|  |  |
| --- | --- |
| Name of the Student | Shubham Bakde |
| Internship Project Title | Superstore Sales Forecasting |
| Name of the Company | TCS ion |
| Name of Industry Mentor | Esmita Gupta |
| Name of the Institute | BK Birla College of Arts, Science & Commerce |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort | Project Environment | Tools Used |
| 15th Oct 2022 | 15th Nov 2022 | 72 hrs | Google Colab | Python |

**Table of Contents**

[Acknowledgements 2](#_Toc119935215)

[Objective 3](#_Toc119935216)

[Introduction 4](#_Toc119935217)

[Internship Activities 5](#_Toc119935218)

[Methodology 6](#_Toc119935219)

[Assumptions 6](#_Toc119935220)

[Exclusions 6](#_Toc119935221)

[Charts 7](#_Toc119935222)

[Algorithms 11](#_Toc119935223)

[Challenges & Opportunities 11](#_Toc119935224)

[Risk vs Rewards 11](#_Toc119935225)

[Reflections on the Internship 11](#_Toc119935226)

[Recommendations 11](#_Toc119935227)

[Conclusion 12](#_Toc119935228)

[Enhancement Scope 12](#_Toc119935229)

[Link to code and executable file 12](#_Toc119935230)

## Acknowledgements

I would like to express my special thanks of gratitude to our head of department and vice principal Mrs. Esmita Gupta ma’am for their guidance and support to completing my project.

I would also like to thank TCS ion for giving me a platform to improve my skills and helping me towards my future.

DATE:

15th November 2022

## Objective

The objective of our project is to find whether there are different patterns in the formats of sales or not and if there are patterns then we have to identify those patterns and predict whether these sales will go upwards downwards or is there noise in our data set or not which will be very helpful for the future predictions.

## Introduction

For our project we have chosen a superstore data set in this data set this file is of an excel format and this file is perfect for as we have the dates and the revenue sales of every category listed so we can find out the patterns in the data set of the categories like furniture in which we can also predict whether the sales are going up or down.

## Internship Activities

Week 1:

* Read the dataset and understood the dataset.
* Got info about all columns in Dataset and their types.

Week 2:

* Creating a sample dataset named furniture in which our data will be only having data of furniture category of all affected columns.
* Plotting our Sales by Year.
* Using rcParams to get Seasonal trends in our dataset to find patterns.

Week 3:

* Creating Matrix format for SARIMAX.
* Finding best combination in our model through ARIMA value.
* Plotting our Results.
* Comparing our Actual v Observed Values.
* Finding Mean Square Error and Root Mean Square Error.
* Plotting Future predictions for furniture sales.

Week 4:

* Creating two new sample datasets from our category of furniture and office supplies.
* Creating a dataset with values of sales of furniture and office supplies to get comparison values.
* Plotting our comparison Graph.

Week 5:

* Creating prophets for our already designed columns with 95% hypothesis.
* Creating a forecasting Dataset for prediction.
* Plotting our Furniture sales prediction.
* Plotting our Office Supplies sales prediction.

## Methodology

Our approach was to first read the pattern if any present in the data set or the sales and for that we need to find the pattern, then our job was to determine a proper SARIMAX matrix and the remark value once we have the ARIMA value we can then predict the future sales based on that value.

## Assumptions

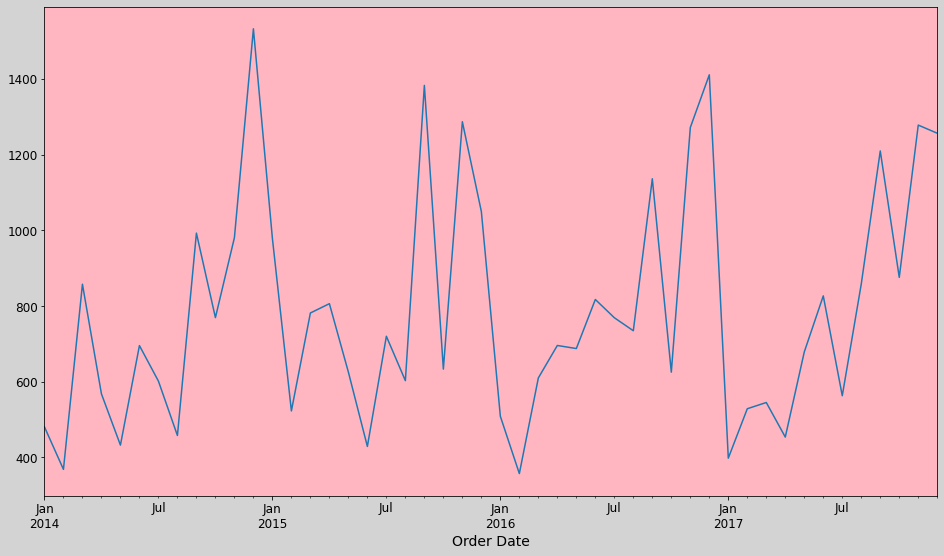
We are assuming that there are patterns in our sales values and there are many factors defining the patterns so after we find the patterns, we assume that the sales will go higher and the noise will be as less as possible.

## Exclusions

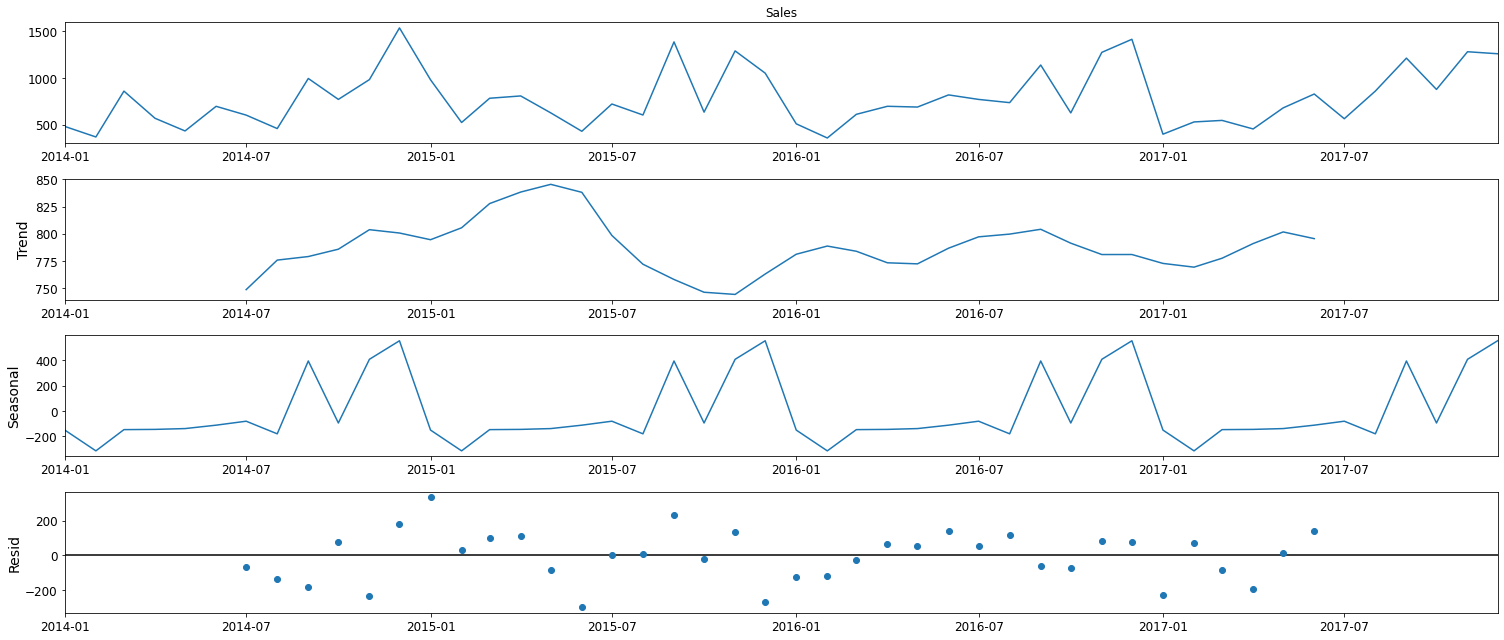
The only exclusion is that we never tried any other model we just tried the ceramics for our predictions.

## Charts

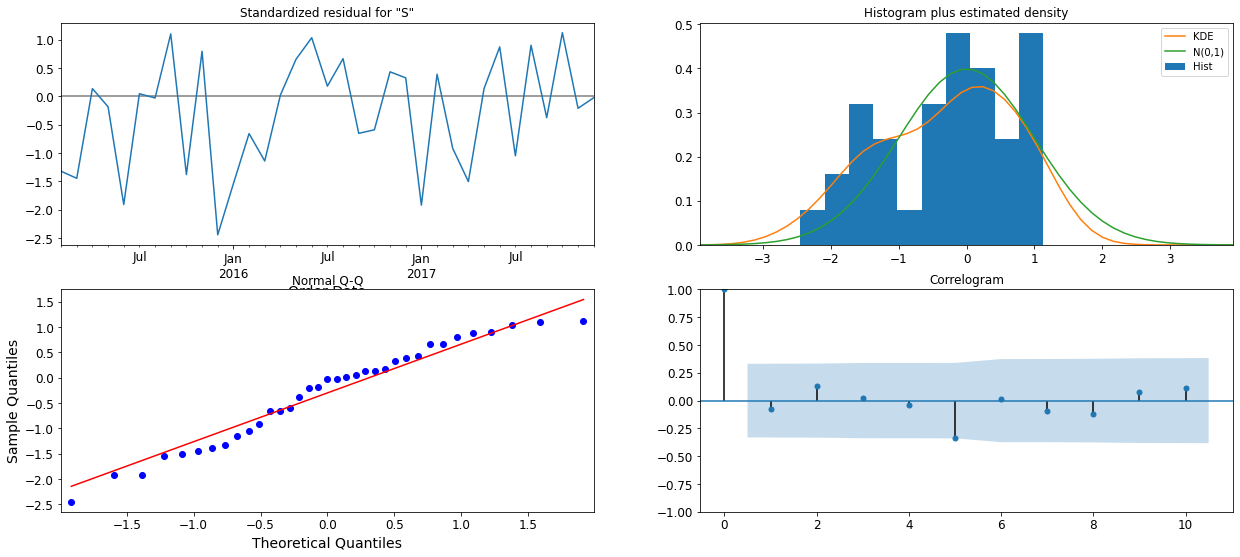
Sales by year:



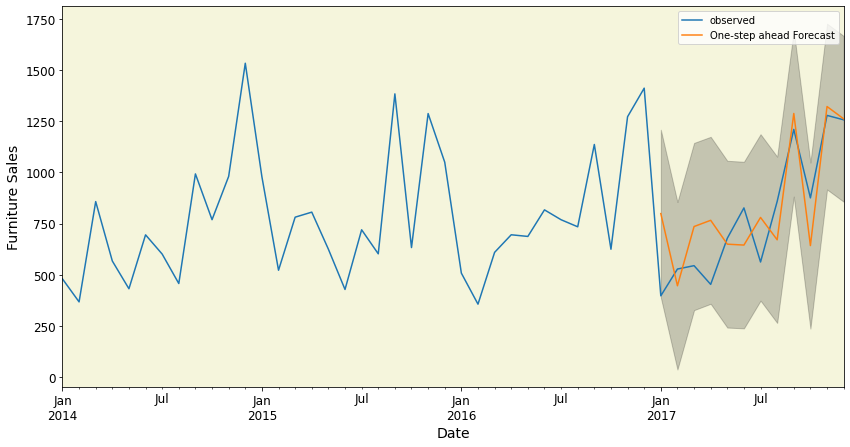
Trends in seasonal sales:



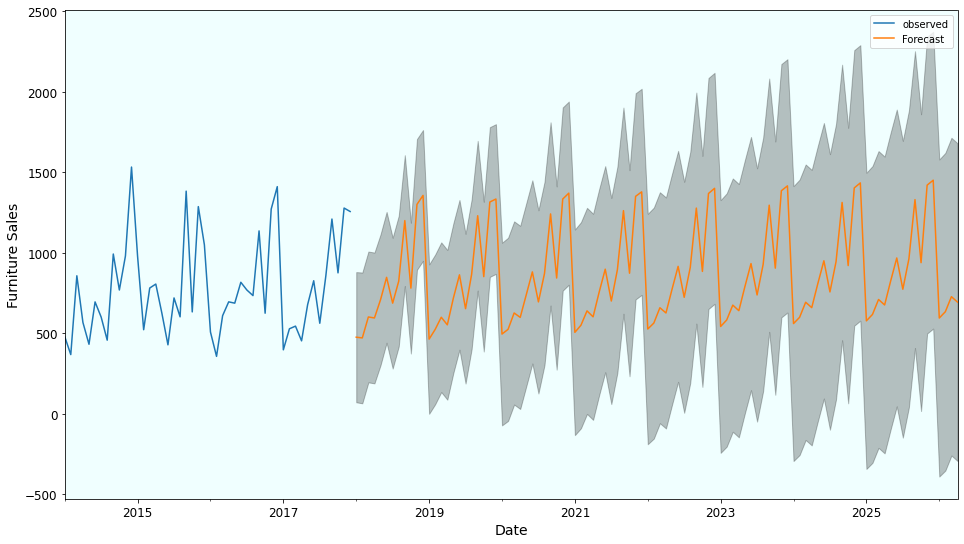
SARIMAX and ARIMA results:



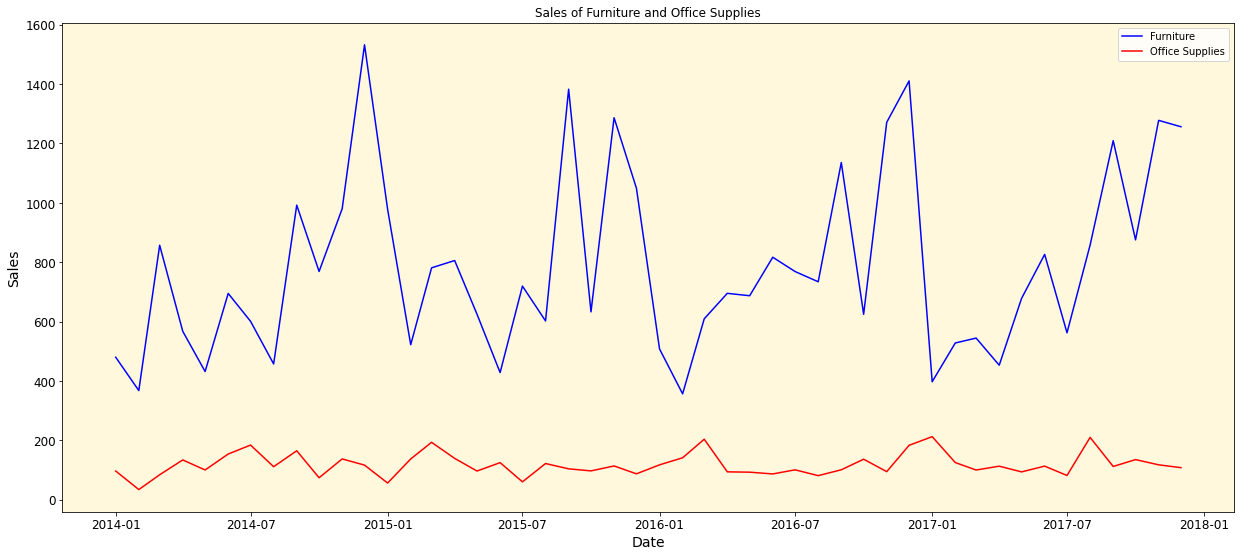
Observed vs One step Ahead forecast:



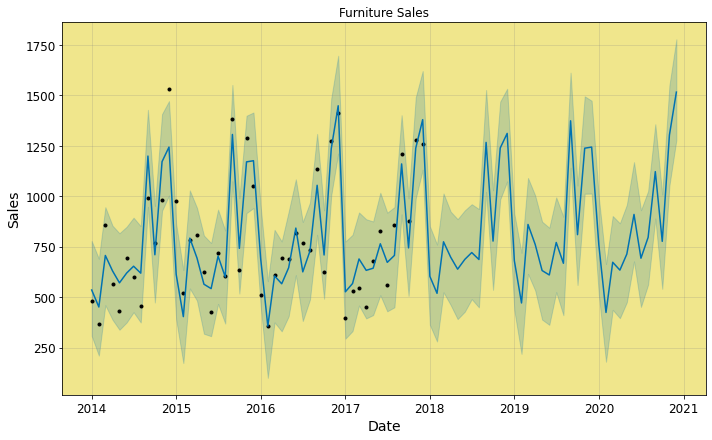
Prediction forecast:



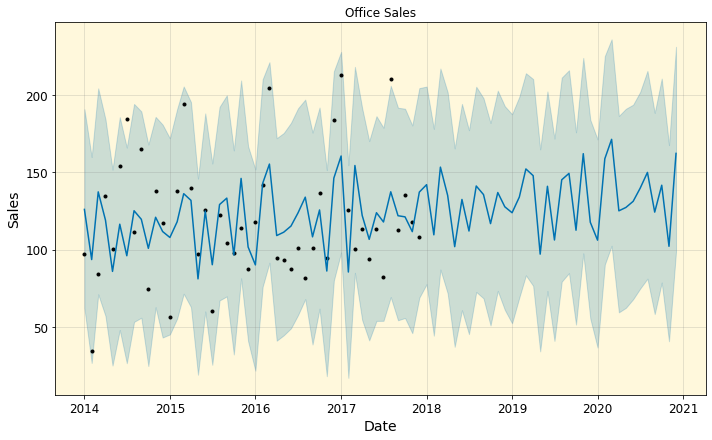
Sales comparison of Furniture vs Office Supplies:



Furniture Sales forecast model:



Office supplies sales forecast model:



## Algorithms

We used a Rama method for our algorithm ARIMA is a method for forecasting the future outcomes based on the historical time series it is based upon the statistical concepts of serial correlation bad past data points actually influence the future data points.

Arima algorithm is especially used for datasets that can be mapped to stationary time series these statistical properties or stationary time series such as autocorrelations are independent of time and in such combination time series, they actually contain signal and noise.

## Challenges & Opportunities

The first of all challenge was to verify that the algorithm will actually work because the sample data was very distinct and there were not many patterns in the data set and the opportunity was to learn more about the time series and how it works in we can see whether there is much noise in our data set and not and this was actually very helpful as we see from our charts that there was not much noise so our hypothesis was actually correct.

## Risk vs Rewards

We took the risk even after the mean square error and root mean square error where on a large number but we got a reward that the signal was actually perfect and there wasn’t much noise in our prediction forecast.

## Reflections on the Internship

It was an amazing platform to work on and to be honest it was really amazing to learn about the time series forecasting model and how it works based on the historical data mainly because this algorithm focuses on signals and noises it is very helpful for a line graph and with a smaller number of noises in our prediction, we can say that our predictions have more accuracy towards the future.

## Recommendations

The positive outcome of this was there was very less noise you know prediction algorithms so the recommendation is that we have to make sure that there is less noise if there is an increase in the noise then the prediction algorithm will have many errors and it will be a waste part time and effort.

## Conclusion

We were successfully able to predict the future sales of our two categories such as furniture and office supplies. Also, we had very less noise and a perfect signal for our future predictions so the result shows that our sales will grow over the years of both furniture’s and office supplies but when we compare them both the furniture sales, they will grow much faster than the office supplies as we have already seen in our historical data of how they have been growing throughout.

## Enhancement Scope

The scope of enhancement is that our data was not much dependent on other factors so if we find that our data was dependent on seasonal variance or any other things then we can find better patterns but our case was strictly towards the furniture’s and office supplies and such categories which are not much dependent on the seasons so we never had much patterns but the time series helped us a lot by giving us proper prediction in less noise based on our historical data.

## Link to code and executable file

[Click Me](https://drive.google.com/file/d/1RuTBCSOP1O29BiXDU5X6tS8qxjY5GqnV/view?usp=share_link)