Walmart Project

Capstone Project 1

Gokul venugopal

8/15/23

DS AI INTELLIPAAT

Table of Contents

Problem Statement

A Retail Company with multiple outlets stores is having poor revenue returns from retail store with most of them facing bankruptcy. This project undertakes to review sales records from the stores with a view to provide useful insights to the company and also to forecast sales outlook for the next 12-weeks.

Project Objective

A Retail Company with multiple outlets across the country are facing issues with inventory management. The task is to come up with useful insights to the company and also to forecast sales outlook for the next 12-weeks.

Data Description

The walmart.csv contains 6435 rows and 8 columns.

Feature Name	Description
Store	Store number
Date	Week of Sales
Weekly_Sales	Sales for the given store in that week
Holiday_Flag	If it is a holiday week
Temperature	Temperature on the day of the sale
Fuel_Price	Cost of the fuel in the region
CPI	Consumer Price Index
Unemployment	Unemployment Rate

From the given data set of the company, it is observed that the data consist of Six thousand four hundred and thirty- five (6345) records with seven features (captured weekly) as follows.

Stores: There are 45 stores and each store has 143 entries with below information.

- Date (Weekly)
- Total sales record for the week
- Holiday Flag (1 or 0)
- Temperature: Average Temperature recorded during the week.
- Fuel Price: Average price of Fuel for the week
- CPI: Consumer Price Index for the week.
- Unemployment: Rate of Unemployment for the week

Data Preprocessing Steps and Inspiration

The Preprocessing steps included the following steps:

Step 1: Load Data
Step 2: Perform Exploratory Data Analysis

a. Check number of records and its distribution
b. Check Data types
c. Check for missing data, invalid entries and duplicates
d. Examine the correlation of the independent features with target (weekly sales) variable.
e. Check for outliers that are known to distort

Step 3: Model Predictions, two approaches of Time series model

a. ARIMA

b. SARIMAX

Step 4: Forecast

Step 5: Compare Results from different model.

Choosing the Algorithm for the Project

Model Selection

Examination of the plot of weekly sales shows continuously time varying data (as shown above). A Time series model (ARIMA, SARIMAX) will be employed for the prediction and forecast.

ARIMA

An Autoregressive Integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. To better comprehend the data or to forecast upcoming series points, both of these models are fitted to time series data.

Assumptions on which ARIMA model is based:

1. Data does not contain anomalies
2. Model Parameter and error term is constant
3. Historic timepoints dictate behaviour of present timepoints
4. Time series is stationary.

SARIMAX

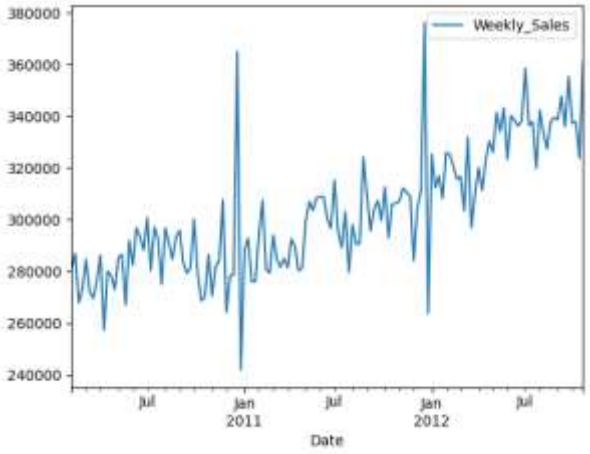
SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors) is an updated version of the ARIMA model. we can say SARIMAX is a seasonal equivalent model like SARIMA and Auto ARIMA. it can also deal with external effects.

Model Technique and Evaluation

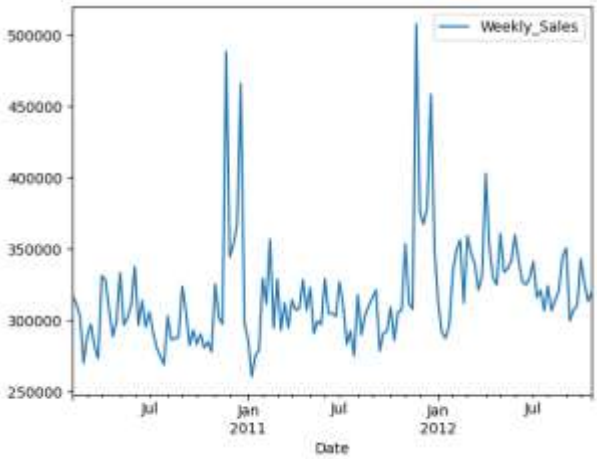
Plot of Weekly Sales against Time for selected stores

Highest 3 Weekly Sales: Store 33

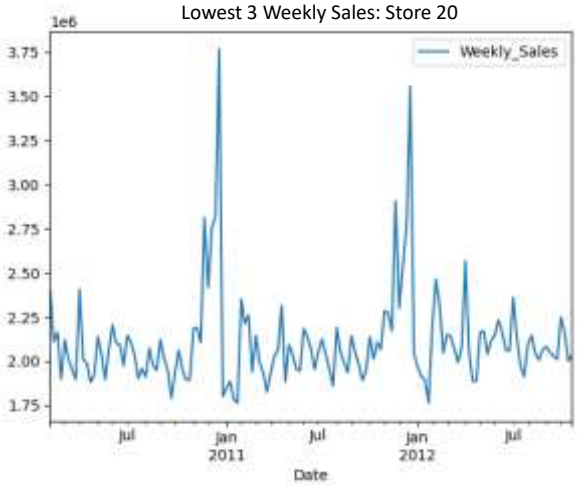
Highest 3 Weekly Sales: Store 4



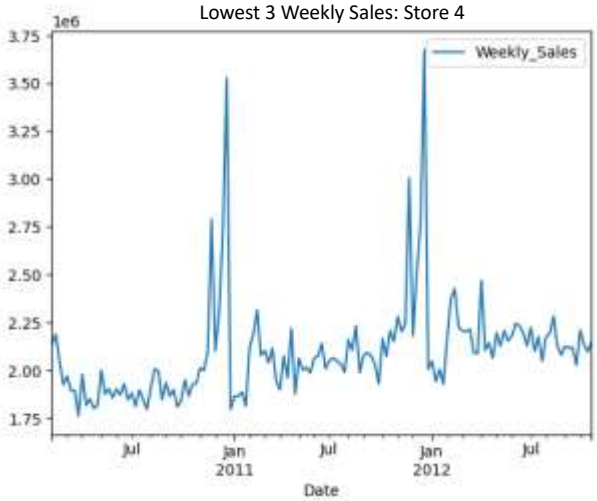
Highest 3 Weekly Sales: Store 5



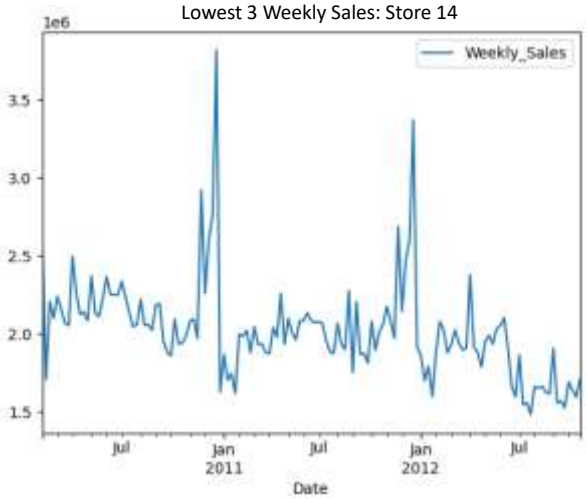
Lowest 3 Weekly Sales: Store 20



Lowest 3 Weekly Sales: Store 4



Lowest 3 Weekly Sales: Store 14



Model Evaluation

The following techniques and steps were involved in the evaluation of the model.

1. Load necessary Libraries
2. Load the dataset
3. Perform Exploratory Data Analysis (EDA) on the dataset.
    a. Find the shape and size of the data
    b. Check for invalid and null entries
    c. Explore Data description

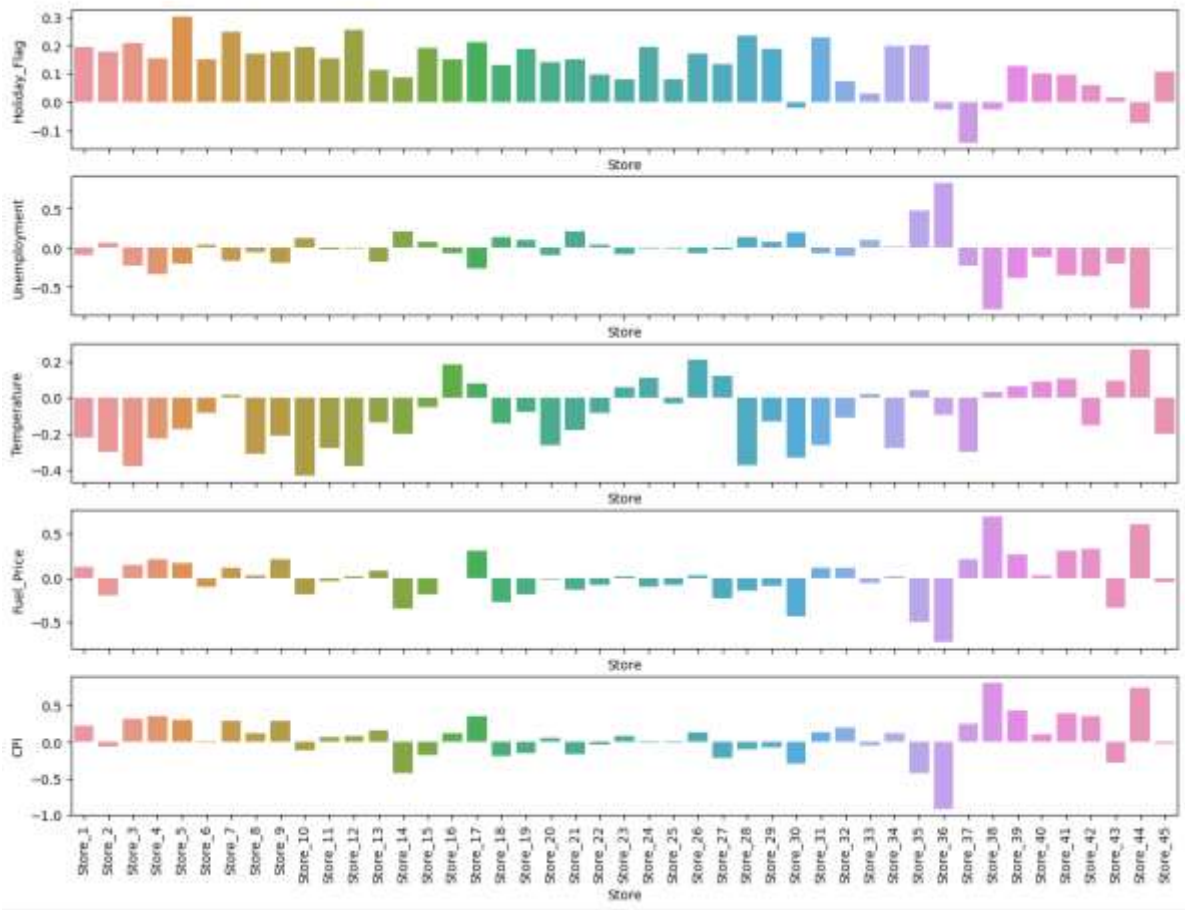
d. Examine the correlation of independent variables to the target variable (Weekly sales)

e. Line plot of the effects of independent variables to the target variable (Weekly sales)

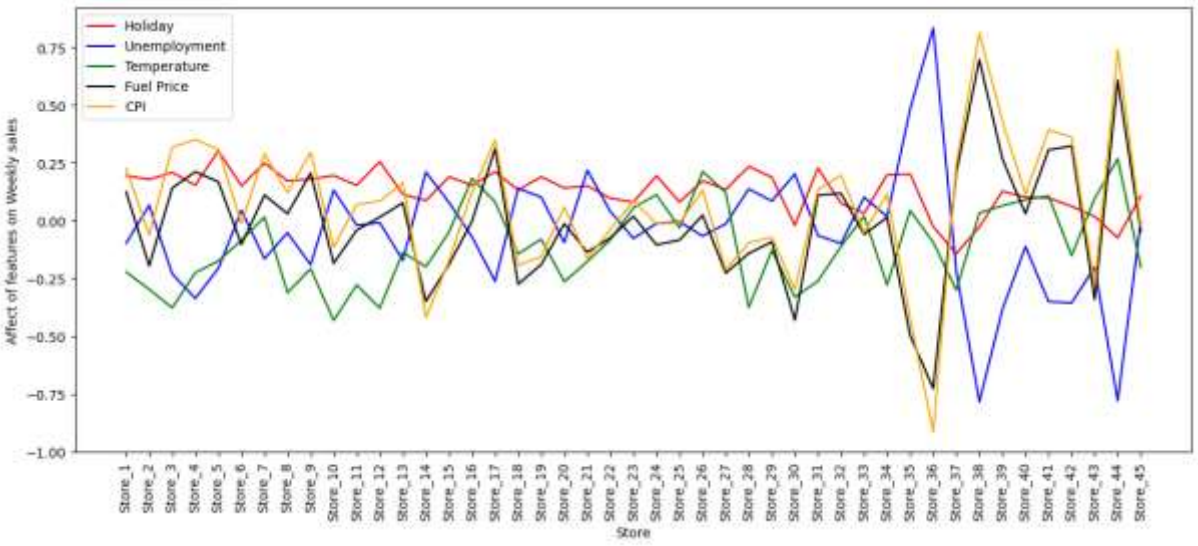
4.Model Prediction

5. Forecast

Model Design



It was observed from the EDA that the effects of the Independent Features on the target variable. Some stores are affected greatly by unemployment, while some affected by Fuel price. This type of fluctuating data will be very hard to build a general model for all stores. As seen in above graph, only 4 features namely had good correlation (Threshold > +/- 0.2) for some of the stores. The 4 features are Unemployment, Fuel Price, CPI and Temperature. Also, CPI was observed to be closely following Fuel price which indicates same effect on the weekly sales by both the features. But both these are uncontrollable factor so cannot be taken into consideration for improving sales. So, 2 features i.e., Temperature and Holiday will be taken into consideration for insights in each store.



Based on the findings above, the decision was to take only 4 features i.e., Temperature, Unemployment, Fuel Price and Holiday will be taken into consideration for insights in each store.

For simplicity and ease of presentation, I have decided to take to limit my predictions from for Top 3 and bottom 3 stores based on weekly sales revenue.

That notwithstanding, the model could be always be used to provide prediction for each store.

Model Approach

Timeseries Model, ARIMA

a. First step is to check the stationarity of the dataset using Adfuller test and KPSS test. Also using Visual method to confirm the stationarity
b. Next is to plot ACF and PACF graph for finding the orders (p, d, q) manually.
c. To find the best ARIMA orders, we will be using AutoARIMA to get the best orders.
d. SARIMAX will be used if seasonality is found the dataset of each respective stores.
e. Predict using SARIMAX and ARIMA and based on lowest error select the best model for the respective store.
f. Forecast 12-week sales using the best model.

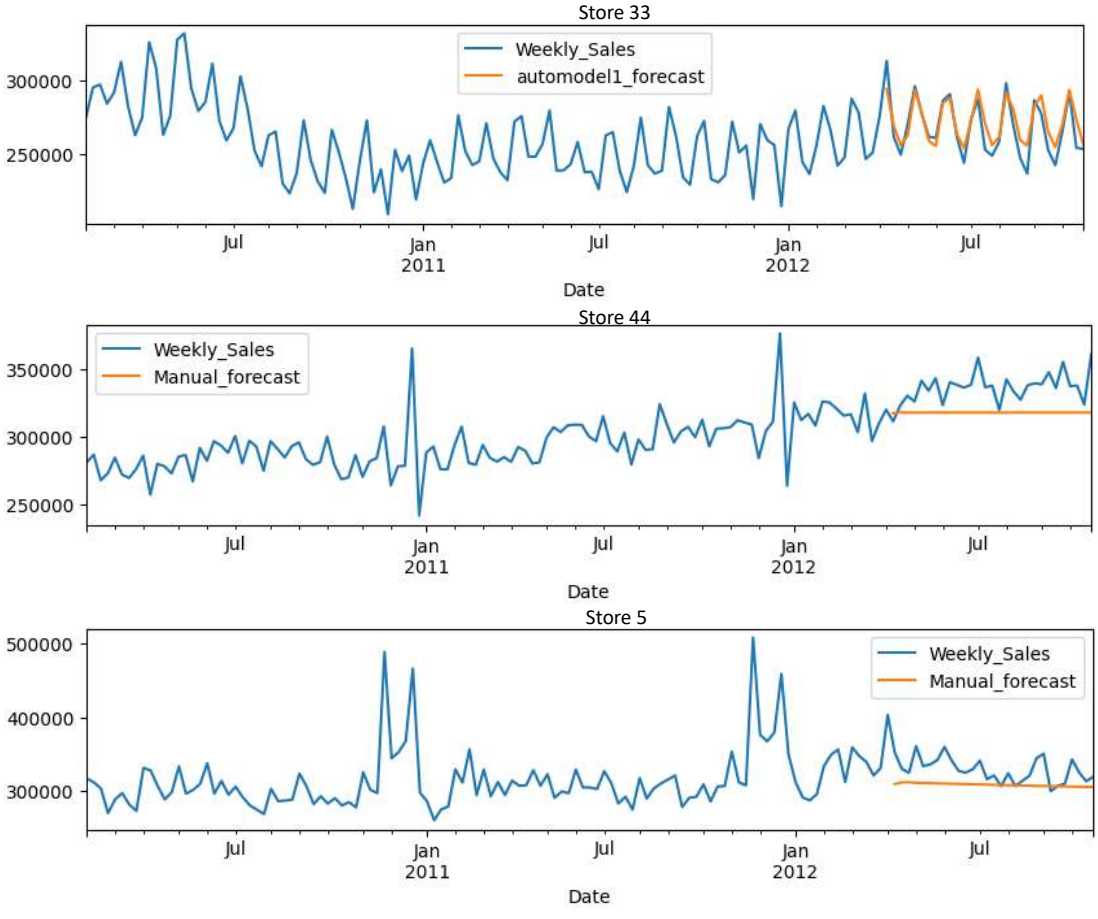
Inferences from the Project

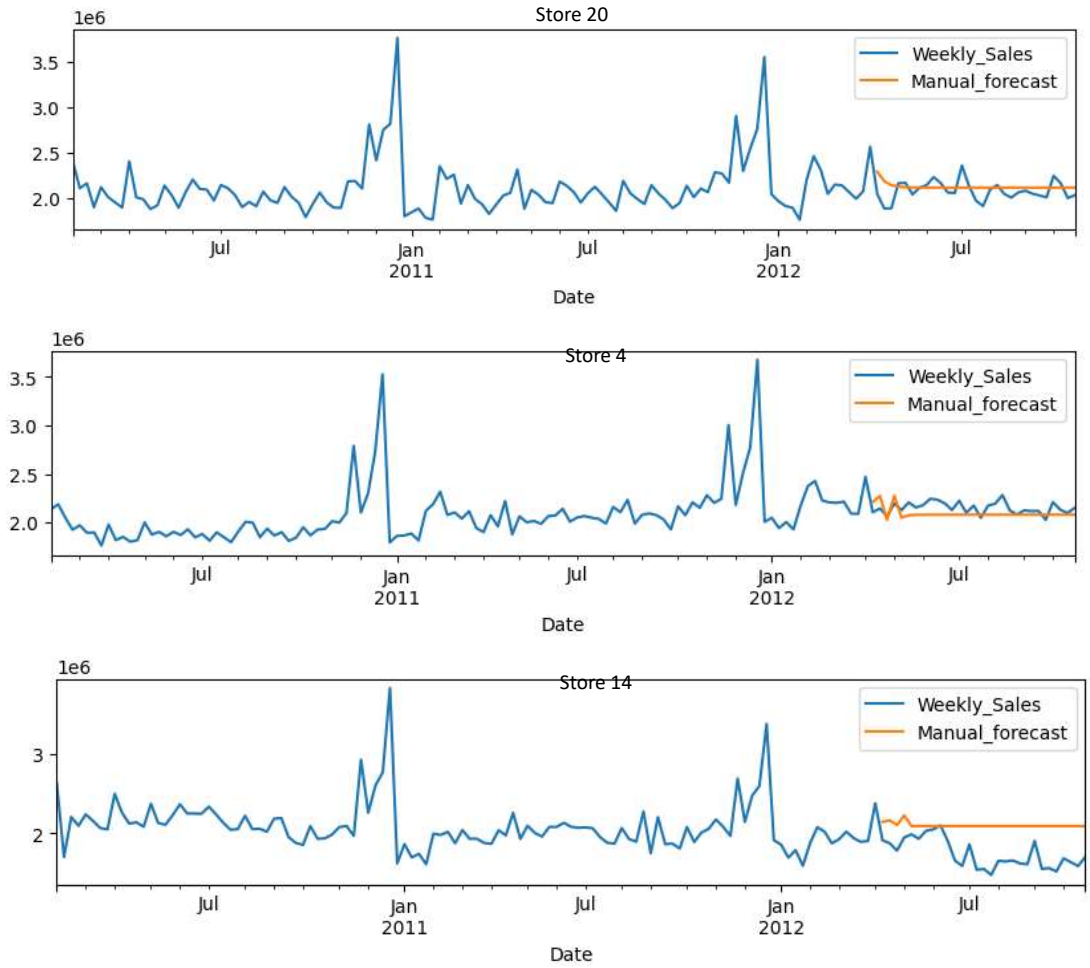
1. ARIMA Model:

a. Predictions

Predictions were performed for 6 stores (Top 3 [33, 44, 5] and Bottom 3 [20, 4, 14]in order of decreasing Weekly sales) using ARIMA. The predictions are summarized in the Table and Graphs below.

Store No	ARIMA		
	MAE	MAPE	RMSE
33	7592.86	0.0289	9434.7
44	19003.03	0.055	21476.02
5	27928.59	0.082	30236.67
20	97959.87	0.0479	125459.2
4	77797.49	0.0356	95336.36
14	356805.3	0.216	396174.1

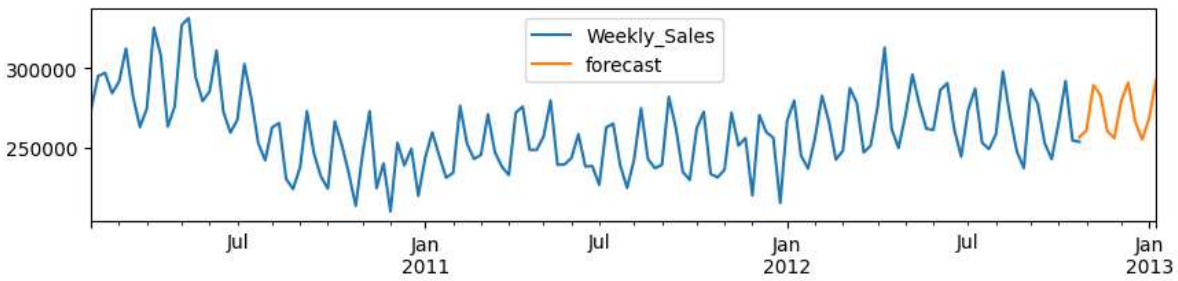




ARIMA was performing poorly for most of stores expect for Store 33 showing that seasonality, trend and noise is affecting the actual forecast which cannot be handled by simple ARIMA model. Due to this SARIMAX was also used to make better fit model for the timeseries data.

b. Forecast

Forecasting for Store 33 using ARIMA model

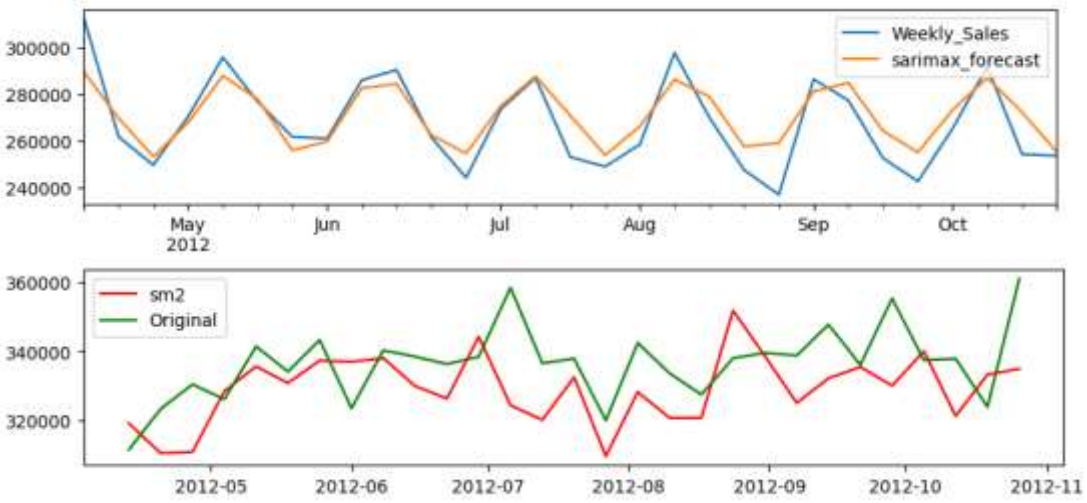


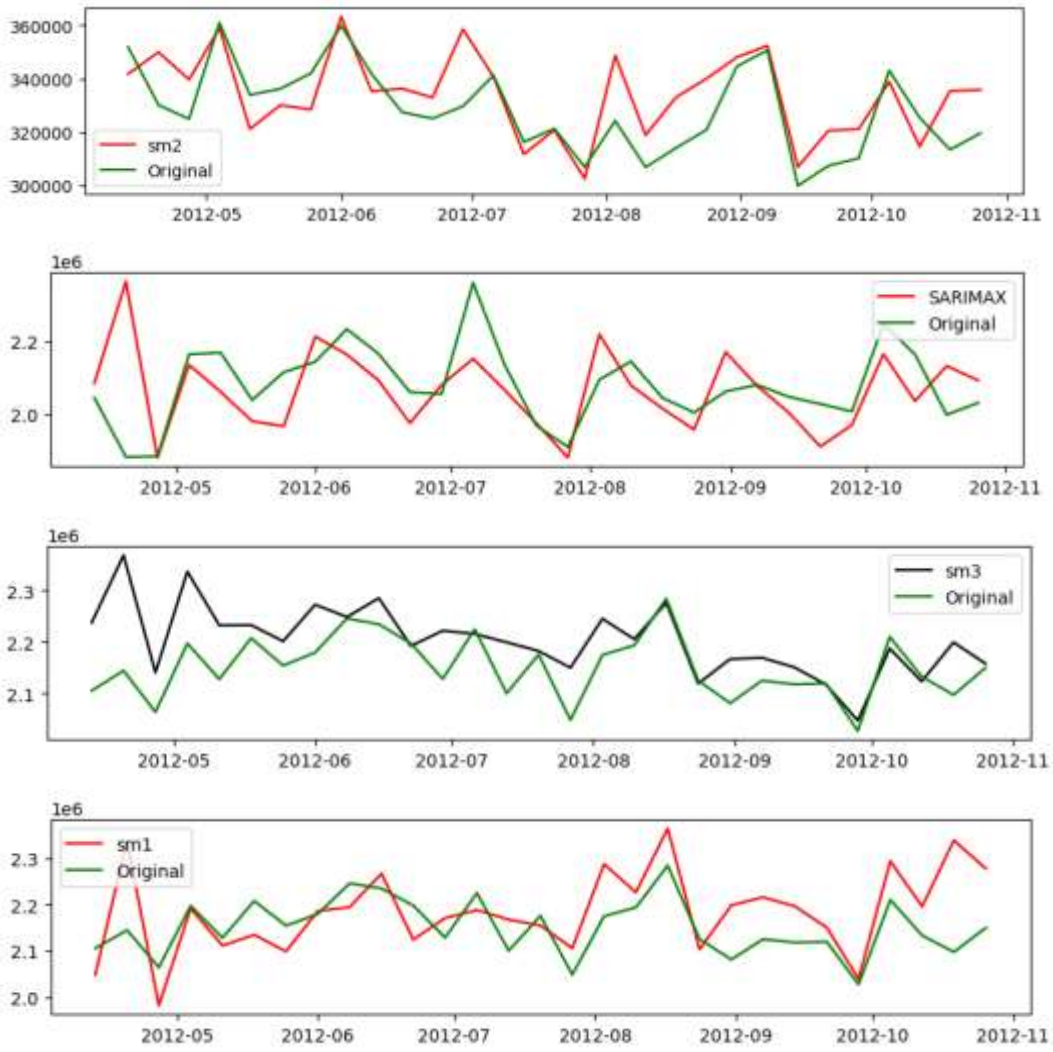
2. SARIMAX Model:

a. Predictions

Predictions were performed for 6 stores (Top 3 [33, 44, 5] and Bottom 3 [20, 4, 14]in order of decreasing Weekly sales) using SARIMAX. The predictions are summarized in the Table and Graphs below

Store No	SARIMAX		
	MAE	MAPE	RMSE
33	7848.55	0.0296	9907.41
44	11143.8	0.0327	13658.97
5	11543.12	0.0356	13232.41
20	84831.59	0.041	121991.9
4	56267.09	0.02639	77335.1
14	92897.16	0.043	107091.3



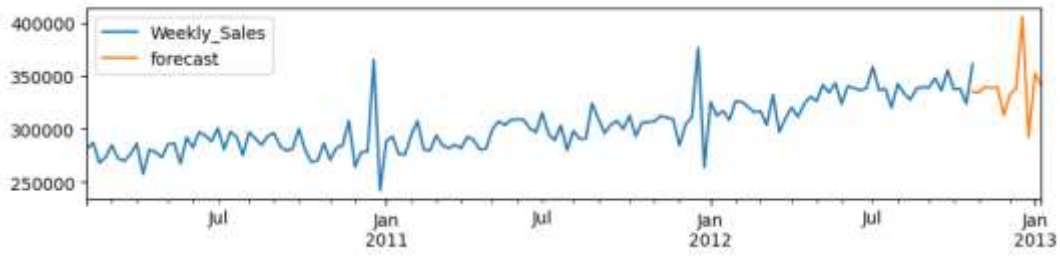


As seen the SARIMAX performs better as compared to ARIMA model and it is clearly evident using the different metrics like RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error).

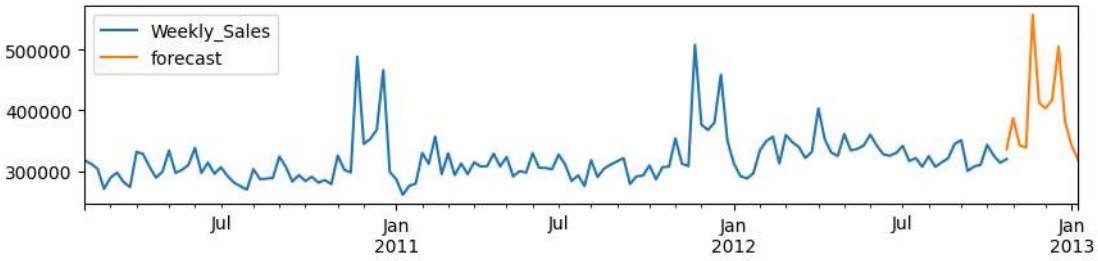
b. Forecast

Forecasting for the remaining 5 stores using SARIMAX model are given below:

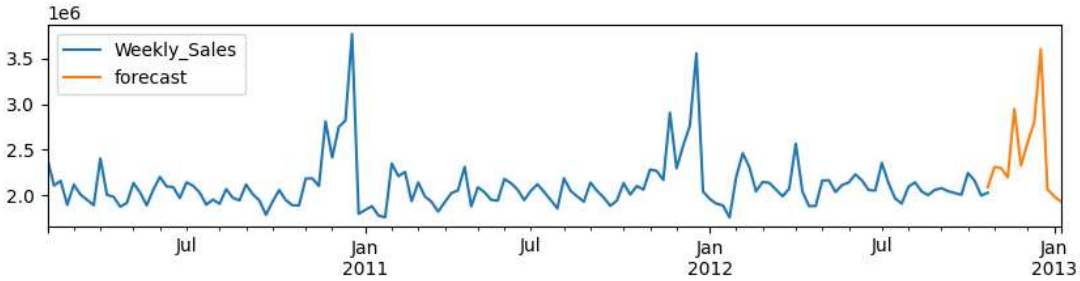
Forecasting for Store 44 using ARIMA model



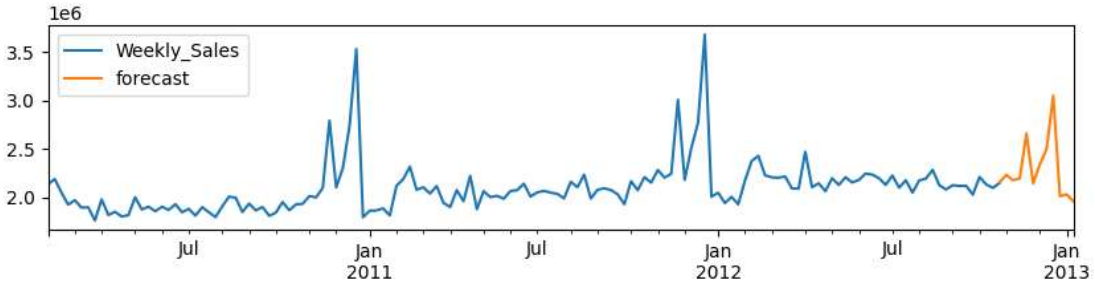
Forecasting for Store 5 using SARIMAX model



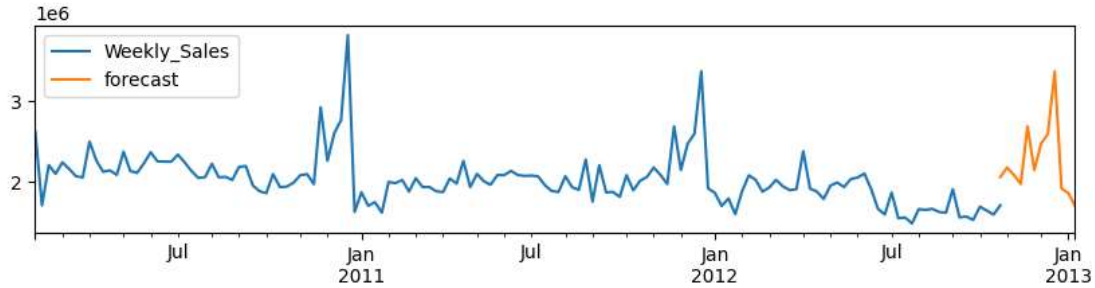
Forecasting for Store 20 using SARIMAX model



Forecasting for Store 4 using SARIMAX model

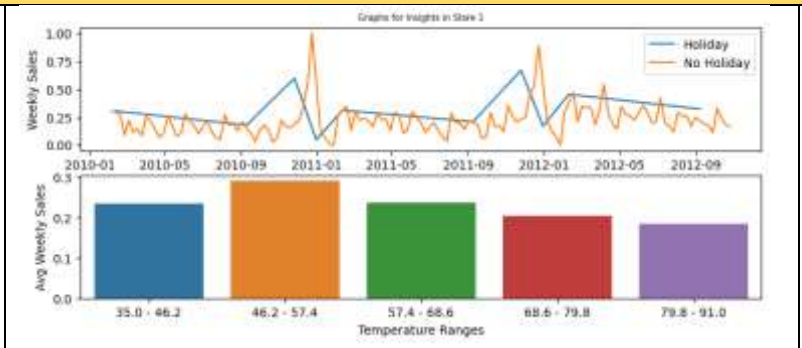
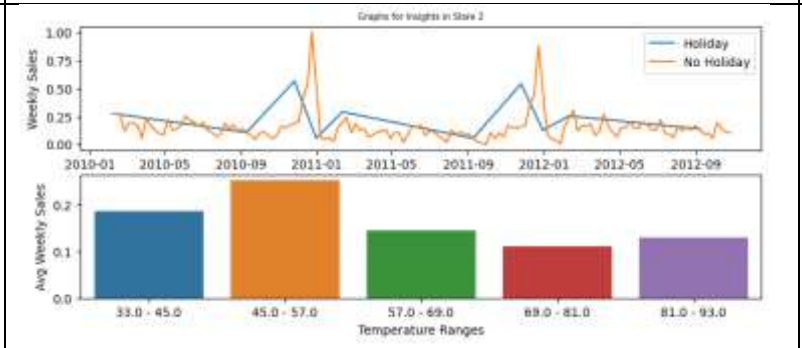
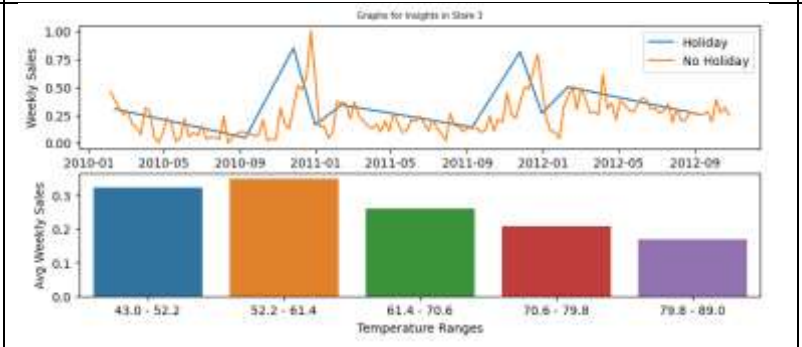
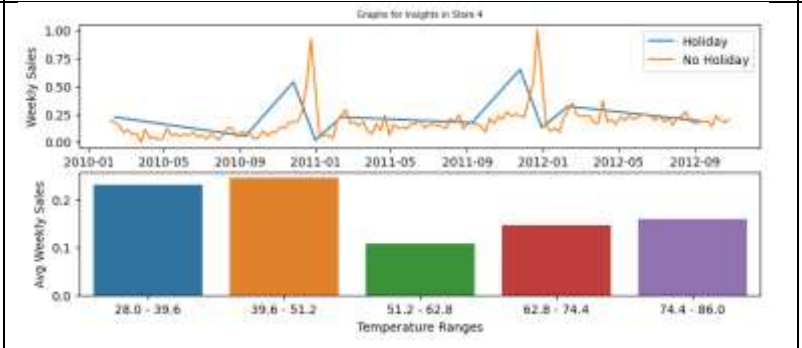
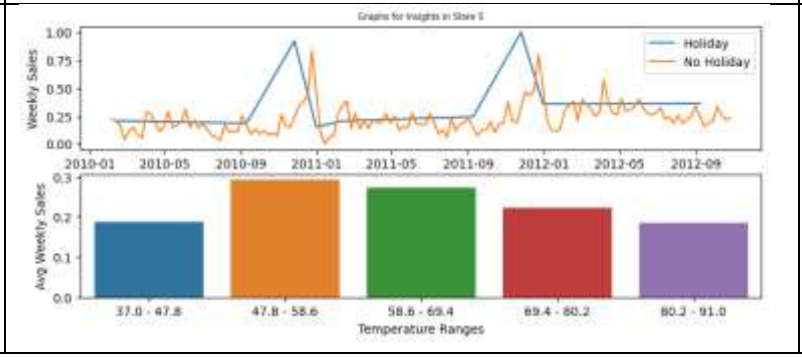


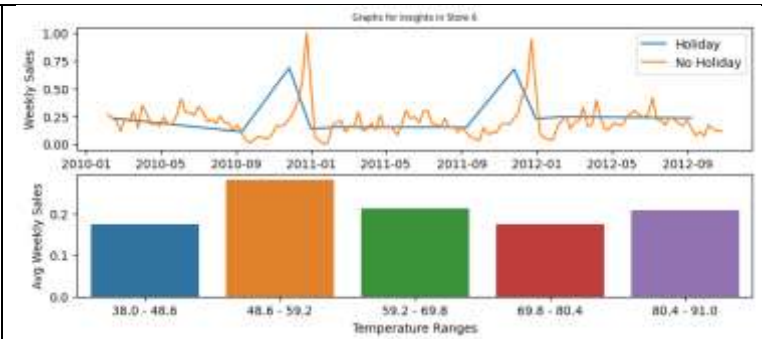
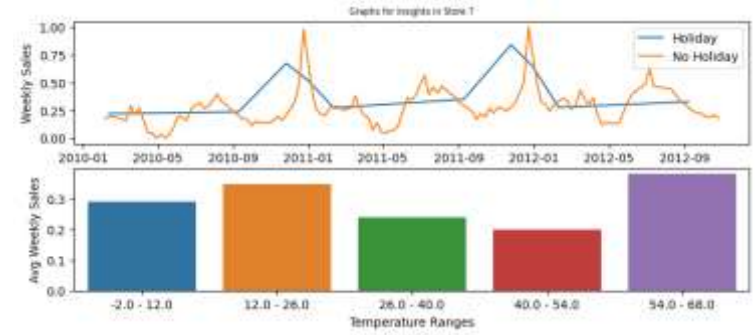
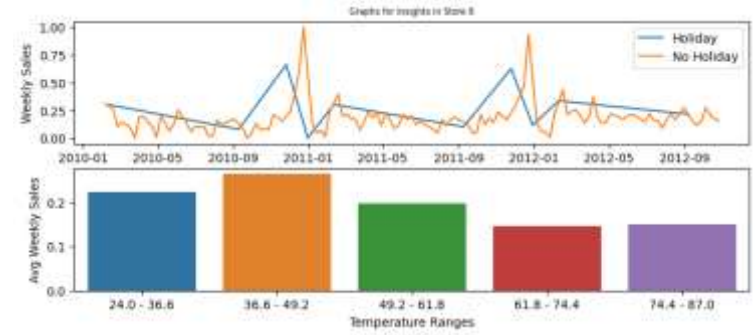
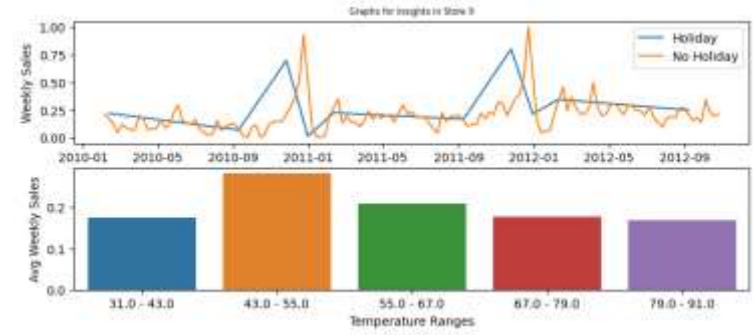
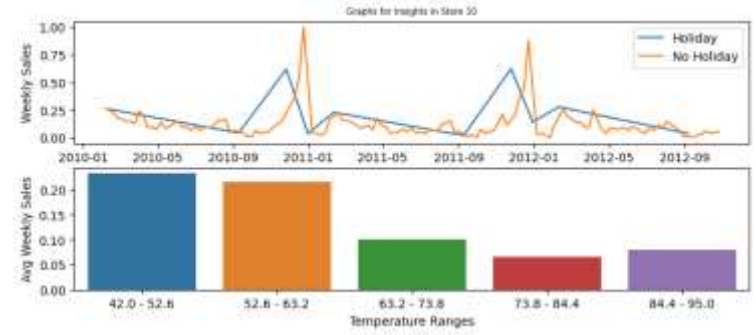
Forecasting for Store 14 using SARIMAX model

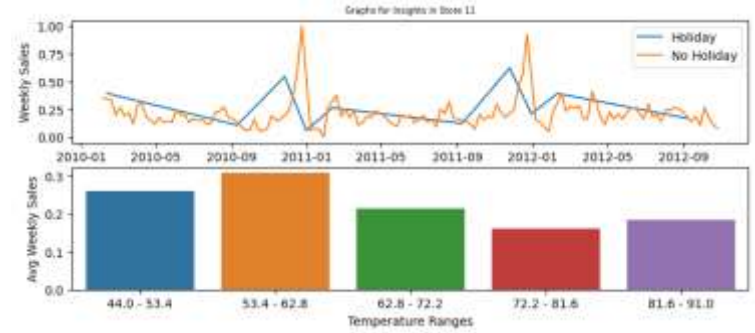
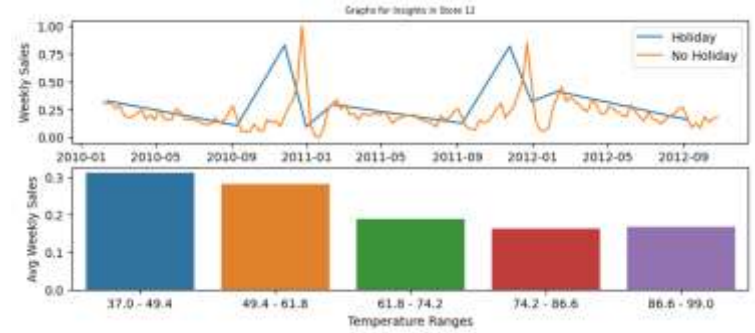
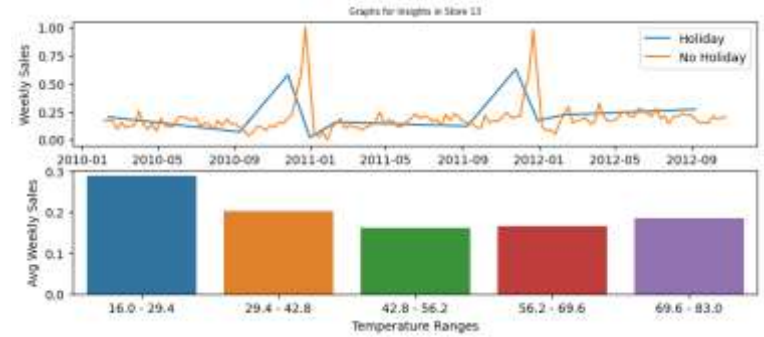
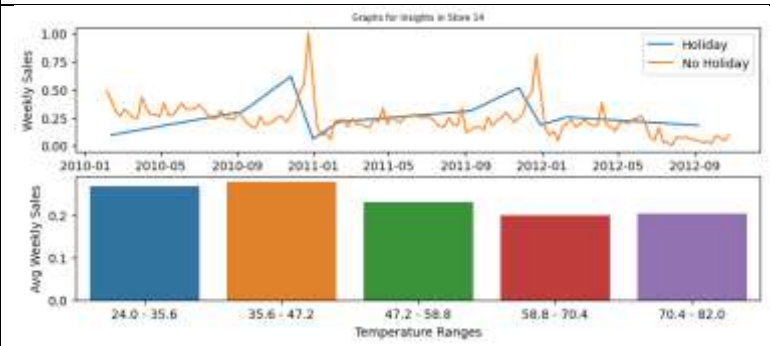
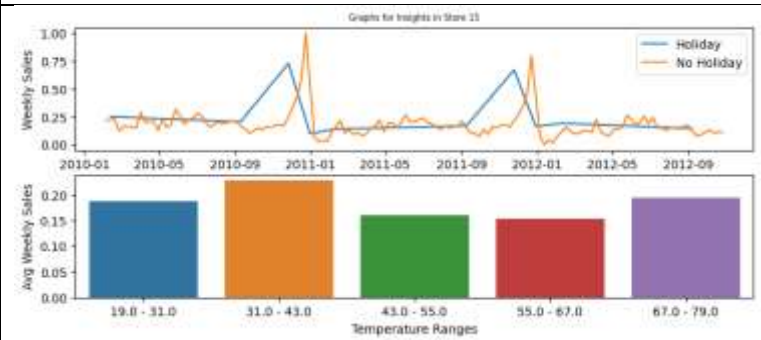


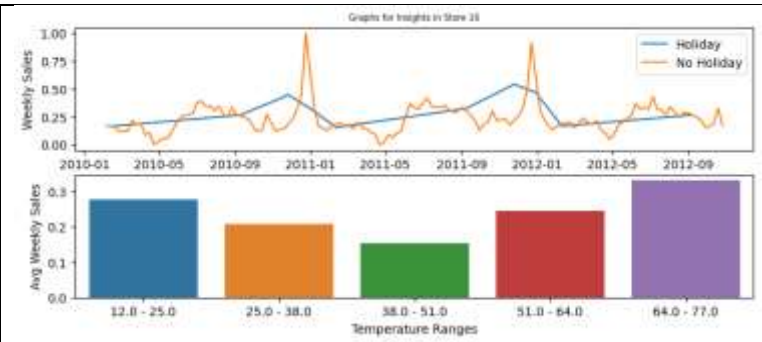
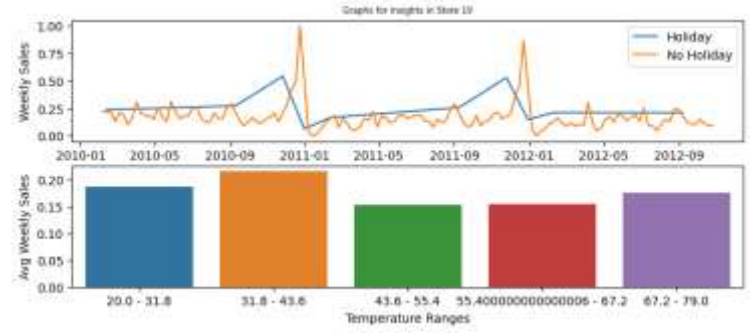
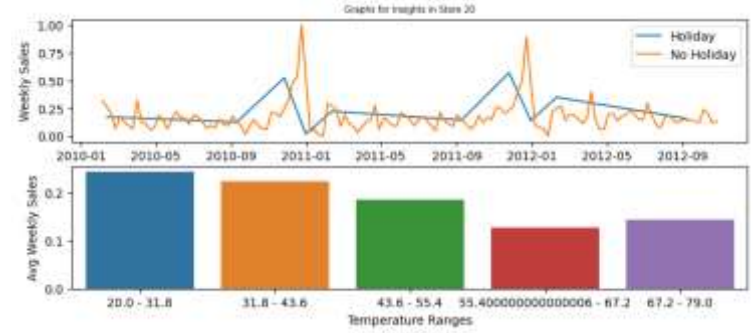
The forecast shows anticipated variabilities as observed in the Weekly sales history. However, overall current sales for stores studied so far will not be enough to save the company from bankruptcy.

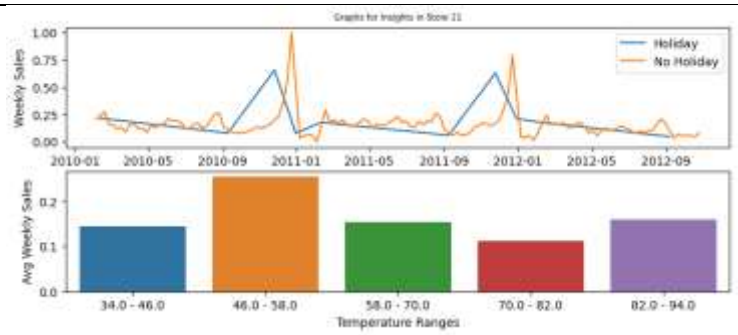
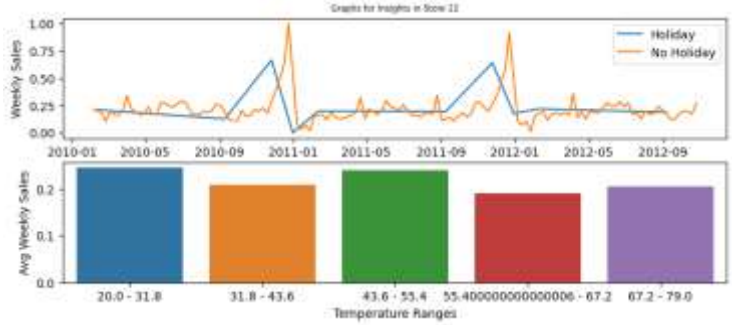
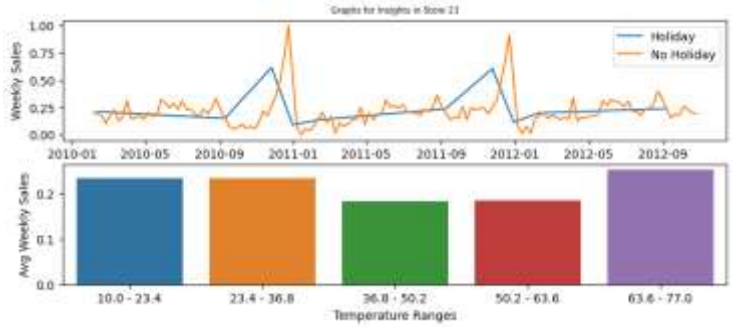
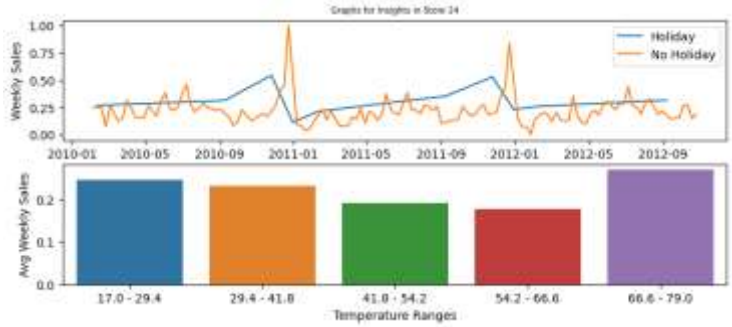
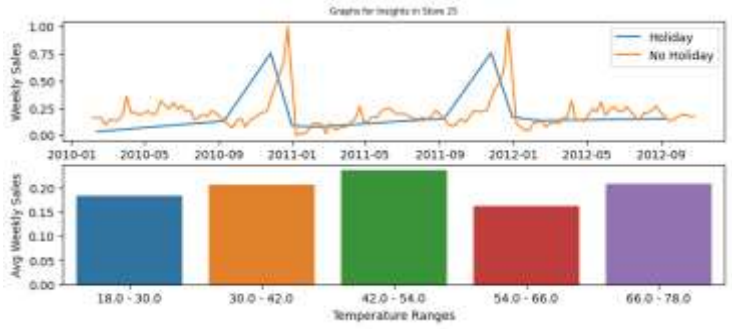
Outlook Recommendations for Stores

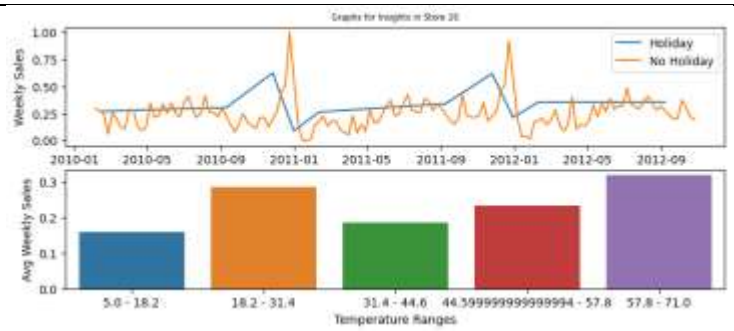
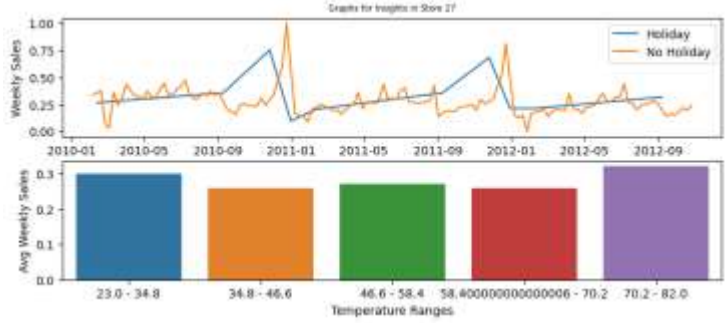
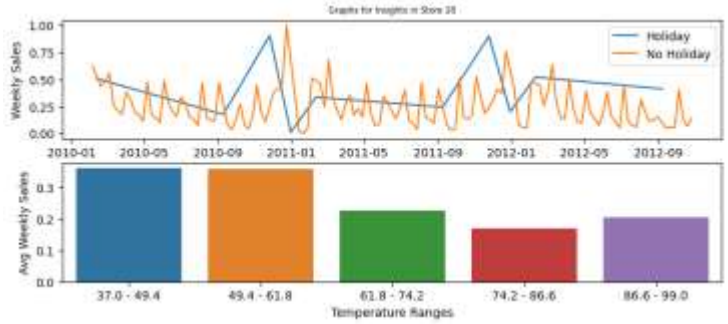
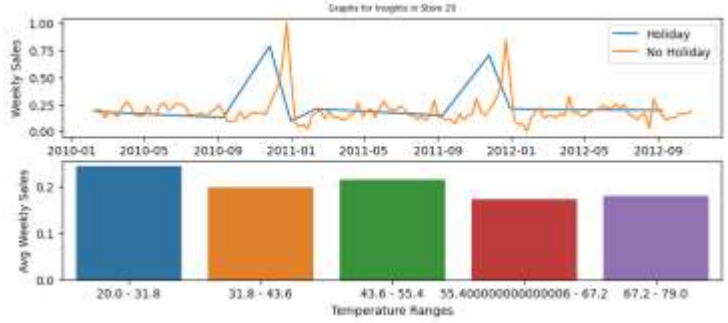
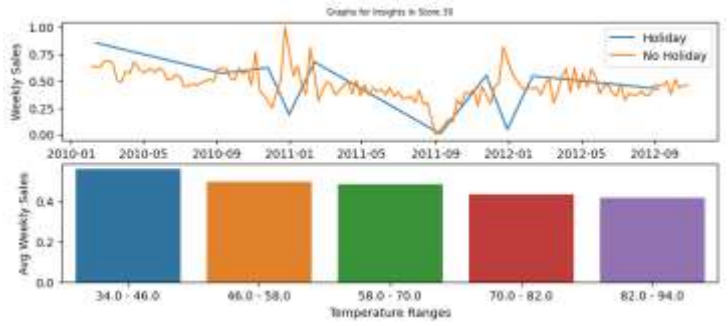
1		• Most Sales during Non-Holidays. • Most Sales during end of year • Increasing Sales during 46.2-57.4-degree Temperature • Stock up inventory for above mentioned time periods
2		• Most Sales during Non-Holidays. • Most Sales during end of year • Increasing Sales during 45-57.0-degree Temperature • Stock up inventory for above mentioned time periods
3		• Most Sales during both holidays and non-Holidays. • Most Sales during end of year • Increasing Sales during 43.0-61.4-degree Temperature • Stock up inventory for above mentioned time periods
4		• Most Sales during Non-Holidays. • Most Sales during end of year • Increasing Sales during 28.0-51.2-degree Temperature • Stock up inventory for above mentioned time periods
5		• Most Sales during Holidays. • Most Sales during end of year • Increasing Sales during 47.0-58.6-degree Temperature • Stock up inventory for above mentioned time periods

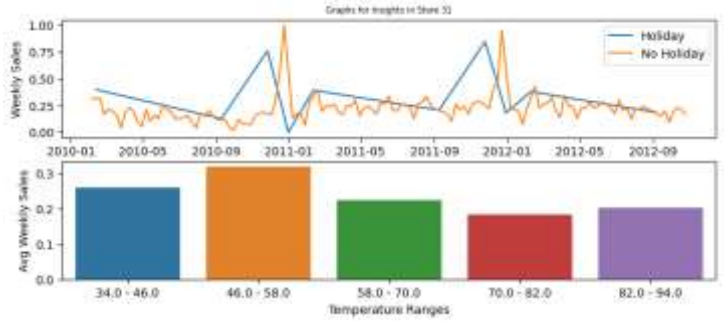
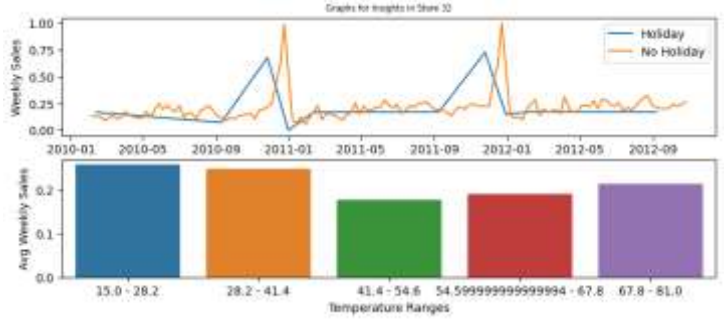
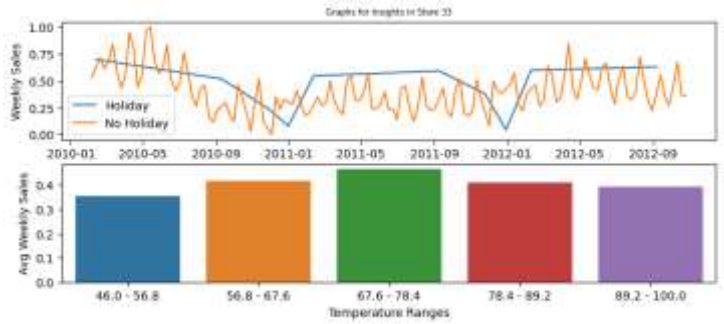
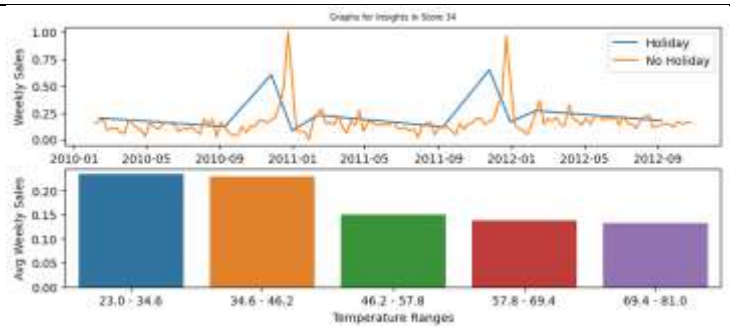
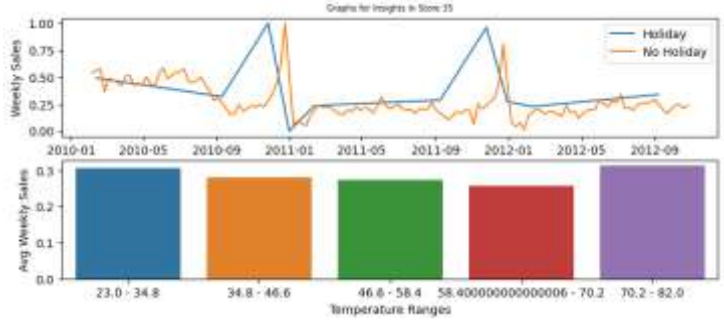
6		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 48.6-59.2-degree Temperature • Stock up inventory for above mentioned time periods
7		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 54.0-68.0-degree Temperature • Stock up inventory for above mentioned time periods
8		• Most Sales during Non-Holidays. • Most Sales during end of year • Increasing Sales during 36.6-49.2-degree Temperature • Stock up inventory for above mentioned time periods
9		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 43.0-55.0-degree Temperature • Stock up inventory for above mentioned time periods
10		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 42.0-52.6-degree Temperature • Stock up inventory for above mentioned time periods

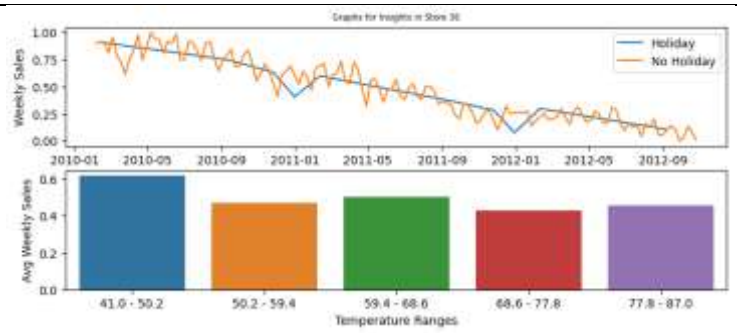
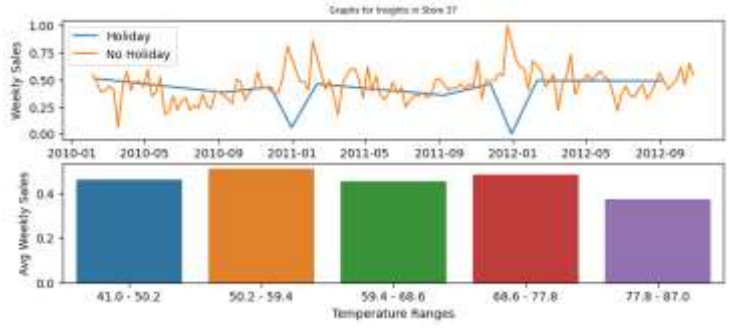
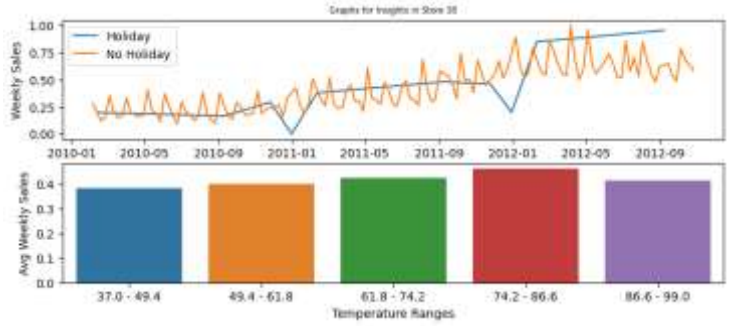
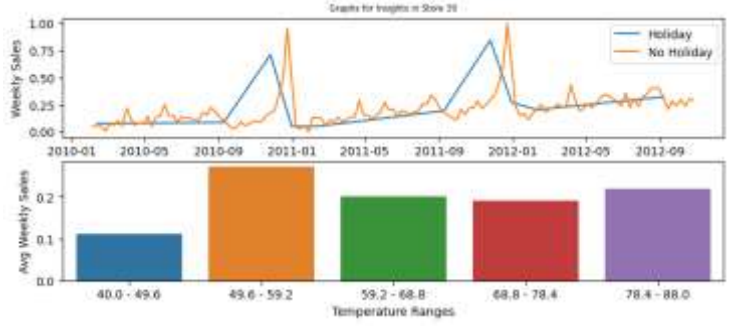
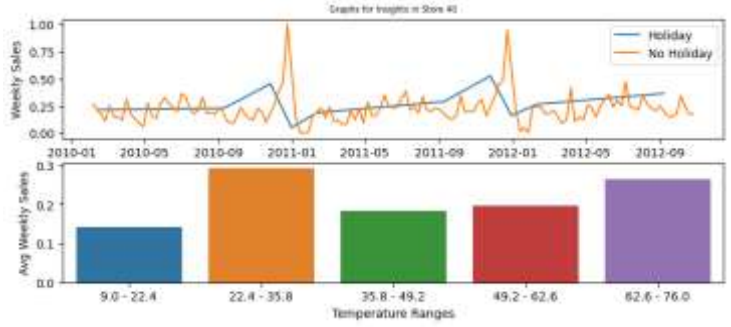
11		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 53.4-62.8-degree Temperature • Stock up inventory for above mentioned time periods
12		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Increasing Sales during 37.0-49.4-degree Temperature • Stock up inventory for above mentioned time periods
13		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 16.0-29.4-degree Temperature • Stock up inventory for above mentioned time periods
14		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 24.0-47.2-degree Temperature • Stock up inventory for above mentioned time periods
15		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 31.0-43.0-degree Temperature • Stock up inventory for above mentioned time periods

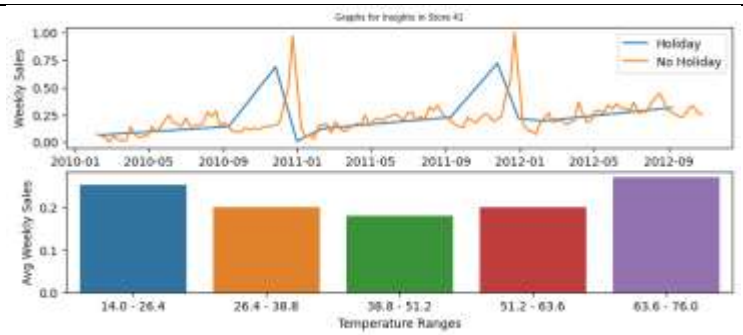
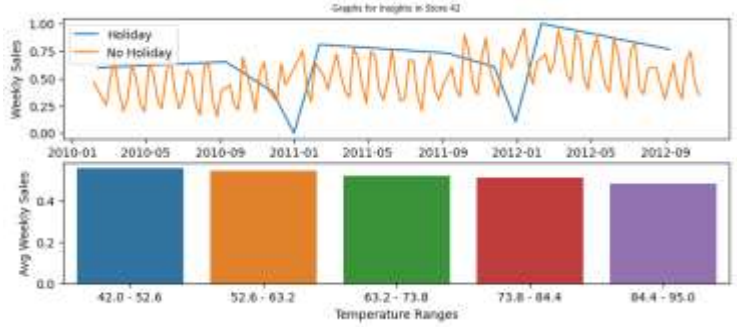
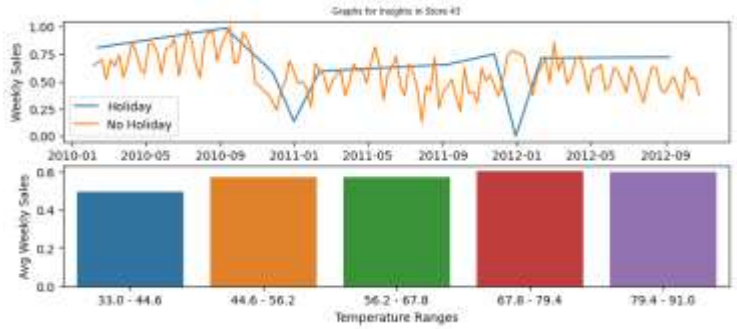
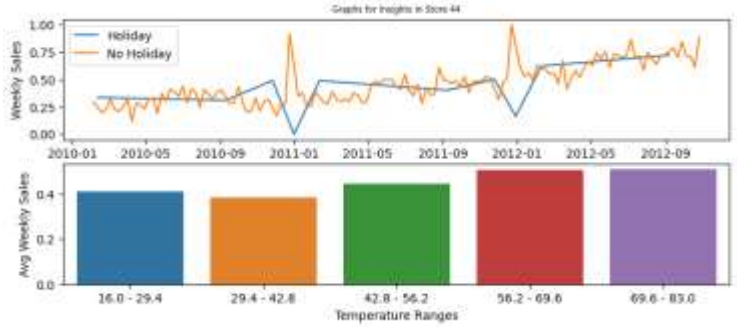
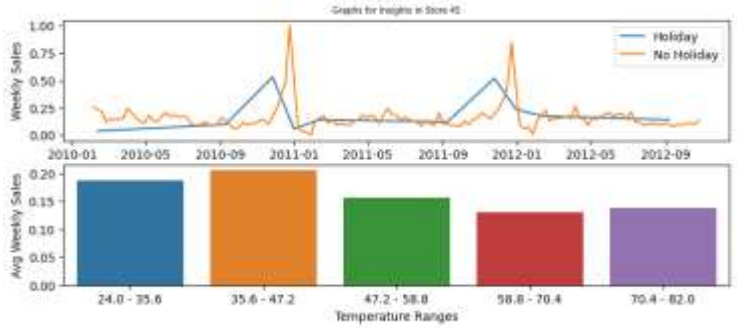
16		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 64.0-77.0-degree Temperature • Stock up inventory for above mentioned time periods
17		• Most Sales during Holidays. • Most Sales throughout the year • Increasing Sales during 6.0-19.6-degree and 46.8-74.0-degree Temperature • Stock up inventory for above mentioned time periods
18		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 27.0-53.0-degree Temperature • Stock up inventory for above mentioned time periods
19		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 31.8-43.6-degree Temperature • Stock up inventory for above mentioned time periods
20		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 20.0-31.8-degree Temperature • Stock up inventory for above mentioned time periods

21		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 46.0-58.0-degree Temperature • Stock up inventory for above mentioned time periods
22		• Most Sales during Non - Holidays. • Most Sales during end of year • Stock up inventory for above mentioned time periods
23		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 63.6-77.0-degree Temperature • Stock up inventory for above mentioned time periods
24		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 66.6-79.0-degree Temperature • Stock up inventory for above mentioned time periods
25		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Increasing Sales during 42.0-54.0-degree Temperature Stock up inventory for above mentioned time periods

26		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 57.8-71.0-degree Temperature • Stock up inventory for above mentioned time periods
27		• Most Sales during Non - Holidays. • Most Sales during end of year • Stock up inventory for above mentioned time periods
28		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Increasing Sales during 37.0-61.8-degree Temperature • Stock up inventory for above mentioned time periods
29		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 42.0-52.6-degree Temperature • Stock up inventory for above mentioned time periods
30		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Most Sales during end of year • Increasing Sales during 42.0-52.6-degree Temperature • Stock up inventory for above mentioned time periods

31		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Increasing Sales during 46.0-58.0-degree Temperature • Stock up inventory for above mentioned time periods
32		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 15.0-41.4-degree Temperature • Stock up inventory for above mentioned time periods
33		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Most Sales during end of year • Increasing Sales during 67.6-78.4-degree Temperature • Stock up inventory for above mentioned time periods
34		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 23.0-46.2-degree Temperature • Stock up inventory for above mentioned time periods
35		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Stock up inventory for above mentioned time periods

36		• Downtrend in sales observed irrespective of day. Do not Stock in excess only as per the demand. • Increasing Sales during 41.0-50.2-degree Temperature • Stock up inventory for above mentioned time periods
37		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Most Sales during start of year • Increasing Sales during 42.0-52.6-degree Temperature • Stock up inventory for above mentioned time periods
38		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Increasing Sales required frequent restocking • Stock up inventory for above mentioned time periods
39		• Most Sales during both Holidays and Non -Holidays. • Most Sales during end of year • Increasing Sales during 49.6-59.2-degree Temperature • Stock up inventory for above mentioned time periods
40		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 22.4-35.8-degree Temperature • Stock up inventory for above mentioned time periods

41		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 63.6-76.0-degree Temperature • Stock up inventory for above mentioned time periods
42		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Almost constant high sales observed, restock frequently to meet the demand during non-holidays
43		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less.
44		• Dip in Sales seen at end of every year during Holidays. Do not stock inventory in excess as demand is less. • Most Sales during end of year • Increasing Sales during 56.2-83.0-degree Temperature • Stock up inventory for above mentioned time periods
45		• Most Sales during Non - Holidays. • Most Sales during end of year • Increasing Sales during 35.6-47.2-degree Temperature • Stock up inventory for above mentioned time periods

Future Possibilities

Machine learning has shown significant promise in the field of sales prediction and forecasting. By analysing historical data, identifying patterns, and making predictions, machine learning algorithms can help businesses optimize their sales strategies, inventory management, and overall decision-making process.

It's important to note that successful implementation of machine learning for sales prediction requires quality data, appropriate algorithm selection, and continuous refinement of models. Additionally, domain expertise is crucial to interpret the predictions accurately and make informed business decisions.

As technology continues to advance, machine learning's role in sales prediction is likely to become even more sophisticated and integral to sales strategy development.

Conclusion

The project undertook a study of retail company with 45 outlets store with 2 years' worth of data for each store. Some important findings form the report include the following.

1. Sales revenue projection for the next 12 weeks down for most of the stores
2. Some stores have very weak or sales activities during some period of the year.
3. To improve the sales revenue, the following steps are recommended:
    a. Develop targeted marketing campaigns for each store's local audience, considering factors like local culture and preferences
    b. Concerted efforts by the company to find out through local market surveys and past sales record what products are in high demand by the local population at any given period of year and make efforts to replenish stocks.
    c. Offer competitive prices while maintaining profit margins. Implement dynamic pricing strategies to adjust prices based on demand and competition through commercial outreach, social media, can improve sales
    d. Identify slow-moving products and implement strategies to reduce excess inventory and minimize losses.
    e. Establish an online presence to cater to customers who prefer online shopping. Offer features like online ordering, in-store pickup, and seamless online-offline experiences.
4. Collect feedback from customers is key step to identify pain points and areas for improvement.
5. Each store only has 2-year data collecting even more data can help to study sales patter even better and plan effectively.
6. Some Stores have to be shutdown if the sales revenue does not improve.

References

1. https://github.com/krishnaik06/ARIMA-And-Seasonal-ARIMA/blob/master/Untitled.ipynb.
2. https://pandas.pydata.org/docs/reference/api/pandas.tseries.offsets.DateOffset.html
3. https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/
4. https://matplotlib.org/stable/gallery/subplots_axes_and_figures/subplots_demo.html
5. https://docs.oracle.com/en/cloud/saas/planning-budgeting-cloud/pfusu/insights_metrics_MAPE.html#GUID-C33B0F01-83E9-468B-B96C-413A12882334
6. https://www.youtube.com/watch?v=2XGSIlgUBDI
7. https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html
8. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp#:~:text=ARIMA%20is%20a%20method%20for,points%20influence%20future%20data%20points.