Predicting Commodity Prices Using Time Series Analysis

Group 4:

1. Piyush Sunil Borse 25PGAI0026
2. Prateek Majumder 25PGAI0027
3. Bhawana Thawarani 25PGAI0137
4. Prajwal Wagh 25PGAI0109
5. Yuvraj Singh Srinet 25PGAI0019

Problem Statement

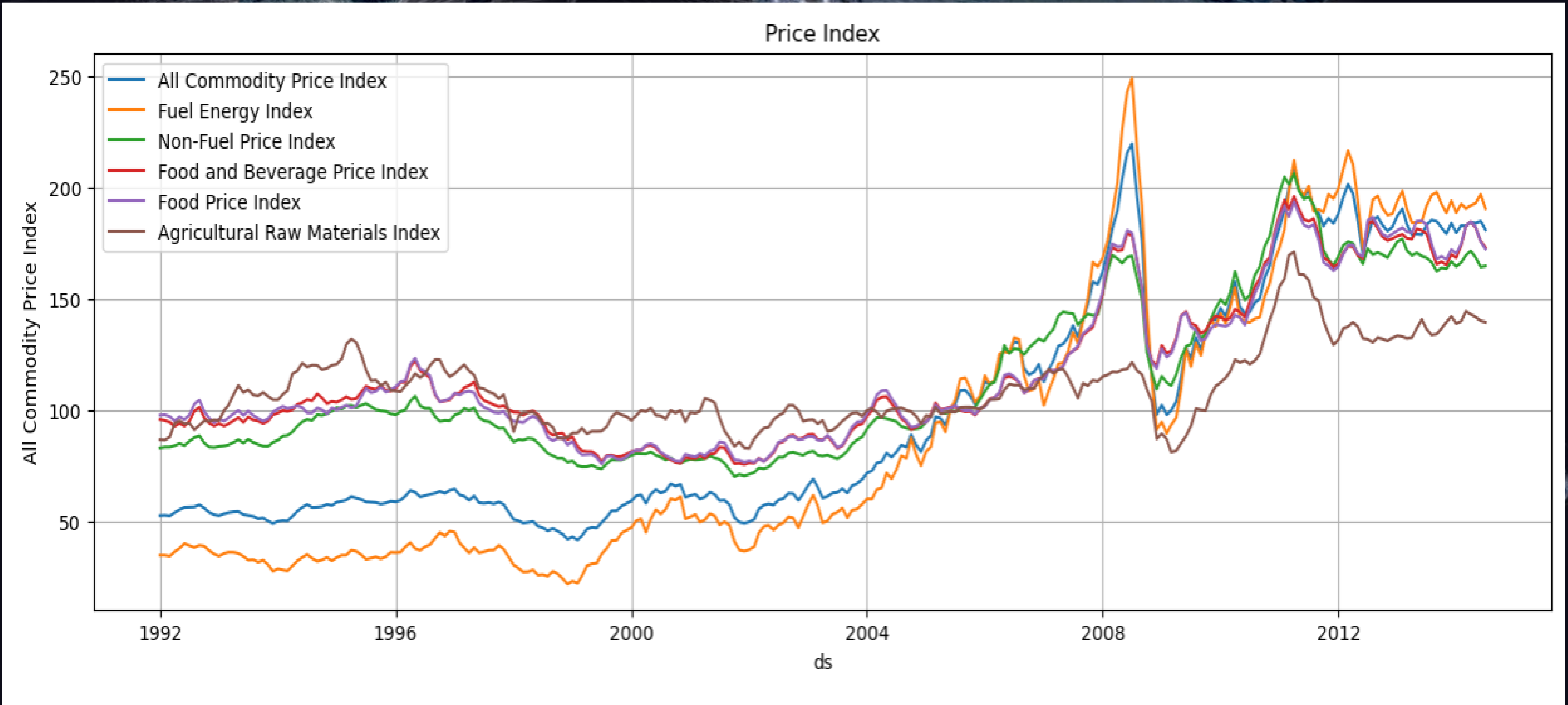
- The objective of this project is to develop a robust time series model that can predict future prices of key global commodities.
- This prediction model will leverage historical price data, to forecast future price movements.
- In our implementation, we have predicted Fuel Price Index.

DATASET

- We got our dataset from Kaggle.
- The dataset includes monthly prices and indexes for 53 commodities from 1992 to 2014, with 2005 as the reference year for indexes (value = 100). Key categories in the dataset include All Commodity, Food and Beverage, Food, Beverage and Fuel Energy. Each category is represented by its respective price index.

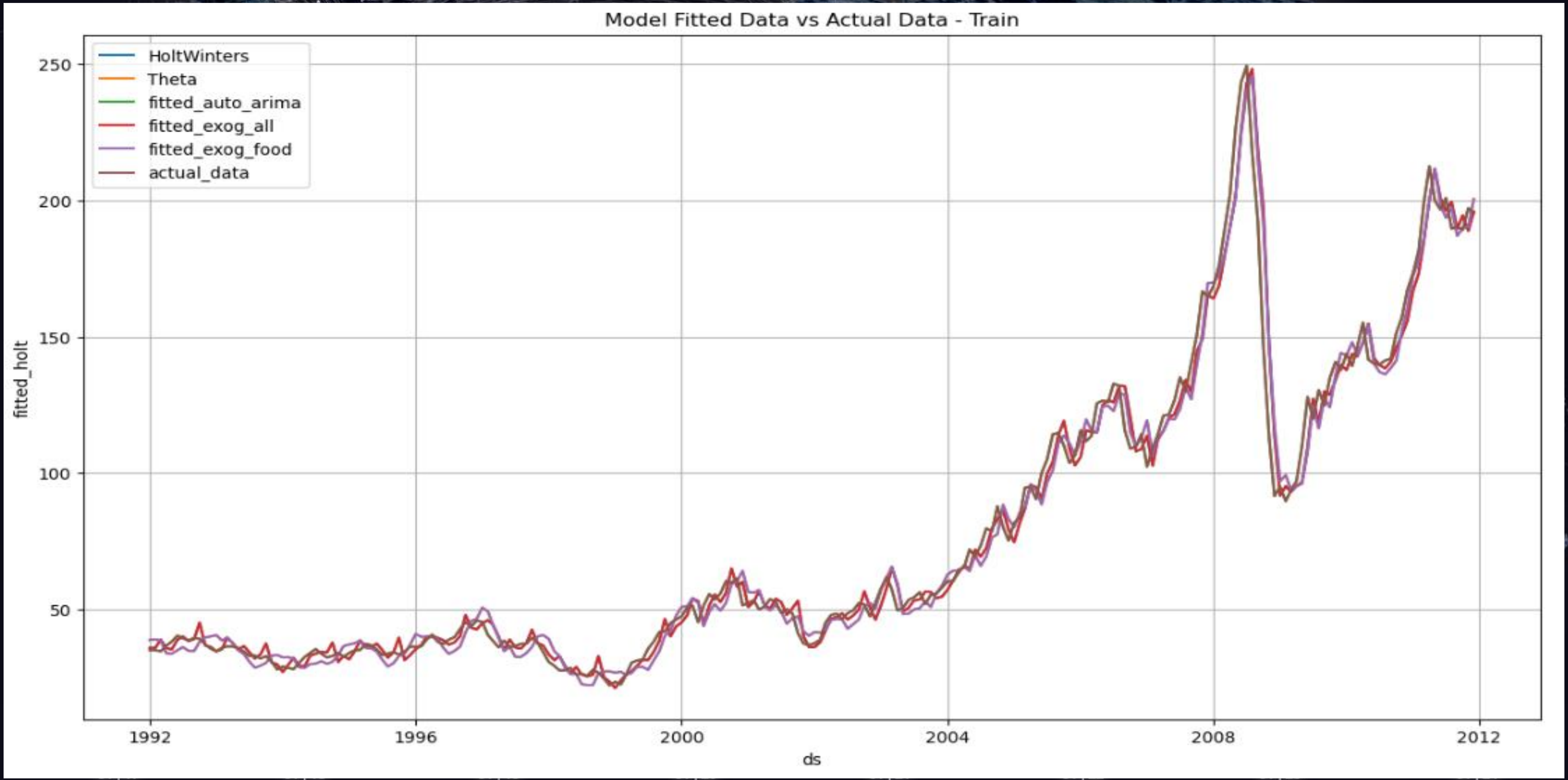
#	Column	Non-Null Count	Dtype
0	All Commodity Price Index	271 non-null	float64
1	Non-Fuel Price Index	271 non-null	float64
2	Food and Beverage Price Index	271 non-null	float64
3	Food Price Index	271 non-null	float64
4	Agricultural Raw Materials Index	271 non-null	float64
5	Fuel Energy Index	271 non-null	float64
6	ds	271 non-null	datetime64[ns]
7	unique_id	271 non-null	object

Data Analysis



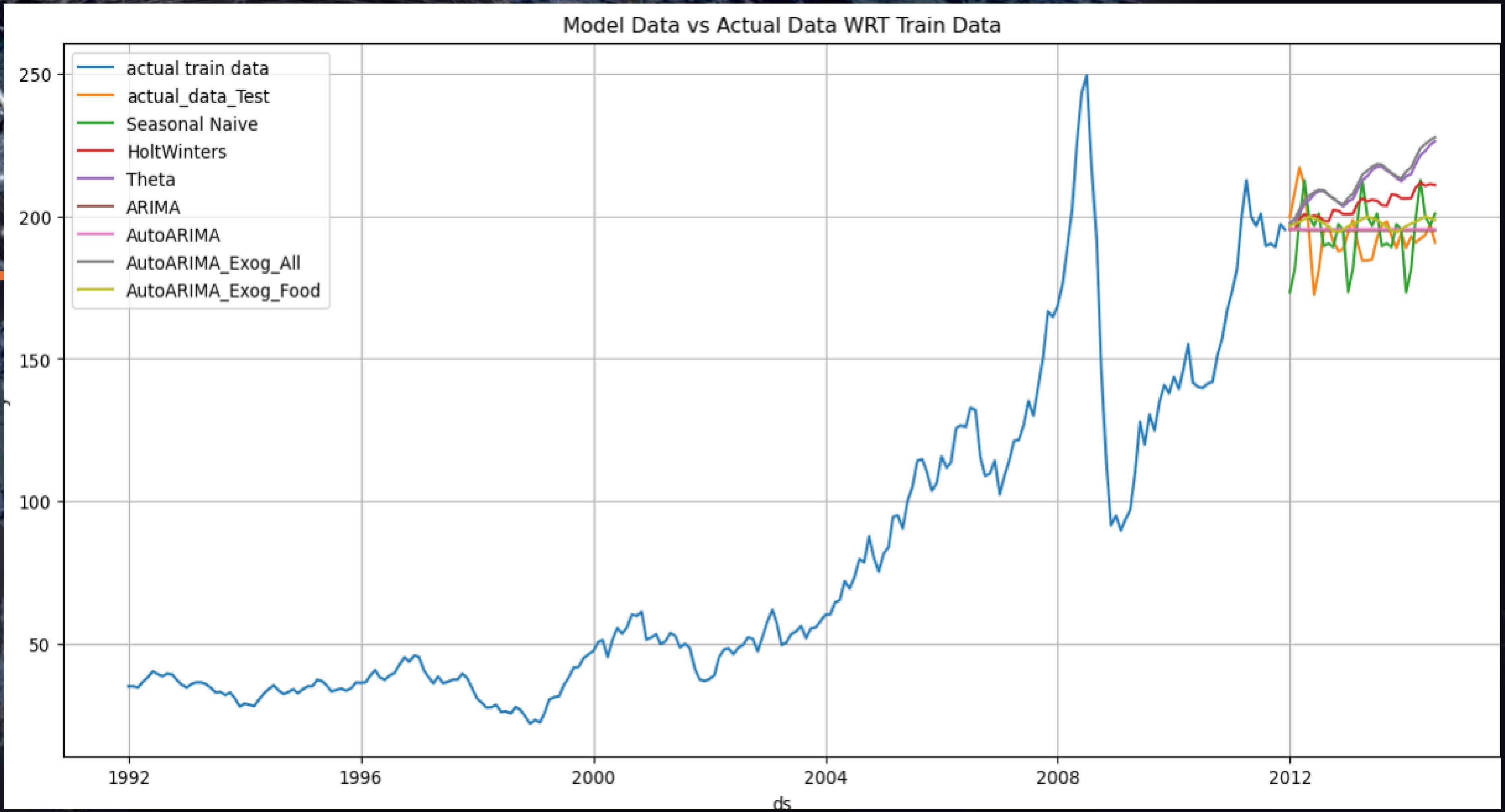
Dip in Data after 2008

Time Series Modeling



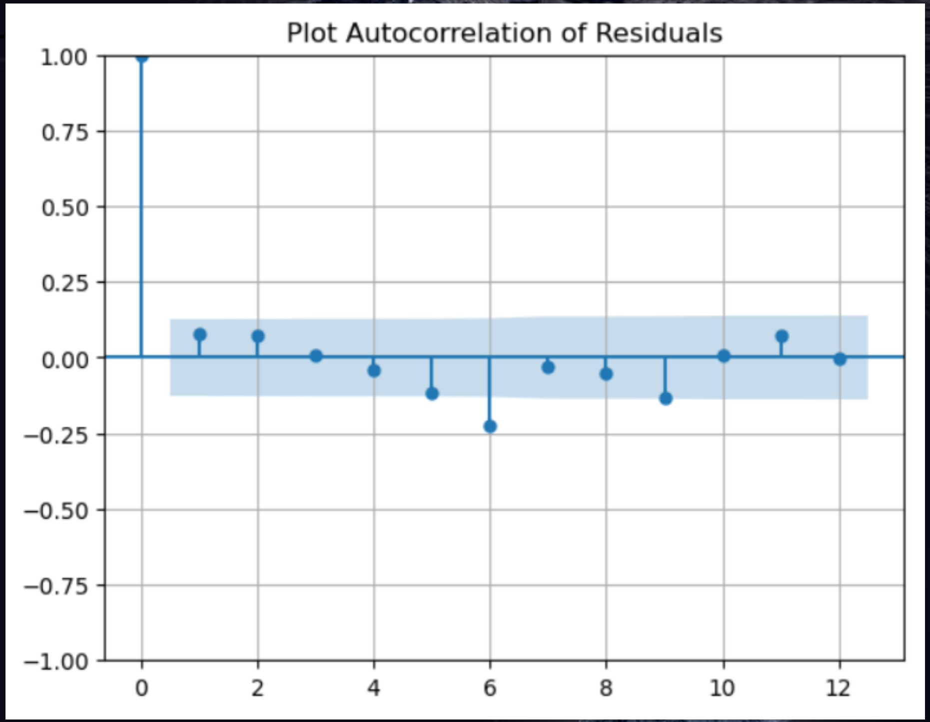
Time Series Forecasting

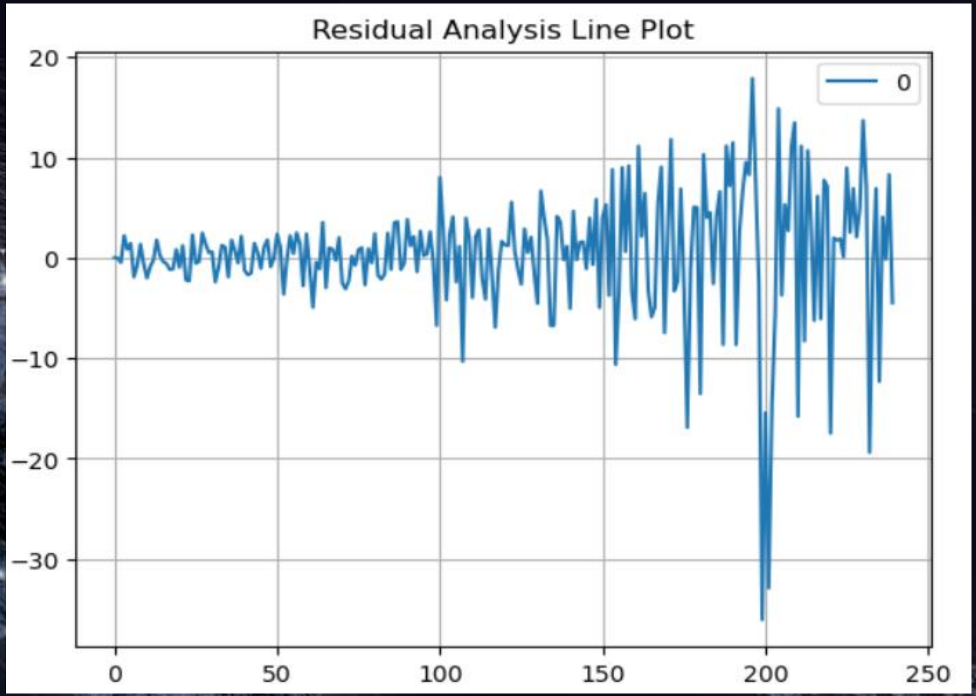
- Seasonal Naïve
- HoltWinters
- Theta
- ARIMA
- AutoARIMA
- AutoARIMA with Exogenous Variables



Residual Diagnostics

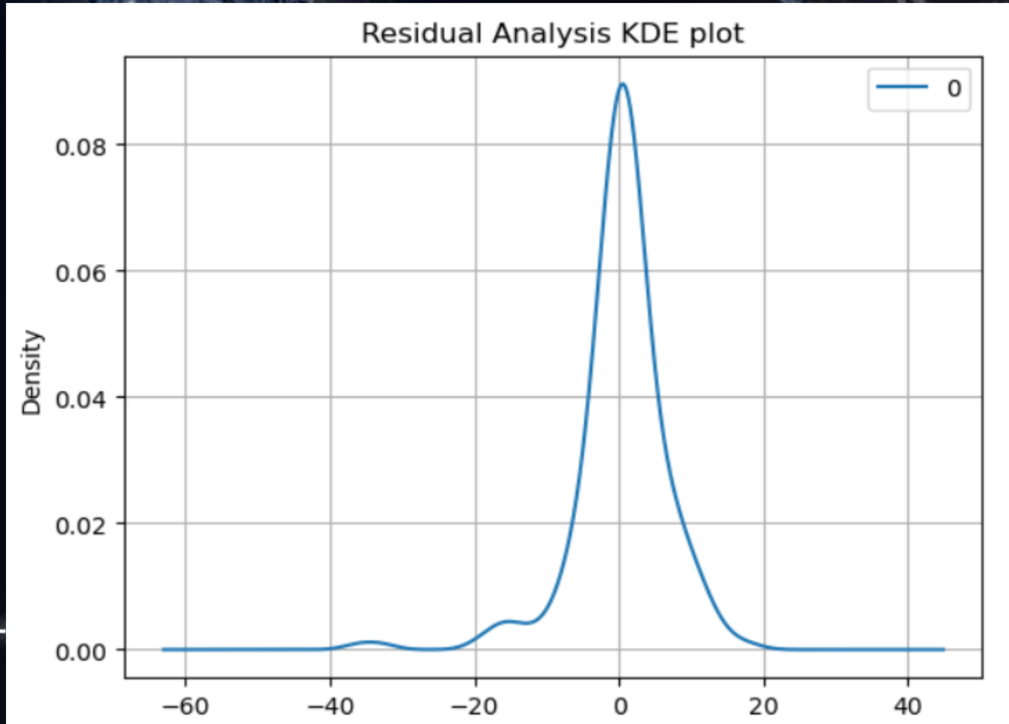
AU






```
Summary statistics of residuals:
                0
count  240.000000
mean     0.362093
std      6.332664
min    -36.052396
25%     -1.894083
50%      0.680979
75%      2.913590
max     17.890489
```

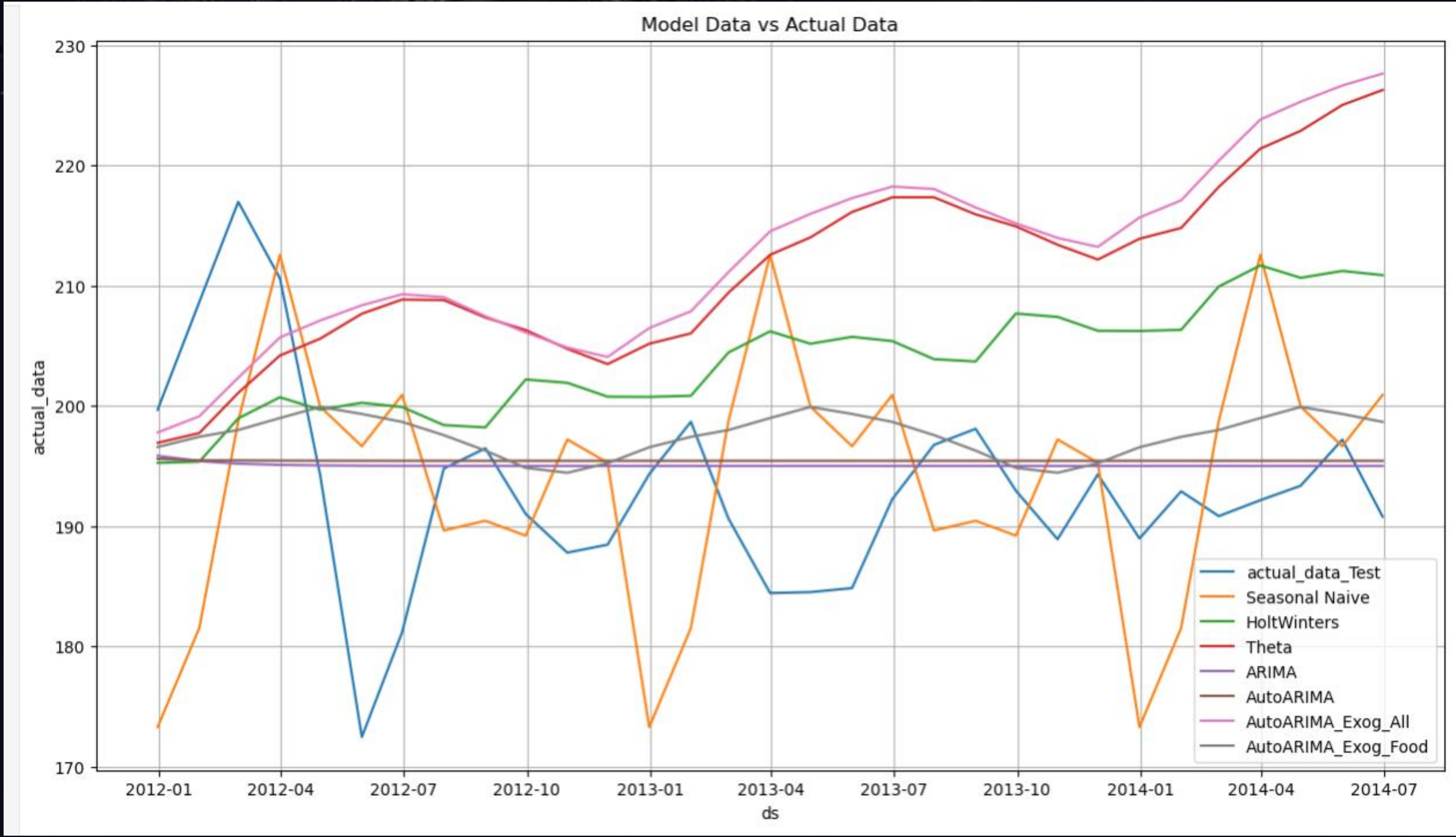
01 | 20



01 | 24

Results:

AUTO A



Error Reporting:

	Seasonal Naive	HoltWinters	Theta	ARIMA	AutoARIMA	AutoARIMA_Exog_All	AutoARIMA_Exog_Food
MAE	11.72	13.49	20.4	6.15	6.37	21.26	7.44
MAPE	6.1	7.07	10.7	3.20	3.32	11.16	3.91
MSE	200.52	223.18	489.28	71.98	74.47	536.49	93.18

The ARIMA model outperforms other models in terms of MAE (6.15), MAPE (3.20), and MSE (71.98). This indicates that ARIMA provides the most accurate forecasts among the models compared.

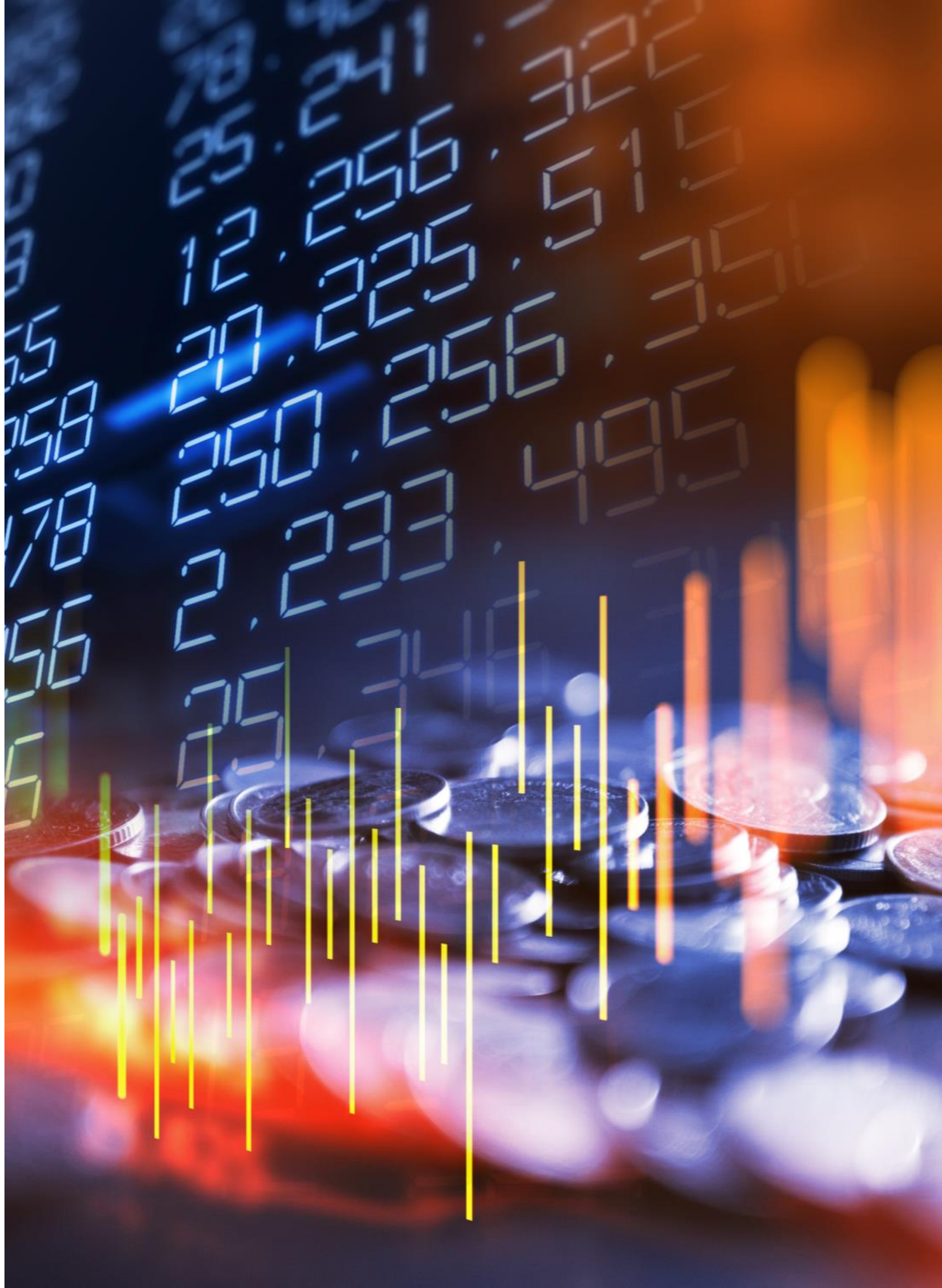
The AutoARIMA model that includes only food-related exogenous variables (AutoARIMA_Exog_Food) also performs well, especially in terms of MAPE (3.91) and MAE (7.44). Its performance is better than the basic AutoARIMA model but not as good as the basic ARIMA model.

Limitations of the Chosen Dataset and Models

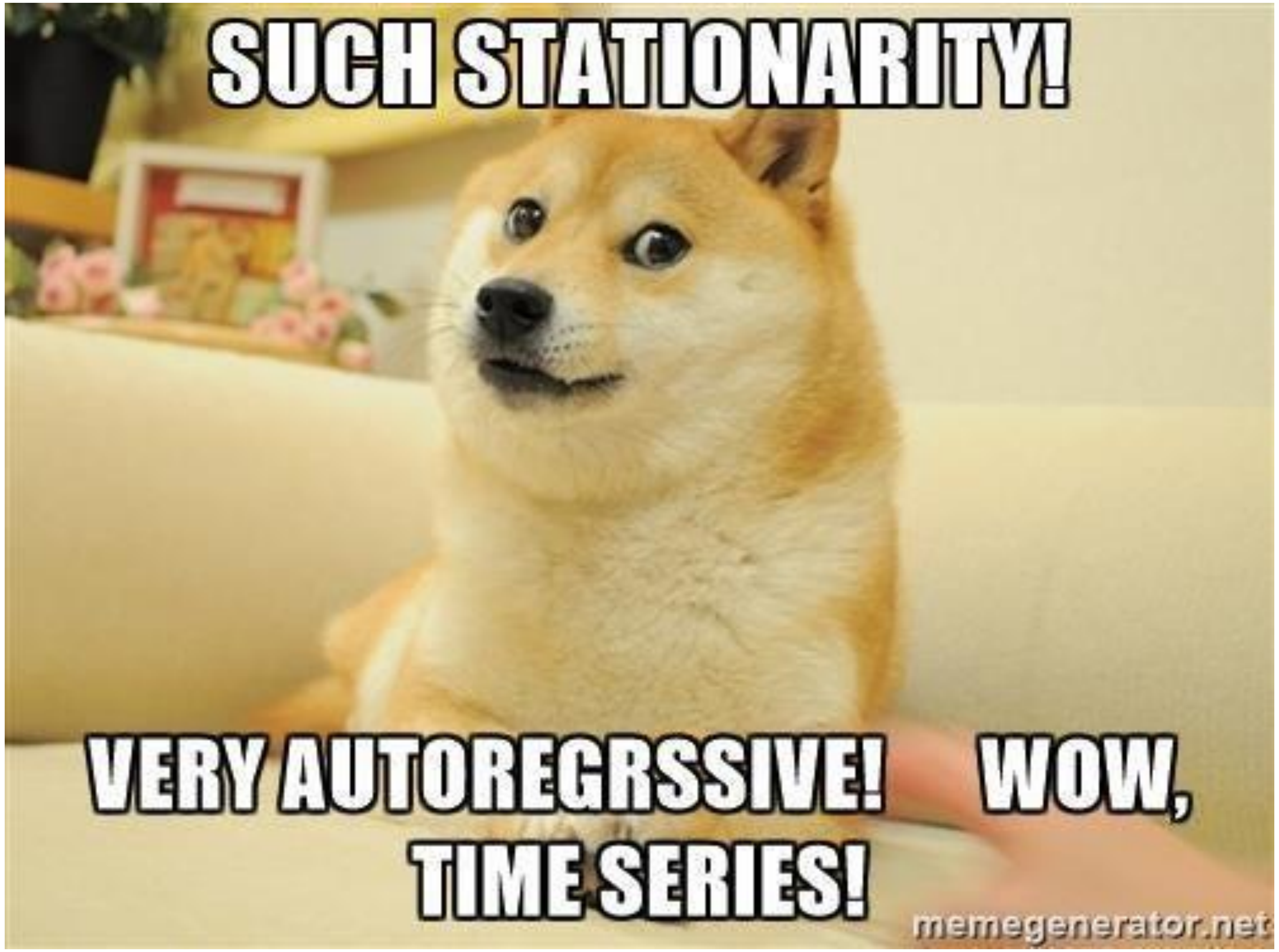
- The dataset used includes some exogenous variables, but the selection might not be comprehensive.
- Models like ARIMA and AutoARIMA assume linear relationships and might not capture complex nonlinear patterns in the data.
- There's a risk that models, especially more complex ones like AutoARIMA with multiple exogenous variables, might overfit to the training data, thus performing poorly on unseen data.
- While models like Holt-Winters and ARIMA can capture seasonality and trend, they may not be sufficient for datasets with more complex seasonal patterns or multiple seasonal cycles.

Potential Areas for Improvement or Further Analysis

- Incorporate More Relevant Exogenous Variables.
- Updating the dataset to include more recent data can help capture current trends and improve model robustness.
- Employ more advanced models like Long Short-Term Memory (LSTM) networks, Prophet by Facebook, or other deep learning techniques that can capture complex nonlinear relationships and multiple seasonalities.



Thank You


SUCH STATIONARITY!
VERY AUTOREGRSSIVE!
WOW, TIME SERIES!
memegenerator.net