Assignment 8 – k-means implementation

Implement the k-means algorithm on the given dataset.
Some guidelines:

1. Perform an Exploratory Data Analysis ("EDA") on the dataset, including visualization and descriptive statistics. Interpret the results and describe them in your own words.

   Note that the CSV file is not in a standard format – part of the assignment is to properly read and process the data.
2. Implement the K-means algorithm from scratch, i.e. using only numpy and built-in python functionality. Refer to the presentation (session 17) for details; you may find the Wikipedia page useful.
3. Implement different distance functions and use them in the algorithm.
4. Try different methods to find the appropriate number of clusters.
5. Visualize the results and cluster centroids.
6. Describe the results in your own words.
7. Solve the same problem by K-means algorithm using Scikit-learn, and visualize the results. Compare your implementation to Scikit-learn's.