DATA SCIENCE

PROJECT REPORT

(Project Semester January-April 2025)

**Exploratory Data Analysis on Agricultural Commodity Prices in India**

Submitted by

Amit Krishna

Registration No: 12326830

Programme and Section: DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING KM005

Course Code: INT375

Under the Guidance of

Maneet Kaur

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

# DECLARATION

I, Amit Krishna, student of DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 2025-04-11

Signature Amit Krishna  
Registration No. 12326830  
  
Name of the student: Amit Krishna

# CERTIFICATE

This is to certify that Mr. Amit Krishna bearing Registration No. 12326830 has completed INT375 project titled, “Exploratory Data Analysis on Agricultural Commodity Prices in India” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor  
Maneet Kaur  
  
Designation of the Supervisor  
School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab.

Date: 2025-04-11

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentor Ms. Maneet Kaur for her valuable guidance throughout this project. I also thank Lovely Professional University for providing the necessary resources and support.

# TABLE OF CONTENTS

1. 1. Introduction
2. 2. Source of Dataset
3. 3. EDA Process
4. 4. Analysis on Dataset

* i. General Description
* ii. Specific Requirements, Functions and Formulas
* iii. Analysis Results
* iv. Visualizations

1. 5. Conclusion
2. 6. Future Scope
3. 7. References

# 1. Introduction

This Exploratory Data Analysis (EDA) project aims to dissect and understand agricultural price patterns using a large and diverse dataset containing real-world market data across various Indian states and markets.

The dataset includes attributes such as **Commodity**, **State**, **Market**, **Variety**, **Grade**, **Minimum Price**, **Maximum Price**, **Modal Price**, and **Arrival Date.**

This report explores relationships between different price points, identifies regions and commodities with the highest and lowest price ranges, and highlights pricing stability

You may visit my LinkedIn using link: <https://tinyurl.com/EDAagri>

# 2. Source of Dataset

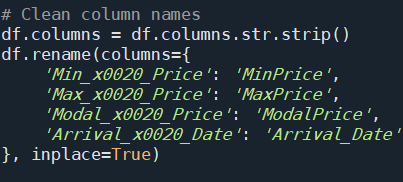
Dataset sourced from data.gov.in containing fields like Commodity, State, Market, Variety, Grade, Min Price, Max Price, Modal Price, and Arrival Date

Link: <https://www.data.gov.in/resource/current-daily-price-various-commodities-various-markets-mandi>

# 3. EDA Process

This phase involves cleaning and preparing the dataset for analysis. The steps included:

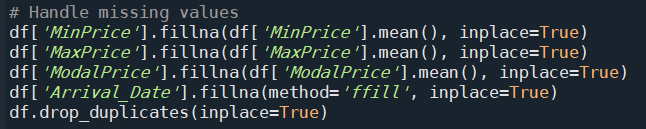
**Column Renaming**: To make columns readable and accessible for analysis.



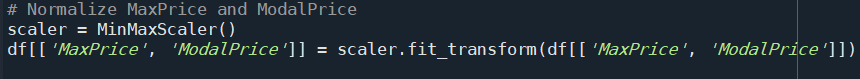
**Date Conversion**: Ensuring date columns are in datetime format for time-based analysis.



**Handling Missing Values**: Filling or removing null values to maintain data integrity.



**Normalization**: Scaling price columns to a consistent range.

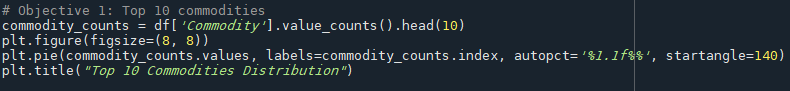


# 4. Analysis on Dataset

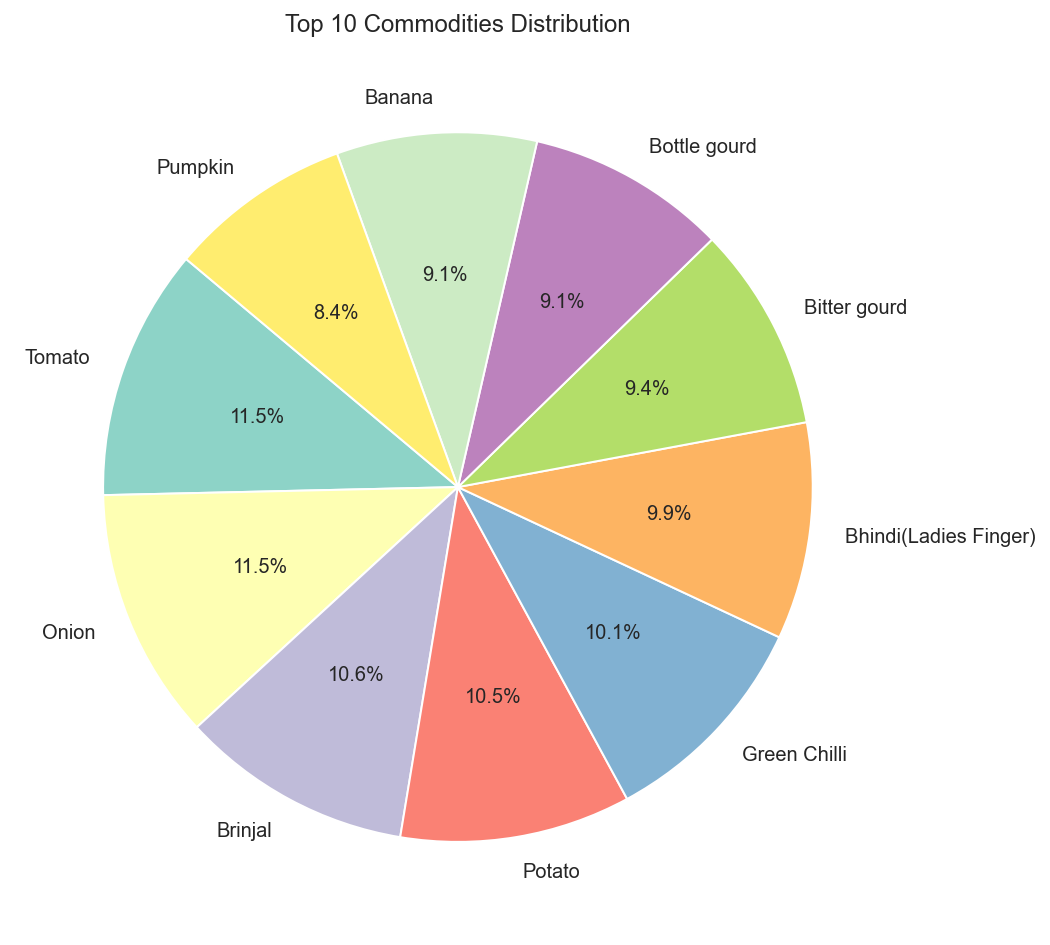
**Objective 1: Top 10 Commodities by Count**

i. **General Description**  
Identify the most frequently occurring commodities in the dataset.

ii. **Specific Requirements, Functions and Formulas**  
Function used: value\_counts(), sns.barplot()



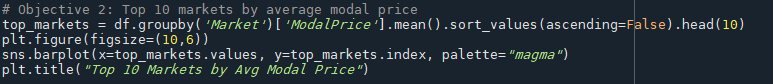
iii. **Analysis Results**  
Onion, Tomato, Potato, and Green Chilli are among the most recorded commodities.

iv. **Visualizations**  


**Objective 2: Top 10 Markets by Avg Modal Price**

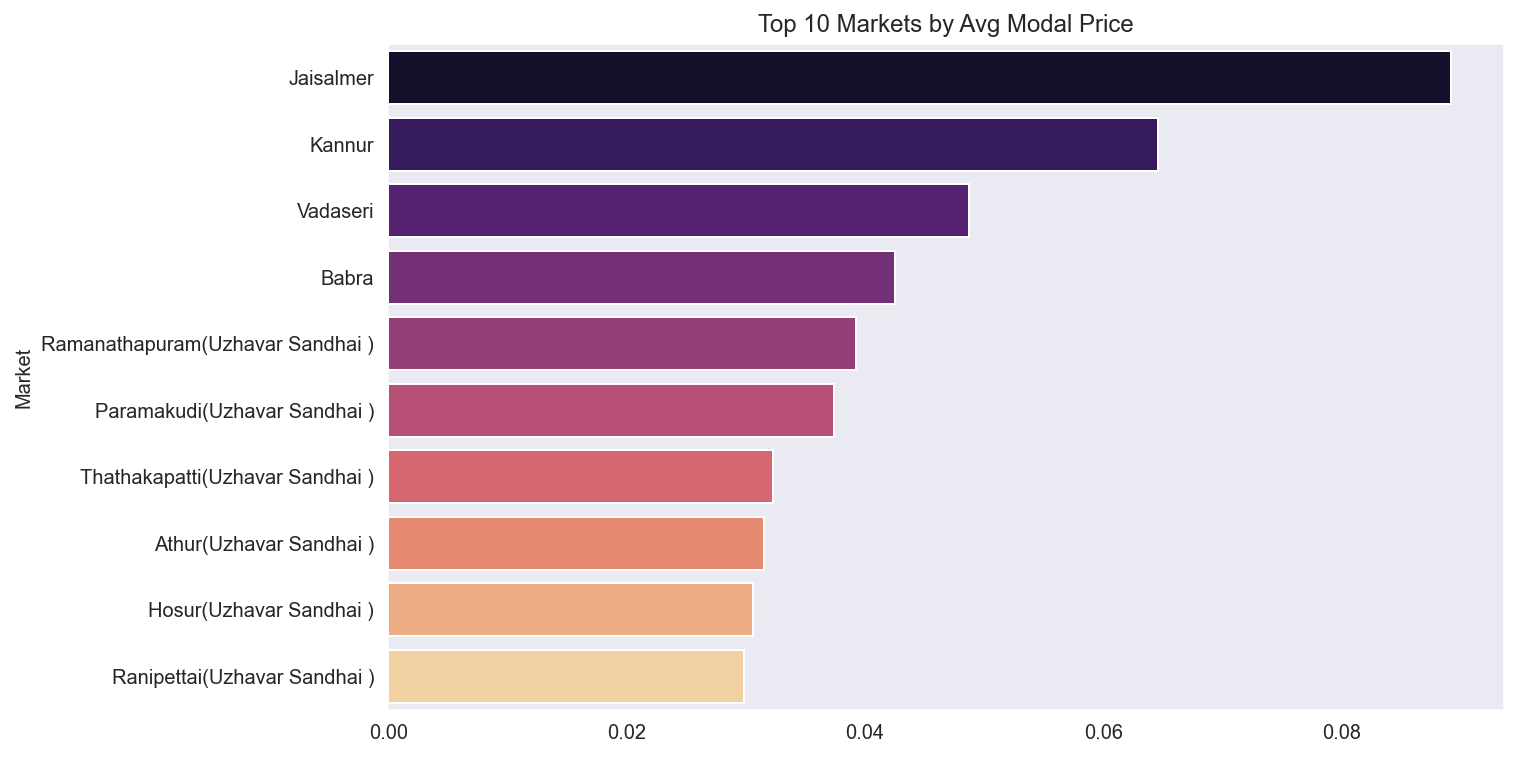
i. **General Description**  
Analyzing markets that report the highest average modal price.

ii. **Specific Requirements, Functions and Formulas**  
Function used: groupby(), mean(), nlargest(), sns.barplot()



iii. **Analysis Results**  
A few markets consistently offer higher average prices.

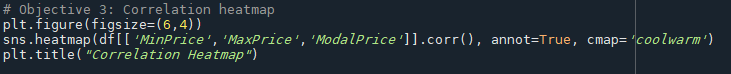
iv. **Visualizations**



**Objective 3: Correlation Between Prices**

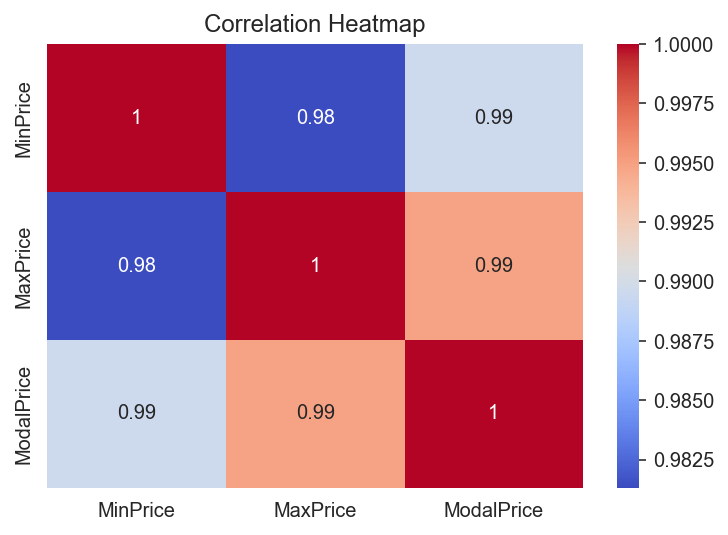
i. **General Description**  
Understand how Min, Max, and Modal prices are correlated.

ii. **Specific Requirements, Functions and Formulas**  
Function used: corr(), sns.heatmap()



iii. **Analysis Results**  
Modal price is positively correlated with Max price.

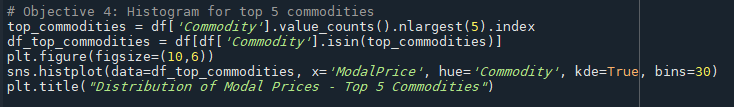
iv. **Visualizations**



**Objective 4: Histogram of Modal Prices for Top Commodities**

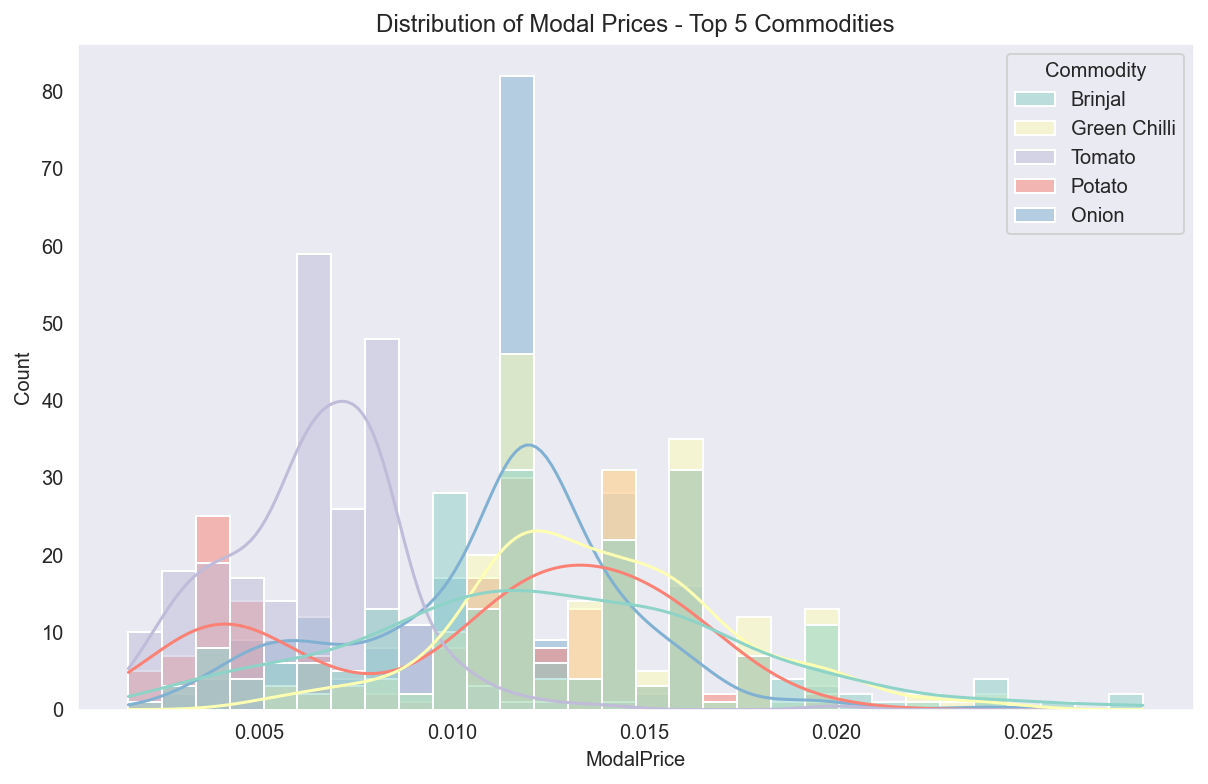
i. **General Description**  
Visualize distribution of modal prices for most frequent commodities.

ii. **Specific Requirements, Functions and Formulas**  
Function used: histplot(), value\_counts(), isin()



iii. **Analysis Results**  
Different commodities show different modal price distributions. With peak going to onion.

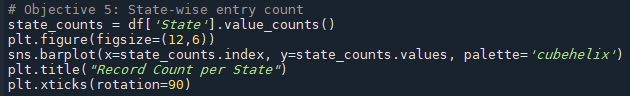
iv. **Visualizations**



**Objective 5: Entry Count per State**

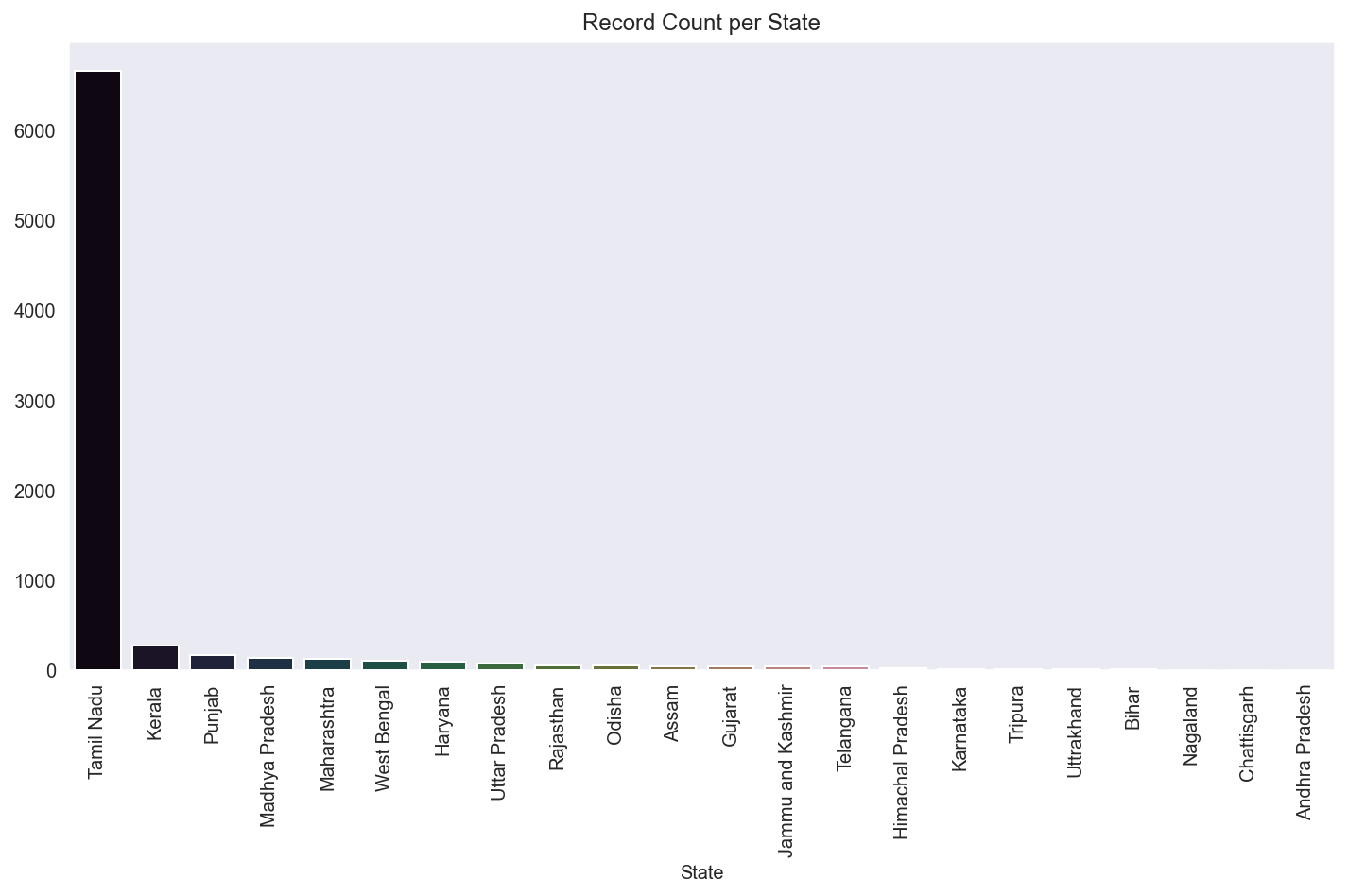
i. **General Description**  
Identify state-wise contribution of records.

ii. **Specific Requirements, Functions and Formulas**  
Function used: value\_counts(), sns.barplot()



iii. **Analysis Results**  
Punjab, Kerela, and Tamil Nadu are the most represented with Tamin Nadu being the most.

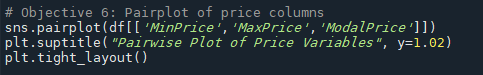
iv. **Visualizations**



**Objective 6: Pairplot of Price Columns**

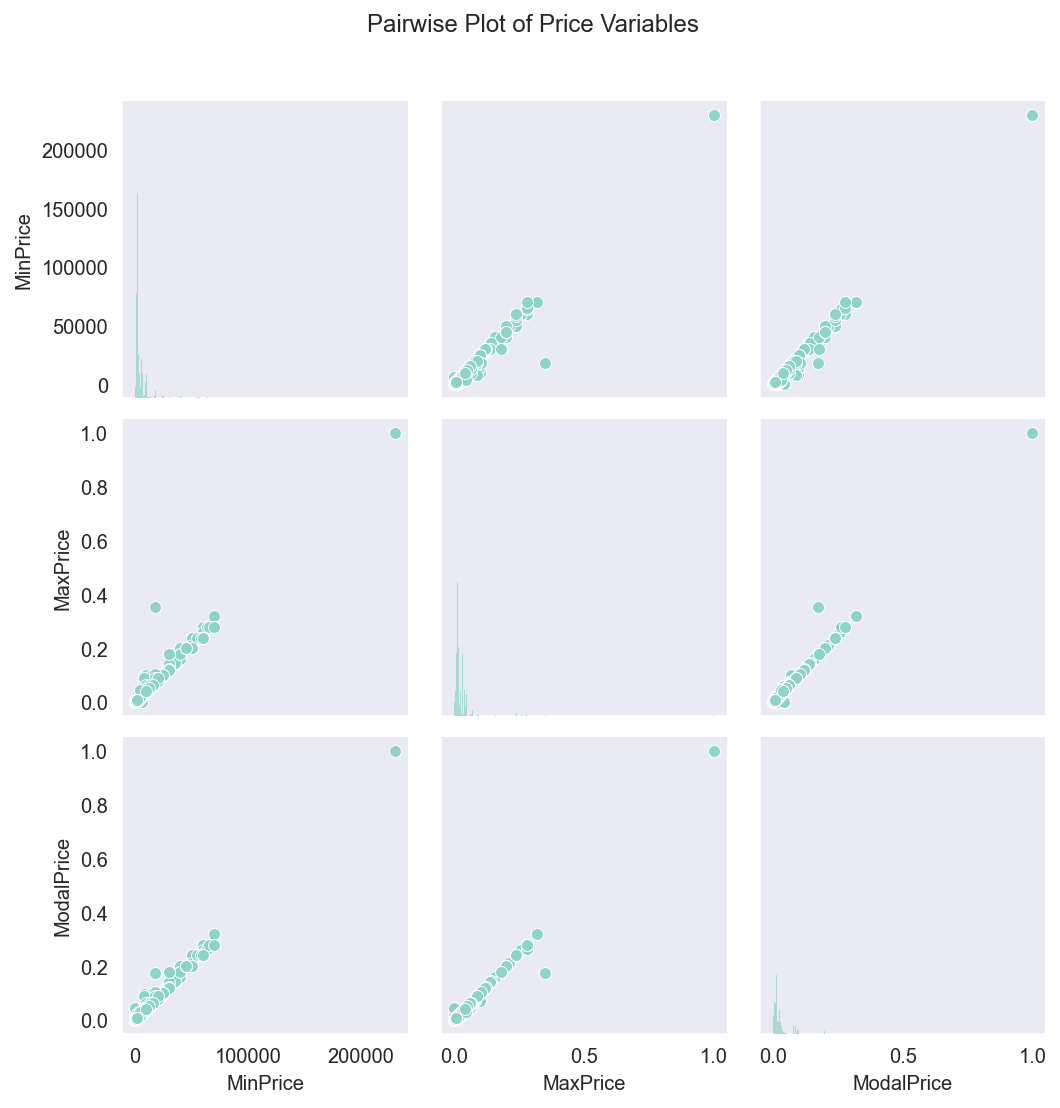
i. **General Description**  
Visualize pairwise relationships between price columns.

ii. **Specific Requirements, Functions and Formulas**  
Function used: pairplot()



iii. **Analysis Results**  
Positive linear trends are visible between price columns.

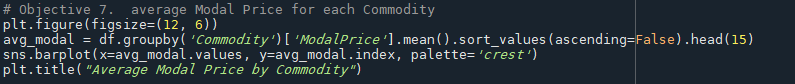
iv. **Visualizations**



**Objective 7: Avg Modal Price by Commodity**

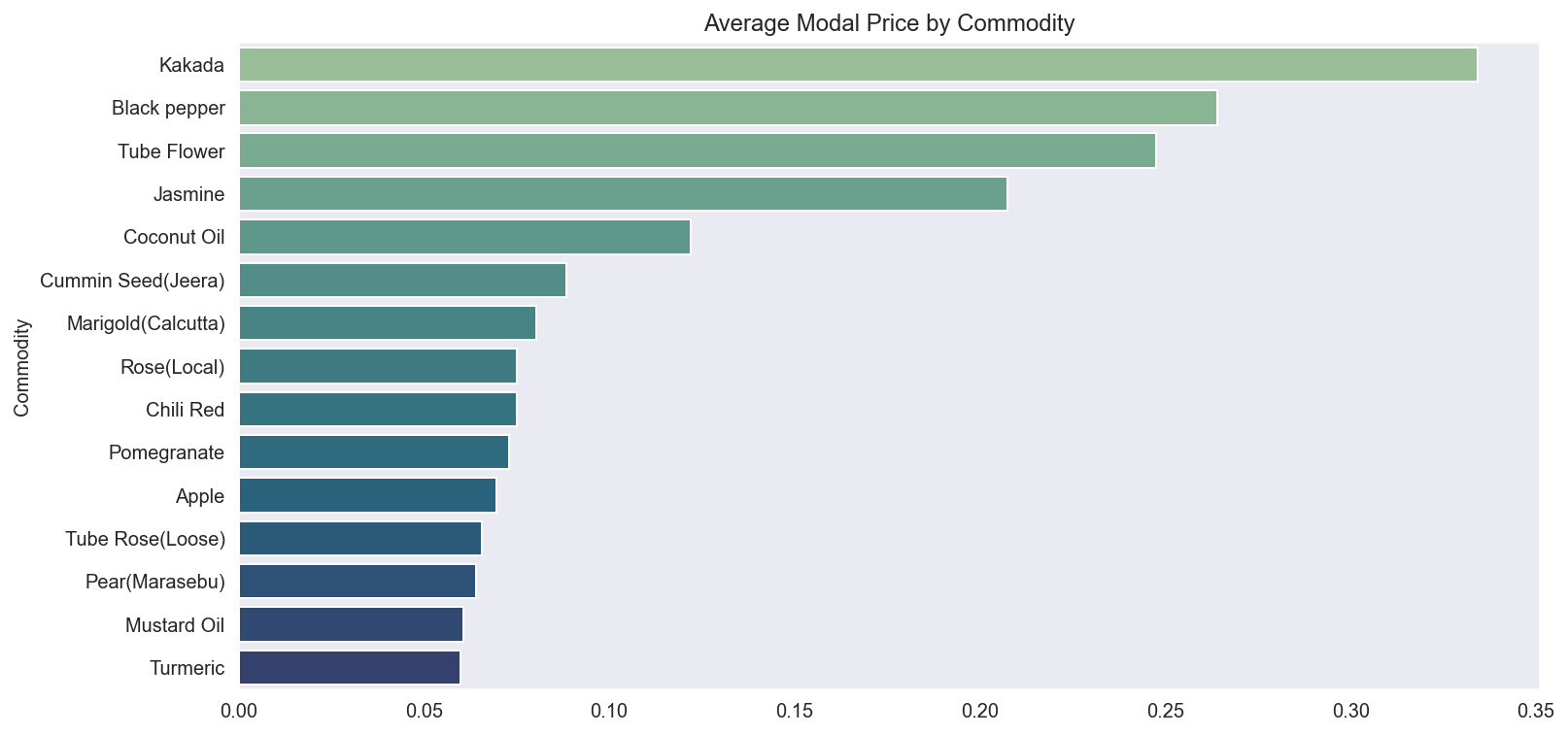
i. **General Description**  
Identify commodities with the highest average modal price.

ii. **Specific Requirements, Functions and Formulas**  
Function used: groupby(), mean(), sort\_values()



iii. **Analysis Results**  
Kakada, Black Pepper and Tube Flower have consistently higher modal prices.

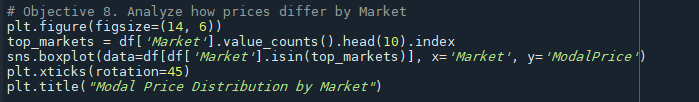
iv. **Visualizations**



**Objective 8: Market-wise Modal Price Distribution**

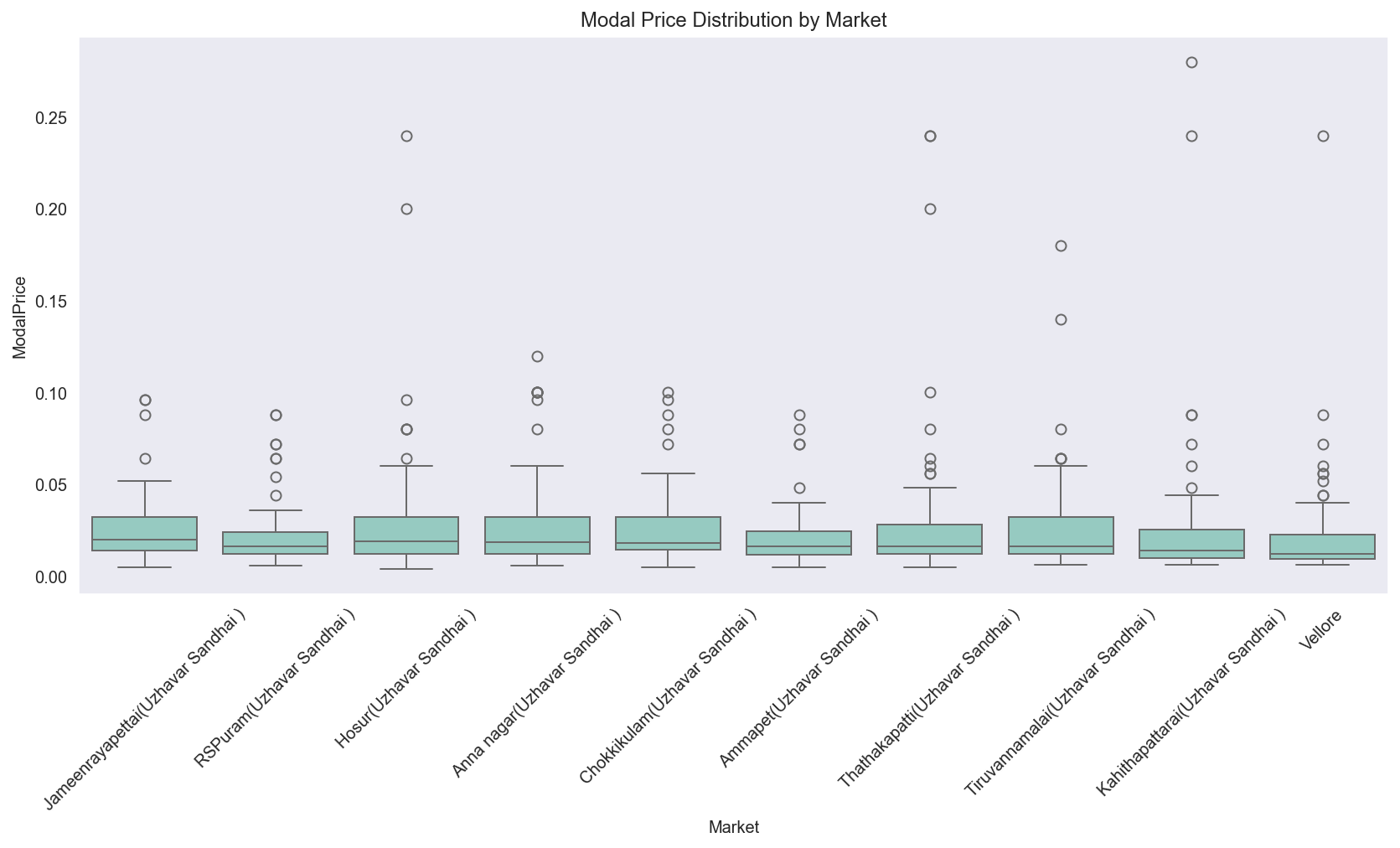
i. **General Description**  
Compare price distributions across key markets.

ii. **Specific Requirements, Functions and Formulas**  
Function used: value\_counts(), boxplot()



iii. **Analysis Results**  
High variability exists among some markets.

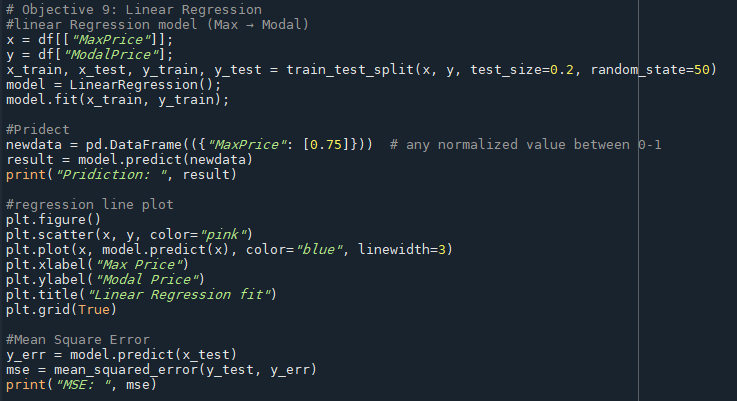
iv. **Visualizations**



**Objective 9: Linear Regression between Max and Modal Prices**

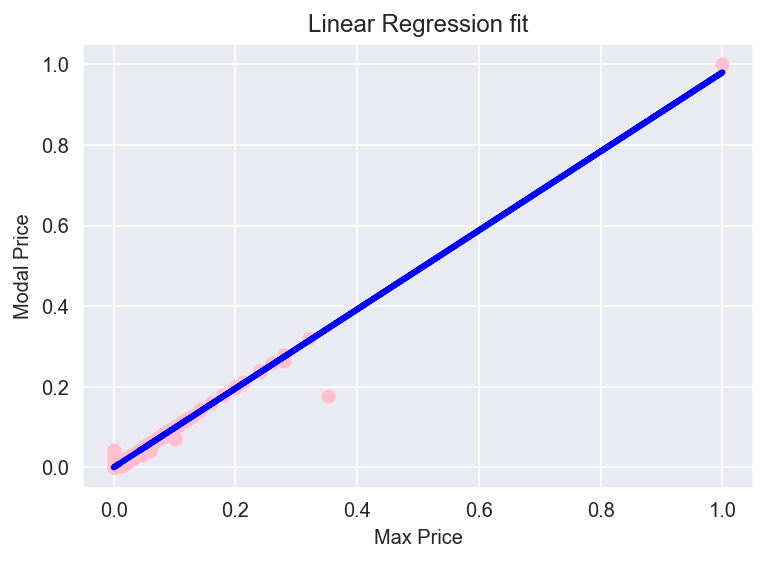
i. **General Description**  
Explore linear relationship and predict modal price from max price.

ii. **Specific Requirements, Functions and Formulas**  
Function used: LinearRegression(), train\_test\_split(), mean\_squared\_error



iii. **Analysis Results**  
The model shows a good fit; low MSE indicates strong correlation.

iv. **Visualizations**  
Regression plot with line of best fit:



# 5. Conclusion

Working on this project gave me an experience of what it's like to work with raw, unstructured data. I started with a messy dataset, cleaned it up, and slowly uncovered meaningful insights using visual tools and basic modeling. The patterns that I found such as:

1)The exploratory analysis on agricultural commodity pricing reveals that a handful of commodities such as Onion, Tomato, and Potato dominate the records.

2) Certain states like Tamil Nadu and Kerela contribute the majority of the data entries.

were not only interesting but also practical. Overall, the dataset provided valuable information on agricultural pricing trends across India.

# 6. Future Scope

Future scope includes:

* Incorporate **time-series forecasting** to predict future price trends.
* Apply **machine learning models** for price classification and anomaly detection.
* Develop **interactive dashboards** using tools like Power BI or Tableau.
* Expand the dataset by integrating **weather or seasonality data** for deeper insights.
* Conduct **market efficiency studies** to identify potential improvements in distribution and pricing.

# 7. References

[1] Dataset: <https://www.data.gov.in/resource/current-daily-price-various-commodities-various-markets-mandi>   
[2] pandas documentation: <https://pandas.pydata.org/>  
[3] seaborn documentation: <https://seaborn.pydata.org/>  
[4] scikit-learn documentation: <https://scikit-learn.org/stable/>  
[5] matplotlib documentation: <https://matplotlib.org/>