## **Avocado Price Prediction**

#### **Abstract**

The purpose of this project is to build a model to help avocado lovers to predict the future price of avocado-based on the product number and assign if the avocado is conventional or organic.

First, I applied a pairplot to measure the relations between features.

Then, I applied a correlation between each type of labeled avocado and the total volume of avocado.

Using a Linear regression for numerical prediction and Random Forest, then evaluated each model.

## Design

Reviews the avocado prices in detail, it presents to the people who love avocado and make it easier for them to enjoy with cheap avocados.

#### Data

dataset containing 18249 observations on 14 variables. the numerical column names refer to price lookup codes.

- 1. Small Hass.
- 2. Large Hass.
- 3. Extra-large Hass

#### Features:

#	Feature Name	Description	Data Type
1	Date	The date of the observation.	Object
2	AveragePrice	The average price of a single avocado.	Float
3	Туре	The type of avocado (Conventional or Organic).	Object
4	Year	The year of the observation.	Integer
5	Region	The city or region of the observation.	Object
6	Total Volume	Total number of avocados sold.	Float
7	4046	Total number of avocados with PLU* 4046 sold.	Float

8	4225	Total number of avocados with PLU 4225 sold.	Float
9	4770	Total number of avocados with PLU 4770 sold.	Float

# Algorithm

Model Used: Linear Regression, Random Forest Regressor, XGB Regressor.

Model	Accuracy Score
Linear Regression	0.60
Random Forest Regressor	0.90
XGB Regressor	0.90

## **Tools**

Numpy and Pandas for data processing.

Scikit-learn for modeling.

Matplotlib and Seaborn for visualization.

## Communication

Presentation that shows that the geography influences the price, and the organic is more expensive than the conventional.

