**Clustering Of Customers**

Under this title, we made analysis according to the tax amount which paid by customers. The main logic of this clustering analysis to divide customers after ‘Subtotal’ and ‘Tax’ amount.

**Realization of Clustering Analysis**

At first the data set is prepared to analysis. We named it as ‘df\_subtotal’.

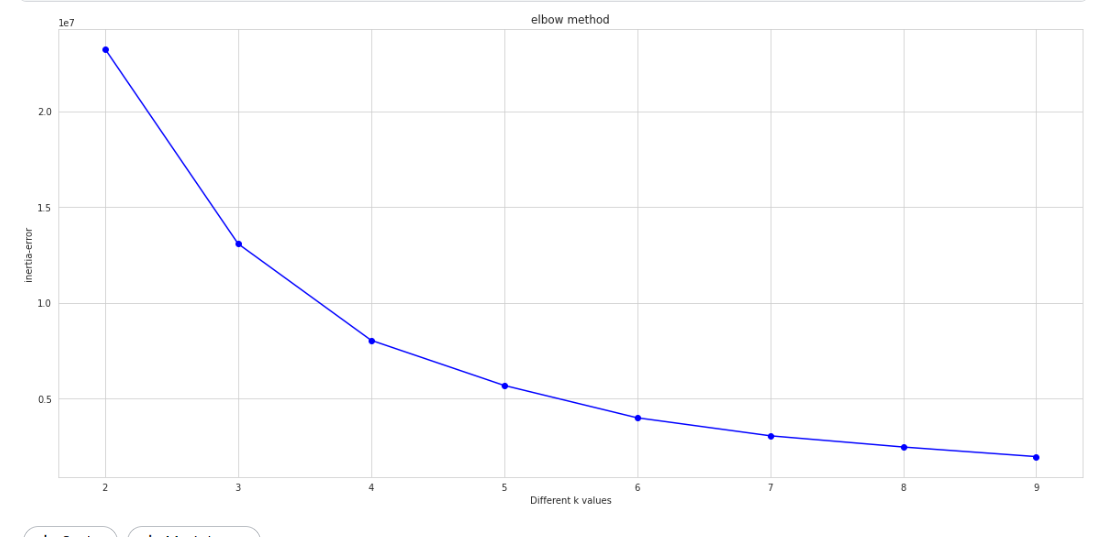
df2.set\_index('CUSTOMER\_ID',inplace=True)  
df\_subtotal = df2[['SUBTOTAL','TAX']]

In this step the target variable ‘Tax’ is dropped.

X = df\_subtotal.drop('TAX',axis=1)

In this step, the k means model determines the best clustering value. At the end of the clustering the elbow is

**from** sklearn.cluster **import** KMeans  
ssd = []  
K = range(2,10)  
**for** k **in** K:  
 model = KMeans(n\_clusters =k, random\_state=42)  
 model.fit(X)  
 ssd.append(model.inertia\_)



With this code below the model predicted the ‘Tax’ class of the customers.

**from** sklearn.cluster **import** KMeans  
  
K\_means\_model = KMeans(n\_clusters=4, random\_state=42)  
K\_means\_model.fit\_predict(X)

With the code below and the graph it is seen that the more paid the customers that paid much more tax.

plt.scatter(df\_subtotal["SUBTOTAL"],df\_subtotal['TAX'],c = df\_subtotal.predicted\_clusters,cmap='viridis');

