

The Wolf of Wall Street

You are hired as data analytics Guy/Gal at the world-class merchant and investment banking firm, *WallStockTM*. Due to COVID-19-related analytics showing that people are practicing the necessary physical distancing measures, the disease is getting under control. This promises that clients will be willing buy stocks at a massive rate.

To reach out to potential buyers in the best possible way, the firm has collected data about the investment behavior of potential clients (buyers) and stockbrokers must work hard to study the financial situation and behavior of their potential buyers. However, it will take a very long time for a stockbroker to study the behavior of people from very different walks of life.

Task:

Your task is split the potential buyers data into an appropriate number of groups, where each group contains *similar* people. This way each stockbroker will be assigned to a *similar* group of people to study and reach out to. The data of potential buyers is provided to you after being preprocessed and compressed to 2-dimensions in a csv file format.

Deliverables: Write a report reflecting on the following items

- Provide the appropriate number of groups to best split the data (potential buyers) with a justification plot of why you chose that number.
- Provide a color-coded plot that shows the resultant groups, where each group of similar potential buyers has a color.

Hints:

- You can use elbow plot to argue the appropriate number of groups to split the data into.
- You can use K-means clustering to split the data into groups.