**Demand Forecasting in Retail Domain**

**Vellore Institute of Technology, Chennai campus**

***School of Computer Science and Engineering (SCOPE)***

***M.Tech CSE with Specialization in Big Data Analytics***

**ABSTRACT:**

Forecasting is a useful technique that can help to understand how historical data influences the future. This is done by looking at past data, defining the patterns, and producing short or long-term predictions. Forecasting is used in multiple areas within retail Such as in Supply chainwe use demand forecasting where we have enough inventory available in the store to meet future demand, prevent out of stock and Workforce planning similarly in Financewe have sales forecasting where we use for budget planning and goal setting. Here we use AR model for forecasting.

**TIME SERIES ANALYSIS:**

Any time series can be broken down into its individual components:

**Trend:** Increase or decrease in the series of data over longer a period .A newly launched product can be expected to have a positive slope as we can expect sales to keep increasing over time

**Seasonality:** Fluctuations in the pattern due to seasonal determinants over a period such as a day, week, month, season. For example --Ice creams sell well during summer seasons

**Cyclical variations:** Occurs when data exhibit rises and falls at irregular intervals. eCommerce example - we can expect low sales after events (big billion day) as a lot of people would have made purchases

**Random or irregular variations:** Instability due to random factors that do not repeat in the pattern. Competitor increased the price for a couple of weeks because of which our demand got increased

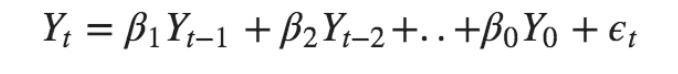
**SYSTEM REQUIREMENT:**

* Windows 10
* Collab from google for running python

**ALGORITHM:**

**ARMODEL**

* Autoregressive models operate under the premise that past values have an effect on current values, which makes the statistical technique popular for analyzing nature, economics, and other processes that vary over time
* We forecast the variable of interest using a linear combination of past values of the variable. The term auto regression indicates that it is a regression of the variable against itself. Thus, an autoregressive model of order p can be written as



**METHODOLOGY:**

Dataset description

Loading the data

Merging the data into single dataframe

Exploratory data analysis

Forecasting

Findings

**DATASET DESCRIPTION:**

**Sales data-set.csv:**

Anonymized information about the 45 stores, indicating the type and size of store.

**Stores data-set.csv:**

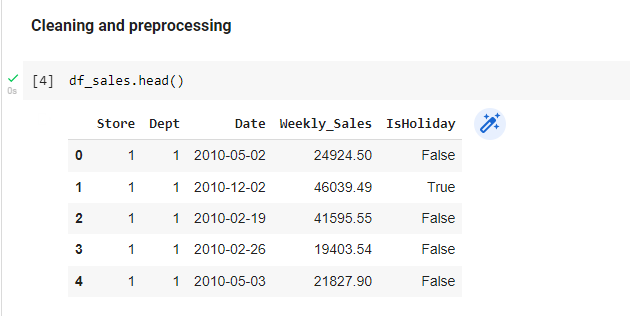
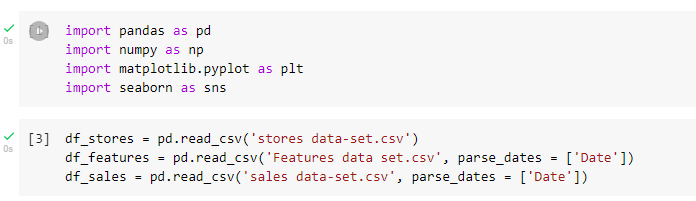
Historical sales data, which covers to 2010-02-05 to 2012-11-01.

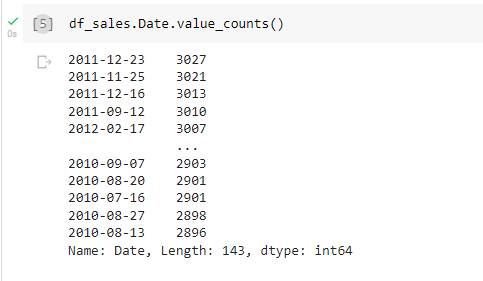
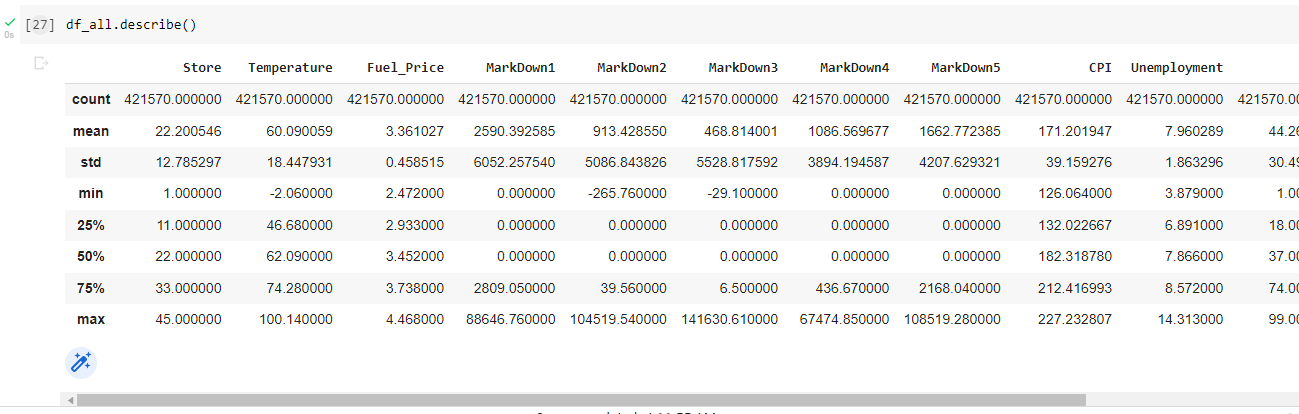
**Features data-set.csv:**

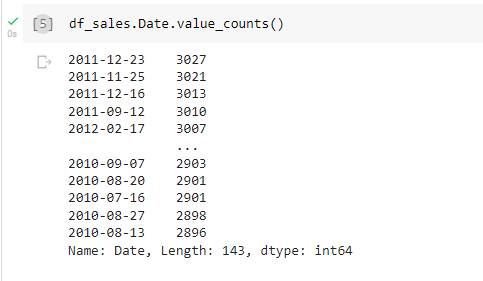
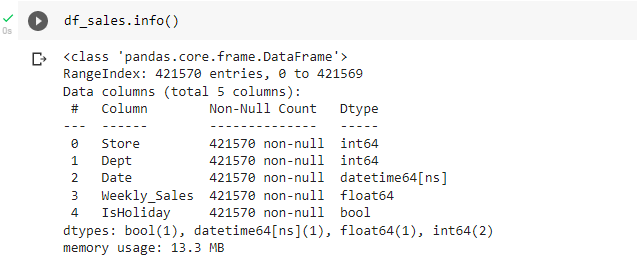
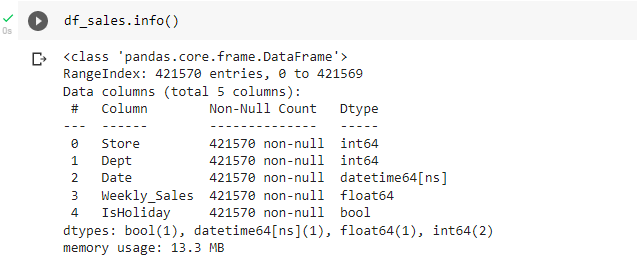
Contains additional data related to the store, department, and regional activity for the given dates.

|  |  |  |  |
| --- | --- | --- | --- |
| **Column name** | **Description** | **Unique value** | **Missing value** |
| Type | Type – type of the sales | Type- 45 | Type- 0 |
| Store | Store- the store number | Store – 45 | Store – 0 |
| Size | Size- size of the sales | Size – 45 | Size – 0 |
| Store | Store - the store number | Store - 143 | Store - 0 |
| Dept | Dept - the department number | Dept - 143 | Dept - 0 |
| Date | Date - the week | Date - 143 | Date - 0 |
| Weekly Sales | Weekly Sales - sales for the given department in the given store. | Weekly Sales – 143 | Weekly Sales- 0 |
| IsHoliday | IsHoliday - whether the week is a special holiday week. | IsHoliday - 143 | IsHoliday -0 |
| Store | Store - the store number | Store - 182 | Store - 0 |
| Date | Date - the week | Date – 182 | Date - 0 |
| Temperature | Temperature - average temperature in the region | Temperature - 182 | Temperature - 0 |
| Fuel\_Price | Fuel\_Price - cost of fuel in the region | Fuel\_Price - 182 | Fuel\_Price - 0 |
| MarkDown1-5 | MarkDown1-5 - anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA | MarkDown1-5 – 182 | MarkDown1 - 51%  Markdown 2 -64%  Markdown 3- 56%  Markdown 4- 58%  Markdown 5- 51% |
| CPI | CPI - the consumer price index | CPI - 182 | CPI - 7% |
| Unemployment | Unemployment - the unemployment rate | Unemployment-182 | Unemployment-7% |
| IsHoliday | IsHoliday - whether the week is a special holiday week. | IsHoliday - 182 | IsHoliday - 0 |

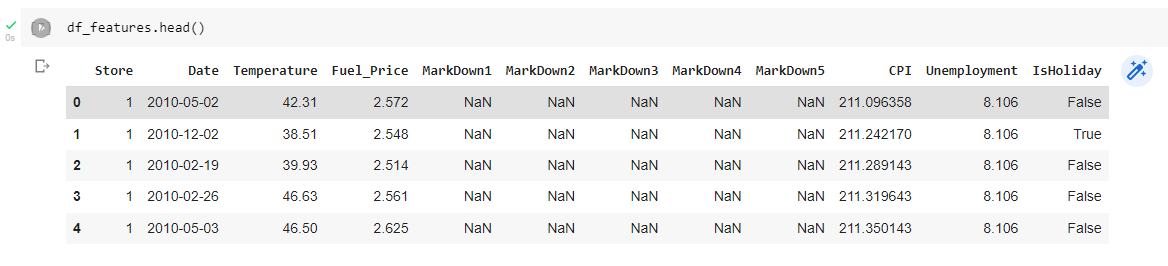
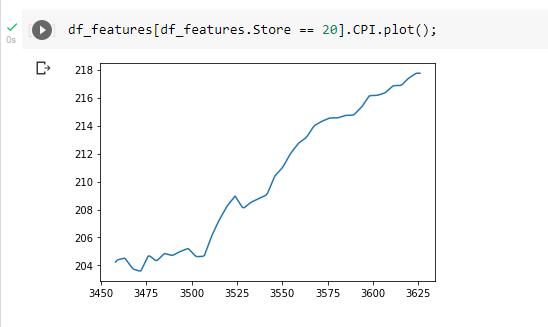
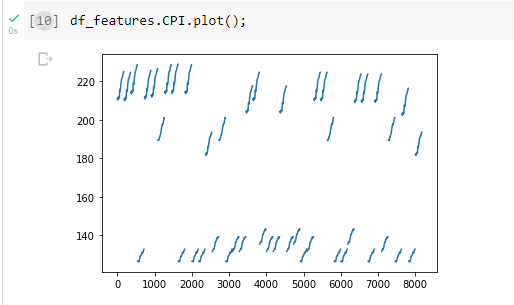
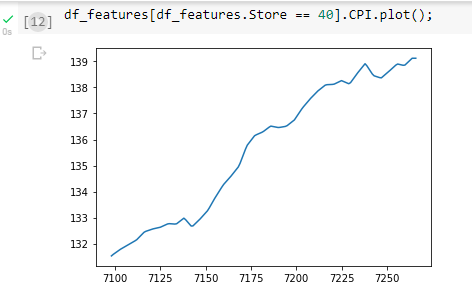
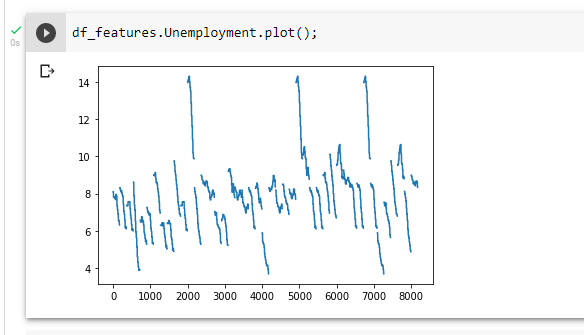
**Exploring the data set**

****

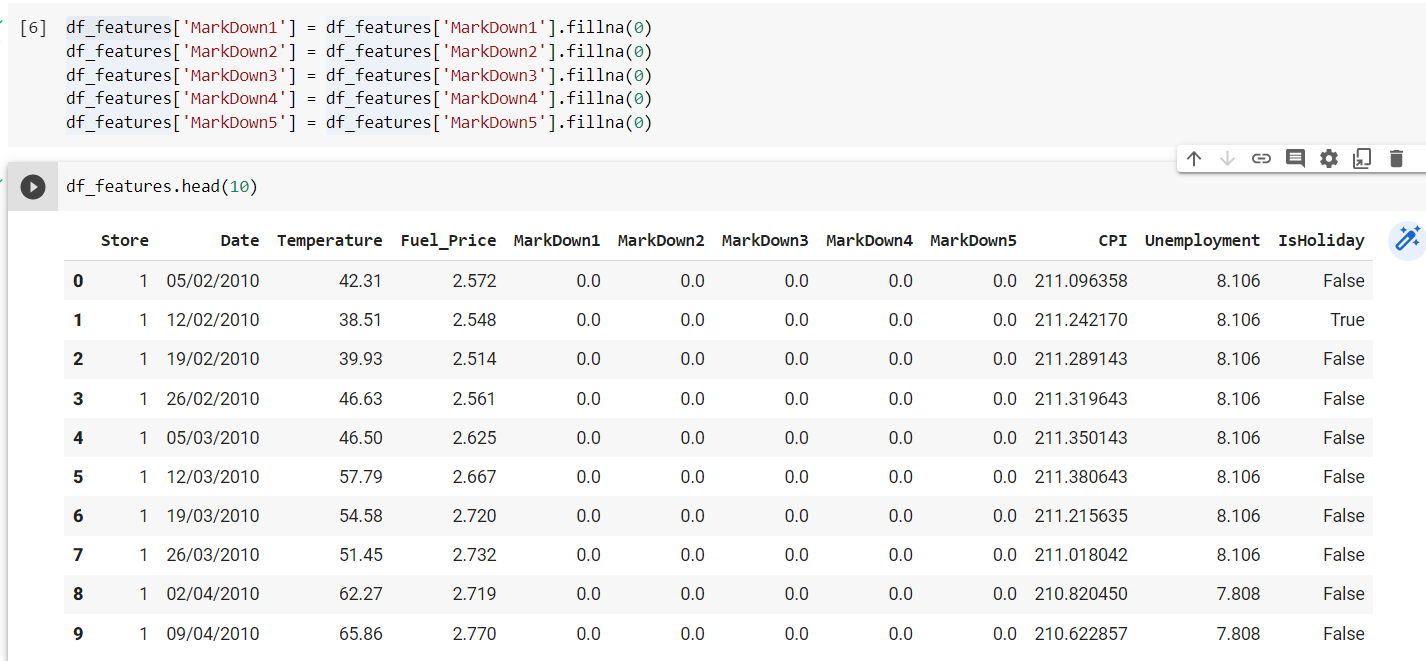
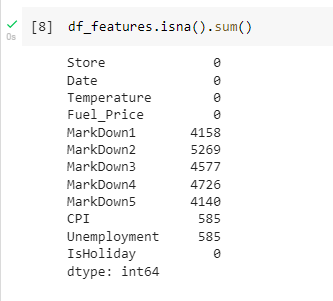
****

****

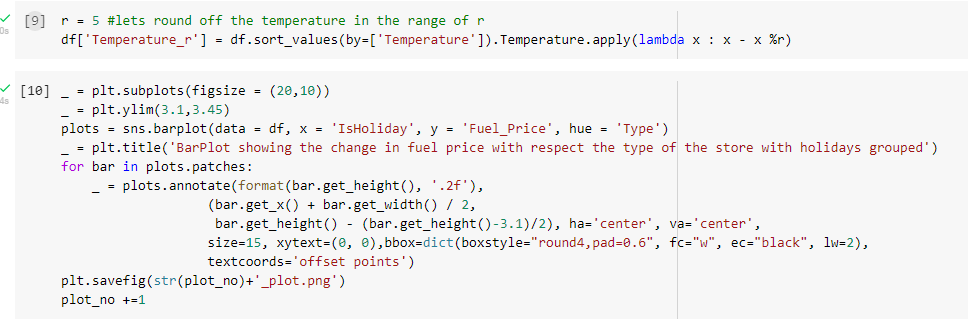
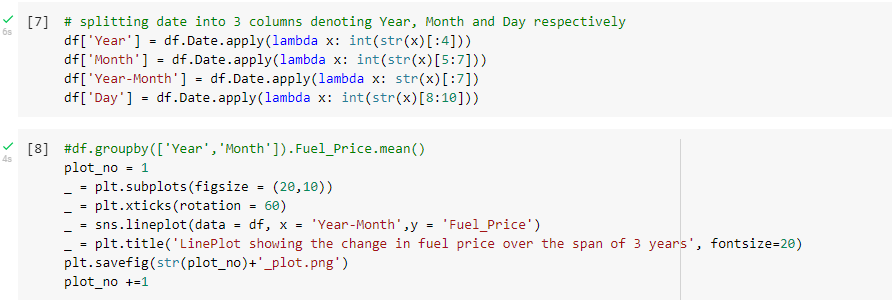
**Visual Insights**

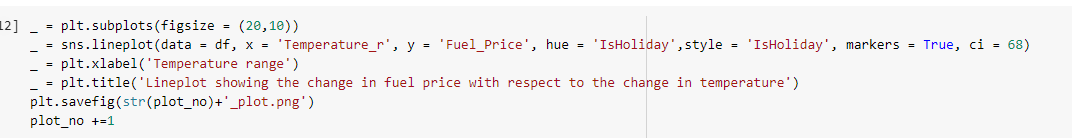
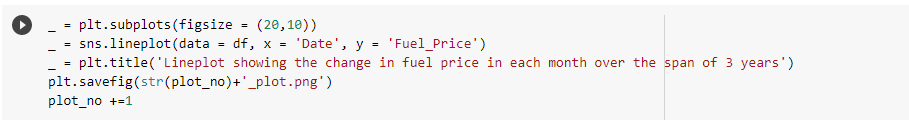
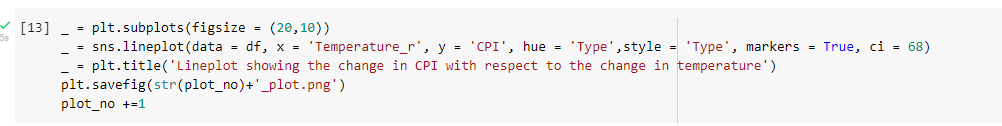
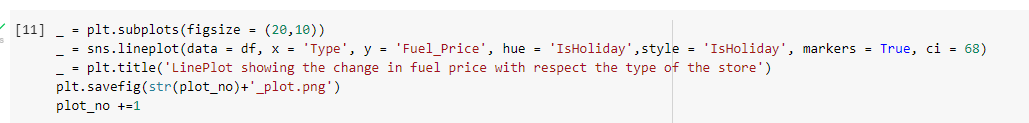
****

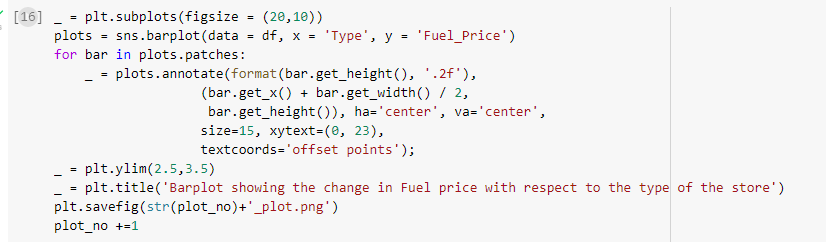
**Missing value Imputation**

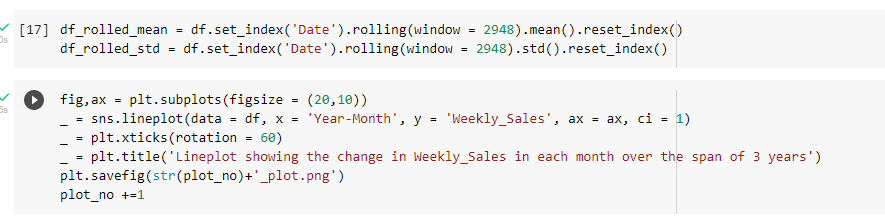
****

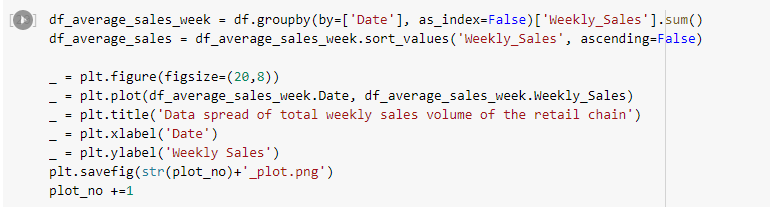
**SAMPLE CODE:**



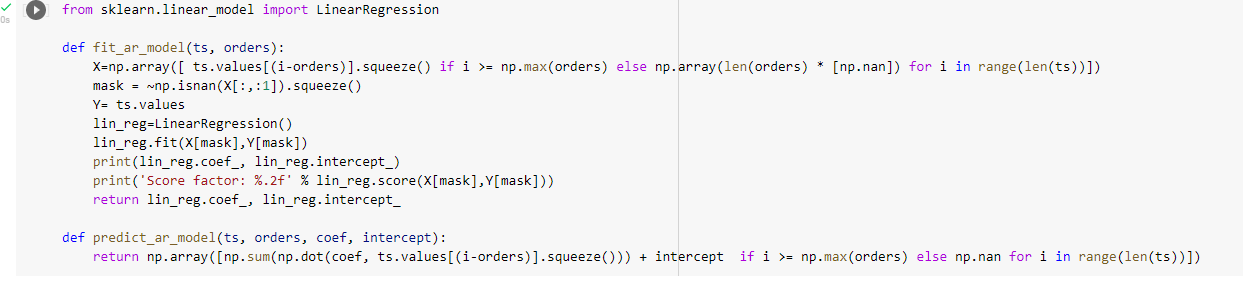
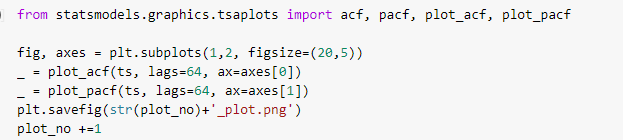


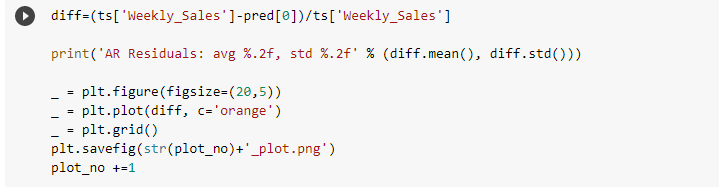
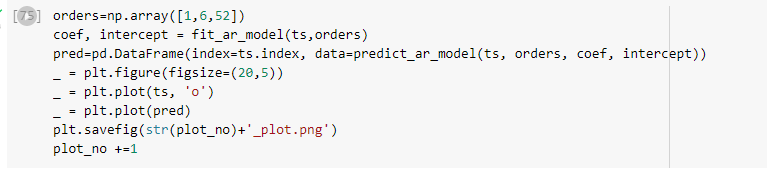


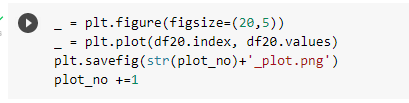
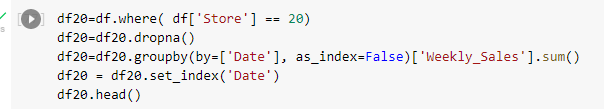




**Forecasting:**

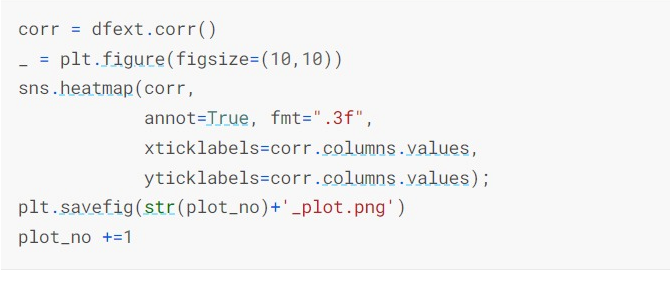




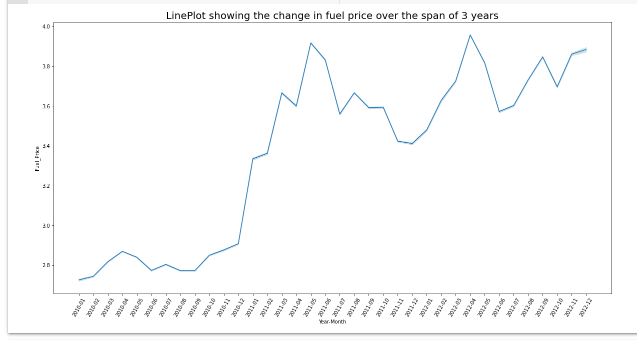
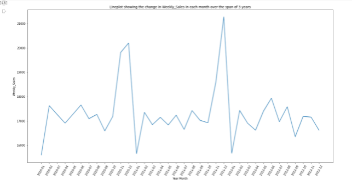
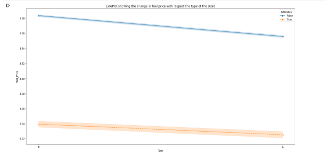
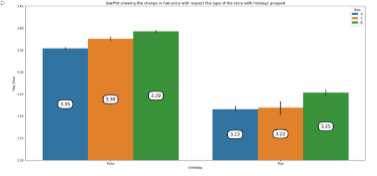


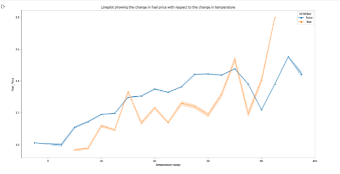
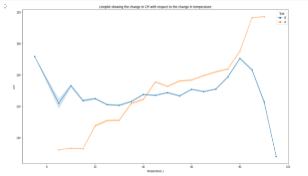
**GETTING THE TEMPERATURE FUEL PRICE AND CPI FOR THE STORE 20**

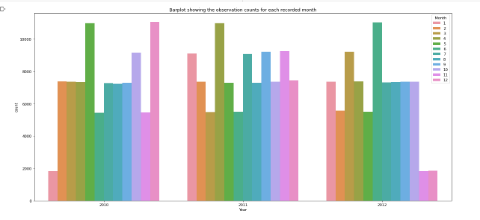
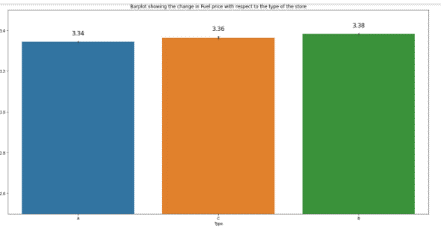
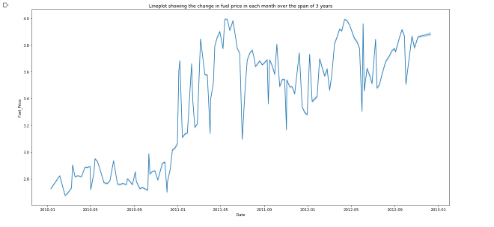
****

**CORRELATION **

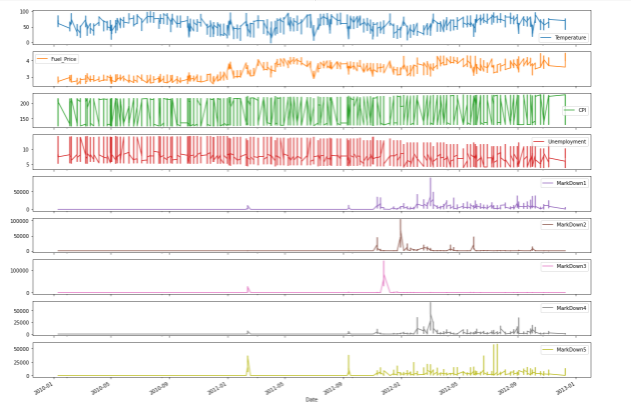
**OUTPUT GRAPH:**

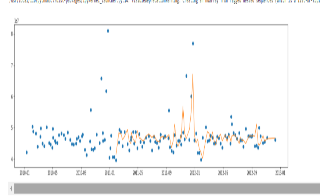
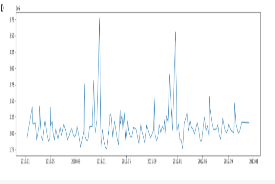


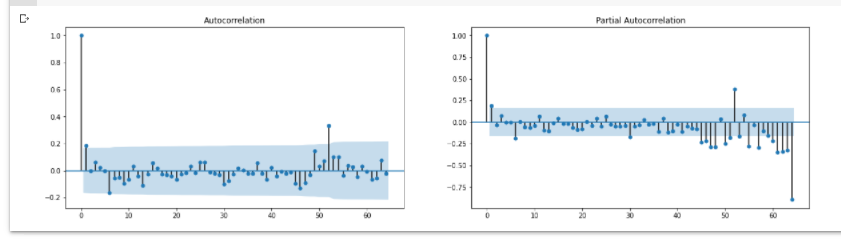
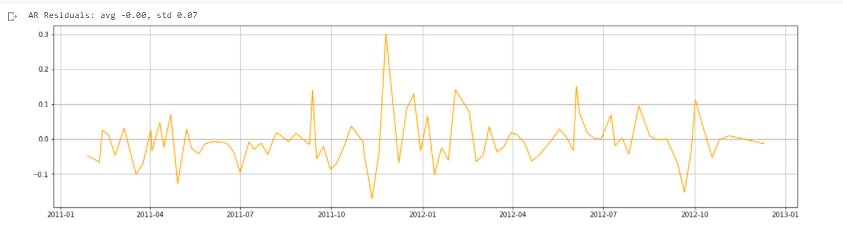




**GAINING INSIGHTS OFMARKDOWN1-5,UNEMPLOYMENT,CPI,FUELPRICE,TEMPERATURE OVER THE PERIOD OF TIME**

****

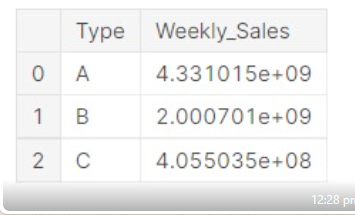
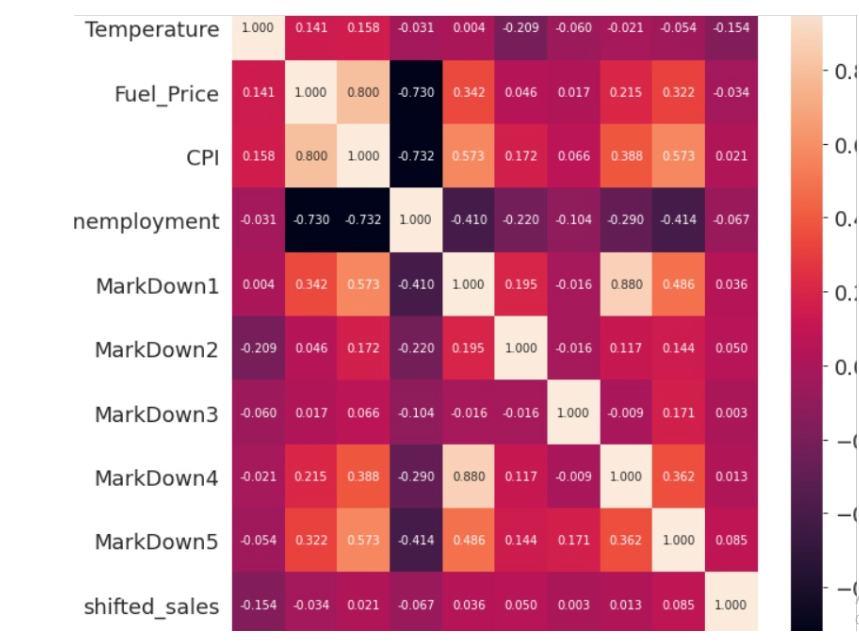
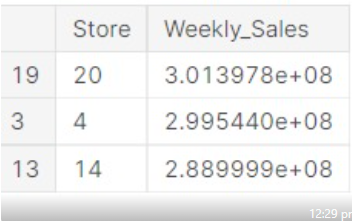
**Forecasting- Graphs:**

**GETTING THE TEMPERATURE FUEL PRICE AND CPI FOR THE STORE 20**

****

**CORRELATION TOP PERFORMING STORE IN TERMS OF SALES**

****

**FINDINGS:**

### The fuel price increases with increase in temperature steadily during workdays and unevenly during holidays

### There is no significant pattern in the data points spread each month in the dataset. However, one noticeable cue is that no sales data is recorded / happened during the month of September in 2013.

### There was a peak during the end of the years 2010 and 2011 but not during 2012. This might be due to comparatively very less observations during the last 2 months in 2012.

**INFERENCE:**

From the basic analytics and forecasting we can suggest the shop owner that they can stock up the high selling products of various season in the low temperature period itself as fuel price tends to increase as the temperature increases which in turn will reflect in the stock prices. Similarly Store number 19 has the highest weekly sales from which all the other stores can take up same trend in order to increase the sales.