

Indian commodity market price comparative study of forecasting methods

A case study on onion, potato and tomato



Presented by:
Suresha H P
SRN:R19MBA10
Batch: BA05

Mentor :
Krishna Kumar Tiwari

Agenda

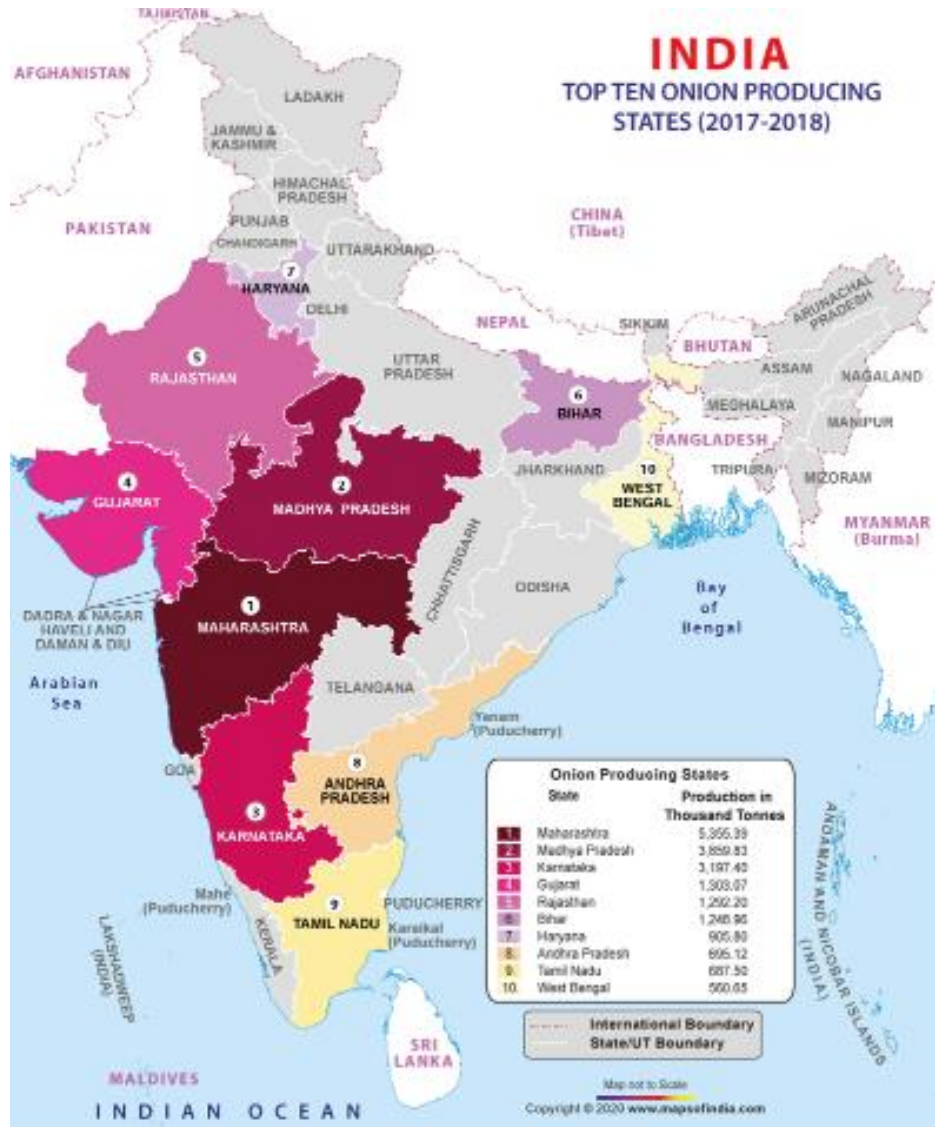
1. Introduction
2. Project Methodology
3. Business Understanding
4. Data Understanding
5. Data Preparation
6. Data Modeling
7. Data Evaluation
8. Deployment
9. Analysis and Results
10. Conclusions and Recommendations for future work
11. Publications in a Conference

1. Introduction

- Onion, tomato and potato are an important agricultural commodity
- Retail Price fluctuations is an important area of study in forecasting
- comparison of results will help in selecting the suitable algorithm's to predict retail prices
- I studied several mandi's across India based on arrival volumes and diesel
- Data considered from different Mandi's between Jan 2015 to July 2020

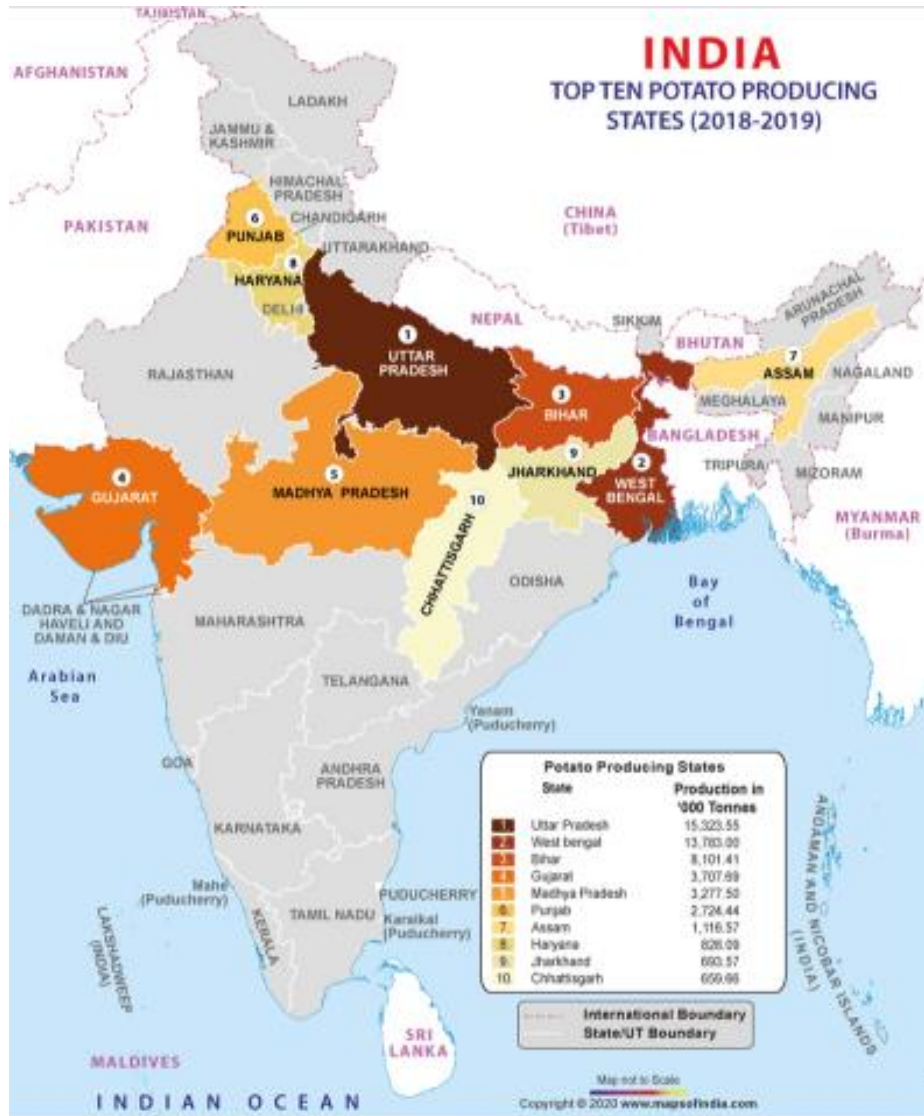


1. Introduction – Onion agricultural Commodity in India



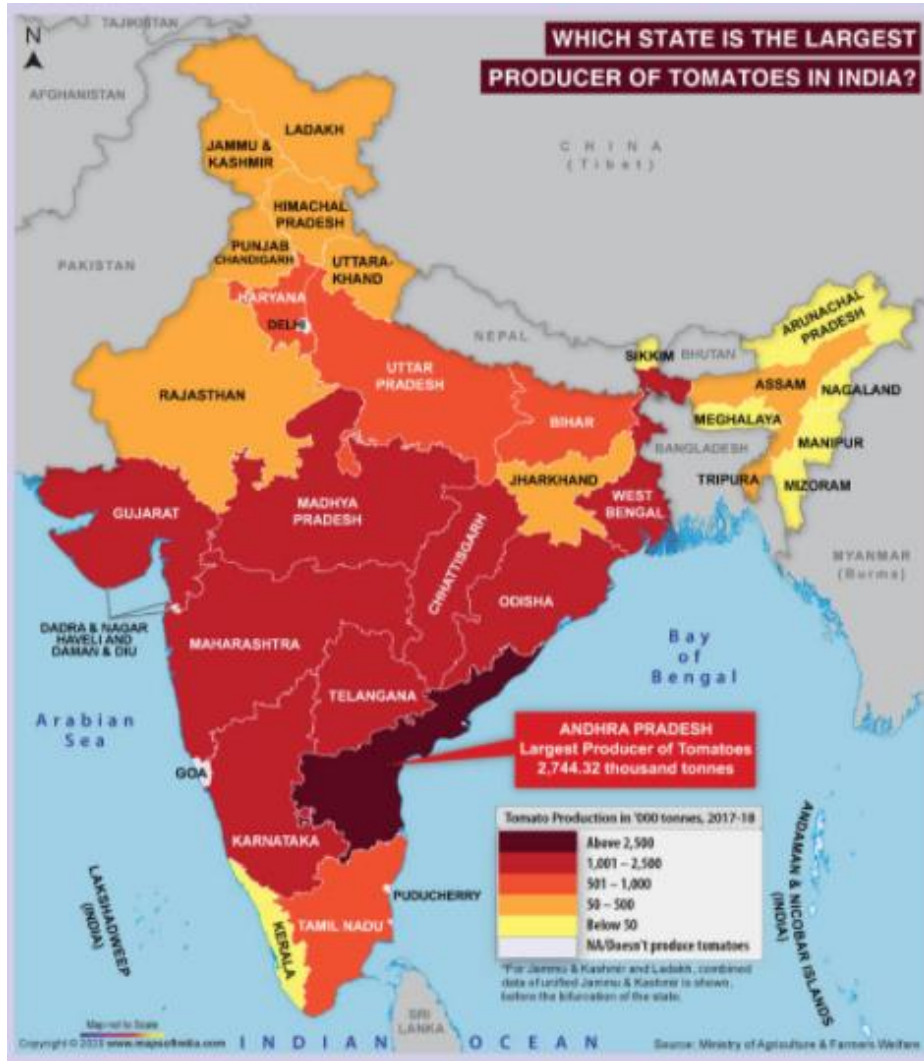
- **Onion** (*Allium cepa*) is one of the most important commercial crop
- Onion crop is grown in about 1.20-million-hectare area
- Annual production of 19.40 million tons
- Top 3 onion producing states are Maharashtra, Madhya Pradesh and Karnataka
- Top 3 mandi's for onion are Bengaluru, Mumbai and Hyderabad based on arrival volume in metric tonnes

1. Introduction – Potato agricultural Commodity in India



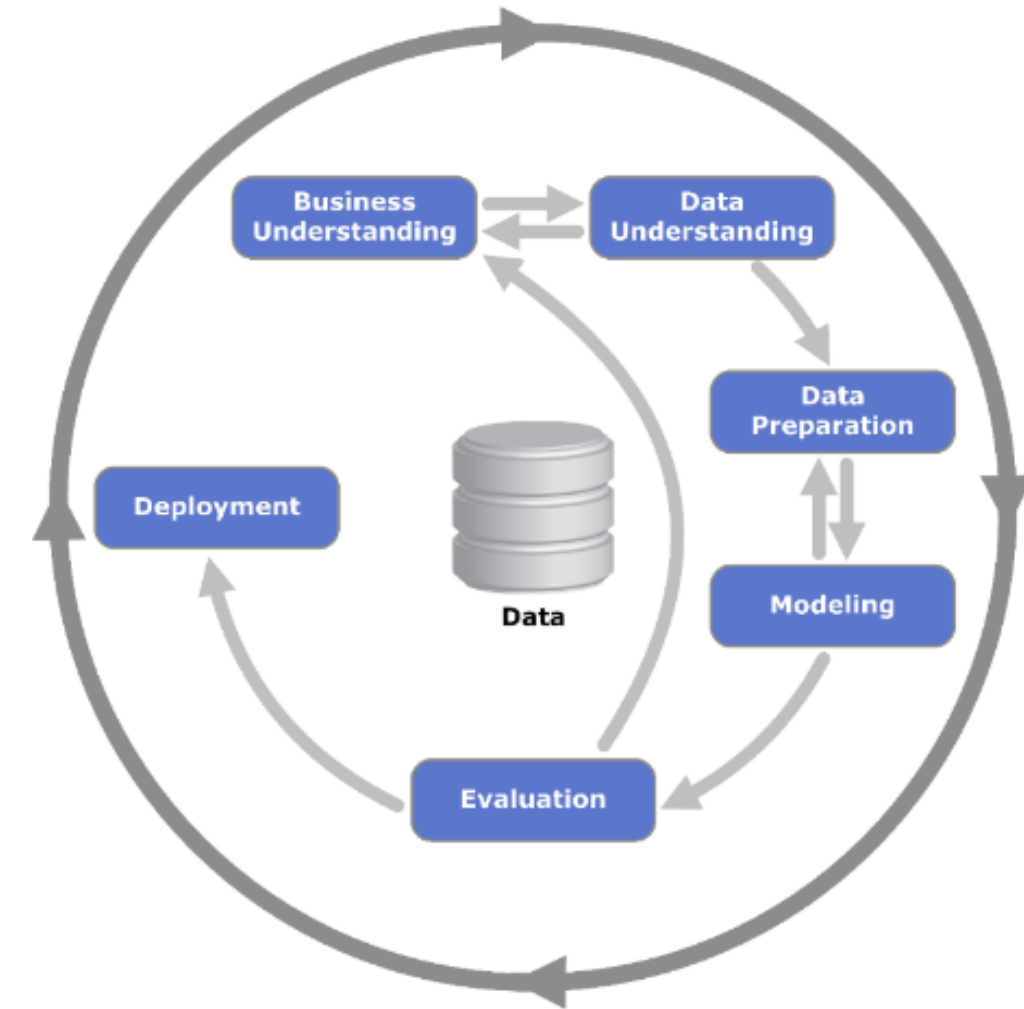
- **Potato** (*Solanum tuberosum*) is the most important food crop
- Potato crop is grown in about 21.8 lakh hectares
- Annual production of 52.5 million tons
- Top 3 potato producing states are Uttar Pradesh, west Bengal and Bihar
- Top 3 mandi's for Potato are Mumbai, Delhi and Bengaluru based on arrival volume in metric tonnes

1. Introduction – Tomato agricultural Commodity in India



- **Tomato** (*Lycopersicon esculentum*) is the most important food crop
- Tomato crop is grown in about 3,50,000 hectares
- Annual production of 53,00,000 tons
- Top 3 Tomato producing states are Andhra Pradesh, Madhya Pradesh and Karnataka
- Top 3 mandi's for Tomato are Delhi, Hyderabad and Mumbai based on arrival volume in metric tonnes

2. Project Methodology



The Methodology followed in this project is The “**CR**oss Industry **S**tandard **P**rocess for **D**ata **M**ining (*CRISP-DM*) is a process model with six phases that naturally describes the data science life cycle:

- Business understanding
- Data understanding
- Data preparation
- Modelling
- Evaluation
- Deployment

3. Business Understanding



- India ranks second worldwide in farm outputs
- As per 2018, agriculture employed more than 50% of the Indian work forces
- Farming contributed 17–18% to India's GDP
- Agriculture is primary source of livelihood for 58% population in India
- Onion 2415.75 thousand tons is exported from India
- Onion exports which outputs value of 3,10,650.09 Rs. Lakhs (in 2017)
- Tomato is short duration crop and gives high yield
- Potato is more suitable for subtropical conditions in India

4. Data Understanding

Date	AHMEDABAD_WholesalePrice	AHMEDABAD_RetailPrice
1/1/2015	1359	2317
2/1/2015	1569	2300
3/1/2015	1389	2285
4/1/2015	1099	2300
5/1/2015	1140	2300
6/1/2015	1582	2486
7/1/2015	1957	3375
8/1/2015	3838	5571
9/1/2015	4180	6762
10/1/2015	2757	5958
11/1/2015	1881	4023
12/1/2015	1283	2889

- Agricultural Data from National horticultural website and Diesel price from gatti website
- Downloaded **onion**, **tomato** and **potato** data from January 2015 till July 2020
- Combined data for commodity prices, volume and diesel prices into onion, tomato and potato using panda's data frames, respectively
- Data is having missing values <15%, total 67 rows of data and for different mandis for onion, tomato and potato respectively.

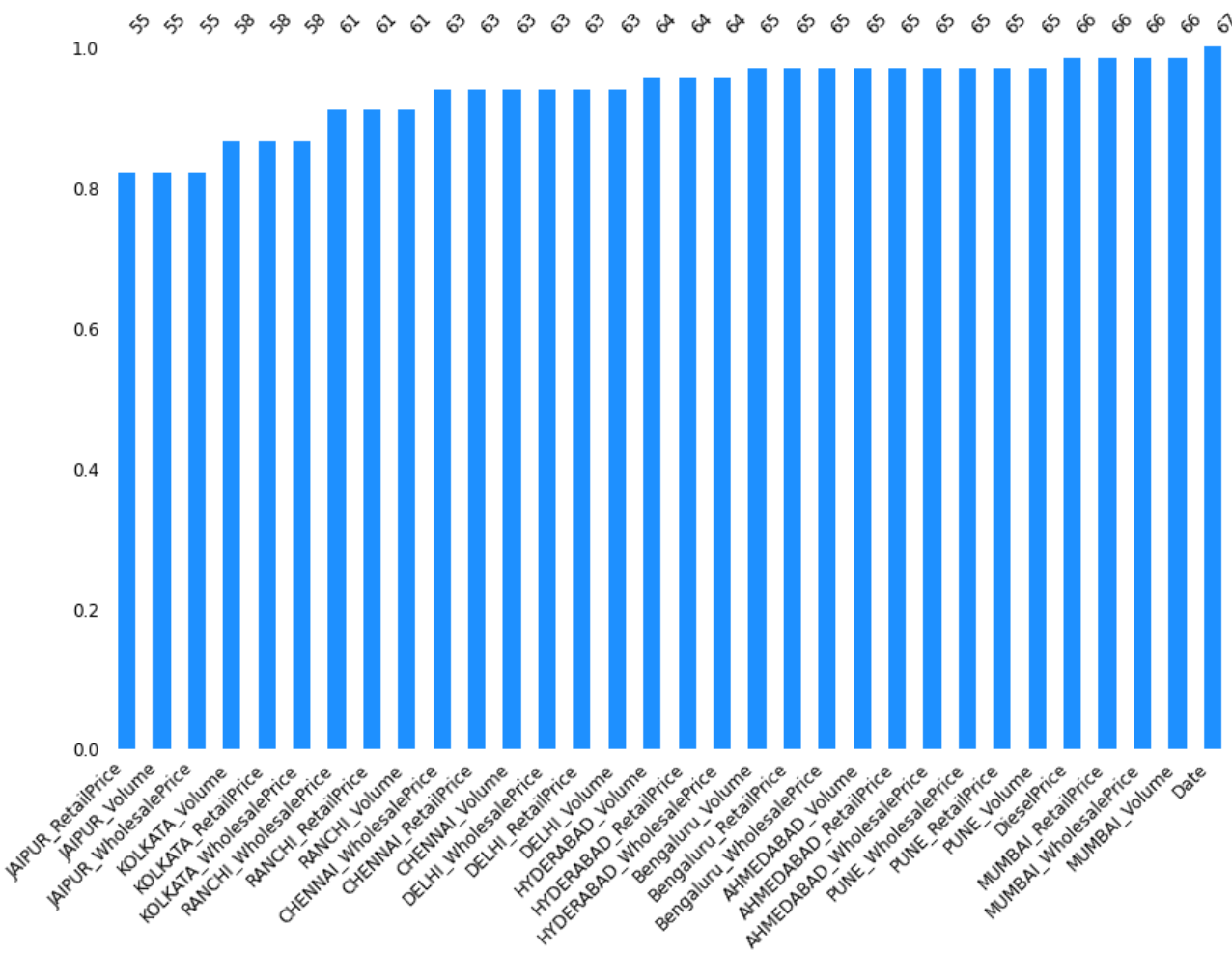
5. Data Preparation

Following are the activities during data preparation:

- Missing value Imputation
- Identify Top 10 mandis of India for Onion, tomato and potato based on volume
- Retail price trend analysis for onion , tomato and potato
- Data Distribution
- Anomalies and Outliers

5. Data Preparation

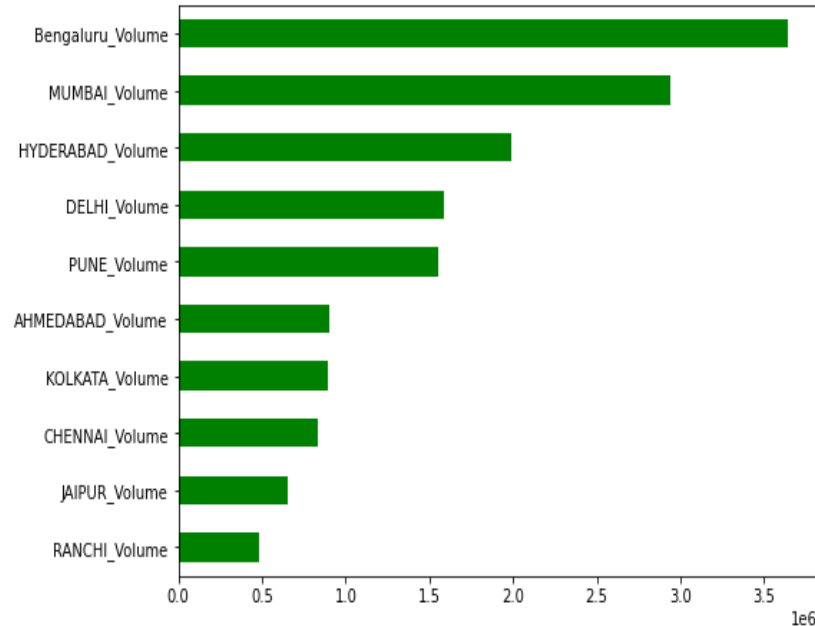
● Missing value Imputation



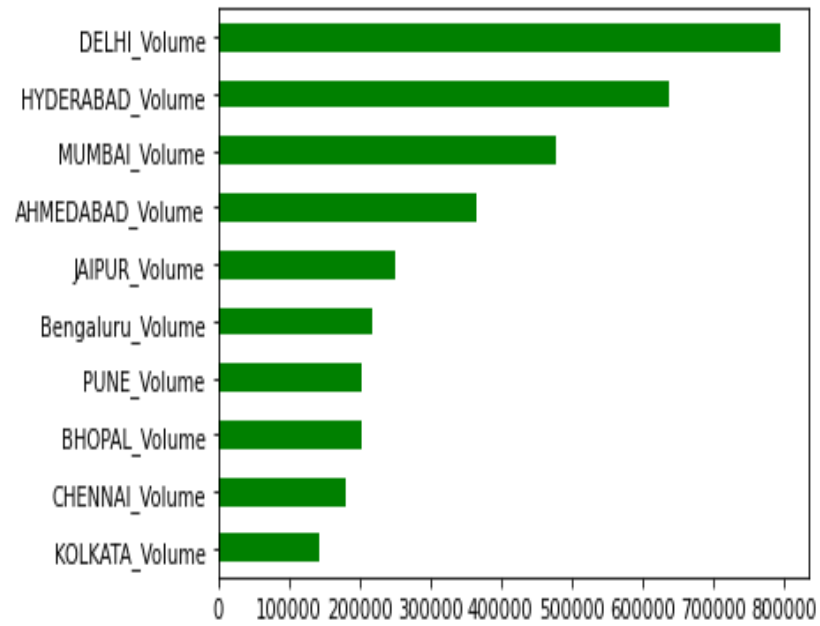
- Observed missing value in onion, tomato, potato and diesel data.
- missing value imputation using replacing missing value by mean

5. Data Preparation

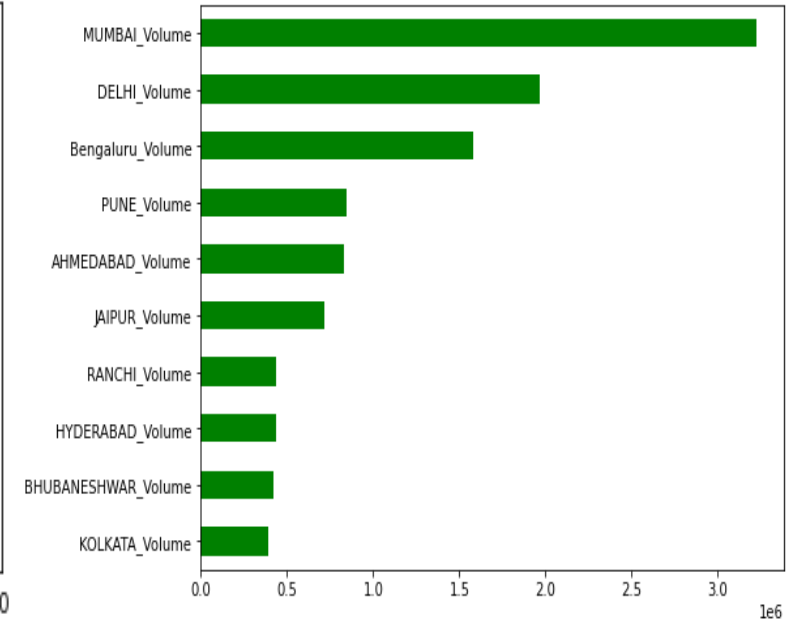
- Top 10 mandis of India for Onion, tomato and potato based on volume



Onion



Tomato

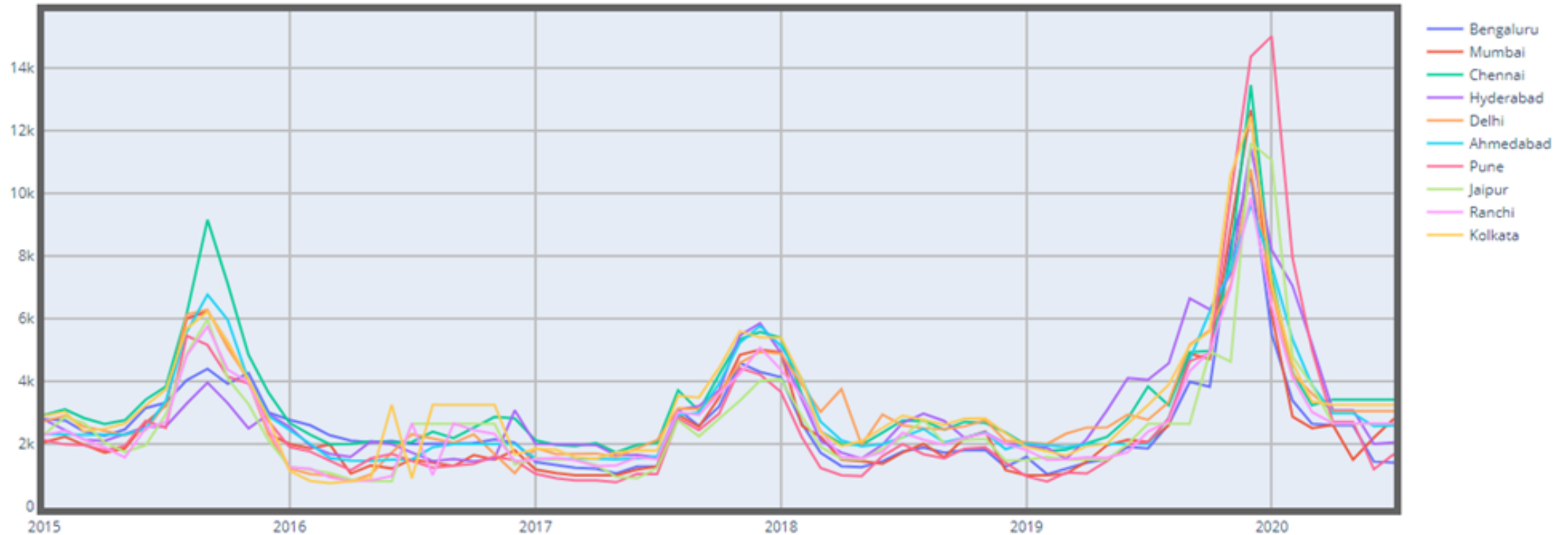


Potato

Top mandis for onion is Bengaluru, Tomato is Delhi and for potato is Mumbai

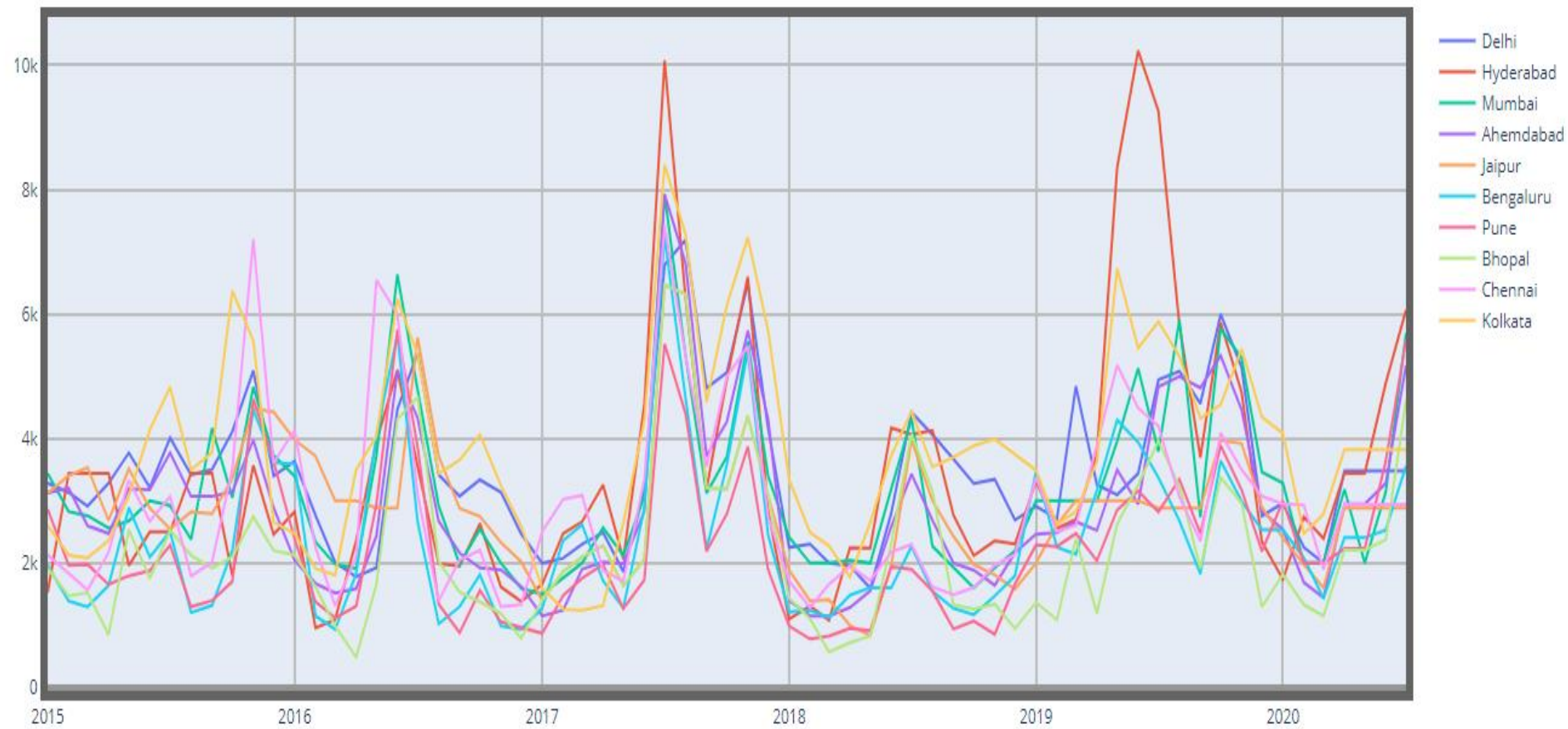
5. Data Preparation

- Trend chart of 10 mandis of India for Onion retail price



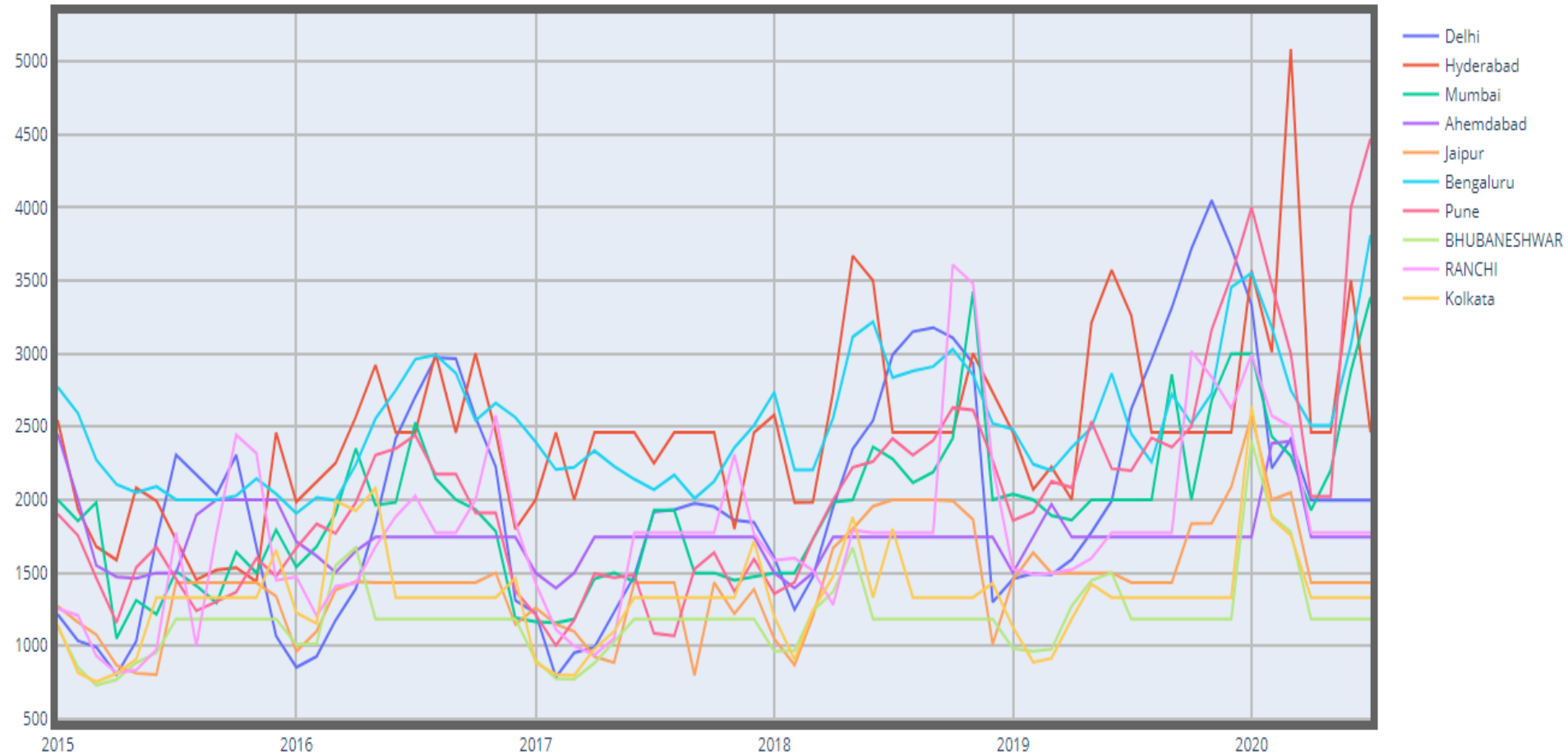
5. Data Preparation

- Trend chart of 10 mandis of India for Tomato retail price



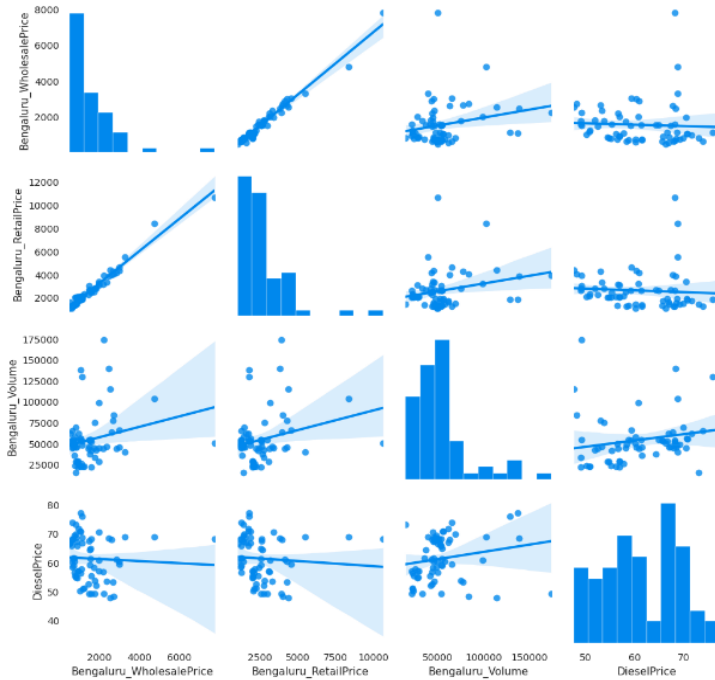
5. Data Preparation

- Trend chart of 10 mandis of India for Potato retail price

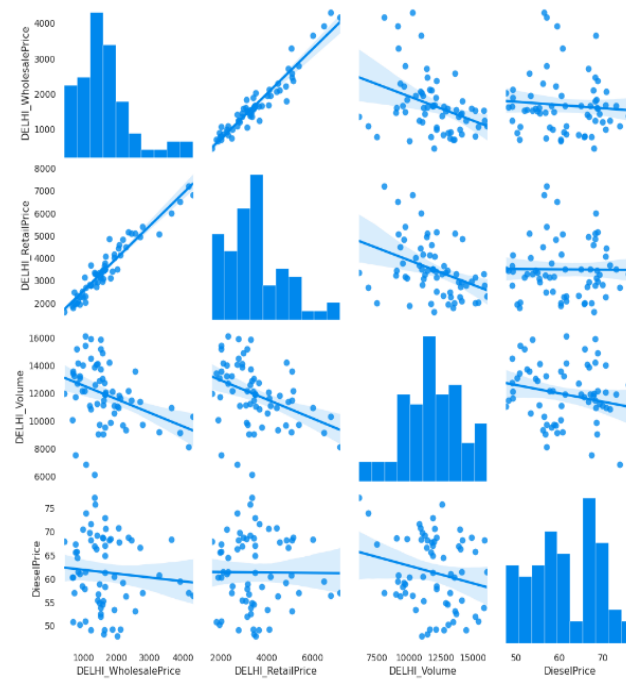


5. Data Preparation

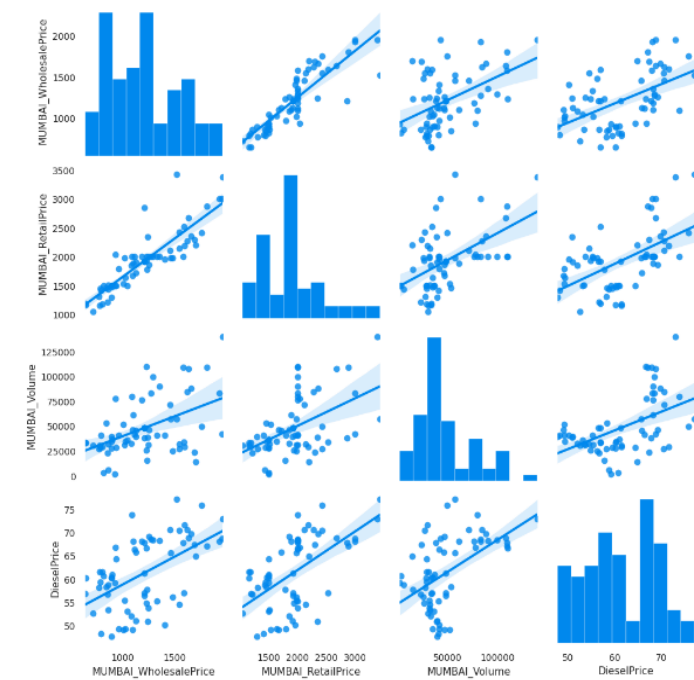
- Data Distribution



Bengaluru – Onion Pair plot



Delhi – Tomato Pair plot

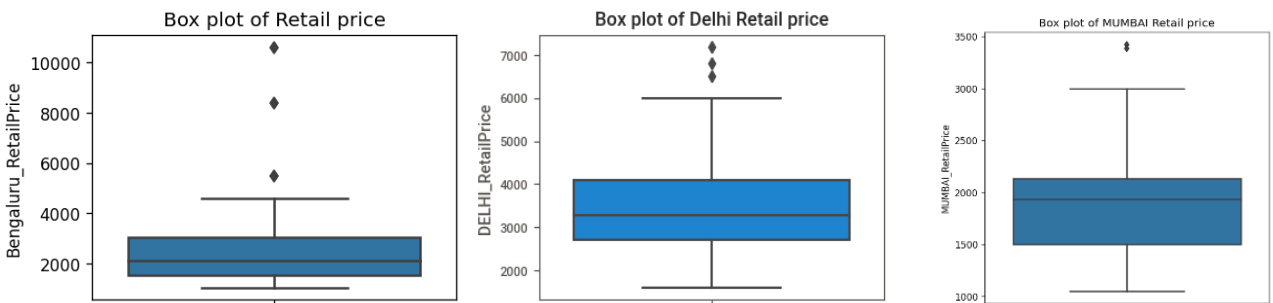


Mumbai – Potato Pair plot

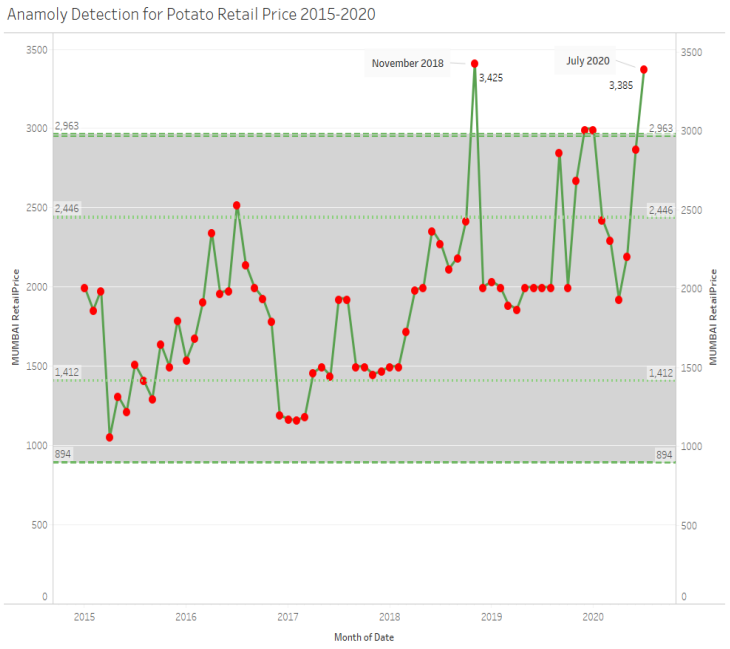
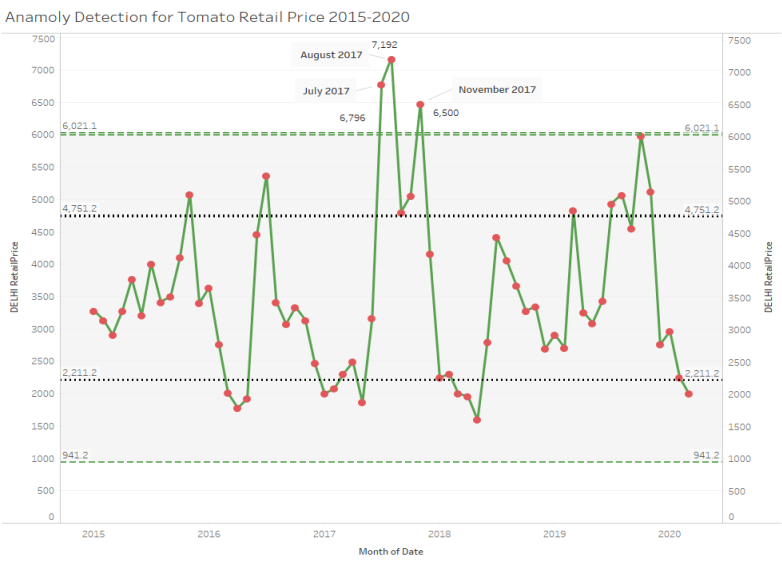
Good Correlation observed between retail price, volume and diesel price for onion, Tomato and for potato

5. Data Preparation

• Anomalies and Outliers



- Identified Anomalies and outliers in onion, tomato and potato



5. Data Preparation - Summary

Performed the following activities and feature engineering steps:

Data Cleaning

Identified top 10 mandi's for onion, tomato and potato

Missing value imputation for all the missing value by mean

Feature Engineering

Understanding of data distribution

Understanding of trend for Retail prices of onion, potato and tomato

Identified anomalies and outliers in the top mandi's – Bengaluru, Delhi and Mumbai for onion, tomato and potato respectively

Output

Clean dataset for Bengaluru (Onion), Delhi (Tomato) and Potato (Mumbai) exported to do data modeling and further machine learning activities

6. Data Modeling

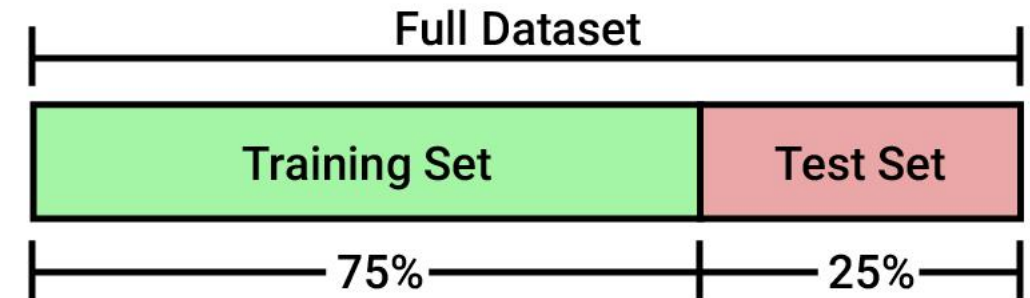
Performed the following activities:

Statistical tests

1. Testing Causation using Granger's Causality Test
2. Cointegration Test
3. Augmented Dickey-Fuller Test (ADF Test)
4. Durbin Watson test
5. Auto correlation and seasonality function analysis

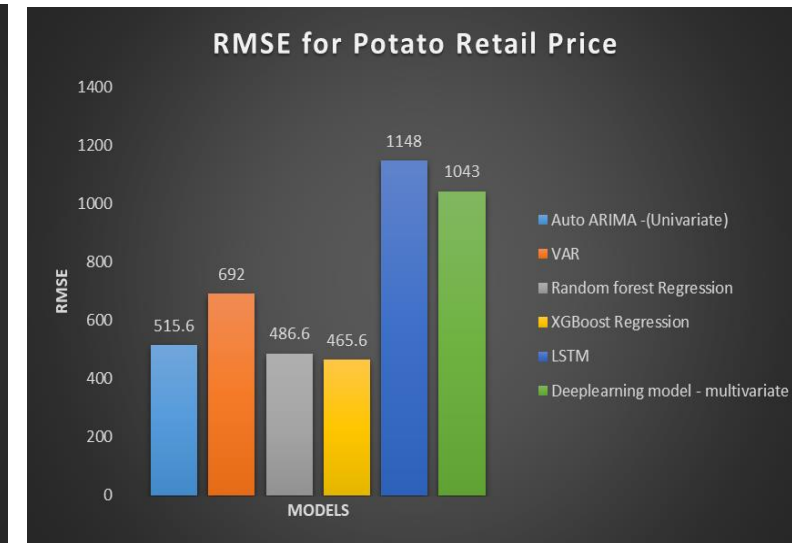
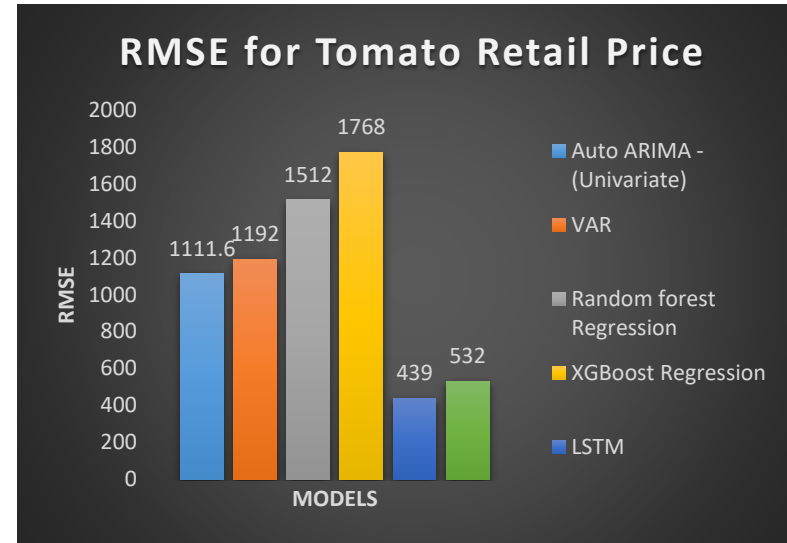
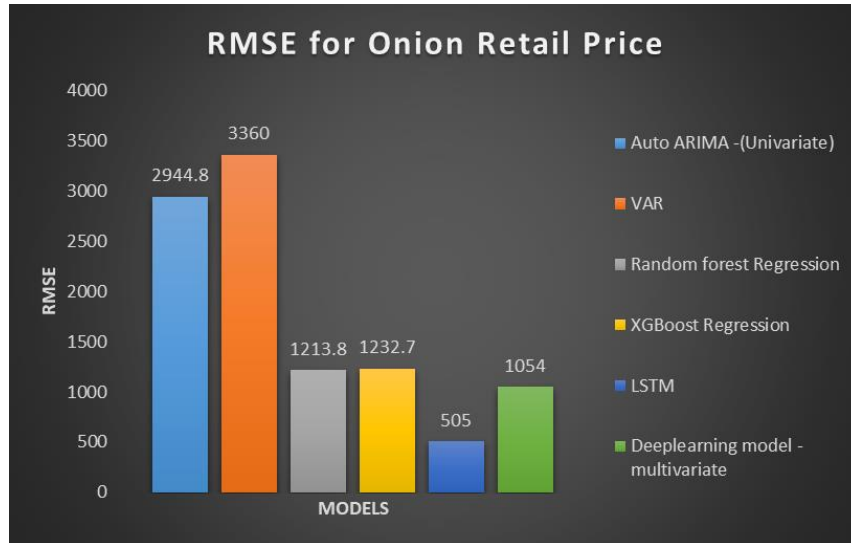
Modeling to predict Retail prices of Onion, tomato and Potato

1. Auto ARIMA (Univariate)
2. VAR
3. Random forest regression
4. XGBoost Regression
5. LSTM
6. Deep learning model (Multivariate)



7. Data Evaluation

RMSE – Root Mean Square Error is the key metrics used to evaluate and compare different prediction models for onion, tomato and potato retail price



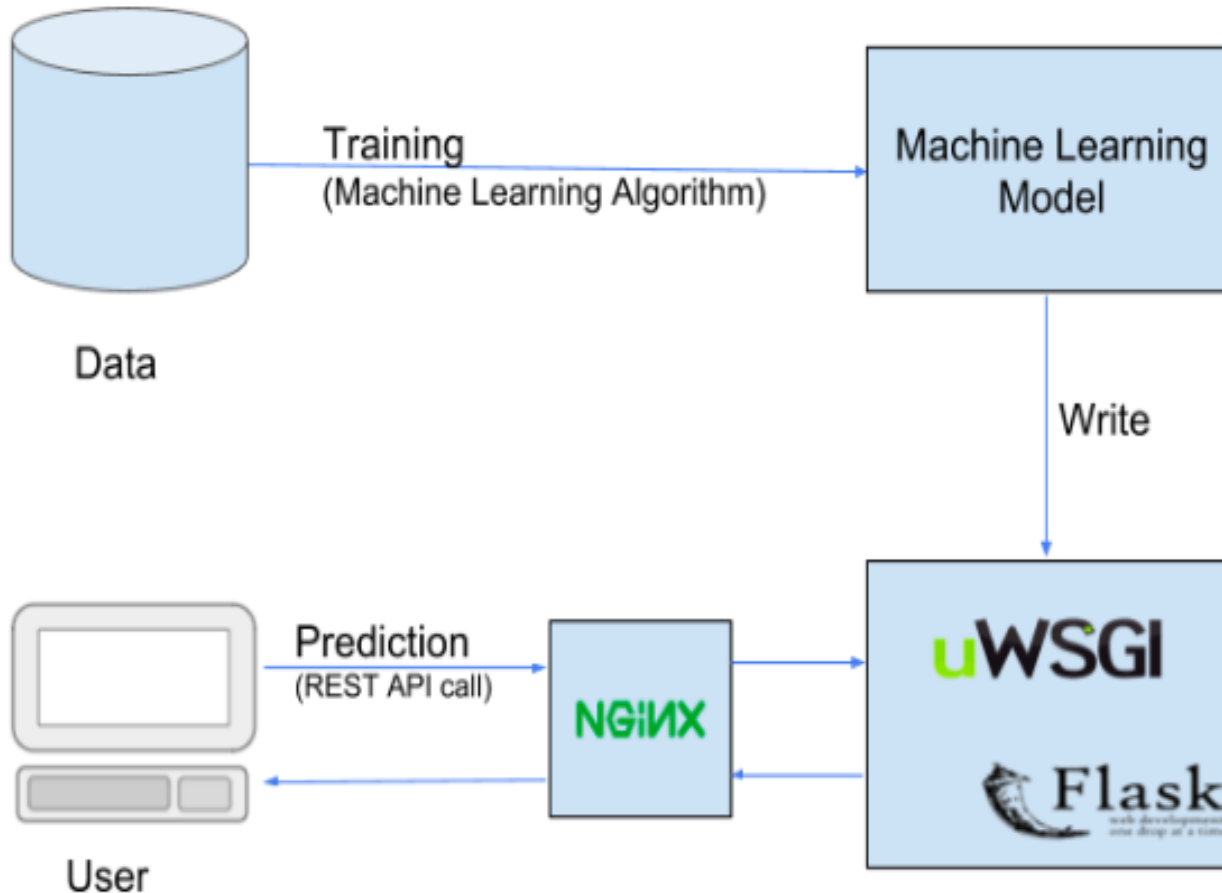
Best 3 Models for onion, tomato and potato retail price predictions are as follows:

Onion Retail Price prediction: LSTM, Deep learning and Random forest

Tomato Retail Price prediction: LSTM, Deep learning and Random forest

Potato Retail Price prediction: XG Boost, Random forest and Auto ARIMA (Univariate)

7. Deployment (Proposed for future work)



Deployment activities from this current project will carry it forward as a continuous activity in the next project.

8. Analysis and Results

Observed linear relationship between wholesale and retail price of the onion, tomato and potato.
I took a decision to drop whole sale price variable from study to avoid overfitting.

There are few major anomalies due to key events are as follows:

Onion Anomalies identified from the data during this project which happened during November 2019 and December 2019. “Heavy unseasonal rainfall has locked latest onion produce in wet fields throughout onion growing states of Maharashtra, Karnataka and Telangana. As a ripple effect, onion prices increased everywhere in the country” (Times, 2019)

Tomato Anomalies identified from the data during this project which happened during July 2017, August 2017 and November 2017. “Between July 2017 to November 2017 there was huge reduction in arrival of volume in different mandi’s in India” (Hindu, 2019)

Potato Anomalies identified from the data during this project which happened during Nov 2018 and July 2020.
“Potato price increased continuously from July till December 2018” (Express, 2019)
“Supply shortage, rise in demand push potato prices in India up by 40% in major cities” (News, 2019)

9. Conclusion and Recommendation for future work

- Agricultural commodity onion, tomato and potato sales flows through the mandis in India.
- This study help to understand the geological location of different mandi's
- Understood the relationship of retail price with wholesale price, volume and diesel price
- Identified Multivariate models is performing better than conventional univariate times series models.
- Potential opportunity to continue this work to further do deployment, rest API, web apps and also to include other variable which is affecting Retail prices of agricultural commodities
- My key contribution in this Project is to emphasize on the role that of multivariate Machine learning prediction and other forecasting methods.
- **GitHub code repository :**

<https://github.com/hpsuresh12345/Agricultural-commodity-forecasting-methods-onion-potato-and-tomato>

10. Publication in Conference

Selected for CST 2020, Zurich Switzerland

Dear Author,

First of all, thank you very much for submitting your paper to CST 2020 to be held in Zurich, Switzerland, November 21 ~ 22, 2020. Based upon the reviewer's reports, we are pleased to inform you that your paper has been **ACCEPTED** by the conference and will be included in the proceedings published by Computer Science Conference Proceedings in Computer Science & Information Technology (CS & IT) series.

Congratulations on your excellent work!

In order to achieve the highest quality proceedings, we urge you to carefully consider the reviewer's comments, if any, when preparing the final version of your paper.

1. Please read the following Information carefully to prepare a final manuscript of your paper <https://cst2020.org/submission/template.doc>.

The maximum number of pages without extra payment is 20 (CCSP format). For each extra page you have pay 50 USD additionally.

2. Submit your final camera-ready version of paper (.doc version+ .pdf version) and filled CR form cst@cst2020.org (or) cst.conf@yahoo.com

3. Final Manuscript Submission Details:

a) When submitting your final manuscript, please ensure that you send us all source files such as .doc and pdf.



