Project 3

CPI Prediction and Analysis

Erfan Hajiamouassar

Dan Lord

Scott Phillips

TOC

Project Overview

Using the LSTM machine learning model, we will determine how closely the Consumer Price Index can be predicted during two of the most turbulent periods in recent history (2008-2010 and 2020-2022), and if we can predict individual CPI components (i.e. energy, food, shelter).

We will also use sentiment analysis to validate/contradict our predictions and/or recommendations.

Data: Consumer Price Index and Components

*Percent change - monthly

Components

- Food - General
- Milk, fresh, whole, fortified per gallon
- Eggs, grade A, large, per doz
- Sugar, white
- Flour
- Malt beverages
- Avg Energy
- Fuel oil #2, per gallon
- Electricity per KWH
- Automotive diesel fuel, per gallon
- Gasoline, all types per gallon
- Utility (piped) gas per therms
- Shelter
- Prescription Drugs
- Medical Care

Data

Source: US Bureau of Labor Statistics (BLS)

Process:

- Identify data source - available dates
- Identify relevant CPI components
- Utilize the BLS API and Excel extracts to download monthly data on the Consumer Price Index (CPI) and several of the CPI components
- Create dataframes and concatenate data
- Fill NaN values
- Create test samples: size of 16 and window size 30
- Create train and test data
- 4 layers, 800 epochs, 18MM parameters
- Run model

Challenges:

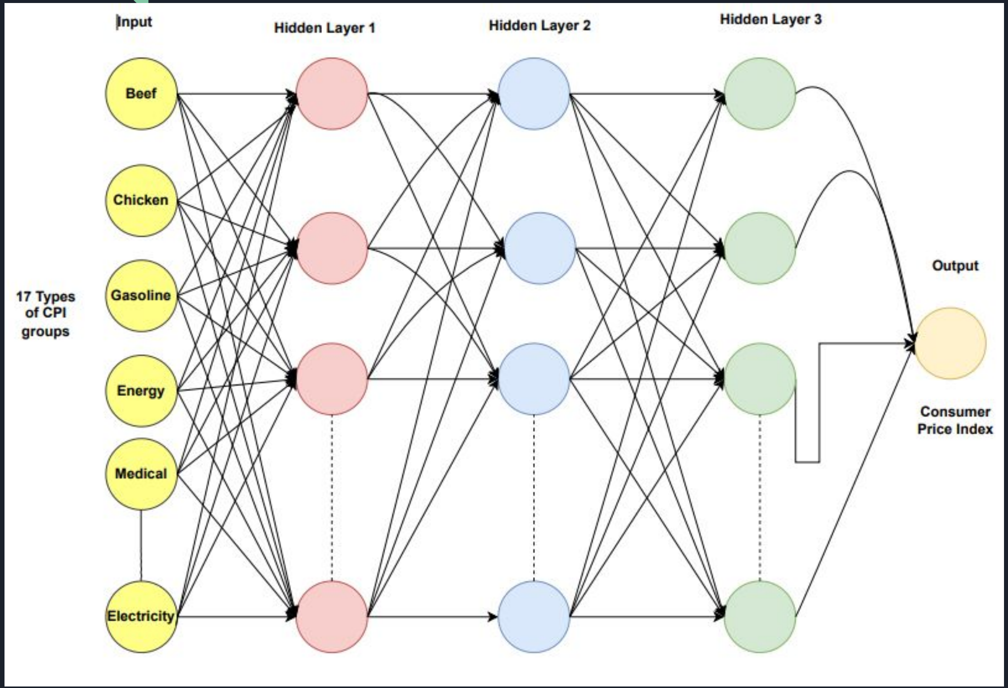
- Data for CPI components - start dates, and missing values
- Many… many iterations tuning the model

Model

Long Short Term Memory (LSTM):

- Based on the python programming languages and multiple libraries such as tensorflow, keras, scikit, pandas and numpy for deep learning algorithms.
- The LSTM model had a set up of multiple layers based on the different CPI that we pulled from the Bureau of Labor Statistics.
- The CPI datasets were used as the input variables, which then would be allocated to a hidden layer consisting of various neurons to be moved to the output dense layer, which will have 1 variable.

LSTM Model - CPI



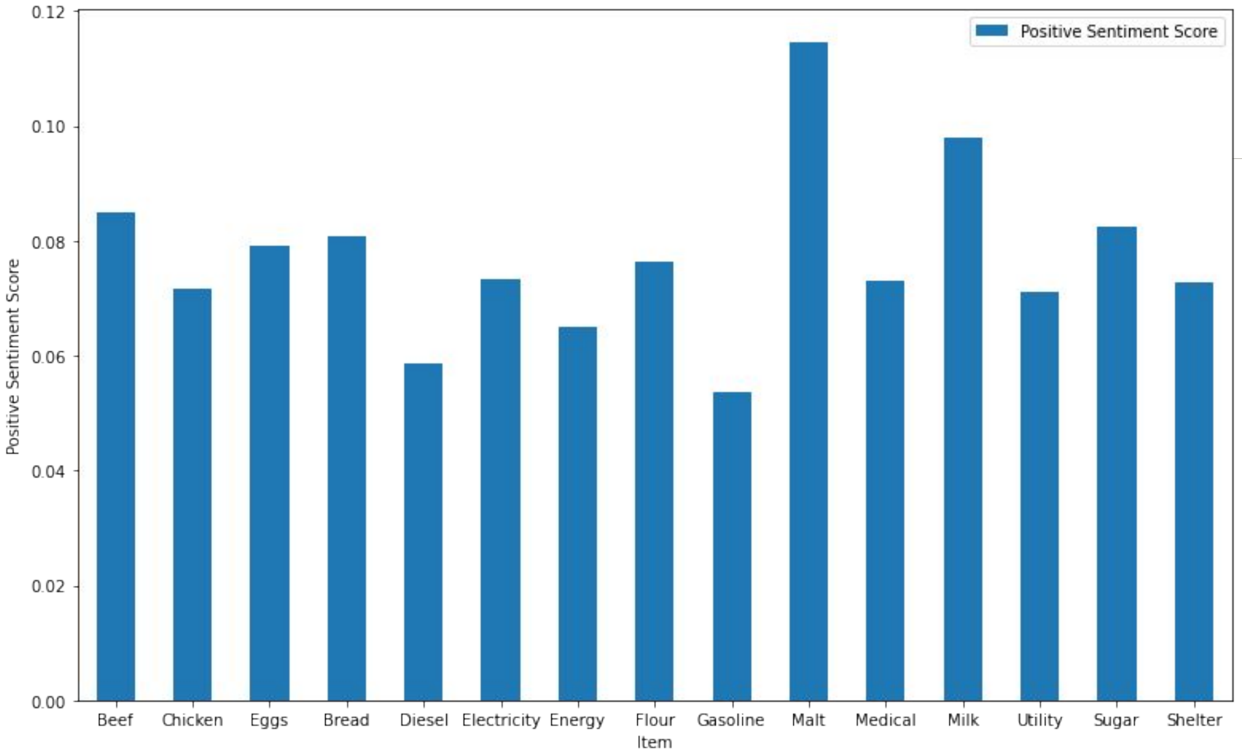
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 800)	2566400
dropout (Dropout)	(None, 30, 800)	0
lstm_1 (LSTM)	(None, 30, 800)	5123200
dropout_1 (Dropout)	(None, 30, 800)	0
lstm_2 (LSTM)	(None, 30, 800)	5123200
dropout_2 (Dropout)	(None, 30, 800)	0
lstm_3 (LSTM)	(None, 800)	5123200
dropout_3 (Dropout)	(None, 800)	0
dense (Dense)	(None, 1)	801

Total params: 17,936,801
Trainable params: 17,936,801
Non-trainable params: 0

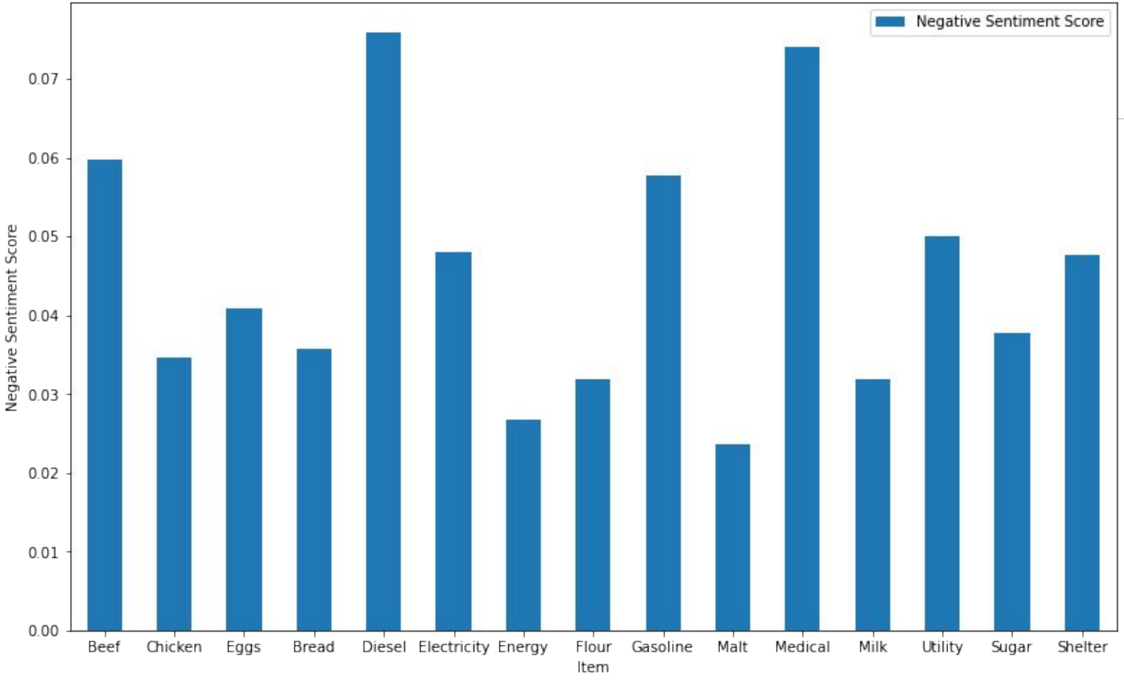
Sentiment Analysis

- Based on the python programming languages and multiple libraries such as newsapi and nltk for deep learning algorithms.
- The sentiment model utilized multiple items included based on the different CPI categories that we pulled from the Bureau of Labor Statistics.
- The categories were used as the input variables, which would be used as headlines for newsapi. The sentiment scores are then concatenated into specific tables for positive and negative sentiments

Sentiment - Positive



Sentiment - Negative



Tableau

Demo

Conclusions / Predictions

LSTM model: A good predictor of CPI and the various components during most of the period from 1990-current. However, CPI predictions were much more accurate during 2008-2010 than during 2020-current. Assumption is the pandemic and subsequent monetary easing and now tightening have not been seen since before this period. Also, certain components such as CPI and staples price value changes in certain months is something that our LSTM model cannot predict at the moment. The model wants to smooth out the data more than reality

Notes:

- We could have developed a basic Recurrent Neural Network(RNN), however the necessary input data and gate function that is associated with LSTM made it a better choice of practice for this particular calculation of CPI prediction.
- We only used 16 components (out of 200) for the CPI model. Using additional components could have yielded better results, but we were limited in processing power and time.

Conclusions / Predictions

Sentiment analysis: Beef, diesel, gas, and medical have the most negative sentiments, as we would have predicted based on current market prices. These four categories have caused direct hardships to average income households because of the pandemic and inflation.

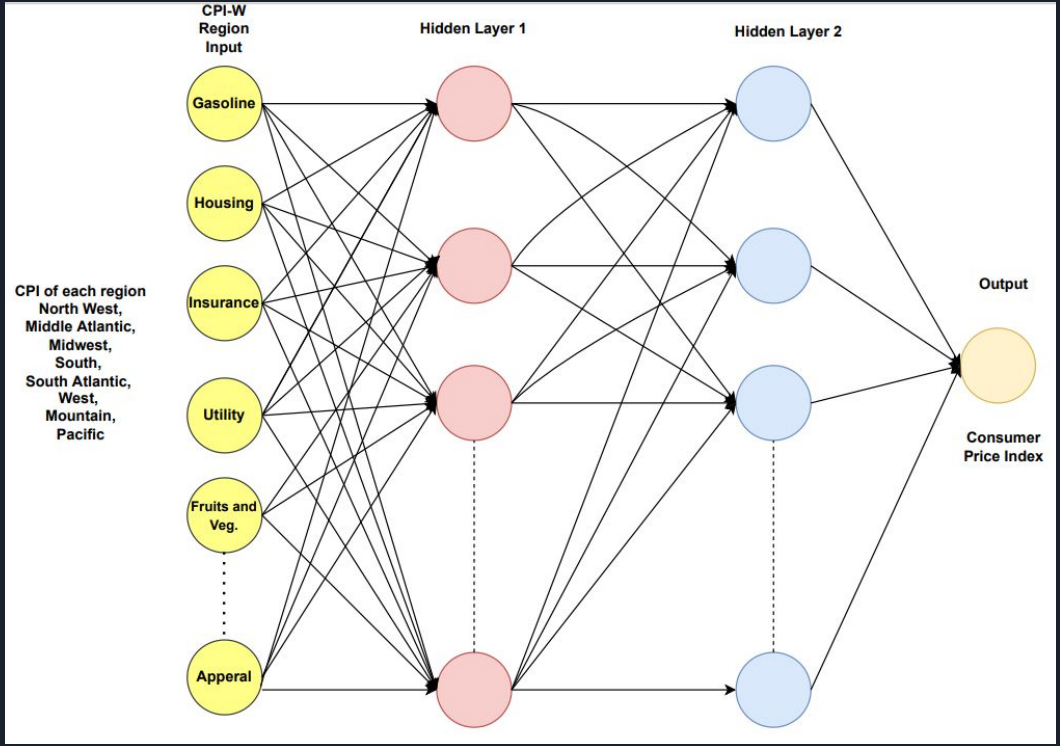
Milk and malt have the most positive sentiments. Beef also has the third highest positive sentiment. The cattle industry is slowly making improvements towards a more competitive market.

Possible improvement: Given a longer period to analyze news, it would be interesting to compare monthly graphs to see how sentiment improves or decreases over time to get more of a trend.

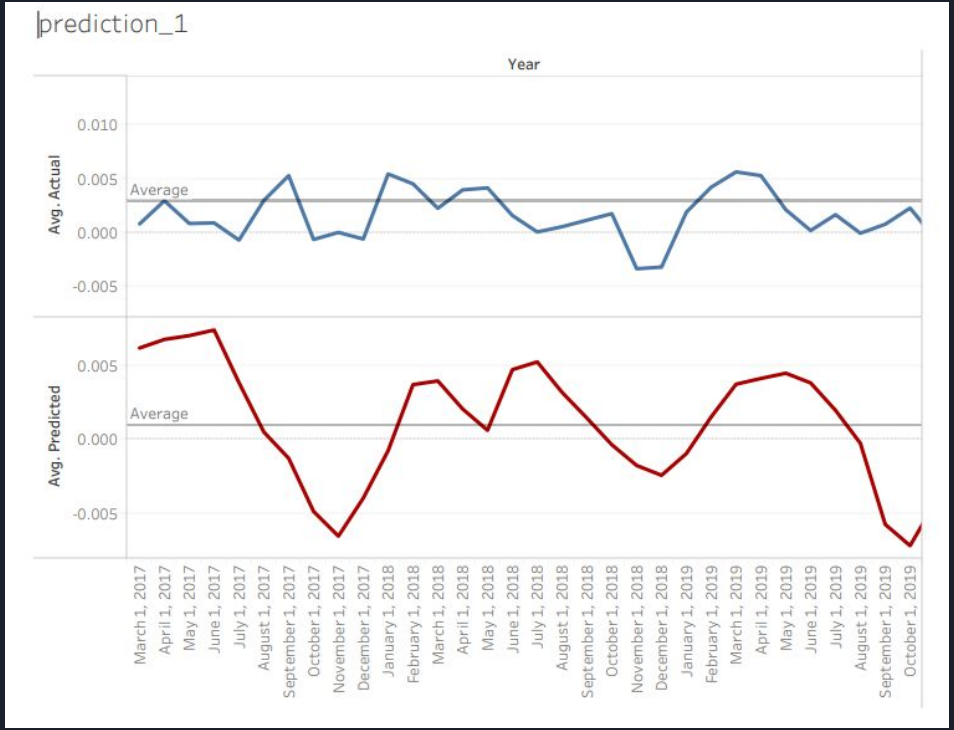
Project Challenges

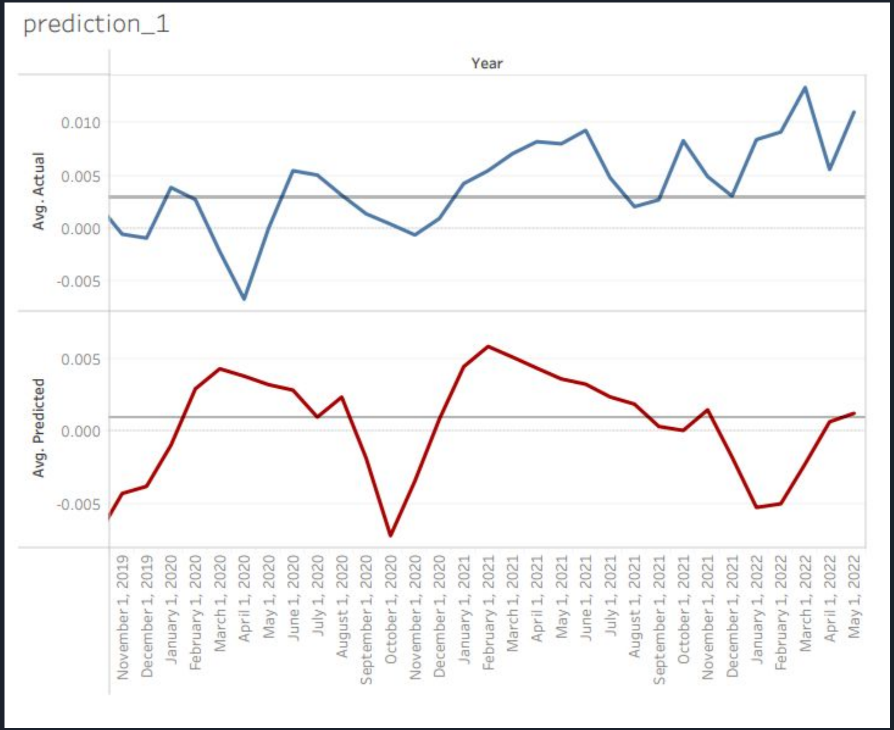
- Data: API vs. Excel download
- Modeling took many iterations, each time required a lot of time
- News API did not work correctly the first time, needed to recode the API
- NewsAPI and Tensorflow require previous Python versions (changing between environments across all hardware)
- Tableau: need to pay for enhanced features

LSTM Model - Regional CPI

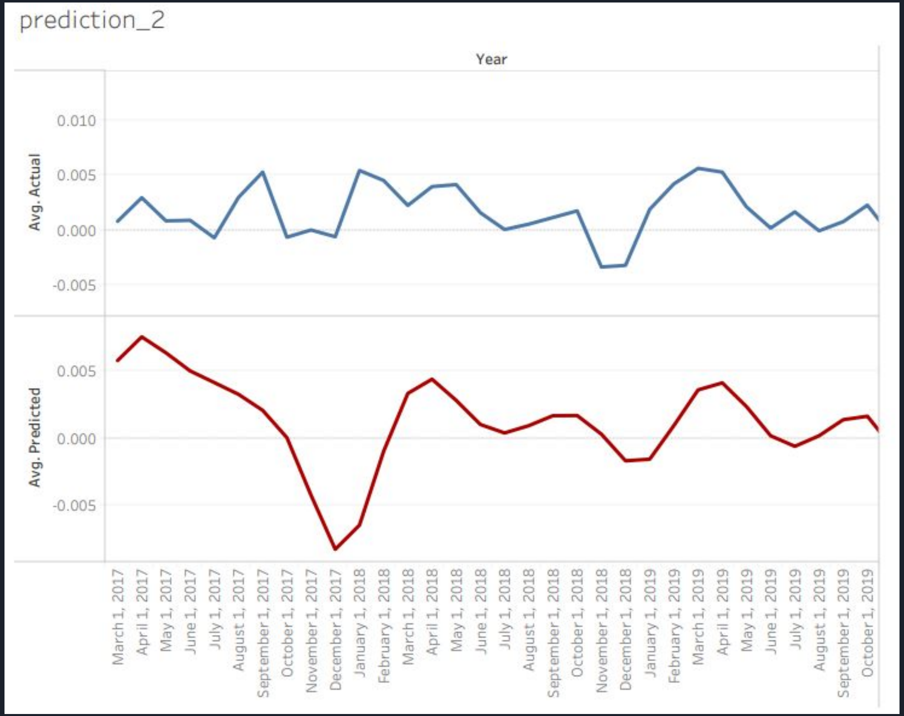


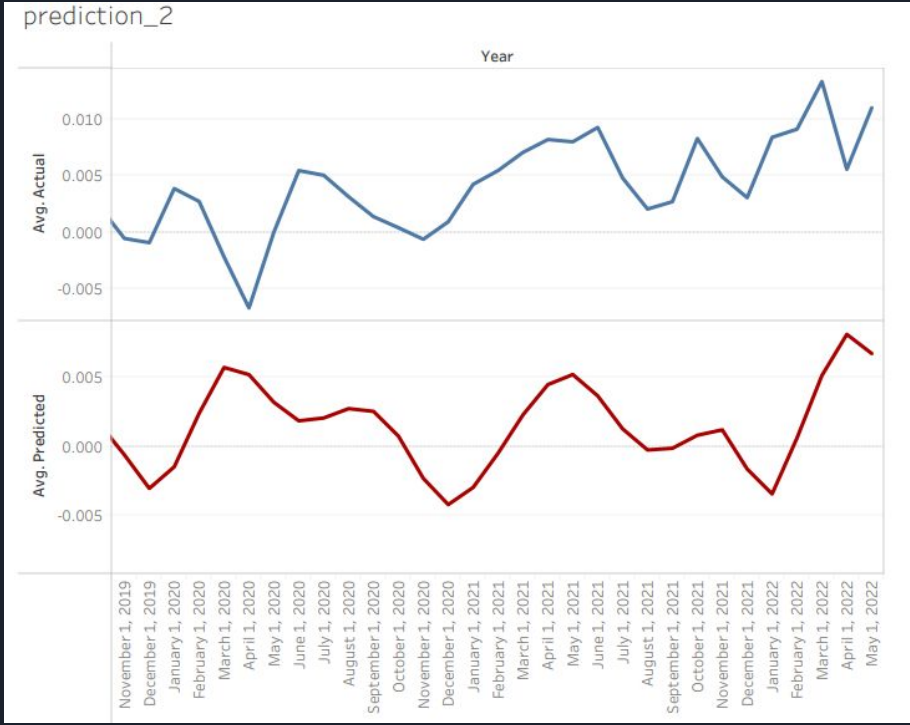
CPI - Prediction 1



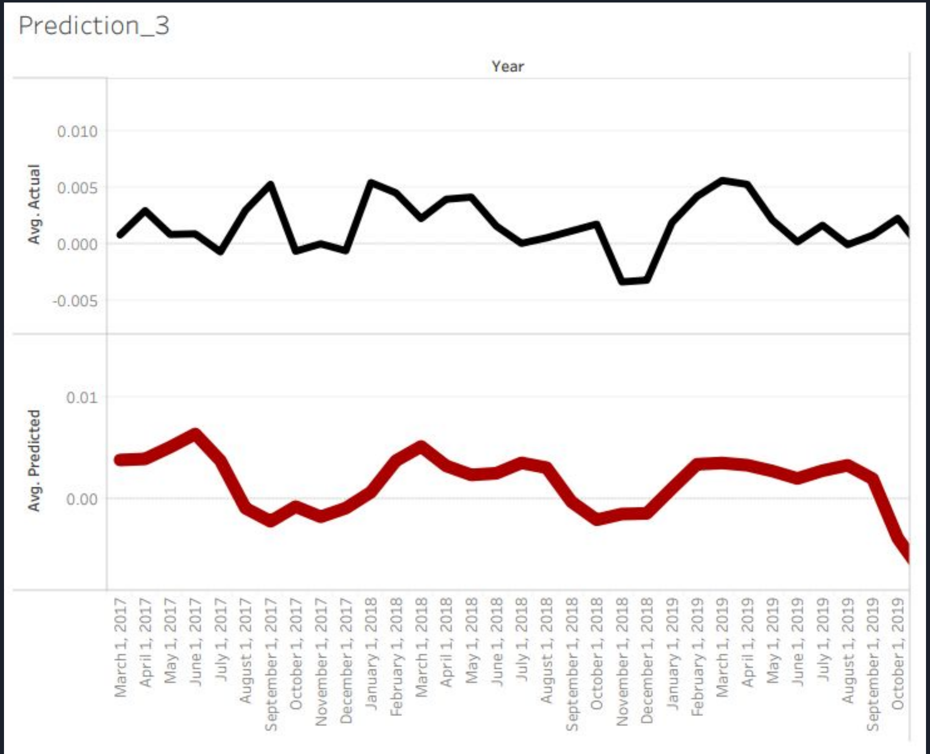


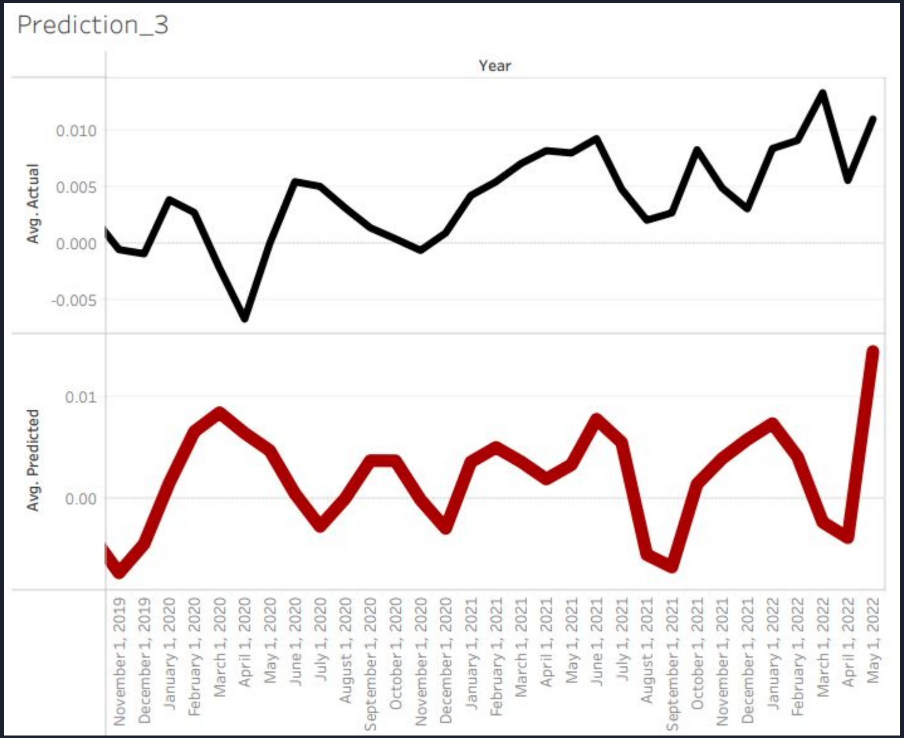
CPI - Prediction 2



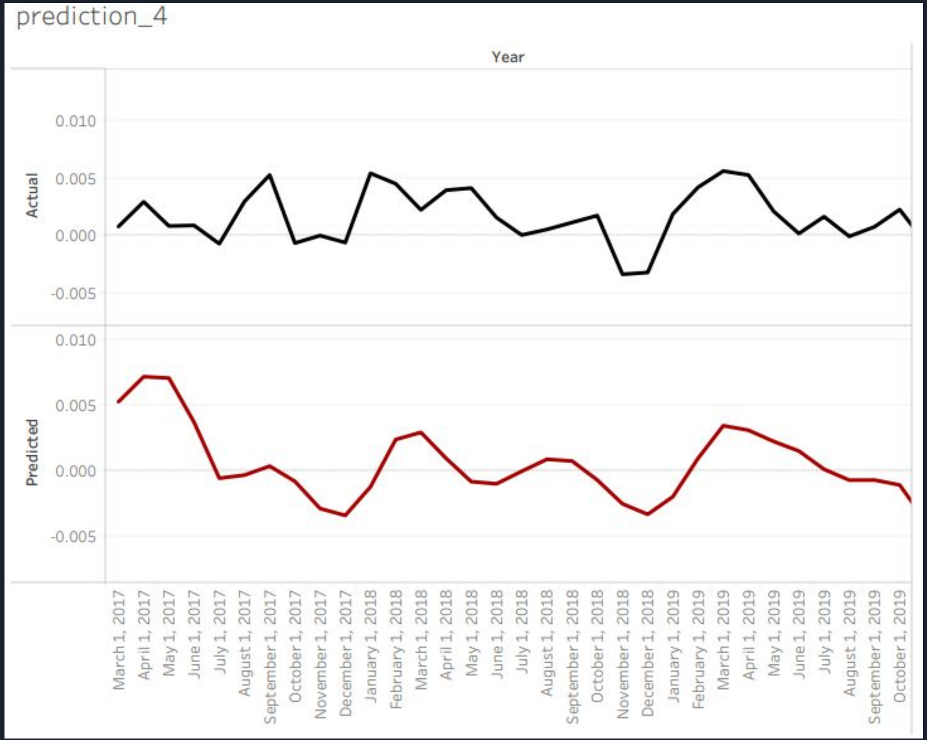


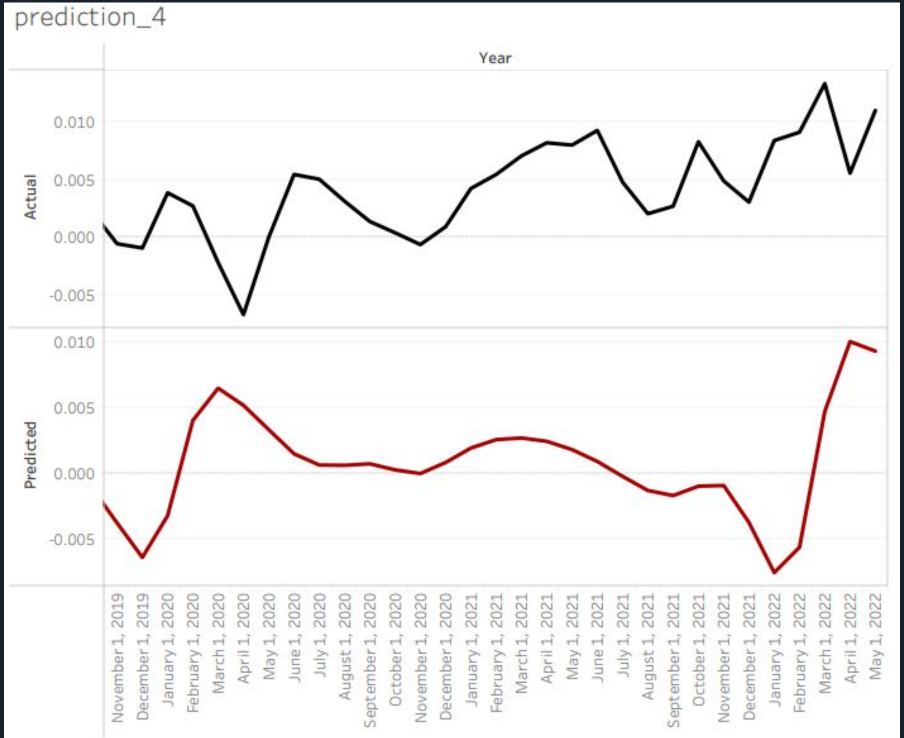
CPI - Prediction 3



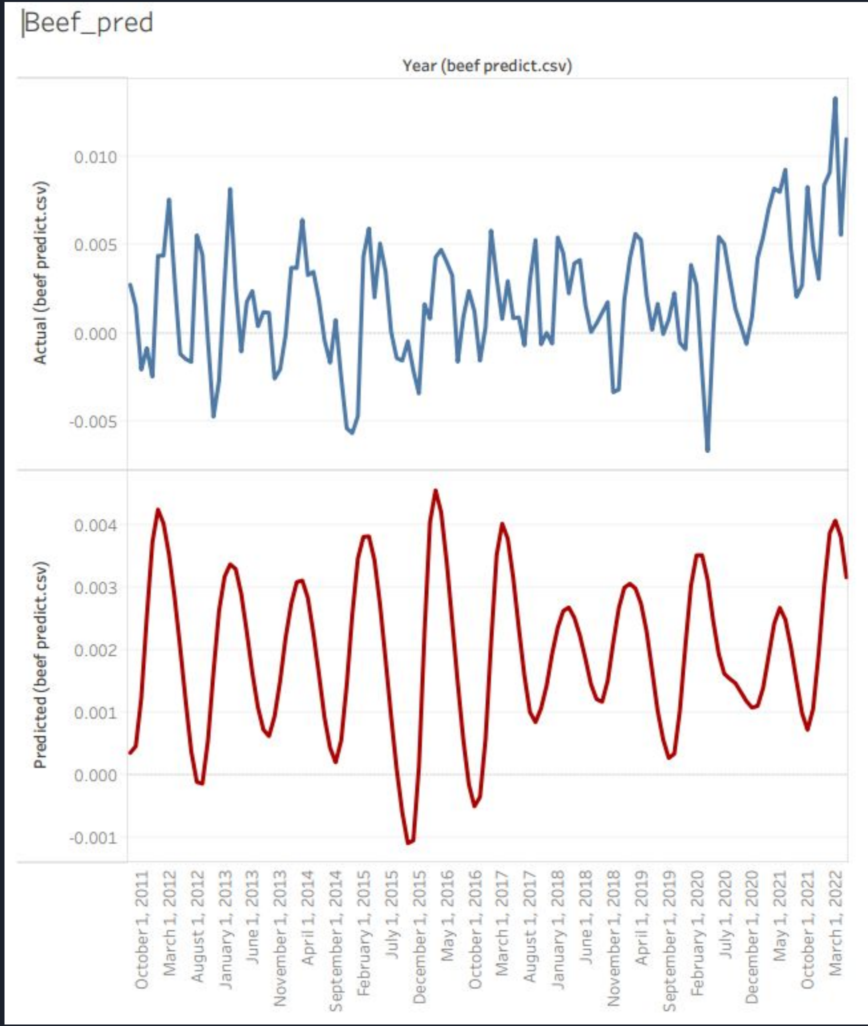


CPI - Prediction 4

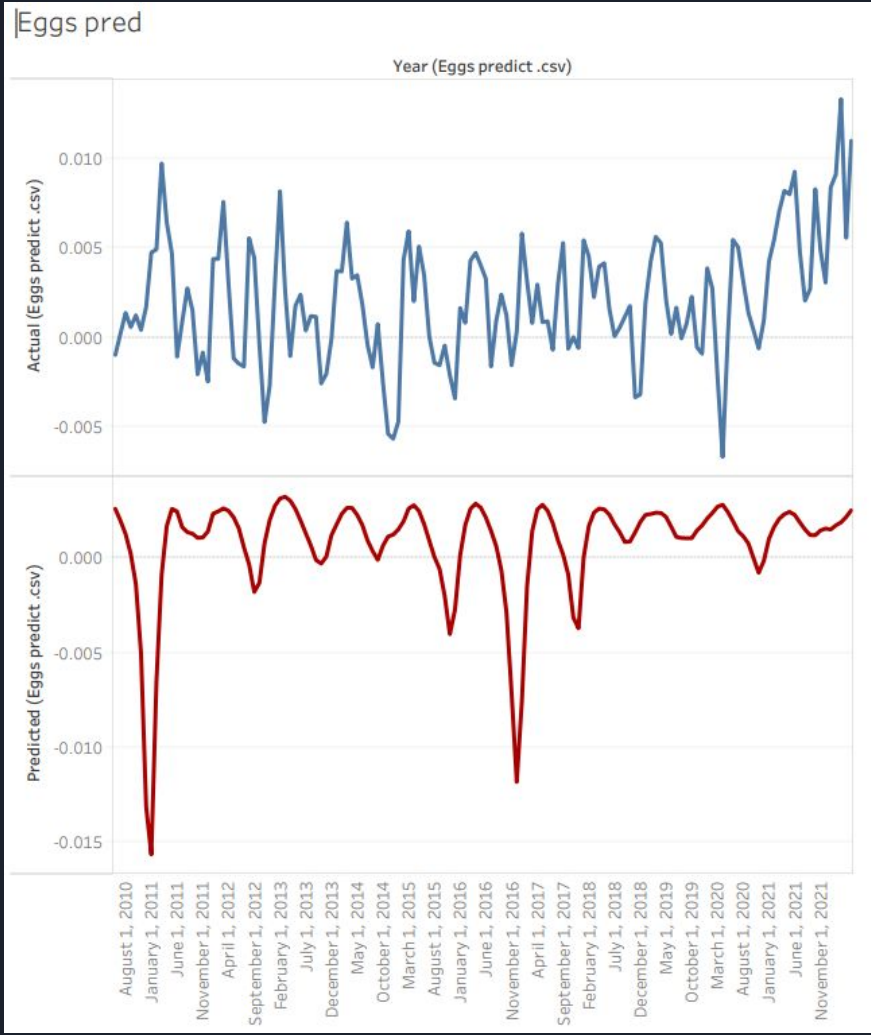




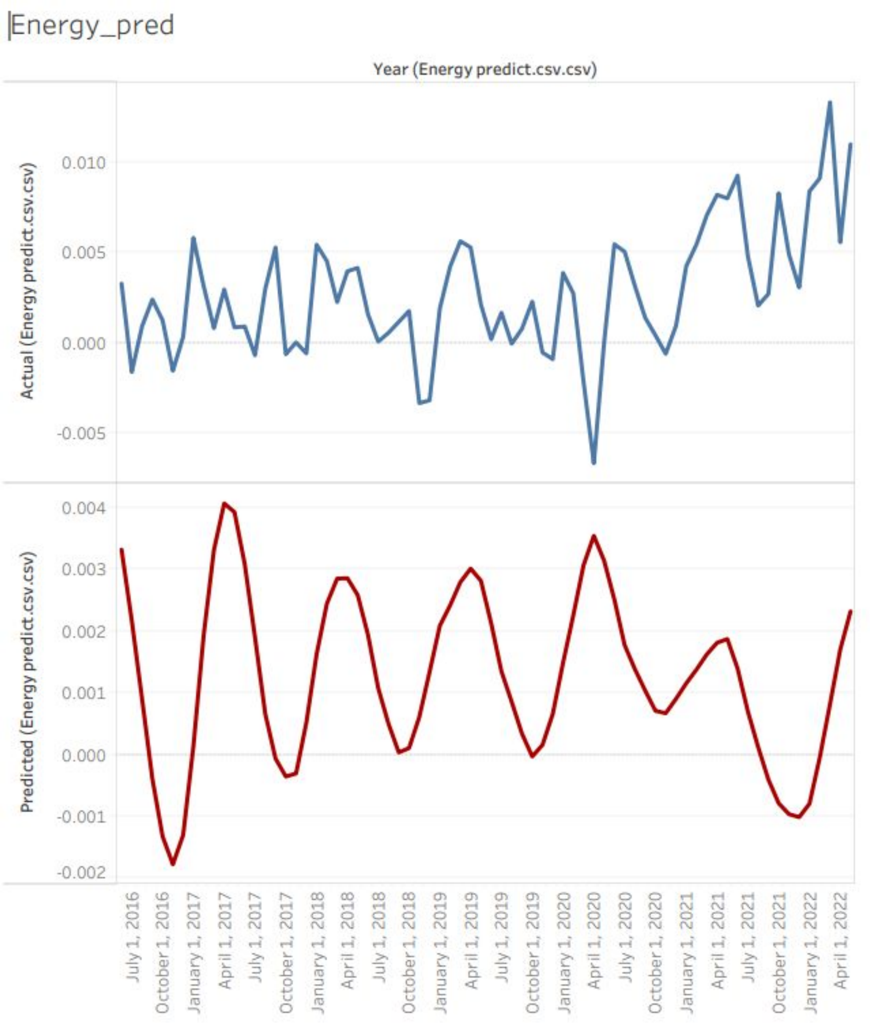
CPI Beef - Prediction



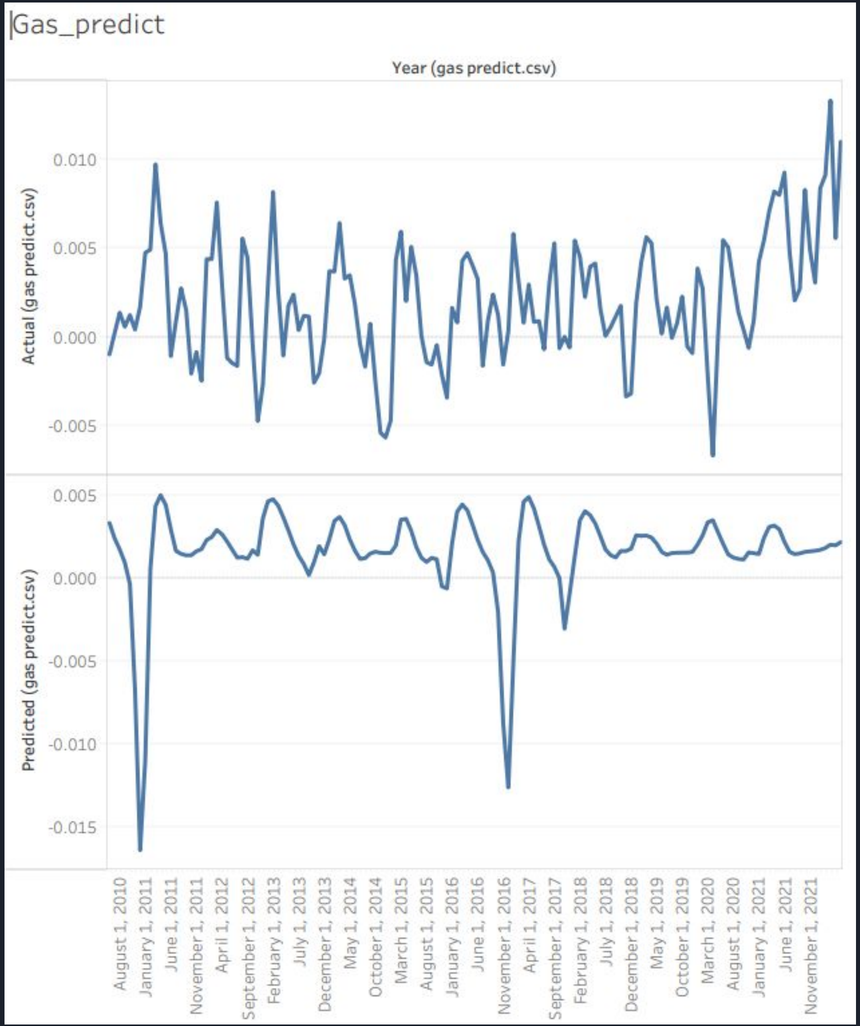
CPI Eggs Prediction



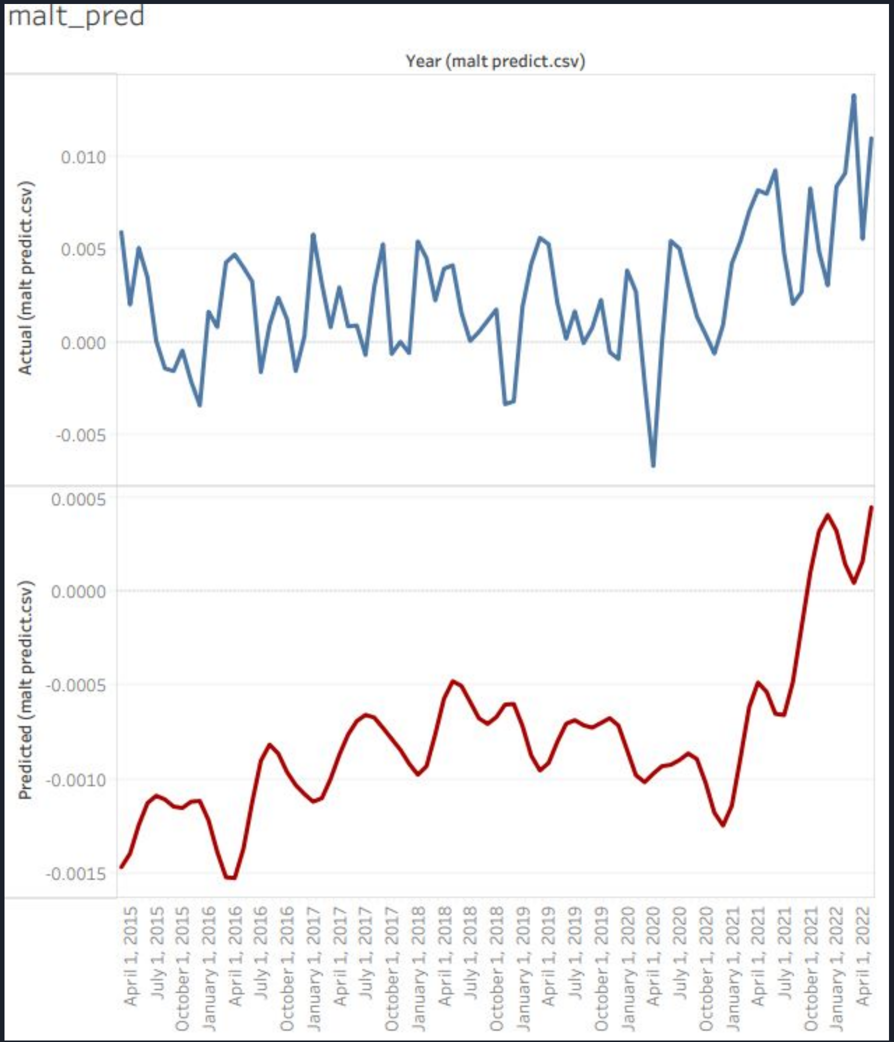
CPI Energy Prediction



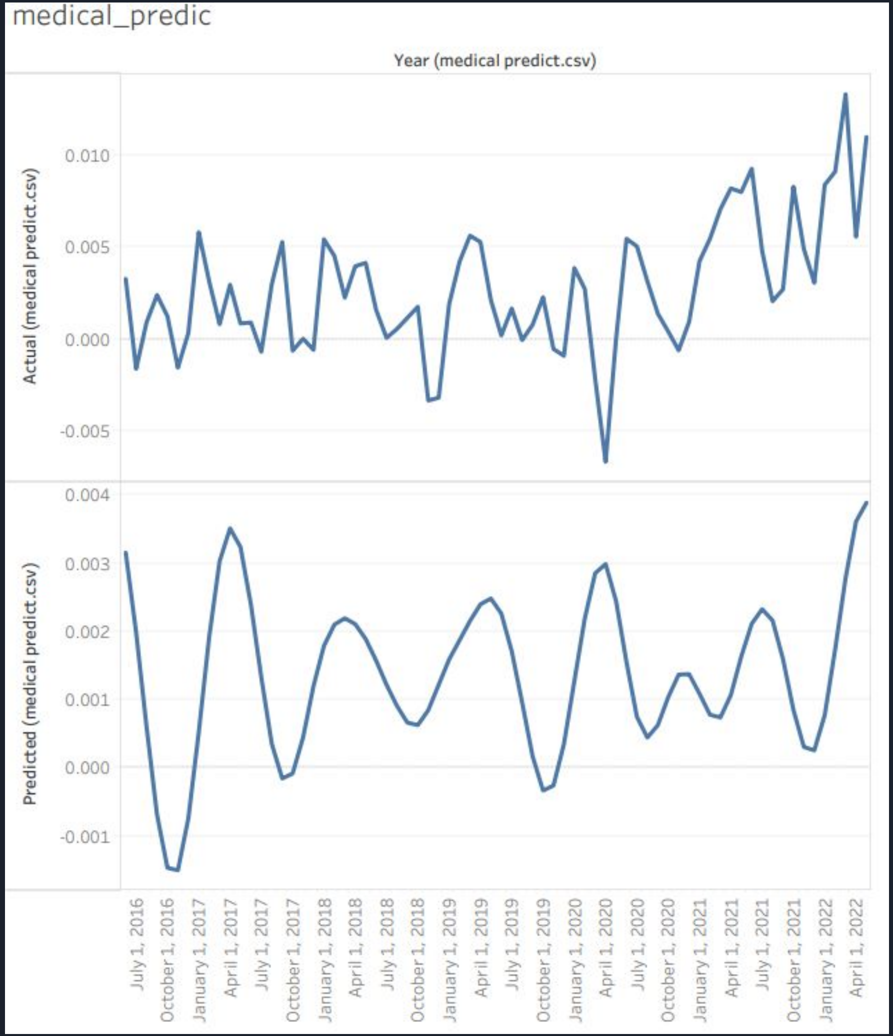
CPI Gas Prediction



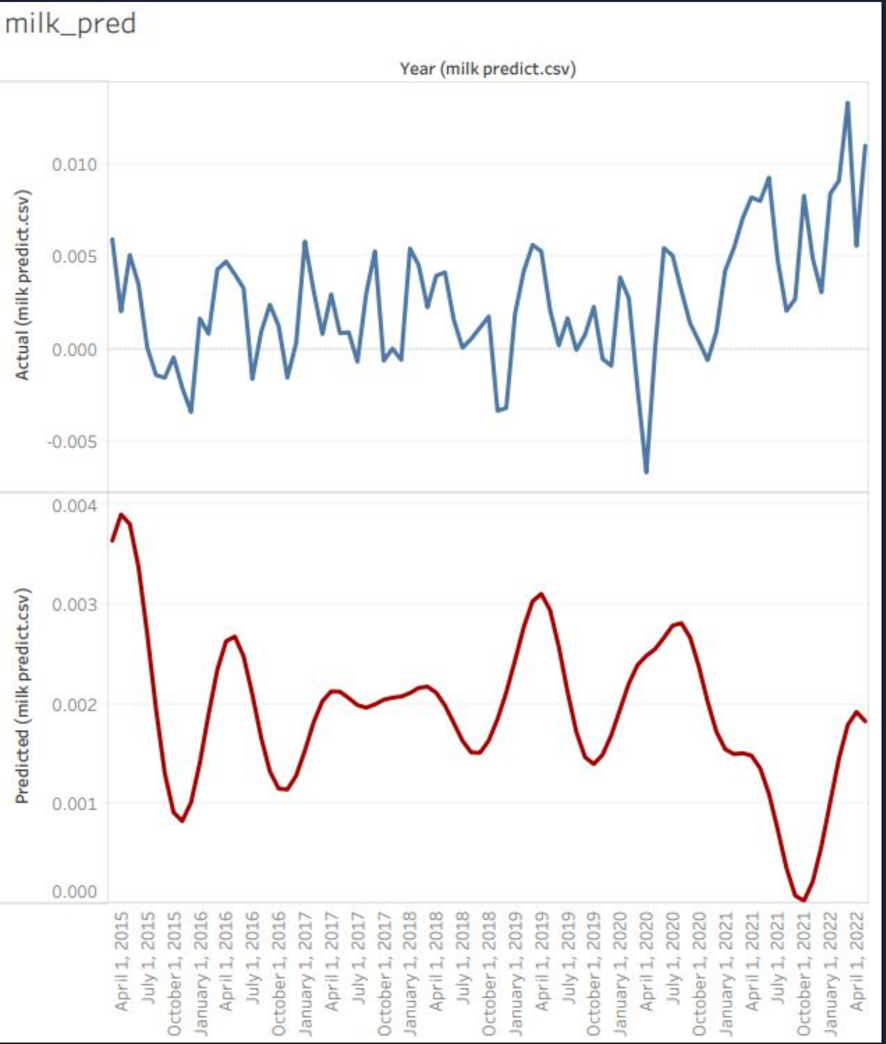
CPI Malt Prediction



CPI Medical Prediction



CPI Milk Prediction



Resources

Class Notes, videos, code, homework code

Consumer price index prediction using Long Short Term Memory (LSTM) based cloud computing - IOPscience

Predicting consumer price index cities and districts in East Java with the gaussian-radial basis function kernel - IOPscience

Acquire and Visualize US Inflation Data with the BLS API, Python, and Tableau | by Randy Runtsch | Towards Data Science