NAME: NIAHAO DONG
NUMBER: 201476606

All the programs are in the file "CA2_201476606.py". In each block of code, the number of problems is indicated. About the implementation of the code, details. Has been written in the code comments.
Original development environment: Python 3, Jupyter Notebook

If you use other IDE, you can open them one by one, You can see Q1 ~ Q6, six .py files.

Q1
In the Main function, which is 75 lines of code. Please modify the data set read path. It's the **fruit** data set.
Line 86, modify the data set read path. It's the **animals** data set.
Line 95, modify the data set read path. It's the **countries** data set.
Line 104, modify the data set read path. It's the **veggies** data set.

CA2 Q1 does not have an explicit k value. Default here is 4. Run it and you will see the following result.

In K-means the centroids are :
[[xxx xxx···xxx]
[xxx xxx···xxx]
[xxx xxx···xxx]
[xxx xxx···xxx]]

Q2
In block2. In the Main function, which is 48 lines of code. Please modify the data set read path. It's the **fruit** data set.
Line 57, modify the data set read path. It's the **animals** data set.
Line 66, modify the data set read path. It's the **countries** data set.
Line 75, modify the data set read path. It's the **veggies** data set.

CA2 Q2 also does not have an explicit k value. Default here is 4. Run it and you will see the following result.

In K-median the centroids are :
[[xxx xxx···xxx]
[xxx xxx···xxx]
[xxx xxx···xxx]
[xxx xxx···xxx]]

Q3

In block3
Line 111, modify the data set read path. It's the **animals** data set.
Line 120, modify the data set read path. It's the **animals** data set.
Line 129, modify the data set read path. It's the **countries** data set.
Line 138, modify the data set read path. It's the **veggies** data set.

When k is greater than 4, the result is too long, and the thumbnail version is displayed., Run it and you will see the following result.

When k is:
x
the centroids are :
[[xxx xxx…xxx]
[xxx xxx…xxx]
[xxx xxx…xxx]
[xxx xxx…xxx]]
B-CUBED precision is:
xxx
B-CUBED recall is:
xxx
B-CUBED F_score is:
xxx

At the **end of the output**, you will see the plot.

Q4
In block4
Line 106, modify the data set read path. It's the **animals** data set.
Line 115, modify the data set read path. It's the **animals** data set.
Line 124, modify the data set read path. It's the **countries** data set.
Line 133, modify the data set read path. It's the **veggies** data set.
The output form is the same as the third question.

Q5
In block5
Line 105, modify the data set read path. It's the **animals** data set.
Line 114, modify the data set read path. It's the **animals** data set.
Line 123, modify the data set read path. It's the **countries** data set.
Line 132, modify the data set read path. It's the **veggies** data set.
The output form is the same as the third question.

Q6
In block6
Line 105, modify the data set read path. It's the **animals** data set.

Line 114, modify the data set read path. It's the **animals** data set.
Line 123, modify the data set read path. It's the **countries** data set.
Line 132, modify the data set read path. It's the **veggies** data set.
The output form is the same as the third question.

Q7
Please see “CA2_Answe.pdf”file.