**KMeans Implementaion:**

Took help from this website: https://towardsdatascience.com/apply-rfm-principles-to-cluster-customers-with-k-means-fef9bcc9ab16

1)DATA PREPROCESSING:

* Read the data using pandas.read\_excel function.
* Removed the null and negative customer-IDs
* Created a recency, monetary and frequency (RMF) table using invoice time as recent transactions, total revenue as monetary value and number of transactions as frequency.
* In order to scale the large values, Calculated the f\_score, r\_score and m\_score based on the RMF table by first sorting the data on each column and then dividing the values into 5 equal parts.

2)KMeans:

* Initialized the random centroid values, based on the maximum value of each of the column of RMF table
* Initialized the number of clusters
* Calculated the error function # Error func. - Distance between new centroids and old centroids
* In a loop, assigned each value to it’s closest cluster
* Calculated the new centroids by taking the average value
* Plotted the 3d plot and demonstrated each cluster with different colours
* Repeated the algorithm for number of clusters=5,7,10

Running time:

* Using time python library to determine the running
* Plotted the graph with k on x axis and running time on y\_axis