Mini Project Report

on

Exploratory data analysis and visualization of superstore sales data

By

Saaz Bharadwaj (Reg. No.- 201700072)

Sharadha Bhardwaj (Reg. No.- 201700061)

Dushyant Singh (Reg. No.- 201700066)

Group ID: 3

In partial fulfilment of the requirements for the award of degree in Bachelor of Technology in Computer Science and Engineering

(2020)

Under Project Guidance of

Ms. Bijoyeta Roy

Assistant Professor

Sikkim Manipal Institute of Technology, Majitar

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent college of Sikkim Manipal University)

MAJITAR, RANGPO, EAST SIKKIM - 737136

PROJECT COMPLETION CERTIFICATE

This is to certify that below mentioned students of Sikkim Manipal Institute of Technology have worked under my supervision and guidance from **8th January 2020** to **15th May 2020** and have successfully completed the project entitled "**Exploratory data analysis and visualization of superstore sales data"** in partial fulfilment of the requirements for the award of degree in Bachelor of Technology in Computer Science and Engineering.

University Registration No.	Name of Student	Course
201700061	**Shradha Bhardwaj**	**B. Tech (CSE)**
201700072	**Saaz Bhardwaj**	**B. Tech (CSE)**
201700066	**Dushyant Singh**	**B. Tech (CSE)**

Ms. Bijoyeta Roy

Assistant Professor

Department of Computer Science and Engineering

Sikkim Manipal institute of Technology

Majitar, Sikkim – 737136

PROJECT REVIEW CERTIFICATE

This is to certify that the work recorded in this project report entitled **"Exploratory data analysis and visualization of superstore sales data"** has been jointly carried out by **Saaz Bhardwaj (Reg. 201700072), Shradha Bhardwaj (Reg. 201700061) and Dushyant Singh (Reg. 201700066)** of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology in partial fulfilment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering. This report has been duly reviewed by the undersigned and recommended for final submission for Mini Project Viva Examination.

Ms. Bijoyeta Roy
Assistant Professor
Department of Computer Science and Engineering
Sikkim Manipal Institute of Technology
Majitar, Sikkim – 737136

CERTIFICATE OF ACCEPTANCE

This is to certify that the below mentioned students of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology (SMIT) have worked under the supervision of **Ms. Bijoyeta Roy** Assistant Professor, Department of Computer Science and Engineering from **8th January 2020 to 15th May 2020** on the project entitled **"Exploratory data analysis and visualization of superstore sales data".**

The project is hereby accepted by the Department of Computer Science & Engineering, SMIT in partial fulfilment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

University Registration No.	Name of Student	Project Venue
201700061	**Shradha Bhardwaj**	**SMIT**
201700072	**Saaz Bhardwaj**	**SMIT**
201700066	**Dushyant Singh**	**SMIT**

Prof. (Dr.) Kalpana Sharma
Professor & Head of Department
Computer Science & Engineering Department
Sikkim Manipal Institute of Technology
Majitar, Sikkim – 737136

DECLARATION

We, the undersigned, hereby declare that the work recorded in this project report entitled "**Exploratory data analysis and visualization of superstore sales data**" in partial fulfilment for the requirements of award of B.Tech (CSE) from Sikkim Manipal Institute of Technology (A constituent college of Sikkim Manipal University) is a faithful and bonafide project work carried out at "**SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY**" under the supervision and guidance of **Ms. Bijoyeta Roy,** Assistant Professor, Department of Computer Science and Engineering.

The results of this investigation reported in this project have so far not been reported for any other Degree or any other Technical forum.

The assistance and help received during the course of the investigation have been duly acknowledged.

Saaz Bhardwaj (Reg. No.-201700072)

Shradha Bhardwaj (Reg. No.-201700061)

Dushyant Singh (Reg. No.-201700066)

ACKNOWLEDGMENT

We place on record and warmly acknowledge the continuous encouragement, invaluable supervision, timely suggestions and inspired guidance offered by our guide, **Ms. Bijoyeta Roy**, Assistant Professor, Department of Computer Science and Engineering in bringing this project to a successful completion.

We are grateful to **Prof.(Dr.) Kalpana Sharma** , Professor and Head of Department , Computer science and Engineering for permitting us to make use of the facilities available in the department to carry out the project successfully.

We are obliged to our project coordinates **Dr. Sandeep Gurung, Mr. Santanu Kr. Misra, Mr. Biraj Upadhyaya** and **Ms. Nitisha Pradhan** for evlevating, inspiring and kind supervision in completion of our project.

We are deeply grateful to all the staff members of Computer Science and Engineering department for supporting us in all aspects.

Saaz Bhardwaj(201700072)
Dushyant Singh(201700066)
Shradha Bhardwaj(201700061)

DOCUMENT CONTROL SHEET

1.	Report No	CSE/Mini Project/Internal/B.Tech/A/3/2020
2.	Title of Report	Exploratory data analysis and visualization of superstore sales data
3.	Type of Report	Technical
4.	Author	Saaz Bhardwaj Dushyant Singh Shradha Bhardwaj
5.	Organising Unit	Sikkim Manipal Institiute Of Technology
6.	Language of Document	English
7.	Abstract	
8.	Security Classification	General
9.	Distribution Statement	General

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Several organizations use supply chain management / inventory management software for efficient logistic handling, failing in which would incur losses to the company due to over – stocking, over – production etc. Most present-day software operate at a terminal / node level i.e. at the retail points, warehouses etc. independently, hence the overall analysis of sales / production at different scopes (city wise, province wise etc.) & for different classes of products could be a tiresome task which needs to be done as it can uncover better insights.

An interface, that could integrate the various terminal sales data and allow the user to select the different scopes and product categories according to their need which would target very specific queries (e.g. Present / expected sales of Furniture in a province) and produce results with the help of the collected data in conjugation with the analytics-based methods and machine learning based techniques, should be very useful.

Various terminals / set of terminals of the organization can be monitored on a single platform and efficient data driven decisions can be made from the derived insights. Future trends and demands can be predicted beforehand and hence assist in better logistic and inventory management. Therefore, selecting and applying the best suited machine learning model for sales forecast along with the integrating platform is intended in our project.

The following application we will be developed in python and the interface shall be developed using HTML.

INTRODUCTION

Sales logs over a span are collected at the various establishments of the Superstores. These logs include date, quantity ordered, sales, profit and many other attributes which can be used to predict revenues, future demands and sales trends with the help of analytics-based methods in conjugation with machine learning models. The user can select the product category or the scope at which predictions are desired (city wise, province wise etc.). Strategists / business planners / store chain managers for the organization would find it very useful.

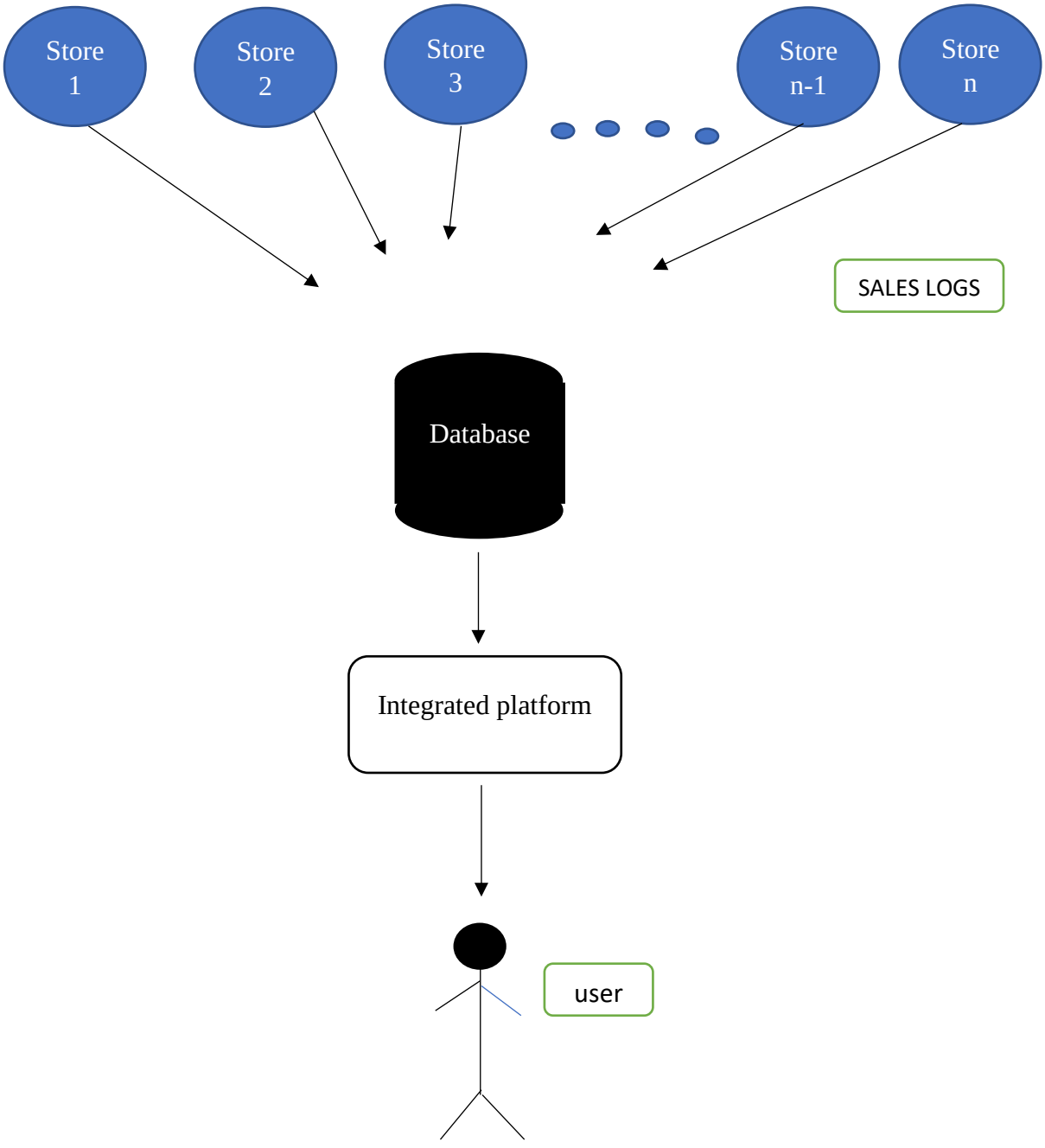

Fig 1.1: Basic structure of the project

For this project we have taken data over a span of 4 years from 2010 to 2013 of sales logs throughout all establishments. The dataset has been taken from Kaggle (https://www.kaggle.com/). The integrated platform has several sub modules that need to be developed, which would work in conjugation with each other to give the desired outcome.

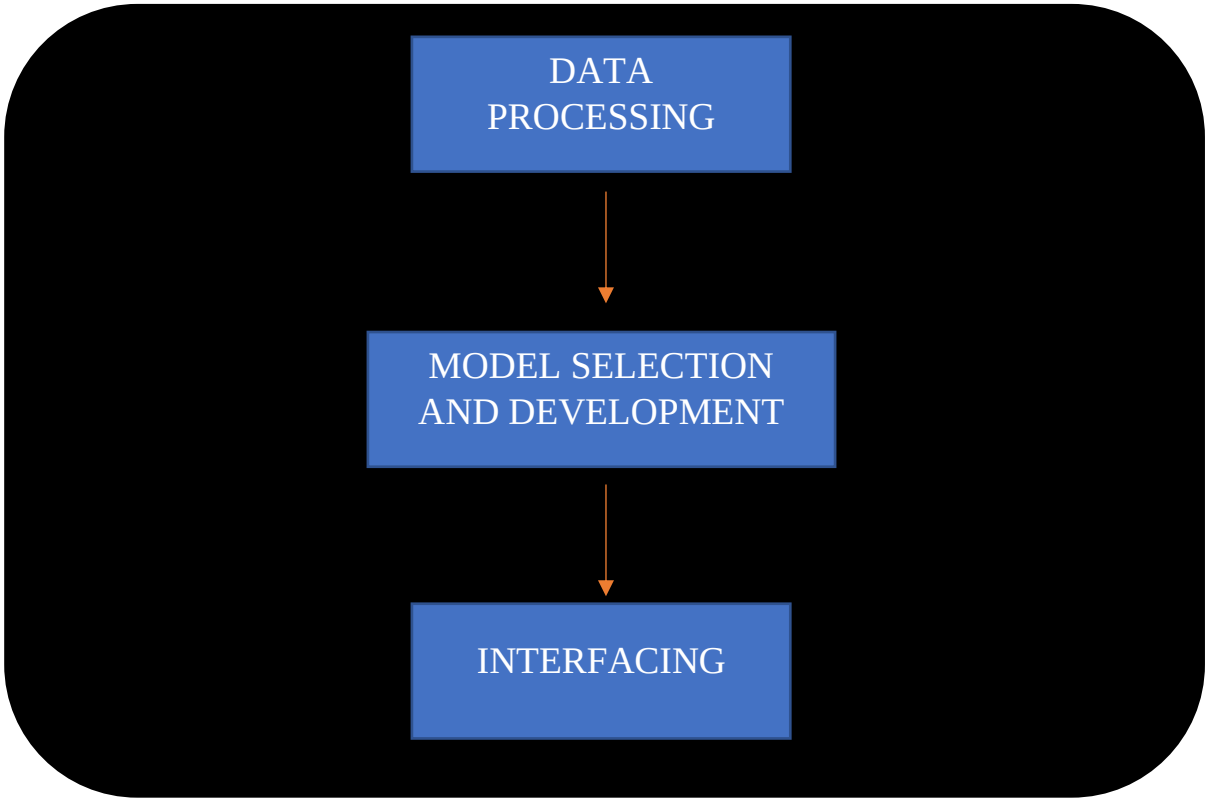

Fig 1.2: Integrated platform and it's sub-modules

For this project we assume that we do not need to provide any module for integrating the data in the database. We are only concerned with utilization of the integrated data to develop our model for time series forecasting.

Data Processing: Before developing or deploying any model we need to process the data to remove any ambiguities, unusual / high variations and deal with missing values.

Model selection and development: Only after processing the data can we manipulate it and develop our model for training with the data set. There are several machine learning models that can be deployed but we need to select the one with the closest predictions. The RMS (Root mean square) values are used as the evaluation metric of the models.

Interfacing: An interface for the user to select the scope as well as the different categories for which the predictions are desired.

The application module will be coded in python due to the several useful libraries available (like Scikit-learn, Matplotlib, Statsmodels etc.) which are useful in data processing, manipulation as well as model development, training and evaluation.

The interface module is intended to be developed on HTML. It is planned as a web application for a wide access horizon.

LITERATURE SURVEY

Sl.No.	Paper & Author Details	Findings	Research gap	Where it was published	Relevance to the project
1	Time-series sales forecasting for an Enterprise Resource Planning system (2019) Toni Malila	Three different algorithms namely ARIMA, LTSM, NN are used for product demand forecast, as well as some hybrid models RMSE is used for the evaluation of the developed models Results show that ARIMA emerges as the best suited model for data predictions	1 year	http://urn.fi/URN:NBN:fi:amk-201906031433 5	We also need to handle large amount of data in our project and make predictions based on previous consumption data The objectives of our projects are same(assist logistics management) This project can be applied only on the terminals of sales, whereas our project can help monitor and manage multiple terminals of sale
2	Time series forecasting using a ARIMA and Neural network model. (2001) G. Peter Zhang	Researches related to ANN's suggest that they can replace traditional linear methods in time series forecasts A hybrid of ANN and ARIMA models is proposed to utilize their individual strengths to improve accuracy	19 years	https://www.journals.elsevier.com/neurocomputing	The ANN and ARIMA models yield different quality results pertaining to the whether the data needs linear or complex fitting, or if the model needs to be adaptive, etc. We need to find the best result yielding and adaptive model for our forecast.

Sl.No.	Paper & Author Details	Findings	Research gap	Where it was published	Relevance to the project
3.	Forecasting aggregate retail sales:A comparison of artificial neural networks and traditional methods (2001) Ilan Alon , Min Qi , Robert J. Sadowski	It compares traditional methods like winter expo. smoothening, Box-Jenkins ARIMA model & multivariate regression with ANNs ANNs were found more favourable followed by ARIMA models but Winters method was seen more viable for multi step forecasts under stable economic conditions To conclude, ANNs were more effective in capturing the dynamic and non seasonal trends	19 years	journal ISSN : 0969-6989 Publisher :Elsevier Science	
4.	A moving average filter based hybrid ARIMA–ANN model for forecasting time series data C. Narendra Babu B.Eswara Reddy		6 years	https://www.journals.elsevier.com/applied-soft-computing	

S.No	Paper & Author Details	Findings	Research gap	Where it was published	Relevance to the project
5	Statistical and Machine Learning forecasting methods: Concerns and ways forward (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward (2018) Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos.	This article shows why ML models are less accurate than statistical ones More objective and unbiased methods are required for performance evaluations of forecast methods	2 years	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194889	

PROBLEM DEFINITION

Due to non-data driven decisions made by inventory / sales management organizations owning stores and warehouses incur heavy losses due to poor logistics management or even over-production. Thus, statistics-based analysis in conjugation with machine learning techniques / models must be used to derive future predictions to avoid the above-mentioned losses.

There are many machine learning models developed for time series forecast but we need to choose the one which gives the closest prediction.

Present day supply chain management / inventory software are mostly developed as terminal / nodal applications. Overall reviewing and management require high manpower for integration of the intelligence derived by these terminal applications. Thus, an integrated platform must be provided for selecting the scope as well as category for future predictions.

PROPOSED SOLUTION STRATEGY

The solution is to have an integrated platform which gives the users the option to select the scope (city wise, province wise etc.) for the desired categories (sales, quantity ordered, furniture, office supplies etc.) for the future predictions to be made by the deployed model.

For this project we already have the integrated sales log from all stores over a span of 4 years (2010-2013).

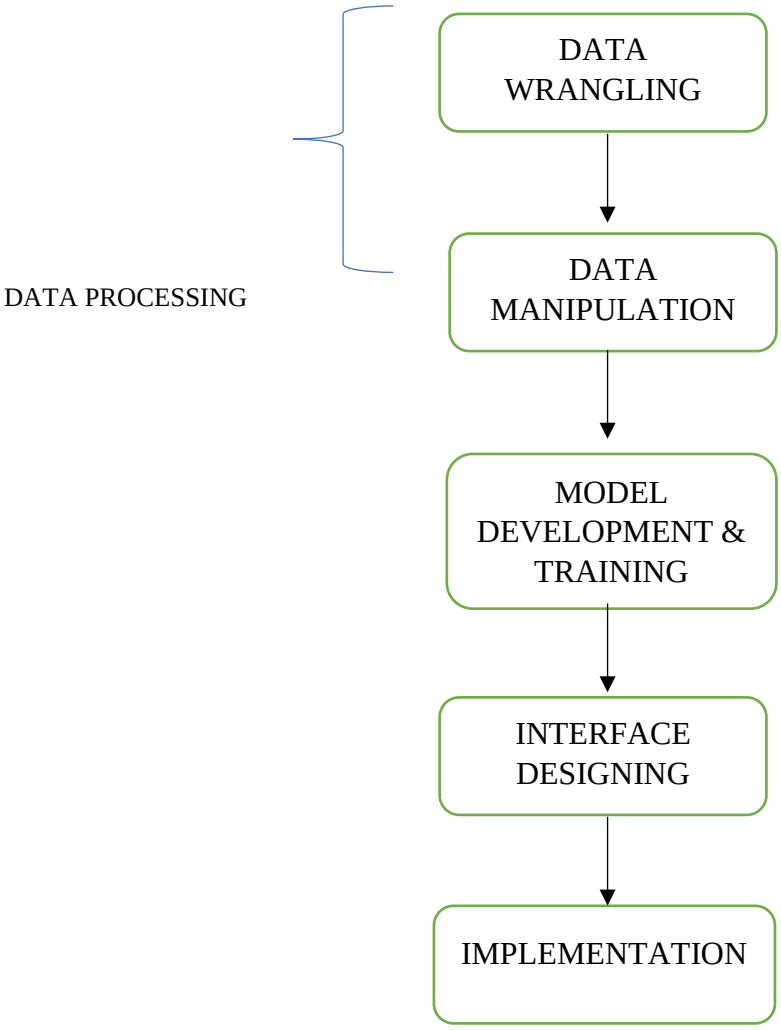

Fig 1.3: Integrated platform development steps / stages

Data Wrangling

Before we even think about developing a model for time-series forecast we need to clean the data else the efficiency of the outcome of the model may be sabotaged or even their functioning could be hindered. This process is called Data Wrangling. The following are the issues we expect to encounter along with the measures we took to remedy them.

1. Data-info mismatch: When the object type described, and entries of the column are not of the same type for e.g. numeric data type described as strings (pin code as strings etc.) Soln.: We used astype() attribute to amend / change the column data type.
2. Wrong type entry in a column: A string entered in a numeric type column is a big ambiguity. It may be a numeric in a string format therefore we use exception handling mechanism in python to convert them into the numeric form if they can be else if they are characters, we raise an exception and replace them with 'NULL'
3. 'NULL' / missing values: We remove the whole row entry if the missing / 'NULL' value is in any of the independent or target variables / columns.

Data Manipulation

To deploy / train / test our model we need to need to manipulate our data set in order to get some derived values which can be used as inputs / specifications into the model i.e. independent variables and dependent / target variables.

For the next part we need to understand what is time series data?

A time order indexed series of data points is known as time series. A time series can be broken down into 3 components:

- Trend: Upward and downward movement of data with time over a large period e.g. house appreciation.
- Seasonality: Seasonal variance. e.g. increase in demand of ice cream in summers or woollen clothes in winters.
- Noise: Spikes and troughs at random intervals.

We need to ensure the stationarity of a time series before applying any statistical modelling (machine learning model). For a time series to be stationary the following must be kept in mind:

1.The mean of the series should not be a function of time

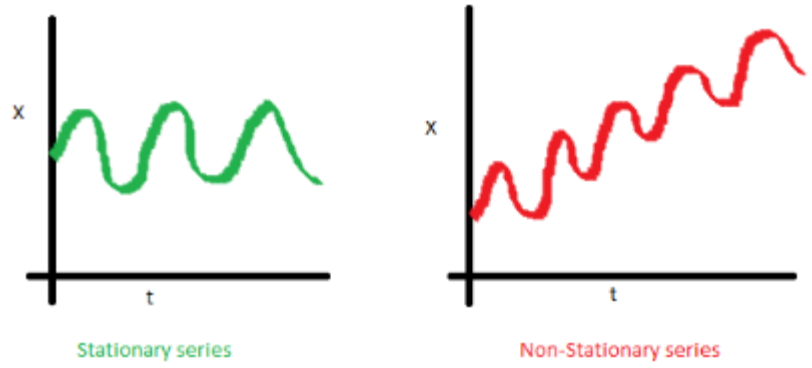

Fig 1.4 : Explanatory figures for constant and variable mean

2.The variance of the series should not be a function of time

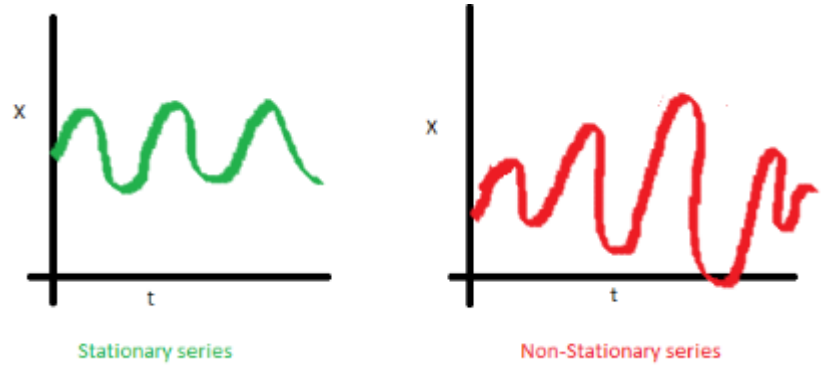

Fig 1.5 : Explanatory figures for constant and variable variance

3.The co-variance of the i^{th} term and $(i+1)^{th}$ term should not be a function of time. Sometimes the spread varies with time which causes this condition.

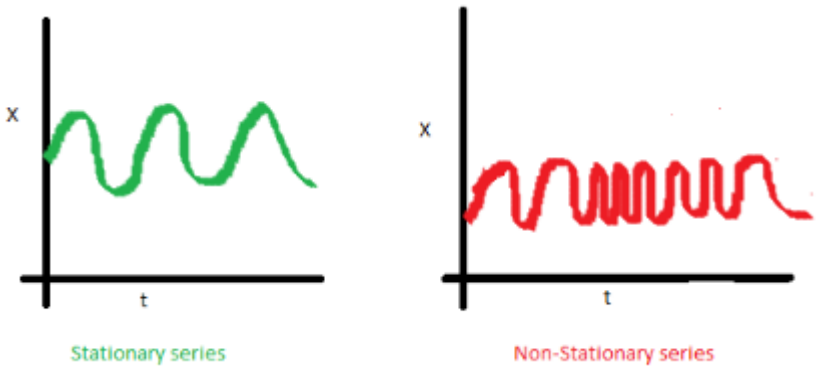

Fig 1.6 : Explanatory figures for even and uneven std. distribution

A stationary time series over a given period having a particular behaviour can be safely assumed to have the same behaviour at some later point in time. It's required by most of the statistical modelling methods that the time series be stationary.

For determining the stationarity of a time series there are two primary methods:

- Rolling statistics method: It's more of a visual method. We plot the rolling mean and rolling standard deviation. The time series is said to be stationary if the plot seems constant with time(straight and parallel to x-axis)
- Augmented Dickey-Fuller test: According to null hypothesis time series is considered stationary if the p-value is low and the critical values at 1%, 5% and 10% confidence intervals are as close as possible to the ADF - statistics. P-threshold is around 0.05

Taking the log() of the dependent variable is a simple way of lowering the rate at which rolling mean increases and hence make the time series stationary
To render a time series stationary we can also subtract the rolling mean, apply exponential decay OR when applying time shift we subtract every point by the one that preceded it
($x_i - x_{i-1}$).

Model Development and Training

Considering the various models we proceed with ARIMA (Auto-regressive integrated moving average) model. As we can notice it consists of the integration of two models (Auto-regressive and Moving average). We also explore it's extensions such as Seasonal ARIMA (SARIMA)
There is a collection techniques for manipulating and interpreting variables that depend on time. ARIMA is included among them which can remove the trend component in order to accurately predict future values.

Auto regressive model operates under the premise that past values have an effect on current values. AR models are commonly used in analysing nature, economics, and other time-varying processes. As long as the assumption holds, we can build a linear regression model that attempts to predict value of a dependent variable today, given the values it had on previous days. The order of AR model (p) incorporates the number of days to be incorporated in the formula.

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p}$$

Moving average model assumes that the value of the dependent variable on the current day depends on the previous day error terms. The model can be expressed as

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

We may also come across the equation,

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where μ is the mean of the series, the $\theta 1$, …, θq are the parameters of the model and the εt, $\varepsilon t-1$,…, $\varepsilon t-q$ are white noise error terms. The value of q is called the order of the MA model.

ARMA model: It is the combination of AR and MA models

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

ARIMA (Auto-regressive integrated moving average model)

ARIMA model (Box-Jenkins model) adds differencing to an ARMA model. Differencing subtracts the current value from the previous and can be used to make a time series stationary. First-order differencing addresses linear trends, and employs the transformation $z_i = y_i - y_{i-1}$. Second-order differencing addresses quadratic trends and employs a first-order difference of a first-order difference, namely $z_i = (y_i - y_{i-1}) - (y_{i-1} - y_{i-2})$, and so on.

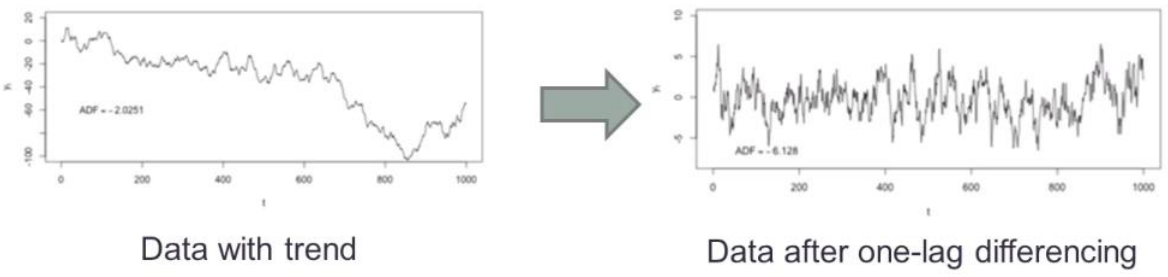

Fig 1.7: ARIMA model effect on a time series

Three integers (p, d, q) are required as parameters for the ARIMA model

- p: Number of Auto-regressive terms (AR-order)
- q: Number of non-seasonal differences (differencing order)
- r: Number of moving average terms (MA-order)

Auto Correlation Function (ACF)

The correlation between the observations at the current point in time and the observations at **all previous points in time**. We can use ACF to determine the optimal number of **MA** terms. The number of terms determines the order of the model.

Partial Auto Correlation Function (PACF)

As the name implies, PACF is a subset of ACF. PACF expresses the correlation between observations made at **two points in time** while accounting for any influence from other data points. We can use PACF to determine the optimal number of terms to use in the **AR** model. The number of terms determines the order of the model.

For example:

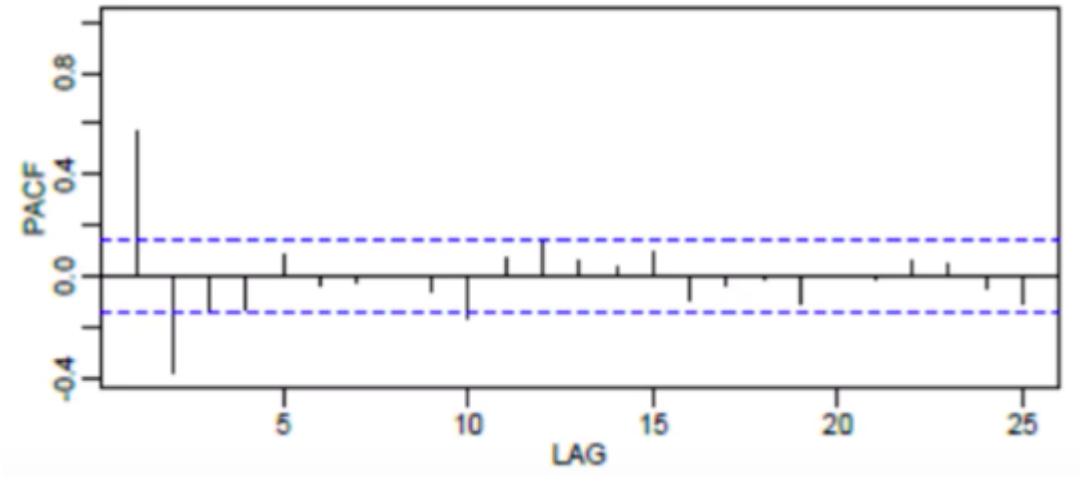

Fig 1.8: A PACF plot of a random time series

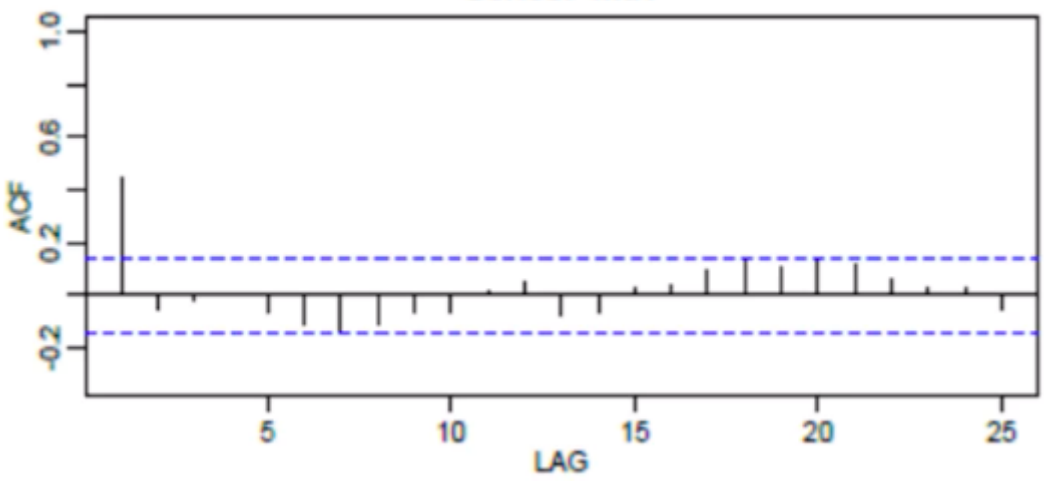

Fig 1.9: An ACF plot of a random time series

Let the horizontal blue lines represent the significance thresholds and the vertical lines be representing the ACF and PACF values at any point in time. Only those vertical lines which exceed the threshold are considered significant. Therefore, form the PACF example plot we derive that we would use only previous two days for the auto regression equation (AR-order =2) and for the ACF example plot we only use yesterday in the moving average equation (MA order = 1). We can set d at any value we desire for the number of days of differencing we want.

Then, we can see how the model compares to the original time series.

An extension of the ARIMA model is SARIMA model which includes one more component as a part of it's parameters

P: Seasonal autoregressive order.
D: Seasonal difference order.
Q: Seasonal moving average order.
m: The number of time steps for a single seasonal period.

ARIMA(p, d, q)

SARIMA(p, d, q)(P, D, Q)m

Eg: An m of 12 for monthly data suggests a yearly seasonal cycle

Another important component that we need to calculate to apply the SARIMA (Seasonal ARIMA) model is the AIC value (Akaike Information Criteria). It is a measure to compare statistical models as it quantifies the goodness of fit and parsimony/simplicity of a model into a single statistic. To weigh the parsimony and goodness of a model for different parameter values is the objective of the AIC value. The one with the least value is generally considered the best one.

IMPLEMENTATION

Before deploying the model we needed to establish the parameters. To do so we first select a range to be expected and acquired by the parameters. The following are the possibilities for a range of (0, 2).

```
p = d = q = range(0, 2)
pdq = list(itertools.product(p, d, q))
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))
```

```
Examples of parameter combinations for Seasonal ARIMA...
SARIMAX: (0, 0, 1) x (0, 0, 1, 12)
SARIMAX: (0, 0, 1) x (0, 1, 0, 12)
SARIMAX: (0, 1, 0) x (0, 1, 1, 12)
SARIMAX: (0, 1, 0) x (1, 0, 0, 12)
```

Fig 1.10: Generating possible parameter combinations

The AIC values for each of the combinations are generated using a loop and then the least yielding combination is selected.

```
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = st.tsa.statespace.SARIMAX(y,
                                            order=param,
                                            seasonal_order=param_seasonal,
                                            enforce_stationarity=False,
                                            enforce_invertibility=False)

            results = mod.fit()

            print('ARIMA{}x{}12 - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

Fig 1.11: Generating AIC values for all the parameter combinations

```
ARIMA(0, 0, 0)x(0, 0, 0, 12)12 - AIC:404.6159252042328
ARIMA(0, 0, 0)x(0, 0, 1, 12)12 - AIC:1475.02143190285

c:\users\sdjsr\appdata\local\programs\python\python37\lib\site-packages\statsmodels\base\model.py:568: ConvergenceWarning: Ma
ximum Likelihood optimization failed to converge. Check mle_retvals
  "Check mle_retvals", ConvergenceWarning)
c:\users\sdjsr\appdata\local\programs\python\python37\lib\site-packages\statsmodels\tsa\statespace\sarimax.py:868: UserWarnin
g: Too few observations to estimate starting parameters for seasonal ARMA. All parameters except for variances will be set to
zeros.
  ' zeros.' % warning_description)


ARIMA(0, 0, 0)x(0, 1, 0, 12)12 - AIC:247.77159581577172
ARIMA(0, 0, 0)x(0, 1, 1, 12)12 - AIC:156.38962142969527
ARIMA(0, 0, 0)x(1, 0, 0, 12)12 - AIC:254.84016907562062

c:\users\sdjsr\appdata\local\programs\python\python37\lib\site-packages\statsmodels\base\model.py:568: ConvergenceWarning: Ma
ximum Likelihood optimization failed to converge. Check mle_retvals
  "Check mle_retvals", ConvergenceWarning)
```

Fig 1.12: AIC values generated for all the parameter combinations

We deployed the model to predict the sales quantity of furniture in the Southern circle region. We trained the model to fit on 2012 - 2013 sales data on the basis of the sales in the years 2010-2011 and the following result was obtained.

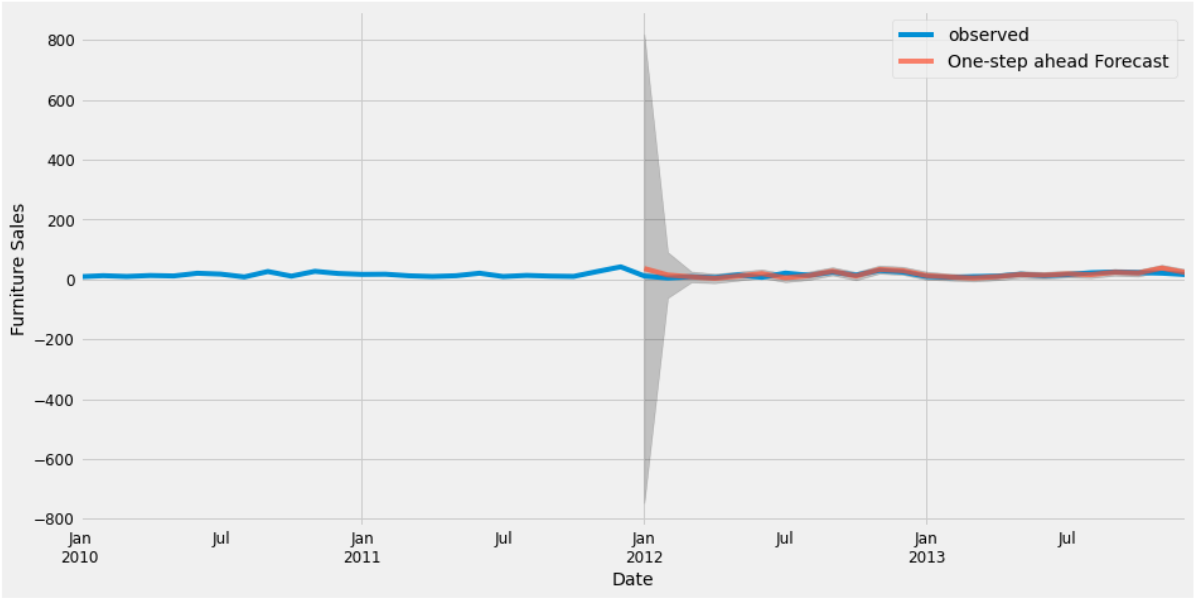

Fig 1.13: Time series of furniture quantity sales for 2013 (original / predicted)

To evaluate the model's efficiency we used RMS error method and the following results were yielded

Interface Designing:

The user interface is made using HTML. It consists of two main pages. The first page of the Superstore Sales Analysis page consists of two drop down menu buttons. The initial value of both the dropdown menu will be general i.e. the overall database will be considered unless a particular option has been selected. The two drop down menu buttons are further named as:

1.Region Selector: This button helps us to select different regions among the four: East, West, Central and South..

2.Product Category Selection: This consists of the office supplies, technology and furniture.

On selecting a particular region/particular product, two graphs will be plotted according to the overall database first. The 1st graph shows us the sales of that chosen region/product in series of time and the 2nd graph shows us the profit earned and the expected profit in the time series i.e. the previous and predicted in both the cases.

Note- We cannot choose both at once i.e. the region and the product category.

Once we are redirected to the second page, we can that it consists of two dropdown boxes again where a 3rd graph will be plotted according to the quantity ordered for the chosen product category i.e. previous and predicted. If region wasn't selected in the first page, the dropdown box in the 2nd page will show region and if it was selected the dropdown box will show States as an option.

EXPERIMENTAL RESULTS

```
y_forecasted = pred.predicted_mean
y_truth = y['2012-01-01':]

mse = ((y_forecasted - y_truth) ** 2).mean()

print('The Root Mean Squared Error of our forecasts is {}'.format(round(np.sqrt(mse), 2)))
```

```
The Root Mean Squared Error of our forecasts is 8.16
```

Fig 1.14 : RMS error

The RMS error value came equivalent to 8.16 which is scarce and that is a good sign for a forecast model for furniture sales quantity in the southern circle in the year 2012-2013.

Besides these we can also examine past sales features such as trends, seasonality, noise distribution etc. For the same dataframe (Southern circle) we obtain the following:

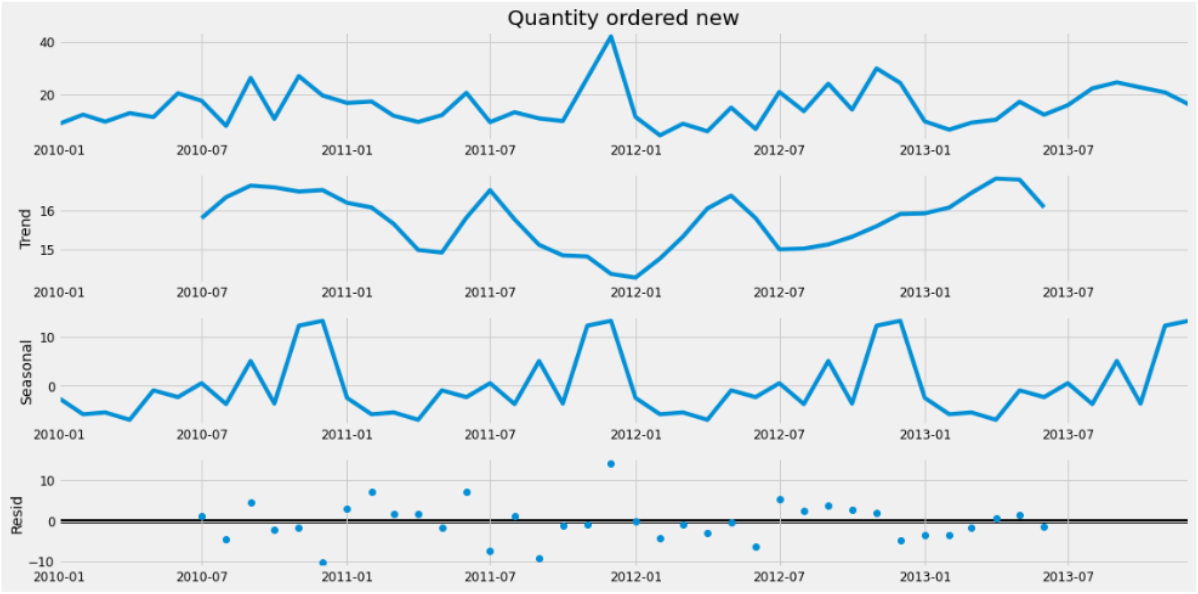

Fig 1.15 Trends and seasonal sales of furniture in the southern circle

REFERENCES

[1] Name : Time-series sales forecasting for an Enterprise Resource Planning system (2019)

Author : Toni Malila

https://www.theseus.fi/handle/10024/173086

[2] Name : Time series forecasting using a ARIMA and Neural network model. (2001)

Author : G. Peter Zhang

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.3756&rep=rep1&type=pdf

[3] Name : Forecasting aggregate retail sales:A comparison of artificial neural networks and traditional methods (2001)

Author : Ilan Alon , Min Qi , Robert J. Sadowski

https://www.researchgate.net/publication/222541007_Forecasting_aggregate_retail_sales_A_comparison_of_artifcial_neural_networks_and_traditional_methods

[4] Name : A moving average filter based hybrid ARIMA–ANN model for forecasting time eries data (2014)

Author : C. Narendra Babu , B.Eswara Reddy

https://www.researchgate.net/publication/263427920_A_moving-average_filter_based_hybrid_ARIMA-ANN_model_for_forecasting_time_series_data

[5] Name : Statistical and Machine Learning forecasting methods: Concerns and ways forward (2018)

Author : Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos.

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194889