A Project Report on

Indian Agriculture Commodity Price Forecasting & Anomaly Discovery,

A Case Study on "Onion, Potato and Tomato" commodities

Submitted in partial fulfilment for award of degree of

Master of Business Administration

In Business Analytics

Submitted by

Mutturaj UPPALADINNI

R19MBA04

Under the Guidance of

Krishna Kumar Tiwari

Data Architect

REVA Academy for Corporate Excellence

REVA University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka,

Bangalore – 560064

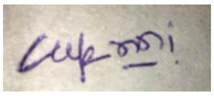
October 2020

Candidate's Declaration

I, Mutturaj UPPALADINNI hereby declare that I have completed the project work towards the Master of Business Administration in Business Analytics at, REVA University on the topic entitled **'Indian Agriculture Commodity Price Forecasting & Anomaly Discovery, A Case Study on "Onion, Potato and Tomato" commodities"** under the supervision of Krishna Kumar Tiwari, Data Architect. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2020.

Place: Bengaluru

Mutturaj UPPALADINNI



Date: 28th Oct 2020

Certificate

This is to Certify that the PROJECT work entitled **“Indian Agriculture Commodity Price Forecasting & Anomaly Discovery, A Case Study on “Onion, Potato and Tomato” commodities”** carried out by Mutturaj UPPALADINNI with SRN R19MBA04, is a bonafide student of REVA University, is submitting the project report in fulfilment for the award of Master of Business Administration in Business Analytics during the academic year 2019. The Project report has been tested for plagiarism and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of PROJECT work prescribed for the said Degree.

Krishna Kumar Tiwari
Guide

Dr. Shinu Abhi
Director, Corporate Training

External Viva

Names of the Examiners

1. <Name> <Designation> <Signature>
2. <Name> <Designation> <Signature>

Place: Bengaluru

Date: 28th Oct 2020

REVA UNIVERSITY
Bengaluru, India

Acknowledgement

I am grateful to **Dr. Shinu Abhi,** Director, Corporate Training, for their assistance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the capstone project.

I would like to thank my project guide **Mr. Krishna Kumar Tiwari** for his valuable guidance to complete the capstone project

It is my gratefulness towards **Dr. Jay Bharateesh Simha,** Chief Mentor and all other mentors for the valuable guidance and suggestions in learning various data science aspects and for the support.

I am thankful for my classmates for their aspiring guidance, invaluably constructive feedback, and friendly advice during the project work; also, I would like to thank Dishani Haldar for her support to complete the project

I would like to acknowledge the support provided by Hon'ble Chancellor, Dr. P Shayma Raju, Vice Chancellor, Dr. K. Mallikharjuna Babu, and Registrar, Dr. M. Dhanamjaya.

It is sincere thanks to all members of program office of RACE who were always supportive in all requirements from the program office.

It is my sincere gratitude towards my parents, and my family for their kind co-operation and inspiration which helped me in completion of this project.

Place: Bengaluru
Date: 28th Oct 2020

Similarity Index Report

This is to certify that this project report titled **"Indian Agriculture Commodity Price Forecasting & Anomaly Discovery, A Case Study on "Onion, Potato and Tomato" commodities"** was scanned for similarity detection. Process and outcome are given below.

Software Used: Turnitin

Date of Report Generation:

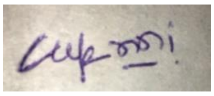
Similarity Index in %: 12

Total word count: 5981

Name of the Guide: **Mr. Krishna Kumar Tiwari**

Place: Bengaluru

Mutturaj UPPALADINNI



Date: 28th Oct 2020

Verified by:

Signature

Dr. Shinu Abhi,

Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	“LSTM “	Long short-term Memory
2	“ARIMA”	Auto-Regressive Integrated Moving Averages.
3	“SARIMA”	Seasonal Autoregressive Integrated Moving Average
4	“CNN”	Convolution neural network

List of Figures

List of Tables

Abstract

Food price fluctuations may affect buyers and producers, they also compounded by trade groups in developed commercial system, wherever marketplaces will not be running in optimal situations, competition for different theoretical grounds.

And the Indian commodity trade and the assessment of a platform for market forecasting and identification of anomalies classification scheme for identifying hoarding cases commercial stock. our dataset consists of Time series Wholesale rates, and farm delivery volumes commodities and shopping at many village marketplaces commodity rates in the town centers. We also make provision a detailed qualitative study of such effects event series like storing, climate disruptions, and outside surprises. The results appear optimistic which tend to the potential of building a model for agricultural pricing.

Agricultural products which could be used for the inconsistency of information reduction and the detection of abnormalities, which could help to control agriculture sector would need to operate more consistently.

Key words: LSTM, RNN, CNN, ARIMA, SARIMA, Information retrieval, Data analytics, Agriculture, Anomaly, Seasonality, weather

Contents

Chapter 1: Introduction

“A significant consideration is price volatility in agricultural commodities, economics, and growth area of research. strong value for money, growing retail consumer spending despite, low prices reduce farmer product incomes. Rainfall in India is an essential source of price volatility since most of it is weather conditions, instead dependant on agricultural production strong rinsing systems. weak, or else extreme, or inconsistent could kill yields, and is particularly harmful toward smallholder cultivators whose cash flows were seriously strained small buffer for handling such disruptions [5]. Various goods also affected for exports such as cotton and oilseeds economic dynamics like speculative ones”.(Madaan et al., 2019)

“MSP (Minimum Support Price) to purchase these products directly from state-owned enterprises to give farmers a reasonable amount of pay, or else thru giving farmers debt waivers .similarly , government responses towards high-price incidents has to be limit exports through enforcing “a high-enough minimum export price (MEP)” This allows suppliers to sell at local and lower consumer inflation, and to import and sell the goods at a discounted rate. Such steps often appear to be inefficient and sensitive and have their own collection of disadvantages. during resolving market volatility in a structural manner. Market oriented methods such as the sale of goods”.(Madaan et al., 2019)

thus, the domestic price fluctuations continue to be regular, and to be accentuated by acts local traders who encounter weather disturbances or other events create opportunities to make a profit. The agricultural sector India's marketplace comprises a wide network of more than 7500 Local marketplaces governed by the Government (called mandis) where farmers sell traders their own products. These stores bring goods to other states or town canters, and market them to shopkeepers. However, the truth of such supply chains is quite demanding. Second, smallholder farmers often find themselves unable to do so take their goods to the mandis themselves because of transport costs are more than 60% for most goods these short ones however, prices are not beneficial to the smaller farmers because further hoarding occurs in the supply chain, and It really harms certain low-income consumers who really need to purchase higher retail prices. My third argument is that rural wealth is strongly complex where farmers often raise their debts to buy farm inputs from the same merchant class to those who distribute their commodity and thus control limited negotiating power., so that local rates can be higher merchants. These are mainly poor farmers, therefore, agricultural production is not just at its highest risk but also gain smaller share of benefit, and inequality can to persevere in yielding farmers themselves, we are developing a model of hoarding detection for strengthen regulatory frameworks for agricultural operations in-country markets. We function with both programmes look at the resources of Onions, Potato and Tomato to plant, these are important nationally grown crops domestic consumption, and potentially even simpler crops analysis. These are not protected by “MSPs” offered through the state, and all come up with extensive shelf life; potatoes and Tomatoes needs cold storage in short span, but Onions can longer hoarded in interim protections which require a medium dry well as cold. therefore, the costs are predicted can only have being influenced by rain, production area along with efficiency in manufacture, handling with regional merchants, and additional occasional activities. Use monthly “retail and mandi-price data, and arrival amounts at mandis”, gathered to longer duration we experiment with various market forecasts for more than a decade models can achieve good performance in the end using a regression model of several variables. then we get an input of news report which contain hoarding evidence related events: we train classifiers to see this as a positive collection to point storing applying price data from time series and the comebacks. We do the two-stage classification: when the incident happened towards whether, we can define properly any additional details asymmetries to a proper functioning of these economies. Our approach proposed will help both the farmers and the supermarket clients.

“Next, we define work related to this field, an initiation to Onion, Potato and Tomato growing context and promoting in India. an overview of our methods of predicting along with grouping and concluding together with discussing promising future work in that area”.(Madaan et al., 2019)

Chapter 2: Literature Review

[1] India's Long Journey to Agricultural Market Transformation: Lessons from With Karnataka. Tech Research. Global Leading centre

This paper explores Karnataka 's ground-breaking agricultural production marketing reforms with the twin goals of evaluating the state and implementation challenges and drawing lessons from the experience of Karnataka for India's e-National Agricultural Market (e-NAM). From the ground study of ten APMC's across the India, we found that Delhi, Mumbai and Bangalore regularly forced through changes, the challenges of deeper reforms are important within the sense of deeply entrenched ties between farmers , traders and commission agents. "(Nidhi Aggarwal, Sargam Jain, and Sudha Narayanan. 2016)"

[2]" Why MSP at 1.5 Times Cost Is Another Empty Promise for Farmers. Retrieved from The Wire https://goo.gl/1pcGtR. "80% of Farmers' Crop Insurance Claims Unpaid Past Payment Deadline By Kabir Agarwal thewire.in — New Delhi: According to information collected by The Wire through the Right to Information Act (RTI), farmers were paid just 20 percent of their registered crop insurance claims for the 2019-20 Rabi season a month and a half after they were due. The total claims reported under the two core crop insurance schemes-Pradhan Mantri Fasal Bima Yojana (PMFBY) and RWBCIS (Restructured Weather Crop Insurance Scheme)- amount to Rs 3,750 (Kabir Agarwal, 2018)

[3] "The seeds of discontent. Retrieved from The Indian Express https://goo.gl/6wxeF1."(Madaan et al., 2019)

Between the images of demonstrating growers in Madhya Pradesh, it was noticeable to see how many youthful people were wore in jeans and shirts — they were clearly unhappy with jobless growth, not all farmers, but also farmers' sons. Today we are facing not only a crisis of farmers but also a crisis of farming families whose children want non-farm jobs. (Bina Agarwal, 2017)

[4] In this paper we estimate the effect of India's demonetization exercise on domestic agricultural exchange commodities that invalidated 86 per cent of the circulating currency. Using data on arrivals and costs for 35 major agricultural commodities from nearly 3000 controlled markets in India for the sum 2011-2017, we look for short-term effects up to three months after demonetization, monitoring both impact and recovery. These 35 commodities constitute an impressive proportion of land under cultivation and production value and are thus additionally indicative of Indian agriculture in one sense. Using earlier years as years of reference we use a mixture of differences techniques

We note that demonetization has disrupted domestic agricultural trade driven markets by more than 15% in the short run, settling at 7% after recovery at the top of the 90-day cycle following demonetization. Exchange perishables were displaced within a week of demonetization to an extent of 23 per cent. It marginally recovered by the 90-day turn but was still 18 per cent lower than the same old. Much of this downturn is attributed to the recent price declines rather than deliveries, which seem to have stabilised over a three-month period. There are major variations between commodities, but most of these variations are expected.(Aggarwal & Narayanan, 2017)

[5] "building the evidence base with farming families in Mozambique, Tanzania,
and Pakistan. Consultative Group to Assist the Poor (CGAP) (2016)".

[6]" Reuters Market Light, Creating Efficient Markets. Retrieved from UNDP Growing
Inclusive Markets http://growinginclusivemarkets.org/media/cases/
India_RML_2010.pdf"."(Anderson & Ahmed, 2016)"(Madaan et al., 2019)

[7] "Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry."
"International Journal of Operations Research and Information Systems 7
(04 2016), 1–21".(Arunraj et al., 2016)

[8] "Unchecked manipulations, price–volume relationship and market
efficiency: Evidence from emerging markets. Research in International
Business and Finance 30 (2014), 51–71".(Azad et al., 2014)

[9] "Farmer suicides in India and the weather god.
Procedia Computer Science 122 (2017), 10–16".(Banik, 2017)

[10] Market Abuse Detection: A Methodology Based on Financial Time Series. Statistica Applicata 18, 4 (2006).(Barucci, 2006)

[11] Predicting Socio- Economic Indicators using News Events. In Proceedings of the 22nd
ACM SIGKDD International Conference on Knowledge Discovery and
Data Mining. ACM, 1455–1464.(Chakraborty et al., 2016)

[12] Ramesh Chand. 2012. Development policies and agricultural markets.
Economic and Political Weekly (2012), 53–63.(Chand, 2016)

[13] LONG SHORT-TERM MEMORY.(Hochreiter & Schmidhuber, 1997)

[14] Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System. (Jha & Sinha, 2013)

[15] Short-Term Price Forecasting For Agro-products Using Artificial Neural Networks.(Li et al., 2010)

Chapter 3: Problem Statement

“Food price fluctuations may affect buyers and producers, they also compounded by trade groups in developed commercial system, wherever marketplaces will not be running in optimal situations, competition for different theoretical grounds”. (Madaan et al., 2019)

“We focus on Onion, Potato and Tomato market in India and demonstrate the assessment of a market forecast model and the discovery and understanding of a deviation to spot stock hoarding incidents by traders”.(Madaan et al., 2019)

“We are trying to address two problems in this diverse environment aimed at empowering small-scale farmers and low-income buyers”.(Madaan et al., 2019)

First, we create a market forecast model that can predict monthly prices ~ 3 months into the future and can help farmers make a better decision about when to sell their products to help farmers get a competitive price for the product.

“Second, we are developing a hoarding recognition model to support low-income consumers who are badly impacted by high prices and who also happen to be small - scale farmers themselves, in order to improve regulatory frameworks in the domestic economic market operations.”(Madaan et al., 2019)

Our dataset consists of time series of wholesale prices and volumes of arrival of agricultural commodities at several markets at the regional level and retail prices of commodities at the facilities of the city.

Chapter 4: Objectives of the Study

We are trying to address two problems in this diverse environment aimed at empowering small-scale farmers and low-income buyers.

“First, we create a market forecast model that can predict monthly prices ~ 3 months into the future and can help farmers make a better decision about when to sell their products to help farmers get a competitive price for the product.”(Madaan et al., 2019)

“Second, we are developing a hoarding recognition model to support low-income consumers who are badly impacted by high prices and who also happen to be small - scale farmers themselves, in order to improve regulatory frameworks in the domestic economic market operations.

Our dataset consists of time series of wholesale prices and volumes of arrival of agricultural commodities at several markets at the regional level and retail prices of commodities at the facilities of the city.”(Madaan et al., 2019)

We are identifying the influence factors for price volatility, like is the weather, season, inflation are contributing anything in price volatility.
We are using the CRISP-DM approach to address this problem and we are using the traditional forecasting models ARIMA, SARIMA and comparing the models with machine learning LSTM models for price forecasting.

Chapter 5: Project Methodology

“We have applied the CRISP-DM approach in our project, CRISP-DM is the powerful methodology, and it is helpful to solve the problems in organized way”.

CRISP-DM is consisting of six steps:

- Business understanding,
- Data understanding,
- Data preparation,
- Modeling,
- Evaluation
- Deployment

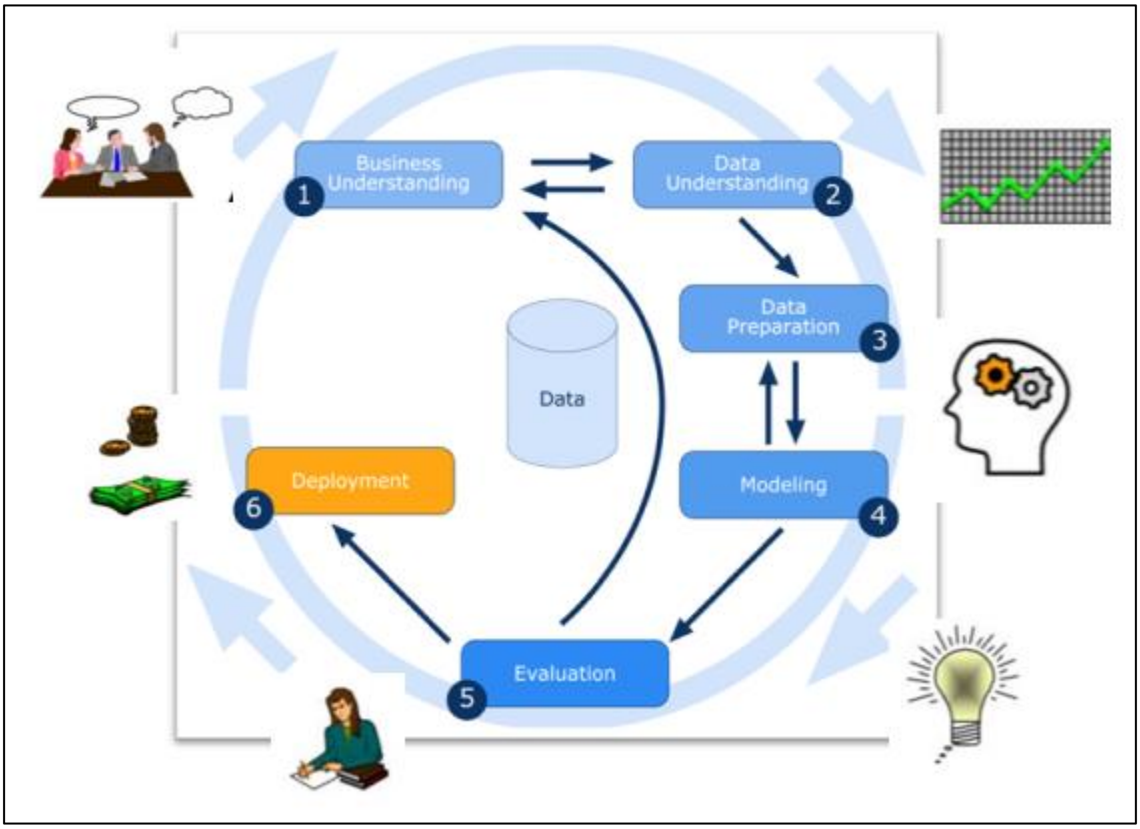


Figure No. 5.1: CRISP-DM methodology

Different forecasting Models used our study
1 ARIMA
2 SARIMA
3 LSTM

1. ARIMA Model

“As explained in otexts ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.”(Haines et al et al., 2019)

2. Seasonal ARIMA (SARIMA)

“As explained in otexts So far, we have restricted our attention to non-seasonal data and non-seasonal ARIMA models. However, ARIMA models are also capable of modelling a wide range of seasonal data”.(Haines et al et al., 2019)

“A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as follows”

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

where
m = number of observations per year. We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

“The seasonal part of the model consists of terms that are like the non-seasonal components of the model but involve backshifts of the seasonal period. For example, an ARIMA (1,1,1) (1,1,1) 4 model (without a constant) is for quarterly data (m=4), and can be written as

(1−ϕ1B) (1−Φ1B4) (1−B) (1−B4) yt=(1+θ1B) (1+Θ1B4)εt.

The additional seasonal terms are simply multiplied by the non-seasonal terms”.(Haines et al et al., 2019)

3. LSTM

“As per Graves book The LSTM architecture consists of a set of recurrently connected subnets, known as memory blocks. These blocks can be thought of as a differentiable version of the memory chips in a digital computer. Each block contains one or more self-connected memory cells and three multiplicative units—the input, output and forget gates—that provide continuous analogues of write, read, and reset operations for the cells”.(Graves, n.d.)

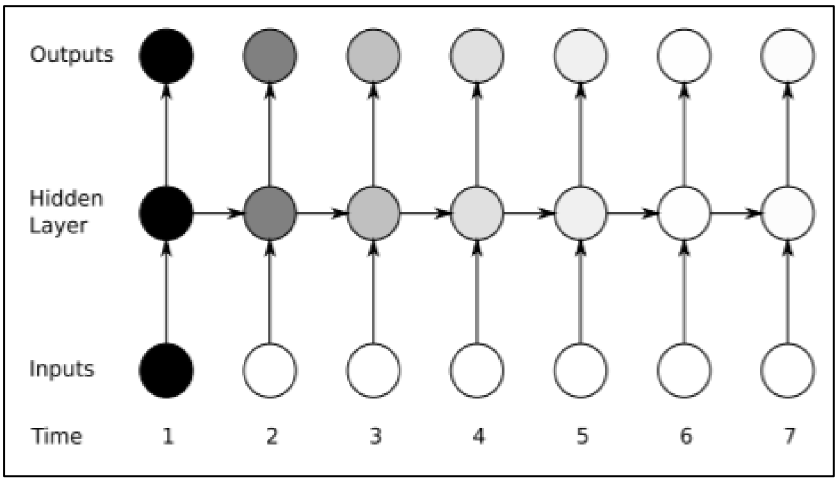


Figure No.5.2: The LSTM architecture

Chapter 6: Business Understanding

“Food price fluctuations may affect buyers and producers, they also compounded by trade groups in developed commercial system, wherever marketplaces will not be running in optimal situations, competition for different theoretical grounds”. (Madaan et al., 2019)

“We focus on Onion, Potato and Tomato market in India and demonstrate the assessment of a market forecast model and the discovery and understanding of a deviation to spot stock hoarding incidents by traders”.(Madaan et al., 2019)

For building a forecast model we are considering the historical data from National Horticultural board for Onion, Potato and Tomato commodities and Weather atlas website will be used for getting the real time weather forecasting data for specific region.

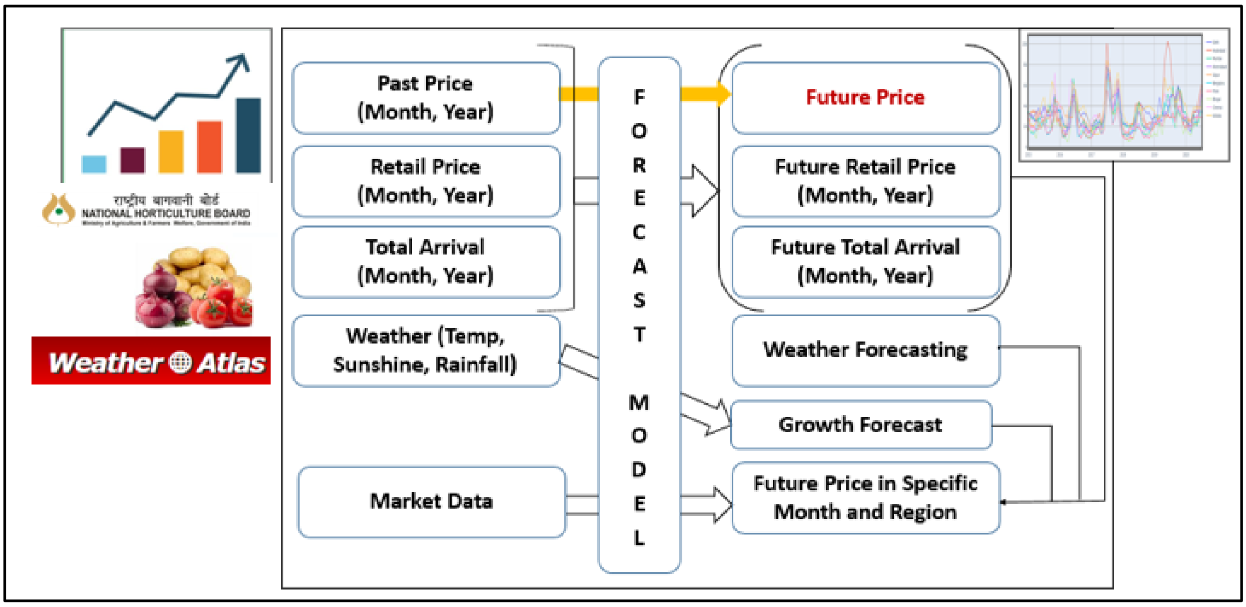


Figure No.6.1: Process Flow chart

Chapter 7: Data Understanding

1. **NHB** - This is the website of National Horticultural Research & Development Foundation and maintains a database on Market Arrivals and Price, Area and Production and Export Data for three commodities - Onion, Potato and Tomatoes. We are in luck! It also has data from 2000 onwards and has only got one form to fill to get the data in a tabular form. Further it also has production and export data. Excellent. Let us use this. Here is the best link to got to get all that is available(National horticultural board, n.d.)

So let us fill the form to get data and test our scraping process.(National horticultural board, n.d.)

- Crop Name: Onion, Potato and Tomato
- Month: Jan
- Market: All
- Year: Jan 2015 to July 2020

“The Agmarknet (Agricultural Marketing Information Network) website maintained by Indian government and accessible to all public, the website includes the major APMC commodity data in terms of wholesale price, retail price and volume arrivals by daily, weekly and monthly bases”.

“We have collected the data for Onion, Potato and Tomato from major APMC’s for almost 5.5years duration data (Jan-2015-July2020), the collected data were having missing values for many APMC’s , we restricted our study for Top10 APMC’s, which we had complete data available and no missing values for continues 2 months”.

We manually sorted all news reports about something to try and do with Onion, Potato and Tomato prices. This helped us to establish reality about the events

2. **Weather Data-** We have collected the weather data from(Atlas, n.d.) for all Top 10 Centers The Weather data-ind.com maintains the global weather data by days/Months and years, we have done the manual scraping and added one extra column to our data set.

We want to first **visually explore** the data to see if we can confirm some of our initial hypotheses as well as make new hypothesis about the problem we are trying to solve.

For this we will start by loading the data and understanding the data structure of the data frame we have.

Data Structure: So, we have eight columns in our dataset. Let us understand what each one is. Two are about the location of the Wholesale Market where Onion, Potato and Tomato where sold.

Center Name: This is the city in India (AHMEDABAD, BANGALORE and so on)

Commodities: This is the commodity name (Onion, Potato and Tomato)

Two are related to the

month: Month in January, February and so on.

year: Year in YYYY representation

Four are about quantity and price in these wholesale market.
Total Arrival: The quantity of Onion arriving in the market in that month in quintals (100 kg)
W. sale Avg.Price: The Avg price in the month in Rs. /quintal
Retail Avg. Price: The Avg price in the month in Rs. /quintal
Weather data: centre wise Avg. Temperature (°C)
We would expect the following the columns to be of the following type
CATEGORICAL: state, city, market
TIME INTERVAL: month, year, date
QUANTITATIVE: quantity, W. sale Avg. Price, Retail Avg price, Weather data

Table No.7.1: RAW data for Onion, Potato& Tomato

	B	C	D	E	F
1	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice	AHMEDABAD_Volume	AHMEDABAD_Avg. Temperature	Date
2	1359	2317	13405	20.1	1/1/2015
3	1569	2300	13009	22.8	2/1/2015
4	1389	2285	11983	27.1	3/1/2015
5	1099	2300	14303	31.2	4/1/2015
6	1140	2300	13349	33.4	5/1/2015
7	1582	2486	11567	32.6	6/1/2015
8	1957	3375	11451	29.4	7/1/2015
9	3838	5571	7103	28.1	8/1/2015
10	4180	6762	7425	28.6	9/1/2015
11	2757	5958	13974	28.4	10/1/2015

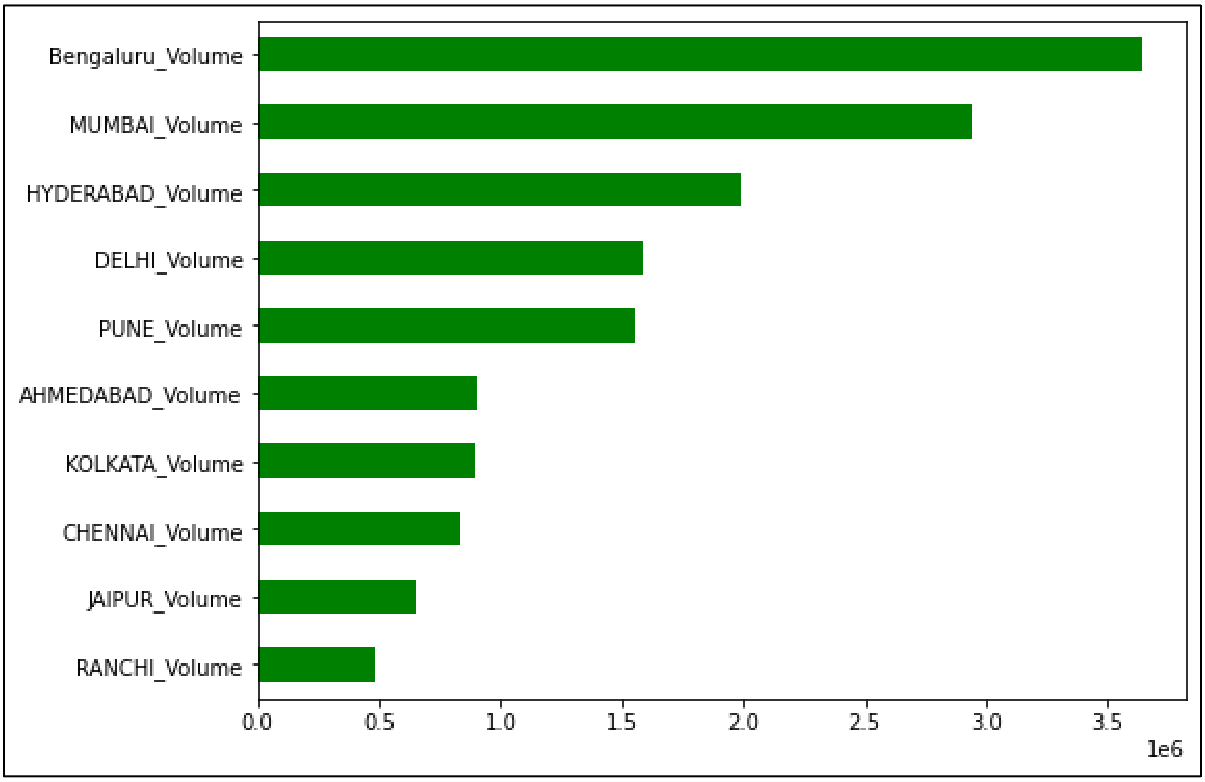
	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice	AHMEDABAD_Volume	AHMEDABAD_Avg. Temperature	Date
2	815.00	2450.00	22231.00	20.1	2015-01-01 00:00:00
3	576.00	2000.00	17116.00	22.8	2015-02-01 00:00:00
4	348.00	1550.00	22966.00	27.1	2015-03-01 00:00:00
5	318.00	1473.00	7729.00	31.2	2015-04-01 00:00:00
6	381.00	1463.00	8255.00	33.4	2015-05-01 00:00:00
7	513.00	1500.00	5256.00	32.6	2015-06-01 00:00:00
8	419.00	1500.00	9134.00	29.4	2015-07-01 00:00:00
9	434.00	1895.00	10844.00	28.1	2015-08-01 00:00:00
10	482.00	2000.00	9793.00	28.6	2015-09-01 00:00:00

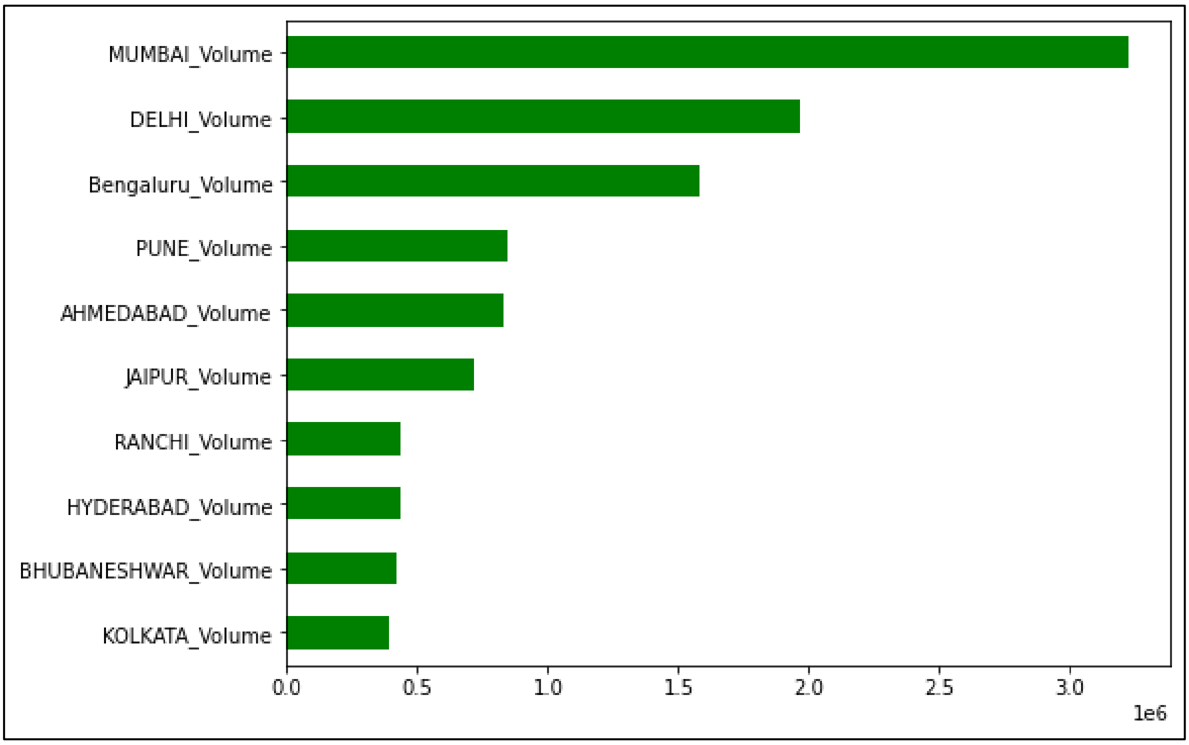
	AK	AL	AM	AN	AO
1	PUNE_WholesalePrice	PUNE_RetailPrice	PUNE_Volume	PUNE_Avg. Temperatu	Date
2	2207	2867	2843	21.3	2015-01-01 00:00:00
3	1465	1965	3163	23.1	2015-02-01 00:00:00
4	1350	1980	3285	26.3	2015-03-01 00:00:00
5	1178	1655	3040	29	2015-04-01 00:00:00
6	1153	1795	2793	29.6	2015-05-01 00:00:00
7	1309	1877	1839	27.3	2015-06-01 00:00:00
8	1589	2287	1567	24.8	2015-07-01 00:00:00
9	915	1298	1456	24.5	2015-08-01 00:00:00

Chapter 8: Data Preparation

Due to unavailability of the complete data from all the APMC's we have considered the top 10 Volume contributor of APMC's from Jan 2015- July 2020 - for Onion, Potato and Tomato

Table No.8.1: Top10 volume for Onion, Potato and Tomato





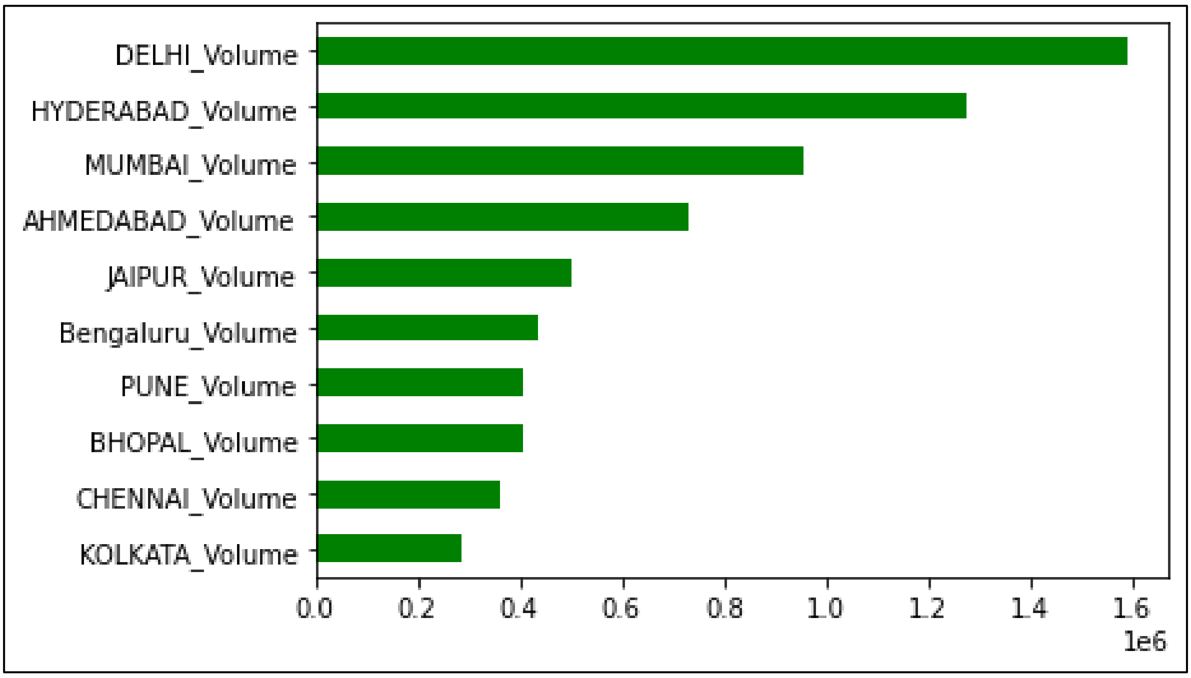
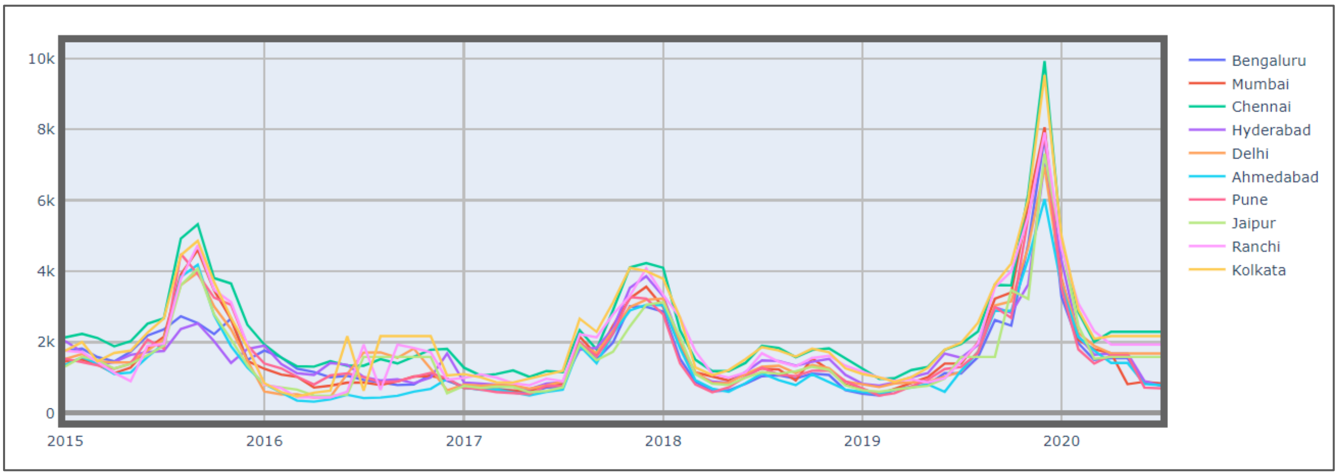
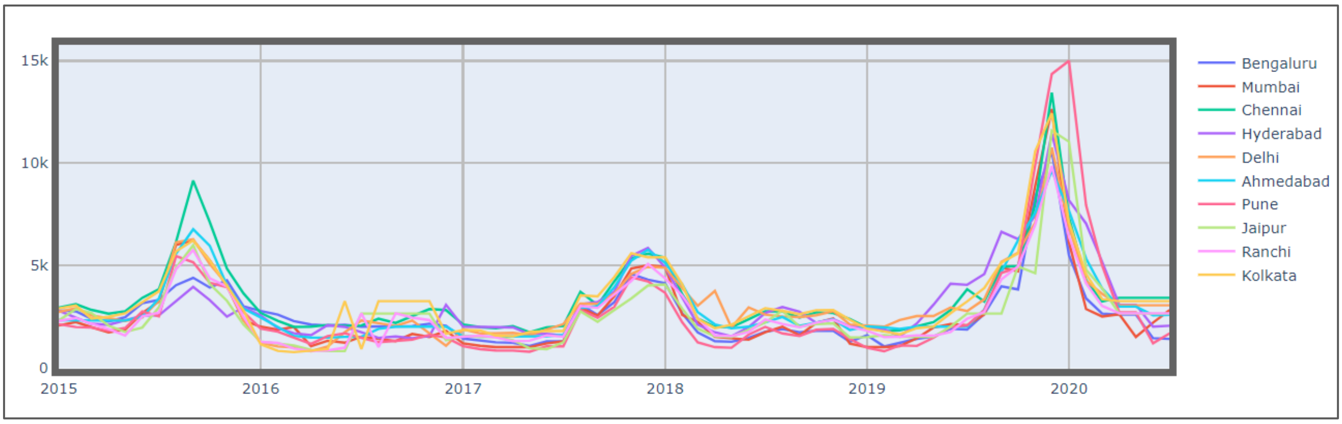


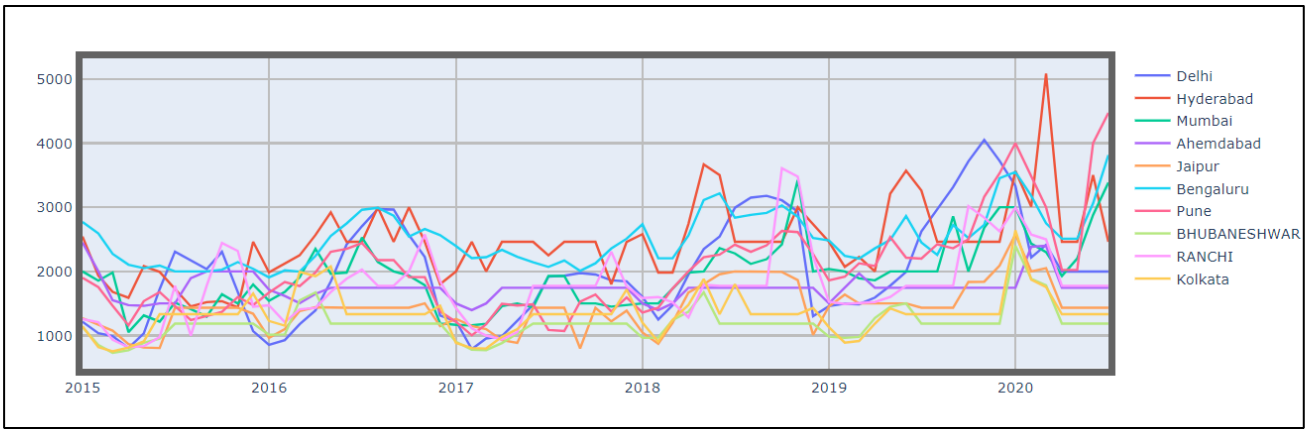
Figure No.8.2: Wholesale and Retail price Trend for Onion, Potato and Tomato

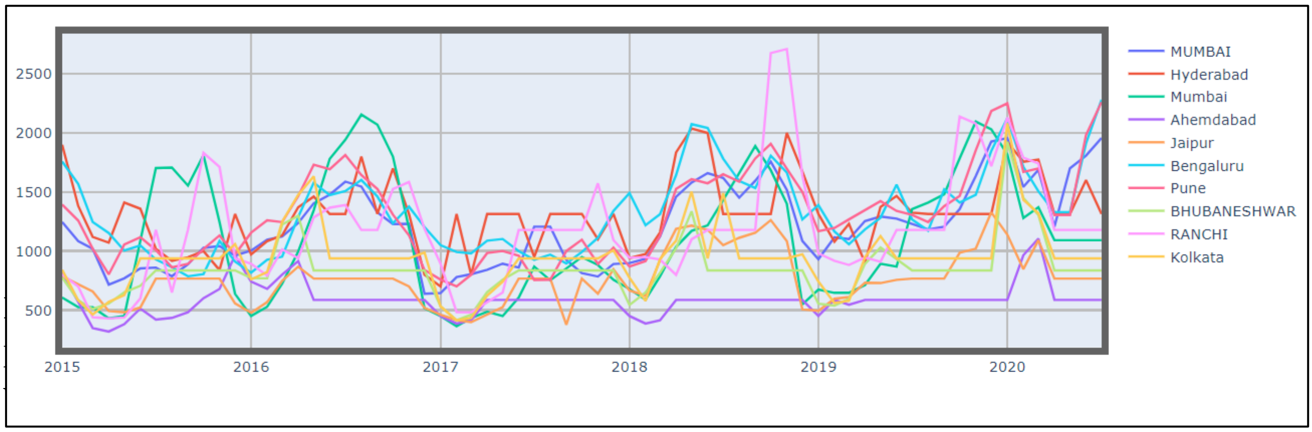
Here we can observe, the Whole sale price and retail price trend from all the APMC's in the period of Jan 2015 – July 2020 for Onion commodity and we have understood the price variations in each year

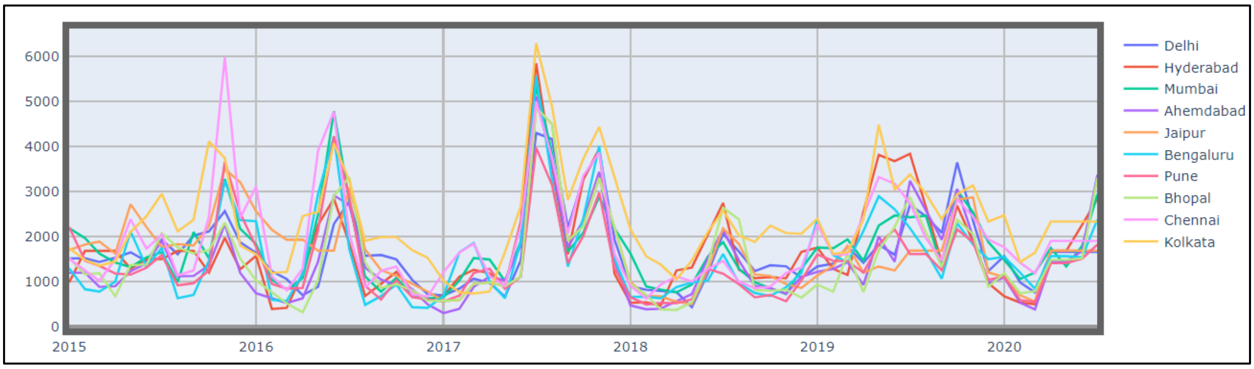




Here we can observe, the Whole sale price and retail price trend from all the APMC's in the period of Jan 2015 – July 2020 for Potato commodity and we have understood the price variations in each year







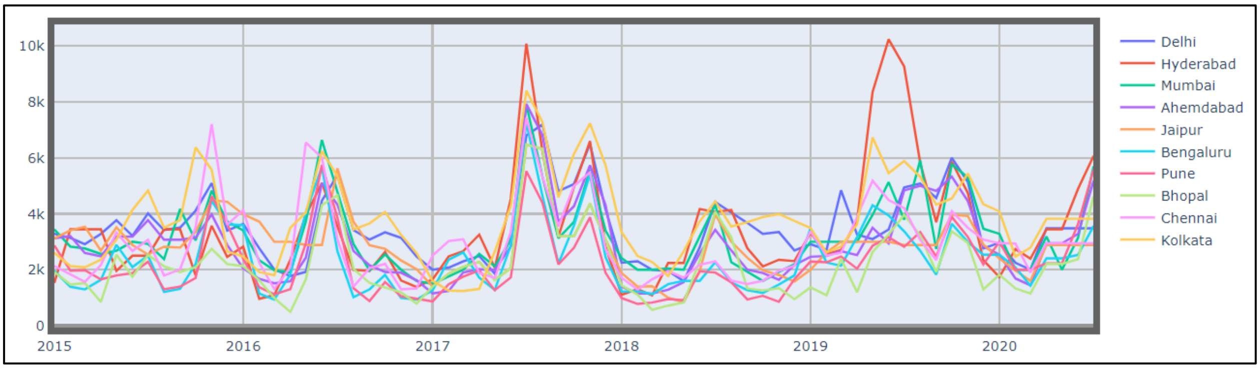
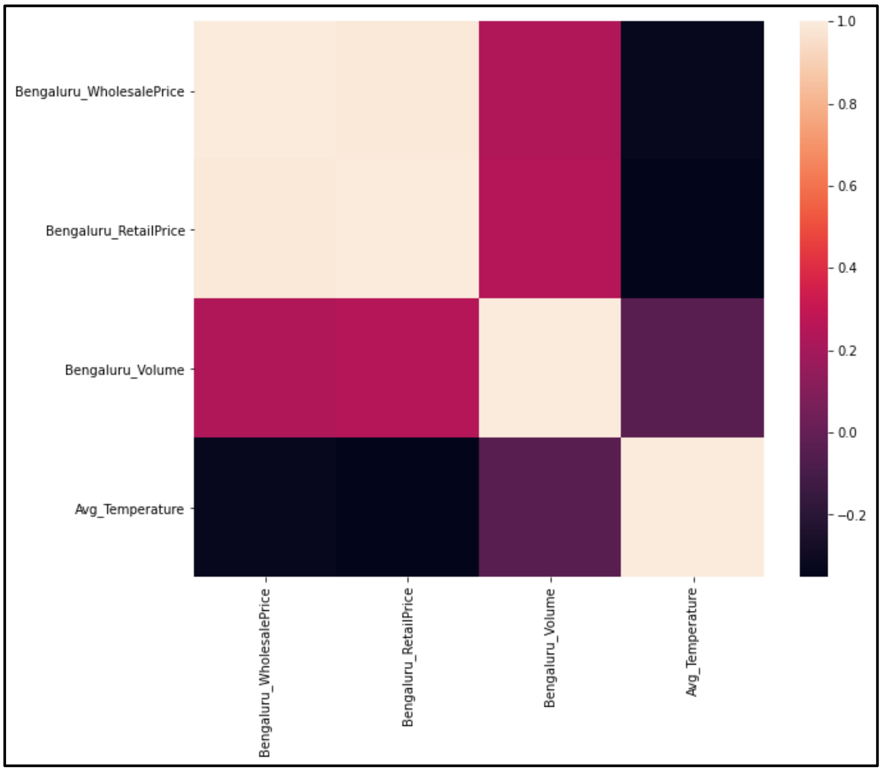
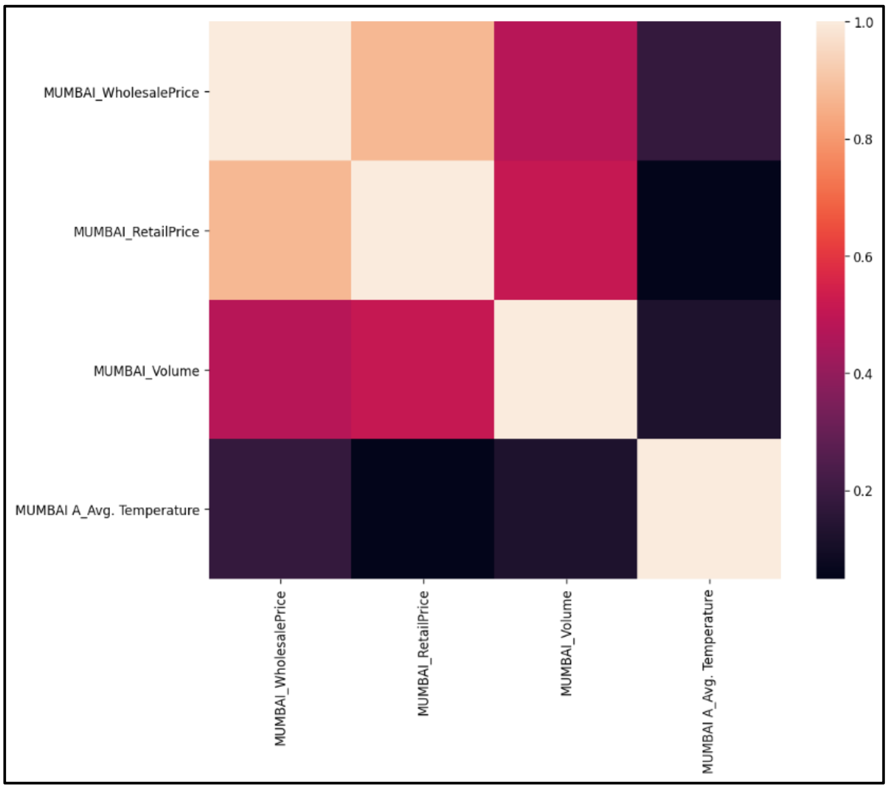


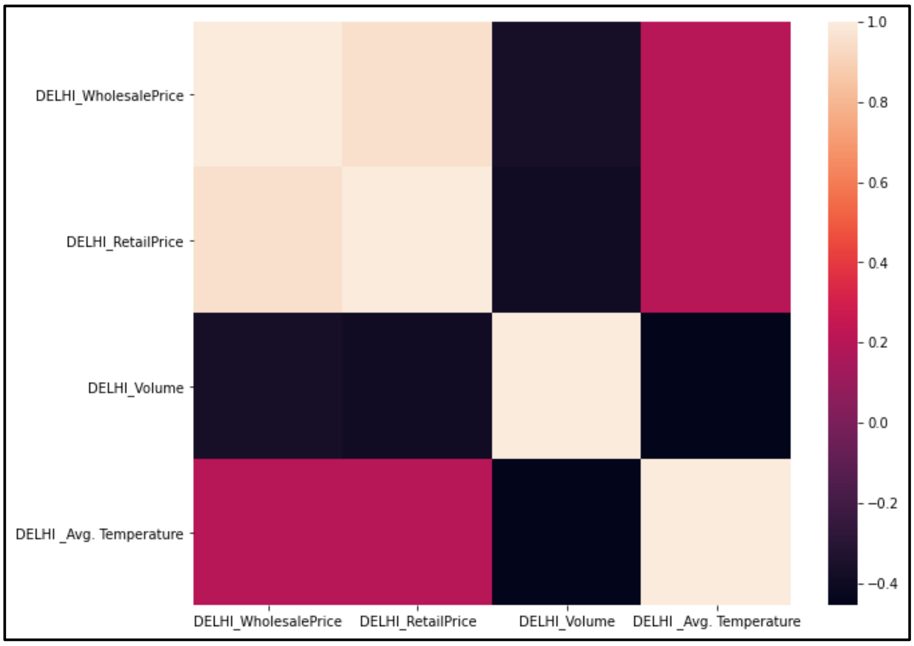
Figure No.8.3: Correlation matrix - Onion, Potato & Tomato



Temperature is the strong correlation with Wholesale price and retail price Volume is having good correlation with Wholesale price and retail price

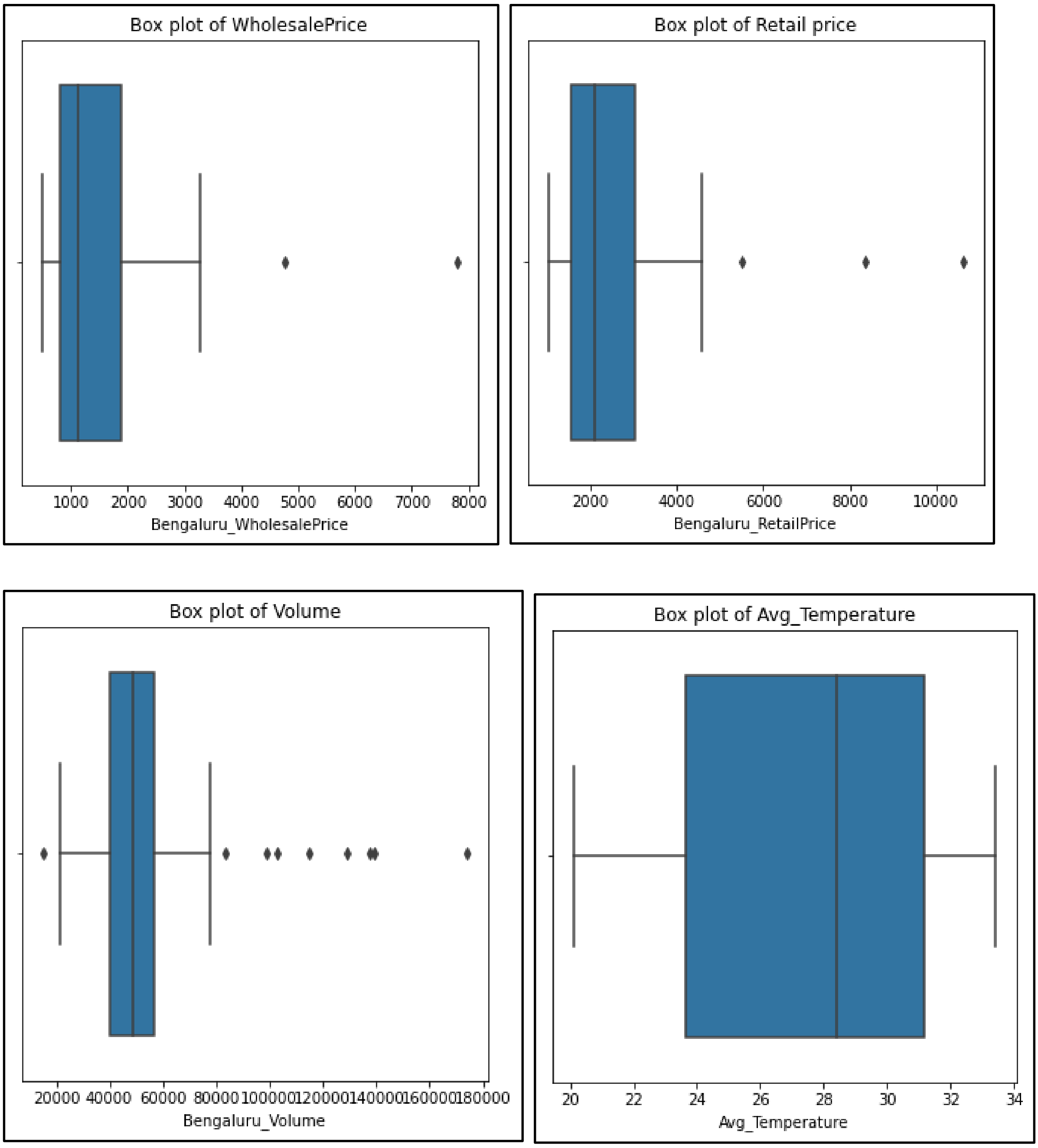


Temperature is having strong correlation with retail price and, Temperature is having good correlation with Wholesale price and Volume, Volume is having good correlation with Wholesale price and retail price

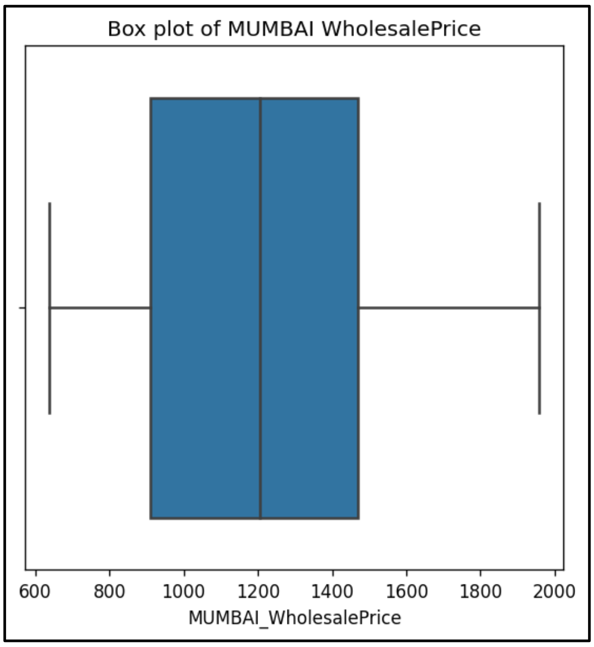


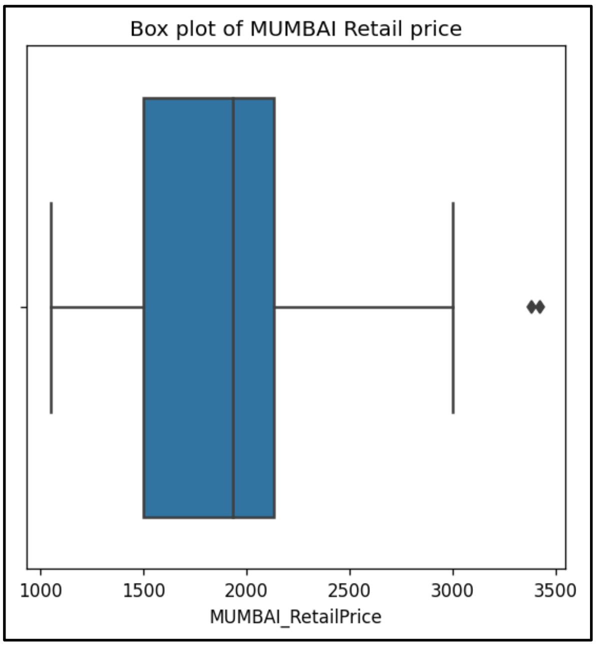
volume is the strong correlation with Wholesale price and retail price
Temperature is having good correlation with Wholesale price and retail price

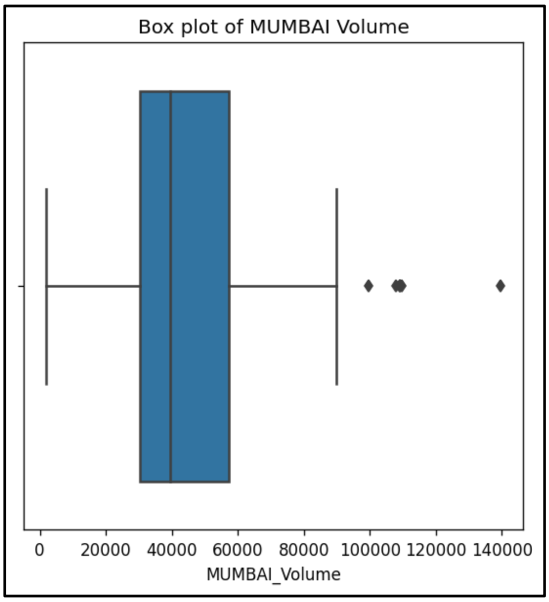
Figure No.8.4: Outliers for Onion, Potato& Tomato – Jan 2015 to Jul 2020

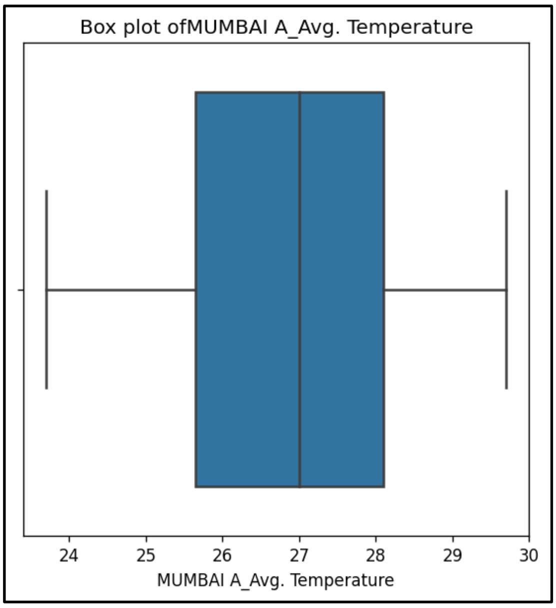


The data distribution is normal and Looking at the histogram, we can see that data values for wholesale price is between 1000 Rs to 4000 Rs, 1000Rs were most frequent. In retail price is between 1000Rs to 4000Rs 1000Rs is the most frequent Volume count is between 20000 to 75000, 50000 are contributing more, Temperature range between 20 to 32 Deg

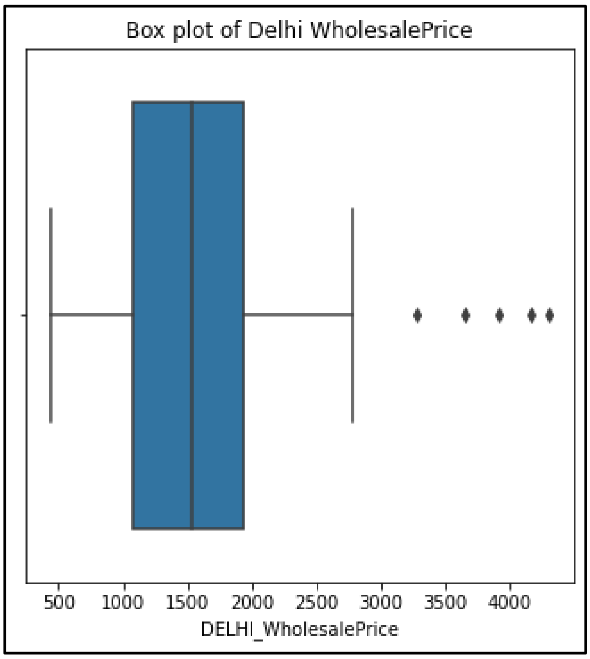


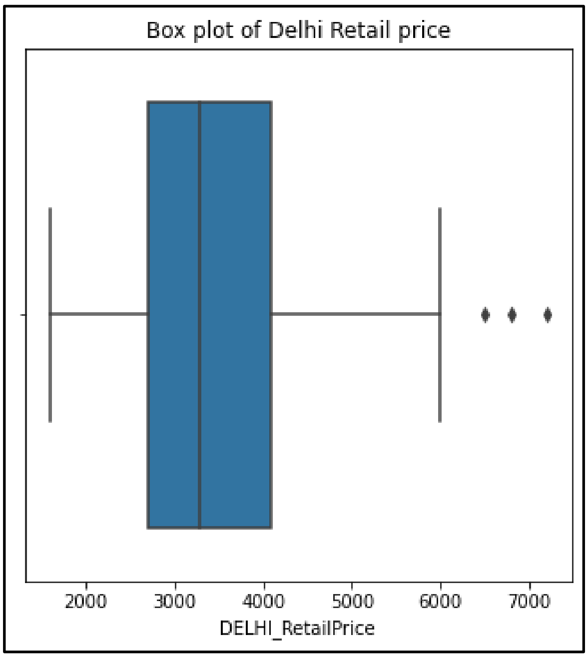


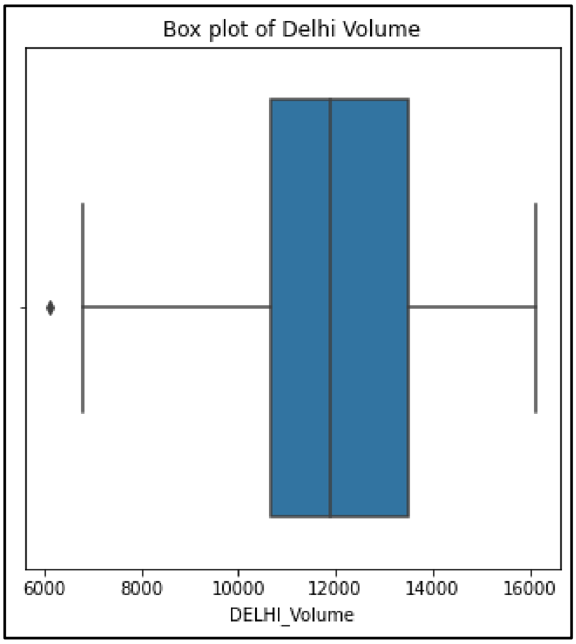


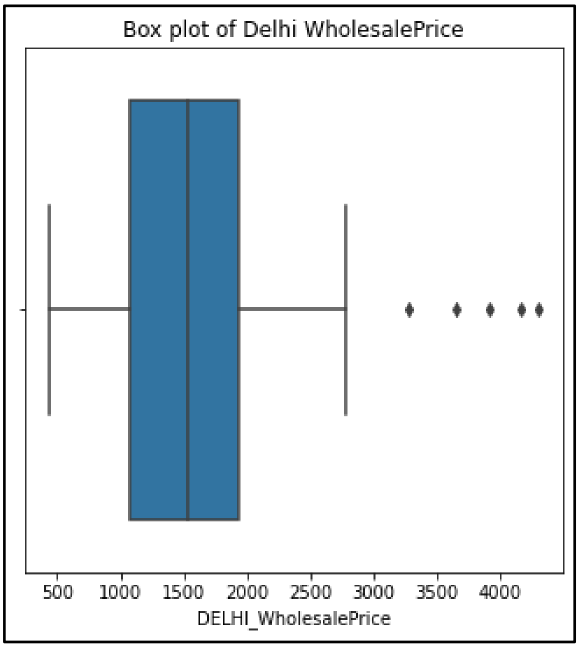


The data distribution is normal and Looking at the histogram, we can see that data values for wholesale price is between 2500 Rs to 3500 Rs, 2000Rs were most frequent. In retail price the 1000Rs to 1500Rs is the most frequent Volume count is between 25000 to 50000 are contributing more, Temperature range between 24 to 29 Deg









The data distribution is normal and Looking at the histogram, we can see that data values for wholesale price is between 500 Rs to 3000 Rs, 1500Rs were most frequent. In retail price the 1500Rs to 5000Rs is the most frequent

Chapter 9: Data Modeling

Now we are building the models for our top 10 volume contributors from each APMC's.

ARIMA

In the predicted model we can see the slight autocorrelation

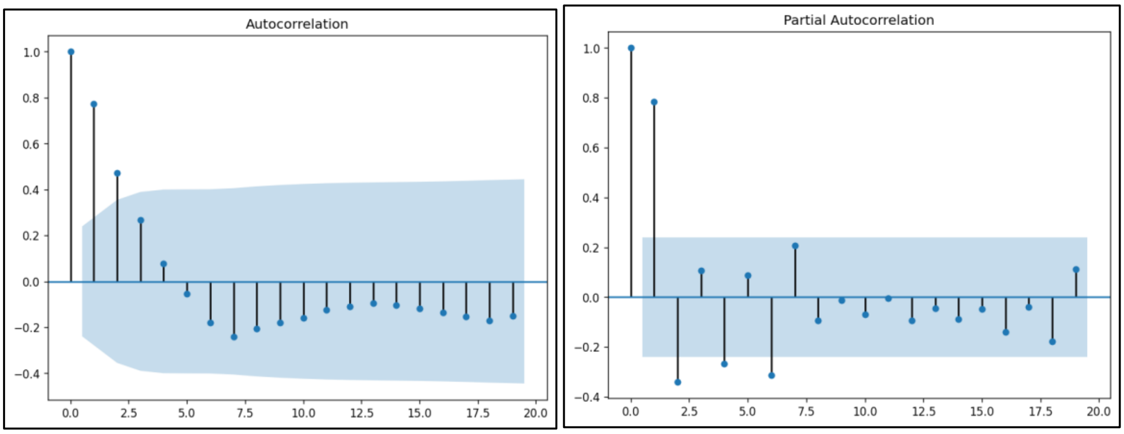


Figure No.9.1: "ACF and PACF"

As you see, in certain months, there are substantial increases in the volume of commodities. That should explain why price cost too much during the mid of every year! The trend is increasing consistently, i.e., session is impacting over time. The seasonality plot shows us certain periods are associated with hikes and this is consistent over the course of the years. Now, let us talk about autocorrelation.

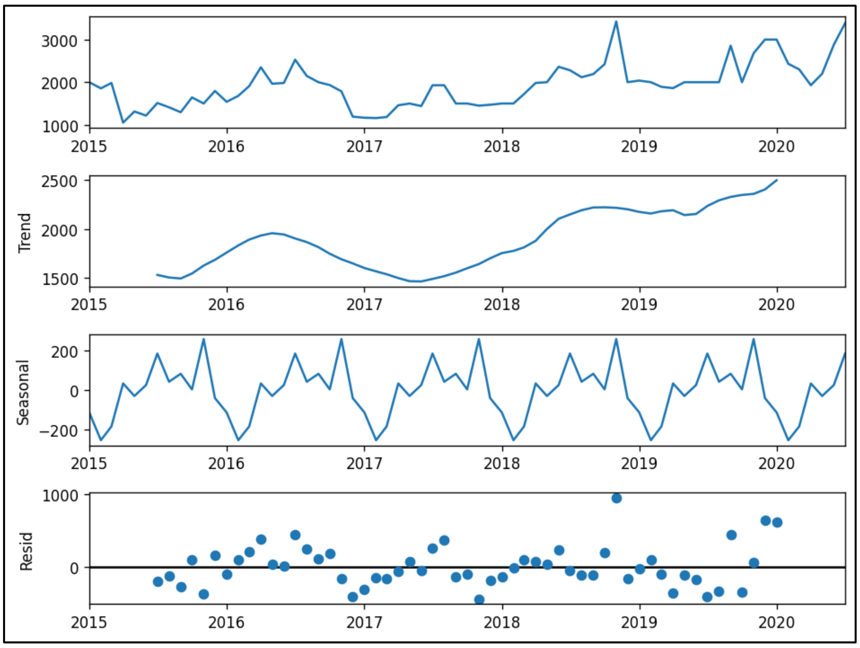


Figure No.9.2: SARIMA Trend, Seasonal& Residual

Looking at the forecasting we are pretty good at prediction, but it is not that much promising as LSTM model, accuracy is about ~67% and RMSE is 2818 RS, which is pretty high.

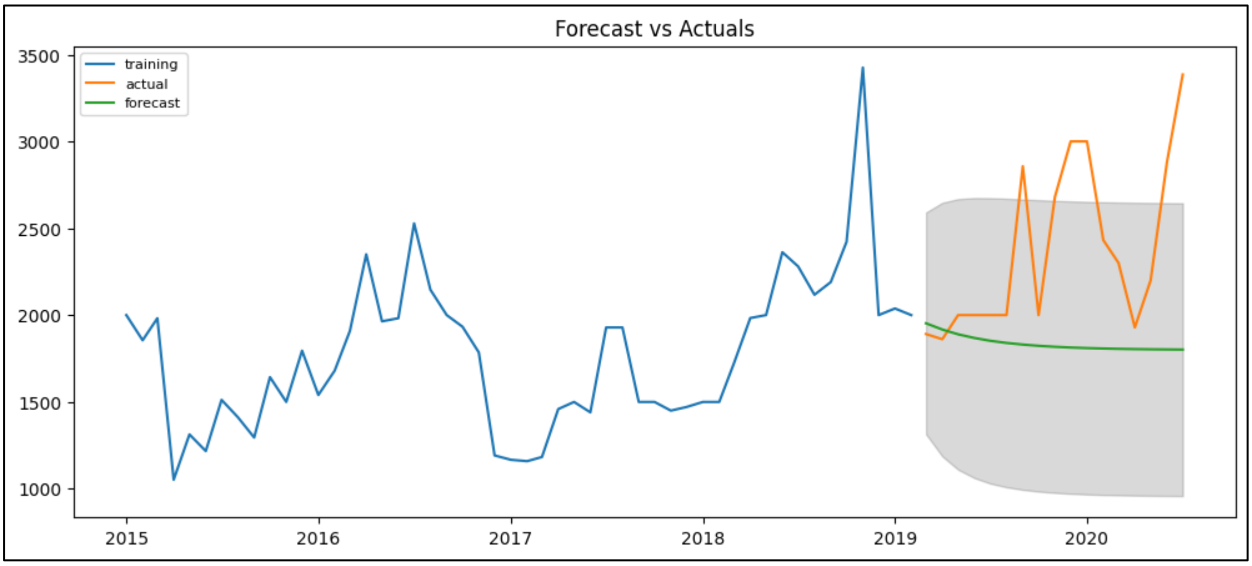


Figure No.9.3: Price forecast using ARIMA

SARIMA

The past prediction is pretty good, and model is failing to consider the anomaly events, we can observe in 2019 year end the anomalies are not following.

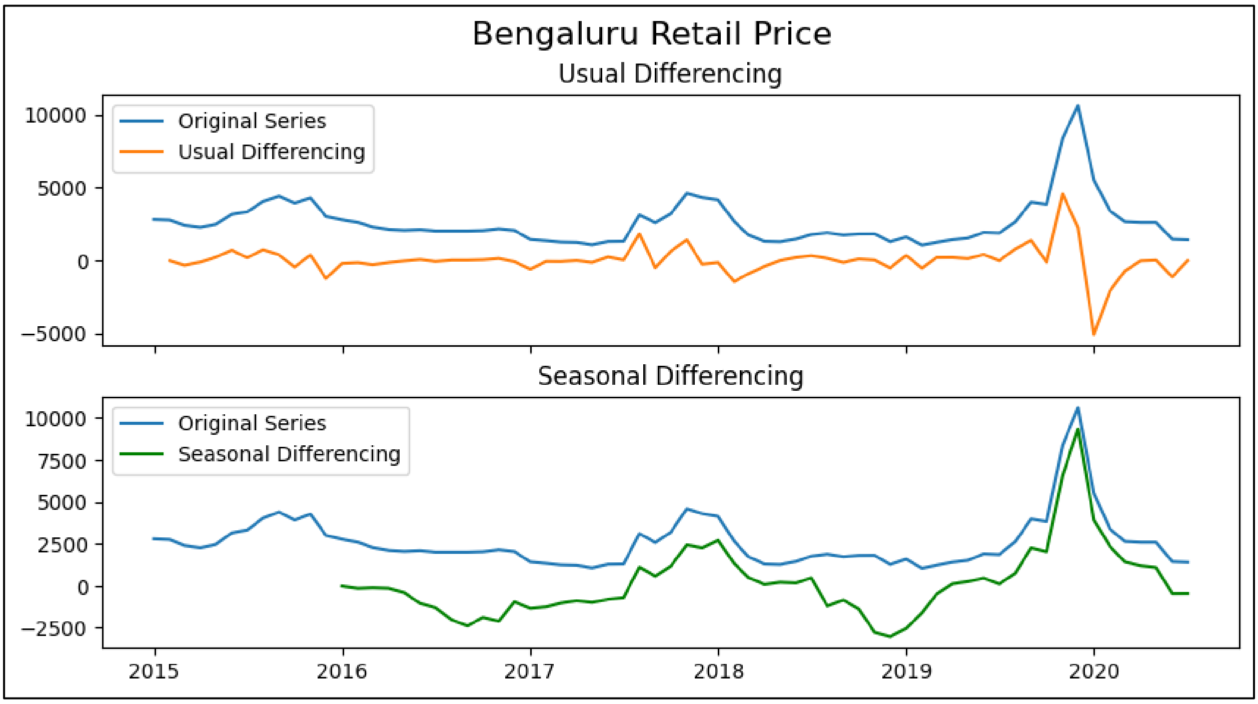


Figure No.9.4: Seasonality differencing for retail price

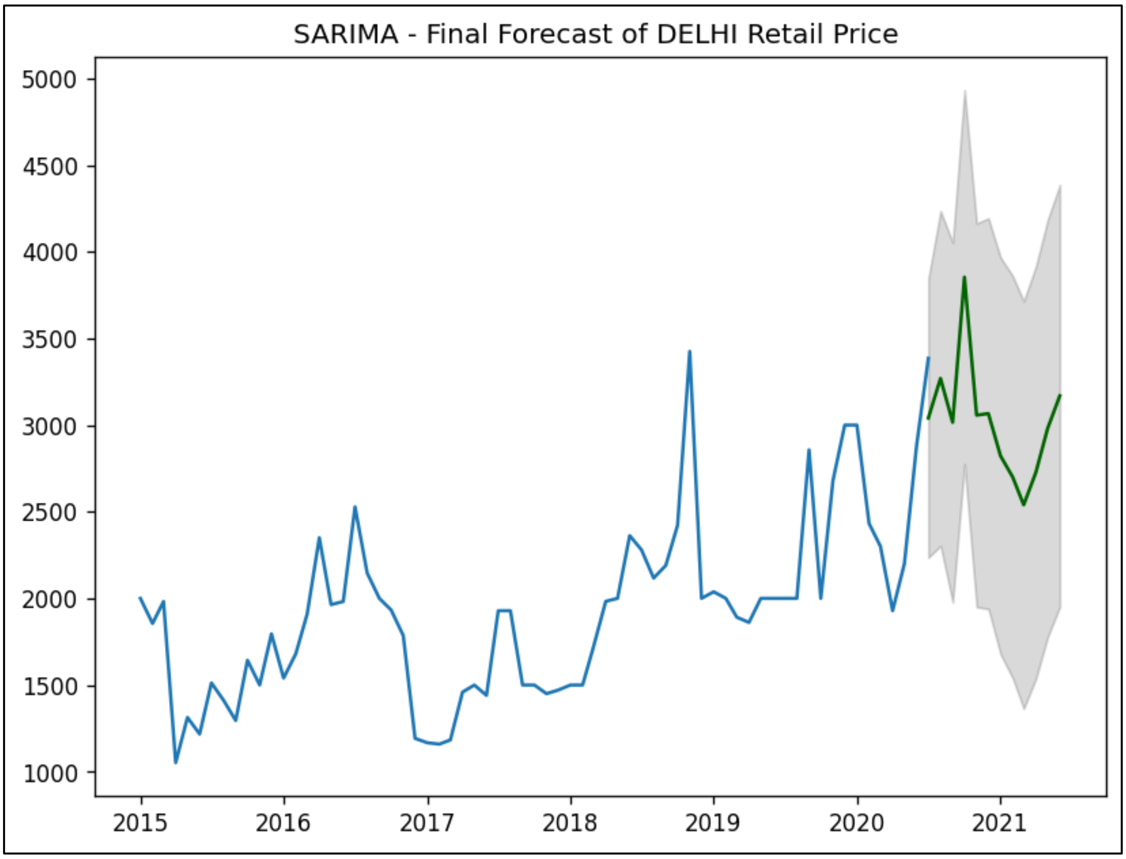


Figure No.9.5: Price forecast using SARIMA

Looking at the forecasting we are pretty good at prediction, but it is not that much promising as LSTM model, accuracy is about ~67% and RMSE is 2818 RS, which high.

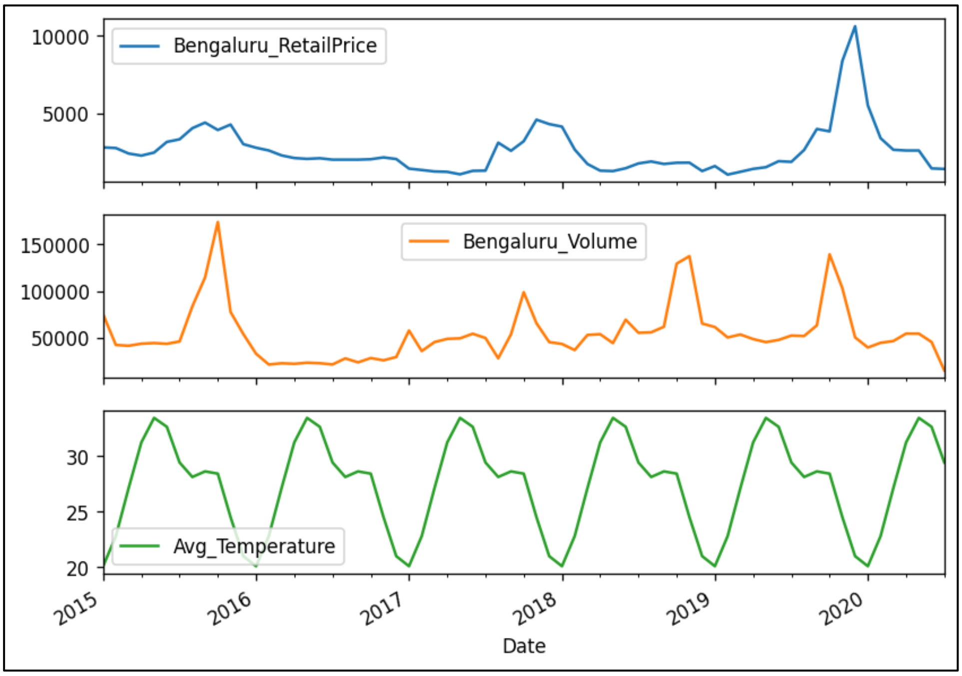
We have used the Tensor flow multivariate time series models (Multi-output models such as Baseline, Linear, Dense Multi step dense, Convolution and LSTM), ARIMA and SARIMA for our model building and predictions

.

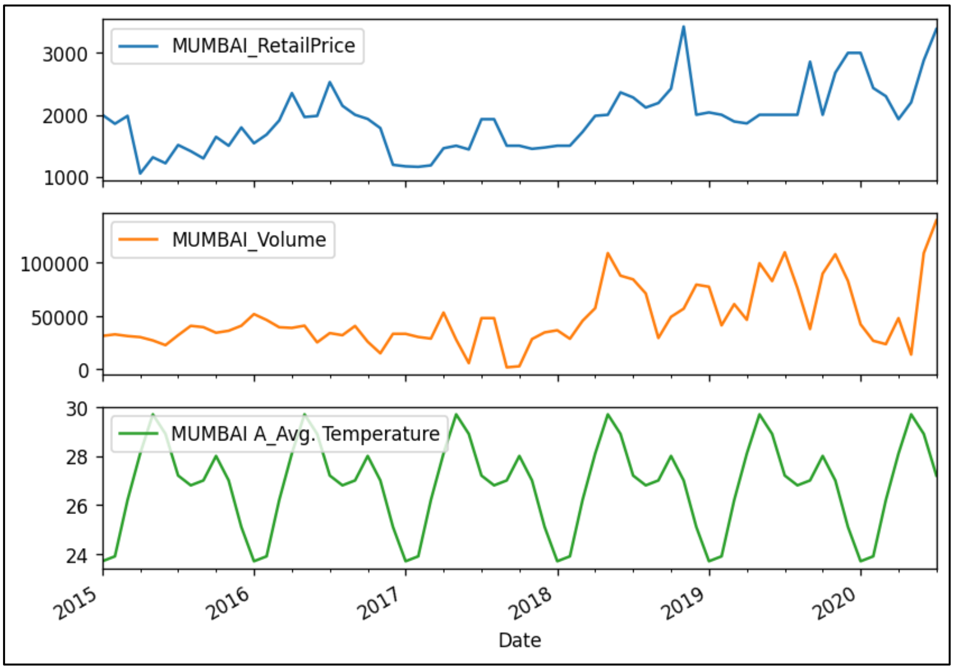
Now will discuss on Tensor flow multivariate time series models building steps
Window Generator is as follows input width=34, label width=33, shift=1

Figure No.13: Trend distribution for retail price, Volume and Temperature for Bangalore APMC's in the period of Jan 2015 – July 2020 for Onion commodity and we can see the clear indication data variations in each year

Figure No.9.6: Trend distribution for retail price, Volume and Temperature



Trend distribution for retail price, Volume and Temperature for Mumbai APMC's in the period of Jan 2015 – July 2020 for Potato commodity and we can see the clear indication data variations in each year



Trend distribution for retail price, Volume and Temperature for Delhi APMC's in the period of Jan 2015 – July 2020 for Tomato commodity and we can see the clear indication data variations in each year

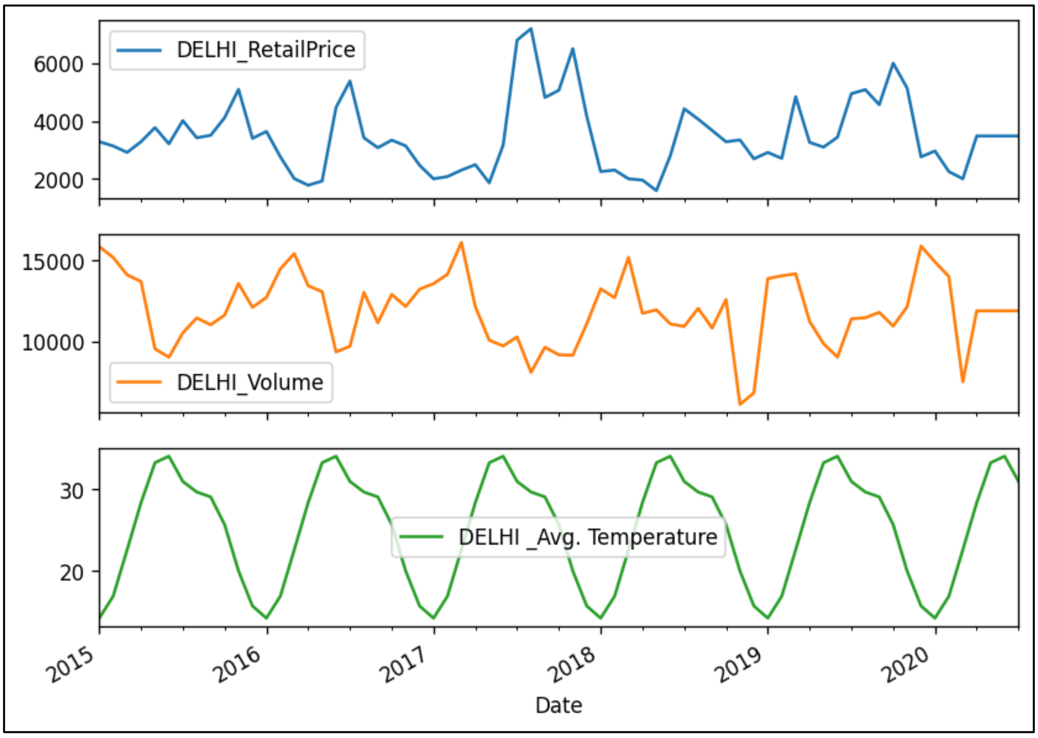
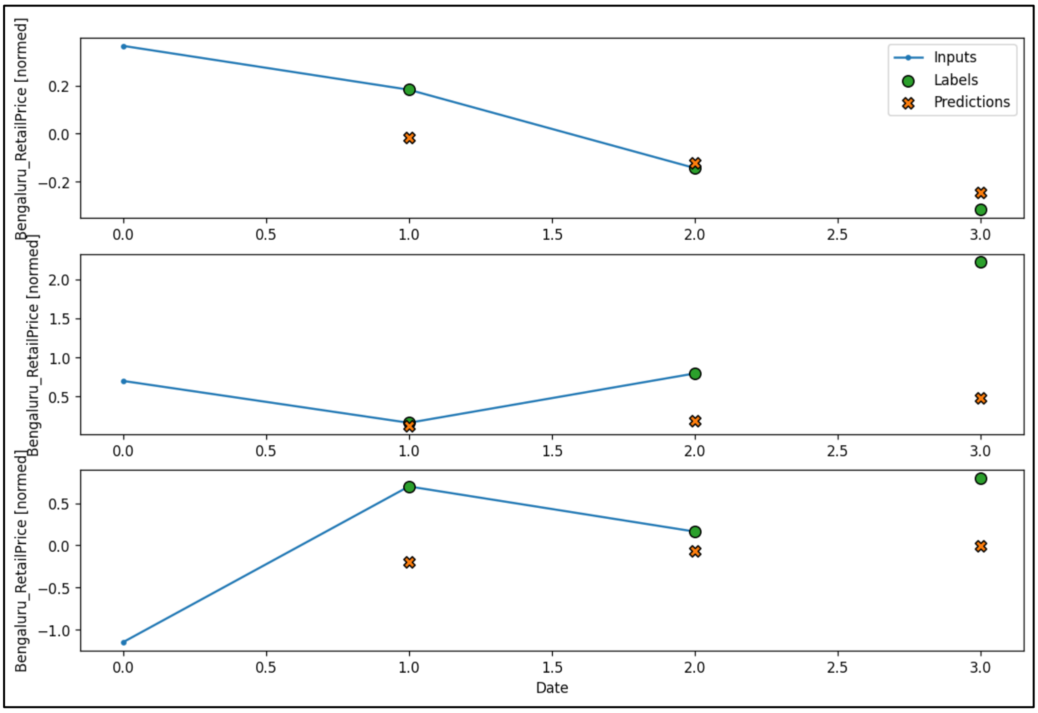


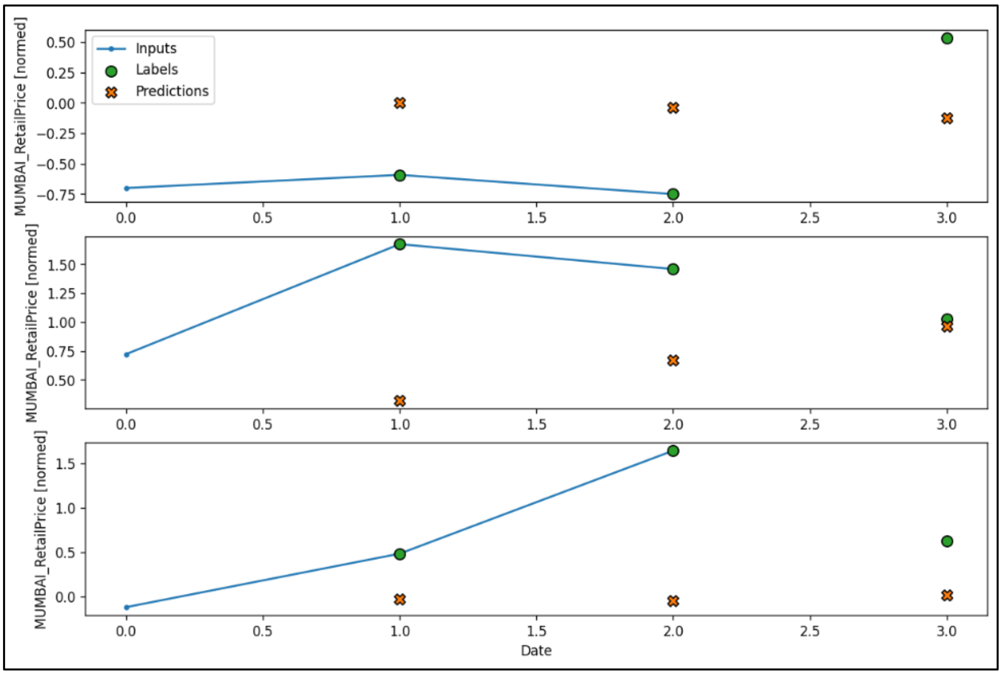
Figure No.9.7: LSTM Model prediction

For the Bengaluru retail price, we have considered the wide window = Window Generator input width=3, label width=3, shift=1.We built model for Baseline, Linear, Dense Multi step dense, Convolution and LSTM, here we have captured the details on LSTM model, since LSTM model has given the promising output.



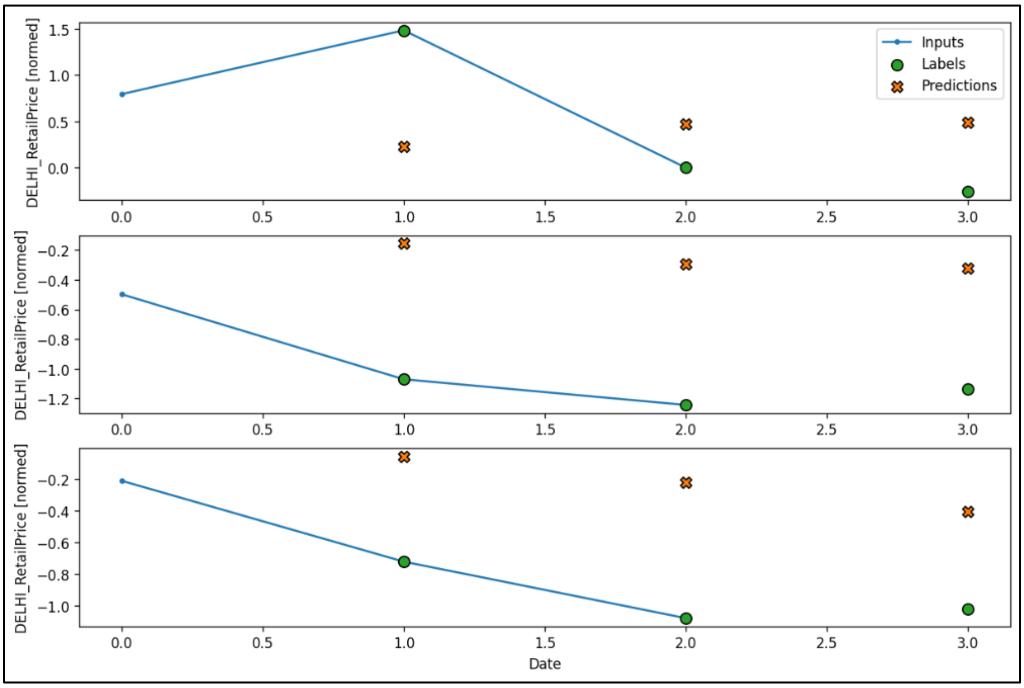
For the Mumbai retail price, we have considered the wide window = Window Generator input width=3, label width=3, shift=1.

We built model for Baseline, Linear, Dense Multi step dense, Convolution and LSTM, here we have captured the details on LSTM model, since LSTM model has given the promising output also Linear model is closer to LSTM.



For the Delhi retail price, we have considered the wide window = Window Generator input width=3, label width=3, shift=1.

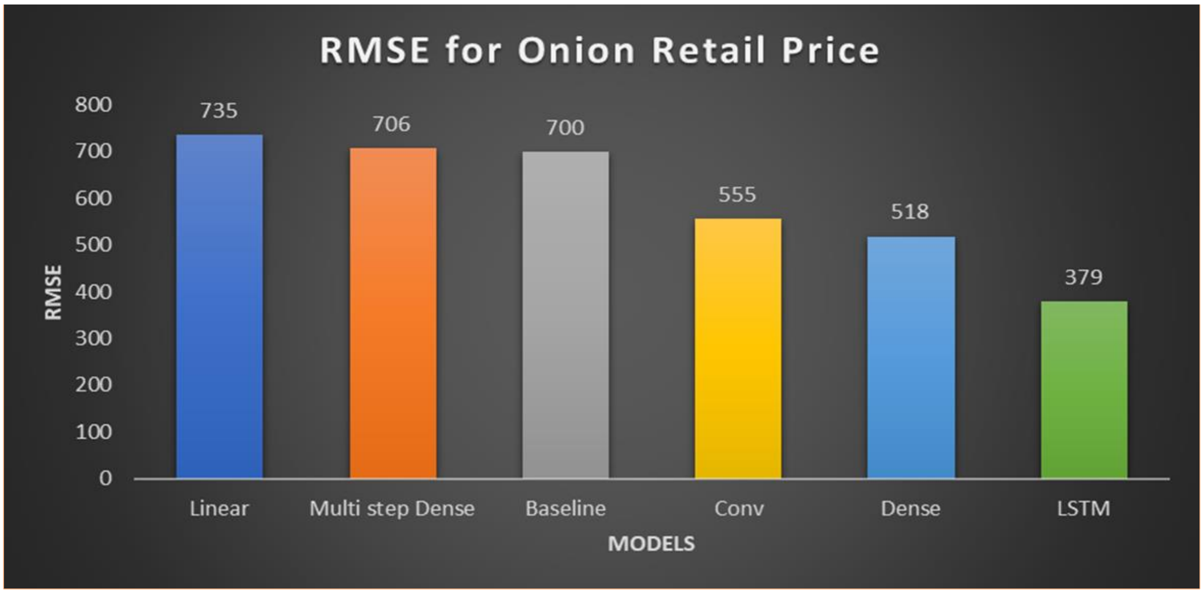
We built model for Baseline, Linear, Dense Multi step dense, Convolution and LSTM, here we have captured the details on multi step dense model, since multi step dense model has given the promising output also Conv and LSTM models are given the closer results



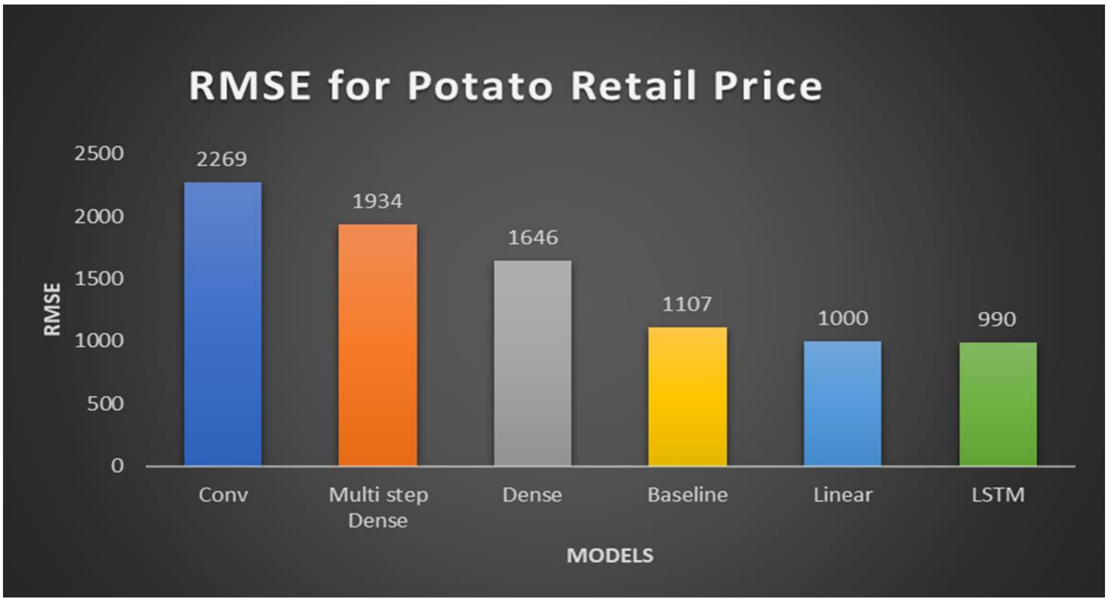
Chapter 10: Data Evaluation

Price Forecasting Model Comparison (RMSE Values reported are the average of RMSE values over the whole time series)

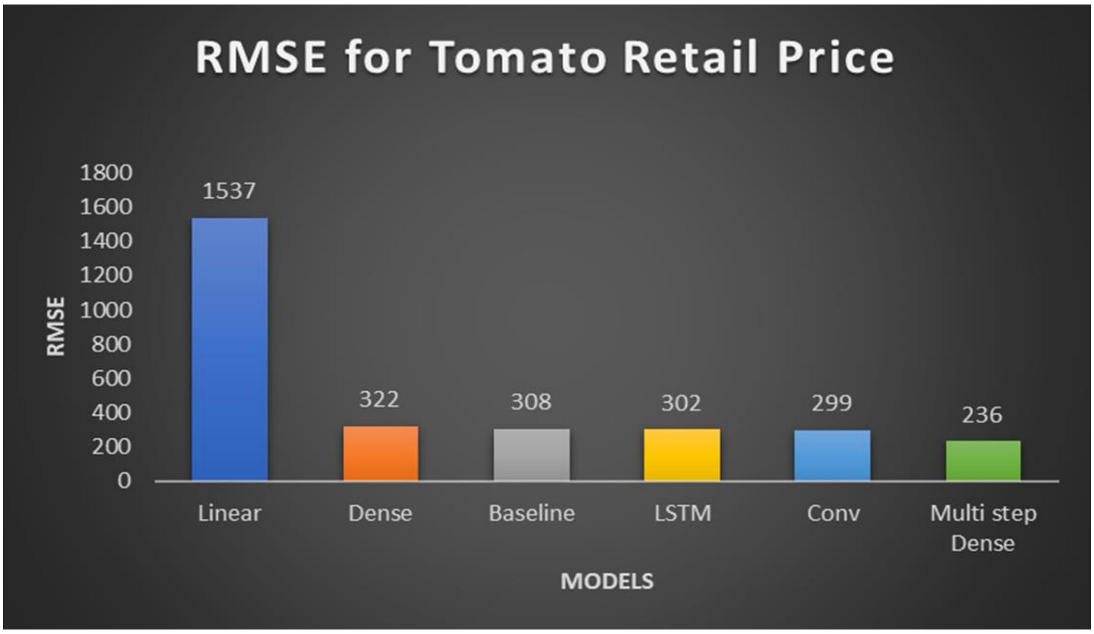
Figure No.10.1: Price Forecasting Model Comparison



LSTM model has given the promising RMSE output for Bengaluru retail price for Onion commodity.



LSTM model has given the promising RMSE output for Mumbai retail price for Potato commodity, also Linear model is closer to LSTM

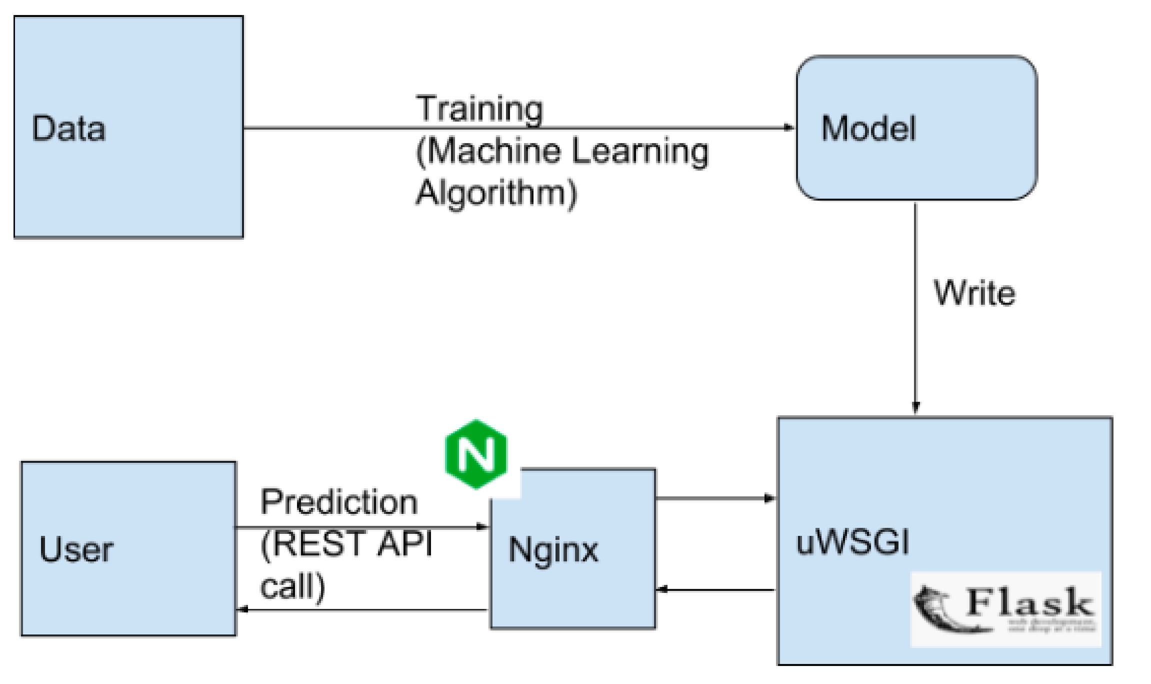


Multi step dense model has given the promising RMSE output for Delhi retail price for Tomato commodity, also Conv and LSTM models are given the closer results

Chapter 11: Deployment

As a future work I am planning to build a front-end model using flask.

Figure No.11.1: Front end proposals

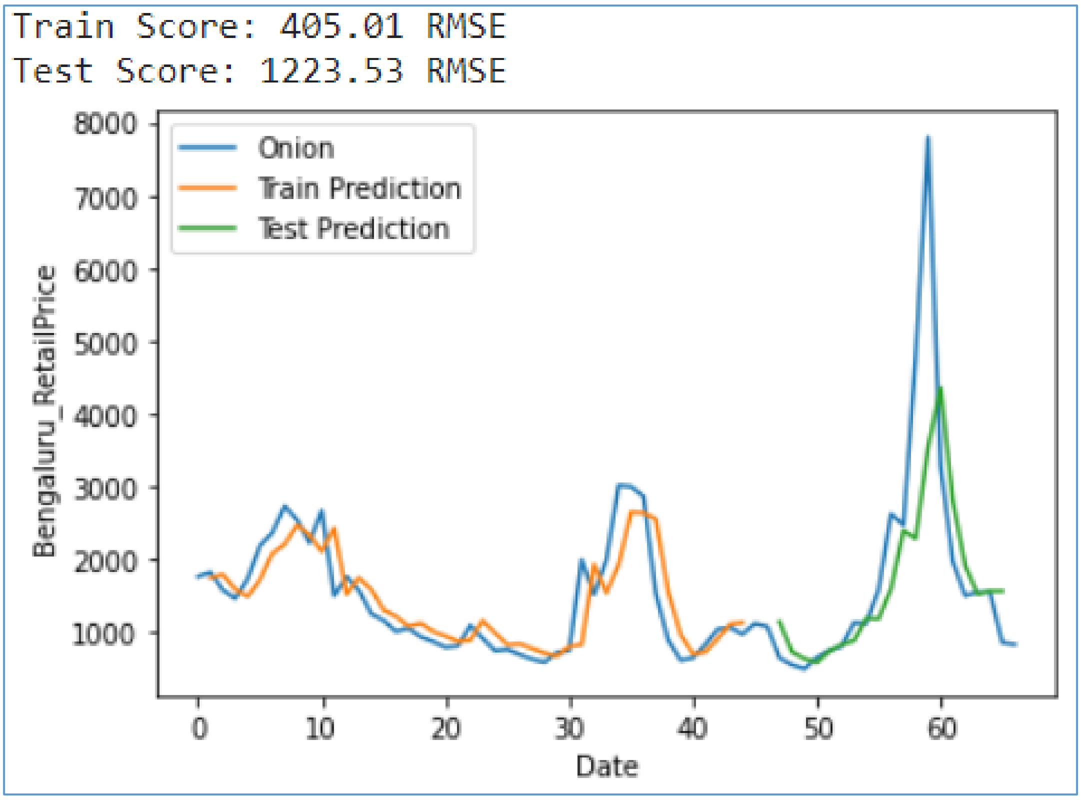


Building and deploying REST service (for ML model) on production

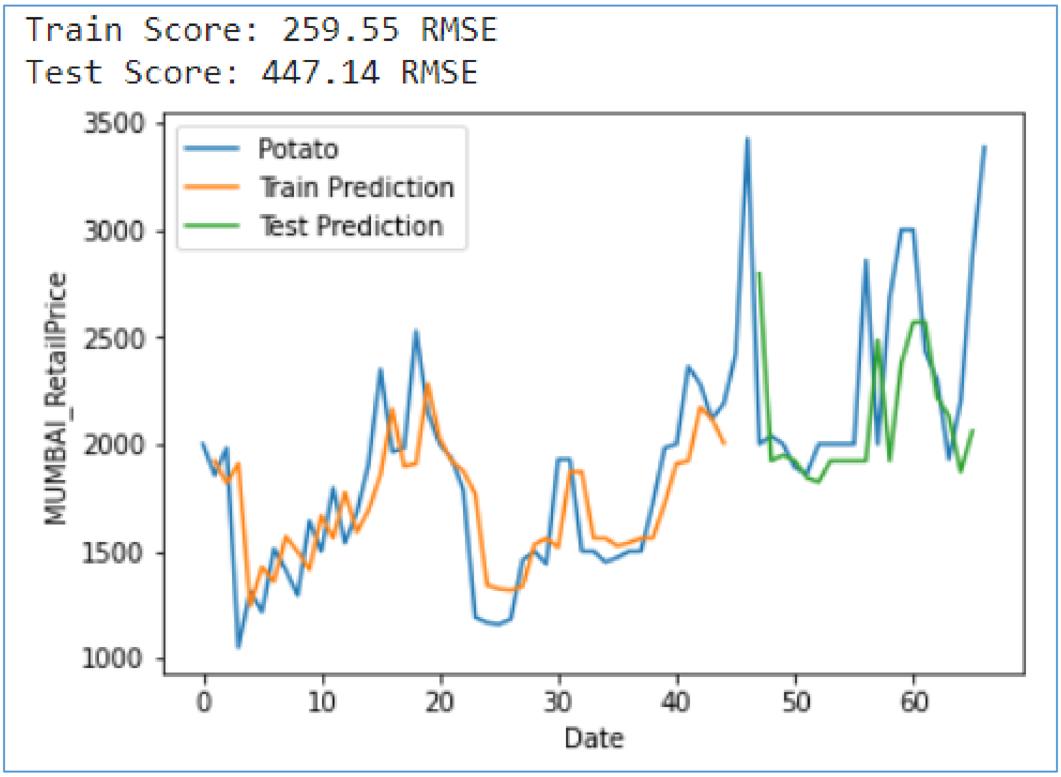
Chapter 12: Analysis and Results

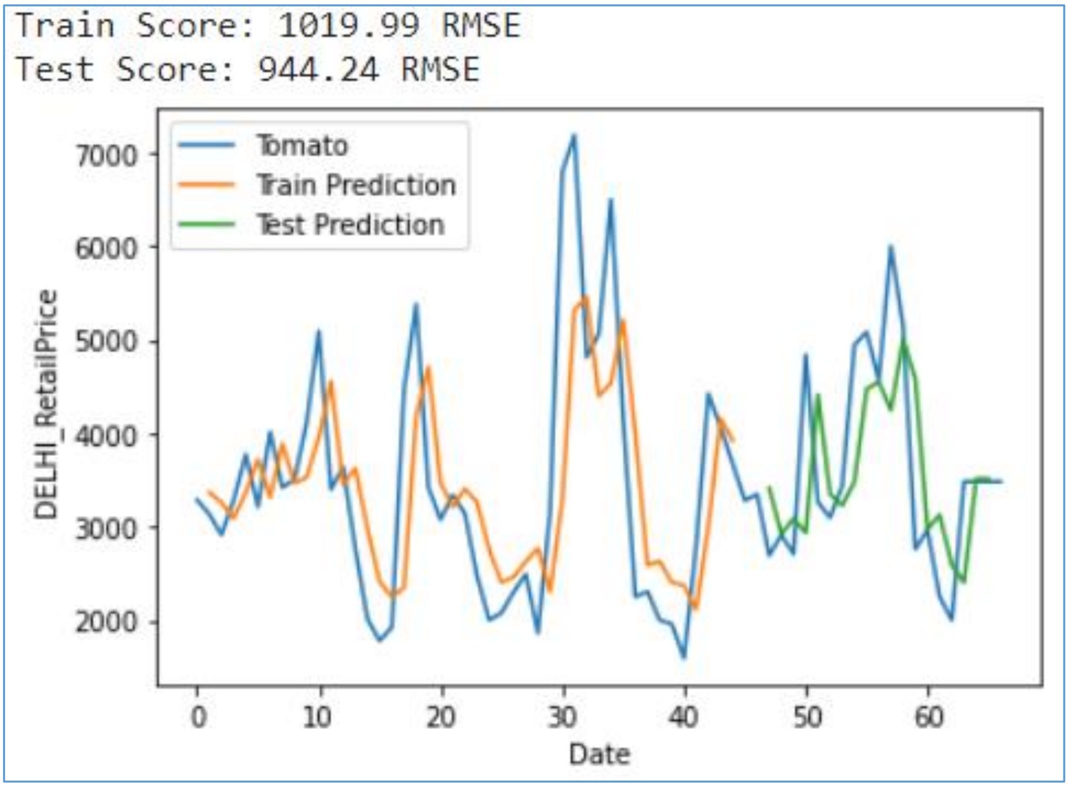
After analyzing and building the traditional and deep learning models we have learnt that the deep learning models are very accurate and suitable for in for solving the multivariate, and also it's easy to consider the each and every events that occurs in the particular time period and predict the three months commodity price

Figure No12.1-: LSTM Forecasting Model



.





We can see from the trend that the actual retail price went up, while our model also anticipated that the retail price will rise. This clearly demonstrates how good LSTMs are for the analysis of time series and concurrent details.

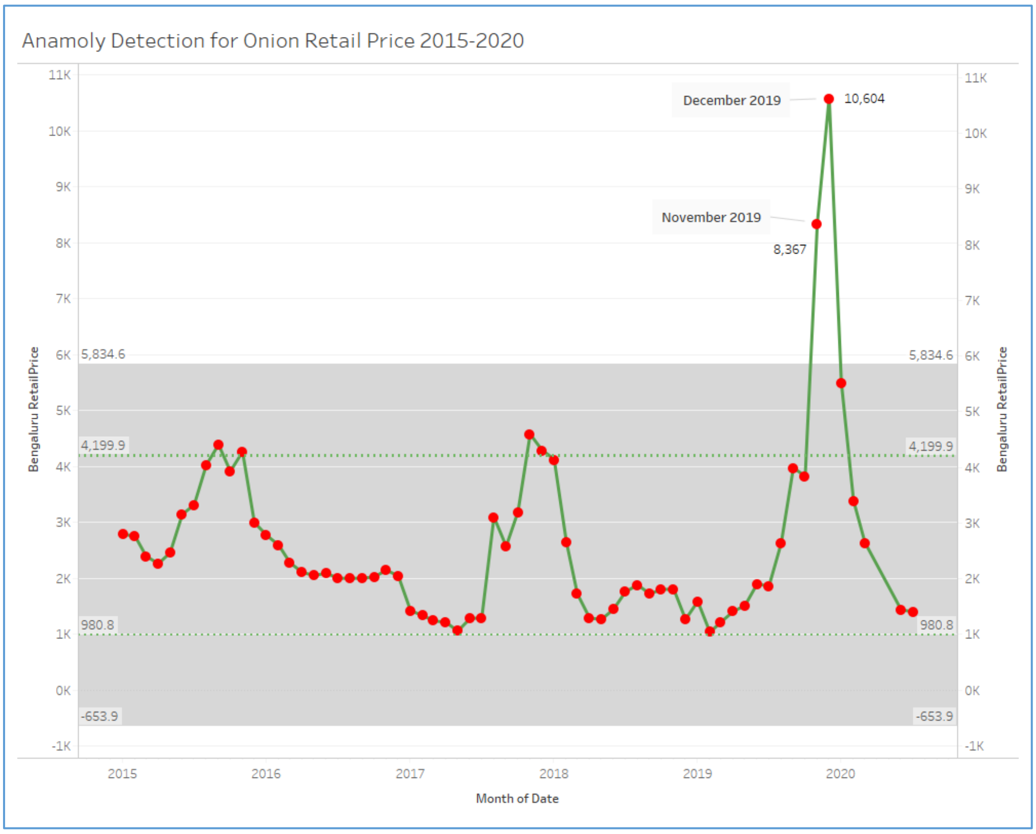
Anomaly discovery:

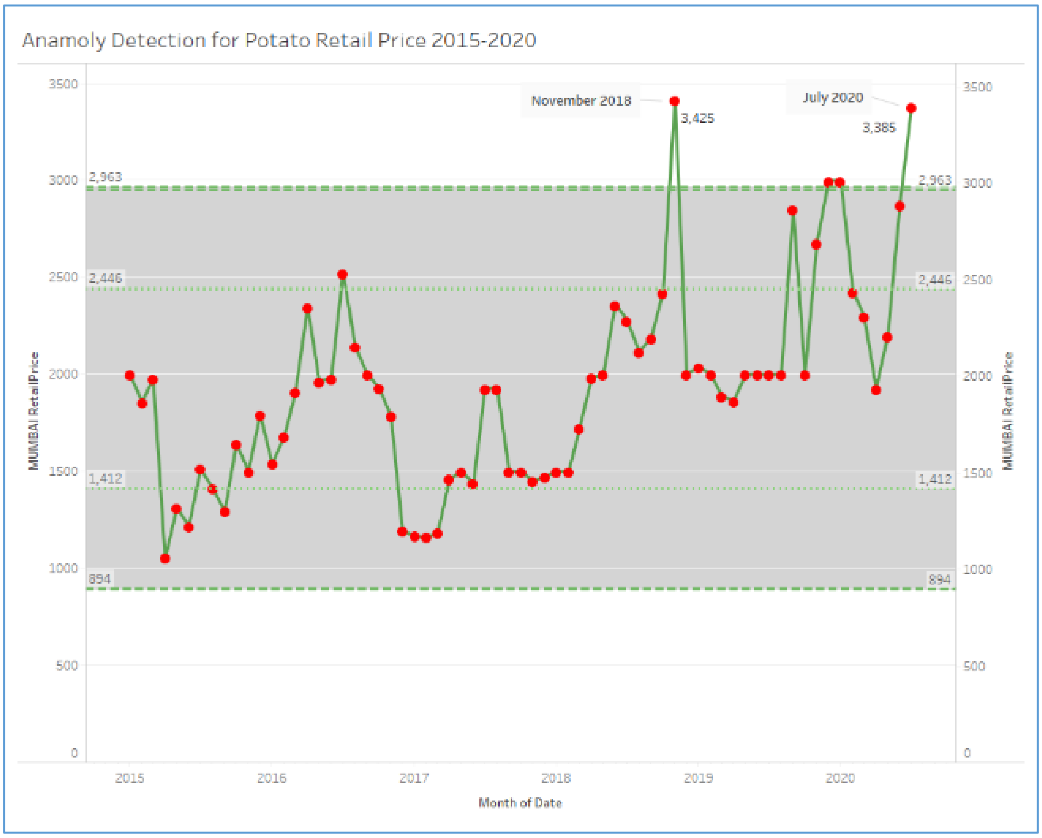
We are now moving to the second issue of identifying trading malpractices, such as hoarding. using newspaper reports we initially recognize circumstances of hoarding and climate-associated abnormalities.

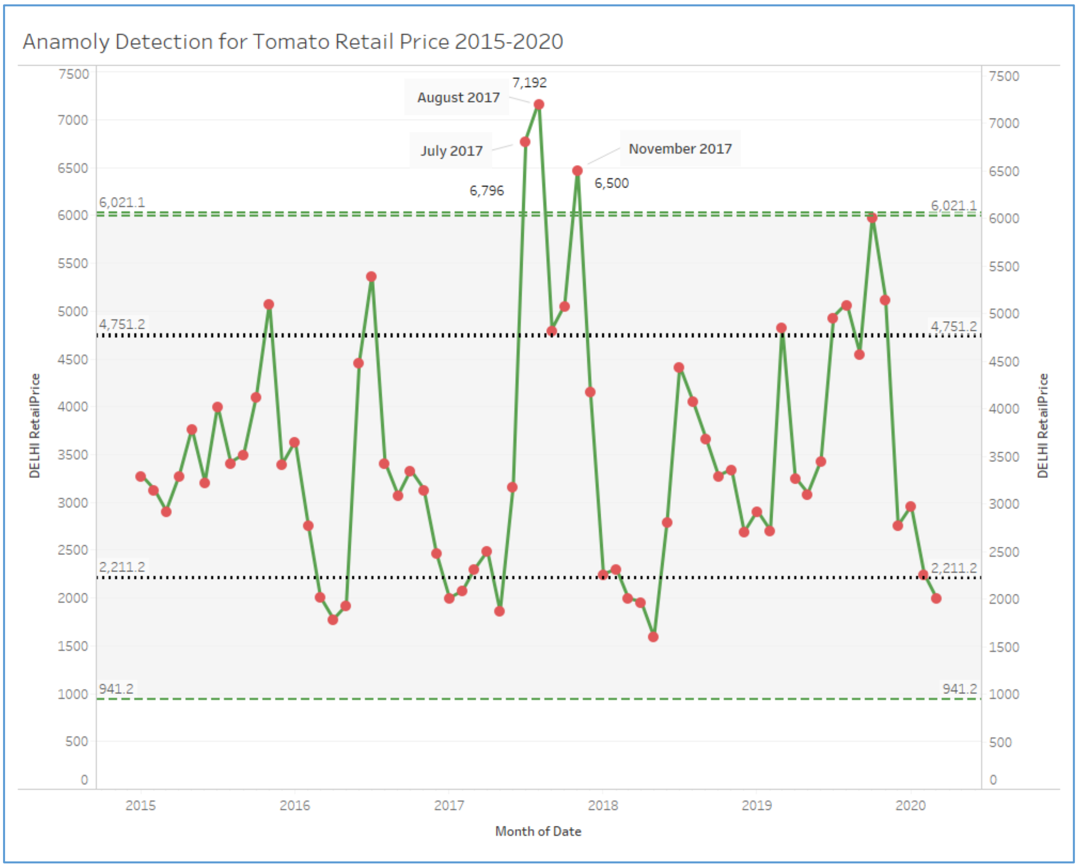
We separate articles into hoarding instances regardless of the climate, and climate incidents while no hoarding occurs.

Real time examples on anomaly discovery with existing data set.

Figure No.12.2: Anomaly discovery for retail price







“Onion prices skyrocket again, close to Rs 100/kg in some states”(India Today Web Desk, 2019)
The great onion crisis of 2019(Parija, 2019)

Anomaly discovery for Potato in the month of Nov 2018 and July 2020
Potato prices likely to double SECTIONS(Ghosal, 2018)
Supply shortage, rise in demand push potato prices in India up by 40% in major cities(Potatonewstoday, 2020)
Anomaly discovery for Tomato in the month of July, Aug, and Nov 2017
“Monsoon puts tomatoes on fire, likely to remain pricey for another 2 weeks(Monsoon puts tomatoes on fire, 2017)”

Chapter 13: Conclusions and Recommendations for future work

Our next phase is to categorise the anomaly type, i.e. If it is really a hoarding incident, or a weather incident, where there was no hoarding. We will be able to boost this with a greater number of datasets with many different classifiers.
As a future work to improvise our model and get the precise results, we are planning to gather the information’s from farmers by getting them on-board and conduct awareness sessions on how trades work in agricultural market.

APMCs (Agriculture Produce Market Committee) are operated by elected representatives with the purpose of conducting APMC business in a visible approach so that both farmers and buyers are will not abused with intermediaries and brokers. The mandis we mentioned all through the document are operated by “APMCs (Agriculture Produce Market Committee)”. The APMCs grant commission agents’ licence to run stores and agriculturalists can get products, weigh it, and auction it to dealers. Sellers require operating licences, In principle, agriculturalists can approach any seller and farmers should get best price for an auctions conducted by the agent, but this is generally desecrated by cartelization between brokers and traders, since agriculturalists might have binding contacts with the broker, a more suitable solution, proposes enhancing state structures to control markets. Indeed, in 2003, the “APMC “The law aimed at abolishing commission agents was proposed, to offer professional facilities to determine the condition and capacity of the goods have being offered, to permit cross-mandi exchange, etc.,

Bibliography

Nidhi Aggarwal, Sargam Jain, and Sudha Narayanan. 2016. The Long road to transformation of agricultural markets in India: Lessons from Karnataka. Technical Report. Indira Gandhi Institute of Development Research, Mumbai, India.

Bina Agarwal. 2017. The seeds of discontent. Retrieved from The Indian Express https://goo.gl/6wxeF1.

Kabir Agarwal. 2018. Why MSP at 1.5 Times Cost Is Another Empty Promise for Farmers. Retrieved from The Wire https://goo.gl/1pcGtR.

Aggarwal, N., & Narayanan, S. (2017). Impact of India's Demonetization on Domestic Agricultural Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3066042

Anderson, J., & Ahmed, W. (2016). *Smallholder Diaries: Building the evidence base with farming families in Mozambique, Tanzania, and Pakistan*. *2*. http://www.cgap.org/sites/default/files/CGAP_Persp2_Apr2016-R.pdf

Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, *7*(2), 1–21. https://doi.org/10.4018/ijoris.2016040101

Atlas. (n.d.). *Weather data*. https://www.weather-ind.com/

Azad, A. S. M. S., Azmat, S., Fang, V., & Edirisuriya, P. (2014). Unchecked manipulations, price-volume relationship and market efficiency: Evidence from emerging markets. *Research in International Business and Finance*, *30*(1), 51–71. https://doi.org/10.1016/j.ribaf.2013.05.003

Banik, N. (2017). Farmer Suicides in India and the Weather God. *Procedia Computer Science*, *122*, 10–16. https://doi.org/10.1016/j.procs.2017.11.335

Barucci, E. (2006). Market Abuse Detection : a Methodology Based on Financial Time Series. *Statistica Applicata*, *18*(4), 559.

Chakraborty, S., Venkataraman, A., Jagabathula, S., & Subramanian, L. (2016). Predicting socio-economic indicators using news events. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 1455–1464. https://doi.org/10.1145/2939672.2939817

Chand, R. (2016). E-platform for national agricultural market. *Economic and Political Weekly*, *51*(28), 15–18.

Ghosal, S. (2018). *Potato prices likely to double SECTIONS*. https://economictimes.indiatimes.com/markets/commodities/news/potato-prices-likely-to-double/articleshow/64299435.cms?from=mdr

Graves, A. (n.d.). *Neural Networks*.

Haines et al, 2019, goleman, daniel; boyatzis, Richard; Mckee, A., Haines et al, 2019, goleman, daniel; boyatzis, Richard; Mckee, A., Haines et al, 2019, & goleman, daniel; boyatzis, Richard; Mckee, A. (2019). *otexts*. Journal of Chemical Information and Modeling. https://otexts.com/fpp2/arima.html

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

India Today Web Desk. (2019). *Onion prices skyrocket again, close to Rs 100/kg in some states*. https://www.indiatoday.in/business/story/onion-prices-skyrocket-close-to-rs-100-kg-in-some-states-1615771-2019-11-05

Jha, G. K., & Sinha, K. (2013). Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System. *Agricultural Economics Research Review*, *26*(2), 229--239. http://ageconsearch.umn.edu/bitstream/162150/2/8-GK-Jha.pdf

Li, G. Q., Xu, S. W., & Li, Z. M. (2010). Short-term price forecasting for agro-products using artificial neural networks. *Agriculture and Agricultural Science Procedia*, *1*, 278–287. https://doi.org/10.1016/j.aaspro.2010.09.035

Madaan, L., Sharma, A., Khandelwal, P., Goel, S., Singla, P., & Seth, A. (2019). Price forecasting & anomaly detection for agricultural commodities in India. *COMPASS 2019 - Proceedings of the 2019 Conference on Computing and Sustainable Societies*, 52–64. https://doi.org/10.1145/3314344.3332488

Monsoon puts tomatoes on fire, likely to remain pricey for another 2 weeks. (2017). *No Title*. https://www.indiatoday.in/india/story/tomatoes-prices-monsoon-rain-1027211-2017-07-31

National horticultural board. (n.d.). *NHB*. http://nhb.gov.in/OnlineClient/MonthlyPriceAndArrivalReport.aspx?enc=3ZOO8K5CzcdC/Yq6HcdIxJ4o5jmAcGG5QGUXX3BlAP4=

Parija, P. (2019). *The great onion crisis of 2019 may be coming to an end*. https://www.livemint.com/news/india/the-great-onion-crisis-of-2019-may-be-coming-to-an-end-11576812495855.html

Potatonewstoday. (2020). *Supply shortage, rise in demand push potato prices in India up by*

40% in major cities. https://www.potatonewstoday.com/2020/07/28/supply-shortage-rise-in-demand-push-potato-prices-in-india-up-by-40-in-major-cities/

Appendix

Plagiarism Report[1]

Indian Agriculture Commodity Price Forecasting & Anomaly Discovery, A Case Study on "Onion, Potato and Tomato" commodities

by Mutturaj Uppaladinni

Submission date: 24-Oct-2020 04:38PM (UTC+0530)
Submission ID: 1425127987
File name: Analytics_Capstone_Project_Report_Mutturaj_UPPALADINNI.docx (2.5M)
Word count: 5981
Character count: 33795

[1] Turnitin report to be attached from the University.

Indian Agriculture Commodity Price Forecasting & Anomaly Discovery, A Case Study on “Onion, Potato and Tomato” commodities

ORIGINALITY REPORT

12%	8%	3%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.caedes.net Internet Source	3%
2	Lovish Madaan, Ankur Sharma, Praneet Khandelwal, Shivank Goel, Parag Singla, Aaditeshwar Seth. "Price forecasting & anomaly detection for agricultural commodities in India", Proceedings of the Conference on Computing & Sustainable Societies - COMPASS 19, 2019 Publication	2%
3	Submitted to Symbiosis International University Student Paper	1%
4	Submitted to Sogang University Student Paper	1%
5	ideas.repec.org Internet Source	1%
6	www.termpaperwarehouse.com Internet Source	1%

7	course1.winona.edu Internet Source	<1 %
8	Submitted to northcap Student Paper	<1 %
9	Submitted to hotelschool Student Paper	<1 %
10	eprints.utem.edu.my Internet Source	<1 %
11	Submitted to University of Reading Student Paper	<1 %
12	jacobbradjohnson.com Internet Source	<1 %
13	Submitted to Regional Centre for Biotechnology Student Paper	<1 %
14	www.potatonewstoday.com Internet Source	<1 %
15	Rajib Bhattacharyya. "chapter 7 Demonetization and Movement Towards Digital Cashless Indian Economy", IGI Global, 2021 Publication	<1 %
16	businesseconomics.in Internet Source	<1 %
17	www.inmybangalore.com Internet Source	<1 %

18	Submitted to Lovely Professional University Student Paper	<1 %
19	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	<1 %
20	Submitted to Thapar University, Patiala Student Paper	<1 %

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

Publications in a Journal/Conference Presented/White Paper[2]

NA

Any Additional Details

[2] URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided.