MN_mode_Clustering

May 3, 2024

```
[1]: !pip install scikit-learn-extra
```

```
Requirement already satisfied: scikit-learn-extra in
c:\users\tky12\anaconda3\lib\site-packages (0.3.0)
Requirement already satisfied: numpy>=1.13.3 in
c:\users\tky12\anaconda3\lib\site-packages (from scikit-learn-extra) (1.26.4)
Requirement already satisfied: scipy>=0.19.1 in
c:\users\tky12\anaconda3\lib\site-packages (from scikit-learn-extra) (1.11.4)
Requirement already satisfied: scikit-learn>=0.23.0 in
c:\users\tky12\anaconda3\lib\site-packages (from scikit-learn-extra) (1.2.2)
Requirement already satisfied: joblib>=1.1.1 in
c:\users\tky12\anaconda3\lib\site-packages (from scikit-learn>=0.23.0->scikit-
learn-extra) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\tky12\anaconda3\lib\site-packages (from scikit-learn>=0.23.0->scikit-
learn-extra) (2.2.0)
```

```
[2]: import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     import numpy as np
     from mpl_toolkits.mplot3d import Axes3D
     from scipy.stats import multivariate_normal
     from scipy.stats import norm
     import plotly.express as px
     from sklearn.preprocessing import MinMaxScaler
     from sklearn.metrics import silhouette_score
     import pickle
```

```
[3]: # Surpress warnings
     def warn(*args, **kwargs):
         pass
     import warnings
     warnings.warn = warn
```

```
[4]: pd.DataFrame.iteritems = pd.DataFrame.items
```

```
[5]: # Retrieve dataset and read first 5 rows
     macroNutrient_mode = pd.read_csv(r"../Dataset/Dataset_for_EDA/
       ↪macroNutrient_mode.csv", encoding= 'unicode_escape')
     macroNutrient_mode
```

```
[5]:         No.                                       Description        Category  \
     0     15155                          ABALONE,MIXED SPECIES,RAW         ABALONE
     1     15156                          ABALONE,MXD SP,CKD,FRIED         ABALONE
     2      9427                                       ABIYUCH,RAW         ABIYUCH
     3      9002                                 ACEROLA JUICE,RAW   ACEROLA JUICE
     4      9001                 ACEROLA,(WEST INDIAN CHERRY),RAW         ACEROLA
     ...     ...                                               ...             ...
     8785   1119     YOGURT,VANILLA,LOFAT,11 GRAMS PROT PER 8 OZ          YOGURT
     8786   1220  YOGURT,VANILLA,LOFAT,11 GRAMS PROT PER 8 OZ,FO…          YOGURT
     8787   1295                          YOGURT,VANILLA,NON-FAT          YOGURT
     8788  16004               YOKAN,PREP FROM ADZUKI BNS & SUGAR           YOKAN
     8789   3217                                          ZWIEBACK        ZWIEBACK

           Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
     0               105             6.01       17.10            0.76
     1               189            11.05       19.63            6.78
     2                69            17.60        1.50            0.10
     3                23             4.80        0.40            0.30
     4                32             7.69        0.40            0.30
     ...             ...              ...         ...             ...
     8785             85            13.80        4.93            1.25
     8786             85            13.80        4.93            1.25
     8787             78            17.04        2.94            0.00
     8788            260            60.72        3.29            0.12
     8789            426            74.20       10.10            9.70

           Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
     0                              0.107                           0.104
     1                              2.741                           1.676
     2                              0.000                           0.000
     3                              0.082                           0.090
     4                              0.082                           0.090
     ...                              ...                             ...
     8785                           0.343                           0.036
     8786                           0.343                           0.036
     8787                           0.000                           0.000
     8788                           0.011                           0.026
     8789                           4.244                           2.073

           Saturated Fatty Acids(g)
     0                        0.149
     1                        1.646
```

```
2                      0.014
3                      0.068
4                      0.068
…                       …
8785                   0.806
8786                   0.806
8787                   0.000
8788                   0.043
8789                   2.525

[8790 rows x 10 columns]
```

[6]: `macroNutrient_mode.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   No.                             8790 non-null   int64
 1   Description                     8790 non-null   object
 2   Category                        8790 non-null   object
 3   Energy (Kcal)                   8790 non-null   int64
 4   Carbohydrate(g)                 8790 non-null   float64
 5   Protein(g)                      8790 non-null   float64
 6   Total Lipid(g)                  8790 non-null   float64
 7   Monounsaturated Fatty Acids(g)  8790 non-null   float64
 8   Polyunsaturated Fatty Acids(g)  8790 non-null   float64
 9   Saturated Fatty Acids(g)        8790 non-null   float64
dtypes: float64(6), int64(2), object(2)
memory usage: 686.8+ KB
```

[7]: `macroNutrient_mode.describe()`

[7]:
```
                No.  Energy (Kcal)  Carbohydrate(g)   Protein(g)  \
count   8790.000000    8790.000000      8790.000000  8790.000000
mean   15663.495222     226.317634        22.127710    11.342849
std     9251.413586     169.877539        27.270822    10.530474
min     1001.000000       0.000000         0.000000     0.000000
25%     9086.250000      91.000000         0.050000     2.380000
50%    14427.500000     191.000000         9.340000     8.000000
75%    20142.750000     337.000000        34.910000    19.880000
max    93600.000000     902.000000       100.000000    88.320000


       Total Lipid(g)  Monounsaturated Fatty Acids(g)  \
count     8790.000000                     8790.000000
mean        10.553725                        3.771797
std         15.814842                        6.825884
```

```
min         0.000000                              0.000000
25%         0.950000                              0.074250
50%         5.140000                              1.600000
75%        13.720000                              4.759000
max       100.000000                             83.689000


          Polyunsaturated Fatty Acids(g)  Saturated Fatty Acids(g)
count                    8790.000000               8790.000000
mean                        2.115612                  3.434338
std                         5.050857                  6.426162
min                         0.000000                  0.000000
25%                         0.134250                  0.149000
50%                         0.591500                  1.435500
75%                         1.880000                  4.173750
max                        74.623000                 95.600000
```

[8]: `macroNutrient_mode.head()`

[8]:
```
      No.                   Description       Category  Energy (Kcal)  \
0  15155        ABALONE,MIXED SPECIES,RAW        ABALONE           105
1  15156         ABALONE,MXD SP,CKD,FRIED        ABALONE           189
2   9427                     ABIYUCH,RAW        ABIYUCH            69
3   9002               ACEROLA JUICE,RAW  ACEROLA JUICE            23
4   9001  ACEROLA,(WEST INDIAN CHERRY),RAW        ACEROLA            32


   Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
0             6.01       17.10            0.76
1            11.05       19.63            6.78
2            17.60        1.50            0.10
3             4.80        0.40            0.30
4             7.69        0.40            0.30


   Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
0                            0.107                           0.104
1                            2.741                           1.676
2                            0.000                           0.000
3                            0.082                           0.090
4                            0.082                           0.090


   Saturated Fatty Acids(g)
0                     0.149
1                     1.646
2                     0.014
3                     0.068
4                     0.068
```

```
[9]: float_columns = [x for x in macroNutrient_mode.columns if x not in ['No.',
      ↪'Description', 'Category']]
```

```
[10]: skew_columns = (macroNutrient_mode[float_columns]
                      .skew()
                      .sort_values(ascending=False))

      skew_columns = skew_columns.loc[skew_columns > 0.75]
      print("{} of the 7 columns are skewed with the vast majority being heavily
       ↪skewed".format(len(skew_columns)))
      skew_columns
```

7 of the 7 columns are skewed with the vast majority being heavily skewed

```
[10]: Saturated Fatty Acids(g)          6.664663
      Polyunsaturated Fatty Acids(g)    6.357650
      Monounsaturated Fatty Acids(g)    4.687548
      Total Lipid(g)                    3.309724
      Protein(g)                        1.166368
      Energy (Kcal)                     1.148610
      Carbohydrate(g)                   1.127598
      dtype: float64
```

```
[11]: # Perform log transform on skewed columns
      for col in skew_columns.index.tolist():
          macroNutrient_mode[col] = np.log1p(macroNutrient_mode[col])
```

```
[12]: macroNutrient_mode[float_columns]
```

```
[12]:       Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
      0          4.663439         1.947338    2.895912        0.565314
      1          5.247024         2.489065    3.026746        2.051556
      2          4.248495         2.923162    0.916291        0.095310
      3          3.178054         1.757858    0.336472        0.262364
      4          3.496508         2.162173    0.336472        0.262364
      ...             ...              ...         ...             ...
      8785       4.454347         2.694627    1.780024        0.810930
      8786       4.454347         2.694627    1.780024        0.810930
      8787       4.369448         2.892592    1.371181        0.000000
      8788       5.564520         4.122608    1.456287        0.113329
      8789       6.056784         4.320151    2.406945        2.370244

            Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
      0                           0.101654                        0.098940
      1                           1.319353                        0.984323
      2                           0.000000                        0.000000
      3                           0.078811                        0.086178
      4                           0.078811                        0.086178
```

```
...                           ...                               ...
8785                     0.294906                          0.035367
8786                     0.294906                          0.035367
8787                     0.000000                          0.000000
8788                     0.010940                          0.025668
8789                     1.657085                          1.122654

        Saturated Fatty Acids(g)
0                       0.138892
1                       0.973049
2                       0.013903
3                       0.065788
4                       0.065788
...                          ...
8785                    0.591114
8786                    0.591114
8787                    0.000000
8788                    0.042101
8789                    1.259880

[8790 rows x 7 columns]
```

[13]: `macroNutrient_mode.describe()`

```
[13]:                No.   Energy (Kcal)   Carbohydrate(g)    Protein(g)  \
       count   8790.000000     8790.000000       8790.000000   8790.000000
       mean   15663.495222        5.065342          2.137177      2.047259
       std     9251.413586        0.997539          1.615109      1.068937
       min     1001.000000        0.000000          0.000000      0.000000
       25%     9086.250000        4.521789          0.048790      1.217876
       50%    14427.500000        5.257495          2.336020      2.197225
       75%    20142.750000        5.823046          3.581016      3.038792
       max    93600.000000        6.805723          4.615121      4.492225

              Total Lipid(g)   Monounsaturated Fatty Acids(g)  \
       count     8790.000000                      8790.000000
       mean         1.770891                         1.045088
       std          1.183944                         0.944070
       min          0.000000                         0.000000
       25%          0.667829                         0.071623
       50%          1.814823                         0.955511
       75%          2.689207                         1.750764
       max          4.615121                         4.438986

              Polyunsaturated Fatty Acids(g)   Saturated Fatty Acids(g)
       count                     8790.000000                8790.000000
       mean                         0.712797                   1.016352
```

```
std                      0.758623              0.895071
min                      0.000000              0.000000
25%                      0.125972              0.138892
50%                      0.464677              0.890152
75%                      1.057790              1.643598
max                      4.325760              4.570579
```

[14]:
```
scaler = MinMaxScaler()
macroNutrient_mode[float_columns] = scaler.
 ↪fit_transform(macroNutrient_mode[float_columns])

macroNutrient_mode.describe()
```

[14]:
```
                 No.  Energy (Kcal)  Carbohydrate(g)   Protein(g)  \
count   8790.000000    8790.000000      8790.000000  8790.000000
mean   15663.495222       0.744277         0.463082     0.455734
std     9251.413586       0.146574         0.349960     0.237953
min     1001.000000       0.000000         0.000000     0.000000
25%     9086.250000       0.664410         0.010572     0.271107
50%    14427.500000       0.772511         0.506167     0.489117
75%    20142.750000       0.855610         0.775931     0.676456
max    93600.000000       1.000000         1.000000     1.000000

       Total Lipid(g)  Monounsaturated Fatty Acids(g)  \
count     8790.000000                     8790.000000
mean         0.383715                        0.235434
std          0.256536                        0.212677
min          0.000000                        0.000000
25%          0.144705                        0.016135
50%          0.393234                        0.215254
75%          0.582695                        0.394406
max          1.000000                        1.000000

       Polyunsaturated Fatty Acids(g)  Saturated Fatty Acids(g)
count                     8790.000000               8790.000000
mean                         0.164780                  0.222368
std                          0.175373                  0.195833
min                          0.000000                  0.000000
25%                          0.029121                  0.030388
50%                          0.107421                  0.194757
75%                          0.244533                  0.359604
max                          1.000000                  1.000000
```

[15]:
```
X = macroNutrient_mode[float_columns]
X
```

```
[15]:       Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
       0        0.685223         0.421947    0.644650        0.122492
       1        0.770972         0.539328    0.673774        0.444529
       2        0.624253         0.633388    0.203973        0.020652
       3        0.466968         0.380891    0.074901        0.056849
       4        0.513760         0.468498    0.074901        0.056849
       ...           ...              ...         ...             ...
       8785     0.654500         0.583869    0.396246        0.175712
       8786     0.654500         0.583869    0.396246        0.175712
       8787     0.642026         0.626764    0.305234        0.000000
       8788     0.817624         0.893283    0.324179        0.024556
       8789     0.889955         0.936086    0.535802        0.513582

             Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
       0                           0.022900                        0.022872
       1                           0.297219                        0.227549
       2                           0.000000                        0.000000
       3                           0.017754                        0.019922
       4                           0.017754                        0.019922
       ...                             ...                             ...
       8785                        0.066435                        0.008176
       8786                        0.066435                        0.008176
       8787                        0.000000                        0.000000
       8788                        0.002465                        0.005934
       8789                        0.373303                        0.259528

             Saturated Fatty Acids(g)
       0                     0.030388
       1                     0.212894
       2                     0.003042
       3                     0.014394
       4                     0.014394
       ...                        ...
       8785                  0.129330
       8786                  0.129330
       8787                  0.000000
       8788                  0.009211
       8789                  0.275650

       [8790 rows x 7 columns]
```
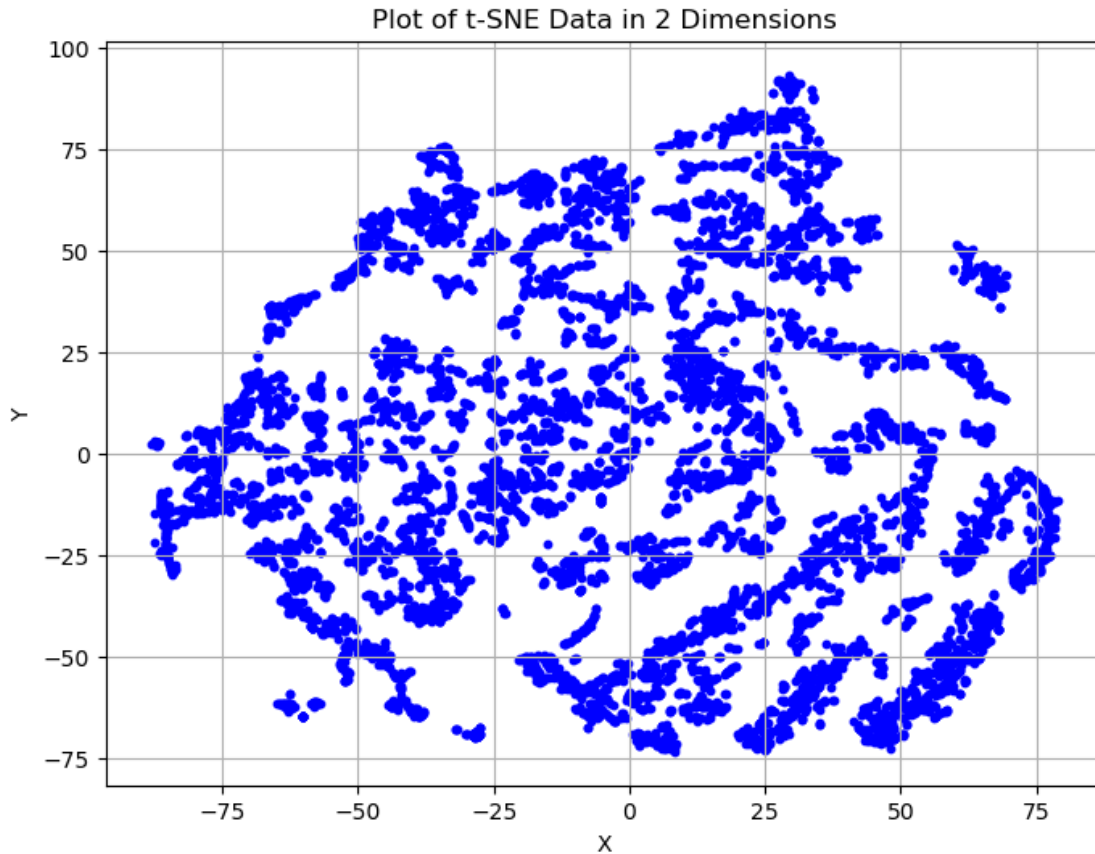
0.1 Dimensionality Reduction

```
[16]: from clustering_function import dimensionality_reduction,scatter_plot_clustering
```

```
[17]: # Reduced to 2 dimensions using tsne
```

```
X_tsne_reduced, tsne_x_data, tsne_y_data, reduction_method =␣
 ↪dimensionality_reduction(X, 't-SNE', n_components=2)
```

[18]:
```
scatter_plot_clustering(tsne_x_data, tsne_y_data, reduction_method)
```

Plot of t-SNE Data in 2 Dimensions

[19]:
```
X_tsne_reduced
```

[19]:
```
array([[-10.804307 , -27.390484 ],
       [ 13.875143 ,  -1.8643672],
       [-58.51878  ,   2.296319 ],
       ...,
       [-52.1724   ,   7.284499 ],
       [-49.443638 ,  45.170883 ],
       [ 10.488729 ,  43.41403  ]], dtype=float32)
```

[20]:
```
pd.DataFrame(X_tsne_reduced).to_csv("../Dataset/Dataset_for_reduced_data/
 ↪mode_reduced.csv")
```

```
[21]: # Reduced to 2 dimensions using pca
      X_pca_reduced, pca_x_data, pca_y_data, reduction_method =␣
      ↪dimensionality_reduction(X, 'PCA', n_components=2)
```

```
[22]: scatter_plot_clustering(pca_x_data, pca_y_data, reduction_method)
```


Plot of PCA Data in 2 Dimensions
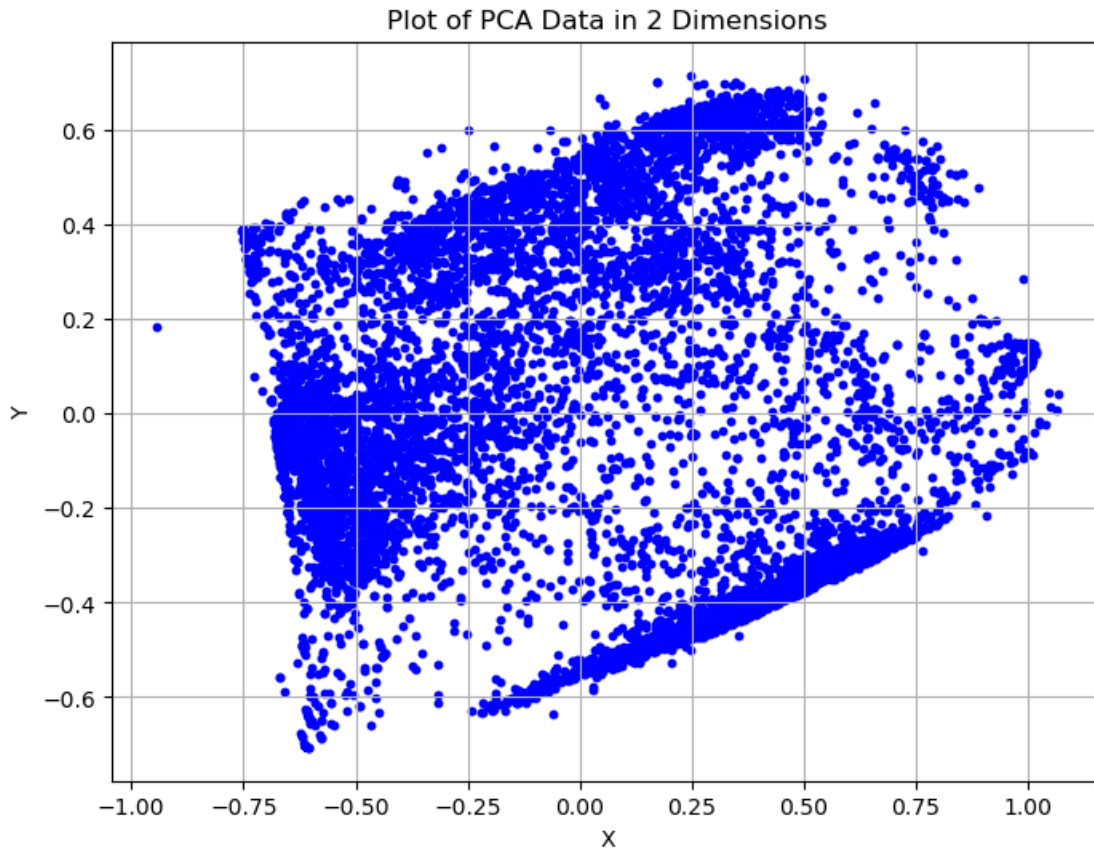
```
[23]: X_pca_reduced
```

```
[23]: array([[-0.28421393, -0.22019718],
             [ 0.13014079,  0.07826331],
             [-0.60331131, -0.0189122 ],
             ...,
             [-0.57450877, -0.03985653],
             [-0.59777671,  0.243115  ],
             [ 0.08221506,  0.51777063]])
```
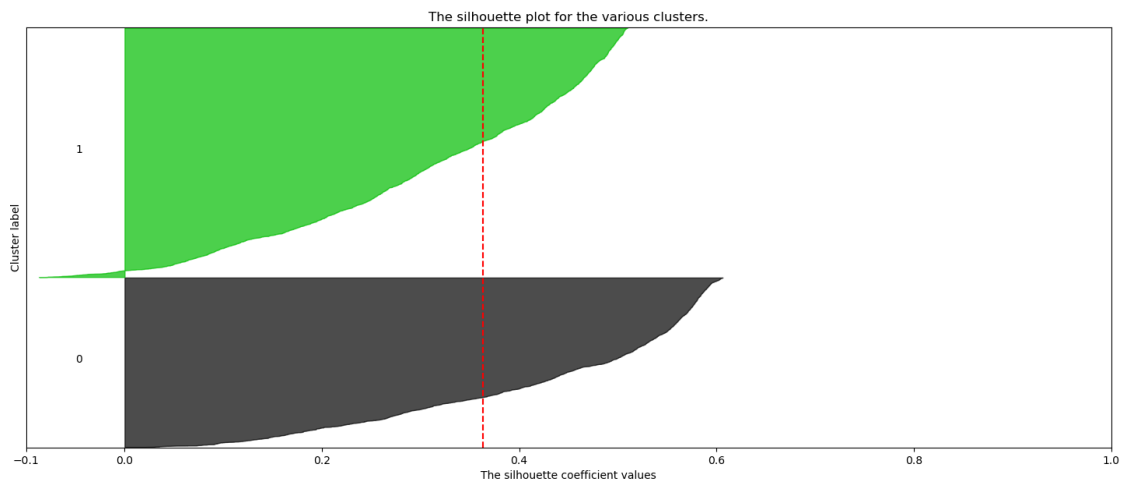
0.2 K Means: Optimal number of clusters

Silhoutte Analysis

```
[24]: from clustering_function import plot_silhouette_analysis
```

```

```
[25]: tsne_list_num_clusters = list(range(2,8))
      plot_silhouette_analysis(X_tsne_reduced, tsne_list_num_clusters)
```

For n_clusters = 2 The average silhouette_score is : 0.36329898
For n_clusters = 3 The average silhouette_score is : 0.43077832
For n_clusters = 4 The average silhouette_score is : 0.39268178
For n_clusters = 5 The average silhouette_score is : 0.3649345
For n_clusters = 6 The average silhouette_score is : 0.40257114
For n_clusters = 7 The average silhouette_score is : 0.39454266



**Silhouette analysis for KMeans clustering on sample data with n_clusters = 2**



**Silhouette analysis for KMeans clustering on sample data with n_clusters = 3**

**Silhouette analysis for KMeans clustering on sample data with n_clusters = 4**

The silhouette plot for the various clusters.



**Silhouette analysis for KMeans clustering on sample data with n_clusters = 5**

The silhouette plot for the various clusters.

**Silhouette analysis for KMeans clustering on sample data with n_clusters = 6**

The silhouette plot for the various clusters.

**Silhouette analysis for KMeans clustering on sample data with n_clusters = 7**

The silhouette plot for the various clusters.

## Elbow Method

```
[26]: from clustering_function import elbow_method_analysis
```

```
[27]: elbow_method_analysis(X_tsne_reduced)
```

The Elbow Method for t-SNE

0.3 Clustering

```
[28]: # Trackers throughout each model
      scores = {} # to track the silhoutte score of the tuned model
      food_groups = {} # to track the counts of each group
```

```
[29]: X_with_labels = X.copy()
      X_with_labels
```

```
[29]:        Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
      0           0.685223         0.421947    0.644650        0.122492
      1           0.770972         0.539328    0.673774        0.444529
      2           0.624253         0.633388    0.203973        0.020652
      3           0.466968         0.380891    0.074901        0.056849
      4           0.513760         0.468498    0.074901        0.056849
      ...              ...              ...         ...             ...
      8785        0.654500         0.583869    0.396246        0.175712
      8786        0.654500         0.583869    0.396246        0.175712
      8787        0.642026         0.626764    0.305234        0.000000
      8788        0.817624         0.893283    0.324179        0.024556
```

14

```
8789          0.889955          0.936086      0.535802          0.513582
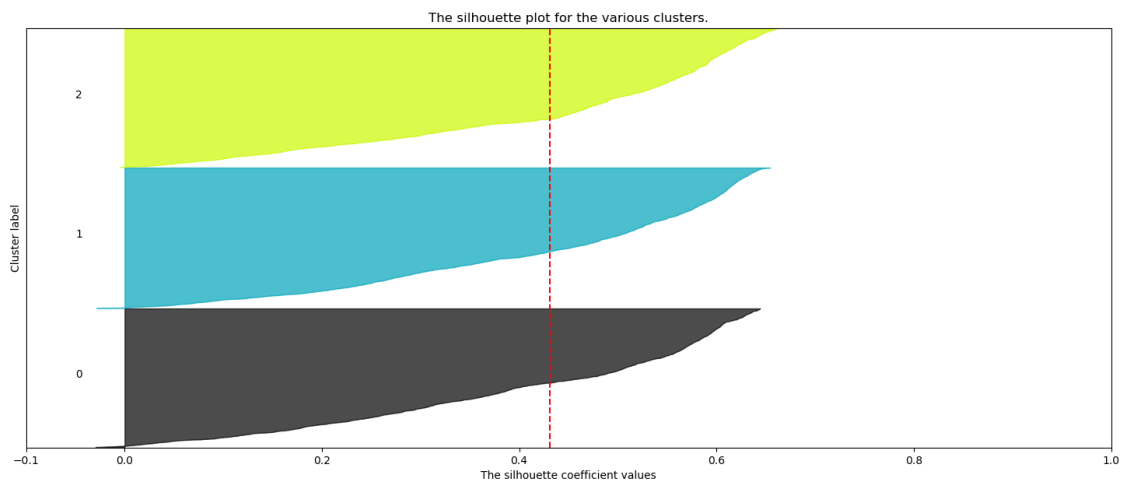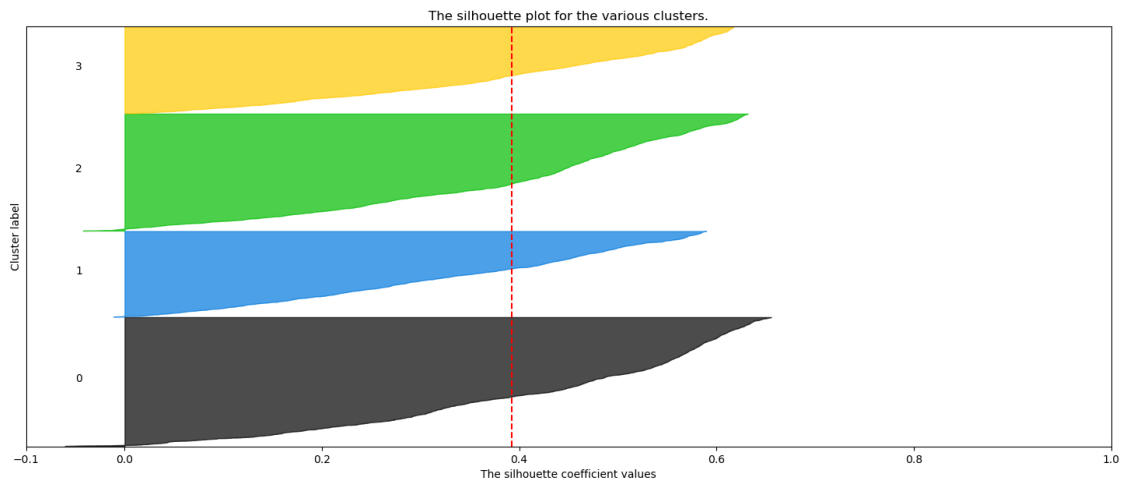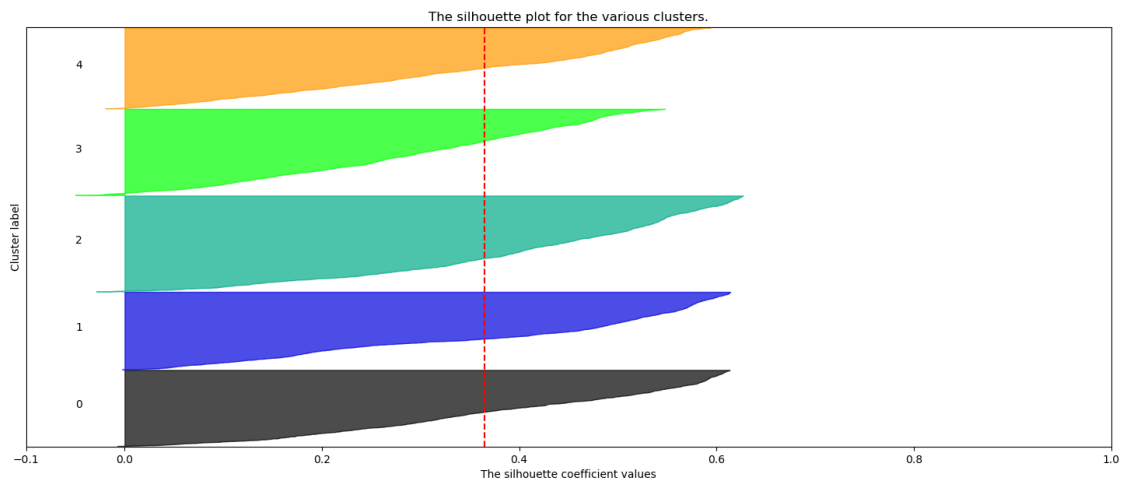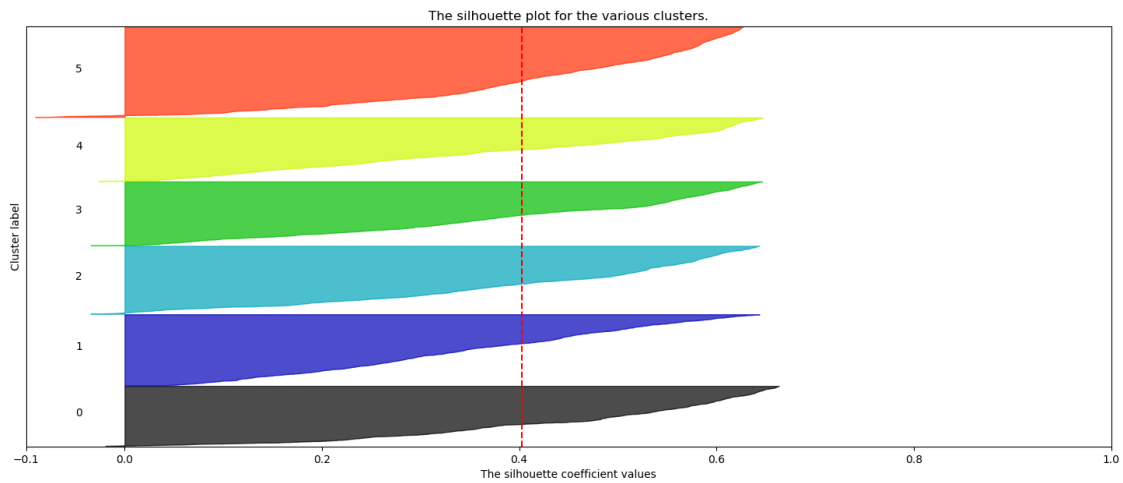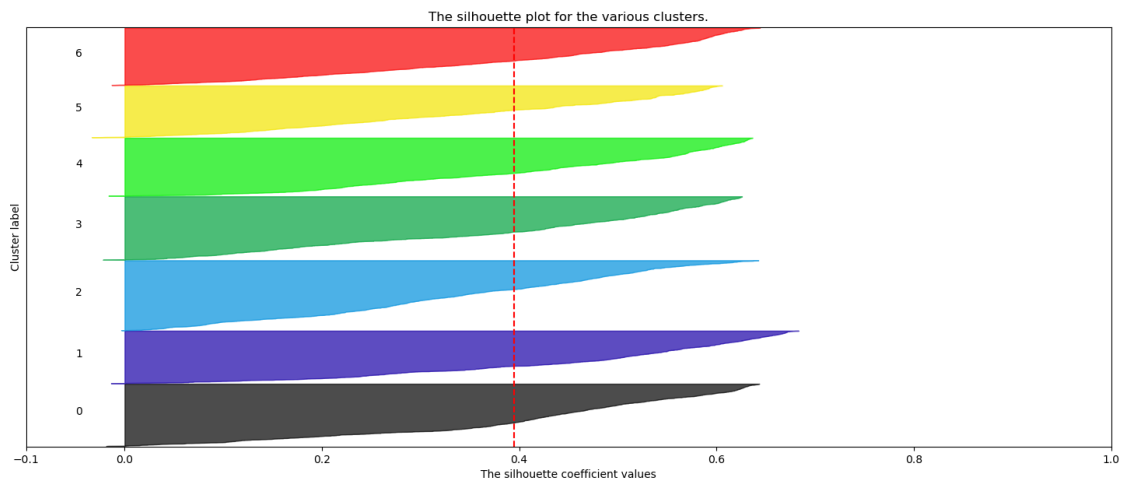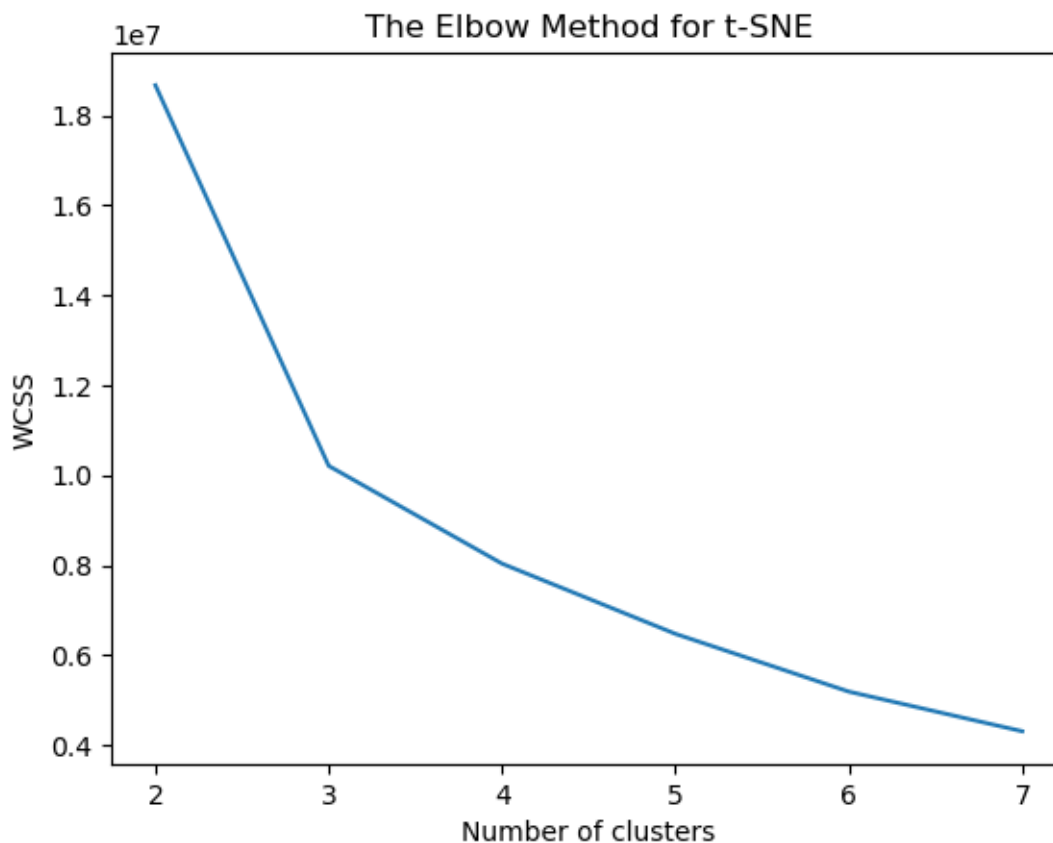
      Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
0                           0.022900                        0.022872
1                           0.297219                        0.227549
2                           0.000000                        0.000000
3                           0.017754                        0.019922
4                           0.017754                        0.019922
...                              ...                             ...
8785                        0.066435                        0.008176
8786                        0.066435                        0.008176
8787                        0.000000                        0.000000
8788                        0.002465                        0.005934
8789                        0.373303                        0.259528

      Saturated Fatty Acids(g)
0                     0.030388
1                     0.212894
2                     0.003042
3                     0.014394
4                     0.014394
...                        ...
8785                  0.129330
8786                  0.129330
8787                  0.000000
8788                  0.009211
8789                  0.275650

[8790 rows x 7 columns]
```

0.4 K Means Clustering

```python
[30]: from clustering_function import perform_kmeans_clustering, plot_clusters
```

```python
[31]: tsne_clust_labels, kmean_model = perform_kmeans_clustering(X_tsne_reduced, 3)
      plot_clusters(X_tsne_reduced, tsne_clust_labels, 'K-Means')
```

t-SNE with K-Means
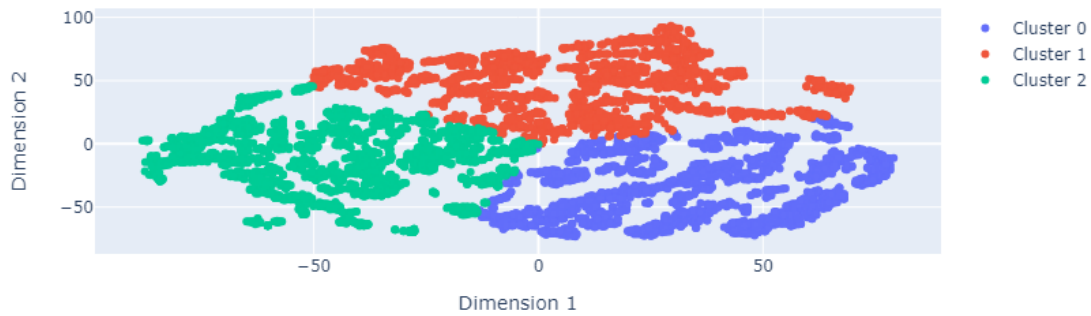

```
[32]: # export model
      with open('../Model_fitted/Kmeans_model_mode_pkl', 'wb') as files:
          pickle.dump(kmean_model, files)
```

```
[33]: X_with_labels['kmeans_tsne'] = tsne_clust_labels
      scores['kmeans_tsne'] = (silhouette_score(X_tsne_reduced,␣
       ↪X_with_labels['kmeans_tsne']))


      X_with_labels
```

```
[33]:        Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
      0           0.685223         0.421947    0.644650        0.122492
      1           0.770972         0.539328    0.673774        0.444529
      2           0.624253         0.633388    0.203973        0.020652
      3           0.466968         0.380891    0.074901        0.056849
      4           0.513760         0.468498    0.074901        0.056849
      ...              ...              ...         ...             ...
      8785        0.654500         0.583869    0.396246        0.175712
      8786        0.654500         0.583869    0.396246        0.175712
      8787        0.642026         0.626764    0.305234        0.000000
      8788        0.817624         0.893283    0.324179        0.024556
      8789        0.889955         0.936086    0.535802        0.513582


             Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
      0                            0.022900                        0.022872
      1                            0.297219                        0.227549
      2                            0.000000                        0.000000
      3                            0.017754                        0.019922
      4                            0.017754                        0.019922
      ...                               ...                             ...
```

```
8785                        0.066435                        0.008176
8786                        0.066435                        0.008176
8787                        0.000000                        0.000000
8788                        0.002465                        0.005934
8789                        0.373303                        0.259528

        Saturated Fatty Acids(g)   kmeans_tsne
0                       0.030388             2
1                       0.212894             0
2                       0.003042             2
3                       0.014394             2
4                       0.014394             2
…                            …             …
8785                    0.129330             2
8786                    0.129330             2
8787                    0.000000             2
8788                    0.009211             1
8789                    0.275650             1

[8790 rows x 8 columns]
```

[34]: ```python
from clustering_function import get_food_groups, plot_cluster_distribution
```

[35]: ```python
food_groups, value_counts = get_food_groups(X_with_labels['kmeans_tsne'],
    'KMeans_tsne',food_groups)
```

[36]: ```python
plot_cluster_distribution(value_counts, 'K-Means')
```

K-Means using t-SNE classes distribution

[37]: ```python
kmeans_tsne = X_with_labels.groupby('kmeans_tsne')
```
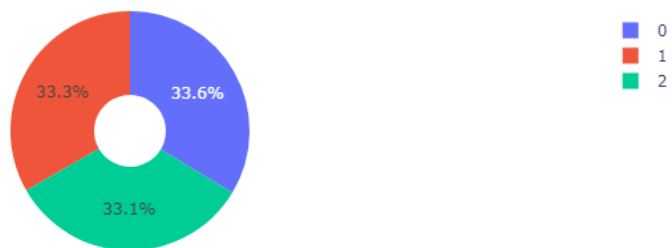
```
[38]: n = kmeans_tsne['kmeans_tsne'].count().count() # number of cluster
      for i in range(0,n):
          display(kmeans_tsne.get_group(i))
```

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
1	0.770972	0.539328	0.673774	0.444529	
14	0.816493	0.139076	0.695060	0.644735	
100	0.786376	0.000000	0.756312	0.513785	
101	0.717455	0.000000	0.681411	0.382380	
103	0.999184	0.000000	0.000000	0.998925	
...	...	...	...	...	
8731	0.776288	0.412147	0.653097	0.543778	
8749	0.769417	0.000000	0.762047	0.442852	
8750	0.733270	0.000000	0.708751	0.396735	
8771	0.632419	0.346116	0.532774	0.232190	
8773	0.673693	0.347863	0.512571	0.388237	

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g)	\
1	0.297219	0.227549	
14	0.511857	0.352783	
100	0.351960	0.293664	
101	0.175770	0.224155	
103	0.861886	0.577201	
...	...	...	
8731	0.282218	0.443141	
8749	0.000000	0.000000	
8750	0.246740	0.204303	
8771	0.089229	0.016934	
8773	0.257480	0.088905	

	Saturated Fatty Acids(g)	kmeans_tsne
1	0.212894	0
14	0.395637	0
100	0.254692	0
101	0.180514	0
103	0.807114	0
...	...	...
8731	0.209057	0
8749	0.000000	0
8750	0.180322	0
8771	0.175470	0
8773	0.267429	0

[2956 rows x 8 columns]

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
5	0.913731	0.870851	0.476131	0.745258	
6	0.670663	0.503637	0.457547	0.324453	

```
7                0.916054          0.866962       0.491577            0.753711
8                0.875881          0.808581       0.437893            0.696246
12               0.904365          0.529205       0.512571            0.822403
...                   ...               ...            ...                 ...
8733             0.822060          0.189696       0.614397            0.684976
8745             0.859894          0.897421       0.719463            0.181414
8748             0.858196          0.937465       0.463454            0.219193
8788             0.817624          0.893283       0.324179            0.024556
8789             0.889955          0.936086       0.535802            0.513582

      Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
5                           0.676104                        0.443583
6                           0.222082                        0.060118
7                           0.684742                        0.451553
8                           0.626129                        0.398092
12                          0.632893                        0.671998
...                              ...                             ...
8733                        0.000000                        0.000000
8745                        0.024320                        0.103539
8748                        0.080576                        0.140080
8788                        0.002465                        0.005934
8789                        0.373303                        0.259528

      Saturated Fatty Acids(g)  kmeans_tsne
5                     0.348734            1
6                     0.180322            1
7                     0.355775            1
8                     0.308817            1
12                    0.494853            1
...                        ...          ...
8733                  0.478288            1
8745                  0.063870            1
8748                  0.000000            1
8788                  0.009211            1
8789                  0.275650            1

[2926 rows x 8 columns]
      Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
0          0.685223         0.421947    0.644650        0.122492
2          0.624253         0.633388    0.203973        0.020652
3          0.466968         0.380891    0.074901        0.056849
4          0.513760         0.468498    0.074901        0.056849
9          0.721842         0.757620    0.153183        0.055176
...             ...              ...         ...             ...
8783       0.556031         0.463708    0.351950        0.035864
8784       0.656199         0.583869    0.396246        0.175712
8785       0.654500         0.583869    0.396246        0.175712
```

```
8786        0.654500       0.583869   0.396246          0.175712
8787        0.642026       0.626764   0.305234          0.000000


      Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
0                           0.022900                        0.022872
2                           0.000000                        0.000000
3                           0.017754                        0.019922
4                           0.017754                        0.019922
9                           0.000000                        0.000000
...                              ...                             ...
8783                        0.010777                        0.001153
8784                        0.061346                        0.007953
8785                        0.066435                        0.008176
8786                        0.066435                        0.008176
8787                        0.000000                        0.000000


      Saturated Fatty Acids(g)  kmeans_tsne
0                     0.030388            2
2                     0.003042            2
3                     0.014394            2
4                     0.014394            2
9                     0.000000            2
...                        ...          ...
8783                  0.024012            2
8784                  0.122689            2
8785                  0.129330            2
8786                  0.129330            2
8787                  0.000000            2

[2908 rows x 8 columns]
```

0.4.1 Insights

- In Cluster 0, on average is Slightly High in Energy and Protein, with a moderate level of Total Lipid.
- In Cluster 1, on average is High in Energy and Carbohydrate, with a moderate level of protein and lipid.
- In Cluster 2, on average is Slightly High in Energy with a moderate level of Carbohydrate.

0.5 Agglomerative Clustering

```
[39]: from clustering_function import tuning_agglomerative, perform_agg_clustering
```

```
[40]: agg_param_grid = {
          'n_clusters': [2, 3, 4, 5, 6],
          'linkage': ['ward'],
          'affinity' : ['euclidean']
      }
```

```
[41]: tuning_agglomerative(X_tsne_reduced, agg_param_grid)
```

Best silhouette score: 0.4223424
Best parameters: {'affinity': 'euclidean', 'linkage': 'ward', 'n_clusters': 3}

```
[42]: Agg_tsne_clust_labels,agg_model = perform_agg_clustering(X_tsne_reduced, 3)
```

```
[43]: # export model
      with open('../Model_fitted/Agg_model_mode_pkl', 'wb') as files:
          pickle.dump(agg_model, files)
```

```
[44]: plot_clusters(X_tsne_reduced, Agg_tsne_clust_labels, "Agglomerative Clustering")
```
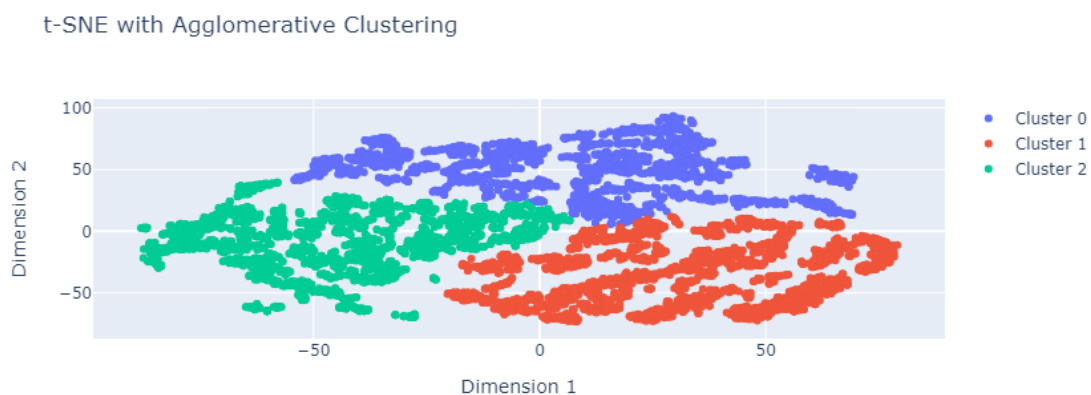


```
[45]: X_with_labels['Agg_tsne'] = Agg_tsne_clust_labels
      scores['Agg_tsne'] = (silhouette_score(X_tsne_reduced,␣
       ↪X_with_labels['Agg_tsne']))
```

```
[46]: food_groups, value_counts = get_food_groups(X_with_labels['Agg_tsne'],␣
       ↪'Agglomerative_tsne',food_groups)
```

```
[47]: plot_cluster_distribution(value_counts, 'Agglomerative Clustering')
```

Agglomerative Clustering using t-SNE classes distribution


```
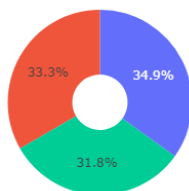[48]: Aggtsne = X_with_labels.groupby('Agg_tsne')
```

```
[49]: n = Aggtsne['Agg_tsne'].count().count() # number of cluster
      for i in range(0,n):
          display(Aggtsne.get_group(i))
```

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
5	0.913731	0.870851	0.476131	0.745258	
7	0.916054	0.866962	0.491577	0.753711	
8	0.875881	0.808581	0.437893	0.696246	
10	0.857339	0.957418	0.221928	0.113698	
12	0.904365	0.529205	0.512571	0.822403	
...	...	...	...	...	
8733	0.822060	0.189696	0.614397	0.684976	
8745	0.859894	0.897421	0.719463	0.181414	
8748	0.858196	0.937465	0.463454	0.219193	
8788	0.817624	0.893283	0.324179	0.024556	
8789	0.889955	0.936086	0.535802	0.513582	

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g)	\
5	0.676104	0.443583	
7	0.684742	0.451553	
8	0.626129	0.398092	
10	0.000000	0.000000	
12	0.632893	0.671998	
...	...	...	
8733	0.000000	0.000000	
8745	0.024320	0.103539	
8748	0.080576	0.140080	
8788	0.002465	0.005934	
8789	0.373303	0.259528	

	Saturated Fatty Acids(g)	kmeans_tsne	Agg_tsne
5	0.348734	1	0

22

```
7                      0.355775            1           0
8                      0.308817            1           0
10                     0.000000            2           0
12                     0.494853            1           0
...                      ...             ...         ...
8733                   0.478288            1           0
8745                   0.063870            1           0
8748                   0.000000            1           0
8788                   0.009211            1           0
8789                   0.275650            1           0

[2793 rows x 9 columns]

       Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
0          0.685223         0.421947    0.644650        0.122492
1          0.770972         0.539328    0.673774        0.444529
14         0.816493         0.139076    0.695060        0.644735
100        0.786376         0.000000    0.756312        0.513785
101        0.717455         0.000000    0.681411        0.382380
...             ...              ...         ...             ...
8750       0.733270         0.000000    0.708751        0.396735
8769       0.628393         0.478249    0.514124        0.043087
8771       0.632419         0.346116    0.532774        0.232190
8772       0.601603         0.330664    0.537600        0.071353
8773       0.673693         0.347863    0.512571        0.388237

       Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
0                            0.022900                        0.022872
1                            0.297219                        0.227549
14                           0.511857                        0.352783
100                          0.351960                        0.293664
101                          0.175770                        0.224155
...                               ...                             ...
8750                         0.246740                        0.204303
8769                         0.013127                        0.002529
8771                         0.089229                        0.016934
8772                         0.011634                        0.002758
8773                         0.257480                        0.088905

       Saturated Fatty Acids(g)  kmeans_tsne  Agg_tsne
0                      0.030388            2         1
1                      0.212894            0         1
14                     0.395637            0         1
100                    0.254692            0         1
101                    0.180514            0         1
...                         ...          ...       ...
8750                   0.180322            0         1
8769                   0.035842            2         1
```

```
8771                        0.175470          0        1
8772                        0.024208          2        1
8773                        0.267429          0        1

[3069 rows x 9 columns]

      Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
2          0.624253         0.633388    0.203973        0.020652
3          0.466968         0.380891    0.074901        0.056849
4          0.513760         0.468498    0.074901        0.056849
6          0.670663         0.503637    0.457547        0.324453
9          0.721842         0.757620    0.153183        0.055176
...             ...              ...         ...             ...
8783       0.556031         0.463708    0.351950        0.035864
8784       0.656199         0.583869    0.396246        0.175712
8785       0.654500         0.583869    0.396246        0.175712
8786       0.654500         0.583869    0.396246        0.175712
8787       0.642026         0.626764    0.305234        0.000000


      Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
2                           0.000000                        0.000000
3                           0.017754                        0.019922
4                           0.017754                        0.019922
6                           0.222082                        0.060118
9                           0.000000                        0.000000
...                              ...                             ...
8783                        0.010777                        0.001153
8784                        0.061346                        0.007953
8785                        0.066435                        0.008176
8786                        0.066435                        0.008176
8787                        0.000000                        0.000000


      Saturated Fatty Acids(g)  kmeans_tsne  Agg_tsne
2                     0.003042            2         2
3                     0.014394            2         2
4                     0.014394            2         2
6                     0.180322            1         2
9                     0.000000            2         2
...                        ...          ...       ...
8783                  0.024012            2         2
8784                  0.122689            2         2
8785                  0.129330            2         2
8786                  0.129330            2         2
8787                  0.000000            2         2

[2928 rows x 9 columns]
```

0.5.1 Insights

- In Cluster 0, on average is High in Energy and Carbohydrate, with a moderate level of protein and lipid.
- In Cluster 1, on average is Slightly High in Energy with a moderate level of Protein and low level of Carbohydrate.
- In Cluster 2, on average is Slightly High in Energy and Carbohydrate, with a Slight Low level of Protein.

0.6 GMM

```python
from clustering_function import gmm_bic_score, perform_gmm_clustering
from sklearn.mixture import GaussianMixture
from sklearn.model_selection import GridSearchCV
```

```python
gmm_param_grid = {
    "n_components": range(1, 6),
    "covariance_type": ["spherical", "tied", "diag", "full"],
}
grid_search = GridSearchCV(
    GaussianMixture(), param_grid=gmm_param_grid, scoring=gmm_bic_score
)
```

```python
grid_search.fit(X_tsne_reduced)
Gmm_tsne_results = grid_search.cv_results_
```

```python
Gmm_tsne_df = pd.DataFrame(Gmm_tsne_results)[
    ["param_n_components", "param_covariance_type", "mean_test_score"]
]
Gmm_tsne_df["mean_test_score"] = -Gmm_tsne_df["mean_test_score"]
Gmm_tsne_df = Gmm_tsne_df.rename(
    columns={
        "param_n_components": "Number of Components",
        "param_covariance_type": "Type of Covariance",
        "mean_test_score": "BIC Score",
    }
)
Gmm_tsne_df.sort_values(by="BIC Score").head()
```

[53]:

	Number of Components	Type of Covariance	BIC Score
3	4	spherical	36198.642583
2	3	spherical	36204.807595
7	3	tied	36209.727408
12	3	diag	36232.551749
9	5	tied	36250.230146

```python
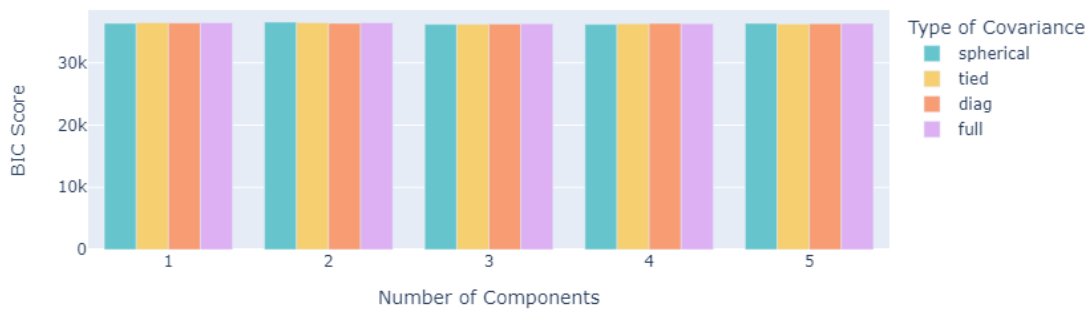fig = px.bar(Gmm_tsne_df, x="Number of Components", y="BIC Score",
             color="Type of Covariance", barmode="group",
```

```
                title="BIC Score by Number of Components and Type of␣
   ↪Covariance",
                color_discrete_sequence=px.colors.qualitative.Pastel)
fig.show()
```

BIC Score by Number of Components and Type of Covariance

[55]:
```
Gmm_tsne_labels,gmm_model = perform_gmm_clustering(X_tsne_reduced, 3,␣
   ↪'spherical')
```

[56]:
```
# export model
with open('../Model_fitted/Gmm_model_mode_pkl', 'wb') as files:
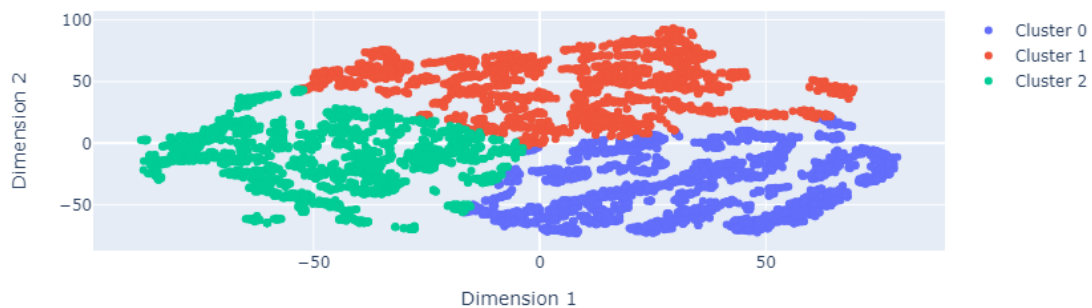    pickle.dump(gmm_model, files)
```

[57]:
```
plot_clusters(X_tsne_reduced, Gmm_tsne_labels, "GMM")
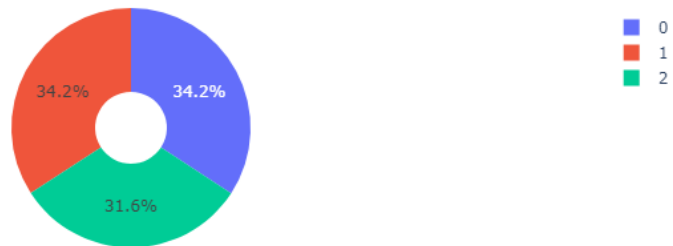```

t-SNE with GMM

```
[58]: X_with_labels['Gmm_tsne'] = Gmm_tsne_labels
      scores['Gmm_tsne'] = (silhouette_score(X_tsne_reduced,␣
       ↪X_with_labels['Gmm_tsne']))
```

```
[59]: food_groups, value_counts = get_food_groups(X_with_labels['Gmm_tsne'],␣
       ↪'GMM_tsne',food_groups)
```

```
[60]: plot_cluster_distribution(value_counts, 'GMM')
```

GMM using t-SNE classes distribution


```
[61]: GmmTsne = X_with_labels.groupby('Gmm_tsne')
```

```
[62]: n = GmmTsne['Gmm_tsne'].count().count() # number of cluster
      for i in range(0,n):
          display(GmmTsne.get_group(i))
```

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
1	0.770972	0.539328	0.673774	0.444529	
14	0.816493	0.139076	0.695060	0.644735	
100	0.786376	0.000000	0.756312	0.513785	
101	0.717455	0.000000	0.681411	0.382380	
103	0.999184	0.000000	0.000000	0.998925	
…	…	…	…	…	
8731	0.776288	0.412147	0.653097	0.543778	
8749	0.769417	0.000000	0.762047	0.442852	
8750	0.733270	0.000000	0.708751	0.396735	
8771	0.632419	0.346116	0.532774	0.232190	
8773	0.673693	0.347863	0.512571	0.388237	

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g)	\
1	0.297219	0.227549	
14	0.511857	0.352783	

100	0.351960	0.293664
101	0.175770	0.224155
103	0.861886	0.577201
...	...	...
8731	0.282218	0.443141
8749	0.000000	0.000000
8750	0.246740	0.204303
8771	0.089229	0.016934
8773	0.257480	0.088905

	Saturated Fatty Acids(g)	kmeans_tsne	Agg_tsne	Gmm_tsne
1	0.212894	0	1	0
14	0.395637	0	1	0
100	0.254692	0	1	0
101	0.180514	0	1	0
103	0.807114	0	0	0
...	...	...	...	...
8731	0.209057	0	1	0
8749	0.000000	0	1	0
8750	0.180322	0	1	0
8771	0.175470	0	1	0
8773	0.267429	0	1	0

[3006 rows x 10 columns]

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
5	0.913731	0.870851	0.476131	0.745258	
6	0.670663	0.503637	0.457547	0.324453	
7	0.916054	0.866962	0.491577	0.753711	
8	0.875881	0.808581	0.437893	0.696246	
10	0.857339	0.957418	0.221928	0.113698	
...	...	...	...	...	
8774	0.686603	0.550199	0.495216	0.296003	
8775	0.682424	0.553930	0.493283	0.275738	
8777	0.670663	0.510318	0.504409	0.271448	
8788	0.817624	0.893283	0.324179	0.024556	
8789	0.889955	0.936086	0.535802	0.513582	

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g)	\
5	0.676104	0.443583	
6	0.222082	0.060118	
7	0.684742	0.451553	
8	0.626129	0.398092	
10	0.000000	0.000000	
...	...	...	
8774	0.125424	0.028458	
8775	0.112813	0.024749	
8777	0.138099	0.042725	

		0.002465			0.005934
8788					
8789		0.373303			0.259528

	Saturated Fatty Acids(g)	kmeans_tsne	Agg_tsne	Gmm_tsne
5	0.348734	1	0	1
6	0.180322	1	2	1
7	0.355775	1	0	1
8	0.308817	1	0	1
10	0.000000	2	0	1
...	...	...	...	...
8774	0.227757	2	2	1
8775	0.208973	2	2	1
8777	0.197391	2	2	1
8788	0.009211	1	0	1
8789	0.275650	1	0	1

[3005 rows x 10 columns]

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g)	\
0	0.685223	0.421947	0.644650	0.122492
2	0.624253	0.633388	0.203973	0.020652
3	0.466968	0.380891	0.074901	0.056849
4	0.513760	0.468498	0.074901	0.056849
9	0.721842	0.757620	0.153183	0.055176
...	...	...	...	...
8783	0.556031	0.463708	0.351950	0.035864
8784	0.656199	0.583869	0.396246	0.175712
8785	0.654500	0.583869	0.396246	0.175712
8786	0.654500	0.583869	0.396246	0.175712
8787	0.642026	0.626764	0.305234	0.000000

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g)	\
0	0.022900	0.022872
2	0.000000	0.000000
3	0.017754	0.019922
4	0.017754	0.019922
9	0.000000	0.000000
...	...	...
8783	0.010777	0.001153
8784	0.061346	0.007953
8785	0.066435	0.008176
8786	0.066435	0.008176
8787	0.000000	0.000000

	Saturated Fatty Acids(g)	kmeans_tsne	Agg_tsne	Gmm_tsne
0	0.030388	2	1	2
2	0.003042	2	2	2
3	0.014394	2	2	2

```
4                        0.014394           2        2        2
9                        0.000000           2        2        2
...                          ...          ...      ...      ...
8783                     0.024012           2        2        2
8784                     0.122689           2        2        2
8785                     0.129330           2        2        2
8786                     0.129330           2        2        2
8787                     0.000000           2        2        2
```

```
[2779 rows x 10 columns]
```

0.6.1 Insights

- In Cluster 0, on average is Slightly High in Energy, with a moderate level of Protein and low level of Carbohydrate.
- In Cluster 1, on average is Very High in Energy, High level in Carbohydrate with a moderate level of Protein.
- In Cluster 2, on average is Slightly High in Energy, with a moderate level of Carbohydrate and slightly low level of Protein.

0.7 K-Medoids

```python
[63]: from clustering_function import tuning_kmedoids, perform_kmd_clustering
```

```python
[64]: kmd_param_grid = {
          'n_clusters': [3, 4],
          'method': ['alternate', 'pam'],
          'init' : ['random', 'heuristic', 'k-medoids++','build']
      }
```

```python
[65]: tuning_kmedoids(X_tsne_reduced, kmd_param_grid)
```

```
Best silhouette score: 0.43088004
Best parameters: {'init': 'random', 'method': 'alternate', 'n_clusters': 3}
```

```python
[66]: Kmd_tsne_labels,kmd_model = perform_kmd_clustering(X_tsne_reduced, 3, 'random',
      ↪'alternate')
```

```python
[67]: # export model
      with open('../Model_fitted/Kmd_model_mode_pkl', 'wb') as files:
          pickle.dump(kmd_model, files)
```

```python
[68]: plot_clusters(X_tsne_reduced, Kmd_tsne_labels, "K-Medoids")
```

t-SNE with K-Medoids
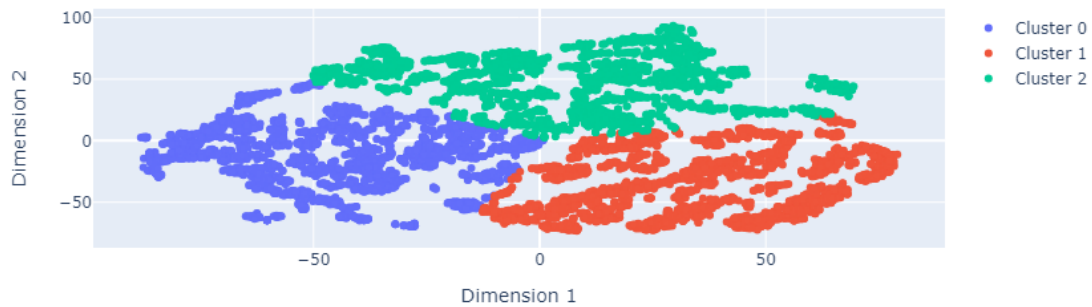

```
[69]: X_with_labels['Kmd_tsne'] = Kmd_tsne_labels
      scores['Kmd_tsne'] = (silhouette_score(X_tsne_reduced,
       ↪X_with_labels['Kmd_tsne']))
```

```
[70]: food_groups, value_counts = get_food_groups(X_with_labels['Kmd_tsne'],
       ↪'KMedoids_tsne',food_groups)
```

```
[71]: plot_cluster_distribution(value_counts, 'K-Medoids')
```

K-Medoids using t-SNE classes distribution


```
[72]: Kmdtsne = X_with_labels.groupby('Kmd_tsne')
```

```
[73]: n = Kmdtsne['Kmd_tsne'].count().count() # number of cluster
      for i in range(0,n):
          display(Kmdtsne.get_group(i))
```

```
      Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
0         0.685223         0.421947    0.644650        0.122492
2         0.624253         0.633388    0.203973        0.020652
3         0.466968         0.380891    0.074901        0.056849
4         0.513760         0.468498    0.074901        0.056849
6         0.670663         0.503637    0.457547        0.324453
...            ...              ...         ...             ...
8784      0.656199         0.583869    0.396246        0.175712
8785      0.654500         0.583869    0.396246        0.175712
8786      0.654500         0.583869    0.396246        0.175712
8787      0.642026         0.626764    0.305234        0.000000
8788      0.817624         0.893283    0.324179        0.024556

      Monounsaturated Fatty Acids(g)  Polyunsaturated Fatty Acids(g)  \
0                           0.022900                        0.022872
2                           0.000000                        0.000000
3                           0.017754                        0.019922
4                           0.017754                        0.019922
6                           0.222082                        0.060118
...                              ...                             ...
8784                        0.061346                        0.007953
8785                        0.066435                        0.008176
8786                        0.066435                        0.008176
8787                        0.000000                        0.000000
8788                        0.002465                        0.005934

      Saturated Fatty Acids(g)  kmeans_tsne  Agg_tsne  Gmm_tsne  Kmd_tsne
0                     0.030388            2         1         2         0
2                     0.003042            2         2         2         0
3                     0.014394            2         2         2         0
4                     0.014394            2         2         2         0
6                     0.180322            1         2         1         0
...                        ...          ...       ...       ...       ...
8784                  0.122689            2         2         2         0
8785                  0.129330            2         2         2         0
8786                  0.129330            2         2         2         0
8787                  0.000000            2         2         2         0
8788                  0.009211            1         0         1         0

[2987 rows x 11 columns]
      Energy (Kcal)  Carbohydrate(g)  Protein(g)  Total Lipid(g)  \
1         0.770972         0.539328    0.673774        0.444529
14        0.816493         0.139076    0.695060        0.644735
100       0.786376         0.000000    0.756312        0.513785
101       0.717455         0.000000    0.681411        0.382380
103       0.999184         0.000000    0.000000        0.998925
...            ...              ...         ...             ...
```

	0	1	2	3
8731	0.776288	0.412147	0.653097	0.543778
8749	0.769417	0.000000	0.762047	0.442852
8750	0.733270	0.000000	0.708751	0.396735
8771	0.632419	0.346116	0.532774	0.232190
8773	0.673693	0.347863	0.512571	0.388237

	Monounsaturated Fatty Acids(g)	Polyunsaturated Fatty Acids(g) \
1	0.297219	0.227549
14	0.511857	0.352783
100	0.351960	0.293664
101	0.175770	0.224155
103	0.861886	0.577201
…	…	…
8731	0.282218	0.443141
8749	0.000000	0.000000
8750	0.246740	0.204303
8771	0.089229	0.016934
8773	0.257480	0.088905

	Saturated Fatty Acids(g)	kmeans_tsne	Agg_tsne	Gmm_tsne	Kmd_tsne
1	0.212894	0	1	0	1
14	0.395637	0	1	0	1
100	0.254692	0	1	0	1
101	0.180514	0	1	0	1
103	0.807114	0	0	0	1
…	…	…	…	…	…
8731	0.209057	0	1	0	1
8749	0.000000	0	1	0	1
8750	0.180322	0	1	0	1
8771	0.175470	0	1	0	1
8773	0.267429	0	1	0	1

[2928 rows x 11 columns]

	Energy (Kcal)	Carbohydrate(g)	Protein(g)	Total Lipid(g) \
5	0.913731	0.870851	0.476131	0.745258
7	0.916054	0.866962	0.491577	0.753711
8	0.875881	0.808581	0.437893	0.696246
12	0.904365	0.529205	0.512571	0.822403
13	0.862406	0.577933	0.329815	0.756303
…	…	…	…	…
8701	0.834115	0.883149	0.529703	0.198541
8733	0.822060	0.189696	0.614397	0.684976
8745	0.859894	0.897421	0.719463	0.181414
8748	0.858196	0.937465	0.463454	0.219193
8789	0.889955	0.936086	0.535802	0.513582

Monounsaturated Fatty Acids(g) Polyunsaturated Fatty Acids(g) \

```
5                                    0.676104                            0.443583
7                                    0.684742                            0.451553
8                                    0.626129                            0.398092
12                                   0.632893                            0.671998
13                                   0.615989                            0.517992
...                                     ...                                ...
8701                                 0.040132                            0.110093
8733                                 0.000000                            0.000000
8745                                 0.024320                            0.103539
8748                                 0.080576                            0.140080
8789                                 0.373303                            0.259528

        Saturated Fatty Acids(g)  kmeans_tsne  Agg_tsne  Gmm_tsne  Kmd_tsne
5                       0.348734            1         0         1         2
7                       0.355775            1         0         1         2
8                       0.308817            1         0         1         2
12                      0.494853            1         0         1         2
13                      0.473315            1         0         1         2
...                          ...          ...       ...       ...       ...
8701                    0.051085            1         0         1         2
8733                    0.478288            1         0         1         2
8745                    0.063870            1         0         1         2
8748                    0.000000            1         0         1         2
8789                    0.275650            1         0         1         2

[2875 rows x 11 columns]
```

0.7.1 Insights

- In Cluster 0, on average is moderate level in Energy and Carbohydrate, with a slightly low level of protein.
- In Cluster 1, on average is Slightly High in Energy, low level in Carbohydrate with a moderate level of Protein.
- In Cluster 2, on average is Very High in Energy and Carbohydrate, with a moderate level of protein and lipid and slightly low level of Saturated Fatty Acids.

[74]: `scores`

[74]:
```
{'kmeans_tsne': 0.4307466,
 'Agg_tsne': 0.4223424,
 'Gmm_tsne': 0.4299104,
 'Kmd_tsne': 0.43088004}
```

The K-Medoids model was shown to be the most effective one when the dataset's missing values were filled in using the mean approach.

[ ]: