

Milestone Project 3

Machine Learning and Avocado Prices
Sarah Patterson

Table of Contents

- Key Takeaways
- Research Question
- Data
- EDA
- Modeling
- Evaluation
- Results
- Conclusion



Key Takeaways

Avocado Pricing

- My stakeholders are grocery store owners who need to know what dictates their avocado prices.
- Through my modeling, I was able to show that avocado type is the best indicator of avocado price. This can inform stakeholders on which type of avocado they want to sell going into the future.

Research question:

**What are the most important factors
in determining avocado price?**

Data

The data includes number, type, bag size, date, price, and location of avocado sales in the US from 2015 to 2018.

Data Source : [HassAvocadoBoard.com](https://hassavocado.board.com)
CSV File: [Kaggle.com](https://www.kaggle.com)

Original: 14 columns x 18,249 rows

Cleaned: 13 columns x 14,196 rows

Modeling: 9 columns x 14,196 rows



Data

Unnamed: 0		Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Bags	type	year	region
0	0	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	8696.87	8603.62	93.25	0.0	conventional	2015	Albany
1	1	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	9505.56	9408.07	97.49	0.0	conventional	2015	Albany
2	2	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	8145.35	8042.21	103.14	0.0	conventional	2015	Albany
3	3	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	5811.16	5677.40	133.76	0.0	conventional	2015	Albany
4	4	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	6183.95	5986.26	197.69	0.0	conventional	2015	Albany

	4046	4225	4770	Small Bags	Large Bags	XLarge Bags	type	year	region
0	1036.74	54454.85	48.16	8603.62	93.25	0.0	conventional	2015	Albany
1	674.28	44638.81	58.33	9408.07	97.49	0.0	conventional	2015	Albany
2	794.70	109149.67	130.50	8042.21	103.14	0.0	conventional	2015	Albany
3	1132.00	71976.41	72.58	5677.40	133.76	0.0	conventional	2015	Albany
4	941.48	43838.39	75.78	5986.26	197.69	0.0	conventional	2015	Albany

Exploratory Data Analysis (EDA)

To answer the research question, the Average Price is my target variable. It is a numerical and continuous feature so I will need to use Regression models to run the data.

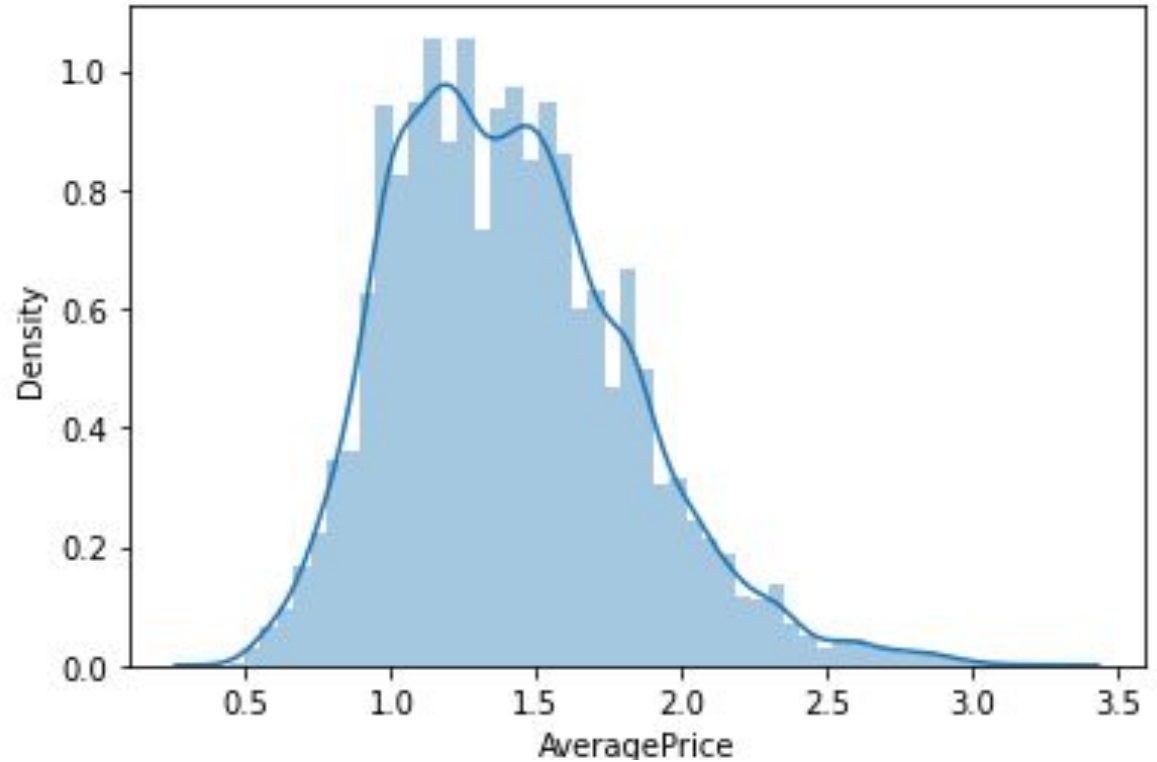
I chose a few models but found the best was the KNeighborsRegressor model with the highest accuracy score

Before running the Machine Learning Models, I wanted to visualize the data to gain better understanding...

Distribution of Average Price

(Price is per one avocado)

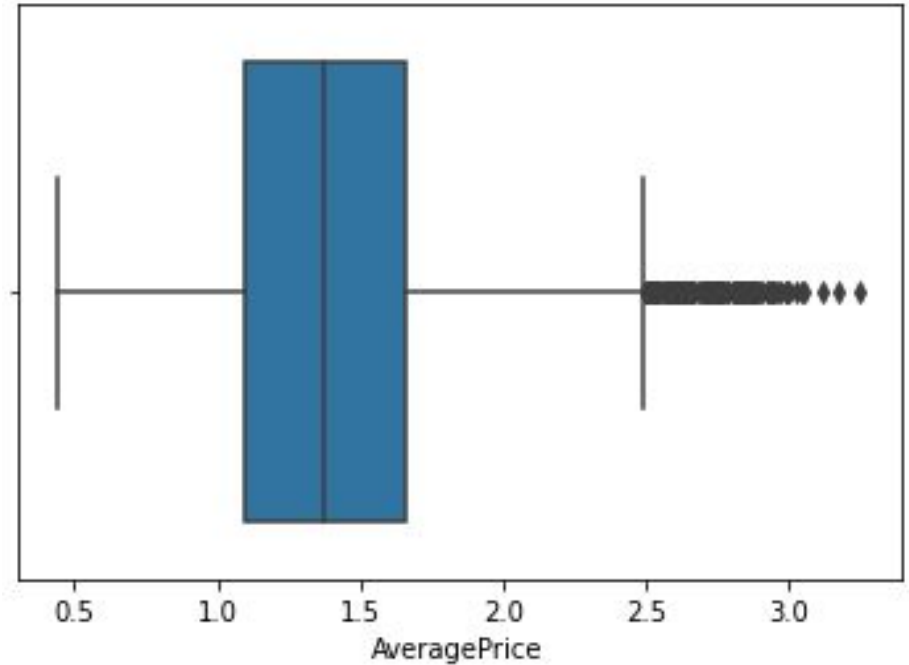
This helps visualize the skewness and kurtosis



Box Plot of Average Price

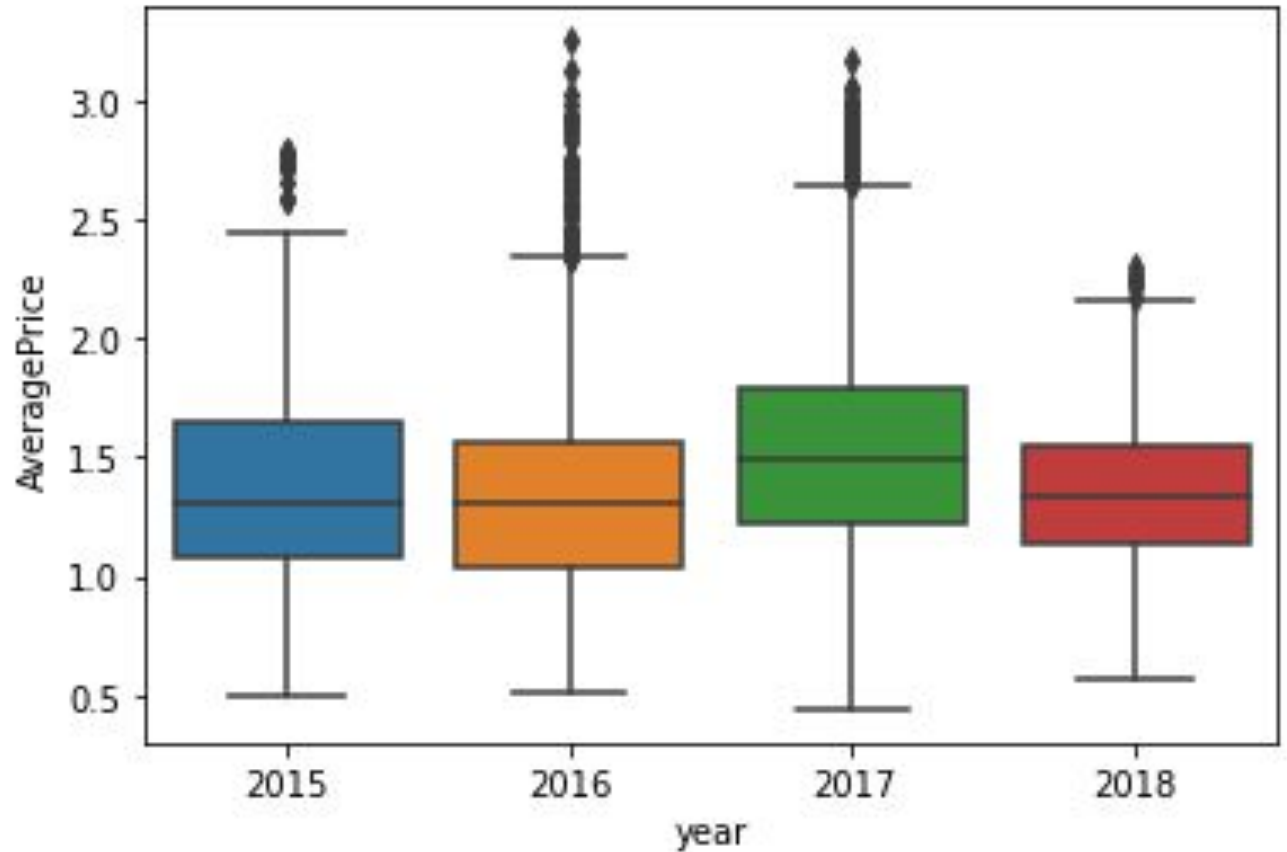
(Price is per one
avocado)

Helps visualize range
of data and outliers.



Boxplot by year of avocado price

This is my favorite visual because it is one of the most informative graphs as it sums the general trends of prices by year.



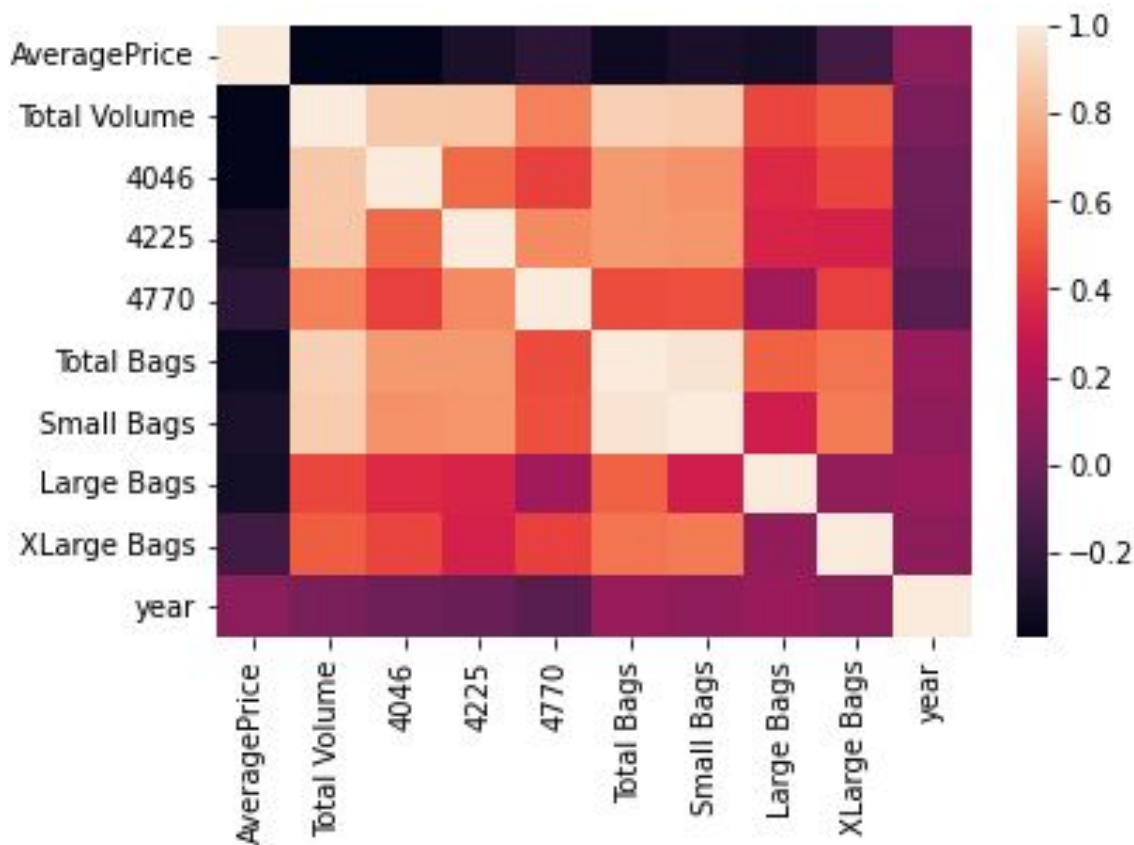
Heat Map of Dataset

Only includes
numerical features

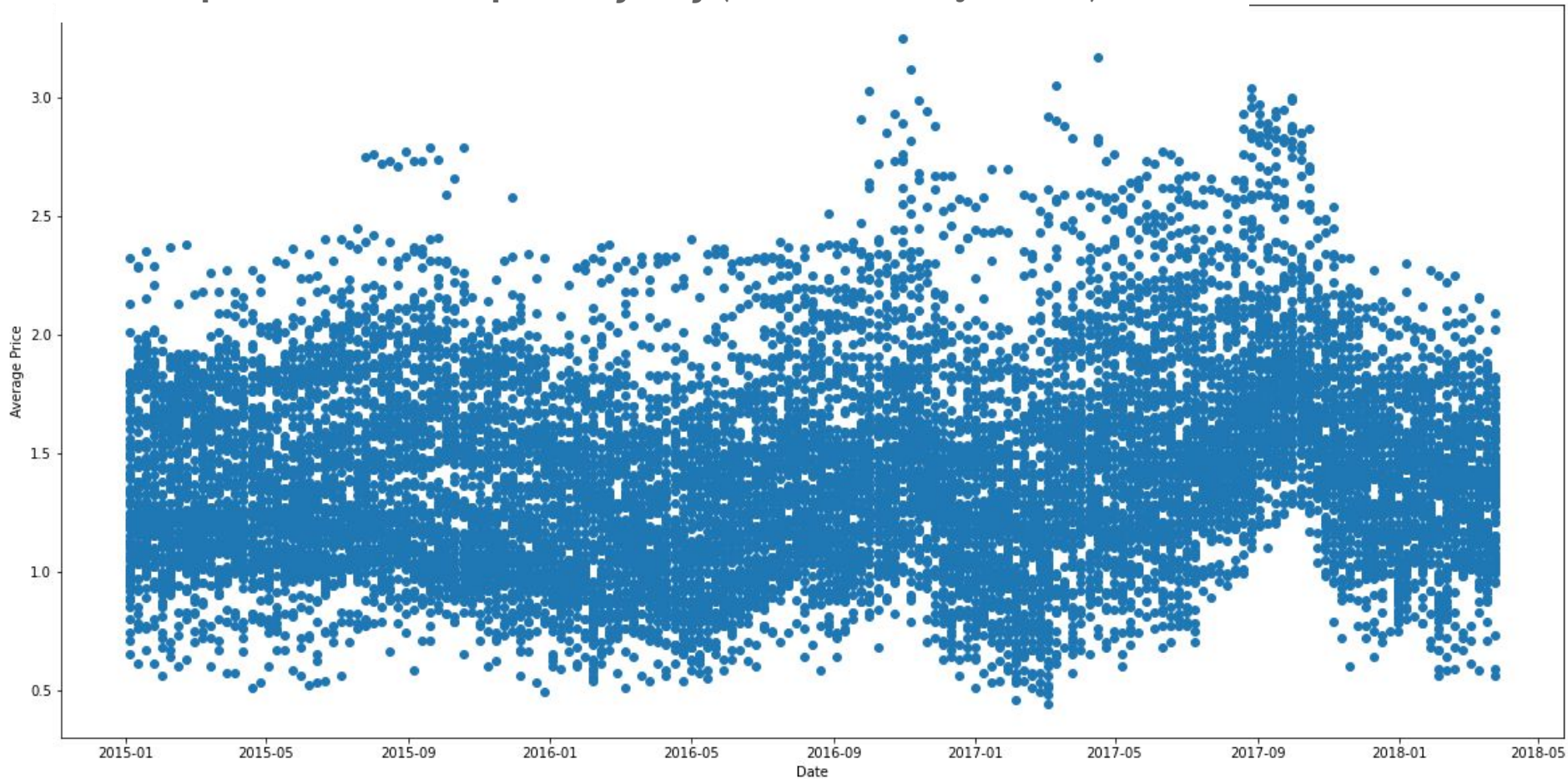
1 = strong correlation

0 = no correlation

-1 = negative correlation



Scatterplot of avocado price by day (includes all major cities)



Modeling

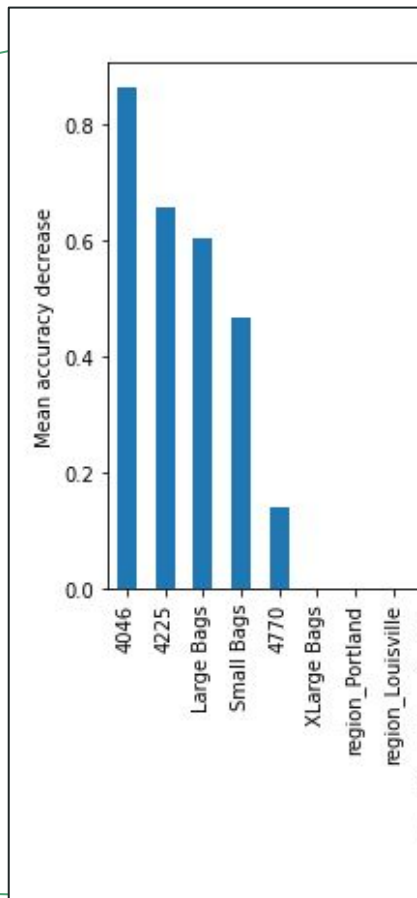
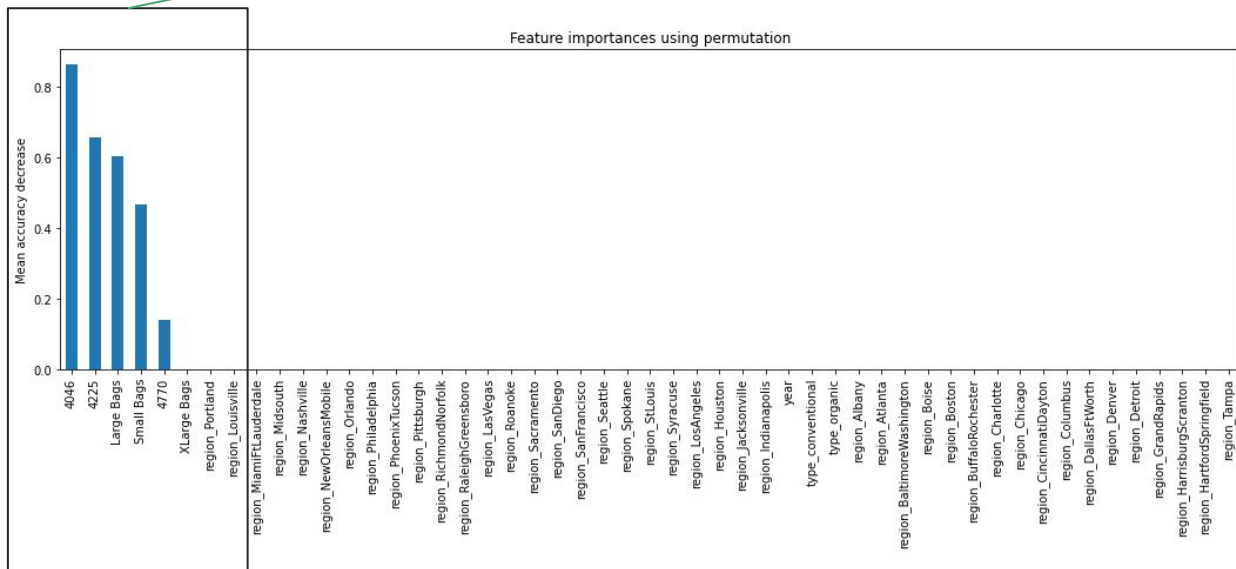
LinearRegression (default hyperparameters)

KNeighborsRegressor (n_neighbors = 5)

RandomForestRegressor (max_depth=2, random_state=0)

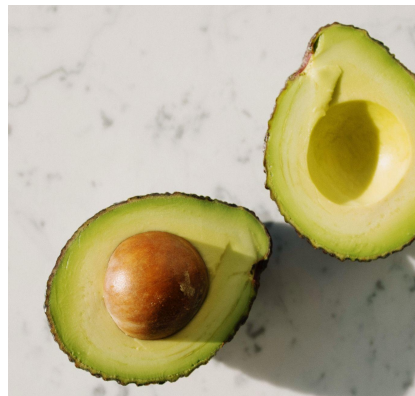
AdaBoostRegressor (random_state=0, n_estimators=100)

Evaluation Metric



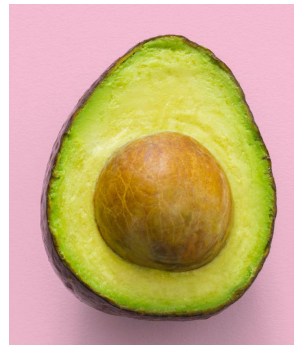
Results

Training Accuracy	Test Accuracy
0.6323	0.7425



The KNeighborsRegressor model yielded an accuracy score of **0.63** on the training data and **0.74** on the test data.

Conclusion



Avocado PLU/size and bag size were the biggest indicators of Price!

- Visualizing the data was my favorite part because it helped me understand it better.
- I had difficulties in preparing the data and running the model. I've practiced on categorical data thus far and had to learn how to use continuous data in ML
- To improve the model, I would try different hyperparameters

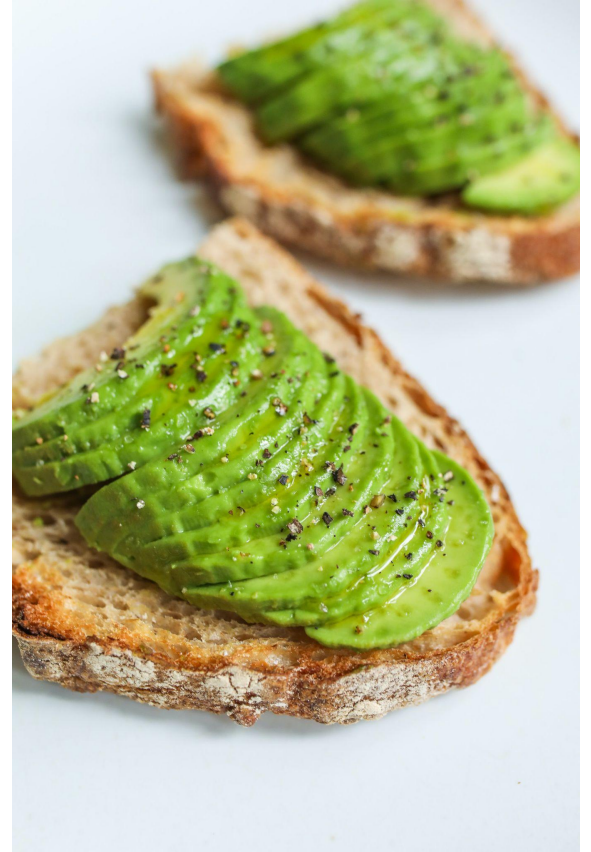
Avocado Toast Recipe

Ingredients:

- 1 ripe Avocado
- 2 slices of preferred bread
- Salt & Pepper

Instructions:

- Toast the bread
- Cut the avocado in half, then in slices
- When toast is ready, scoop avocado half on each slice of toast.
- Spread evenly
- Add salt and pepper to taste
- Enjoy!



Sources

All images from: <https://www.pexels.com/royalty-free-images/>

(Accessed OCT 2022)

Raw data from: <https://hassavocadoboard.com/retail/volume-and-price-data>

CSV file from: <https://www.kaggle.com/datasets/neuromusic/avocado-prices>

(Accessed OCT 2022)