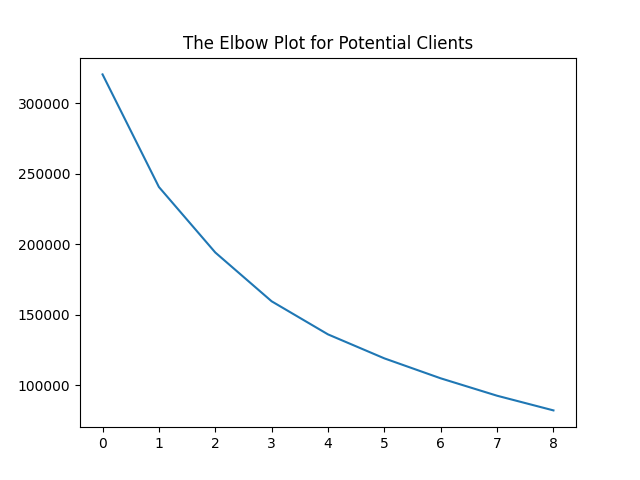
**Lab 10**

**Appropriate number of groups (clusters):**

Empirical Method: We have 5000 rows in our dataset. Therefore, we can use the following formula to calculate the number of clusters:

Number of clusters =

Thus, we obtain **50 clusters** using this formula.

Elbow Plot Method: If we generate the Elbow plot diagram for the provided dataset, we get the following trend:

The x-axis here represents the number of clusters, while the y-axis indicates the sum of square error within clusters. A lower value on the y-axis (Error) is preferable.

We can see a significant drop in error as the number of clusters increases from 0 to 1 or from 1 to 2. However, beyond 4 clusters, the reduction in error becomes marginal, suggesting diminishing returns. While lower error values are desirable, it's also important to balance this with the practicality of deploying resources. Our goal here is to minimize the number of clusters to optimize resource allocation, such as stockbrokers assigned to study and reach out to specific groups**. Therefore, the most appropriate number of clusters to split the data is determined to be 4.**

**Appropriate number of groups (clusters): 4, as we prefer Elbow Plot Method rather than Empirical Method.**

**Color Coded Plot:**

