```
In [ ]: # Import Libraries
In [1]: import numpy as np
               import pandas as pd
               import matplotlib.pyplot as plt
               import seaborn as sns
               from sklearn.decomposition import PCA
               from yellowbrick.cluster import KElbowVisualizer
               from sklearn.cluster import AgglomerativeClustering
               from sklearn.cluster import KMeans
               from sklearn.cluster import DBSCAN
               from kmodes.kprototypes import KPrototypes
               from sklearn.cluster import KMeans
In [2]: # import the dataset and Removing the customer id column from the dataset as it provides no useful data, no split for k-mean
In [3]: | dataset = pd.read_csv("marketing_campaign.csv", sep='\t')
               dataset.head()
Out[3]:
                         ID Year_Birth
                                               Education Marital_Status
                                                                                                 \label{lem:control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_c
                                                                                     Income
                0 5524
                                                                         Single 58138.0
                                                                                                                                   04-09-2012
                                                                                                                                                                          635 ...
                                                                                                                                                                                                                                       n
                                     1957 Graduation
                                                                                                                                                            58
                                                                         Single 46344.0
                1 2174
                                                                                                                                   08-03-2014
                                                                                                                                                            38
                                                                                                                                                                                                                5
                                     1954 Graduation
                                                                                                            1
                                                                                                                            1
                                                                                                                                                                            11 ...
                                                                                                                                                                                                                                       0
                2 4141
                                     1965 Graduation
                                                                       Together 71613.0
                                                                                                            n
                                                                                                                            n
                                                                                                                                  21-08-2013
                                                                                                                                                                                                                 4
                                                                                                                                                                                                                                       0
                                                                                                                                                            26
                                                                                                                                                                          426
                                                                       Together 26646.0
                                                                                                                            n
                                                                                                                                   10-02-2014
                3 6182
                                     1984 Graduation
                                                                                                            1
                                                                                                                                                            26
                                                                                                                                                                            11 ...
                                                                                                                                                                                                                 6
                                                                                                                                                                                                                                       0
                4 5324
                                     1981
                                                       PhD
                                                                        Married 58293.0
                                                                                                                            0
                                                                                                                                   19-01-2014
                                                                                                                                                            94
                                                                                                                                                                          173 ...
                                                                                                                                                                                                                 5
                                                                                                                                                                                                                                       0
               5 rows × 29 columns
In [4]: n_row, n_col = dataset.shape
               print("The number of rows in dataset are {0} and the number of columns are {1}".format(n_row,n_col))
               The number of rows in dataset are 2240 and the number of columns are 29
In [5]: # Information on dataset before removing unnecessary columns
In [6]: dataset.info()
               <class 'pandas.core.frame.DataFrame'>
               RangeIndex: 2240 entries, 0 to 2239
               Data columns (total 29 columns):
                       Column
                                                              Non-Null Count
                #
                                                                                           Dtvpe
               ---
                0
                        ID
                                                              2240 non-null
                                                                                            int64
                1
                        Year Birth
                                                              2240 non-null
                                                                                            int64
                        Education
                                                              2240 non-null
                                                                                            object
                2
                        Marital_Status
                3
                                                              2240 non-null
                                                                                            object
                4
                        Income
                                                              2216 non-null
                                                                                            float64
                        Kidhome
                                                              2240 non-null
                 5
                                                                                            int64
                        Teenhome
                                                              2240 non-null
                 6
                                                                                            int64
                        Dt Customer
                                                              2240 non-null
                                                                                            obiect
                 7
                8
                        Recency
                                                              2240 non-null
                                                                                            int64
                 9
                        MntWines
                                                              2240 non-null
                                                                                            int64
                10
                        MntFruits
                                                              2240 non-null
                                                                                            int64
                        MntMeatProducts
                                                              2240 non-null
                 11
                                                                                            int64
                                                              2240 non-null
                12
                        MntFishProducts
                                                                                            int64
                 13
                        {\tt MntSweetProducts}
                                                              2240 non-null
                                                                                            int64
                                                               2240 non-null
                 14
                        MntGoldProds
                                                                                            int64
                15
                        NumDealsPurchases
                                                              2240 non-null
                                                                                            int64
                        NumWebPurchases
                                                              2240 non-null
                16
                                                                                            int64
                17
                        NumCatalogPurchases
                                                              2240 non-null
                                                                                            int64
                18
                        NumStorePurchases
                                                               2240 non-null
                                                                                            int64
                 19
                        NumWebVisitsMonth
                                                              2240 non-null
                                                                                            int64
                       AcceptedCmp3
                                                              2240 non-null
                 20
                                                                                            int64
                 21
                        AcceptedCmp4
                                                              2240 non-null
                                                                                            int64
                22
                        AcceptedCmp5
                                                              2240 non-null
                                                                                            int64
                                                              2240 non-null
                 23
                        AcceptedCmp1
                                                                                            int64
                        AcceptedCmp2
                                                              2240 non-null
                 24
                                                                                            int64
                 25
                        Complain
                                                              2240 non-null
                                                                                            int64
                 26
                        Z_CostContact
                                                              2240 non-null
                                                                                            int64
                 27
                        Z Revenue
                                                              2240 non-null
                                                                                            int64
                28 Response
                                                              2240 non-null
                                                                                            int64
               dtypes: float64(1), int64(25), object(3)
               memory usage: 507.6+ KB
```

```
In [7]: dataset.describe(include='all')
 Out[7]:
                                                                                            Teenhome Dt_Customer
                                Year_Birth Education Marital_Status
                                                                                                                               MntWines ... NumWe
                           ID
                                                                       Income
                                                                                 Kidhome
                                                                                                                     Recency
                  2240.000000
                              2240.000000
                                              2240
                                                           2240
                                                                   2216.000000 2240.000000
                                                                                          2240.000000
                                                                                                            2240
                                                                                                                 2240.000000
                                                                                                                             2240.000000
            count
                                                              8
                                                                                                             663
           unique
                         NaN
                                     NaN
                                                 5
                                                                         NaN
                                                                                     NaN
                                                                                                NaN
                                                                                                                        NaN
                                                                                                                                    NaN
                         NaN
                                     NaN
                                         Graduation
                                                          Married
                                                                         NaN
                                                                                     NaN
                                                                                                NaN
                                                                                                       31-08-2012
                                                                                                                        NaN
                                                                                                                                    NaN
             top
                                     NaN
                                              1127
                                                            864
                                                                         NaN
                                                                                     NaN
                                                                                                NaN
                                                                                                              12
                                                                                                                        NaN
                                                                                                                                    NaN
             freq
                         NaN
                  5592.159821
                              1968.805804
                                                                  52247.251354
                                                                                 0.444196
                                                                                             0.506250
                                                                                                                   49.109375
                                                                                                                              303.935714
                                               NaN
                                                            NaN
                                                                                                             NaN
            mean
                                11.984069
                                                                  25173.076661
                                                                                 0.538398
                                                                                             0.544538
                                                                                                                   28.962453
                                                                                                                              336.597393
             std
                  3246.662198
                                               NaN
                                                            NaN
                                                                                                             NaN
             min
                     0.000000
                              1893.000000
                                               NaN
                                                            NaN
                                                                   1730.000000
                                                                                 0.000000
                                                                                             0.000000
                                                                                                             NaN
                                                                                                                     0.000000
                                                                                                                                0.000000
             25%
                  2828.250000 1959.000000
                                               NaN
                                                            NaN
                                                                  35303.000000
                                                                                 0.000000
                                                                                             0.000000
                                                                                                             NaN
                                                                                                                   24.000000
                                                                                                                               23.750000
             50%
                  5458.500000
                              1970.000000
                                               NaN
                                                            NaN
                                                                  51381.500000
                                                                                 0.000000
                                                                                             0.000000
                                                                                                             NaN
                                                                                                                    49.000000
                                                                                                                               173.500000
                  8427.750000 1977.000000
                                                                  68522.000000
                                                                                 1.000000
                                                                                             1.000000
                                                                                                                    74.000000
                                                                                                                              504.250000
             75%
                                               NaN
                                                                                                             NaN
            max 11191,000000 1996,000000
                                                                 666666.000000
                                                                                 2,000000
                                                                                             2,000000
                                               NaN
                                                            NaN
                                                                                                             NaN
                                                                                                                    99.000000
                                                                                                                             1493.000000
          11 rows × 29 columns
 In [8]: # The columns for Z_CostContact and Z_Revenue provide no useful information for my analysis, therefore I remove them.
 In [9]: dataset.drop(columns = ["ID","Z_CostContact","Z_Revenue"], inplace = True)
In [10]: # Information on dataset after removing unnecessary columns
In [11]: dataset.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 2240 entries, 0 to 2239
          Data columns (total 26 columns):
           #
               Column
                                      Non-Null Count
                                                        Dtype
           0
                                      2240 non-null
               Year Birth
                                                        int64
                                      2240 non-null
           1
               Education
                                                        object
           2
               Marital_Status
                                      2240 non-null
                                                        object
                                       2216 non-null
                                                        float64
                Income
           4
               Kidhome
                                      2240 non-null
                                                        int64
           5
                                       2240 non-null
               Teenhome
                                                        int64
           6
               Dt Customer
                                      2240 non-null
                                                        object
                Recency
                                       2240 non-null
                                                        int64
           8
               MntWines
                                       2240 non-null
                                                        int64
                                       2240 non-null
           9
               MntFruits
                                                        int64
           10
               MntMeatProducts
                                       2240 non-null
                                                        int64
           11
               MntFishProducts
                                      2240 non-null
                                                        int64
                                       2240 non-null
           12
               MntSweetProducts
                                                        int64
           13
               MntGoldProds
                                       2240 non-null
                                                        int64
                                       2240 non-null
           14
               NumDealsPurchases
                                                        int64
           15
               NumWebPurchases
                                      2240 non-null
                                                        int64
           16
               NumCatalogPurchases
                                      2240 non-null
                                                        int64
           17
               NumStorePurchases
                                       2240 non-null
                                                        int64
                                       2240 non-null
           18
               NumWebVisitsMonth
                                                        int64
                                       2240 non-null
           19
               AcceptedCmp3
                                                        int64
           20
               AcceptedCmp4
                                       2240 non-null
                                                        int64
           21
               AcceptedCmp5
                                       2240 non-null
                                                        int64
           22
               AcceptedCmp1
                                       2240 non-null
                                                        int64
                                       2240 non-null
           23
               AcceptedCmp2
                                                        int64
           24
               Complain
                                      2240 non-null
                                                        int64
                                       2240 non-null
                                                        int64
               Response
          dtypes: float64(1), int64(22), object(3)
          memory usage: 455.1+ KB
In [12]: # we have a mix data type, Dt_Customer is data type object
          # Education and Marital_Status are categorical features
          # Income has missing values
```

```
In [13]: # exploring the categorial features
         print(dataset["Marital_Status"].value_counts())
         Married
                     864
         Together
                     580
         Single
                     480
         Divorced
                     232
         Widow
                      77
         Alone
                       3
         Absurd
                       2
         YOLO
                       2
         Name: Marital_Status, dtype: int64
In [14]: print(dataset["Education"].value_counts())
         Graduation
                       1127
                        486
         PhD
         Master
                        370
         2n Cvcle
                        203
         Basic
                         54
         Name: Education, dtype: int64
In [15]: # Feature Engineering
In [16]: |dataset["Marital_Status"].replace({'Alone': 'Single', 'Absurd': 'Single', 'YOLO': 'Single'}, inplace=True)
In [17]: # replace date of birth with age
In [18]: dataset['Year_Birth']= 2022-dataset['Year_Birth']
         dataset.rename(columns={"Year_Birth": "Age"},inplace=True)
In [19]: dataset.head()
Out[19]:
                 Education Marital_Status Income Kidhome Teenhome Dt_Customer Recency MntWines MntFruits ... NumCatalogPurchases NumStorePurchases
                                                               04-09-2012
          0
             65 Graduation
                                Single 58138.0
                                                   0
                                                            0
                                                                              58
                                                                                     635
                                                                                               88 ...
                                                                                                                   10
                                                                                                                                     4
             68 Graduation
                                                               08-03-2014
                                                                              38
                                                                                      11
                                                                                               1 ...
                                                                                                                                     2
                                Single 46344.0
                                                            1
                                                                                                                    1
                                                           0
                                                                                               49 ...
             57 Graduation
                              Together 71613.0
                                                   0
                                                               21-08-2013
                                                                             26
                                                                                     426
                                                                                                                    2
                                                                                                                                    10
          3
             38
                Graduation
                              Together 26646.0
                                                           0
                                                               10-02-2014
                                                                             26
                                                                                      11
                                                                                               4 ...
                                                                                                                    0
                                                                                               43 ...
             41
                     PhD
                               Married 58293.0
                                                            0
                                                               19-01-2014
                                                                              94
                                                                                     173
         5 rows × 26 columns
In [20]: #Total spendings on all items
         dataset["Total_Spent"] = dataset["MntWines"]+ dataset["MntFruits"]+ dataset["MntMeatProducts"]+ dataset["MntFishProducts"]+
In [21]: # creating a new feature showing the number of days of customer engagement
In [22]: | dataset['Dt_Customer'] = pd.to_datetime(dataset.Dt_Customer)
         newest_customer = dataset['Dt_Customer'].max()
         dataset['newest_customer'] = newest_customer
         dataset['days_engaged'] = (dataset['newest_customer'] - dataset['Dt_Customer']).dt.days
         print(dataset['days_engaged'])
         dataset.drop(columns=['Dt_Customer', 'newest_customer'],inplace=True)
         0
                 971
         1
                 125
         2
                 472
         3
                  65
                 321
         4
         2235
                 541
         2236
                  61
         2237
                 315
         2238
                 316
         2239
                 782
         Name: days_engaged, Length: 2240, dtype: int64
         C:\Users\yasha\AppData\Local\Temp\ipykernel_31952\4042051918.py:1: UserWarning: Parsing '21-08-2013' in DD/MM/YYYY forma
         t. Provide format or specify infer_datetime_format=True for consistent parsing.
           dataset['Dt_Customer'] = pd.to_datetime(dataset.Dt_Customer)
         C:\Users\yasha\AppData\Local\Temp\ipykernel_31952\4042051918.py:1: UserWarning: Parsing '19-01-2014' in DD/MM/YYYY forma
         {\tt t. \ Provide \ format \ or \ specify \ infer\_date time\_format = True \ for \ consistent \ parsing.}
           dataset['Dt_Customer'] = pd.to_datetime(dataset.Dt_Customer)
```

```
In [23]: dataset.describe()
Out[23]:
                                                                                                  MntFruits MntMeatProducts MntFishProducts MntSweetProducts
                         Age
                                     Income
                                                Kidhome
                                                           Teenhome
                                                                         Recency
                                                                                     MntWines
                  2240.000000
                                2216.000000
                                             2240.000000
                                                         2240.000000
                                                                     2240.000000
                                                                                  2240.000000
                                                                                               2240.000000
                                                                                                               2240.000000
                                                                                                                               2240.000000
                                                                                                                                                 2240.000000
            count
                    53.194196
                                52247.251354
                                                0.444196
                                                             0.506250
                                                                        49.109375
                                                                                   303.935714
                                                                                                 26.302232
                                                                                                                166.950000
                                                                                                                                 37.525446
                                                                                                                                                   27.062946
            mean
                                                                                                 39.773434
                    11.984069
                               25173.076661
                                                0.538398
                                                             0.544538
                                                                        28.962453
                                                                                   336.597393
                                                                                                                225.715373
                                                                                                                                 54.628979
                                                                                                                                                   41.280498
             std
                    26.000000
                                1730.000000
                                                0.000000
                                                             0.000000
                                                                         0.000000
                                                                                     0.000000
                                                                                                  0.000000
                                                                                                                  0.000000
                                                                                                                                  0.000000
                                                                                                                                                    0.000000
             min
             25%
                    45.000000
                               35303.000000
                                                0.000000
                                                             0.000000
                                                                        24.000000
                                                                                    23.750000
                                                                                                  1.000000
                                                                                                                  16.000000
                                                                                                                                  3.000000
                                                                                                                                                    1.000000
             50%
                    52.000000
                               51381.500000
                                                0.000000
                                                             0.000000
                                                                        49.000000
                                                                                   173.500000
                                                                                                  8.000000
                                                                                                                 67.000000
                                                                                                                                 12.000000
                                                                                                                                                    8.000000
             75%
                    63.000000
                               68522.000000
                                                1.000000
                                                             1.000000
                                                                        74.000000
                                                                                   504.250000
                                                                                                 33.000000
                                                                                                                232.000000
                                                                                                                                 50.000000
                                                                                                                                                   33.000000
             max
                   129.000000 666666.000000
                                                2.000000
                                                             2.000000
                                                                        99.000000
                                                                                  1493.000000
                                                                                                199.000000
                                                                                                               1725.000000
                                                                                                                                259.000000
                                                                                                                                                  263.000000
          8 rows × 25 columns
In [24]: # Data Cleaning process in order to target NAN values
In [25]: dataset.isna().sum()
          dataset.dropna(inplace = True)
In [26]: # checking on the relevant features
          plt.figure()
In [27]:
           cols_to_plot = ['Income', 'Age', 'Total_Spent']
           sns.pairplot(dataset[cols_to_plot], diag_kind='kde', diag_kws={'color':'g'}, plot_kws={'color':'y'})
          plt.show()
           <Figure size 800x550 with 0 Axes>
               600000
               400000
            Income
               200000
                     0
                   120
                  100
                    80
                    60
                    40
                 2500
                 2000
              Total Spent
                  1500
                  1000
                  500
                     0
                         0
                              200000 400000 600000
                                                              50
                                                                          100
                                                                                          0
                                                                                                 1000
                                                                                                         2000
                                                                                                                 3000
                                                                                               Total_Spent
                                   Income
                                                                    Age
```

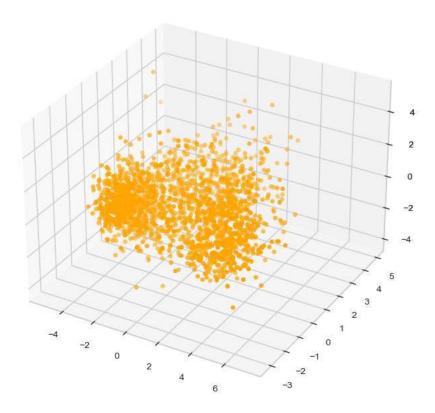
In [28]: # Dropping the outliers

```
In [29]: dataset = dataset[(dataset["Age"]<100)]</pre>
           dataset = dataset[(dataset["Income"]<600000)]</pre>
In [30]: # num of samples after cleaning
           print(len(dataset))
           2212
In [31]: # Show casing the correlation among the features
In [32]: plt.figure(figsize=(16,9))
           sns.heatmap(dataset.corr(), cmap='viridis', annot = True)
           plt.title('Correlation Matrix')
           plt.show()
                                                                              Correlation Matrix
                                                                                                                                                          1.0
                                  0.2 -0.24 0.36 0.016 0.16 0.0130.0340.0410.022 0.06 0.066 0.16 0.13 0.14 -0.12-0.061 0.07-0.0190.0120.0078.00490.0210.12-0.02
                                   1 0.510 0350 008 0.69 0.51 0.69 0.52 0.52 0.39 0.11 0.46 0.7 0.63 0.65 0.0150 .22 0.4 0.33 0.1 0.0280 16 0.79 0.02
                              Kidhome
                                                                                                                                                          0.8
                               ) 36 0 0350 035<mark> 1 0</mark> 0 014) 00390 18 0 26 0 21 0 160 0190 39 0 16 0 110 049 0 13 0 0430 038 0 19 0 15 0 016 00770 15 0 140 008
                     Teenhome
                               .0160.0080.0110.014 1 0.0160.0058.026.00076.0250.0180.0026.0057.024.00046.0190.0320.018.00028.0240.0018.0057-0.2 0.02.0.031
                     MntWines
                              0.16 <mark>0.69 -0.50</mark> 0039.016<mark>-1 | 0.39 0.57 0.4 0.39 0.390.00910.55 0.63 0.64 -0.320.061 0.37 0.47 0.35 0.21-0.0360.25 <mark>0.89 0</mark>.15</mark>
                                                                                                                                                          0.6
                              .013 0.51 <mark>-0.37 -0.18</mark>0.00530.39 <mark>1 0.55 0.59 0.57 0.39 -0.13 0.3 0.49 0.46 -0.420.01£0.00660.21 0.190.009£0.0030.12 <mark>0.61</mark> 0.06</mark>
                     MntFruits
                              MntMeatProducts
                MntFishProducts
                              .041 0.52 <mark>-0.39 -0.20.000790.4 | 0.59 0.57 | 1 | 0.58</mark> 0.43 -0.14 0.3 | 0.53 0.46 <mark>-0.45</mark> 00028016 0.19 0.260 00230 0190.11 <mark>0.64</mark> 0.078
                                                                                                                                                          0.4
                              .022 0.52 -0.38 -0.160.025 0.39 0.57 0.53 0.58 1 0.36 -0.12 0.33 0.49 0.46 -0.420 00170.029 0.26 0.25 0.01-0.0210 12 <mark>0.61</mark> 0.076
               MntSweetProducts
                              0.06 0.39 <mark>-0.35-0.019</mark>0.018 0.39 0.39 0.36 0.43 0.36 <mark>1.</mark> 0.053 0.41 0.44 0.39 <mark>-0.25</mark> 0.13 0.024 0.18 0.17 0.051-0.03 0.14 <mark>0.53</mark> 0.15
                  MntGoldProds
                              0.066-0.11 0.22 0.390.0026.00910.13-0.12-0.14-0.120.053 1 0.24-0.0120.066 0.35-0.0230.016-0.18-0.13-0.038.0037.00320.066 0.2
              NumDealsPurchases
                                                                                                                                                          0.2
                              NumWebPurchases
                              NumCatalogPurchases
                              0.14 0.63 -0.5 0.040.00040.64 0.46 0.49 0.46 0.46 0.39 0.066 0.52 0.52 1 -0.43 0.069 0.18 0.21 0.18 0.0850 0.120.036 0.68 0.1
              NumStorePurchases
                                                                                                                                                          0.0
                             0 12 0 65 0 45 0 13 0 019 0 32 0 42 0 54 0 45 0 42 0 25 0 35 0 052 0 52 0 43 1 0 0 610 029 0 28 0 20 007 5 0 210 002 60 5 0 26
              NumWebVisitsMonth
                  AcceptedCmp3
                             0.0640.0150.0160.0430.0320.0610.0150.018.0002800170.13-0.0230.043.0.1-0.0690.061<mark>-11</mark>-0.080.0810.0960.0720.00960.25.0.0530.006
                 AcceptedCmp4
                              0.07 0.22 -0.16 0.0380.018 0.370.006@.0920.0160.0290.0240.016 0.16 0.14 0.18-0.0290.08 1 0.31 0.24 0.3 -0.0270.18 0.25 0.014
                                                                                                                                                          -02
                             0.019 0.4 -0.2 -0.19.00023047 0.21 0.38 0.19 0.26 0.18 -0.18 0.14 0.32 0.21 -0.28 0.081 0.31 1 0.41 0.220.00840.32 0.47 0.02
                 AcceptedCmp5
                 AcceptedCmp1 0.012 0.33 -0.17 -0.150 0.21 0.35 0.19 0.31 0.26 0.25 0.17 -0.13 0.16 0.31 0.18 -0.2 0.096 0.24 0.41 1 0.18-0.025 0.3 0.38-0.03
                              0.0078 0.1 -0.0820.0160.00140.210.0098.0440.00230.01 0.0510.0380.035 0.1 0.0850.0076.072 0.3 0.22 0.18
                                                                                                                        1 -0.011 0.17 0.14 0.006
                  AcceptedCmp2
                                                                                                                                                          -0.4
                              00440.0280.0370.0070.00570.0360.0020.0240.0190.0210.030.00370.0140.0190.0120.0210.00960.0240.00840.0250.011 11 000044.0340.042
                     Complain
                              0.0210 16-0.078-0.15 -0.2 0.25 0.12 0.24 0.11 0.12 0.140 00320.15 0.22 0.0380.00280.25 0.18 0.32 0.3 0.170 0001
                     Response
                              0.12 0.79 -0.56 -0.14 0.02 0.89 0.61 0.85 0.64 0.61 0.53 0.066 0.53 0.78 0.68 -0.5 0.053 0.25 0.47 0.38 0.14-0.0340.26
                    Total Spent
                                                                                                                                      1
                             0.0210.0280.0580.0090.031 0.15 0.06 0.0710.0780.076 0.15 0.2 0.17 0.091 0.1 <u>0.260.0069.0140.0230.0370.0060.042 0.18 0.14</u>
                  days engaged
                                                                                                                                           ays_engages
                                                                                         nCatalogPurcha
In [33]: # Encoding ordinal features using OrdinalEncoder
In [34]: from sklearn.preprocessing import OneHotEncoder, OrdinalEncoder, StandardScaler, MinMaxScaler
           education_order = ['Basic', '2n Cycle', 'Graduation', 'Master', 'PhD']
           oe = OrdinalEncoder(categories = [education_order], dtype=int)
           education_oe = oe.fit_transform(dataset[['Education']])
           dataset_enc= dataset.assign(Education_encode=education_oe)
           print(dataset enc.shape)
           print(dataset_enc[['Education', 'Education_encode']])
           (2212, 28)
                   Education Education encode
           a
                  Graduation
                                                  2
           1
                  Graduation
                                                  2
           2
                  Graduation
                                                  2
           3
                  Graduation
                          PhD
                                                  4
           4
           2235
                  Graduation
                                                  2
           2236
                          PhD
                                                  4
           2237
                  Graduation
           2238
                       Master
                                                  3
           2239
                          PhD
           [2212 rows x 2 columns]
```

```
In [35]: ohe = OneHotEncoder(sparse=False, dtype='int')
         Marital ohe = ohe.fit transform(dataset[['Marital Status']])
         Marital_ohe = pd.DataFrame(data=Marital_ohe,columns=ohe.get_feature_names(['Marital_Status']), index=dataset.index,)
         dataset_enc = pd.concat([dataset_enc,Marital_ohe],axis=1)
         dataset_enc.drop(columns=['Marital_Status', 'Education'], inplace=True)
         C:\Users\yasha\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is d
         eprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
           warnings.warn(msg, category=FutureWarning)
In [36]: dataset_enc.info()
         <class 'pandas.core.frame.DataFrame'>
         Int64Index: 2212 entries, 0 to 2239
         Data columns (total 31 columns):
                                       Non-Null Count Dtype
          # Column
         ---
              -----
          0
                                       2212 non-null
                                                        int64
              Age
              Income
                                       2212 non-null
                                                       float64
          1
          2
              Kidhome
                                       2212 non-null
                                                       int64
          3
              Teenhome
                                       2212 non-null
                                                       int64
          4
              Recency
                                       2212 non-null
                                                        int64
                                       2212 non-null
              MntWines
                                                        int64
          6
              MntFruits
                                       2212 non-null
                                                        int64
              MntMeatProducts
                                       2212 non-null
                                                        int64
          7
          8
              MntFishProducts
                                       2212 non-null
                                                        int64
          9
              MntSweetProducts
                                       2212 non-null
                                                        int64
          10
              MntGoldProds
                                       2212 non-null
                                                        int64
          11 NumDealsPurchases
                                       2212 non-null
                                                        int64
                                       2212 non-null
          12 NumWebPurchases
                                                       int64
          13
              NumCatalogPurchases
                                       2212 non-null
                                                        int64
                                       2212 non-null
          14 NumStorePurchases
                                                        int64
          15
              NumWebVisitsMonth
                                       2212 non-null
                                                        int64
          16 AcceptedCmp3
                                       2212 non-null
                                                        int64
          17 AcceptedCmp4
                                       2212 non-null
                                                        int64
          18 AcceptedCmp5
                                       2212 non-null
                                                        int64
          19 AcceptedCmp1
                                       2212 non-null
                                                        int64
          20 AcceptedCmp2
                                       2212 non-null
                                                        int64
                                       2212 non-null
          21 Complain
                                                        int64
          22 Response
                                       2212 non-null
                                                        int64
                                       2212 non-null
          23
              Total_Spent
                                                        int64
          24 days engaged
                                       2212 non-null
                                                        int64
          25 Education_encode
                                       2212 non-null
                                                        int32
          26 Marital_Status_Divorced 2212 non-null
                                                        int32
          27 Marital_Status_Married 2212 non-null
                                                        int32
          28 Marital_Status_Single 2212 non-null
29 Marital_Status_Together 2212 non-null
                                                        int32
                                                        int32
          30 Marital_Status_Widow
                                       2212 non-null
                                                       int32
         dtypes: float64(1), int32(6), int64(24)
         memory usage: 501.2 KB
In [37]: binary_columns = ['Marital_Status_Divorced','Marital_Status_Married', 'Marital_Status_Single','Marital_Status_Together','Mar
                          ,'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1','AcceptedCmp2', 'Complain', 'Response']
         dataset_to_scaler = dataset_enc.drop(columns=binary_columns)
         scaler = StandardScaler().fit_transform(dataset_to_scaler)
         scaled_dataset = pd.DataFrame(scaler,columns= dataset_to_scaler.columns )
         binary_series = dataset_enc[binary_columns]
         scaled_dataset = pd.concat([scaled_dataset,binary_series],axis=1)
         scaled_dataset.isna().sum()
         scaled_dataset = scaled_dataset.fillna(0)
In [38]: # Dimensionality reduction with PCA
In [39]: from sklearn.compose import ColumnTransformer
         from sklearn.decomposition import PCA
         pca = PCA(n_components=3)
         pca.fit(scaled_dataset)
         PCA_dataset = pd.DataFrame(pca.transform(scaled_dataset), columns=(["feature1","feature2", "feature3"]))
```

```
In [40]: plt.figure(figsize=(10,8))
    plt.axes(projection='3d').scatter(PCA_dataset["feature1"], PCA_dataset["feature2"], PCA_dataset["feature3"])
    plt.axes(projection='3d').scatter(PCA_dataset["feature1"], PCA_dataset["feature2"], PCA_dataset["feature3"]).set_color('oran
    plt.title("A 3D Projection Of Data In The Reduced Dimension")
    plt.show()
```

A 3D Projection Of Data In The Reduced Dimension



```
In [41]: # Clustering
In [42]: from yellowbrick.cluster import KElbowVisualizer
from sklearn.cluster import KMeans
from sklearn.cluster import AgglomerativeClustering
```

```
In [43]: Elbow_M = KElbowVisualizer(KMeans(), k=(2,11))
Elbow_M.fit(PCA_dataset)
Elbow_M.show()
```

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

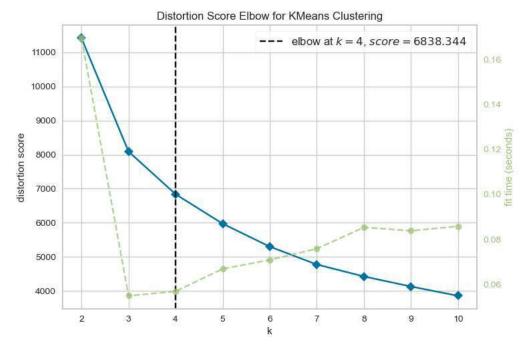
warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(



Out[43]: <AxesSubplot:title={'center':'Distortion Score Elbow for KMeans Clustering'}, xlabel='k', ylabel='distortion score'>

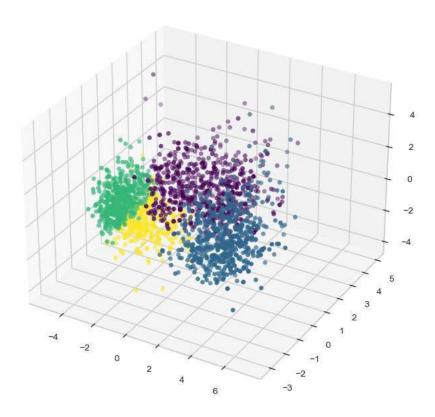
In [44]: # Using elbow curve to find the optimum number of clusters

In [45]: # The Agglomerative Clustering model

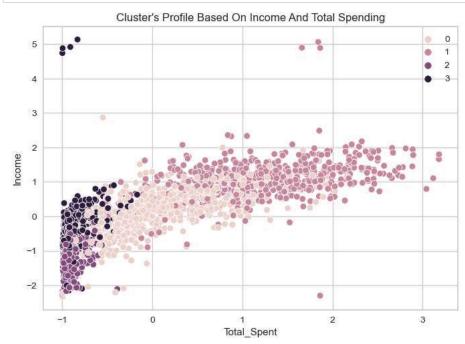
```
In [47]: # Plotting clusters
In [48]: fig = plt.figure(figsize=(10,8))
    plt.axes(projection='3d').scatter(PCA_dataset["feature1"], PCA_dataset["feature2"], PCA_dataset["feature3"], c=PCA_dataset['plt.title("The Clusters by Agglomerative model")

Out[48]: Text(0.5, 0.92, 'The Clusters by Agglomerative model')
```

The Clusters by Agglomerative model







In []: # clustering using K-means model

```
In [59]: kmeans = KMeans(n_clusters =4 , init = 'k-means++', random_state = 50)
# fit model and predict clusters
labels = kmeans.fit_predict(PCA_dataset)
PCA_dataset["Clusters"] = labels
#Adding the Clusters feature to the orignal dataframe.
scaled_dataset["Clusters"]= labels

#Plotting the clusters
fig = plt.figure(figsize=(10,8))
ax = plt.subplot(111, projection='3d', label="bla")
ax.scatter(PCA_dataset["feature1"], PCA_dataset["feature2"], PCA_dataset["feature3"], s=40, c=PCA_dataset["Clusters"], marketax.set_title(" Clustering by K-Means model")
plt.show()
```

C:\Users\yasha\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:1334: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment v ariable OMP_NUM_THREADS=9.

warnings.warn(

Clustering by K-Means model

