

comparisons

May 4, 2020

1 Project 2

this is not a group project

This project will focus on comparing and contrasting the supervised and unsupervised algorithms we have learned so far. **Clearly label each section and number in the notebook**

1.1 Part I

Use the dataset *burgersOrPizza.csv* to build 3 models that predict whether a food is a burger or pizza (you can use any of the predictive models we've learned).

For each model:

0. Explore data (with ggplot)
1. Explain which variables you're using to predict the outcome.
2. Explain which model validation technique you're using and why.
3. Explain why you did or did not choose to standardize your continuous variables.
4. Evaluate how the model performed. Explain.

At the end:

5. Compare the performance of the 3 models using the accuracy, and the confusion matrix (consider things like how many it got correct, which errors it w

```
[49]: #https://docs.bamboolib.8080labs.com/documentation/how-tos/
      ↪ installation-and-setup/install-bamboolib
      #https://docs.bamboolib.8080labs.com/documentation/how-tos/
      ↪ installation-and-setup/install-bamboolib/test-bamboolib
      #https://stackoverflow.com/questions/19913659/
      ↪ pandas-conditional-creation-of-a-series-dataframe-column

import yaml
#import bamboolib as bam
import pandas as pd
import numpy as np
import warnings
from plotnine import *
#import matplotlib as plt
warnings.filterwarnings('ignore')
```

```

from sklearn.preprocessing import StandardScaler #Z-score variables
from sklearn.preprocessing import LabelEncoder

from sklearn.cluster import AgglomerativeClustering, KMeans
from sklearn.mixture import GaussianMixture
from sklearn.metrics import silhouette_score
import scipy.cluster.hierarchy as sch

from sklearn.model_selection import train_test_split # simple TT split cv
from sklearn.linear_model import LogisticRegression # Logistic Regression Model
from sklearn.tree import DecisionTreeClassifier # Decision Tree
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, \
    confusion_matrix, roc_curve, roc_auc_score #model eval

%matplotlib inline
plt.rcParams['figure.figsize'] = (8,25)
pd.options.display.max_columns = None
pd.set_option('display.max_rows', 500)

```

```

[2]: foods = pd.read_csv('data/burgersOrPizza.csv')
foods.head()

```

```

[2]:

```

| | Item_Name \ |
|---|-------------------------------------|
| 0 | Chicken n Cheese Slider |
| 1 | Corned Beef n Cheese Slider |
| 2 | Ham n Cheese Slider |
| 3 | Jalapeno Roast Beef n Cheese Slider |
| 4 | Roast Beef n Cheese Slider |

| | Item_Description | Food_Category | Calories \ |
|---|---|---------------|------------|
| 0 | Chicken n Cheese Slider on Mini Bun w/ Chicken... | Burgers | 290.0 |
| 1 | Corned Beef n Cheese Slider on Mini Bun w/ Cor... | Burgers | 220.0 |
| 2 | Ham n Cheese Slider on Mini Bun w/ Roast Ham &... | Burgers | 230.0 |
| 3 | Jalapeno Roast Beef n Cheese Slider on Mini Bu... | Burgers | 240.0 |
| 4 | Roast Beef n Cheese Slider on Mini Bun w/ Roas... | Burgers | 240.0 |

| | Total_Fat | Saturated_Fat | Trans_Fat | Cholesterol | Sodium | Potassium \ |
|---|-----------|---------------|-----------|-------------|--------|-------------|
| 0 | 12.0 | 3.5 | 0.0 | 25.0 | 720.0 | NaN |
| 1 | 9.0 | 3.5 | 0.0 | 30.0 | 890.0 | NaN |
| 2 | 9.0 | 3.5 | 0.0 | 30.0 | 750.0 | NaN |
| 3 | 11.0 | 4.5 | 0.0 | 30.0 | 670.0 | NaN |
| 4 | 11.0 | 4.5 | 0.0 | 30.0 | 670.0 | NaN |

| | Carbohydrates | Protein | Sugar | Dietary_Fiber | Calories_100g \ |
|---|---------------|---------|-------|---------------|-----------------|
| 0 | 30.0 | 15.0 | 1.0 | 1.0 | 293 |

| | | | | | |
|---|------|------|-----|-----|-----|
| 1 | 21.0 | 14.0 | 1.0 | 1.0 | 242 |
| 2 | 22.0 | 13.0 | 3.0 | 1.0 | 253 |
| 3 | 21.0 | 14.0 | 1.0 | 1.0 | 245 |
| 4 | 21.0 | 14.0 | 1.0 | 1.0 | 264 |

| | Total_Fat_100g | Saturated_Fat_100g | Trans_Fat_100g | Cholesterol_100g | \ |
|---|----------------|--------------------|----------------|------------------|---|
| 0 | 12 | 4 | 0.0 | 25 | |
| 1 | 10 | 4 | 0.0 | 33 | |
| 2 | 10 | 4 | 0.0 | 33 | |
| 3 | 11 | 5 | 0.0 | 31 | |
| 4 | 12 | 5 | 0.0 | 33 | |

| | Sodium_100g | Potassium_100g | Carbohydrates_100g | Protein_100g | Sugar_100g | \ |
|---|-------------|----------------|--------------------|--------------|------------|---|
| 0 | 727 | NaN | 30 | 15 | 1.0 | |
| 1 | 978 | NaN | 23 | 15 | 1.0 | |
| 2 | 824 | NaN | 24 | 14 | 3.0 | |
| 3 | 684 | NaN | 21 | 14 | 1.0 | |
| 4 | 736 | NaN | 23 | 15 | 1.0 | |

| | Dietary_Fiber_100g |
|---|--------------------|
| 0 | 1.0 |
| 1 | 1.0 |
| 2 | 1.0 |
| 3 | 1.0 |
| 4 | 1.0 |

```
[3]: foods_string = foods.select_dtypes('object')
```

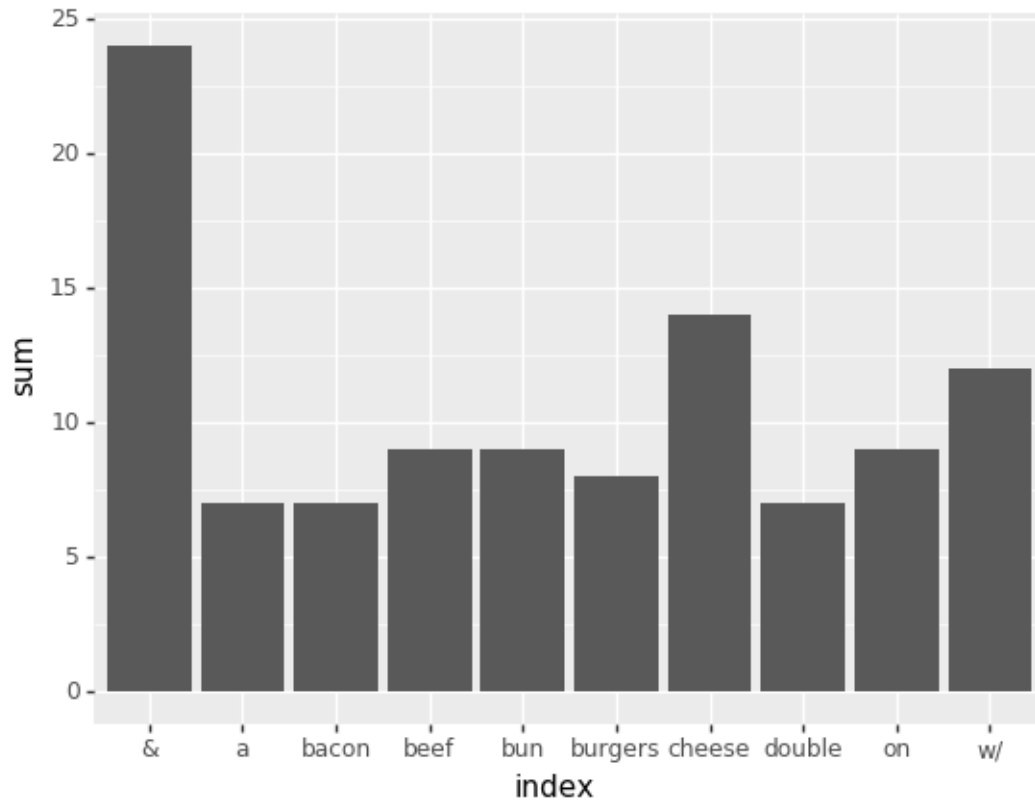
```
[4]: conditions = [(foods_string['Food_Category'] == 'Burgers'),
                    (foods_string['Food_Category'] == 'Pizza'),
                    ((foods_string['Food_Category'] != 'Pizza').all() or
                     →(foods_string['Food_Category'] != 'Burgers').all()))
choices = [0,1,2]

foods_string['binary'] = np.select(conditions, choices, default=2)
```

```
[5]: foods_string['Item_Description'].replace(',', '', regex=True, inplace=True)
foods_string['Item_Description'] = foods_string['Item_Description'].str.lower()
```

```
[6]: burger = foods_string[foods_string['binary'] == 0]
words = burger.Item_Description.sample(n = 20).head(20).apply(lambda x: pd.
→value_counts(x.split(" "))).sum(axis = 0).sort_values(ascending=[False])
words = pd.DataFrame(words, columns=['sum'])
words.reset_index(inplace=True)
```

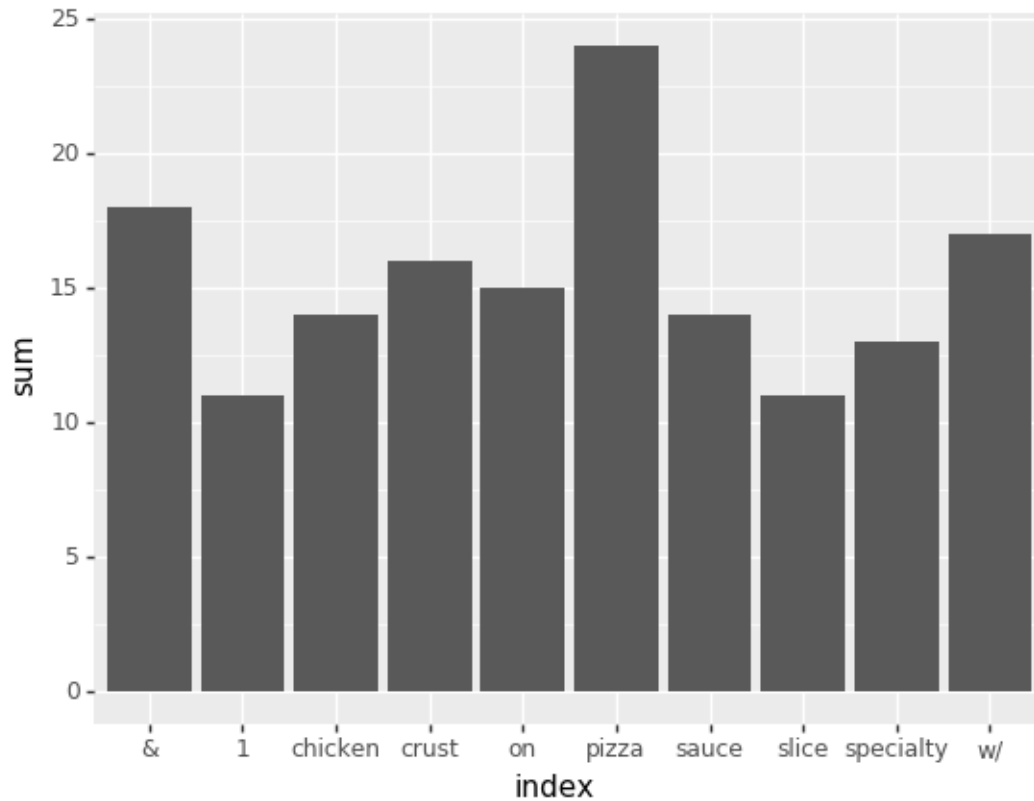
```
[7]: (ggplot(words.head(10),aes(x = 'index', y= 'sum'))+geom_bar(stat='identity'))
```



```
[7]: <ggplot: (-9223371872393059668)>
```

```
[8]: pizza = foods_string[foods_string['binary'] == 1]
words = pizza.Item_Description.sample(n = 20).head(20).apply(lambda x: pd.
    ↪value_counts(x.split(" "))).sum(axis = 0).sort_values(ascending=[False])
words = pd.DataFrame(words, columns=['sum'])
words.reset_index(inplace=True)
```

```
[9]: (ggplot(words.head(10),aes(x = 'index', y= 'sum'))+geom_bar(stat='identity'))
```



```
[9]: <ggplot: (-9223371872392812664)>
```

```
[10]: with open("data/words.yaml") as file:
        words = yaml.load(file, Loader=yaml.FullLoader)

        pizza = words['pizza']
        burger = words['burger']
```

```
[11]: foods_string['pizza_words'] = np.nan
        foods_string['burger_words'] = np.nan
```

```
[12]: for i in range(0, len(foods_string.Item_Description)):
        burger_freq = []
        pizza_freq = []
        for w in foods_string.Item_Description[i].split(" "):
            burger_freq.append(burger.count(w))
            pizza_freq.append(pizza.count(w))

        foods_string.burger_words.loc[i] = sum(burger_freq)
        foods_string.pizza_words.loc[i] = sum(pizza_freq)
```

```
[13]: foods_string.sample(n = 15).head(15)
```

```
[13]:
```

| | Item_Name \ |
|-----|---|
| 173 | Honolulu Hawaiian on Gluten Free Crust, Small |
| 895 | Breakfast Slider w/ Hamburger, Egg & Cheese |
| 456 | Tropical Luau Pizza, Original Crust, Pizza For... |
| 857 | Avocado Bacon Burger Whatameal |
| 477 | BBQ Chicken Bacon Pizza, Pan Crust |
| 13 | Whopper Sandwich w/ Cheese & Mayo |
| 542 | Guinevere's Garden Delight on Original Crust, ... |
| 827 | Green Chile Double Jr |
| 189 | Ultimate Pepperoni on Gluten Free Crust, Small |
| 467 | Chicken & Veggie Pizza, Pan Crust, 1 Slice |
| 180 | Philly Cheese Steak Pizza on Handmade Pan Crus... |
| 312 | Stacked Grilled Cheese Burger Munchie Meal |
| 373 | Tuscan Six Cheese Pizza, Original Crust, Large... |
| 690 | Pepperoni on Pan Crust, Large |
| 272 | Hamburger Meal, Kids |

| | Item_Description | Food_Category | binary \ |
|-----|---|---------------|----------|
| 173 | honolulu hawaiian on gluten free crust w/ robu... | Pizza | 1 |
| 895 | breakfast slider w/ hamburger egg & cheese bre... | Burgers | 0 |
| 456 | tropical luau pizza on original crust w/ julie... | Pizza | 1 |
| 857 | avocado bacon burger whatameal on texas toast ... | Burgers | 0 |
| 477 | bbq chicken bacon pizza on pan crust w/ zesty ... | Pizza | 1 |
| 13 | whopper sandwich w/ cheese & mayo whopper sand... | Burgers | 0 |
| 542 | guinevere's garden delight on original crust w... | Pizza | 1 |
| 827 | green chile double jr on small white toasted b... | Burgers | 0 |
| 189 | ultimate pepperoni on gluten free crust w/ rob... | Pizza | 1 |
| 467 | chicken & veggie pizza on pan crust w/ grilled... | Pizza | 1 |
| 180 | philly cheese steak pizza on handmade pan crus... | Pizza | 1 |
| 312 | stacked grilled cheese burger munchie meal on ... | Burgers | 0 |
| 373 | tuscan six cheese pizza on original crust w/ p... | Pizza | 1 |
| 690 | pepperoni on pan crust large pizzas | Pizza | 1 |
| 272 | hamburger meal kids thickburgers & sandwiches | Burgers | 0 |

| | pizza_words | burger_words |
|-----|-------------|--------------|
| 173 | 6.0 | 1.0 |
| 895 | 0.0 | 1.0 |
| 456 | 4.0 | 0.0 |
| 857 | 1.0 | 3.0 |
| 477 | 4.0 | 0.0 |
| 13 | 0.0 | 0.0 |
| 542 | 4.0 | 0.0 |
| 827 | 0.0 | 5.0 |
| 189 | 7.0 | 1.0 |
| 467 | 5.0 | 0.0 |

| | | |
|-----|-----|-----|
| 180 | 4.0 | 1.0 |
| 312 | 0.0 | 0.0 |
| 373 | 3.0 | 0.0 |
| 690 | 3.0 | 0.0 |
| 272 | 0.0 | 0.0 |

1.1.1 Logistic Regression

I am using the created common words count variables as the predictor variables because the description of the food is much more telling of what a food is than any nutritional information. Also if the data is there, might as well use it. I used a traditional train test split for this model because it is the easiest to implement and think other methods are overkill in this situation. I did not standardize any variables because both variables are on the same scale. The model did very well given the accuracy score of .94 meaning that the model correctly classified 94% of the data. False positives rate was also very high meaning that the outcome predicted is trustable and precise such that that 99% of the times the model predicts a positive (pizza) outcome it is correct. Sensitivity score is fairly high (93%) suggesting that the model will falsely classify 7% of pizzas as burgers.

```
[14]: predictors = {'pizza_words', 'burger_words'}
X_train, X_test, y_train, y_test = train_test_split(foods_string[predictors],
↳ foods_string["binary"], test_size=0.4)
```

```
[15]: model = LogisticRegression()
model.fit(X_train, y_train)

train_pred = model.predict(X_train)
test_pred = model.predict(X_test)
```

```
[16]: foods_string.reset_index(inplace=True)
test_copy = X_test.copy()
test_copy.reset_index(inplace=True)

test_copy = pd.merge(test_copy,
↳ foods_string[['index', 'Item_Description', 'binary']], on='index', how='left')
test_copy['log_preds'] = test_pred
```

```
[17]: print("Accuracy - TP+TN/TP+FP+FN+TN:", accuracy_score(y_test, test_pred))
print("Precision - TP/TP+FP:", precision_score(y_test, test_pred)) #relates to
↳ low false positivity
print("Recall/Sensitivity/TPR - TP/TP+FN:", recall_score(y_test, test_pred))
```

Accuracy - TP+TN/TP+FP+FN+TN: 0.9407008086253369

Precision - TP/TP+FP: 0.9877551020408163

Recall/Sensitivity/TPR - TP/TP+FN: 0.9272030651340997

1.1.2 Decision Tree

I am using the created common words count variables as the predictor variables because the description of the food is much more telling of what a food is than any nutritional information. Also if the data is there, might as well use it. I used a traditional train test split for this model because it is the easiest to implement and think other methods are overkill in this situation. I did not standardize any variables because both variables are on the same scale. The model did very well given the accuracy score of .92 meaning that the model correctly classified 92% of the data. False positives rate was also high meaning that the outcome predicted is trustable and precise such that that 89% of the times the model predicts a positive (pizza) outcome it is correct. Sensitivity score is suspiciously high (100%) suggesting that the model will default to pizza.

```
[18]: tree = DecisionTreeClassifier()
      tree.fit(X_train, y_train)

      y_pred = tree.predict(X_test)
```

```
[19]: test_copy['dt_preds'] = y_pred
```

```
[20]: print("Accuracy - TP+TN/TP+FP+FN+TN:", accuracy_score(y_test, y_pred))
      print("Precision - TP/TP+FP:", precision_score(y_test, y_pred)) #relates to low_
      ↪ false positivity
      print("Recall/Sensitivity/TPR - TP/TP+FN:", recall_score(y_test, y_pred))
```

Accuracy - TP+TN/TP+FP+FN+TN: 0.954177897574124

Precision - TP/TP+FP: 1.0

Recall/Sensitivity/TPR - TP/TP+FN: 0.9348659003831418

1.1.3 Naive-Bayes

I am using the created common words count variables as the predictor variables because the description of the food is much more telling of what a food is than any nutritional information. Also if the data is there, might as well use it. I used a traditional train test split for this model because it is the easiest to implement and think other methods are overkill in this situation. I did not standardize any variables because both variables are on the same scale. The model did very well given the accuracy score of .91 meaning that the model correctly classified 91% of the data. False positives rate was also high meaning that the outcome predicted is trustable and precise such that that 90% of the times the model predicts a positive (pizza) outcome it is correct. Sensitivity score is suspiciously high (99%) suggesting that the model will most likely default to pizza.

```
[21]: nb = GaussianNB()

      nb.fit(X_train, y_train)

      y_pred = nb.predict(X_test)
```

```
[22]: test_copy['nb_preds'] = y_pred
```



```
[23]: print("Accuracy - TP+TN/TP+FP+FN+TN:", accuracy_score(y_test, y_pred))
      print("Precision - TP/TP+FP:", precision_score(y_test, y_pred)) #relates to low
      → false positivity
      print("Recall/Sensitivity/TPR - TP/TP+FN:", recall_score(y_test, y_pred))
```

```
Accuracy - TP+TN/TP+FP+FN+TN: 0.9487870619946092
Precision - TP/TP+FP: 0.9416058394160584
Recall/Sensitivity/TPR - TP/TP+FN: 0.9885057471264368
```

1.1.4 Comparison

Looking at all the models performances in accuracy, you can clearly see the small differences in each model. Both the logistic and decision tree models have high true positives as indicated in the accuracy score as well as 0 false positives. However it is interesting to note that the naive bayes model has a significantly lower false positives meaning that this model is bias and is likely to predict pizza more than burger. In a case where outcome is more important to predict a positive, it might be better to use this model, but since were predicting pizza and burgers id say this is the worse models. Differences in the models can also be seen actually inspecting the output and it is clear that the models have a tough time when word counts are equal. This can be easily solved by adding more descriptive words in the ingredient list such as mayo and triple since these words are more used in describing burgers than pizza.

```
[24]: cnf_matrix = confusion_matrix(test_copy['binary'], test_copy['log_preds'])
      cnf_matrix
```

```
[24]: array([[107,  3],
            [ 19, 242]], dtype=int64)
```

```
[25]: cnf_matrix = confusion_matrix(test_copy['binary'], test_copy['dt_preds'])
      cnf_matrix
```

```
[25]: array([[110,  0],
            [ 17, 244]], dtype=int64)
```

```
[26]: cnf_matrix = confusion_matrix(test_copy['binary'], test_copy['nb_preds'])
      cnf_matrix
```

```
[26]: array([[ 94, 16],
            [  3, 258]], dtype=int64)
```

```
[27]: test_copy[(test_copy['binary'] != test_copy['log_preds']).all() or
              (test_copy['binary'] != test_copy['dt_preds']).all() or
              (test_copy['binary'] != test_copy['nb_preds'])]
```

```
[27]:
```

| | index | burger_words | pizza_words | \ |
|----|-------|--------------|-------------|---|
| 1 | 254 | 0.0 | 0.0 | |
| 24 | 283 | 0.0 | 0.0 | |
| 32 | 454 | 2.0 | 2.0 | |

| | | | |
|-----|-----|-----|-----|
| 53 | 256 | 0.0 | 0.0 |
| 55 | 272 | 0.0 | 0.0 |
| 88 | 327 | 0.0 | 0.0 |
| 113 | 81 | 0.0 | 0.0 |
| 128 | 76 | 0.0 | 0.0 |
| 142 | 301 | 0.0 | 0.0 |
| 198 | 79 | 0.0 | 0.0 |
| 199 | 452 | 2.0 | 2.0 |
| 209 | 255 | 0.0 | 0.0 |
| 240 | 329 | 0.0 | 0.0 |
| 249 | 18 | 0.0 | 0.0 |
| 291 | 698 | 1.0 | 1.0 |
| 324 | 22 | 0.0 | 0.0 |
| 327 | 85 | 0.0 | 0.0 |
| 355 | 84 | 0.0 | 0.0 |
| 367 | 313 | 0.0 | 0.0 |

| | Item_Description | binary | log_preds \ |
|-----|---|--------|-------------|
| 1 | little cheeseburger little burgers | 0 | 0 |
| 24 | 1/2 lb frisco thickburger thickburgers & sandw... | 0 | 0 |
| 32 | the big bonanza on original crust w/ bbq sauce... | 1 | 0 |
| 53 | little bacon cheeseburger | 0 | 0 |
| 55 | hamburger meal kids thickburgers & sandwiches | 0 | 0 |
| 88 | bbq bacon & cheese ba lunch & dinner | 0 | 0 |
| 113 | cheddar butterburger w/ bacon single butterbur... | 0 | 0 |
| 128 | butterburger the original butterburgers kids m... | 0 | 0 |
| 142 | hamburger burgers & more kids & adults menu | 0 | 0 |
| 198 | butterburger cheese triple butterburgers | 0 | 0 |
| 199 | the big bonanza on original crust w/ bbq sauce... | 1 | 0 |
| 209 | little bacon burger | 0 | 0 |
| 240 | chili cheese krystal lunch & dinner | 0 | 0 |
| 249 | bk veggie burger w/out mayo chicken & more | 0 | 0 |
| 291 | tomato pesto flatbreads artisan flatbread full... | 1 | 0 |
| 324 | whopper sandwich w/ cheese w/out mayo whopper ... | 0 | 0 |
| 327 | cheddar butterburger butterburgers triple | 0 | 0 |
| 355 | cheddar butterburger single butterburgers; cho... | 0 | 0 |
| 367 | hella peno burger munchie meal w/ 20 fl oz coc... | 0 | 0 |

| | dt_preds | nb_preds |
|-----|----------|----------|
| 1 | 0 | 1 |
| 24 | 0 | 1 |
| 32 | 1 | 0 |
| 53 | 0 | 1 |
| 55 | 0 | 1 |
| 88 | 0 | 1 |
| 113 | 0 | 1 |
| 128 | 0 | 1 |

| | | |
|-----|---|---|
| 142 | 0 | 1 |
| 198 | 0 | 1 |
| 199 | 1 | 0 |
| 209 | 0 | 1 |
| 240 | 0 | 1 |
| 249 | 0 | 1 |
| 291 | 0 | 0 |
| 324 | 0 | 1 |
| 327 | 0 | 1 |
| 355 | 0 | 1 |
| 367 | 0 | 1 |

1.2 Part II

Use the dataset *KrispyKreme.csv* to build 2 clustering models (you can use any of the clustering models we've learned).

For each model:

0. Explore data (with ggplot)
1. Explain which variables you're using to predict the outcome.
2. Evaluate how the model performed using silhouette scores. Look at different numbers of clusters (like $k = 3, 5, \dots$). Which number of clusters is the best fit?
3. Describe the clusters (what are they like? how are they different)

At the end:

4. Compare the clusters obtained by the two models. Overall are they similar? or really different (i.e. do they contain mostly the same members?)

Please get rid of extra analyses/superfluous code before turning it in. Turn in A PDF on Blackboard.

```
[28]: donuts = pd.read_csv('data/KrispyKreme.csv')
      donuts.head(10)
```

```
[28]:
```

| | Restaurant_Item_Name | restaurant | \ |
|---|---|--------------|---|
| 0 | Krispy Kreme Apple Fritter | Krispy Kreme | |
| 1 | Krispy Kreme Chocolate Iced Cake Doughnut | Krispy Kreme | |
| 2 | Krispy Kreme Chocolate Iced Custard Filled Dou... | Krispy Kreme | |
| 3 | Krispy Kreme Chocolate Iced Glazed Doughnut | Krispy Kreme | |
| 4 | Krispy Kreme Chocolate Iced Glazed Cruller Dou... | Krispy Kreme | |
| 5 | Krispy Kreme Football Doughnut | Krispy Kreme | |
| 6 | Krispy Kreme Chocolate Iced w/ Kreme Filling | Krispy Kreme | |
| 7 | Krispy Kreme Chocolate Iced Glazed Doughnut w/... | Krispy Kreme | |
| 8 | Krispy Kreme Chocolate Iced Raspberry Filled D... | Krispy Kreme | |
| 9 | Krispy Kreme Cinnamon Apple Filled Doughnut | Krispy Kreme | |

| Restaurant_ID | Item_Name | \ |
|---------------|-----------|---------------|
| 0 | 49 | Apple Fritter |

| | | |
|---|----|--|
| 1 | 49 | Chocolate Iced Cake Doughnut |
| 2 | 49 | Chocolate Iced Custard Filled Doughnut |
| 3 | 49 | Chocolate Iced Glazed Doughnut |
| 4 | 49 | Chocolate Iced Glazed Cruller Doughnut |
| 5 | 49 | Football Doughnut |
| 6 | 49 | Chocolate Iced w/ Kreme Filling |
| 7 | 49 | Chocolate Iced Glazed Doughnut w/ Rainbow Spr... |
| 8 | 49 | Chocolate Iced Raspberry Filled Doughnut |
| 9 | 49 | Cinnamon Apple Filled Doughnut |

| | Item_Description | Food_Category | \ |
|---|---|---------------|---|
| 0 | Apple Fritter, Doughnuts | Baked Goods | |
| 1 | Chocolate Iced Cake Doughnut, Doughnuts | Baked Goods | |
| 2 | Chocolate Iced Custard Filled Doughnut, Doughnuts | Baked Goods | |
| 3 | Chocolate Iced Glazed Doughnut, Doughnuts | Baked Goods | |
| 4 | Chocolate Iced Glazed Cruller Doughnut, Doughnuts | Baked Goods | |
| 5 | Football Doughnut, Doughnuts | Baked Goods | |
| 6 | Chocolate Iced w/ Kreme Filling, Doughnuts | Baked Goods | |
| 7 | Chocolate Iced Glazed Doughnut w/ Rainbow Spr... | Baked Goods | |
| 8 | Chocolate Iced Raspberry Filled Doughnut w/ Tw... | Baked Goods | |
| 9 | Cinnamon Apple Filled Doughnut, Doughnuts | Baked Goods | |

| | Serving_Size | Serving_Size_text | Serving_Size_Unit | Serving_Size_household | \ |
|---|--------------|-------------------|-------------------|------------------------|---|
| 0 | 100 | NaN | g | NaN | |
| 1 | 71 | NaN | g | NaN | |
| 2 | 85 | NaN | g | NaN | |
| 3 | 63 | NaN | g | NaN | |
| 4 | 70 | NaN | g | NaN | |
| 5 | 71 | NaN | g | NaN | |
| 6 | 85 | NaN | g | NaN | |
| 7 | 66 | NaN | g | NaN | |
| 8 | 88 | NaN | g | NaN | |
| 9 | 76 | NaN | g | NaN | |

| | Calories | Total_Fat | Saturated_Fat | Trans_Fat | Cholesterol | Sodium | \ |
|---|----------|-----------|---------------|-----------|-------------|--------|---|
| 0 | 350 | 19.0 | 9.0 | 0.0 | 0 | 110 | |
| 1 | 280 | 13.0 | 5.0 | 0.0 | 25 | 310 | |
| 2 | 300 | 15.0 | 7.0 | 0.0 | 0 | 140 | |
| 3 | 240 | 11.0 | 5.0 | 0.0 | 0 | 90 | |
| 4 | 260 | 10.0 | 4.0 | 0.0 | 20 | 270 | |
| 5 | 310 | 16.0 | 7.0 | 0.0 | 0 | 140 | |
| 6 | 350 | 19.0 | 9.0 | 0.0 | 0 | 140 | |
| 7 | 250 | 11.0 | 5.0 | 0.0 | 0 | 90 | |
| 8 | 310 | 15.0 | 7.0 | 0.0 | 0 | 140 | |
| 9 | 270 | 14.0 | 7.0 | 0.0 | 0 | 135 | |

| Potassium | Carbohydrates | Protein | Sugar | Dietary_Fiber | Calories_100g | \ |
|-----------|---------------|---------|-------|---------------|---------------|---|
|-----------|---------------|---------|-------|---------------|---------------|---|

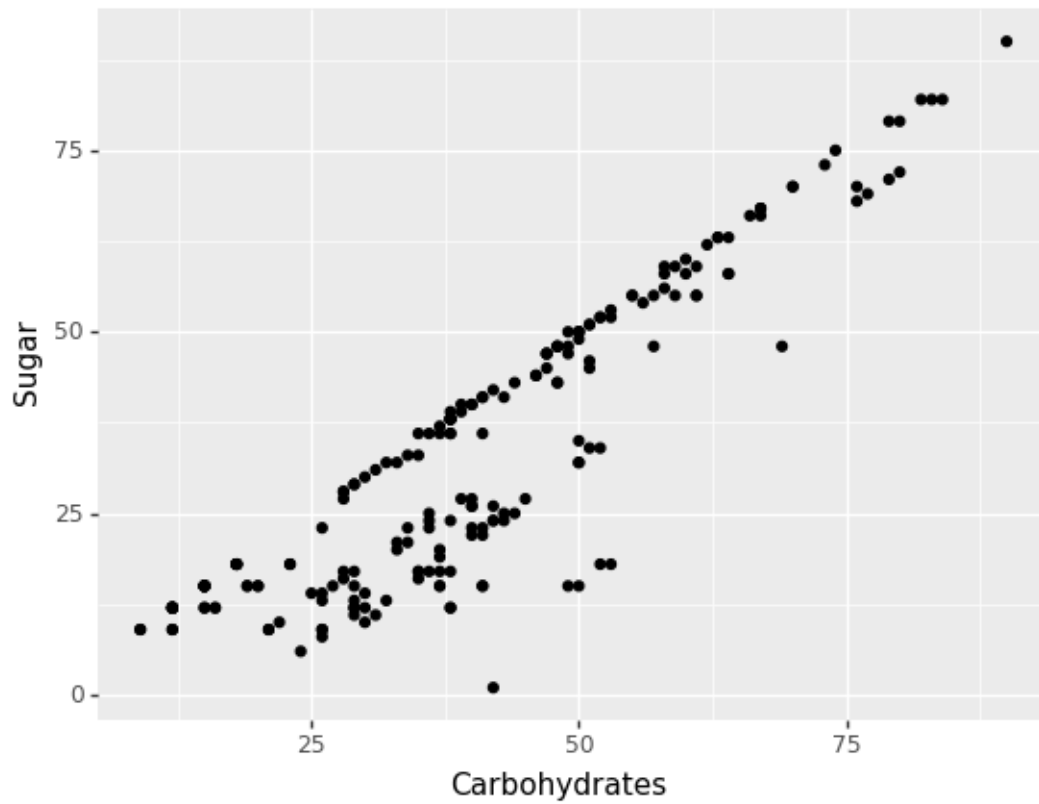
| | | | | | | |
|---|------|----|---|----|-----|-----|
| 0 | 45.0 | 42 | 4 | 26 | 1.0 | 350 |
| 1 | 35.0 | 37 | 3 | 19 | NaN | 394 |
| 2 | 50.0 | 37 | 4 | 17 | 1.0 | 353 |
| 3 | 35.0 | 33 | 3 | 20 | NaN | 381 |
| 4 | 20.0 | 40 | 3 | 26 | NaN | 371 |
| 5 | NaN | 35 | 7 | 16 | 3.0 | 437 |
| 6 | 50.0 | 41 | 4 | 23 | 1.0 | 412 |
| 7 | 35.0 | 36 | 3 | 23 | NaN | 379 |
| 8 | 50.0 | 41 | 4 | 22 | 1.0 | 352 |
| 9 | 50.0 | 31 | 4 | 11 | 1.0 | 355 |

| | Total_Fat_100g | Saturated_Fat_100g | Trans_Fat_100g | Cholesterol_100g | \ |
|---|----------------|--------------------|----------------|------------------|---|
| 0 | 19 | 9 | 0 | 0 | |
| 1 | 18 | 7 | 0 | 35 | |
| 2 | 18 | 8 | 0 | 0 | |
| 3 | 17 | 8 | 0 | 0 | |
| 4 | 14 | 6 | 0 | 29 | |
| 5 | 23 | 10 | 0 | 0 | |
| 6 | 22 | 11 | 0 | 0 | |
| 7 | 17 | 8 | 0 | 0 | |
| 8 | 17 | 8 | 0 | 0 | |
| 9 | 18 | 9 | 0 | 0 | |

| | Sodium_100g | Potassium_100g | Carbohydrates_100g | Protein_100g | Sugar_100g | \ |
|---|-------------|----------------|--------------------|--------------|------------|---|
| 0 | 110 | 45.0 | 42 | 4 | 26 | |
| 1 | 437 | 49.0 | 52 | 4 | 27 | |
| 2 | 165 | 59.0 | 44 | 5 | 20 | |
| 3 | 143 | 56.0 | 52 | 5 | 32 | |
| 4 | 386 | 29.0 | 57 | 4 | 37 | |
| 5 | 197 | NaN | 49 | 10 | 23 | |
| 6 | 165 | 59.0 | 48 | 5 | 27 | |
| 7 | 136 | 53.0 | 55 | 5 | 35 | |
| 8 | 159 | 57.0 | 47 | 5 | 25 | |
| 9 | 178 | 66.0 | 41 | 5 | 14 | |

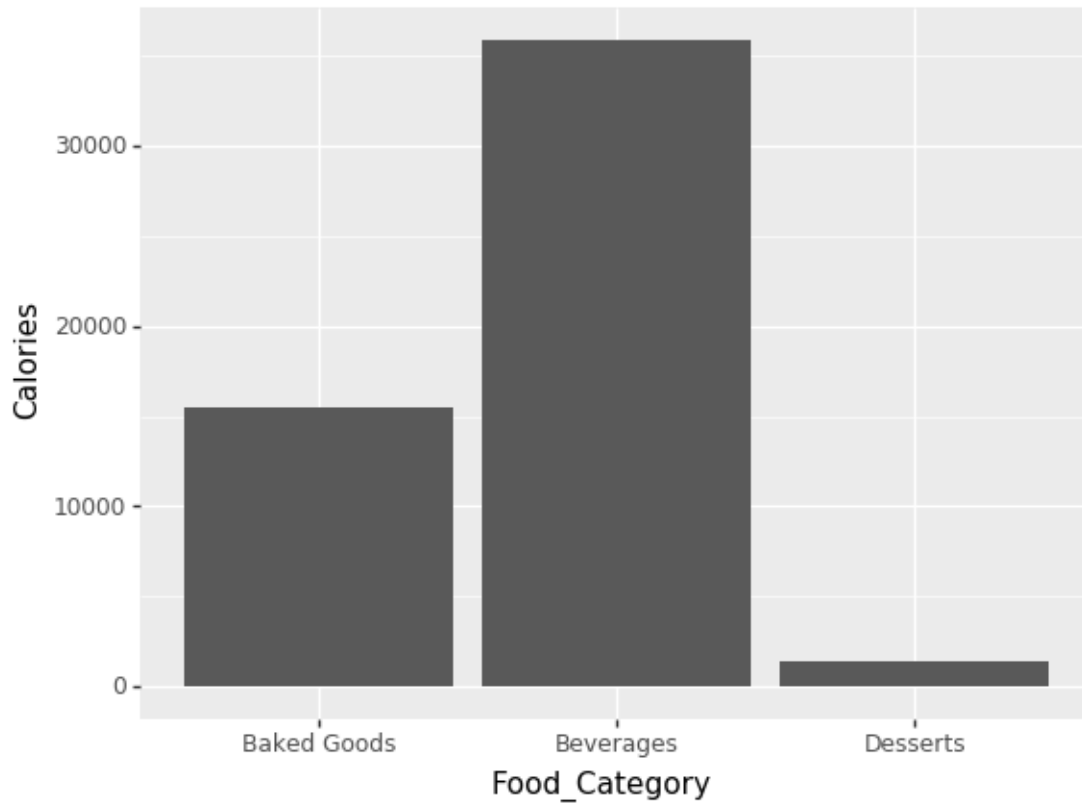
| | Dietary_Fiber_100g |
|---|--------------------|
| 0 | 1.0 |
| 1 | NaN |
| 2 | 1.0 |
| 3 | NaN |
| 4 | NaN |
| 5 | 4.0 |
| 6 | 1.0 |
| 7 | NaN |
| 8 | 1.0 |
| 9 | 1.0 |

```
[29]: (ggplot(donuts, aes(x = 'Carbohydrates', y= 'Sugar'))+geom_point())
```



```
[29]: <ggplot: (-9223371872392741644)>
```

```
[30]: (ggplot(donuts, aes(x = 'Food_Category', y=␣  
↪ 'Calories'))+geom_bar(stat='identity'))
```



```
[30]: <ggplot: (-9223371872392751644)>
```

```
[31]: donuts[donuts['Food_Category'] == 'Desserts'].head(10)
```

```
[31]:
```

| | Restaurant_Item_Name | restaurant | \ |
|-----|---|--------------|---|
| 110 | Krispy Kreme Chocolate Kool Kreme Cone | Krispy Kreme | |
| 111 | Krispy Kreme Chocolate Vanilla Swirl Kool Krem... | Krispy Kreme | |
| 112 | Krispy Kreme Vanilla Kool Kreme Cone | Krispy Kreme | |
| 162 | Krispy Kreme Kool Kreme Original Glazed Doughn... | Krispy Kreme | |

| | Restaurant_ID | Item_Name | \ |
|-----|---------------|--|---|
| 110 | 49 | Chocolate Kool Kreme Cone | |
| 111 | 49 | Chocolate Vanilla Swirl Kool Kreme Cone | |
| 112 | 49 | Vanilla Kool Kreme Cone | |
| 162 | 49 | Kool Kreme Original Glazed Doughnut Sundae | |

| | Item_Description | Food_Category | \ |
|-----|---|---------------|---|
| 110 | Chocolate Kool Kreme Cone, Kool Kreme | Desserts | |
| 111 | Chocolate Vanilla Swirl Kool Kreme Cone, Kool ... | Desserts | |
| 112 | Vanilla Kool Kreme Cone, Kool Kreme | Desserts | |
| 162 | Kool Kreme Original Glazed Doughnut Sundae, Ko... | Desserts | |

| | Serving_Size | Serving_Size_text | Serving_Size_Unit | Serving_Size_household | \ |
|-----|--------------|-------------------|-------------------|------------------------|---|
| 110 | 182 | NaN | g | NaN | |
| 111 | 182 | NaN | g | NaN | |
| 112 | 182 | NaN | g | NaN | |
| 162 | 199 | NaN | g | NaN | |

| | Calories | Total_Fat | Saturated_Fat | Trans_Fat | Cholesterol | Sodium | \ |
|-----|----------|-----------|---------------|-----------|-------------|--------|---|
| 110 | 310 | 9.0 | 6.0 | 0.0 | 30 | 170 | |
| 111 | 310 | 9.0 | 6.0 | 0.0 | 30 | 180 | |
| 112 | 300 | 8.0 | 5.0 | 0.0 | 30 | 180 | |
| 162 | 450 | 17.0 | 9.0 | 0.0 | 20 | 220 | |

| | Potassium | Carbohydrates | Protein | Sugar | Dietary_Fiber | Calories_100g | \ |
|-----|-----------|---------------|---------|-------|---------------|---------------|---|
| 110 | NaN | 52 | 9 | 34 | 2.0 | 170 | |
| 111 | NaN | 51 | 8 | 34 | 1.0 | 170 | |
| 112 | NaN | 50 | 8 | 35 | 0.0 | 165 | |
| 162 | NaN | 69 | 8 | 48 | 1.0 | 226 | |

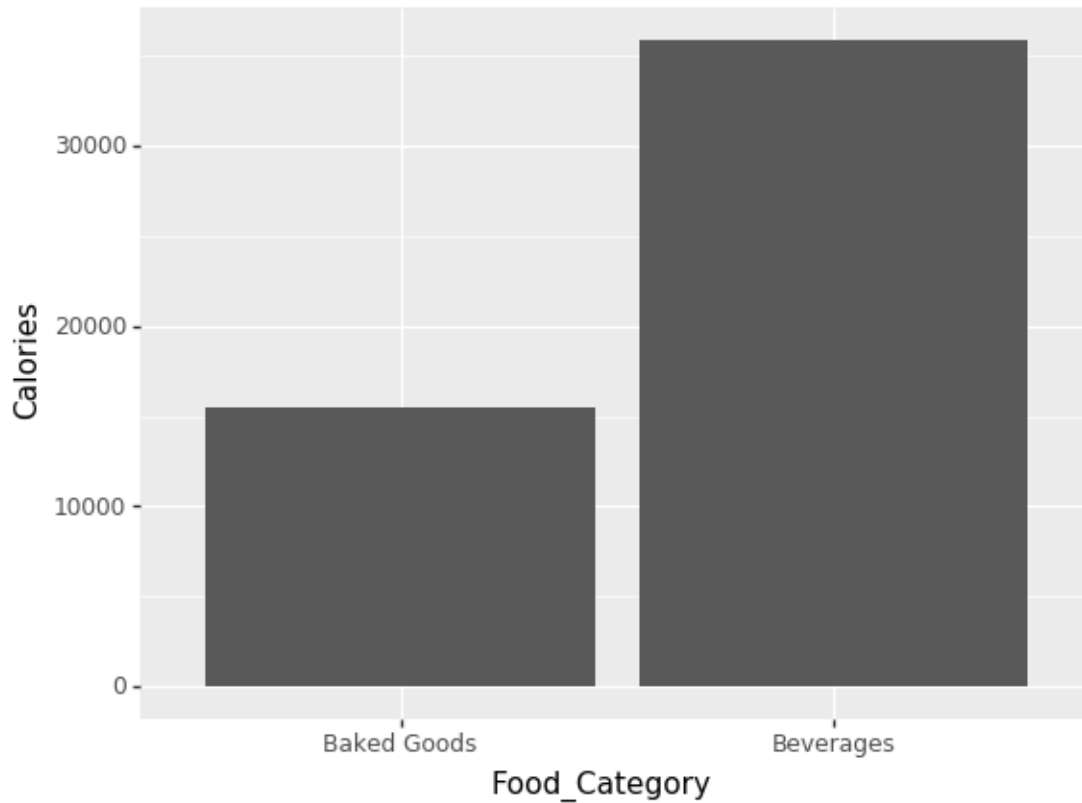
| | Total_Fat_100g | Saturated_Fat_100g | Trans_Fat_100g | Cholesterol_100g | \ |
|-----|----------------|--------------------|----------------|------------------|---|
| 110 | 5 | 3 | 0 | 16 | |
| 111 | 5 | 3 | 0 | 16 | |
| 112 | 4 | 3 | 0 | 16 | |
| 162 | 9 | 5 | 0 | 10 | |

| | Sodium_100g | Potassium_100g | Carbohydrates_100g | Protein_100g | \ |
|-----|-------------|----------------|--------------------|--------------|---|
| 110 | 93 | NaN | 29 | 5 | |
| 111 | 99 | NaN | 28 | 4 | |
| 112 | 99 | NaN | 27 | 4 | |
| 162 | 111 | NaN | 35 | 4 | |

| | Sugar_100g | Dietary_Fiber_100g |
|-----|------------|--------------------|
| 110 | 19 | 1.0 |
| 111 | 19 | 1.0 |
| 112 | 19 | 0.0 |
| 162 | 24 | 1.0 |

```
[32]: donuts.drop(donuts.loc[donuts['Food_Category']=='Desserts'].index, inplace=True)
```

```
[33]: (ggplot(donuts,aes(x = 'Food_Category', y=
↳ 'Calories'))+geom_bar(stat='identity'))
```

```
[33]: <ggplot: (-9223371872392711676)>
```

```
[34]: le = LabelEncoder()

donuts['Food_Label'] = le.fit_transform(donuts['Food_Category'])
```

```
[35]: # for an approximate result, multiply the mass value by 28.35
mask = donuts['Serving_Size_Unit'] == 'oz'
donuts.loc[mask, 'Serving_Size'] = donuts.loc[mask, 'Serving_Size'].
    ↪ apply(lambda oz: oz * 28.38)
```

```
[36]: vars = ['Item_Name', 'Item_Description', 'Food_Label', 'Serving_Size',
    ↪ 'Total_Fat', 'Carbohydrates', 'Protein']
features = ['Food_Label', 'Serving_Size', 'Total_Fat', 'Carbohydrates',
    ↪ 'Protein']
```

```
[37]: donuts = donuts[vars]
```

```
[38]: X = donuts[features]
```

```
[39]: z = StandardScaler()
X[features] = z.fit_transform(X)
```

1.2.1 Model

Im using the scaled serving size, label category, and macro nutrients for each items as my variables because every other variable can be derived from these main variables for example sugar at the end of the day is still accounted for in carbs, and same goes for saturated fat being a type of fat. Overall, models had better silhouette scores around the 7th cluster with the highest being in the hierarchy model. I decided to compare clusters using only 3 cluster groups because its a bit easier to make sense of the data as such.

```
[40]: n_components = [2,3,4,5,6,7]
Xdf = X
```

1.2.2 KM

```
[41]: sils = []

for n in n_components:
    km = KMeans(n_clusters = n)
    km.fit(X)
    colName = str(n) + "KM_assign"
    clusters = km.predict(X)

    Xdf[colName] = clusters

    sils.append(silhouette_score(X, clusters))

print(sils)
```

```
[0.5610192871178572, 0.5614802152714, 0.6024903093907765, 0.6762161744667429,
0.7186735820004582, 0.7732165794943456]
```

1.2.3 GM

```
[42]: sils = []

for n in n_components:
    gmm = GaussianMixture(n_components = n)
    gmm.fit(X)
    colName = str(n) + "GM_assign"
    clusters = gmm.predict(X)

    Xdf[colName] = clusters

    sils.append(silhouette_score(X, clusters))
```

```
print(sils)
```

```
[0.39439745385924735, 0.39881460016296566, 0.5105513755240202,  
0.6484137381700464, 0.7622245540064875, 0.8238416784401827]
```

1.2.4 HAC

```
[43]: sils = []  
  
for n in n_components:  
    hac = AgglomerativeClustering(n_clusters = n,  
                                  affinity = "euclidean",  
                                  linkage = "ward")  
  
    hac.fit(X)  
  
    colName = str(n) + "HAC_assign"  
    clusters = hac.labels_  
  
    Xdf[colName] = clusters  
  
    sils.append(silhouette_score(X, clusters))  
  
print(sils)
```

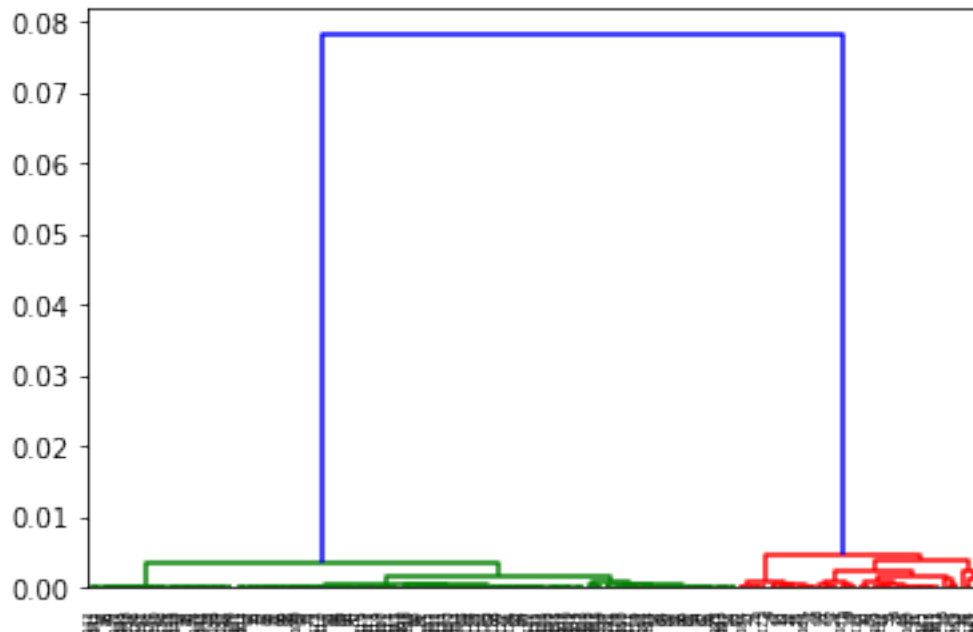
```
[0.36324473479028224, 0.48392684149661913, 0.604769567614849,  
0.7151308149515027, 0.7957302279050571, 0.8483062036393453]
```

```
[44]: groups = ['3GM_assign', '3KM_assign', '3HAC_assign', 'index']
```

```
[45]: donuts.reset_index(inplace = True)  
Xdf.reset_index(inplace = True)
```

```
[46]: donuts = pd.merge(donuts, Xdf[groups], on='index')
```

```
[47]: dendro = sch.dendrogram(sch.linkage(donuts[features], metric = "cosine",  
↪method='average'))
```



1.2.5 Analysis

Overall all models did well distinguishing what was a drink and what was a baked good item most likely from the drastic jump in serving size. Think about this now I think it would have been more interesting to see if it would make the distinction from the macronutrients. I think it would because of the jump in protein between both foods. Although I wanted this model to distinguish between healthy vs non healthy items, i got more of a separation between drinks with skim milk and “skinny” options vs drinks with added confections.

```
[60]: donuts = donuts.sort_values(by=['Carbohydrates'], ascending=[False])
print(donuts[donuts['Food_Label'] == 1].describe())
donuts[donuts['Food_Label'] == 1].head(60)
```

| | index | Food_Label | Serving_Size | Total_Fat | Carbohydrates \ |
|-------|------------|------------|--------------|------------|-----------------|
| count | 146.000000 | 146.0 | 146.000000 | 146.000000 | 146.000000 |
| mean | 113.917808 | 1.0 | 453.302466 | 4.390411 | 42.595890 |
| std | 47.923719 | 0.0 | 92.845084 | 3.786771 | 19.968105 |
| min | 37.000000 | 1.0 | 340.560000 | 0.000000 | 9.000000 |
| 25% | 73.250000 | 1.0 | 340.560000 | 0.000000 | 28.000000 |
| 50% | 111.000000 | 1.0 | 454.080000 | 4.000000 | 42.000000 |
| 75% | 149.750000 | 1.0 | 567.600000 | 7.000000 | 56.750000 |
| max | 204.000000 | 1.0 | 567.600000 | 14.000000 | 90.000000 |

| | Protein | 3GM_assign | 3KM_assign | 3HAC_assign |
|-------|------------|------------|------------|-------------|
| count | 146.000000 | 146.000000 | 146.000000 | 146.000000 |
| mean | 8.589041 | 1.315068 | 0.931507 | 0.616438 |

| | | | | |
|-----|-----------|----------|----------|----------|
| std | 3.512875 | 0.466142 | 1.001086 | 0.926694 |
| min | 0.000000 | 1.000000 | 0.000000 | 0.000000 |
| 25% | 6.000000 | 1.000000 | 0.000000 | 0.000000 |
| 50% | 8.500000 | 1.000000 | 0.000000 | 0.000000 |
| 75% | 10.750000 | 2.000000 | 2.000000 | 2.000000 |
| max | 17.000000 | 2.000000 | 2.000000 | 2.000000 |

| | index | Item_Name \ |
|-----|-------|---|
| 169 | 173 | Frozen Vanilla Latte, 20 oz |
| 166 | 170 | Lemonade Chiller, 16 oz |
| 197 | 201 | Caramel Latte w/ Skim Milk, 20 oz |
| 194 | 198 | Caramel Latte w/ 2% Milk, 20 oz |
| 67 | 67 | Iced Caramel Latte w/ Skim Milk, 20 oz |
| 55 | 55 | Mocha Latte w/ Skim Milk, 20 oz |
| 52 | 52 | Mocha Latte w/ 2% Milk, 20 oz |
| 64 | 64 | Iced Caramel Latte w/ 2% Milk, 20 oz |
| 61 | 61 | Iced Mocha w/ Skim Milk, 20 oz |
| 180 | 184 | Frozen Caramel Latte, 20 oz |
| 58 | 58 | Iced Mocha w/ 2% Milk, 20 oz |
| 168 | 172 | Frozen Vanilla Latte, 16 oz |
| 172 | 176 | Frozen Mocha, 20 oz |
| 121 | 124 | Caramel Mocha w/ 2% Milk, 20 oz |
| 118 | 121 | Caramel Mocha w/ Skim Milk, 20 oz |
| 115 | 118 | Iced Caramel Mocha w/ 2% Milk, 20 oz |
| 196 | 200 | Caramel Latte w/ Skim Milk, 16 oz |
| 112 | 115 | Iced Caramel Mocha w/ Skim Milk, 20 oz |
| 193 | 197 | Caramel Latte w/ 2% Milk, 16 oz |
| 51 | 51 | Mocha Latte w/2% Milk, 16 oz |
| 66 | 66 | Iced Caramel Latte w/ Skim Milk, 16 oz |
| 54 | 54 | Mocha Latte w/ Skim Milk, 16 oz |
| 97 | 97 | Vanilla Latte w/ Skim Milk, 20 oz |
| 63 | 63 | Iced Caramel Latte w/ 2% Milk, 16 oz |
| 94 | 94 | Vanilla Latte w/ 2% Milk, 20 oz |
| 60 | 60 | Iced Mocha w/ Skim Milk, 16 oz |
| 57 | 57 | Iced Mocha w/ 2% Milk, 16 oz |
| 73 | 73 | Hazelnut Latte w/ Skim Milk, 20 oz |
| 103 | 103 | Iced Vanilla Latte w/ Skim Milk, 20 oz |
| 70 | 70 | Hazelnut Latte w/ 2% Milk, 20 oz |
| 100 | 100 | Iced Vanilla Latte w/ 2% Milk, 20 oz |
| 179 | 183 | Frozen Caramel Latte, 16 oz |
| 79 | 79 | Iced Hazelnut Latte w/ Skim Milk, 20 oz |
| 171 | 175 | Frozen Mocha, 16 oz |
| 167 | 171 | Frozen Vanilla Latte, 12 oz |
| 37 | 37 | Hot Chocolate w/ 2% Milk, 12 oz |
| 76 | 76 | Iced Hazelnut Latte w/ 2% Milk, 20 oz |
| 200 | 204 | Jolly Rancher Blue Raspberry Chiller, 20 oz |
| 117 | 120 | Caramel Mocha w/ Skim Milk, 16 oz |
| 120 | 123 | Caramel Mocha w/ 2% Milk, 16 oz |

| | | |
|-----|-----|---|
| 135 | 138 | Skinny Sugar Free Caramel Latte w/ Skim Milk, ... |
| 195 | 199 | Caramel Latte w/ Skim Milk, 12 oz |
| 192 | 196 | Caramel Latte w/ 2% Milk, 12 oz |
| 114 | 117 | Iced Caramel Mocha w/ 2% Milk, 16 oz |
| 134 | 137 | Skinny Sugar Free Caramel Latte w/ 2% Milk, 20 oz |
| 111 | 114 | Iced Caramel Mocha w/ Skim Milk, 16 oz |
| 158 | 161 | Chocolate Iced Coffee, 20 oz |
| 96 | 96 | Vanilla Latte w/ Skim Milk, 16 oz |
| 53 | 53 | Mocha Latte w/ Skim Milk, 12 oz |
| 85 | 85 | Original Kreme Latte w/ Skim Milk, 20 oz |
| 93 | 93 | Vanilla Latte w/ 2% Milk, 16 oz |
| 62 | 62 | Iced Caramel Latte w/ 2% Milk, 12 oz |
| 65 | 65 | Iced Caramel Latte w/ Skim Milk, 12 oz |
| 82 | 82 | Original Kreme Latte w/ 2% Milk, 20 oz |
| 155 | 158 | Iced Sugar Free Skinny Caramel Latte w/ Skim M... |
| 152 | 155 | Iced Sugar Free Skinny Caramel Latte w/ 2% Mil... |
| 69 | 69 | Hazelnut Latte w/ 2% Milk, 16 oz |
| 106 | 106 | Vanilla Iced Coffee, 20 oz |
| 72 | 72 | Hazelnut Latte w/ Skim Milk, 16 oz |
| 91 | 91 | Iced Original Kreme Latte w/ Skim Milk, 20 oz |

| | Item_Description | Food_Label \ |
|-----|---|--------------|
| 169 | Frozen Vanilla Latte w/ Whipped Cream, 20 oz, ... | 1 |
| 166 | Lemonade Chiller, also called Frozen Lemonade,... | 1 |
| 197 | Caramel Latte w/ Skim Milk, Whipped Cream & Ca... | 1 |
| 194 | Caramel Latte w/ 2% Milk, Whipped Cream & Cara... | 1 |
| 67 | Iced Caramel Latte w/ Skim Milk, Whipped Cream... | 1 |
| 55 | Mocha Latte w/ Skim Milk, Whipped Cream & Choc... | 1 |
| 52 | Mocha Latte w/ 2% Milk, Whipped Cream & Chocol... | 1 |
| 64 | Iced Caramel Latte w/ 2% Milk, Whipped Cream &... | 1 |
| 61 | Iced Mocha w/ Skim Milk, Whipped Cream & Choco... | 1 |
| 180 | Frozen Caramel Latte w/ Whipped Cream & Caram... | 1 |
| 58 | Iced Mocha w/ 2% Milk, Whipped Cream & Chocola... | 1 |
| 168 | Frozen Vanilla Latte w/ Whipped Cream, 16 oz, ... | 1 |
| 172 | Frozen Mocha w/ Whipped Cream & Chocolate Driz... | 1 |
| 121 | Caramel Mocha w/ 2% Milk, Whipped Cream & Cara... | 1 |
| 118 | Caramel Mocha w/ Skim Milk, Whipped Cream & Ca... | 1 |
| 115 | Iced Caramel Mocha w/ 2% Milk, Whipped Cream &... | 1 |
| 196 | Caramel Latte w/ Skim Milk, Whipped Cream & Ca... | 1 |
| 112 | Iced Caramel Mocha w/ Skim Milk, Whipped Cream... | 1 |
| 193 | Caramel Latte w/ 2% Milk, Whipped Cream & Cara... | 1 |
| 51 | Mocha Latte w/ 2% Milk, Whipped Cream & Chocol... | 1 |
| 66 | Iced Caramel Latte w/ Skim Milk, Whipped Cream... | 1 |
| 54 | Mocha Latte w/ Skim Milk, Whipped Cream & Choc... | 1 |
| 97 | Vanilla Latte w/ Skim Milk, 20 oz, Coffees | 1 |
| 63 | Iced Caramel Latte w/ 2% Milk, Whipped Cream &... | 1 |
| 94 | Vanilla Latte w/ 2% Milk, 20 oz, Coffees | 1 |
| 60 | Iced Mocha w/ Skim Milk, Whipped Cream & Choco... | 1 |

| | | |
|-----|---|---|
| 57 | Iced Mocha w/ 2% Milk, Whipped Cream & Chocola... | 1 |
| 73 | Hazelnut Latte w/ Skim Milk, 20 oz, Coffees | 1 |
| 103 | Iced Vanilla Latte w/ Skim Milk, 20 oz, Coffees | 1 |
| 70 | Hazelnut Latte w/ 2% Milk, 20 oz, Coffees | 1 |
| 100 | Iced Vanilla Latte w/ 2% Milk, 20 oz, Coffees | 1 |
| 179 | Frozen Caramel Latte w/ Whipped Cream & Caram... | 1 |
| 79 | Iced Hazelnut Latte w/ Skim Milk, 20 oz, Coffees | 1 |
| 171 | Frozen Mocha w/ Whipped Cream & Chocolate Driz... | 1 |
| 167 | Frozen Vanilla Latte w/ Whipped Cream, 12 oz, ... | 1 |
| 37 | Hot Chocolate w/ 2% Milk, Whipped Cream & Choc... | 1 |
| 76 | Iced Hazelnut Latte w/ 2% Milk, 20 oz, Coffees | 1 |
| 200 | Jolly Rancher Blue Raspberry Chiller, 20 oz, D... | 1 |
| 117 | Caramel Mocha w/ Skim Milk, Whipped Cream & Ca... | 1 |
| 120 | Caramel Mocha w/ 2% Milk, Whipped Cream & Cara... | 1 |
| 135 | Skinny Sugar Free Caramel Latte w/ Skim Milk, ... | 1 |
| 195 | Caramel Latte w/ Skim Milk, Whipped Cream & Ca... | 1 |
| 192 | Caramel Latte w/ 2% Milk, Whipped Cream & Cara... | 1 |
| 114 | Iced Caramel Mocha w/ 2% Milk, Whipped Cream &... | 1 |
| 134 | Skinny Sugar Free Caramel Latte w/ 2% Milk, 20... | 1 |
| 111 | Iced Caramel Mocha w/ Skim Milk, Whipped Cream... | 1 |
| 158 | Chocolate Iced Coffee w/ Milk & Chocolate Syru... | 1 |
| 96 | Vanilla Latte w/ Skim Milk, 16 oz, Coffees | 1 |
| 53 | Mocha Latte w/ Skim Milk, Whipped Cream & Choc... | 1 |
| 85 | Original Kreme Latte w/ Skim Milk, 20 oz, Coffees | 1 |
| 93 | Vanilla Latte w/ 2% Milk, 16 oz, Coffees | 1 |
| 62 | Iced Caramel Latte w/ 2% Milk, Whipped Cream &... | 1 |
| 65 | Iced Caramel Latte w/ Skim Milk, Whipped Cream... | 1 |
| 82 | Original Kreme Latte w/ 2% Milk, 20 oz, Coffees | 1 |
| 155 | Iced Sugar Free Skinny Caramel Latte w/ Skim M... | 1 |
| 152 | Iced Sugar Free Skinny Caramel Latte w/ 2% Mil... | 1 |
| 69 | Hazelnut Latte w/ 2% Milk, 16 oz, Coffees | 1 |
| 106 | Vanilla Iced Coffee, 20 oz, Coffees | 1 |
| 72 | Hazelnut Latte w/ Skim Milk, 16 oz, Coffees | 1 |
| 91 | Iced Original Kreme Latte w/ Skim Milk, 20 oz,... | 1 |

| | Serving_Size | Total_Fat | Carbohydrates | Protein | 3GM_assign | 3KM_assign | \ |
|-----|--------------|-----------|---------------|---------|------------|------------|---|
| 169 | 567.60 | 11.0 | 90 | 2 | 2 | 2 | |
| 166 | 454.08 | 0.0 | 84 | 0 | 2 | 2 | |
| 197 | 567.60 | 4.5 | 83 | 17 | 1 | 2 | |
| 194 | 567.60 | 11.0 | 82 | 17 | 1 | 2 | |
| 67 | 567.60 | 4.0 | 80 | 15 | 1 | 2 | |
| 55 | 567.60 | 5.0 | 80 | 15 | 1 | 2 | |
| 52 | 567.60 | 12.0 | 79 | 15 | 1 | 2 | |
| 64 | 567.60 | 10.0 | 79 | 15 | 1 | 2 | |
| 61 | 567.60 | 5.0 | 77 | 13 | 1 | 2 | |
| 180 | 567.60 | 12.0 | 76 | 3 | 2 | 2 | |
| 58 | 567.60 | 11.0 | 76 | 13 | 1 | 2 | |
| 168 | 454.08 | 10.0 | 74 | 2 | 2 | 2 | |

| | | | | | | |
|-----|--------|------|----|----|---|---|
| 172 | 567.60 | 12.0 | 73 | 3 | 2 | 2 |
| 121 | 567.60 | 11.0 | 70 | 12 | 1 | 2 |
| 118 | 567.60 | 4.0 | 70 | 13 | 1 | 2 |
| 115 | 567.60 | 10.0 | 67 | 10 | 1 | 2 |
| 196 | 454.08 | 4.0 | 67 | 14 | 1 | 2 |
| 112 | 567.60 | 4.0 | 67 | 11 | 1 | 2 |
| 193 | 454.08 | 10.0 | 66 | 14 | 1 | 2 |
| 51 | 454.08 | 11.0 | 64 | 12 | 1 | 2 |
| 66 | 454.08 | 4.0 | 64 | 12 | 1 | 2 |
| 54 | 454.08 | 5.0 | 64 | 12 | 1 | 2 |
| 97 | 567.60 | 0.0 | 63 | 12 | 2 | 2 |
| 63 | 454.08 | 9.0 | 63 | 12 | 1 | 2 |
| 94 | 567.60 | 7.0 | 62 | 12 | 1 | 2 |
| 60 | 454.08 | 5.0 | 61 | 10 | 2 | 2 |
| 57 | 454.08 | 10.0 | 61 | 10 | 2 | 2 |
| 73 | 567.60 | 0.0 | 61 | 12 | 2 | 2 |
| 103 | 567.60 | 0.0 | 60 | 10 | 2 | 2 |
| 70 | 567.60 | 7.0 | 60 | 12 | 1 | 2 |
| 100 | 567.60 | 6.0 | 59 | 10 | 2 | 2 |
| 179 | 454.08 | 11.0 | 59 | 2 | 2 | 2 |
| 79 | 567.60 | 0.0 | 58 | 10 | 2 | 2 |
| 171 | 454.08 | 11.0 | 58 | 3 | 2 | 2 |
| 167 | 340.56 | 9.0 | 58 | 1 | 2 | 2 |
| 37 | 340.56 | 14.0 | 57 | 12 | 1 | 2 |
| 76 | 567.60 | 6.0 | 57 | 10 | 2 | 2 |
| 200 | 567.60 | 0.0 | 56 | 0 | 1 | 0 |
| 117 | 454.08 | 4.0 | 55 | 11 | 2 | 2 |
| 120 | 454.08 | 10.0 | 55 | 10 | 2 | 2 |
| 135 | 567.60 | 4.0 | 53 | 12 | 2 | 2 |
| 195 | 340.56 | 4.0 | 53 | 11 | 2 | 2 |
| 192 | 340.56 | 9.0 | 53 | 11 | 2 | 2 |
| 114 | 454.08 | 9.0 | 52 | 8 | 2 | 2 |
| 134 | 567.60 | 11.0 | 52 | 12 | 1 | 2 |
| 111 | 454.08 | 4.0 | 52 | 8 | 2 | 2 |
| 158 | 567.60 | 3.0 | 51 | 6 | 2 | 2 |
| 96 | 454.08 | 0.0 | 51 | 10 | 2 | 2 |
| 53 | 340.56 | 4.5 | 51 | 10 | 2 | 2 |
| 85 | 567.60 | 0.0 | 51 | 12 | 2 | 2 |
| 93 | 454.08 | 6.0 | 50 | 10 | 2 | 2 |
| 62 | 340.56 | 7.0 | 50 | 9 | 2 | 2 |
| 65 | 340.56 | 4.0 | 50 | 9 | 2 | 2 |
| 82 | 567.60 | 7.0 | 50 | 12 | 2 | 2 |
| 155 | 567.60 | 4.0 | 50 | 10 | 2 | 2 |
| 152 | 567.60 | 10.0 | 49 | 10 | 2 | 2 |
| 69 | 454.08 | 6.0 | 49 | 10 | 2 | 2 |
| 106 | 567.60 | 2.0 | 49 | 4 | 1 | 0 |
| 72 | 454.08 | 0.0 | 49 | 10 | 2 | 2 |
| 91 | 567.60 | 0.0 | 48 | 10 | 2 | 2 |

| | 3HAC_assign |
|-----|-------------|
| 169 | 0 |
| 166 | 0 |
| 197 | 0 |
| 194 | 0 |
| 67 | 0 |
| 55 | 0 |
| 52 | 0 |
| 64 | 0 |
| 61 | 0 |
| 180 | 0 |
| 58 | 0 |
| 168 | 0 |
| 172 | 0 |
| 121 | 0 |
| 118 | 0 |
| 115 | 0 |
| 196 | 0 |
| 112 | 0 |
| 193 | 0 |
| 51 | 0 |
| 66 | 0 |
| 54 | 0 |
| 97 | 0 |
| 63 | 0 |
| 94 | 0 |
| 60 | 0 |
| 57 | 0 |
| 73 | 0 |
| 103 | 0 |
| 70 | 0 |
| 100 | 0 |
| 179 | 0 |
| 79 | 0 |
| 171 | 0 |
| 167 | 0 |
| 37 | 0 |
| 76 | 0 |
| 200 | 2 |
| 117 | 0 |
| 120 | 0 |
| 135 | 0 |
| 195 | 0 |
| 192 | 0 |
| 114 | 0 |
| 134 | 0 |
| 111 | 0 |

| | |
|-----|---|
| 158 | 0 |
| 96 | 0 |
| 53 | 0 |
| 85 | 0 |
| 93 | 0 |
| 62 | 0 |
| 65 | 0 |
| 82 | 0 |
| 155 | 0 |
| 152 | 0 |
| 69 | 0 |
| 106 | 2 |
| 72 | 0 |
| 91 | 0 |

```
[56]: donuts = donuts.sort_values(by=['Total_Fat'], ascending=[False])
print(donuts[donuts['Food_Label'] == 0].describe())
donuts[donuts['Food_Label'] == 0]
```

| | index | Food_Label | Serving_Size | Total_Fat | Carbohydrates | \ |
|-------|------------|------------|--------------|-----------|---------------|---|
| count | 55.000000 | 55.0 | 55.000000 | 55.000000 | 55.000000 | |
| mean | 68.781818 | 0.0 | 72.181818 | 14.418182 | 34.709091 | |
| std | 74.488302 | 0.0 | 14.587020 | 4.289051 | 6.935344 | |
| min | 0.000000 | 0.0 | 45.000000 | 7.000000 | 21.000000 | |
| 25% | 13.500000 | 0.0 | 62.000000 | 11.000000 | 29.000000 | |
| 50% | 27.000000 | 0.0 | 71.000000 | 14.000000 | 36.000000 | |
| 75% | 165.500000 | 0.0 | 84.500000 | 18.000000 | 40.000000 | |
| max | 189.000000 | 0.0 | 100.000000 | 24.000000 | 50.000000 | |

| | Protein | 3GM_assign | 3KM_assign | 3HAC_assign |
|-------|-----------|------------|------------|-------------|
| count | 55.000000 | 55.0 | 55.0 | 55.0 |
| mean | 3.600000 | 0.0 | 1.0 | 1.0 |
| std | 0.973729 | 0.0 | 0.0 | 0.0 |
| min | 2.000000 | 0.0 | 1.0 | 1.0 |
| 25% | 3.000000 | 0.0 | 1.0 | 1.0 |
| 50% | 4.000000 | 0.0 | 1.0 | 1.0 |
| 75% | 4.000000 | 0.0 | 1.0 | 1.0 |
| max | 7.000000 | 0.0 | 1.0 | 1.0 |

| | index | Item_Name | \ |
|-----|-------|---|---|
| 182 | 186 | Chocolate Hazelnut Kreme Filled Doughnut | |
| 18 | 18 | Glazed Blueberry Cake | |
| 31 | 31 | Powdered Blueberry Filled | |
| 16 | 16 | Double Chocolate Eclair Doughnut | |
| 183 | 187 | Powdered w/ Chocolate Kreme Filling Doughnut | |
| 176 | 180 | Powdered w/ Strawberry Kreme Filling Doughnut | |
| 22 | 22 | Glazed w/ Kreme Filling Doughnut | |

| | | |
|-----|-----|---|
| 6 | 6 | Chocolate Iced w/ Kreme Filling |
| 173 | 177 | Double Dark Chocolate Doughnut |
| 0 | 0 | Apple Fritter |
| 161 | 165 | Fun Faces Collection, Silly Smile |
| 35 | 35 | Tennis Ball Doughnut |
| 162 | 166 | Fun Faces Collection, Smirk |
| 163 | 167 | Fun Faces Collection, Surprise |
| 165 | 169 | Fun Faces Collection, Winking Face |
| 160 | 164 | Fun Faces Collection, Cheeky, |
| 185 | 189 | New York Cheesecake |
| 181 | 185 | Nutty Cocoa Ring |
| 175 | 179 | Powdered Lemon Kreme Filled Doughnut |
| 17 | 17 | Strawberry Shortcake Eclair Doughnut |
| 15 | 15 | Dulce de Leche Filled Doughnut |
| 5 | 5 | Football Doughnut |
| 174 | 178 | Powdered Cinnamon Cake Doughnut |
| 8 | 8 | Chocolate Iced Raspberry Filled Doughnut |
| 10 | 10 | Glazed Cinnamon Bun Doughnut |
| 2 | 2 | Chocolate Iced Custard Filled Doughnut |
| 23 | 23 | Glazed Lemon Filled Doughnut |
| 33 | 33 | Powdered Strawberry Filled Doughnut |
| 9 | 9 | Cinnamon Apple Filled Doughnut |
| 24 | 24 | Glazed Raspberry Filled Doughnut |
| 1 | 1 | Chocolate Iced Cake Doughnut |
| 29 | 29 | Mini Original Glazed Doughnut |
| 184 | 188 | Strawberry Iced Glazed Doughnut w/ Sprinkles |
| 36 | 36 | Traditional Cake Doughnut |
| 32 | 32 | Powdered Cake Doughnut |
| 122 | 125 | Original Glazed Doughnut Holes, 24 Box |
| 11 | 11 | Cinnamon Twist Doughnut |
| 7 | 7 | Chocolate Iced Glazed Doughnut w/ Rainbow Spr... |
| 30 | 30 | Original Glazed Doughnut |
| 20 | 20 | Glazed Cinnamon Doughnut |
| 3 | 3 | Chocolate Iced Glazed Doughnut |
| 164 | 168 | White Iced Glazed Doughnut w/ Chocolate Drizzle |
| 25 | 25 | Glazed Sour Cream Cake Doughnut |
| 26 | 26 | Maple Iced Glazed Doughnut |
| 177 | 181 | Strawberry Iced Glazed Doughnut |
| 21 | 21 | Glazed Cruller Cake Doughnut |
| 34 | 34 | Sugar Doughnut |
| 19 | 19 | Glazed Chocolate Cake Doughnut |
| 4 | 4 | Chocolate Iced Glazed Cruller Doughnut |
| 159 | 163 | Cinnamon Sugar Doughnut |
| 27 | 27 | Mini Chocolate Iced Glazed Doughnut |
| 28 | 28 | Mini Chocolate Iced Glazed Doughnut w/ Rainbow... |
| 13 | 13 | Glazed Traditional Cake Doughnut Holes, 10 Count |
| 12 | 12 | Glazed Blueberry Cake Doughnut Holes, 24 Count... |
| 14 | 14 | Glazed Chocolate Cake Doughnut Holes, 10 Count |

| | Item_Description | Food_Label \ |
|-----|---|--------------|
| 182 | Chocolate Hazelnut Kreme Filled Doughnut w/ Nu... | 0 |
| 18 | Glazed Blueberry Cake, Doughnuts | 0 |
| 31 | Powdered Blueberry Filled, Doughnuts | 0 |
| 16 | Double Chocolate Eclair Doughnut w/ Chocolate ... | 0 |
| 183 | Powdered w/ Chocolate Kreme Filling Doughnut, ... | 0 |
| 176 | Powdered w/ Strawberry Kreme Filling Doughnut,... | 0 |
| 22 | Glazed w/ Kreme Filling Doughnut, Doughnuts | 0 |
| 6 | Chocolate Iced w/ Kreme Filling, Doughnuts | 0 |
| 173 | Double Dark Chocolate Doughnut, Doughnuts | 0 |
| 0 | Apple Fritter, Doughnuts | 0 |
| 161 | Fun Faces Collection, Silly Smile w/ Kreme Fil... | 0 |
| 35 | Tennis Ball Doughnut, Doughnuts | 0 |
| 162 | Fun Faces Collection, Smirk w/ Kreme Filling, ... | 0 |
| 163 | Fun Faces Collection, Surprise w/ Kreme Fillin... | 0 |
| 165 | Fun Faces Collection, Winking Face w/ Kreme Fi... | 0 |
| 160 | Fun Faces Collection, Cheeky w/ Kreme Filling,... | 0 |
| 185 | New York Cheesecake w/ Cheesecake Filling, Cre... | 0 |
| 181 | Nutty Cocoa Ring w/ Nutella Hazelnut Spread, C... | 0 |
| 175 | Powdered Lemon Kreme Filled Doughnut, Doughnuts | 0 |
| 17 | Strawberry Shortcake Eclair Doughnut w/ Strawb... | 0 |
| 15 | Dulce de Leche Filled Doughnut Coated w/ Granu... | 0 |
| 5 | Football Doughnut, Doughnuts | 0 |
| 174 | Powdered Cinnamon Cake Doughnut, Doughnuts | 0 |
| 8 | Chocolate Iced Raspberry Filled Doughnut w/ Tw... | 0 |
| 10 | Glazed Cinnamon Bun Doughnut, Doughnuts | 0 |
| 2 | Chocolate Iced Custard Filled Doughnut, Doughnuts | 0 |
| 23 | Glazed Lemon Filled Doughnut, Doughnuts | 0 |
| 33 | Powdered Strawberry Filled Doughnut, Doughnuts | 0 |
| 9 | Cinnamon Apple Filled Doughnut, Doughnuts | 0 |
| 24 | Glazed Raspberry Filled Doughnut, Doughnuts | 0 |
| 1 | Chocolate Iced Cake Doughnut, Doughnuts | 0 |
| 29 | Mini Original Glazed Doughnut, Doughnuts | 0 |
| 184 | Strawberry Iced Glazed Doughnut w/ Sprinkles, ... | 0 |
| 36 | Traditional Cake Doughnut, Doughnuts | 0 |
| 32 | Powdered Cake Doughnut, Doughnuts | 0 |
| 122 | Original Glazed Doughnut Holes, 24 Box, Doughn... | 0 |
| 11 | Cinnamon Twist Doughnut, Doughnuts | 0 |
| 7 | Chocolate Iced Glazed Doughnut w/ Rainbow Spri... | 0 |
| 30 | Original Glazed Doughnut, Doughnuts | 0 |
| 20 | Glazed Cinnamon Doughnut w/ Powdered Cinnamon ... | 0 |
| 3 | Chocolate Iced Glazed Doughnut, Doughnuts | 0 |
| 164 | White Iced Glazed Doughnut w/ Chocolate Drizzl... | 0 |
| 25 | Glazed Sour Cream Cake Doughnut, Doughnuts | 0 |
| 26 | Maple Iced Glazed Doughnut, Doughnuts | 0 |
| 177 | Strawberry Iced Glazed Doughnut, Doughnuts | 0 |
| 21 | Glazed Cruller Cake Doughnut, Doughnuts | 0 |

| | | |
|-----|---|---|
| 34 | Sugar Doughnut, Doughnuts | 0 |
| 19 | Glazed Chocolate Cake Doughnut, Doughnuts | 0 |
| 4 | Chocolate Iced Glazed Cruller Doughnut, Doughnuts | 0 |
| 159 | Cinnamon Sugar Doughnut, Doughnuts | 0 |
| 27 | Mini Chocolate Iced Glazed Doughnut, Doughnuts | 0 |
| 28 | Mini Chocolate Iced Glazed Doughnut w/ Rainbow... | 0 |
| 13 | Glazed Traditional Cake Doughnut Holes, 10 Cou... | 0 |
| 12 | Glazed Blueberry Cake Doughnut Holes, 24 Count... | 0 |
| 14 | Glazed Chocolate Cake Doughnut Holes, 10 Count... | 0 |

| | Serving_Size | Total_Fat | Carbohydrates | Protein | 3GM_assign | 3KM_assign | \ |
|-----|--------------|-----------|---------------|---------|------------|------------|---|
| 182 | 89.0 | 24.0 | 40 | 5 | 0 | 1 | |
| 18 | 80.0 | 24.0 | 39 | 2 | 0 | 1 | |
| 31 | 78.0 | 22.0 | 30 | 4 | 0 | 1 | |
| 16 | 100.0 | 21.0 | 50 | 5 | 0 | 1 | |
| 183 | 75.0 | 20.0 | 37 | 5 | 0 | 1 | |
| 176 | 75.0 | 20.0 | 37 | 5 | 0 | 1 | |
| 22 | 84.0 | 19.0 | 40 | 4 | 0 | 1 | |
| 6 | 85.0 | 19.0 | 41 | 4 | 0 | 1 | |
| 173 | 93.0 | 19.0 | 45 | 5 | 0 | 1 | |
| 0 | 100.0 | 19.0 | 42 | 4 | 0 | 1 | |
| 161 | 88.0 | 18.0 | 43 | 4 | 0 | 1 | |
| 35 | 71.0 | 18.0 | 38 | 5 | 0 | 1 | |
| 162 | 87.0 | 18.0 | 42 | 4 | 0 | 1 | |
| 163 | 88.0 | 18.0 | 43 | 4 | 0 | 1 | |
| 165 | 87.0 | 18.0 | 42 | 4 | 0 | 1 | |
| 160 | 88.0 | 18.0 | 44 | 4 | 0 | 1 | |
| 185 | 93.0 | 17.0 | 35 | 5 | 0 | 1 | |
| 181 | 60.0 | 17.0 | 26 | 4 | 0 | 1 | |
| 175 | 76.0 | 17.0 | 32 | 4 | 0 | 1 | |
| 17 | 100.0 | 17.0 | 50 | 4 | 0 | 1 | |
| 15 | 76.0 | 16.0 | 35 | 4 | 0 | 1 | |
| 5 | 71.0 | 16.0 | 35 | 7 | 0 | 1 | |
| 174 | 61.0 | 16.0 | 24 | 3 | 0 | 1 | |
| 8 | 88.0 | 15.0 | 41 | 4 | 0 | 1 | |
| 10 | 66.0 | 15.0 | 29 | 4 | 0 | 1 | |
| 2 | 85.0 | 15.0 | 37 | 4 | 0 | 1 | |
| 23 | 84.0 | 14.0 | 36 | 4 | 0 | 1 | |
| 33 | 76.0 | 14.0 | 30 | 4 | 0 | 1 | |
| 9 | 76.0 | 14.0 | 31 | 4 | 0 | 1 | |
| 24 | 84.0 | 14.0 | 37 | 4 | 0 | 1 | |
| 1 | 71.0 | 13.0 | 37 | 3 | 0 | 1 | |
| 29 | 70.0 | 12.0 | 30 | 4 | 0 | 1 | |
| 184 | 66.0 | 12.0 | 34 | 2 | 0 | 1 | |
| 36 | 57.0 | 12.0 | 26 | 3 | 0 | 1 | |
| 32 | 61.0 | 12.0 | 29 | 3 | 0 | 1 | |
| 122 | 54.0 | 11.0 | 25 | 3 | 0 | 1 | |
| 11 | 54.0 | 11.0 | 26 | 3 | 0 | 1 | |

| | | | | | | |
|-----|------|------|----|---|---|---|
| 7 | 66.0 | 11.0 | 36 | 3 | 0 | 1 |
| 30 | 49.0 | 11.0 | 22 | 3 | 0 | 1 |
| 20 | 56.0 | 11.0 | 29 | 3 | 0 | 1 |
| 3 | 63.0 | 11.0 | 33 | 3 | 0 | 1 |
| 164 | 67.0 | 11.0 | 36 | 2 | 0 | 1 |
| 25 | 71.0 | 11.0 | 40 | 3 | 0 | 1 |
| 26 | 63.0 | 11.0 | 34 | 3 | 0 | 1 |
| 177 | 63.0 | 11.0 | 33 | 3 | 0 | 1 |
| 21 | 55.0 | 10.0 | 29 | 3 | 0 | 1 |
| 34 | 45.0 | 10.0 | 21 | 3 | 0 | 1 |
| 19 | 71.0 | 10.0 | 38 | 3 | 0 | 1 |
| 4 | 70.0 | 10.0 | 40 | 3 | 0 | 1 |
| 159 | 45.0 | 10.0 | 21 | 3 | 0 | 1 |
| 27 | 66.0 | 9.0 | 36 | 3 | 0 | 1 |
| 28 | 70.0 | 9.0 | 40 | 3 | 0 | 1 |
| 13 | 51.0 | 8.0 | 28 | 2 | 0 | 1 |
| 12 | 51.0 | 7.0 | 28 | 2 | 0 | 1 |
| 14 | 51.0 | 7.0 | 27 | 2 | 0 | 1 |

3HAC_assign

| | |
|-----|---|
| 182 | 1 |
| 18 | 1 |
| 31 | 1 |
| 16 | 1 |
| 183 | 1 |
| 176 | 1 |
| 22 | 1 |
| 6 | 1 |
| 173 | 1 |
| 0 | 1 |
| 161 | 1 |
| 35 | 1 |
| 162 | 1 |
| 163 | 1 |
| 165 | 1 |
| 160 | 1 |
| 185 | 1 |
| 181 | 1 |
| 175 | 1 |
| 17 | 1 |
| 15 | 1 |
| 5 | 1 |
| 174 | 1 |
| 8 | 1 |
| 10 | 1 |
| 2 | 1 |
| 23 | 1 |
| 33 | 1 |

| | |
|-----|---|
| 9 | 1 |
| 24 | 1 |
| 1 | 1 |
| 29 | 1 |
| 184 | 1 |
| 36 | 1 |
| 32 | 1 |
| 122 | 1 |
| 11 | 1 |
| 7 | 1 |
| 30 | 1 |
| 20 | 1 |
| 3 | 1 |
| 164 | 1 |
| 25 | 1 |
| 26 | 1 |
| 177 | 1 |
| 21 | 1 |
| 34 | 1 |
| 19 | 1 |
| 4 | 1 |
| 159 | 1 |
| 27 | 1 |
| 28 | 1 |
| 13 | 1 |
| 12 | 1 |
| 14 | 1 |

[]: