Parallel K-Means Clustering

By Cody Mangham, Carson Carpenter and Vivian Tran

K-Means Clustering

- Data Clustering Technique
- One of the more simpler and popular unsupervised machine learning algorithms
- What does it do?
  - Finds groups within a random dataset
  - The number of groups will be represented by the letter, K.

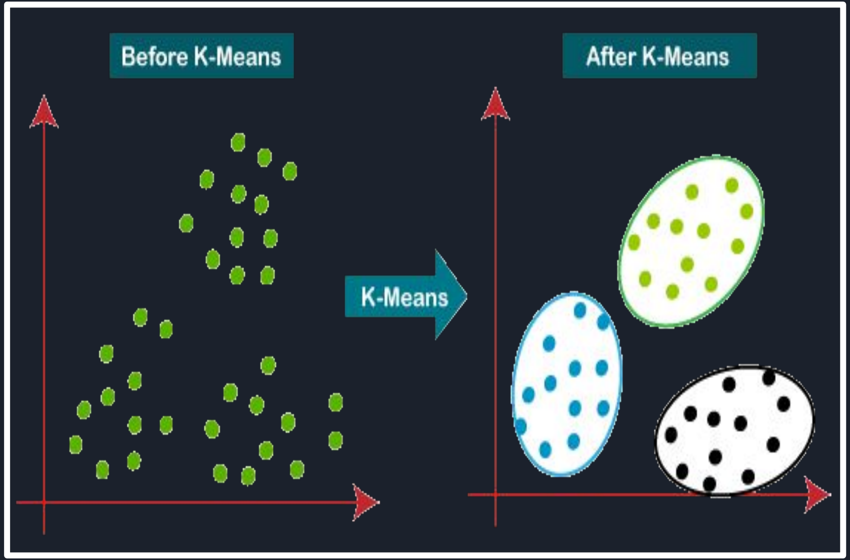

Figure 1: Before and After of K-Means Clustering

Source of Image:
https://www.analyticsvidhya.com/blog/2021/04/k-means-clustering-simplified-in-python/

Problem and Motivation

- Problem: To take a given dataset and separate these observations into a number of K clusters
- Motivation: The motivation to choose this topic is because we were able to implement random datasets in order to create clusters.

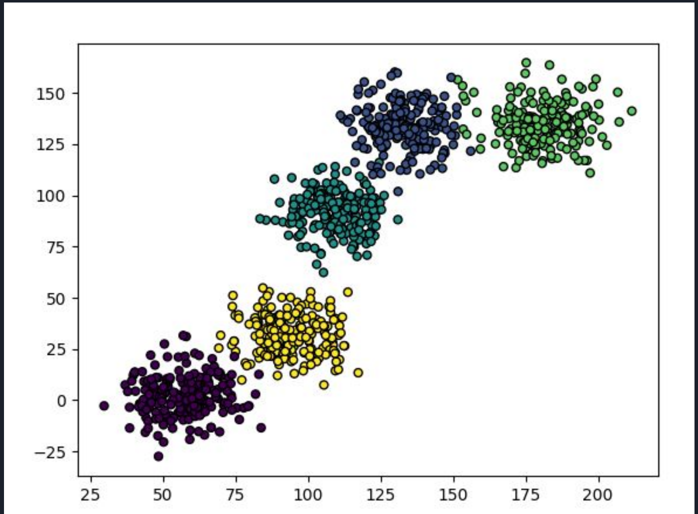

Figure 2: Example of a Clustering Scatter Plot from Project Program

Programming Language Chosen and Used

- C++
- MPI
- Python (matplotlib)

Commands used for the Program

MPI_Init

- Initializes the MPI execution environment

MPI_Comm_rank

- Determines the rank of the calling process in the communicator

MPI_Comm_size

- Determines the size of the group associated with a communicator

MPI_Bcast

- Broadcasts a message from the process with the rank “root” to all other processes of the communication

MPI_Allreduce

- Combines values from all processes and distributes the results back to all processes

MPI_Barrier

- Blocks processes until all other processes are finished

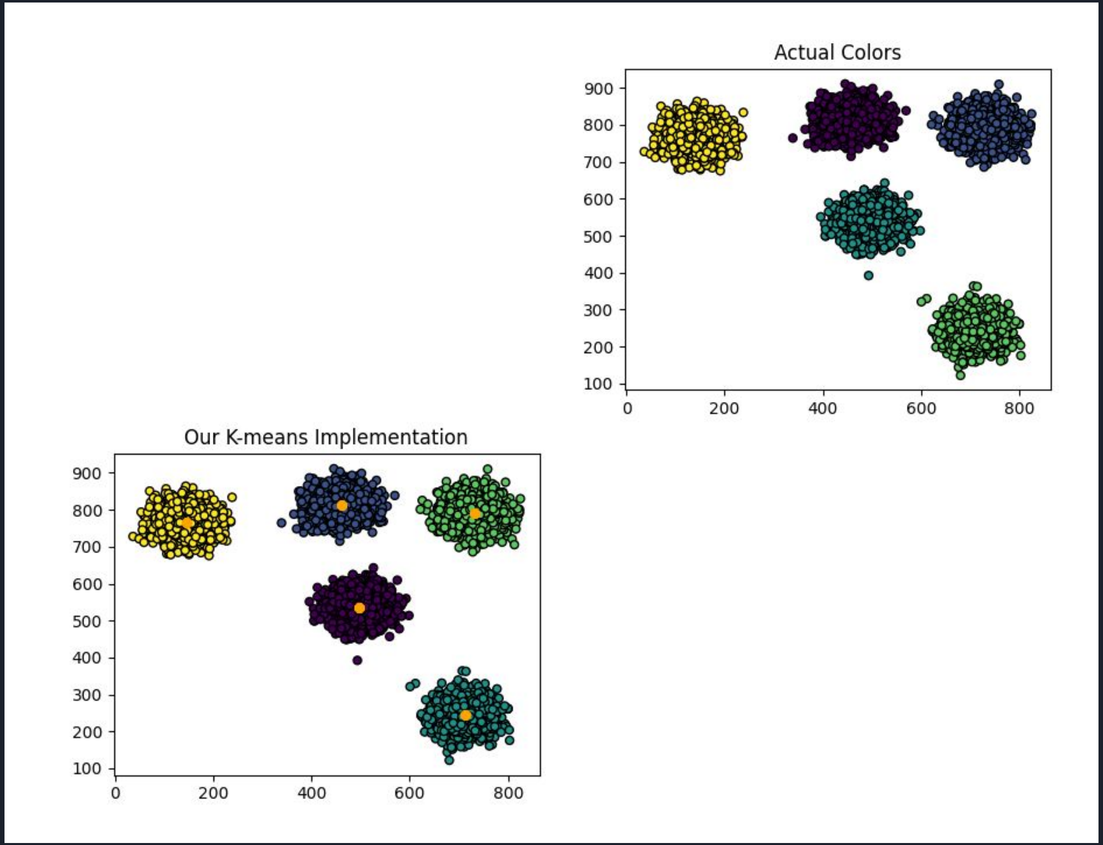

Figure 3: Our successful implementation finding the centroids

Technical Difficulties

- Used a struct
- Bad Initialization

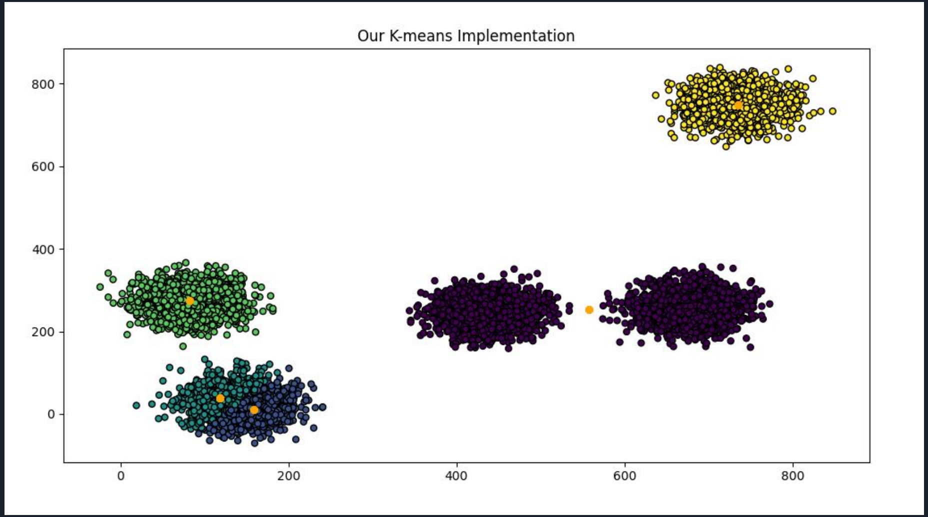

Figure 4: Another Implementation of Program

Difficulties

- Understanding the concept and identifying our own problem and solution to the concept
- Circle and Moon problem

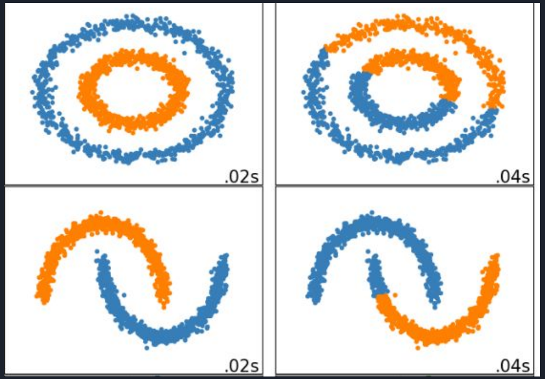

Figure 5: Example of Circle and Moon Problems

Shortcomings

- Better method to initialize cluster points
- We did not test data with more than 2-dimensional data
- Did not test the data on significantly large datasets

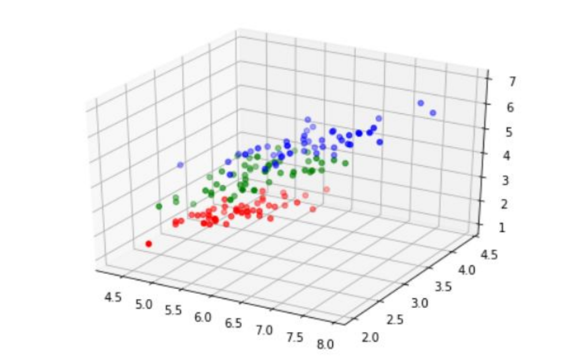

Figure 6: Example of 3D Plot

Thank You