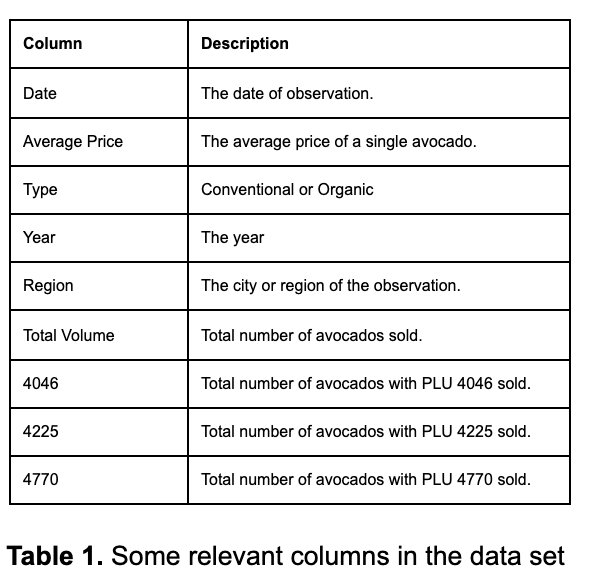
**Term Project - Part 3**

**Introduction:**

Avocado is a food that has become very popular in the United States and around the world. In the US, the local consumption of this fruit in 1985 was 436 million pounds. This number has increased to over 2.6 million pounds consumed by 2020. This food was popularized as a "superfood" as it contains plentiful amounts of healthy fats and fiber. Besides, avocados entered meme status after millionaire Tim Gurner was quoted saying he had gotten rich by not spending money on frivolous things such as avocado toast. A recent survey from Statista showed that the top barrier to buying avocados is its expensive price, according to 78 percent of respondents (Mr.Gurner probably had a point). (Shahbandeh,2020)

In this project, we seek to study avocado consumption relating to the fruit's price, region, and characteristics in the United States. This study would be helpful for producers, retailers, and consumers. With our questions, we want to help determine the preferences of consumers over organic or conventional avocados, and help predict whether the avocados are organic or conventional based on its price, volume sold, and the region it is sold in. These questions could help retailers determine the consumers' preferences according to the area and the price they should sell each avocado. Also, this study can help Millennials find a city with cheap avocados and live out the ‘Millennial American Dream.' 

The data in the study was downloaded from Kaggle.com, and it was originally taken from the Hass Avocado Board website in May of 2018. In the table above, some relevant information on the data set is described.

**Managerial Questions**:

1) Descriptive:

D1) Do customers prefer to buy conventional avocados or organic avocados?

2) Predictive:

P1) Which features in our dataset are most explanatory for predicting the price of avocados?

P2) An avocado trader comes to us with a catalogue of 500 individual pieces of information regarding the amount of avocados sold in certain regions, and the price that they are sold. However, since this trader has gained the avocados through “suspicious” means, they do not have information on whether the avocados are organic or conventional. Based on the models created from the dataset that is analyzed in this project, how many of the trader’s avocados are organic and how many are conventional based on the information given by them?

3) Optimization (Prescriptive):

O1) How can you optimize distribution of organic and conventional avocados between 45 regions/cities (i.e., decide what amount of avocados and what regions each supplier will send to) based on supply/demand data for each city given each city coordinates and assuming that the cost of transporting conventional avocados is 2 times the distance between the cities, and the cost to transport avocados is 5 times the distance between the cities?

**Methodology:**

Data Editing:

Our dataset was relatively clean when we first acquired it. Therefore, in order to implement the data cleaning techniques learned in this course, we needed to apply some edits to make it less clean. To clutter the dataset, we deleted data points at random from our dataset.

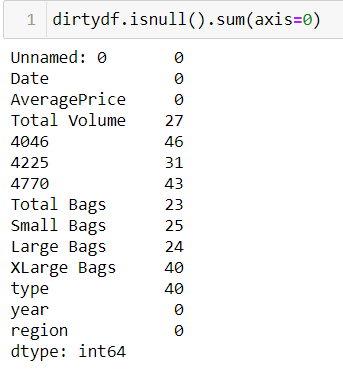


Figure 1: Dirty Data

Data Cleaning:

After cluttering the data, the first step was to impute the missing values. To do this, we first did some data visualization in Power BI and Python. What we noticed is almost all the variables had a very large distribution. For example, when plotting a distplot of ‘4046’ as shown in Figure 2, the range of the tail on the distplot is very high.

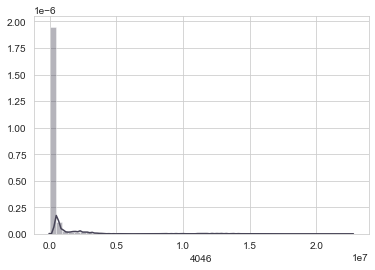


Figure 2: Distplot of 4046

When we want to impute the missing values, this skewed distribution could affect our data. To verify this we checked the describe report from pandas as well, seen in Figure 3.

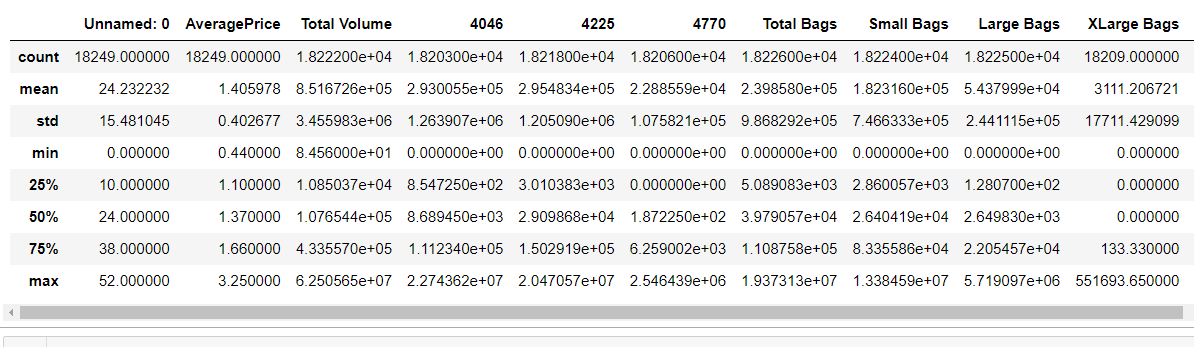


Figure 3: Describe Report of Our Data

We can see that the median and mode can be an order of magnitude off from each other and that the interquartile range is very large as well. To determine the cause of this, we then plotted the data into Power BI for a more clear visualization. We found that the cause for this large distribution is most likely tied to the region, as some regions sold more avocados than others. This is exemplified in Figure 4, where the Boston region only sells 10,000 avocados as compared to Los Angeles with 2 million during the same timeframe.

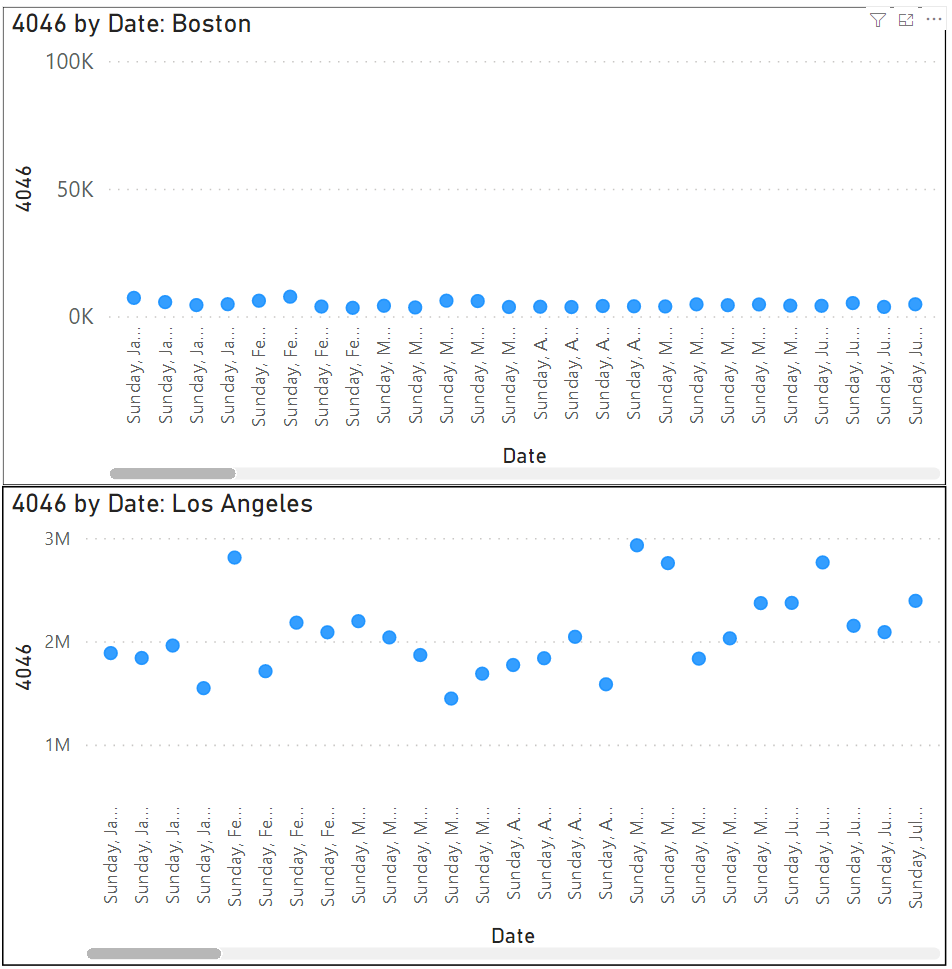


Figure 4: Comparison of Avocado Sales Between Boston and Los Angeles

It was also noted that within each region and year, the amount of avocados sold had only a small amount of fluctuation and was relatively stable. Therefore, it was decided to impute the missing values by using the mean value regional data in that particular year to get an accurate data approximation. For example, if Los Angeles had missing data in 2015 for sales of ‘4046’ avocados, the missing values were filled with the mean value of Los Angeles in 2015 for ‘4046’ avocados. This was applied for all numerical data columns. For our missing categorical columns such as the type of avocado, they were filled with the most common occurrence in that same year and region.

One big complication we noticed during this process is that there were a lot of aggregate regions in our data. For example, there is also a California region in addition to the many cities in California as regions (Los Angeles, San Francisco, San Diego, Sacramento). When we plotted this distribution of regions in Power BI, we noted that the aggregate regions (TotalUS, California, SouthCentral, etc) account for a large portion of our sales, seen in Figure 5. Because we didn’t know exactly how these aggregate regions were created and what subregions could be contained in them, we decided to drop the data that contained our large aggregate regions. This also helped with our problems with large data distributions and would remove possible double counting of the data. The results shown in Figure 6 show that our data is better distributed after removing large aggregate regions. Comparing Figures 5 and 6 also shows that the magnitude of the sales has equalized as well.

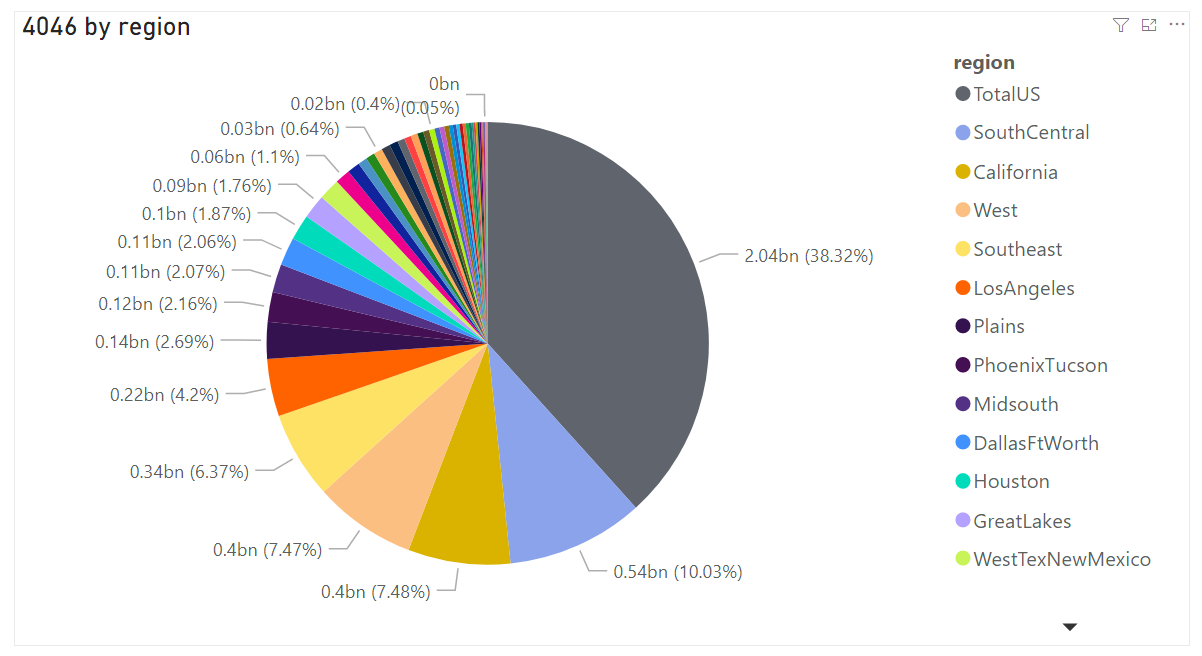


Figure 5: Amount of 4046 sold by region before cleaning

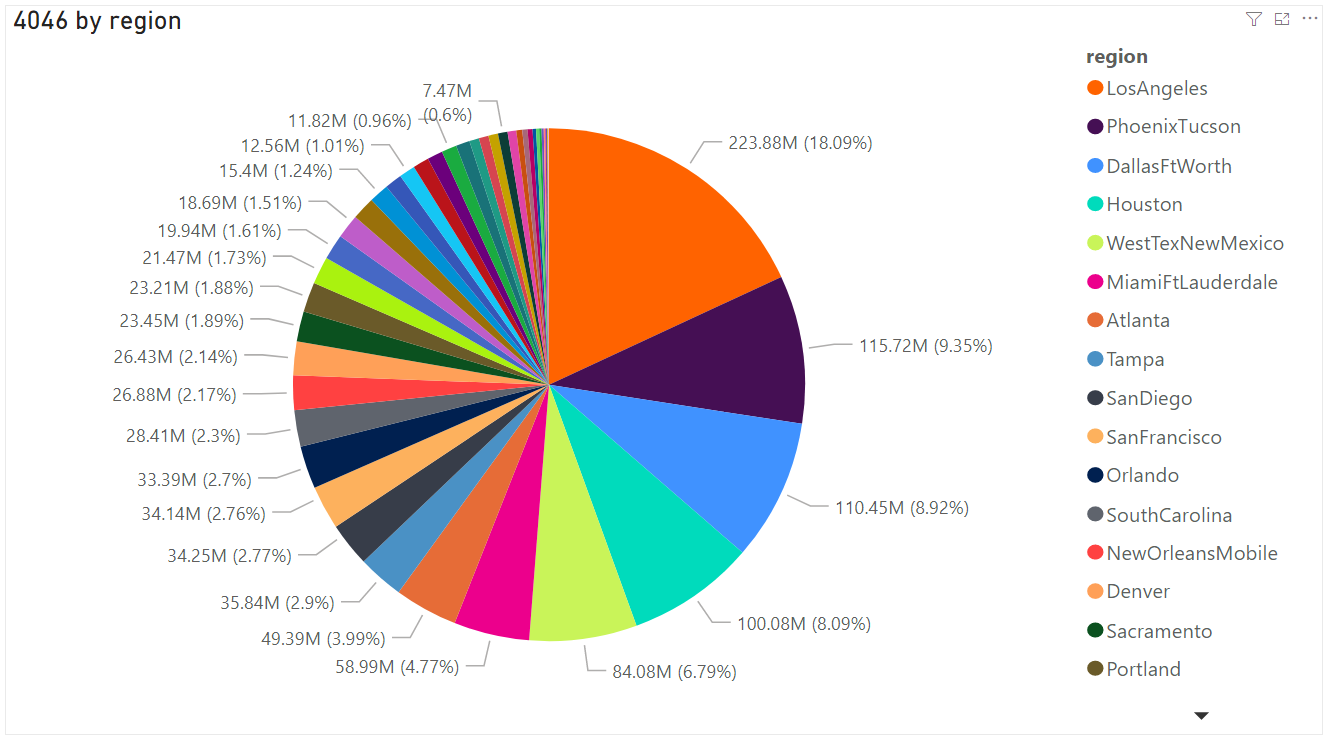


Figure 6: Amount of 4046 sold by region after cleaning

Next we removed columns that were aggregates of other columns. In our case, these were the Total Volume and Total Bags columns. Total Volume was the sum of our 3 types of avocados and Total Bags, while Total Bags was the sum of our 3 types of bag columns. We dropped the unneeded Unnamed:0 column, as it was an index column. Then we dummied the type column, setting organic as 1 and conventional as 0.

From here, 500 rows of data was extracted from the dataset and set aside as the “black market” avocados that will be used for Question 2 of the predictive questions. The “organic” column of the “black market” avocados was removed to simulate a situation where the target variable is not known, and the data will later be used to predict how many of the data points are conventional or organic.

After removing aggregate regions and correlated columns with other columns, outlier values were modified to fit the data better. Columns 4046, 4225, 4770, Small bags, Large bags, and XLarge bags had outliers, as shown in figure 7.

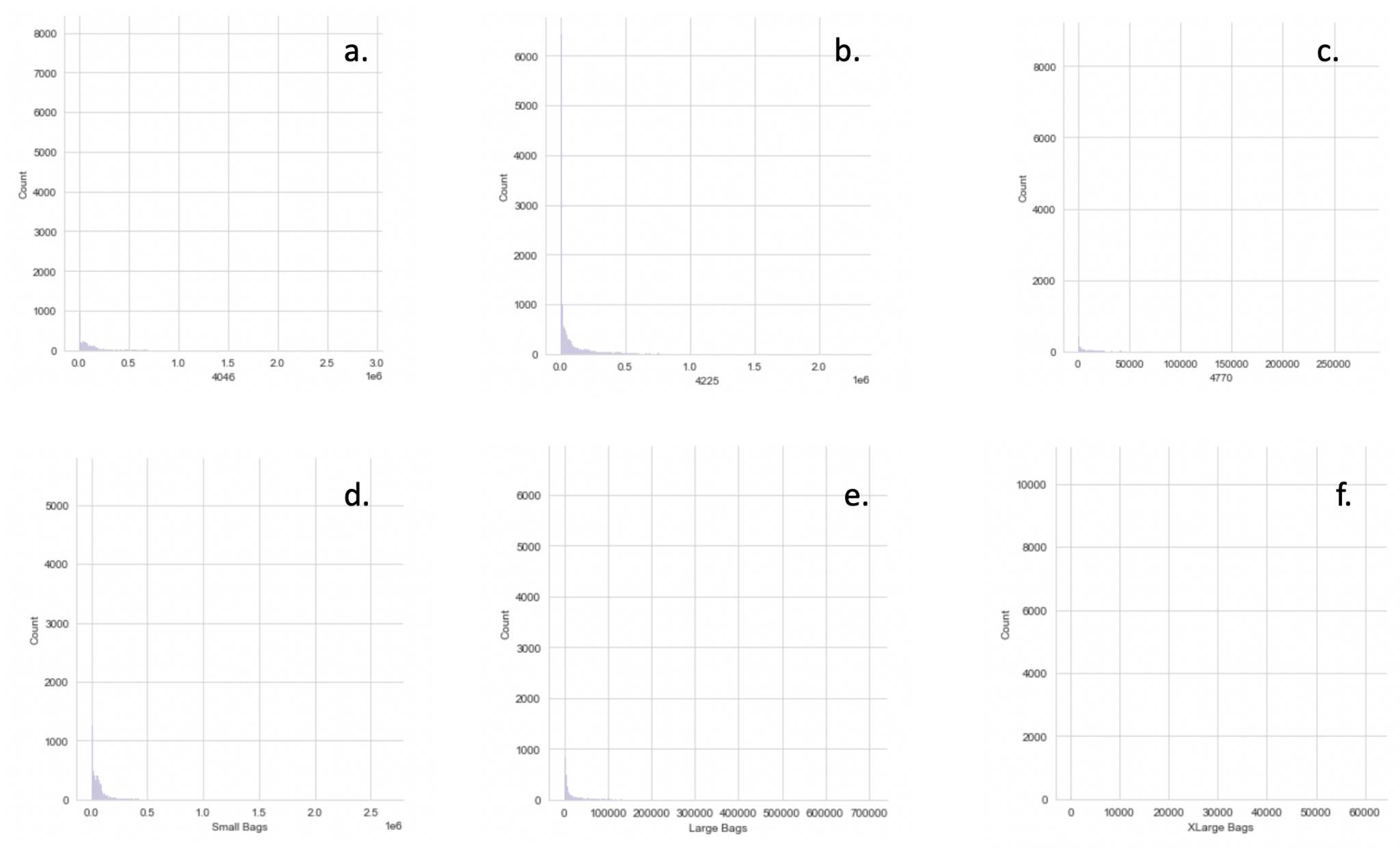


Figure 7. Outliers in columns 4046(a), 4225(b), 4770(c), Small bags(d), Large bags(e.) and XLarge bags(f.).

To examine the outliers for the six columns, we used the percentile-based method, interquartile range method, and standard deviation method. Each method provides a range, and for each method, we had a minimum and a maximum value. The minimum value of the three minimum values obtained by each method is selected to find the outliers. The maximum value of the three maximum values obtained by each method is selected as well. If the column value is out of this range, the data point is identified as an outlier. Since all the outlier values in every column occurred by exceeding the upper bound of the obtained range, we replaced each column's outlier values by the upper bound of the range-1, because these values are still high but not high enough to be identified as outliers.

Then, dummy variables were created for the categorical variables. These variables allow us to represent nominal variables in a numerical way. The only remaining categorical variable is region. From the region column we derived 42 other columns (regions in the data set). We noticed a correlation between the data and year column as part of the cleaning and to avoid this correlation, we dropped the year column. The date data was in year-month-day format which was transformed into decimals to make it a numerical value for our model. Finally, the data was scaled using the StandardScaler function from the scikit-learn package.

Data Partition:

The dependent variable, or Y is the organic column for the classification model, and for the regression model, the dependent variable Y is the average price column. The X variables for each model were determined to be the rest of the columns included in the model. The data was then split into a 30% test set and 70% training set.

Methodology for Predictive Question 1 (P1)**:**

Our first predictive question referred to the prediction of avocado prices, which can be set up as a Regression problem, and which we solved by testing various Linear Regression models. Prior to testing any models, we calculated a correlation matrix among most independent variables and between those variables and our target. For convenience’s sake, we did not include the region dummy variables in the correlation matrix analysis, as there were more than forty of them. Instead, we tested models with and without the region variables to evaluate their impact.

Below is a figure (Figure R1) of the absolute values of the correlation matrix. We noticed here that there are no major correlations amongst the independent variables, therefore all variables will be kept in the model. We also noticed that the variables that appear most correlated to ‘AveragePrice’ are ‘organic’ followed by ‘4046’.

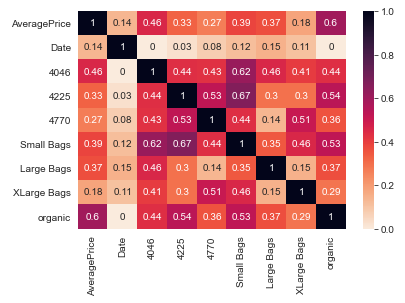


Figure R1: Absolute Values of Correlation Matrix

To predict price, and taking into account which variables had the highest correlation to ‘AveragePrice’, we then tested seven different models that included the following sets of independent variables:

Model 0: (['organic'])

Model 1: (['organic', '4046'])

Model 2: (['organic', '4046', 'Small Bags'])

Model 3: (['organic', '4046', 'Small Bags', '4225', 'Large Bags'])

Model 4: (['organic', '4046', 'Small Bags', '4225', 'Large Bags', '4770', 'XLarge Bags'])

Model 5: (['organic', '4046', 'Small Bags', '4225', 'Large Bags', '4770', 'XLarge Bags', 'Date'])

Model 6: ([all variables])

Where Model 6 used all 52 possible independent variables (i.e., the eight variables from Model 5, plus 44 dummy variables representing different regions.)

The models were evaluated by looping over the different combinations of predictors. We recorded various metrics for each model, including Train R2, Test R2, Train Adjusted R2, Test Adjusted R2, Train RMSE, and Test RMSE, which we use to evaluate the results.

Methodology for Predictive Question 2 (P2)**:**

Regarding the second predictive question, we implemented various classification models with regards to the main question asked: whether or not we could predict which type of avocado is organic or conventional and thus be able to use the trained model to test a real life scenario. To start, we first separated the data with respect to ‘organic’ as our target. The rest of the variables were chosen as the independent variables that influenced the target. Once these categories were defined, they were then split into the training and test sets, which were then used for the various classification models. These models were Logistic Regression, SVM, KNN, Decision Tree and Random Forests.

Once the best classification model was determined, the “black market” avocados data was inputted into the model to predict whether the avocados are conventional or organic.

Methodology for Optimization Question 1 (O1)**:**

Taking our results from the second predictive question (P2), we set up an optimization problem to minimize the cost of delivering avocados between cities. We created 2 optimization models, one for conventional and one for organic avocados. The supply nodes were each region and the values for each were the results from our second predictive question (P2). The demand nodes were the same regions but with randomized values for our demands. Both types of avocados had higher supply than demand. We then generated random distances between the regions in the form of x and y coordinates to use in our objective model. Finally, we set the organic transportation costs to be 5 times the euclidean distance, and the conventional transportation costs to be 2 times the euclidean distance between cities.

Organic Avocado Optimization Model:

1. Decision Variables:

* *xij* = Amount of organic avocados (in tons) to deliver from city *i* to city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.

1. Objective Function: Minimize Total Cost (in US $),

Where (parameters):

* *cxi* and *cxj* = X coordinate of each city *i* and each city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.
* *cyi* and *cyj* = Y coordinate of each city *i* and each city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.

1. Constraints:

Objective Function Subject To:

* C1: At every agency, Supply + Incoming – Outgoing – Demand ≥ 0.

, ⦡*k* = 1..45.

where, parameter:

* *sk* = supply of organic avocados (in tons) at each city *k*, ⦡*k* = 1..45.
* *dk* = demand of organic avocados (in tons) at each city *k*, ⦡*k* = 1..45.

Also Subject To:

* *xij* ≥ 0, ⦡*i* = 1..45 and ⦡*j* = 1..45 (non-negativity constraints).

See Appendix 1 for values for *cxk*, *cyk*, *sk* and *dk*, ⦡*k* = 1..45.

Conventional Avocado Optimization Model:

1. Decision Variables:

* *xij* = Amount of conventional avocados (in tons) to deliver from city *i* to city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.

1. Objective Function: Minimize Total Cost (in US $),

Where (parameters):

* *cxi* and *cxj* = X coordinate of each city *i* and each city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.
* *cyi* and *cyj* = Y coordinate of each city *i* and each city *j*, ⦡*i* = 1..45 and ⦡*j* = 1..45.

1. Constraints:

Objective Function Subject To:

* C1: At every agency, Supply + Incoming – Outgoing – Demand ≥ 0.

, ⦡*k* = 1..45.

where, parameter:

* *sk* = supply of conventional avocados (in tons) at each city *k*, ⦡*k* = 1..45.
* *dk* = demand of conventional avocados (in tons) at each city *k*, ⦡*k* = 1..45.

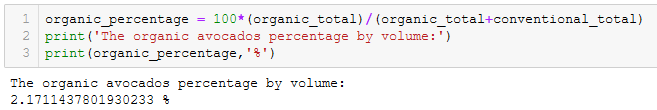
Also Subject To:

* *xij* ≥ 0, ⦡*i* = 1..45 and ⦡*j* = 1..45 (non-negativity constraints).

See Appendix 1 for values for *cxk*, *cyk*, *sk* and *dk*, ⦡*k* = 1..45.

**Computational Results:**

Results for Descriptive Question 1 (D1):

  
 After the data cleaning process, the total volume of conventional avocados and organic avocados were calculated, and the percentage of volume for organic avocados was calculated. This step was done right after the data cleaning process before normalizing the data, before removing outliers, and after removing the 500 data points for the “black market” avocados. It was found that 2.171% of the total volume sold comes from organic avocados, while 97.829% of the volume comes from conventional avocados. This is realistic, since conventional avocados are cheaper to buy and cheaper to produce. Hence, a much larger percentage of the volume of avocados sold comes from conventional avocados.

Results for Predictive Question 1 (P1)**:**

Based on all metrics calculated, including Test Adjusted R2, the best performing model is Model 6, the model that used all variables, including all of the different regions as dummy variables (Table R1 and Figure R2 below.) This leads us to believe that region is an important variable into the model.

We also notice that the performance of this model is not great. The Test Adjusted R2 is 0.567, which is not very high. We noticed that there are large differences in the data between regions. Hence, a better approach to predict price would probably be to create separate models for different regions. However, this is beyond the scope of our project.

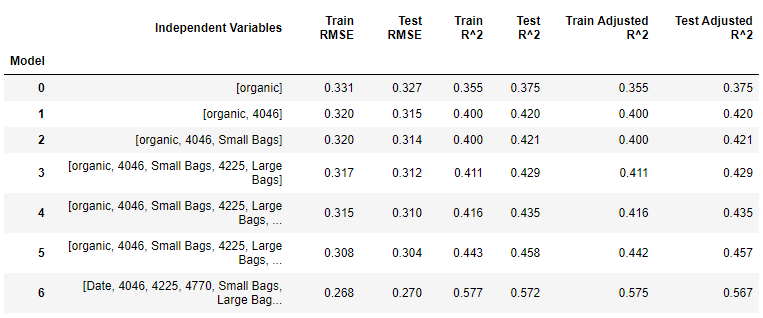


Table R1: Metrics for Various Linear Regression Models Tested to Predict Avocado Price

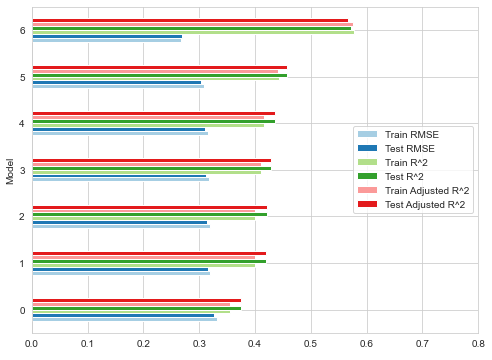


Figure R2: Metrics for Various Linear Regression Models Tested to Predict Avocado Price

Regarding important features to predict avocado price, we noticed that for all of the models that did not include regions (Models 1-5), the driving variable was always the variable for avocado type: 'organic'. We made this conclusion by looking at the coefficients from each of these models, where the coefficient for 'organic' was always above 0.4, while the coefficients for the other variables were at or below 0.1 (see Table R.2 below.)

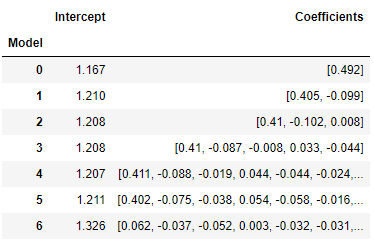


Table R.2: Intercept and Coefficients for Various Linear Models Tested to Predict Avocado Price

For our best model, Model 6 (the one that included all of the region variables), we noticed that 'organic' is still the most important, with a coefficient equal to 0.367. However, we identified that many of the region variables are also similarly important, with coefficients such as 0.362, 0.353, 0.343, 0.342, 0.33, 0.315, etc. This is probably due to the very significant differences that we observed in the data between regions (see Table R.3 below.)

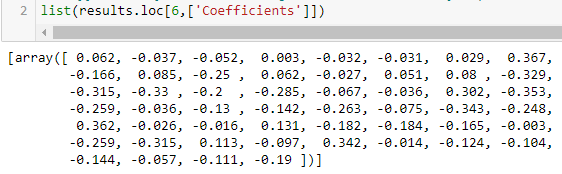


Table R.3: Regression Coefficients for Model 6.

(Notice that coefficients for several regions now become equally important in predicting price)

These results agree with our initial expectations. We understand that in the real world, organic fruits such as avocado tend to be more expensive due to the unique costs of growing them. Therefore, we expected the variable ‘organic’ to be a driver in this model. We also know that fruit prices may vary by region (due to production costs in that region, transportation costs, climate, and perhaps most importantly, demand.) Therefore, it was natural to think that region would also be an important independent variable to predict price.

Finally, we conclude that even though our model is good, it is not a great model. A Test Adjusted R2 of 0.567 is not ideal (even if it is more “realistic” in real-life problems.) A more likely and realistic approach to predict avocado prices would be to create different models for different regions. This is beyond the scope of our project, but it is a recommendation that we would make for future steps.

Results for Predictive Question 2 (P2):

With each model, a confusion matrix and classification report were used to determine the accuracy of the trained model. Because the accuracy of each model was very high for all the different models, a separate accuracy score was generated for each model based on the training and test sets. This was shown to a greater number of decimal points to illustrate the differences between them. In the figure below, we can see that the training and test set accuracy scores for Logistic Regression are fairly high and comparable for both sets of data. This same trend can be seen with the other classification models, as shown below as well.

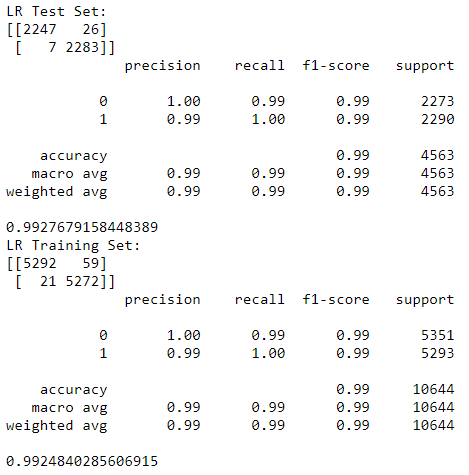


Figure C1: Logistic Regression Statistics

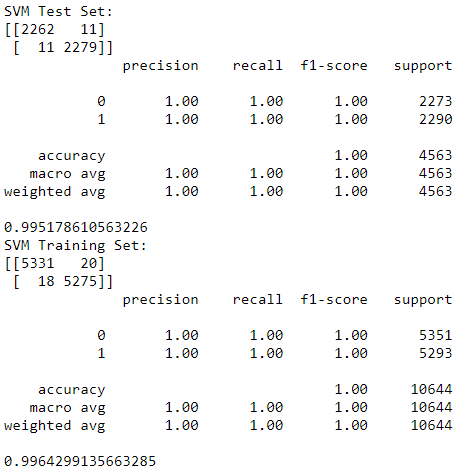


Figure C2: SVM Statistics

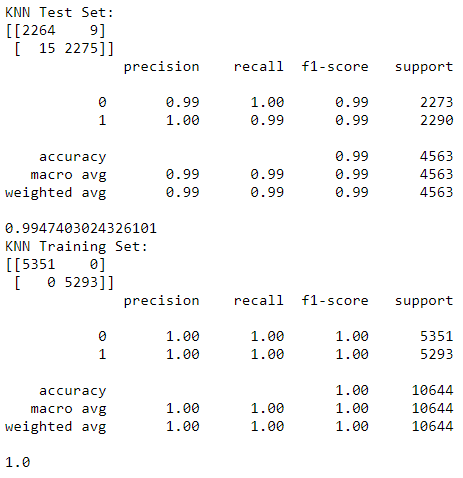


Figure C3.1: KNN Statistics Before Optimization

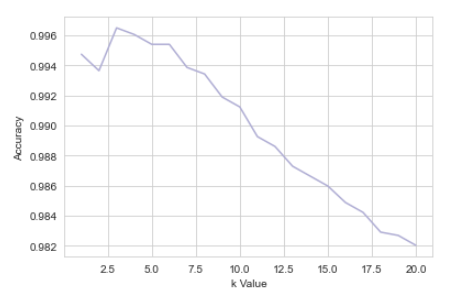


Figure C3.2: Graph Depicting Accuracy Score and K Value for KNN Model

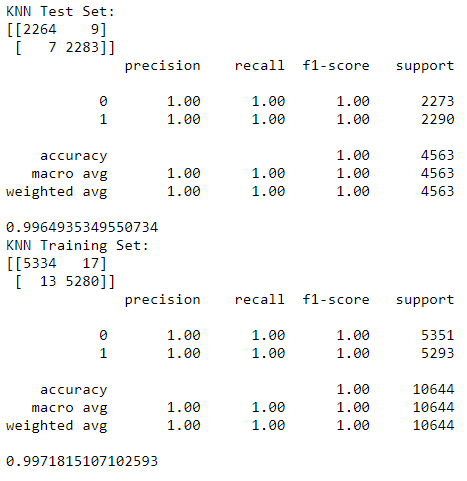


Figure C3.3: KNN Statistics After Optimization

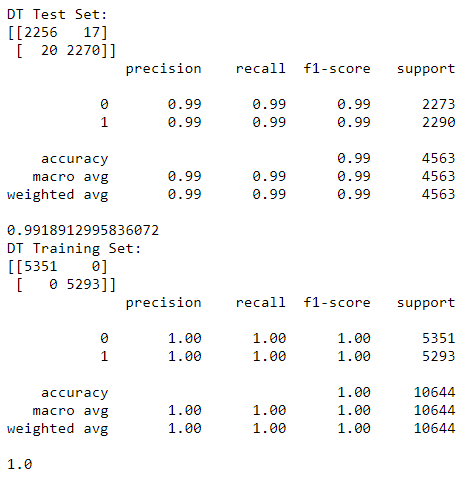


Figure C4: Decision Tree Statistics

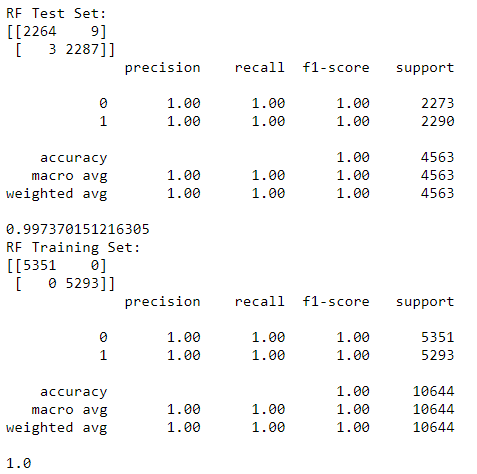


Figure C5.1: Random Forests Statistics Before Optimization

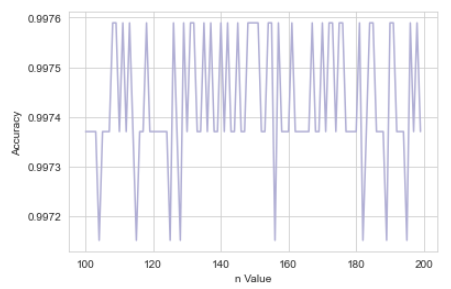


Figure C5.2: Graph Depicting Accuracy Score and N Value for Random Forest Model

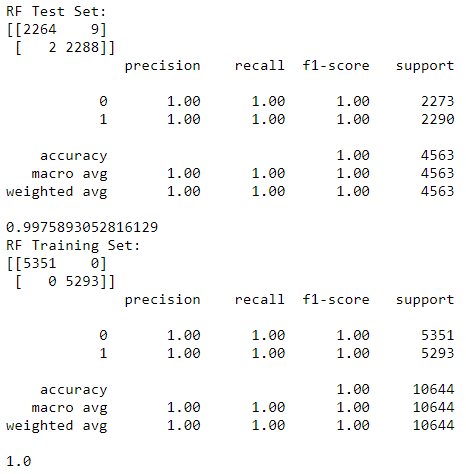


Figure C5.3: Random Forests Statistics After Optimization

As all of the models were used in this portion of the project to determine which model would suit the data and give the best accuracy score, the differences between the distinct models can be shown. While accuracy scores for all the models were very high (all in the 0.99 range), there were a few distinctions. While the number of iterations did not seem to affect the accuracy score for Logistic Regression too much, optimizing both the KNN and the Random Forest model did allow for some improvement, albeit very minor in the case of Random Forest. For KNN, the test accuracy score decreased when optimized but the training score was improved, whereas for Random Forest, the change from optimizing the n value was very minor, a 0.001 improvement. Therefore, our best model for this data would most likely be the Random Forest model, with KNN following it up. With regards to the second predictive question, given how high the accuracy scores are for each model, we are definitely able to predict which type of avocado each region would want.

Therefore in order to answer the second predictive question, the Random Forest model was used to predict the type of avocado based on the information found in the “black market” avocados data. The methodology to clean the “black market” avocados data is the same as the method used on the training dataset, and once inputted, the final answer is as follows:

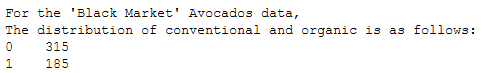


Figure C6: Distribution of conventional and organic avocados for “black market” avocados.

This shows that the model predicted 315 of the data points were conventional, while 185 rows were organic. Although the sample was taken randomly from the original dataset, the split of the predicted type of avocado was not 50/50, but rather weighing heavily on the conventional side. This is what we predicted, since it was found that the split of organic and conventional avocados rows were 50/50. The cause of the non-50/50 split may be due to the method of sampling and error, where more conventional data was taken during the sampling than the organic purely from chance. Another cause may be due to incomplete data cleaning on the “black market” avocados, where the columns are not cleaned the same way as the original dataset. However, from analysis, this is not the case, since the steps taken to clean this data is similar to the original dataset. Lastly, the model used may be the cause of the non-50/50 split. Although the Random Forest model is deemed as the most accurate model, the other models may very well be used to determine the outcome of this split as well. However, with the Random Forest Model, it was found that 315 of the data points were conventional, while 185 rows were organic. Further testing can involve testing the other trained models with this realistic situation and see if these all match up. Overall, while the final output was fairly unbalanced, we expect this output to be accurate, given the accuracy scores of the various models.

Results for Optimization Question 1 (O1):

1. Organic Avocados Optimization:

In total, there were 185 tons of avocados on the supply nodes and 145 on the demand nodes. The regions that supplied avocados to the other areas were: Albany, Boise, Boston, Buffalo, Columbus, Detroit, Indianapolis, Las Vegas, Los Angeles, Miami, New Orleans, Orlando, Pittsburgh, Raleigh, Richmond, San Francisco, and South Carolina. The regions that received avocados were: Atlanta, Baltimore, Cincinnati, Scranton, Houston, Jacksonville, Louisville, New York, Northern New England, Phoenix, Portland, Roanoke, Syracuse, and New Mexico (Table O1). The optimal objective value for the transport is $4394.

| **Supplier** | **Recipient** | **Tons of Avocado Supplied -> Received** |
| --- | --- | --- |
| Albany | PhoenixTucson | 5 |
| Boise | CincinnatiDayton | 2 |
| Boise | Roanoke | 1 |
| Boston | HarrisburgScranton | 2 |
| Boston | NorthernNewEngland | 2 |
| BuffaloRochester | HarrisburgScranton | 2 |
| Columbus | DallasFtWorth | 1 |
| Columbus | WestTexNewMexico | 1 |
| Detroit | BaltimoreWashington | 3 |
| Indianapolis | Portland | 1 |
| LasVegas | Denver | 3 |
| LosAngeles | Jacksonville | 1 |
| MiamiFtLauderdale | HarrisburgScranton | 1 |
| NewOrleansMobile | BaltimoreWashington | 3 |
| Orlando | Jacksonville | 1 |
| Pittsburgh | NewYork | 1 |
| RaleighGreensboro | Portland | 1 |
| RichmondNorfolk | Atlanta | 1 |
| RichmondNorfolk | PhoenixTucson | 1 |
| SanFrancisco | Louisville | 1 |
| SanFrancisco | Syracuse | 1 |
| SouthCarolina | Houston | 1 |
| Tampa | NorthernNewEngland | 1 |

Table O1. Supply and recipient of organic avocados.

1. Conventional Avocados Optimization:

In total, there were 315 tons of avocados on the supply nodes and 314 on the demand nodes. The regions that supplied avocados to the other areas were: Albany, Boise, Boston, Buffalo, Charlotte, Chicago, Cincinnati, Columbus, Detroit, Indianapolis, Louisville, Miami, Nashville,New York, Philadelphia, Portland, San Francisco, South Carolina, Spokane. The regions that received avocados were:GrandRapids, PhoenixTucson, Pittsburgh, Boston, BaltimoreWashington, RaleighGreensboro, Jacksonville, LosAngeles, StLouis, Syracuse, RichmondNorfolk, NewOrleansMobile, Orlando, Tampa, NorthernNewEngland, WestTexNewMexico, SanDiego, Sacramento, Houston(Table O2). The optimal objective value for the transport is $3466.

| **Supplier** | **Recipient** | **Tons of Avocado Supplied -> Received** |
| --- | --- | --- |
| Albany | GrandRapids | 3 |
| Albany | PhoenixTucson | 2 |
| Boise | Pittsburgh | 1 |
| BuffaloRochester | Boston | 3 |
| Charlotte | BaltimoreWashington | 3 |
| Charlotte | RaleighGreensboro | 3 |
| Chicago | Jacksonville | 1 |
| Chicago | LosAngeles | 1 |
| Chicago | StLouis | 1 |
| Chicago | Syracuse | 5 |
| CincinnatiDayton | RichmondNorfolk | 1 |
| Columbus | BaltimoreWashington | 1 |
| Columbus | Boston | 1 |
| Columbus | NewOrleansMobile | 3 |
| Columbus | Orlando | 3 |
| Columbus | Tampa | 2 |
| DallasFtWorth | NorthernNewEngland | 2 |
| DallasFtWorth | WestTexNewMexico | 1 |
| Detroit | SanDiego | 5 |
| Indianapolis | RaleighGreensboro | 4 |
| ​​Louisville | Boston | 2 |
| Louisville | Sacramento | 2 |
| Louisville | Syracuse | 1 |
| MiamiFtLauderdale | HarrisburgScranton | 4 |
| MiamiFtLauderdale | Seattle | 1 |
| Nashville | Houston | 2 |
| NewYork | HartfordSpringfield | 2 |
| Philadelphia | GrandRapids | 2 |
| Philadelphia | RichmondNorfolk | 1 |
| Portland | Denver | 2 |
| Portland | LasVegas | 2 |
| Portland | RaleighGreensboro | 2 |
| Roanoke | RichmondNorfolk | 1 |
| Portland | Denver | 4 |
| Portland | LasVegas | 1 |
| Portland | RaleighGreensboro | 3 |
| Roanoke | RichmondNorfolk | 1 |
| SanFrancisco | Syracuse | 1 |
| SouthCarolina | HartfordSpringfield | 4 |
| SouthCarolina | Houston | 5 |
| Spokane | Boston | 1 |
| Spokane | RichmondNorfolk | 5 |
| Spokane | Seattle | 4 |

Table O1. Supply and recipient of organic avocados.

Bibliography:

Kiggins, J. (2018, June 6). *Avocado prices*. Kaggle. Retrieved September 20, 2021, from https://www.kaggle.com/neuromusic/avocado-prices.

Shahbandeh, M. (2020, November 3). *U.S. avocado Consumption 2018 l Statista*. Statista. Retrieved September 20, 2021, from https://www.statista.com/statistics/591263/average-avocado-consumption-us-per-week/.

Appendix 1:

Input data for optimization model: *cxk* (x coordinate for node/city *k*), *cyk* (y coordinate for node/city *k*), *sk* (supply for node/city *k*) and *dk* (demand for node/city *k*).

| Number | Nodes (Cities) | X Coord | Y Coord | Supply-Conv (Tons) | Supply-Org (Tons) | Demand-Conv (Tons) | Demand-Org (Tons) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Albany | 0 | 20 | 7 | 5 | 2 | 0 |
| 2 | Atlanta | 15 | 25 | 6 | 2 | 6 | 3 |
| 3 | BaltimoreWashington | 50 | 90 | 5 | 3 | 9 | 9 |
| 4 | Boise | 25 | 25 | 7 | 4 | 6 | 0 |
| 5 | Boston | 45 | 45 | 5 | 5 | 12 | 1 |
| 6 | BuffaloRochester | 45 | 40 | 6 | 4 | 3 | 2 |
| 7 | Charlotte | 65 | 75 | 7 | 5 | 1 | 5 |
| 8 | Chicago | 70 | 15 | 9 | 0 | 0 | 0 |
| 9 | CincinnatiDayton | 20 | 25 | 9 | 0 | 8 | 2 |
| 10 | Columbus | 55 | 55 | 11 | 8 | 1 | 6 |
| 11 | DallasFtWorth | 25 | 85 | 11 | 2 | 8 | 3 |
| 12 | Denver | 80 | 70 | 7 | 1 | 9 | 4 |
| 13 | Detroit | 60 | 60 | 10 | 3 | 5 | 0 |
| 14 | GrandRapids | 0 | 5 | 5 | 5 | 9 | 1 |
| 15 | HarrisburgScranton | 0 | 85 | 8 | 7 | 9 | 12 |
| 16 | HartfordSpringfield | 40 | 0 | 2 | 4 | 7 | 0 |
| 17 | Houston | 60 | 15 | 6 | 3 | 9 | 4 |
| 18 | Indianapolis | 90 | 90 | 4 | 3 | 0 | 2 |
| 19 | Jacksonville | 80 | 50 | 8 | 2 | 9 | 4 |
| 20 | LasVegas | 75 | 65 | 6 | 3 | 8 | 0 |
| 21 | LosAngeles | 89 | 45 | 7 | 1 | 8 | 0 |
| 22 | Louisville | 63 | 23 | 8 | 0 | 1 | 1 |
| 23 | MiamiFtLauderdale | 20 | 30 | 9 | 6 | 6 | 5 |
| 24 | Nashville | 67 | 13 | 7 | 2 | 5 | 2 |
| 25 | NewOrleansMobile | 55 | 57 | 3 | 12 | 6 | 9 |
| 26 | NewYork | 46 | 16 | 2 | 8 | 0 | 9 |
| 27 | NorthernNewEngland | 15 | 89 | 4 | 6 | 6 | 9 |
| 28 | Orlando | 66 | 52 | 5 | 1 | 8 | 0 |
| 29 | Philadelphia | 17 | 10 | 9 | 4 | 6 | 3 |
| 30 | PhoenixTucson | 0 | 23 | 7 | 2 | 9 | 8 |
| 31 | Pittsburgh | 37 | 26 | 9 | 13 | 10 | 0 |
| 32 | Portland | 84 | 78 | 12 | 1 | 7 | 3 |
| 33 | RaleighGreensboro | 70 | 78 | 7 | 2 | 15 | 1 |
| 34 | RichmondNorfolk | 12 | 21 | 7 | 4 | 18 | 0 |
| 35 | Roanoke | 23 | 24 | 9 | 1 | 5 | 2 |
| 36 | Sacramento | 60 | 25 | 7 | 10 | 8 | 10 |
| 37 | SanDiego | 69 | 63 | 4 | 2 | 9 | 2 |
| 38 | SanFrancisco | 67 | 34 | 10 | 9 | 9 | 5 |
| 39 | Seattle | 8 | 93 | 8 | 0 | 15 | 0 |
| 40 | SouthCarolina | 58 | 12 | 13 | 2 | 9 | 1 |
| 41 | Spokane | 23 | 27 | 10 | 5 | 0 | 5 |
| 42 | StLouis | 89 | 22 | 3 | 13 | 4 | 8 |
| 43 | Syracuse | 75 | 42 | 2 | 3 | 12 | 4 |
| 44 | Tampa | 35 | 67 | 6 | 3 | 8 | 2 |
| 45 | WestTexNewMexico | 24 | 82 | 8 | 6 | 9 | 7 |