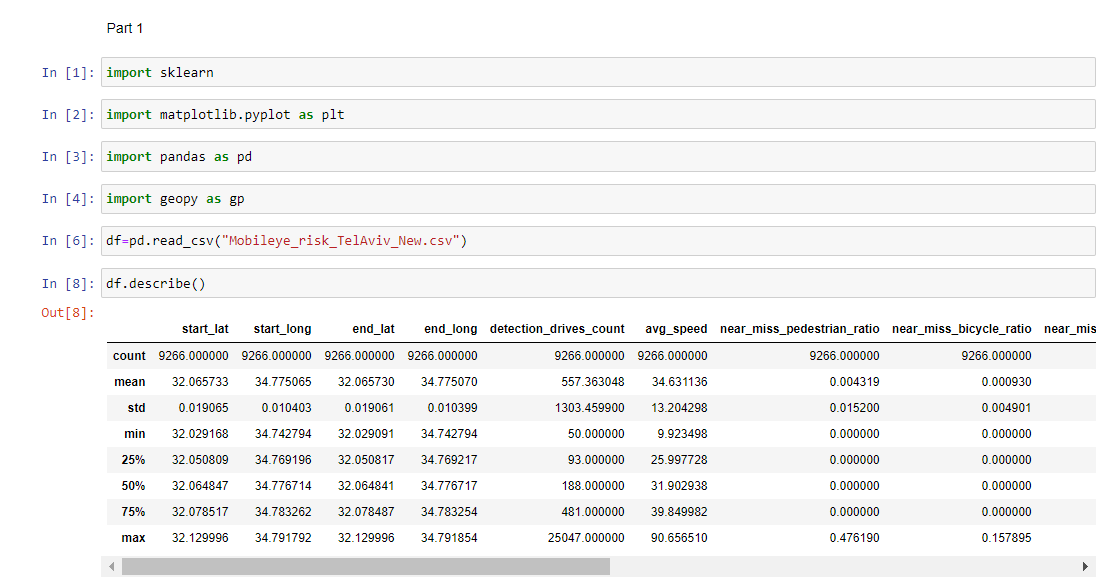
**Data Science EX2- Nadav Vilder 312173776**

**Part 1**

1. הורדתי

Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

**  
3+4**

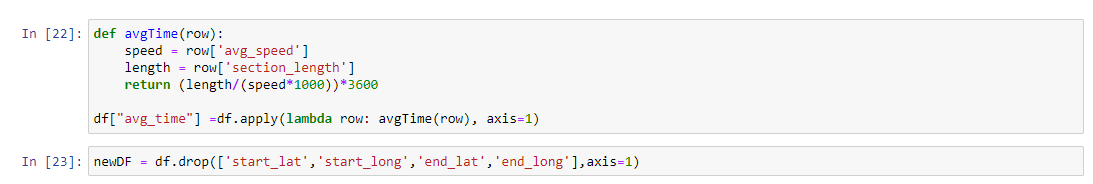
**5.**

**Part 2**

**6**. המרחק הממוצע הוא :42.65 מטר

המרחק הארוך ביותר הוא 449.79 מטר..

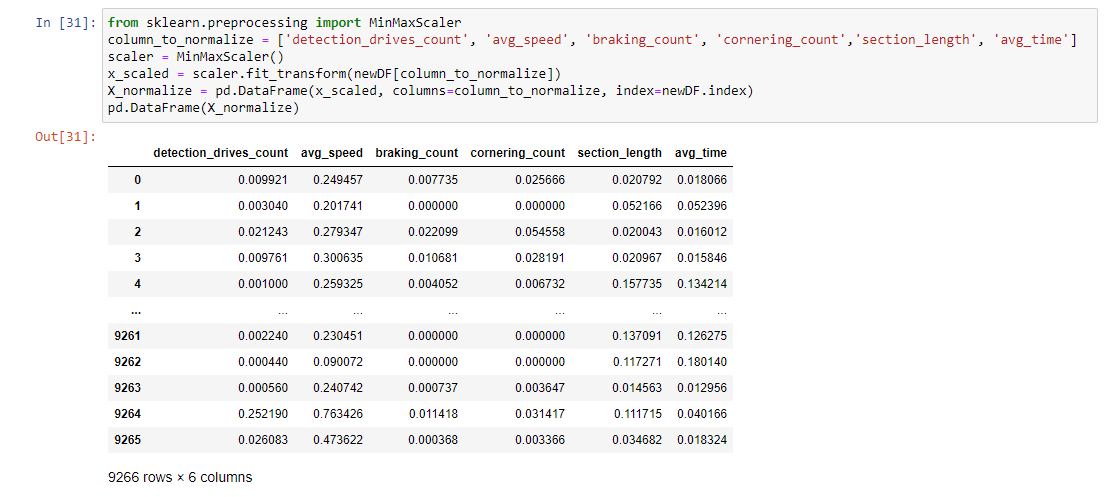
המרחק הקצר ביותר הוא 0.67 מטר.

**7+8**

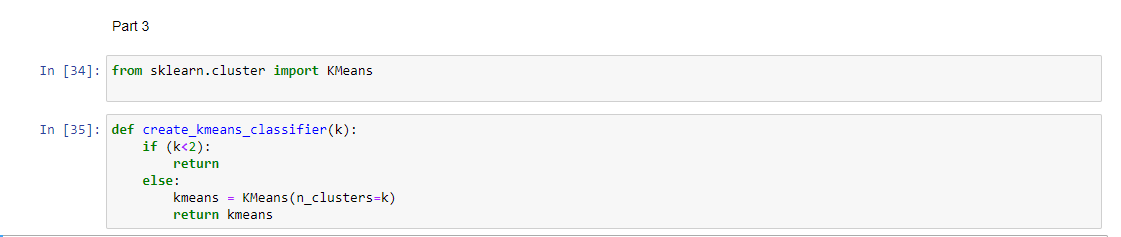
**9**. לא כל המשתנים בדאטה פריים צריכים נרמול. המשתנים שלא צריכים הם:

| near\_miss\_pedestrian\_ratio | near\_miss\_bicycle\_ratio | near\_miss\_vehicle\_ratio | avg\_pedestrian\_on\_road\_volume | avg\_bicycle\_on\_road\_volume |
| --- | --- | --- | --- | --- |

**10.**



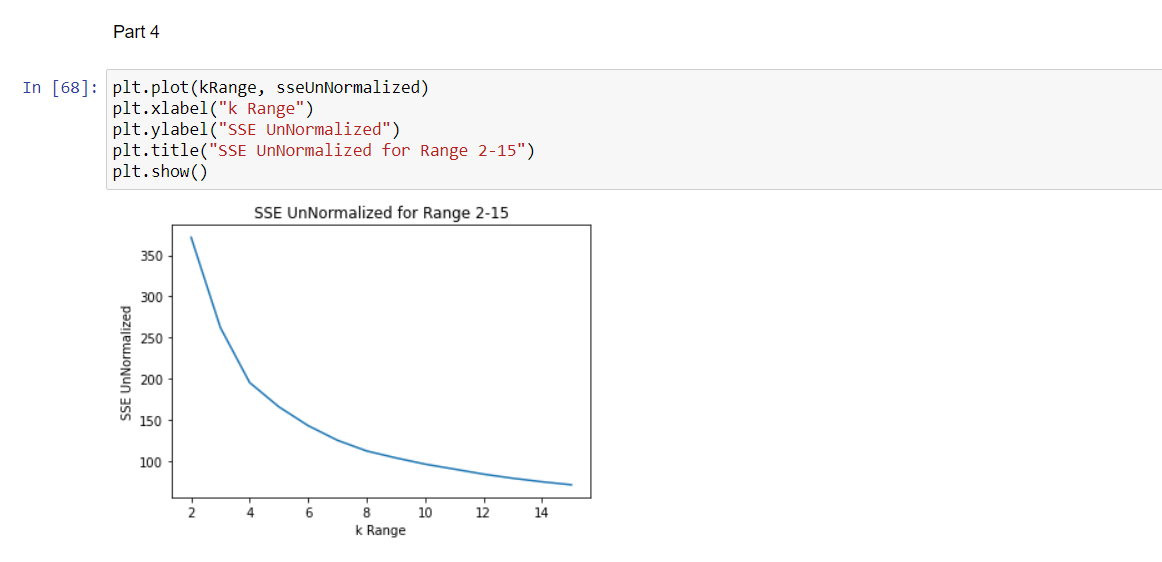
**Part 3**

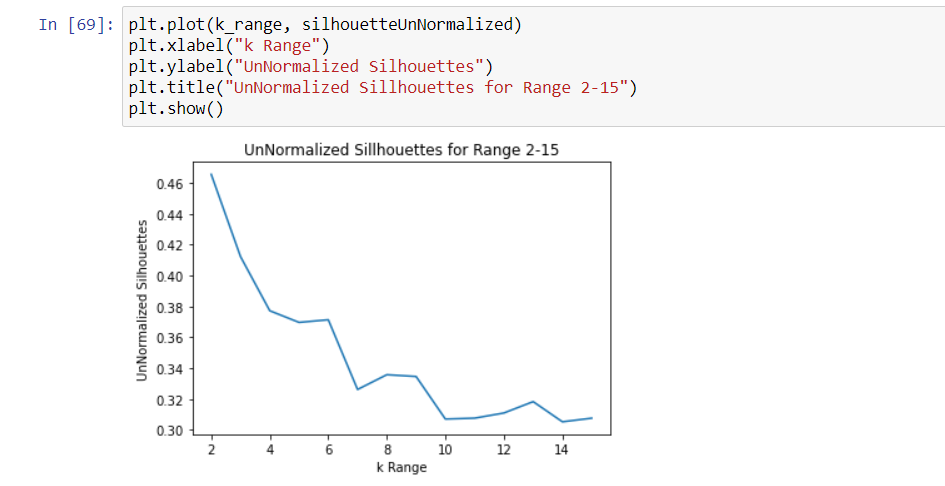
**11.**

**12-15**

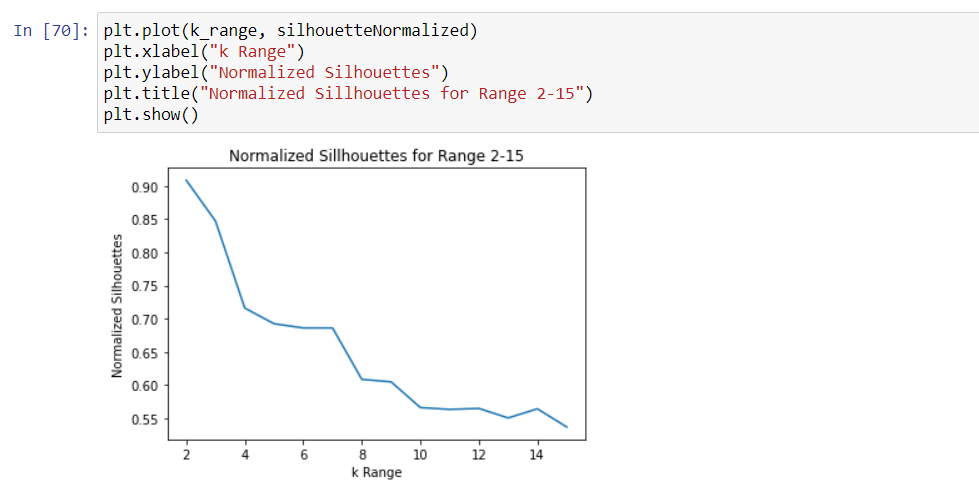
**Part 4**

**16**

**17.**

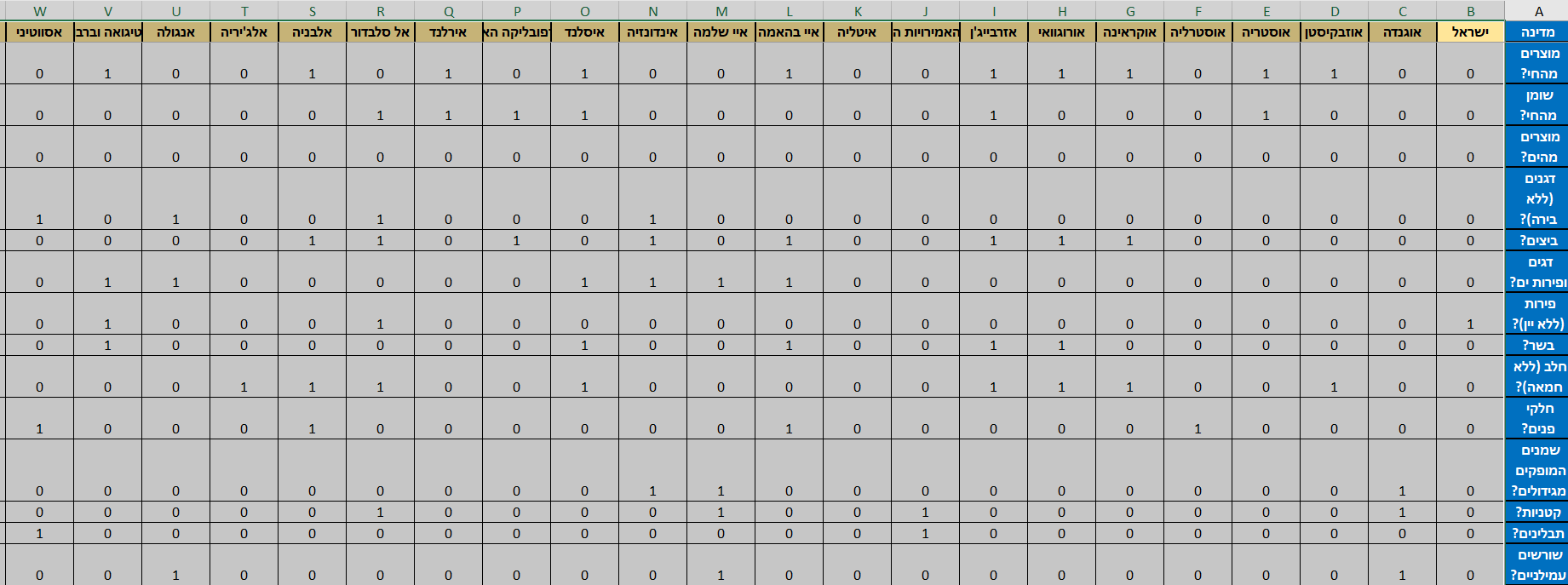


**18.**

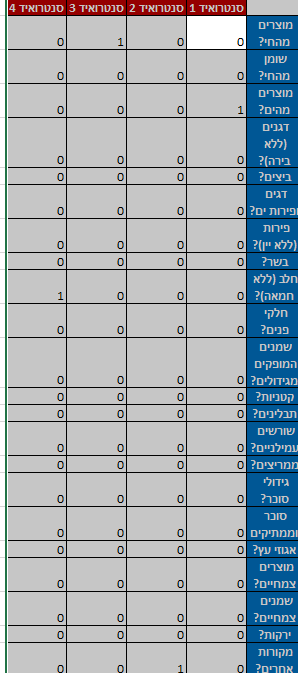


ניתן לראות שלנתונים המנורמלים יש ציוני סילווט גבוהים יותר וטובים יותר.

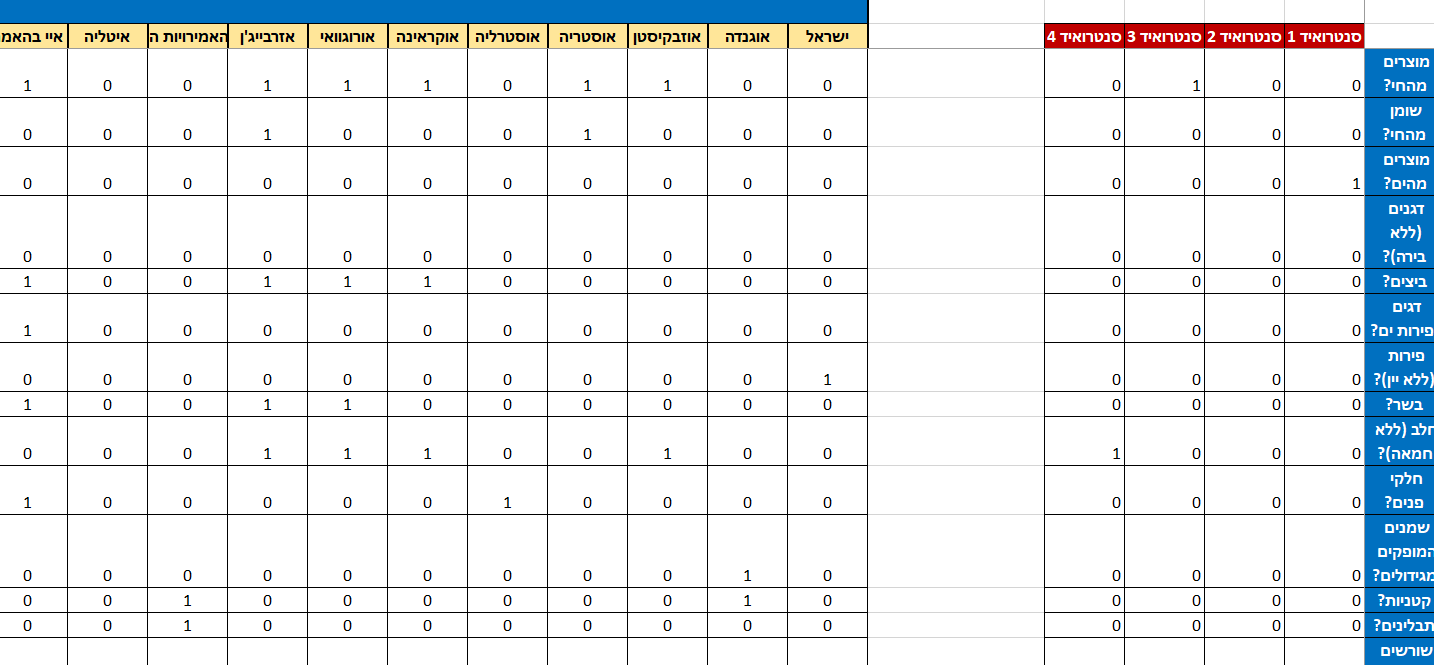
**19.** כל מדינה היא וקטור באורך 22 כאשר לכל ערך יש פרמטר 0 או 1 אשר מבטא אם מקור השומן של אותו ערך נצרך ברמה גבוהה עבור אותה מדינה (1) או לא (0).

**20.** עשיתי פייסט טרנספוז.

**21. הצבתי שרירותית אפסים ואחדות עבור כל סנטרואיד.**

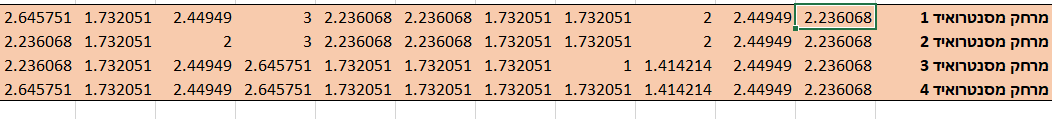


**22.**

 **הדבקתי אחד ליד השני**

**23. חישבתי לפי**



**24. חישבתי לפי**



**25.חישבתי לפי**





**26**

. **המטרה:** להביא למינימום את סך המרחקים הריבועיים של המדינות מהסנטרואידים שלהם

**משתני ההחלטה:** ערכי הסנטרואידים ( 22 ערכים \* 4 סנטרואידים = 88 משתני החלטה בסך הכל)

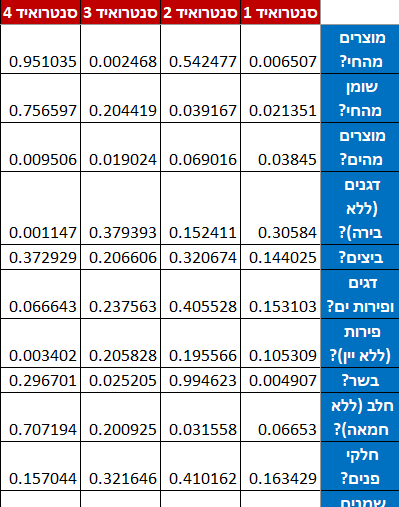
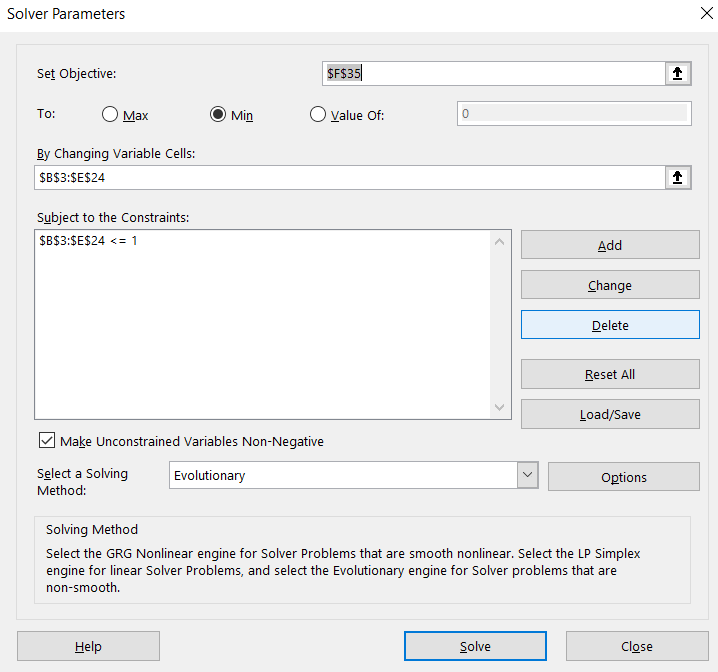
**אילוצים:** ערכי הסנטרואיד נעים בטווח שבין 0 ל- 1 אנו יודעים זאת מראש, מאחר שבהגדרה כל

סנטרואיד מורכב מהממוצע של נקודות נתונים הכוללות ערכים אלו

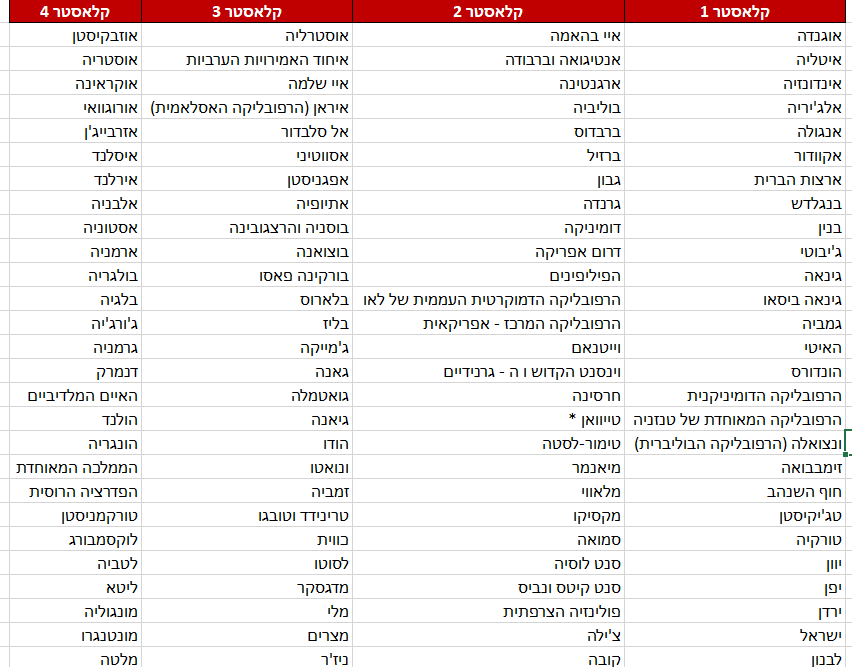
**27. חישבתי לפי**



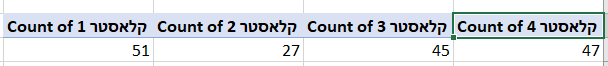
**28.**





**29.** 

**30**.

**משימה 5 : ויזואליזציה**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **סנטרואיד 1** | **סנטרואיד 2** | **סנטרואיד 3** | **סנטרואיד 4** |
| **מוצרים מהחי?** | 0.006507 | 0.542477 | 0.002468 | 0.951035 |
| **שומן מהחי?** | 0.021351 | 0.039167 | 0.204419 | 0.756597 |
| **מוצרים מהים?** | 0.03845 | 0.069016 | 0.019024 | 0.009506 |
| **דגנים (ללא בירה)?** | 0.30584 | 0.152411 | 0.379393 | 0.001147 |
| **ביצים?** | 0.144025 | 0.320674 | 0.206606 | 0.372929 |
| **דגים ופירות ים?** | 0.153103 | 0.405528 | 0.237563 | 0.066643 |
| **פירות (ללא יין)?** | 0.105309 | 0.195566 | 0.205828 | 0.003402 |
| **בשר?** | 0.004907 | 0.994623 | 0.025205 | 0.296701 |
| **חלב (ללא חמאה)?** | 0.06653 | 0.031558 | 0.200925 | 0.707194 |
| **חלקי פנים?** | 0.163429 | 0.410162 | 0.321646 | 0.157044 |
| **שמנים המופקים מגידולים?** | 0.138521 | 0.117063 | 0.288492 | 0.009568 |
| **קטניות?** | 0.261337 | 0.201554 | 0.340492 | 0.016181 |
| **תבלינים?** | 0.123806 | 0.182575 | 0.308553 | 0.054241 |
| **שורשים עמילניים?** | 0.259652 | 0.184421 | 0.17616 | 0.0125 |
| **ממריצים?** | 0.059445 | 0.13984 | 0.081518 | 0.513645 |
| **גידולי סוכר?** | 0.016529 | 0.11012 | 0.068459 | 0.021264 |
| **סוכר וממתיקים** | 0.054937 | 0.157597 | 0.025534 | 0.014692 |
| **אגוזי עץ?** | 0.23516 | 0.077708 | 0.134973 | 0.245332 |
| **מוצרים צמחיים?** | 0.729921 | 0.00709 | 0.314126 | 0.003748 |
| **שמנים צמחיים?** | 0.995152 | 0.038903 | 0.014946 | 0.004014 |
| **ירקות?** | 0.232351 | 0.144293 | 0.218475 | 0.151371 |
| **מקורות אחרים?** | 0.076676 | 0.414104 | 0.299859 | 0.145495 |

**בקלאסטר 1 מובהק שמקור השומן הוא צמחי  
בקלאסטר 2 מבשר  
ומקלאסטר 4 מהחי**