Food Price Prediction

By : Khairiyah Rizkiyah

Table of Contents

#1 Data and Business Understanding

Introduction

The domestic price of essential foods is monitored continuously in every country to make sure the controlled price and inflation.

Data and
Business
Understanding

Dataset used are Indonesia Food Price from the World Food Programme Price Database

This dataset contains Food Prices data for Indonesia and covers foods such as rice, eggs, meat, and sugar. The data are collected from several markets across the country.

Price Data available from January 2007 to March 2020 every month.

All the foods category price available until March 2020, meanwhile the non-food commodity only available until 2013.

Data and
Business
Understanding

Objective Statements

01

Identifying the trend of commodity price represented by the dataset.

02

Comparing the best forecasting model in predicting the future value of the food price data.

#2 Data Preparation

Data Preparation

There are 41.956 rows of data from 2007 until 2020 with 13 Original columns.

Not all columns will be used in further analysis. **There are 8 features selected** from 13 general features available:

No.	Selected Feature	Description
1.	Date	The date of price data recorded from market
2.	Market	All the market in Indonesia chosen to collect data
3	Category	The category of commodity
4	Commodity	The name of commodity
5	Unit	The measurements of commodity
6	Priceflag	There are 2 type of priceflag, aggregate and actual
7	Pricetype	The data collected is in retail price type
8	Price	The price of commodity (IDR)

Data Preparation

Missing Value Handling

There are only a few missing values, then it was decided that the missing value would be deleted (drop).

```
foodprice.isna().sum()
```

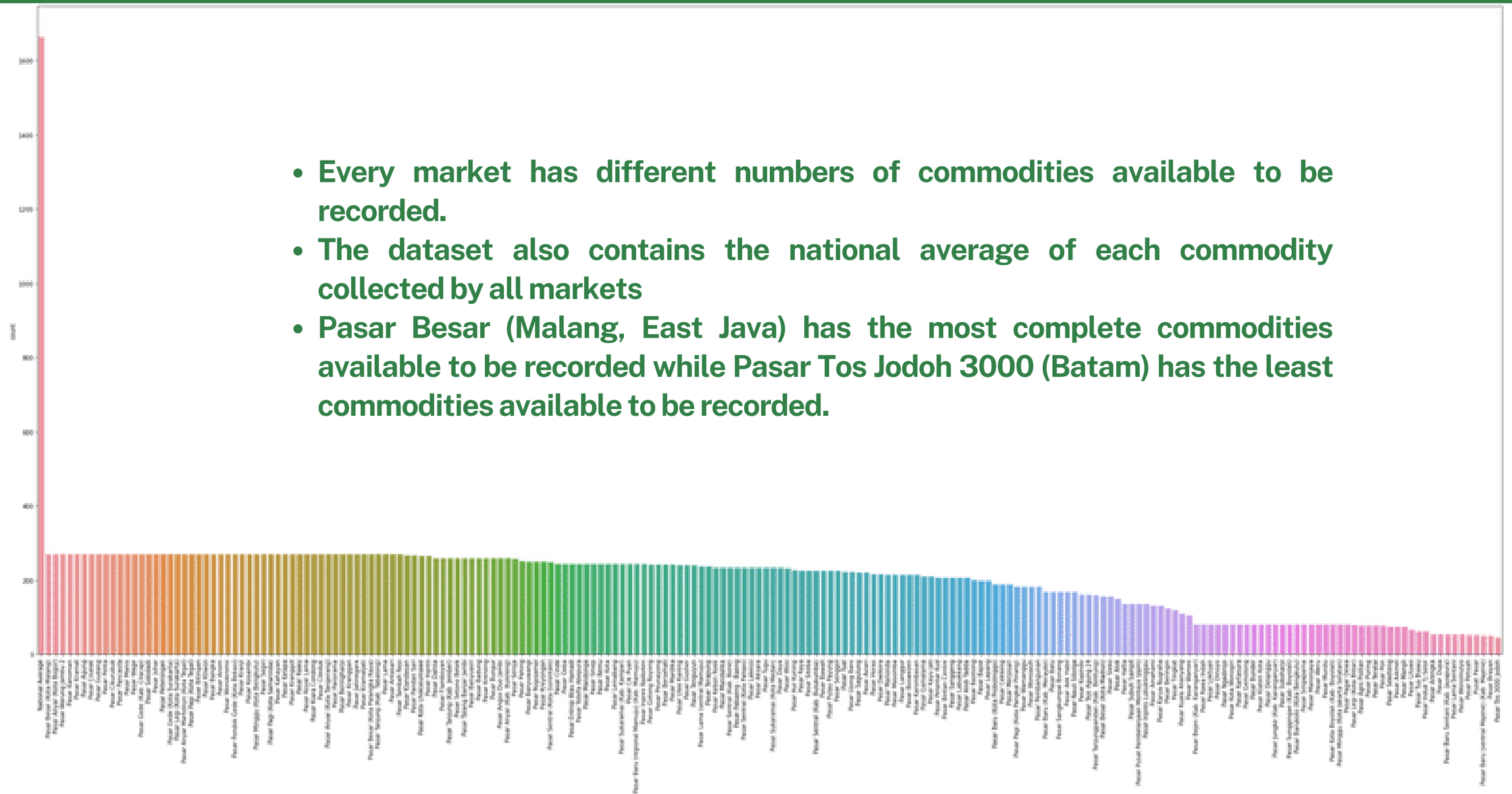
```
date          0
market        0
category      1
commodity     1
unit          1
priceflag     1
pricetype     1
price         1
dtype: int64
```

```
#check missing value again
foodprice2.isna().sum()
```

```
date          0
market        0
category      0
commodity     0
unit          0
priceflag     0
pricetype     0
price         0
dtype: int64
```

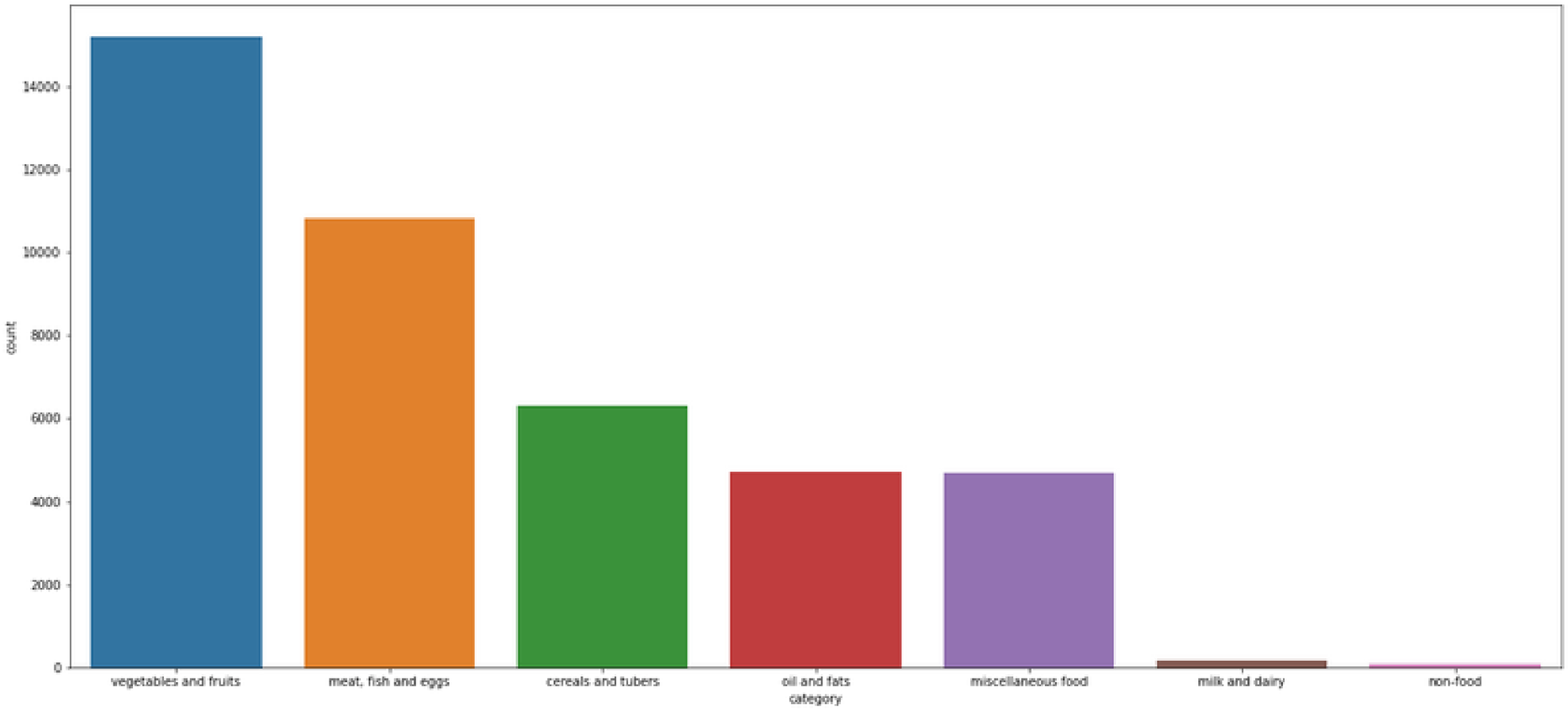
#3 Exploratory Data Analysis (EDA)

The Market



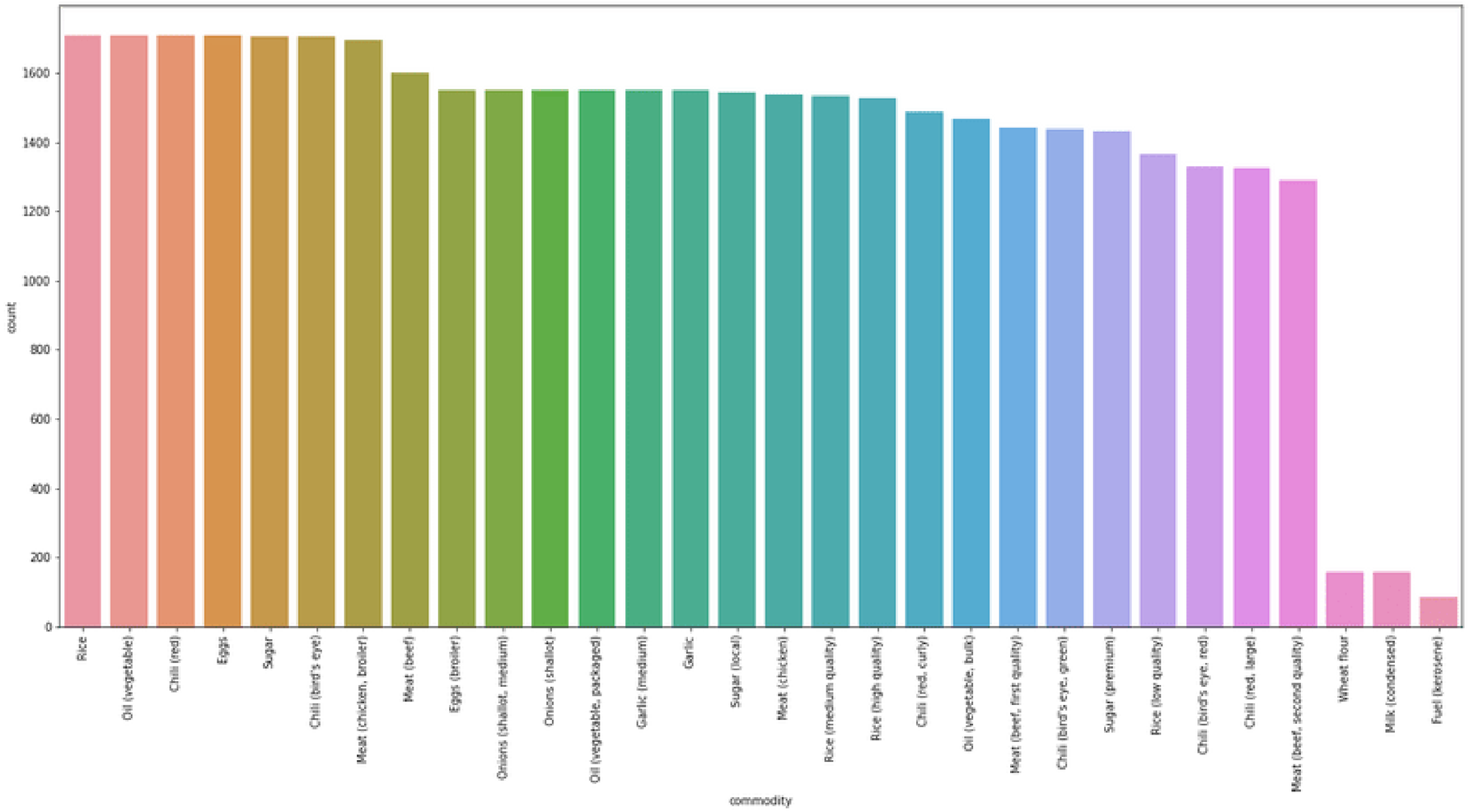
- Every market has different numbers of commodities available to be recorded.
- The dataset also contains the national average of each commodity collected by all markets
- Pasar Besar (Malang, East Java) has the most complete commodities available to be recorded while Pasar Tos Jodoh 3000 (Batam) has the least commodities available to be recorded.

The Category



- **Consist of 7 Category**
- **Not all commodities are recorded in each market, the numbers of data are different in each category.**

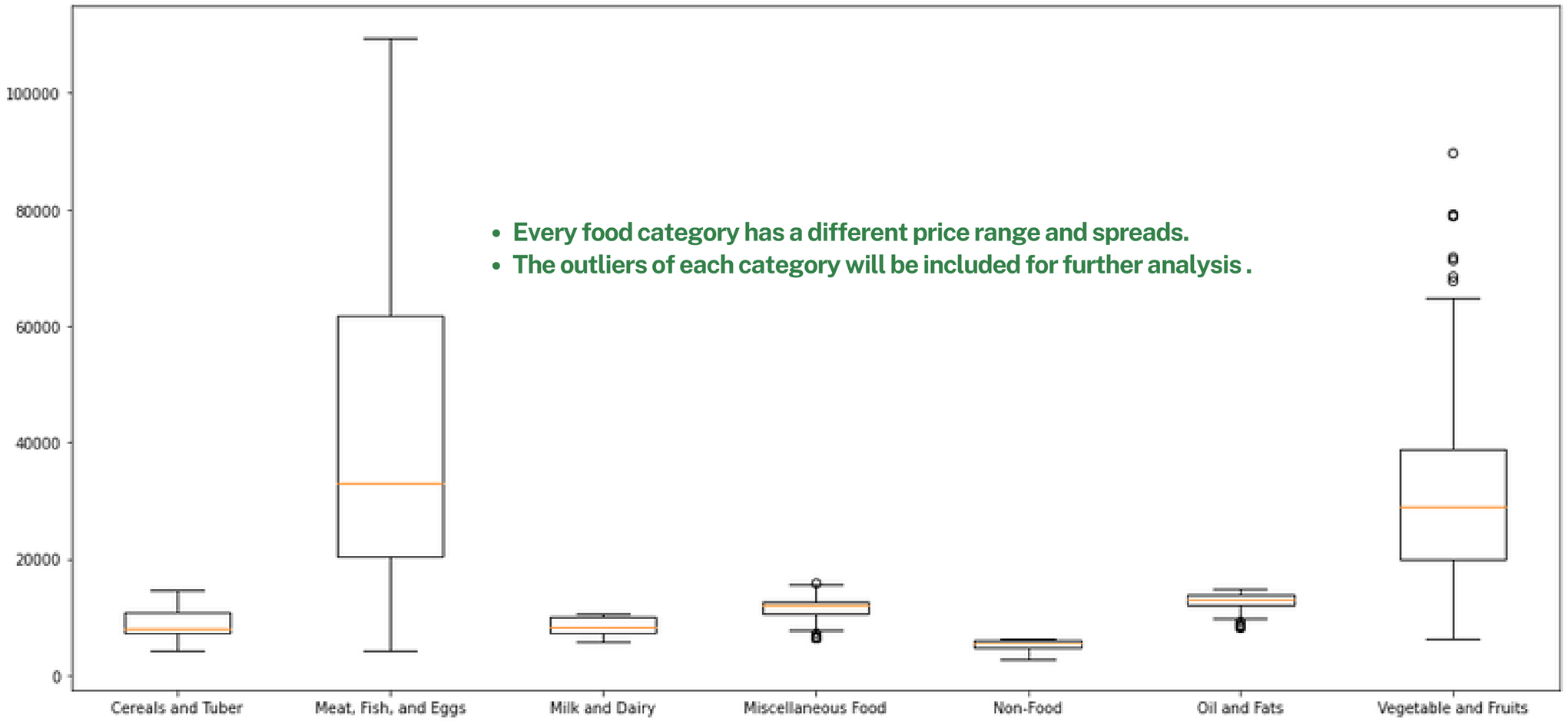
The Commodity



Commodity:

1. Rice
2. Wheat flour
3. Eggs
4. Meat
5. Milk (condensed)
6. Sugar
7. Fuel (kerosene)
8. Oil (vegetable)
9. Chili (bird's eye)
10. Chili (red)
11. Rice (high quality)
12. Rice (low quality)
13. Rice (medium quality)
14. Eggs (broiler)
15. Meat (beef, first quality)
16. Meat (chicken)
17. Sugar (local)
18. Sugar (premium)
19. Garlic
20. Garlic (medium)
21. Onions (shallot)
22. Onions (shallot, medium)
23. Meat (beef, second quality)
24. Chili (bird's eye, red)
25. Chili (red, large)

Price Distribution



- Every food category has a different price range and spreads.
- The outliers of each category will be included for further analysis .

#4 Analysis & Modelling

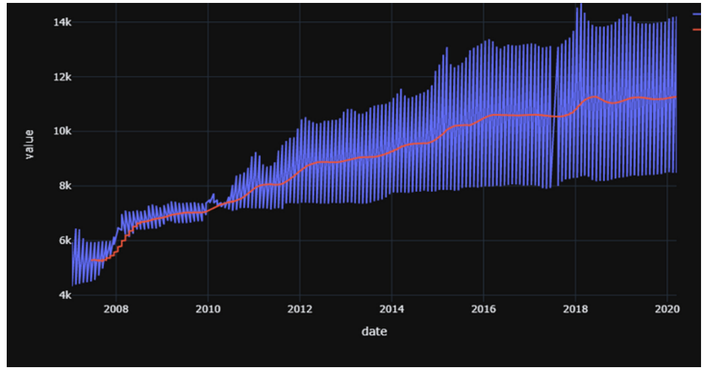
HOW WILL THE DATA BE ANALYZED AND MODELED?

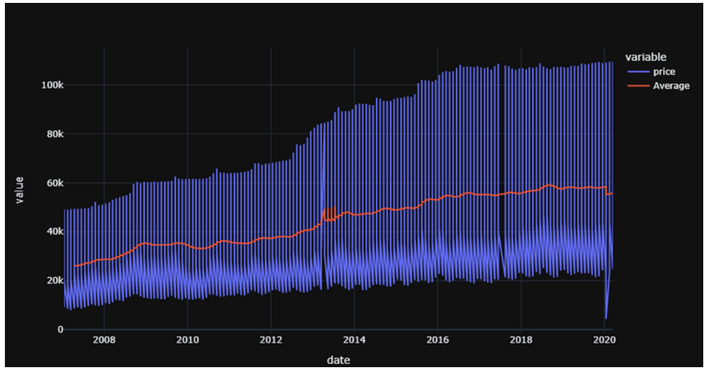
The first step in the process of trend identification is smoothing. In this case, the 12-month **Moving Average** method is used for each category's commodity.

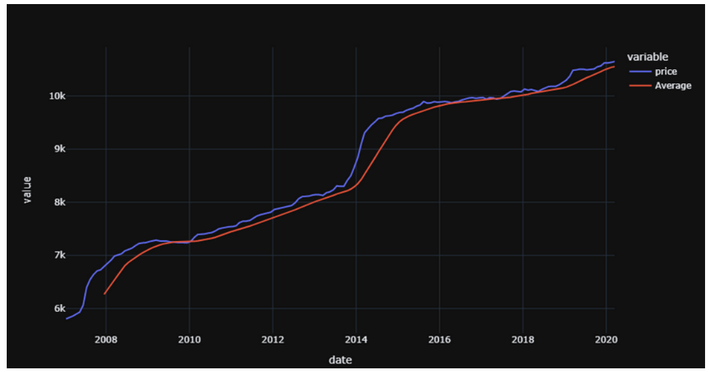
The most common model to forecast time series data is Exponential Smoothing. This case will use **Simple Exponential Smoothing** as the basic model, and **Holt-Winters** (Triple Exponential Smoothing) to take consider the trend and seasonality.

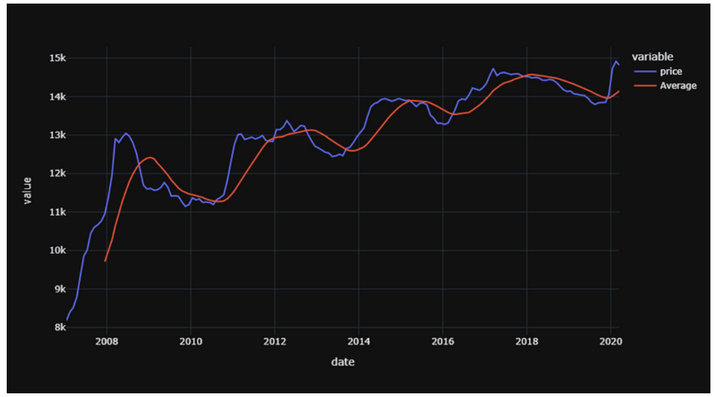
Trend Analysis

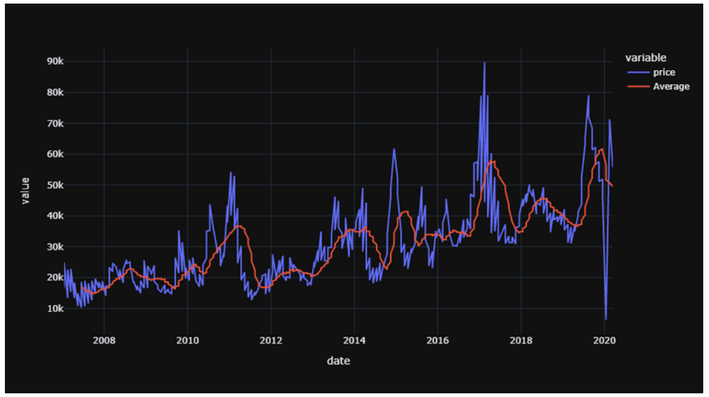
Using 12 Month
Moving Average











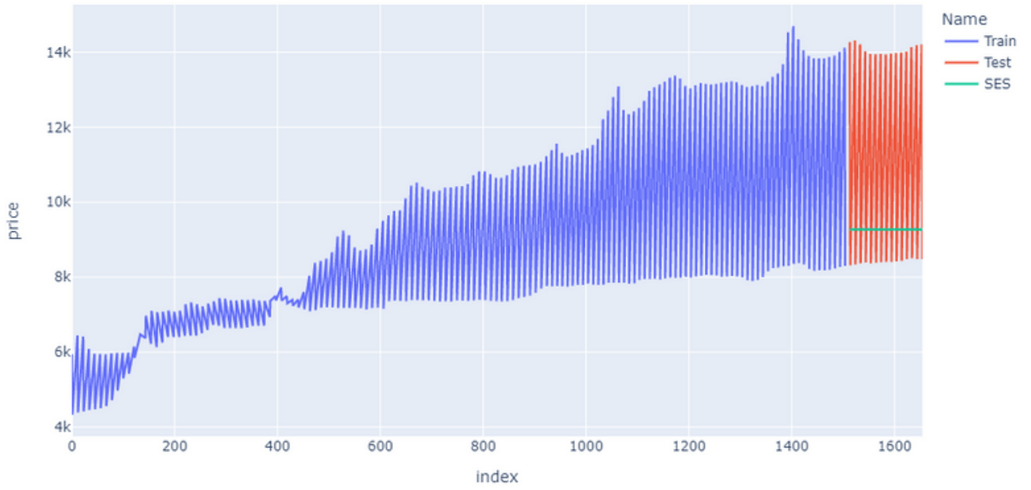
All food categories have a rising trend.

Modeling

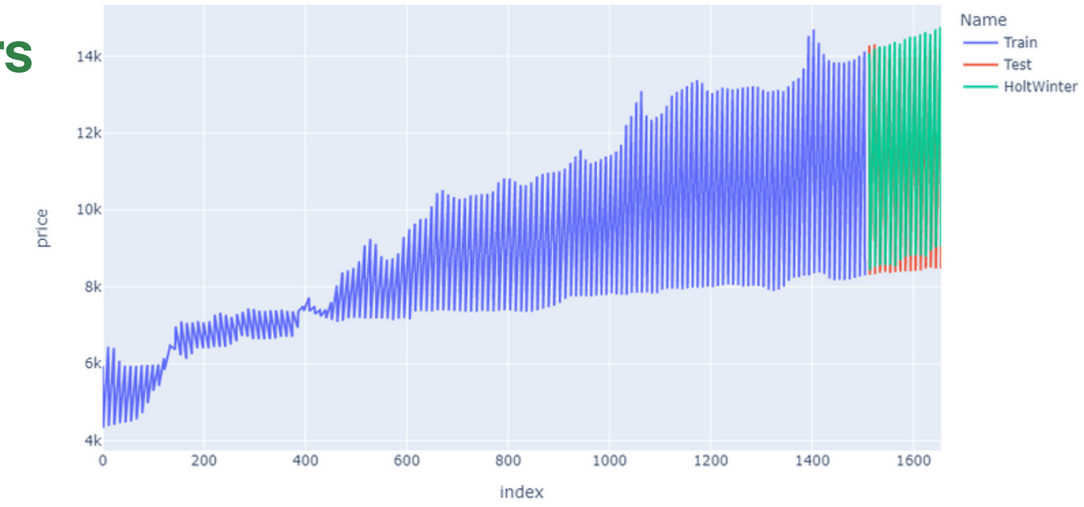
Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #1: Cereal and Tubers

Simple Exponential Smoothing



Holt-Winters

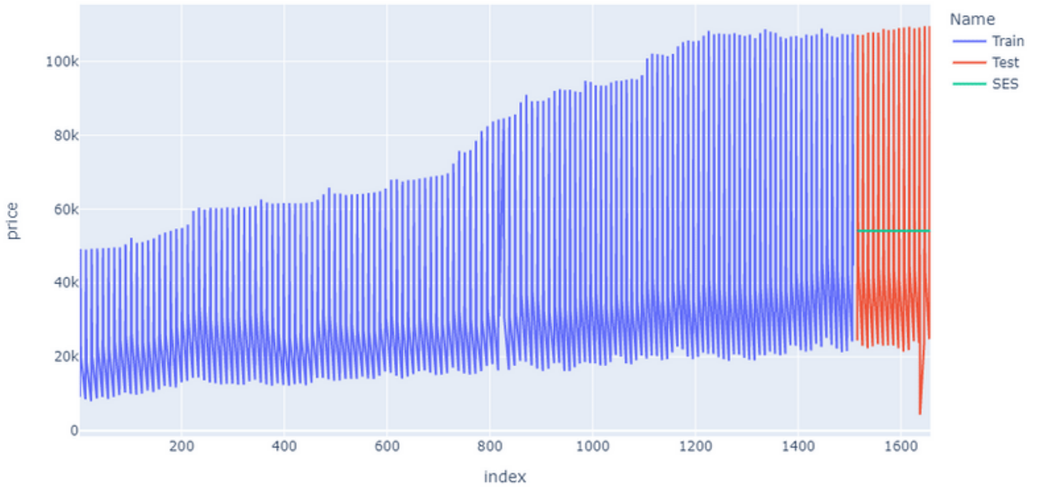


Modeling

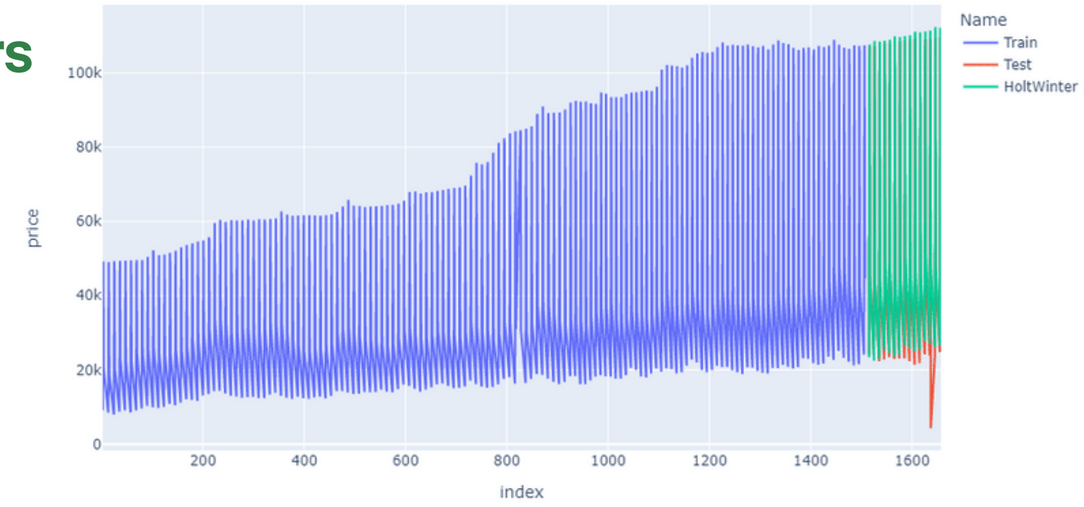
Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #2: Meat, Fish and Eggs

Simple Exponential Smoothing



Holt-Winters

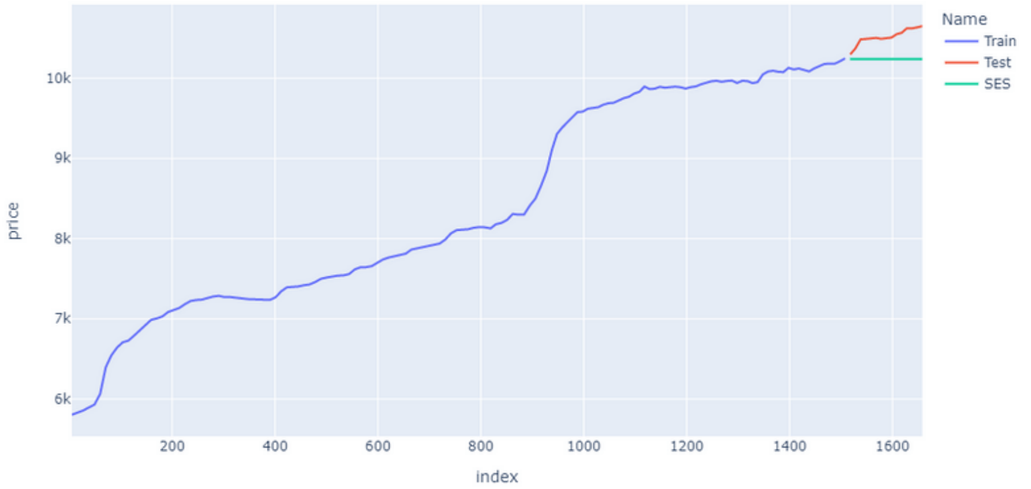


Modeling

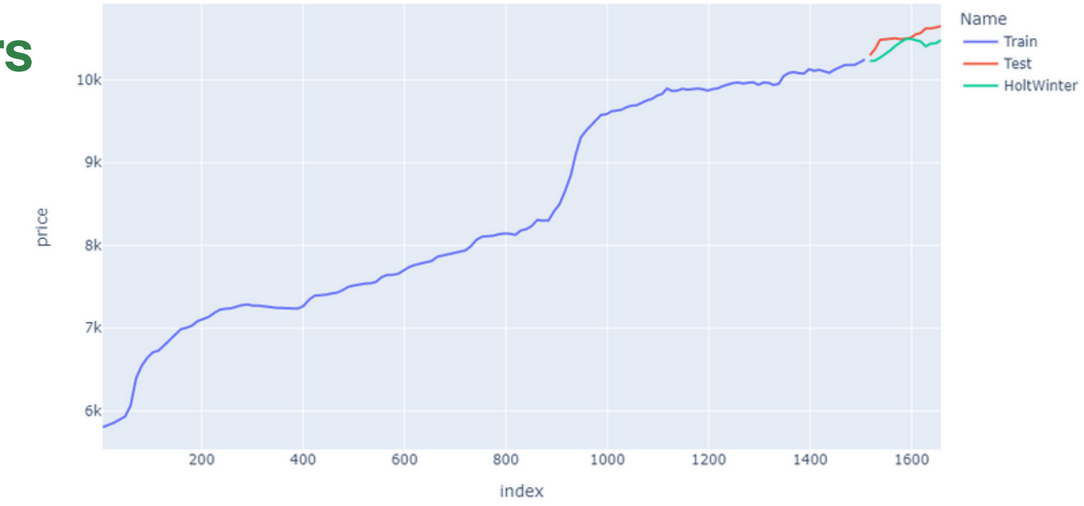
Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #3: Milk and Dairy

Simple Exponential Smoothing



Holt-Winters

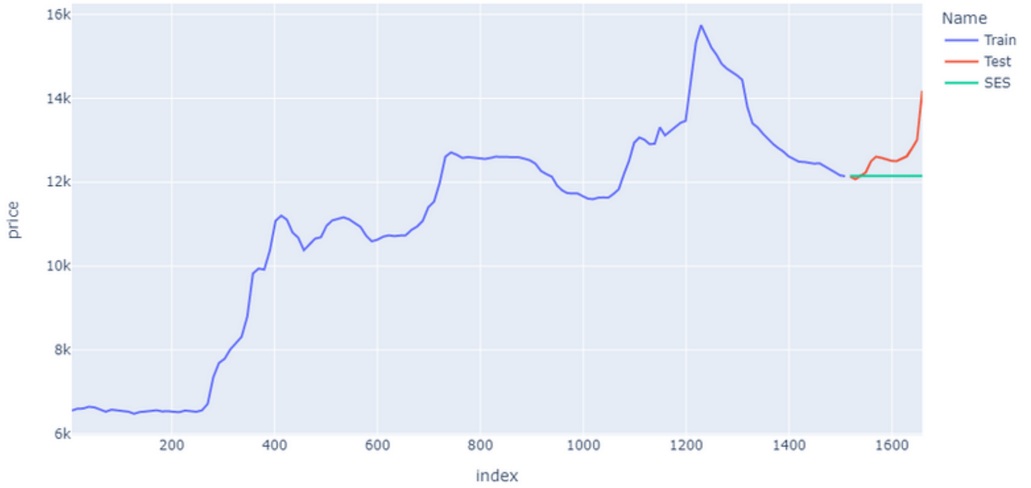


Modeling

Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #4: Miscellaneous Food

Simple Exponential Smoothing



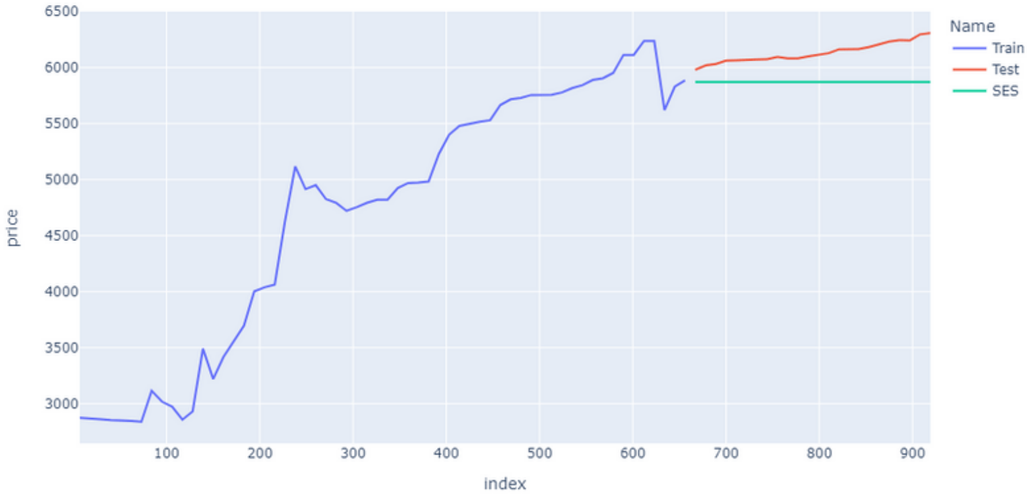
Holt-Winters

Modeling

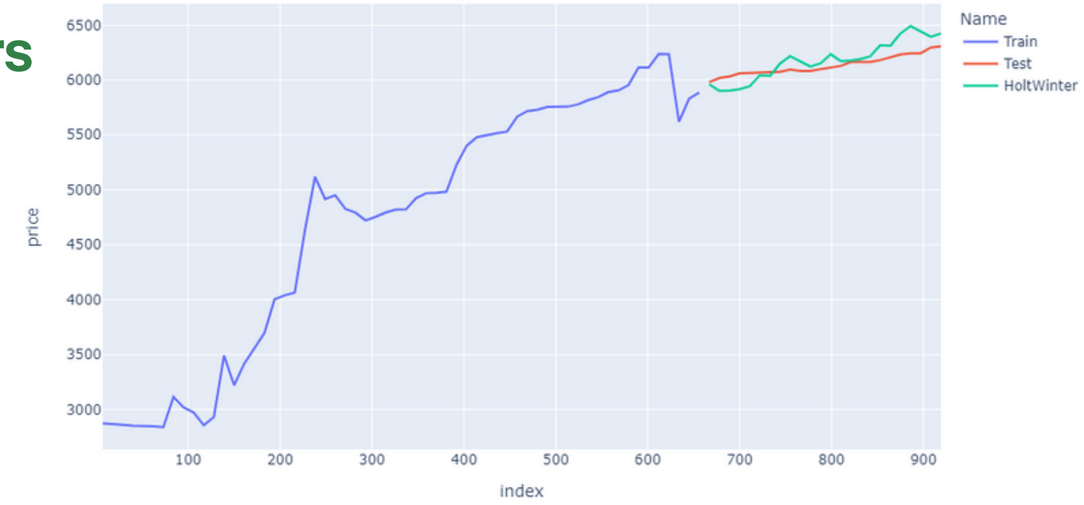
Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #5: Non-Food

Simple Exponential Smoothing



Holt-Winters



Modeling

Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #6: Oil and Fats

Simple Exponential Smoothing

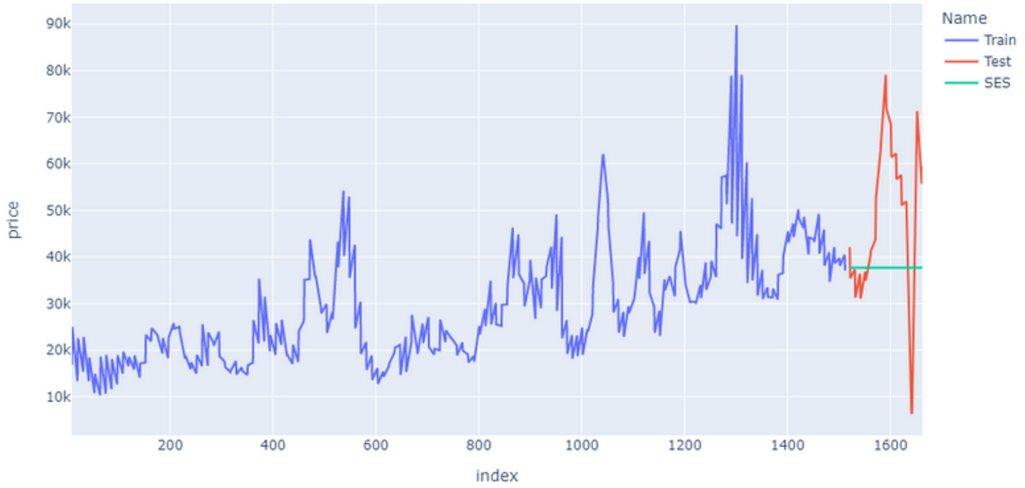
Holt-Winters

Modeling

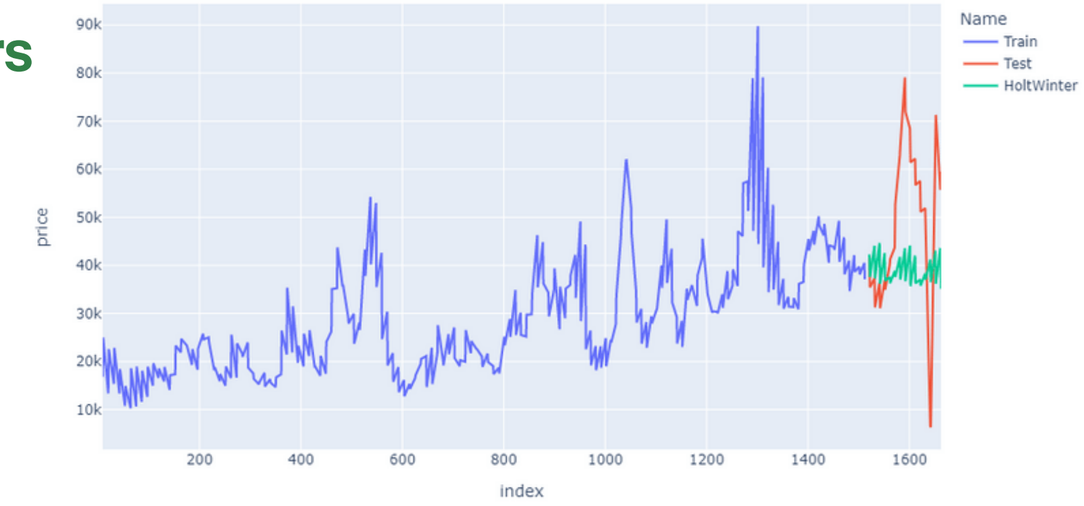
Comparing Between Simple Exponential Smoothing (SES) and HoltWinters Forecasing Model

Category #7: Vegetables and Fruits

Simple Exponential Smoothing



Holt-Winters



Model Evaluation

Holt-Winters gives better results

Based on the Root Mean Square Error (RMSE), Holtwinters Model give better price prediction than Simple Exponential Smoothing (SES) on most category.

No.	Category of Commodity	RMSE	
		SES	Holt-Winters
1.	Cereals and Tubers	3.451	393
2.	Meat, Fish and Eggs	37.431	7.161
3	Milk and Dairy	295	140
4.	Miscellaneous Food	664	890
5.	Non-Food	274	113
6.	Oil and Fats	360	478
7.	Vegetables and Fruits	20.639	19.947

#5 Conclusion and Recommendation

CONCLUSION:

- Domestic food commodity prices have a rising trend every year and show seasonality.
- SES Model always gives a stagnant result. Therefore, it is best uses only to predict the next point of value.
- Food price fit best with Holt-Winter's forecasting modeling than SES.

RECOMMENDATION?

- **For domestic food prices, only use models that take into consideration trends and seasonality for forecasting (VAR, SARIMA, Holt-Winters, etc).**

- **Each commodity has different price nature, better results can be obtained by individual commodity analysis.**

- **More evaluation metrics should be provided.**

Thank you!