מגיש: ערד קוצר ת"ז: 311226203 מסלול: מדמ"ח יזמות <u>חלק 1</u> שאלה 1 ^ Mobileye_risk_TelA....csv •

שאלה 2

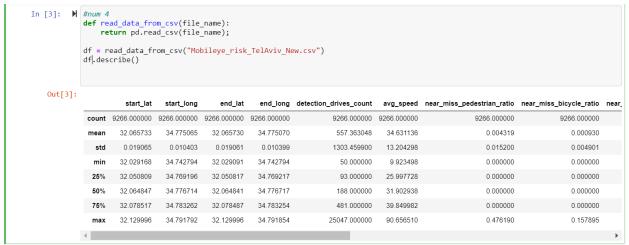
Sklearn היא ספריית למידת מכונה אשר הוקמה בשנת 2007 על ידי דייוויד קורנפאו כפרויקט קיץ של גוגל ומאז תרמו אליה מתנדבים רבים כקוד פתוח. כיום הוא מתוחזק על ידי צוות מתנדבים כקוד פתוח. הספרייה עובדת עבור שפת התכנות Python 3. והיא כוללת אלגוריתמים שונים של סיווג, רגרסיה ואשכולות, כולל יערות אקראיים, הגדלת שיפוע DBSCAN, ו, k-means -והיא נועדה לשתף פעולה עם הספריות המדעיות של Python NumPy ו - SciPy.

```
In [1]: H #num 2 import sklearn
```

שאלה 3

```
In [2]: | #num 3

import pandas as pd
import matplotlib.pyplot as plt
import geopy
```



<u>חלק 2:</u>

5 שאלה

```
address1 = (row['start_lat'], row['start_long'])
address2 = (row['end_lat'], row['end_long'])
return (geodesic(address1, address2).m) #in meters
In [16]: M df['section_length'] = df.apply(lambda x: distance(x), axis=1)
In [17]: ► df.head()
    Out[17]:
               io avg_pedestrian_on_road_volume
                                                  avg_bicycle_on_road_volume braking_count cornering_count harsh_braking_ratio
                                                                                                                                  harsh_cornering_ratio
               .0
                                        0.033557
                                                                     0.033557
               .0
                                        0.071429
                                                                     0.015873
                                                                                          0
                                                                                                           0
                                                                                                                              0.0
                                                                                                                                                             24.101012
                                                                                                                                               0.000000
               .0
                                        0.010327
                                                                     0.000000
                                                                                          60
                                                                                                          389
                                                                                                                              0.0
                                                                                                                                               0.002571
                                                                                                                                                              9.673643
               .0
                                        0.013605
                                                                     0.003401
                                                                                          29
                                                                                                          201
                                                                                                                              0.0
                                                                                                                                               0.003253
                                                                                                                                                             10.088931
               .0
                                                                                          11
                                                                                                                              0.0
                                        0.026667
                                                                     0.000000
                                                                                                                                               0.020833
                                                                                                                                                             71.514146
```

שאלה 6

```
In [19]:

    df.section_length.describe()

   Out[19]: count
                       9266.000000
                         42.657029
              mean
                         44.375733
              std
              min
                          0.671996
              25%
                         11.204812
              50%
                         26.910230
              75%
                         62.054658
                        449.794289
              max
              Name: section_length, dtype: float64
```

הסבר – ניתן לראות שהקטע הארוך ביותר הוא ברוקך 449.794 מטר, בעוד שהכי קטן הוא באורך 0.6719 מטר. כמו כן המרחק הממוצע הוא 42.657 מטר למקטע.

```
In [10]: ▶ def time(row):
                 avg_time_hour = ((row['section_length'])/(row['avg_speed'])*1000)
                 return (avg_time_hour)
In [11]: M df['avg_time'] = df.apply(lambda x: time(x), axis=1)
In [12]: M df.avg_time.describe()
   Out[12]: count
                       9266.000000
                       1448.854185
             mean
             std
                       1687.832432
                         24.307915
             min
                        322.428115
             50%
                        736.008088
             75%
                       2093.566885
                      17109.670249
             Name: avg_time, dtype: float64
In [13]: M df['avg_time_sec'] = df.apply(lambda x: (time(x)*3600), axis=1)
```

8 שאלה

```
In [19]: M df1 = df.drop(['start_lat', 'start_long', 'end_lat', 'end_long'], axis =1)
```

												O -L			
n [17]: H	df1.de	scribe()													
Out[17]:		d-tti d-t					avg_pedestrian_on_road_volume	on road volume	braking count	cornering count	hareh braking ratio	harsh cornering ratio	section length	ava time	avg_time_sec
	_														
	count	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9266.000000	9.266000e+03
	mean	557.363048	34.631136	0.004319	0.000930	0.008162	0.050953	0.026391	66.247788	276.613857	0.000002	0.015083	42.657029	1448.854185	5.215875e+06
	std	1303.459900	13.204298	0.015200	0.004901	0.017387	0.074877	0.043735	138.458602	614.604640	0.000135	0.037954	44.375733	1687.832432	6.076197e+06
	min	50.000000	9.923498	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.671996	24.307915	8.750849e+04
	25%	93.000000	25.997728	0.000000	0.000000	0.000000	0.001731	0.000000	5.000000	38.000000	0.000000	0.000000	11.204812	322.428115	1.160741e+06
	50%	188.000000	31.902938	0.000000	0.000000	0.000000	0.020090	0.010441	22.000000	87.000000	0.000000	0.001331	26.910230	736.008088	2.649629e+06
	75%	481.000000	39.849982	0.000000	0.000000	0.009326	0.067423	0.033782	68.000000	265.000000	0.000000	0.011058	62.054658	2093.566885	7.536841e+06
	max	25047.000000	90.656510	0.476190	0.157895	0.235294	0.780000	0.716418	2715.000000	7130.000000	0.012346	0.636364	449.794289	17109.670249	6.159481e+07
	4						1								F

העמדות הסבר DATA נרמול. לעבור שלנו צריכים המשתנים לא נרמול שצריכים :הם לעבור המשתנים 'detection_drives_count', 'avg_speed', 'avg_pedestrian_on_road_volume', 'avg_ bicycle_on_road_volume', 'braking_count', 'cornering_count', 'section_length', ''avg time

צריכות לעבור נרמול וזאת משום שהערכים שלהם לא בין 0- 1 כפי שאנחנו רוצים אותם. המשתנים שלא צריך לנרמל הן :

0.000220

0.011418

0.031417

0.111715 0

'near_miss_pedestrian_ratio', 'near_miss_bicycle_ratio', 'near_miss_vehicle_ratio', 'harsh_braking_ratio', 'harsh_cornering_ratio'

שאלה 10

```
In [56]:  H from sklearn.preprocessing import MinMaxScaler
                                         column_to_normalize = ['detection_drives_count', 'avg_speed', 'avg_pedestrian_on_road_volume', 'avg_bicycle_on_road_volume',
                                         scaler = MinMaxScaler()
                                         x_scaled = scaler.fit_transform(df1[column_to_normalize])
                                         X_normalize = pd.DataFrame(x_scaled, columns=column_to_normalize, index=df1.index)
                                         X_normalize1 = pd.concat([X_normalize, df2], axis=1)
                                         pd.DataFrame(X_normalize)
           Out[56]:
                                                           detection_drives_count avg_speed avg_pedestrian_on_road_volume avg_bicycle_on_road_volume braking_count cornering_count section_length avg_speed a
                                                  0
                                                                                             0.009921
                                                                                                                         0.249457
                                                                                                                                                                                                             0.043022
                                                                                                                                                                                                                                                                                     0.046840
                                                                                                                                                                                                                                                                                                                            0.007735
                                                                                                                                                                                                                                                                                                                                                                      0.025666
                                                                                                                                                                                                                                                                                                                                                                                                              0.020792 0
                                                                                             0.003040
                                                                                                                           0.201741
                                                                                                                                                                                                              0.091575
                                                                                                                                                                                                                                                                                       0.022156
                                                                                                                                                                                                                                                                                                                            0.000000
                                                                                                                                                                                                                                                                                                                                                                       0.000000
                                                                                                                                                                                                                                                                                                                                                                                                              0.052166 0
                                                                                                                                                                                                                                                                                                                            0.022099
                                                                                             0.021243
                                                                                                                          0.279347
                                                                                                                                                                                                            0.013240
                                                                                                                                                                                                                                                                                                                                                                       0.054558
                                                                                                                                                                                                                                                                                                                                                                                                              0.020043 0
                                                                                             0.009761
                                                                                                                                                                                                             0.017443
                                                                                                                                                                                                                                                                                                                            0.010681
                                                                                             0.001000
                                                                                                                          0.259325
                                                                                                                                                                                                            0.034188
                                                                                                                                                                                                                                                                                      0.000000
                                                                                                                                                                                                                                                                                                                            0.004052
                                                                                                                                                                                                                                                                                                                                                                       0.006732
                                                                                                                                                                                                                                                                                                                                                                                                              0.157735 0
                                           9261
                                                                                             0.002240
                                                                                                                         0.230451
                                                                                                                                                                                                            0.278181
                                                                                                                                                                                                                                                                                      0.276533
                                                                                                                                                                                                                                                                                                                            0.000000
                                                                                                                                                                                                                                                                                                                                                                       0.000000
                                                                                                                                                                                                                                                                                                                                                                                                              0.137091 0
                                           9262
                                                                                             0.000440
                                                                                                                           0.090072
                                                                                                                                                                                                             0.210172
                                                                                                                                                                                                                                                                                       0.114413
                                                                                                                                                                                                                                                                                                                            0.000000
                                                                                                                                                                                                                                                                                                                                                                       0.000000
                                                                                                                                                                                                                                                                                                                                                                                                              0.117271 0
                                                                                                                                                                                                                                                                                                                                                                                                              0.014563 0
                                           9263
                                                                                             0.000560
                                                                                                                         0.240742
                                                                                                                                                                                                            0.060096
                                                                                                                                                                                                                                                                                     0.087240
                                                                                                                                                                                                                                                                                                                            0.000737
                                                                                                                                                                                                                                                                                                                                                                       0.003647
```

0.000000

9264

0.252190

0.763426

<u>חלק 3:</u>

שאלה 11

```
In [122]: ▶ #num 11
                import numpy as np
from sklearn.cluster import KMeans
                from sklearn.preprocessing import LabelEncoder
                from sklearn.preprocessing import MinMaxScaler
import seaborn as sns
                import matplotlib.pyplot as plt
In [173]: ▶ #num 11
                def create_kmeans_classifier(K):
    if ((type(K) == int or float) and (float(K) == int(K)) and (int(K) > 0)):
                          return KMEANS(int(K))
                     else:
                           print("Invalid input!")
                                                                                                                                              שאלה 13 + 12
 In [22]: ► #num 12 +13
                k = range(1, 15)
                error = []
n_error = []
                def KMEANS(K):
    kmeans = KMeans(n_clusters= K)
    kmeans.fit(df1)
    return kmeans.inertia_
                for i in k:
                    error.append(create_kmeans_classifier(i))
In [23]: H #num 12 +13
                def KMEANS(K):
                     kmeans = KMeans(n_clusters= K)
kmeans.fit(X_normalize)
return kmeans.inertia_
                for i in k:
                    n_error.append(create_kmeans_classifier(i))
 שאלה 14
 def KMEANS(K):
    kmeans = KMeans(n_clusters= K)
    kmeans.fit(df1)
                     return kmeans.cluster_centers_
                     Silhouette.append(create_kmeans_classifier(i))
 In [30]: ▶ #num 14
                def KMEANS(K):
    kmeans = KMeans(n_clusters= K)
    kmeans.fit(X_normalize)
```

return kmeans.cluster_centers_

n_Silhouette.append(create_kmeans_classifier(i))

שאלה 15

```
In [27]: | #num 15
| print( "", len(Silhouette), "\n" , len(n_Silhouette), "\n" , len(error), "\n" , len(n_error))

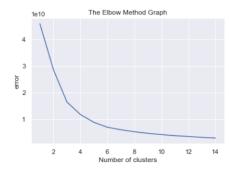
14
14
14
14
```

<u>חלק 4:</u>

שאלה 16

```
In [66]: 
plt.title('The Elbow Method Graph')
plt.xlabel('Number of clusters')
plt.ylabel('error')
plt.plot(k, error)
```

Out[66]: [<matplotlib.lines.Line2D at 0x241806474c0>]



```
In [97]: )
#num 17
from sklearn.detomposition import PCA
#Load Data
data = df1
pca = PCA(2)

#Transform the data
df = pca.fit_transform(data)

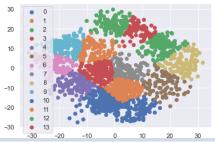
#Import KMeans module
from sklearn.cluster import KMeans

#Initialize the class object
kmeans = KMeans(n_clusters = 14)

#predict the labels of clusters.
label = kmeans.fit_predict(df)

#Getting unique labels
u_labels = np.unique(label)

#plotting the results:
for i in u_labels:
    plt.scatter(df[label == i , 0] , df[label == i , 1] , label = i)
plt.legend()
plt.show()
```



שאלה 18

```
In [98]: ▶ #num 18
             from sklearn.datasets import load digits
             from sklearn.decomposition import PCA
             data = X_normalize
             pca = PCA(2)
             #Transform the data
             df = pca.fit_transform(data)
             #Import KMeans module
             from sklearn.cluster import KMeans
             #Initialize the class object
             kmeans = KMeans(n_clusters= 14)
             #predict the labels of clusters.
             label = kmeans.fit_predict(df)
             #Getting unique labels
             u_labels = np.unique(label)
             #plotting the results:
             for i in u labels:
                 plt.scatter(df[label == i , 0] , df[label == i , 1] , label = i)
             plt.legend()
             plt.show()
               1.0
               0.8
               0.6
               0.4
               0.2
               0.0
               -0.2
                                                            11
12
               -0.4
```

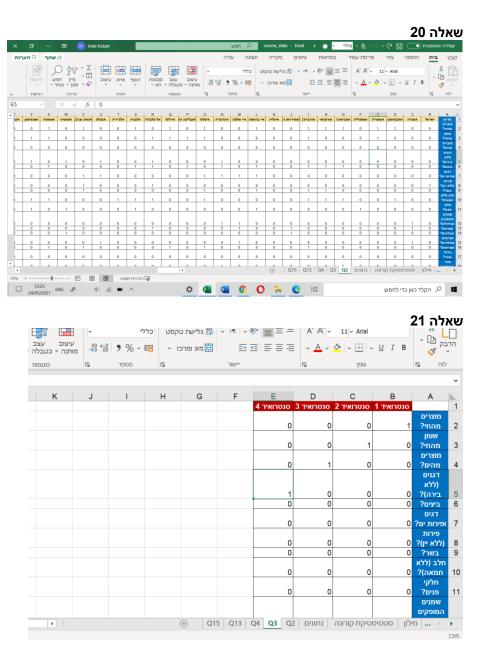
הסבר:

ניתן לראות שכאשר הנתונים מנורמלים אז השוני בין הקבוצות הופך להיות יותר בצורה דו ממדית – ולכן בסכימה של ערכים לכול טווח ערכים ניתן לקטלג אותם יותר בקלות, דבר זה יכול להקל עלינו כאשר הנתונים מאוד קרובים של ערכים לכול טווח ערכים ניתן לקטלג אותם יותר בקלות, דבר זה יכול להקל עלינו כאוריד את וההשוואה בין הקבוצות לא ברורה לנו חשיבות ערכי הנתונים. כלומר בעזרת נרמול אנחנו יכולים גם להוריד את סיבוכיות החישוב בסדר גודל אחד מ $o(n^3) \Rightarrow o(n^2)$ וגם לנו יהיה יותר קל להשוות את זה בעיין כאשר מורידים לנו ממד להשוות.

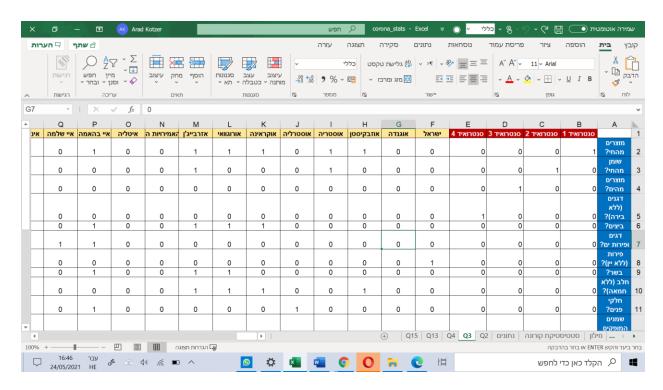
: Excel – 5 חלק

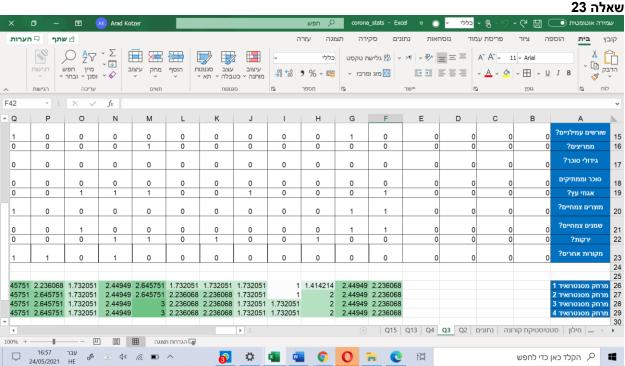
שאלה 19

לכול מדינה נגדיר ווקטור שמורכב ממספרי האחדים וכול אחד "מחזק" את גדול הווקטור ואת הכיוון שלו. לאחר מ כן ניתן לראות כול ווקטור כמזה חח"ע למדינה ולמדוד אותה לפי ערכיו.

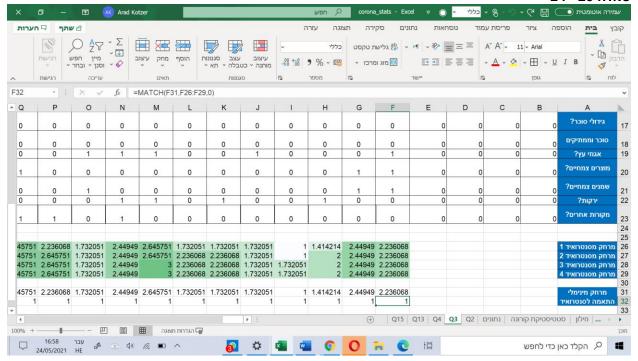


שאלה 22





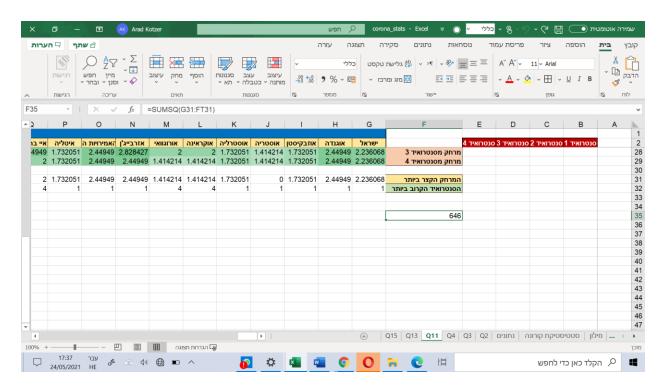
שאלה 25 +24

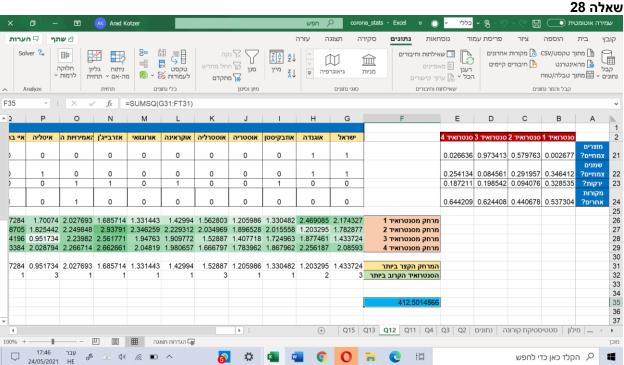


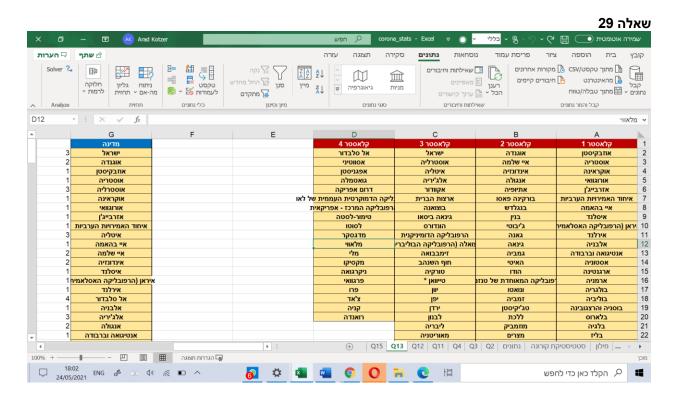
שאלה 26

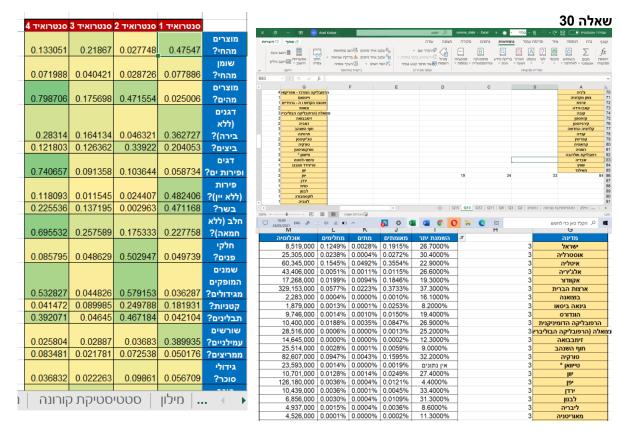
פונקציית המטרה שלנו היא התא של סכום המרחקים = SUM(F31:FS31) שאותה אנחנו רוצים להביא למינימום, כך שפשוט נזיז את כול המיקומים של הסנטרואידים.

האילוצים שנקבעה הם - שמיקומי המרחקים יהיה קטן שווה ל 1. וגם שמבחינת איטרציות שיהיה קטן ממיליון כדי י שהאקסל לא יתקע לנו.









א - ניתן לראות שבקלאסטר 1 - יש 84 מדינות. בקלאסטר 2 - יש 33 מדינות. בקלאסטר 3 - יש 34 מדינות. בקלאסטר 1 - יש 19 מדינות

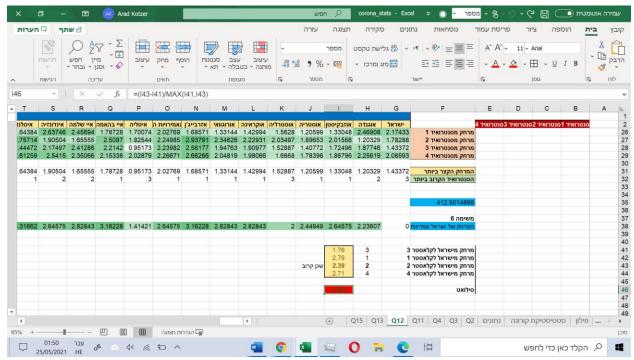
ב- ניתן לראות שבקלאסטר 1 - עיקר השומן מגיע מהחיי ביחד עם ששמנים צמחיים– כמובן שביחס לקבוצות האחרות, כמוכן ניתן לראות שהרבה ממדינות אלה הן מדינות דרום אמריקאיות או ברית המועצות לשעבר, מדינות של רות, כמוכן ניתן לראות שהרבה ממדינות אלה הן מדינות דרום אמריקאיות או ברית המועצות עולם שלישי, אך גם רוב נמנות כמדינות מפותחות במדינות מפותחות כמו המדינות הסקנדינביות צרפת ושו לא מאופיינות כמדינות מצריכת שומן יחסית גבוהה מהחיי.

ניתן לראות שבקלאסטר 2 - עיקר השומן מגיע מאגוזים, תבלינים, שומנים שמגיעים מגידולים, חלקי פנים ומוצרי ם מהים – כמובן שביחס לקבוצות האחרות. ניתן לראות שכול המדינות באופן מוחלט מוגדרות כמדינות עולם שלי שי ולכן מאופינות ע"י בישול מסורתי, יחסית מגוון וכול זאת לצד עוני קשה. כמוכן ניתן לראות שהיה אפשר להסיק בזכות נתוני התמותה והחולי הנמוכים שמזון ושומן מסוג זה יכולים להבטיח עמידות בפני המגפה – אך חשוב לזכ ור שבמדינות אלא תוחלת החיים היא נמוכה בעד 50% ממדינות המערב (ולכן אחוז זקנים נמוך), לצד מערכת ברי אות כושלת וחוסר יכולת לאסוף נתונים בצורה אמינה.

ניתן לראות שבקלאסטר 3 - עיקר השומן ממקורות צמחיים, ומסוכרים ו"ממקורות אחרים" בעוד שאר הערכים יח סית נמוכים – כמובן שביחס לקבוצות האחרות, ניתן לראות שהרבה מדינות מערביות נכללות בקרב קבוצה זאת ד בר שמתאפיין באוכל תעשייתי עשיר בסוכר לצד מוצרים צמחיים. כמוכן ניתן לראות שאחוז השמנת היתר הינו הג בוהה ביותר בקבוצה זאת, לצד תמותה וחולי מקורנה. אבל שוב יש לזכור שלצד נתוני התזונה, תוחלת החיים במ דינות אלה היא מהגבוהות בעולם – דבר המשפיע על אחוז הזקנים ועל היכולת לעקוב אחרי המצב הרפואי של הא וכלוסייה. ולכן ישפיע גם על הנתונים.

ניתן לראות שבקלאסטר 4 - עיקר השומן מגיע ממוצרים מהים, מוצרי חלב, שמנים מגידולים ותבלינים – כמובן ש ביחס לקבוצות האחרות. דבר הדומה קצת למדינות קלאסטר 2 שבהם המדינות מתפתחות לצד זאת שמדינות 4 הן בעיקר מדינות מתפתחות בהם רוב האוכלוסיה שוכנת לצד החופים ואגמים לכן מרבית התזונה בהם היא מאכלי ים. בדומה למדינות 2 מדובר במדינות עם מערכות בריאות כושלות לצד תוחלת חיים יחסית נמוכה. וכמובן לצד יכולת לאסוף נתונים בצורה אמינה ולכן יש לבחון את הנתונים במדינות אלה בערבון מוגבל.

<u>חלק 3</u> שאלה 31



ניתן לראות שהקלאסטר הכי קרוב הוא קלאסטר 2. הסילוואט הוא 0.263 והמדינות שהכי קרובות לנו במרחק הן הרפובליקה הדומיניקנית, גינאה בוסאו ולבנון – לא בטוח לגבי הקריבים אבל לדעתי ניתן להסיק שגם כנראה בלבנון אוהבים חומוס. ☺