataset. In this we can see that what is the season of the champagne where the sales are more.

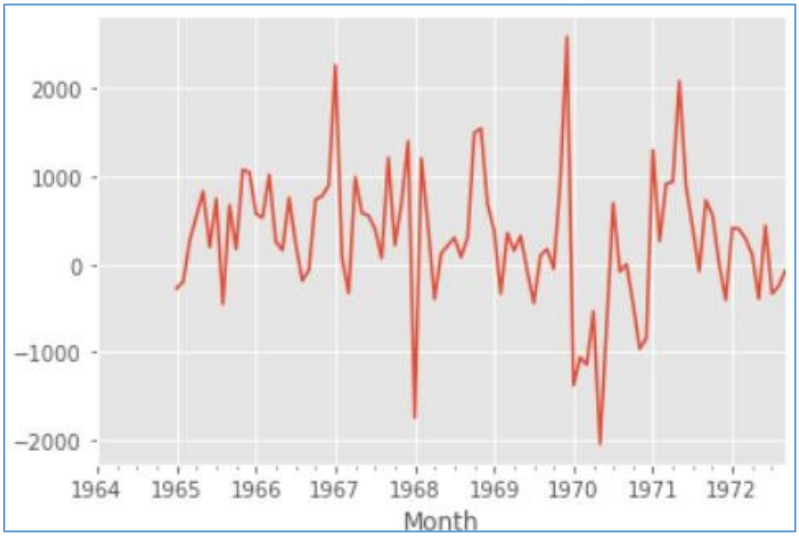

Figure 6 : Seasonal plot

5. Autocorrelation and partial autocorrelation plot - In this figure 6, checked autocorrelation and partial autocorrelation.

```
plt.figure(figsize=(12,8))
plt.subplot(211)
plot_acf(dataset['Seasonal First Difference'].iloc[13:],lags=40,ax=plt.gca(), color ='blue')
plt.subplot(212)
plot_pacf(dataset['Seasonal First Difference'].iloc[13:],lags=40,ax=plt.gca(), color ='red')
plt.show()
```

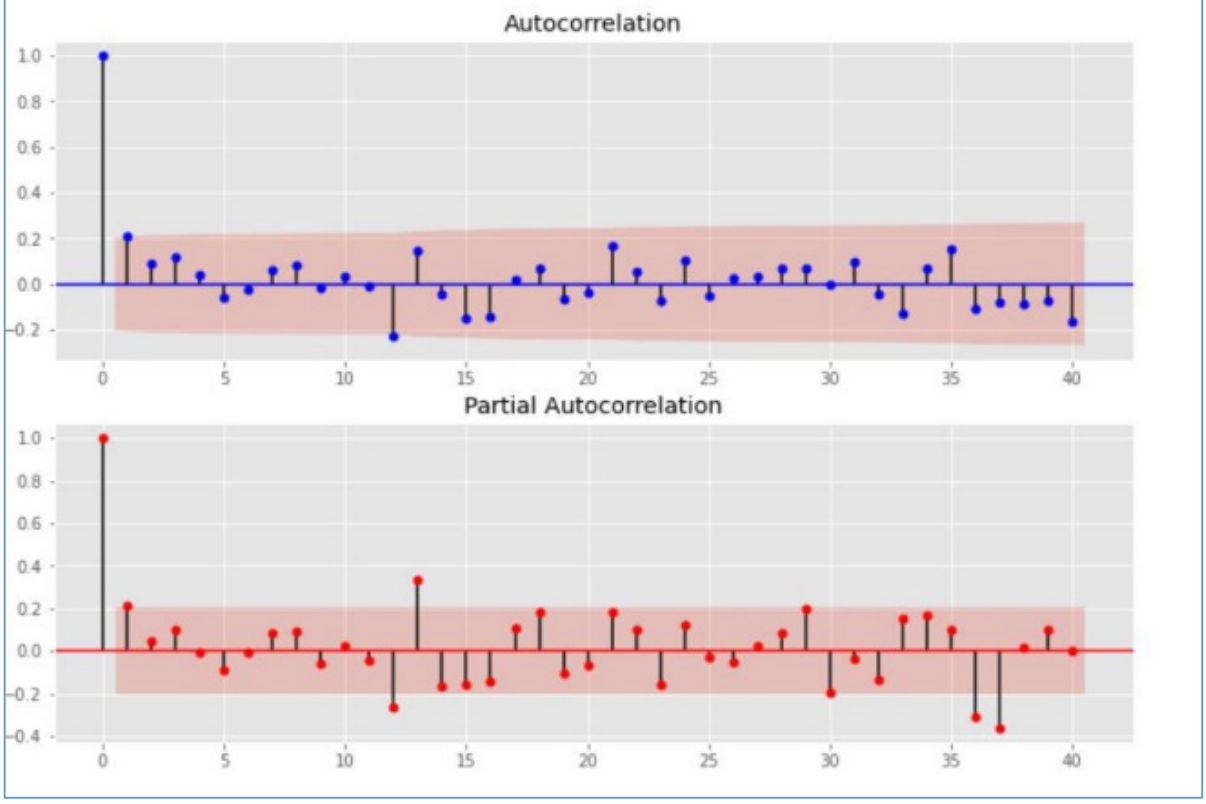

Figure 7 : Autocorrelation and partial autocorrelation

6. ARIMA model - In figure 7, created our ARIMA model and checked the parameter of the model. For example, coefficient, standard error, dependent variable and roots.

```
model=ARIMA(dataset['Sales'],order=(1,1,1))
model_fit=model.fit()
```

```
model_fit.summary()
```

ARIMA Model Results

Dep. Variable:	D.Sales	**No. Observations:**	104
Model:	ARIMA(1, 1, 1)	**Log Likelihood**	-951.126
Method:	css-mle	**S.D. of innovations**	2227.262
Date:	Mon, 22 Jun 2020	**AIC**	1910.251
Time:	10:53:38	**BIC**	1920.829
Sample:	02-01-1964	**HQIC**	1914.536
	- 09-01-1972		

	coef	std err	z	P>|z|	[0.025	0.975]
const	22.7859	12.405	1.837	0.066	-1.528	47.100
ar.L1.D.Sales	0.4343	0.089	4.866	0.000	0.259	0.609
ma.L1.D.Sales	-1.0000	0.026	-38.503	0.000	-1.051	-0.949

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.3023	+0.0000j	2.3023	0.0000
MA.1	1.0000	+0.0000j	1.0000	0.0000

Figure 8*: Parameter of ARIMA model*

7. Forecast plots - In figure 8, checked the prediction of the model whether it is giving good prediction or not. In graph, system is learning to find future data points.

Figure 9 : *Forecasting plots*

8. Final forecast plots - In figure 9, plot shows the prediction for 10 years. We can see that sales can increase in coming years.

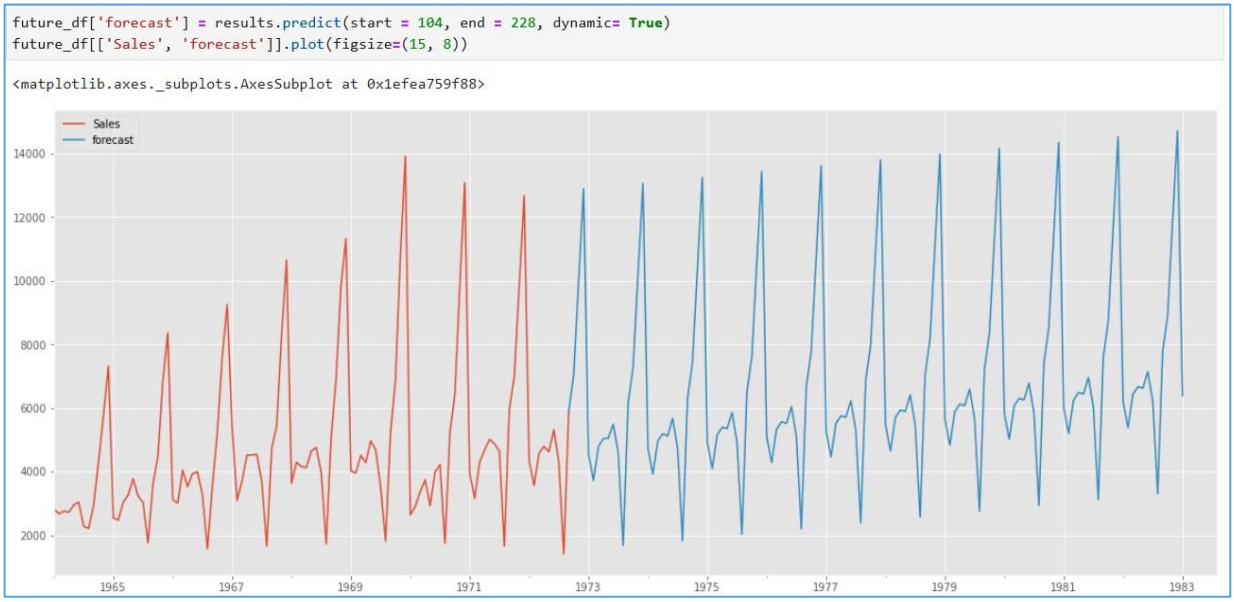

Figure 10 : Final forecasting plot

CONCLUSION

In this research, there was the forecasting of the French champagne sales using time series forecasting. In this, the forecasting model was build using the ARIMA model. Forecasting is essential in many sectors, especially in business sectors, because if a person knows about facing profit or loss in the coming years, it can be easily handled to meet the problem efficiently. For example - a farmer knows that which is the best season to cultivate crops so that it can make more profit and to meet less loss. ARIMA model was used to forecast the sales of champagne. It was giving excellent and accurate results and projected the sales of champagne after ten years.

In future, we can implement this project in many fields such as product sales prediction, weather forecasting, stock prediction as model and algorithm is somewhat similar. It will be very useful to predict whether a organization will face profit and loss in future. By this, it will be useful for the organization to work efficiently.

REFERENCES

[1]. Dannecker, L., Böhm, M., Lehner, W. and Hackenbroich, G., 2011, September. Forcasting evolving time series of energy demand and supply. In East European Conference on Advances in Databases and Information Systems (pp. 302-315). Springer, Berlin, Heidelberg. Accessed on June 2020.

[2]. Roulston, M.S., Kaplan, D.T., Hardenberg, J. and Smith, L.A., 2003. Using medium-range weather forcasts to improve the value of wind energy production. Renewable Energy, 28(4), pp.585-602. Accessed on June 2020.

[3]. Ekpenyong, E.J. and Omekara, C.O., 2008. Application of Fourier Series Analysis To Temperature Data. Global Journal of Mathematical Sciences, 7(1), pp.5-13. Accessed on June 2020.

[4]. Joshi, K., Roehl, N. and Chien, Y.H.A., Sas Institute Inc, 2018. Forcasting interest in an object over a future period of time using a three-stage time-series analysis process. U.S. Patent Application 15/788,238. Accessed on June 2020.

[5]. Das, M., Tan, S. and Loh, N., 1987, April. Adaptive predictive coding of images based upon multiplicative time series modelling. In ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 12, pp. 1386-1389). IEEE. Accessed on June 2020.

[6]. Farrell, L., Morgenroth, E. and Walker, I., 1999. A time series analysis of UK lottery sales: long and short run price elasticities. oxford Bulletin of Economics and Statistics, 61(4), pp.513-526. Accessed on June 2020.

[7]. Kapoor, S.G., Madhok, P. and Wu, S.M., 1981. Modeling and forecasting sales data by time series analysis. Journal of Marketing Research, 18(1), pp.94-100. Accessed on June 2020.

[8]. https://www.kaggle.com/anupamshah/perrin-freres-monthly-champagne-sales Accessed on June 2020

[9]. https://www.statisticssolutions.com/time-series-analysis/#:~:text=Time%20series%20analysis%20is%20a,particular%20time%20periods%20or%20intervals. Accessed on June 2020

[10]. https://en.wikipedia.org/wiki/Time_series Accessed on June 2020

[11]. https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775 Accessed on June 2020

[12]. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/#:~:text=So%20what%20exactly%20is%20an,used%20to%20forecast%20future%20values. Accessed on June2020

END