

## Problem

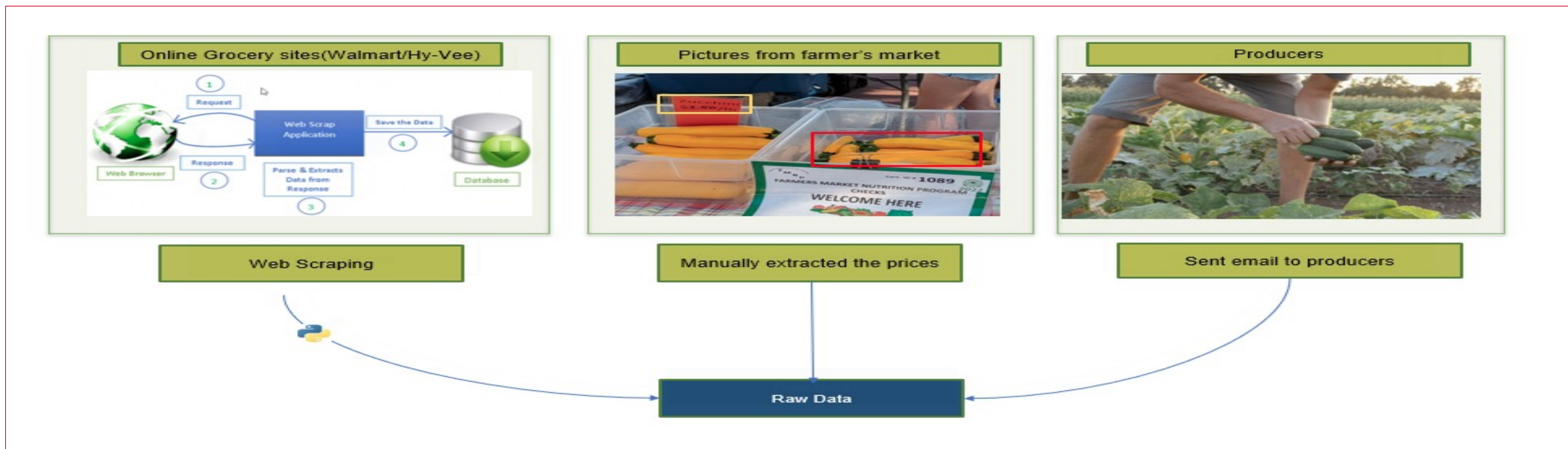
### Background

- The Iowa State Farm Food and Enterprise Development (FEED) is frequently asked for benchmarks on pricing of products both in retail and wholesale spaces
- While data is available from sources like AMS and USDA, there is limited aggregation of sales for these products at the local level
- There is a need for research on the potential sales point for wholesale products when many of the producers are operating in direct-to-consumer retail spaces

### Objective

- Data Discovery: Investigate possibility of scraping data and use Google Trends to identify potential key products by terms being searched(Hits)
- Compare values of collected data with USDA to determine if this helps in better decision making
- Identify potential aspects that could be automated for future data collection
- Investigate potential use of AI to identify impacts to local food markets. For example, weather impacts

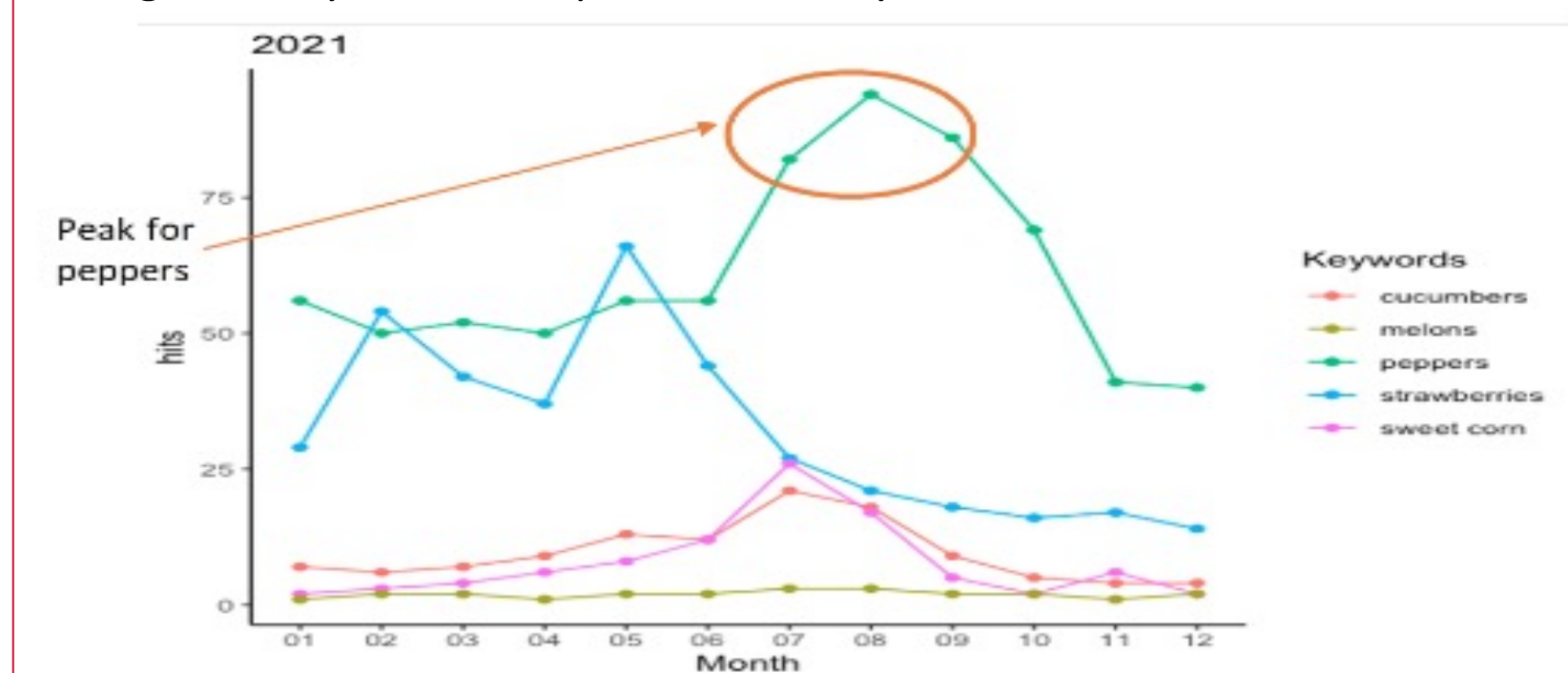
## Data Collection



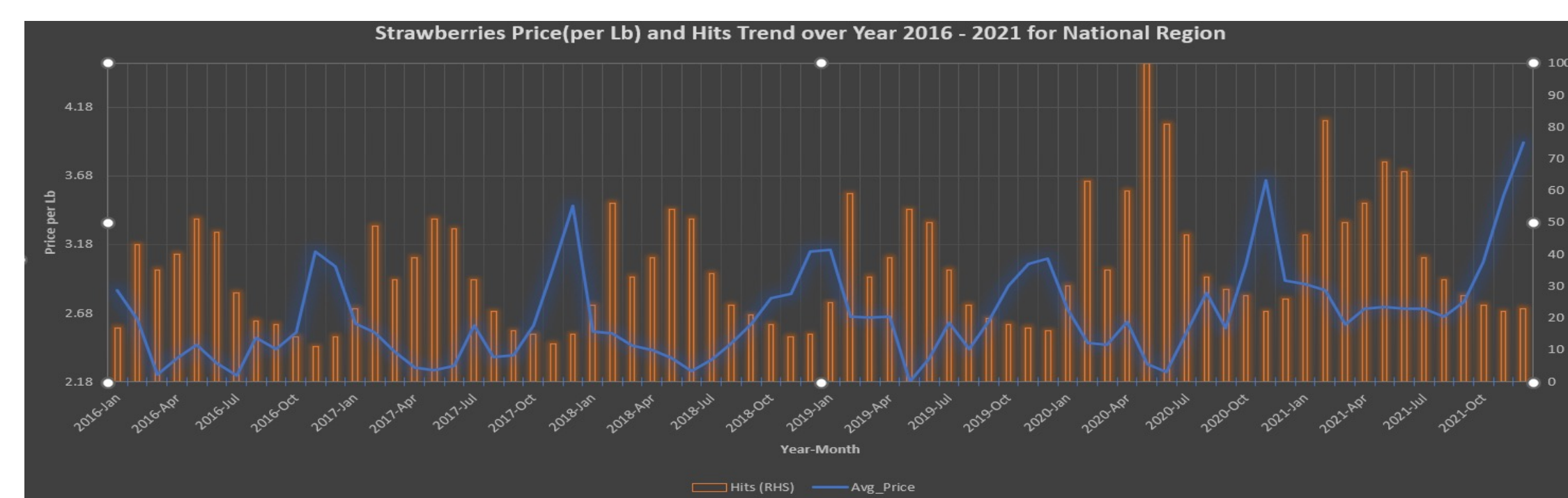
## Approach

### Data Exploration

Google hits by search for products over years

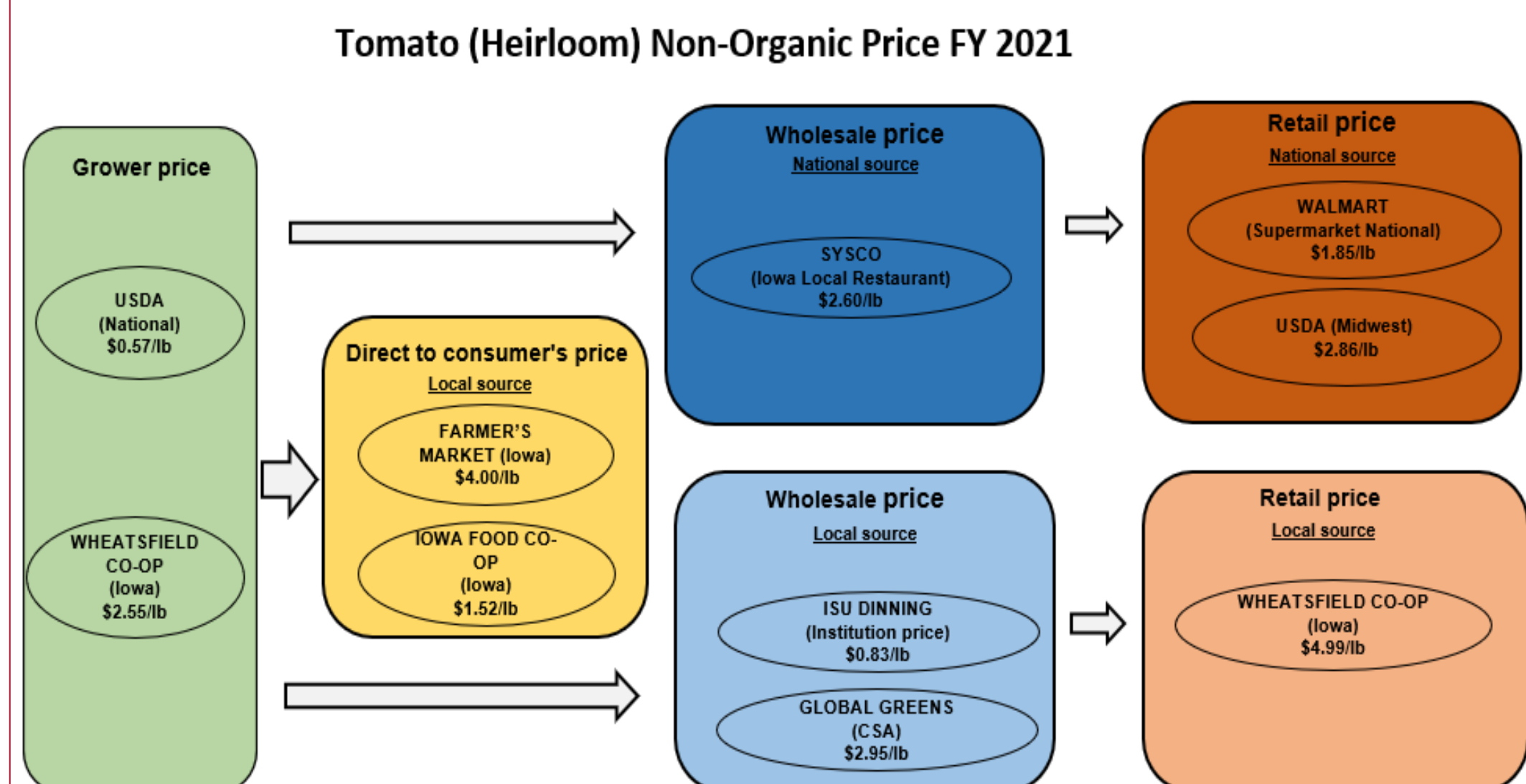


Negative correlation between search trends and prices (Low Price when Google Searches are High)



### Comparison: USDA Vs local/Grocery/Food Hubs

Compare USDA price with the prices from all data sources like Walmart/Institution/Wheatsfield



Price Trend Over Years for Multiple Data Sources

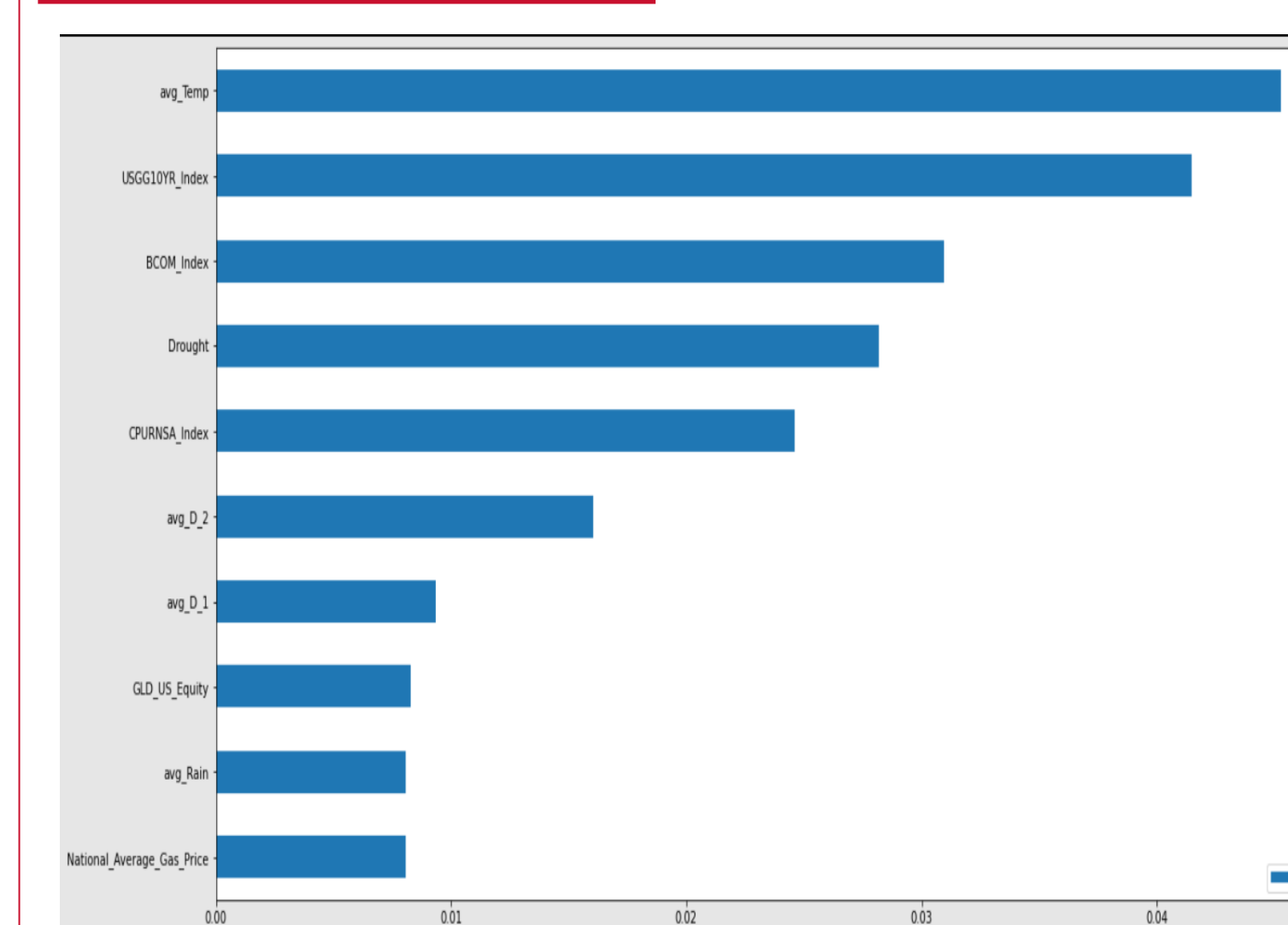


## Result and Analysis

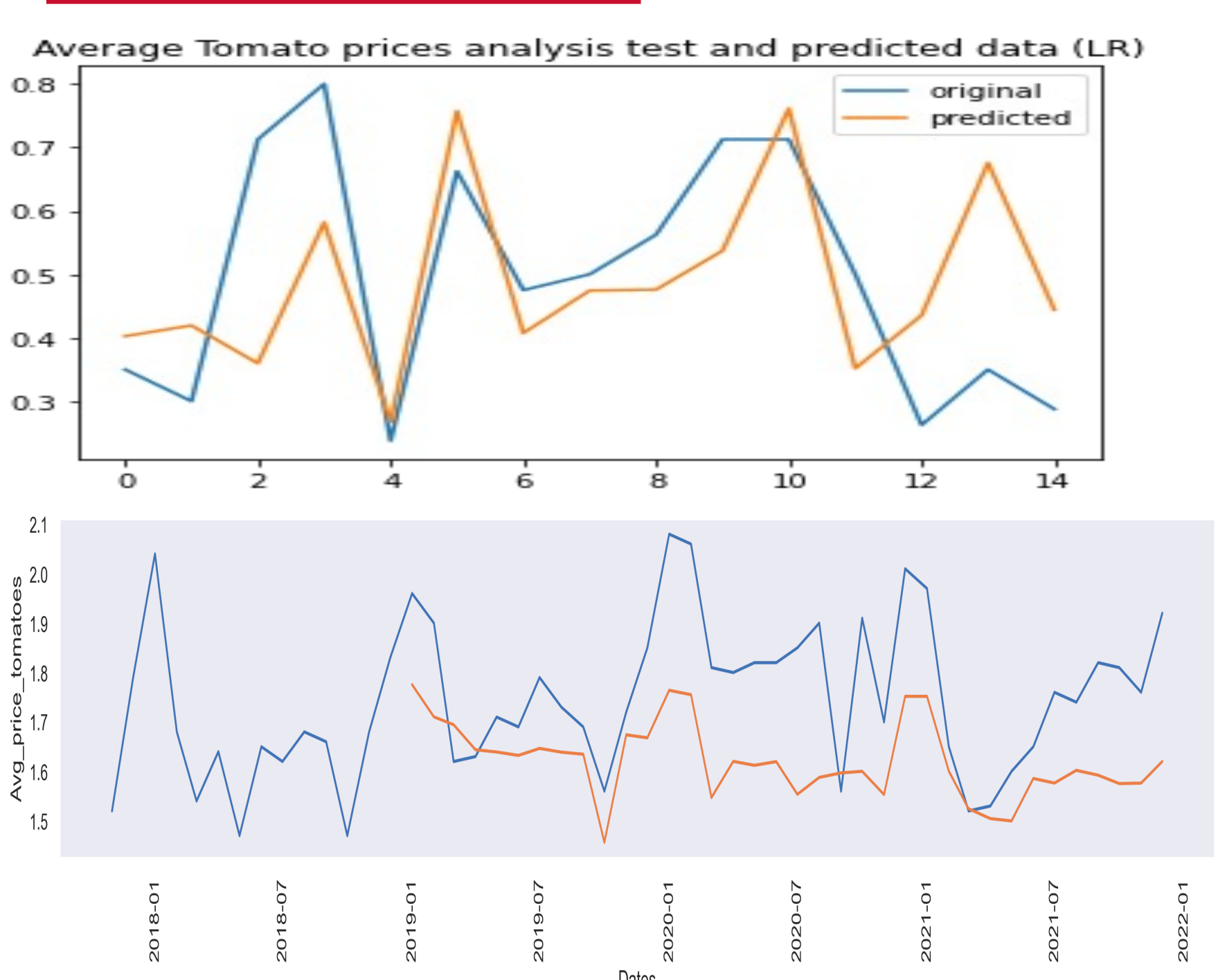
### AI: Predicting Average Price of a Commodity

- Regression and XGBoost models used for prediction (Please refer the codes section for more details)
- Computed average value of each variable at monthly level for years 2016 - 2021. For example: price of a commodity is calculated by taking average of each month for mentioned years
- Indicators used:
  - Historical USDA retail price data of apples and tomatoes in the Midwest region (USDA Market News)
  - Historical precipitation and temperature data from ISU mesonet
  - Drought data in Story County, Iowa (drought.gov)
  - 10 Year US Interest Rate (USO\_US\_Equity)
  - Gold Bullion historical price (GLD\_US\_equity)
  - Crude oil futures (WTI)
  - BCOM Index Bloomberg Commodity Index (BCOM\_Index)

### Top features By the model



### Actual Vs Predicted price graph



### Comparison of model performance metrics (Tomato)

Model	MAE	MSE	RMSE	R^2
Linear Regression	0.133622	0.025319	0.159120	0.257504
XGBoost	0.138326	0.028487	0.168782	0.164596

### Price Markup Percentage compared to Grower Price FY 2021

Products	Grower Price	Direct to Consumer			Wholesale Price			Retail Price		
		IOWA FOOD CO-OP	SYSCO	ISU DINNING	GLOBAL GREENS(CSA)	USDA (Midwest)	WALMART	WHEATSFIELD CO-OP		
Apple	\$0.61	113%	75%	18%	-	97%	111%	228%		
Strawberries	\$1.82	216%	19%	10%	-	46%	24%	174%		
Pears	\$0.28	507%	221%	79%	-	179%	364%	968%		
Tomato	\$0.57	167%	356%	46%	418%	402%	225%	775%		
Green Pepper	\$0.20	325%	500%	130%	325%	335%	210%	600%		

## Future Work, Conclusions, and References

### Where is your money going when you pay \$1 for tomatoes?



### Future Work

- Further parameter tuning and adding more indicators to bring down the margin of error
- One key prospect in the future is to add weather data specific to where that commodity was grown, for example adding historical weather data from Washington to find the correlation in how weather from those producer regions can directly impact the price of that commodity to the consumer
- Use computer vision algorithms to automate data collection by extract prices from pictures(farmer's market)
- Run python scheduler every week to collect the historical prices from websites

### References

- USDA – AMS, National Integrated Drought Information System, Iowa Environmental Mesonet, Farmer's Market Iowa