Members

1. Mahmoud Adel Kamal Ismail (2018170362)
2. Mahmoud Alaa Morsy (2018170363)
3. Mahmoud Mohamed Saeed (2018170365)

Step 1

1. Column names : "zcta" , "sex"

"meanage" , "meaneducation" "meanemployment" , "meanhouseholdincome"

Analysis of the table :

|  |  |  |  |
| --- | --- | --- | --- |
| zcta  meanemployment | sex  meanhouseholdincome | meanage | meaneducation |
| Min. : 601 | Length:64076 | Min. : 0.00 | Min. : 0.00 |
| Min. :0.000 | Min. : 0 |  |  |
| 1st Qu.:27305 | Class :character | 1st Qu.: 36.65 | 1st Qu.:11.91 |
| 1st Qu.:1.542 | 1st Qu.: 37642 |  |  |
| Median :49909 | Mode :character | Median : 39.30 | Median :12.46 |
| Median :1.813 | Median : 44163 |  |  |
| Mean :49801 |  | Mean : 39.68 | Mean :12.53 |
| Mean :1.787 | Mean : 48245 |  |  |
| 3rd Qu.:72007 |  | 3rd Qu.: 42.28 | 3rd Qu.:13.11 |
| 3rd Qu.:2.077 | 3rd Qu.: 54373 |  |  |
| Max. :99950 | Max. | :137.08 Max. | :19.00 Max. :3.000 |
| Max. :250000 |  |  |  |

1. Number of rows : **64076**.
2. There are no duplicates in the table.

Step2



zipCode sex meanage meaneducation

Min. : 601 Length:64076 Min. : 0.00 Min. : 0.00

1st Qu.:27305 Class :character 1st Qu.: 36.65 1st Qu.:11.91

Median :49909 Mode :character Median : 39.30 Median :12.46

Mean :49801 Mean : 39.68 Mean :12.53

3rd Qu.:72007 3rd Qu.: 42.28 3rd Qu.:13.11

Max. :99950 Max. :137.08 Max. :19.00

meanemployment income

Min. :0.000 Min. : 0

1st Qu.:1.542 1st Qu.: 37642

Median :1.813 Median : 44163

Mean :1.787 Mean : 48245

3rd Qu.:2.077 3rd Qu.: 54373

Max. :3.000 Max. :250000

Mean average Income = **48245**

Median average Income = **44163**

1. There are two outliers at 0 and 250000

Text

Description automatically generated

1. New Mean average Income = **48465**

Step 3

Chart

Description automatically generatedChart

Description automatically generated

(3)

1. U should visualize your data in different ways.
2. Visualization makes u understand your data well.

# Advanced Analytics/Methods (K-means)

1.

Table

Description automatically generated with medium confidence

2.

Chart, scatter chart

Description automatically generated

3. Chose K = 3 using elbow method

Chart

Description automatically generated with medium confidence

Chart, scatter chart

Description automatically generated

4. The clustering is not only vertically splitting of the data as it was before. This is because the ‘mean household income’ had larger scale than the ‘mean electricity usage’ and the k-mean was biased towards the feature that has larger magnitude.

Chart, scatter chart

Description automatically generated

5. Yes, I chose K = 5 using elbow method

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated

6. Yes, I noticed outliers in the data.

After removing them I chose K = 4 using the elbow method.

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated