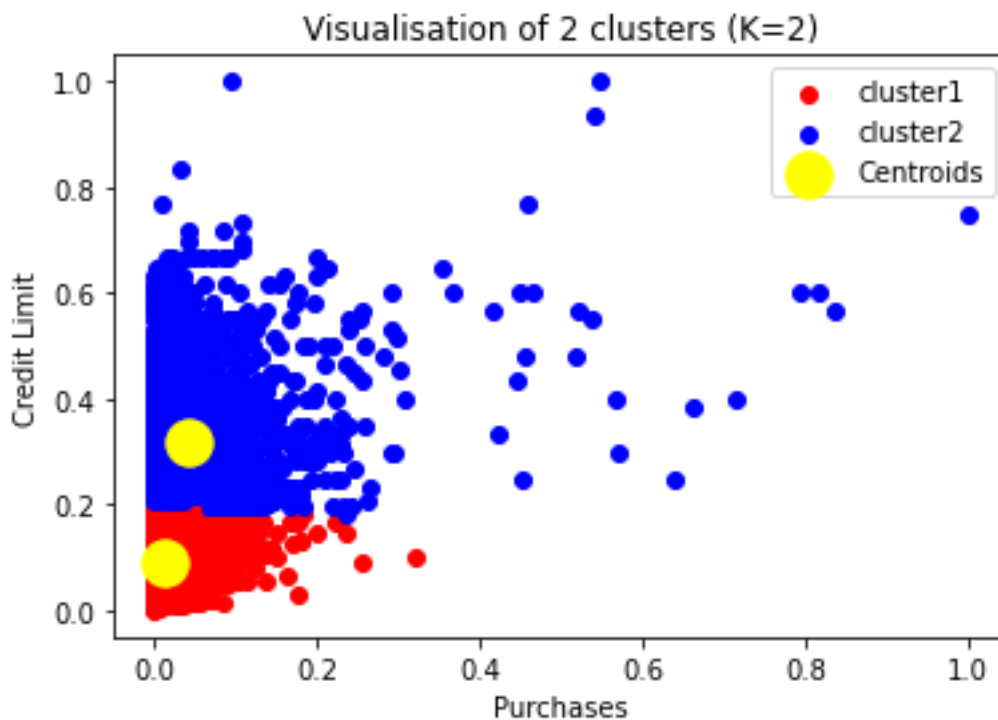


Cluster Interpretation

Q2. Write your own code for the K-means algorithm using only two attributes, PURCHASES and CREDITLIMIT. Take K=2. Plot clusters on a scatter plot with X and Y being the two attributes. Color data points belonging to the first cluster with red and the second cluster with blue. Copy the plot diagram in the word document and interpret the output. [1.5 marks]



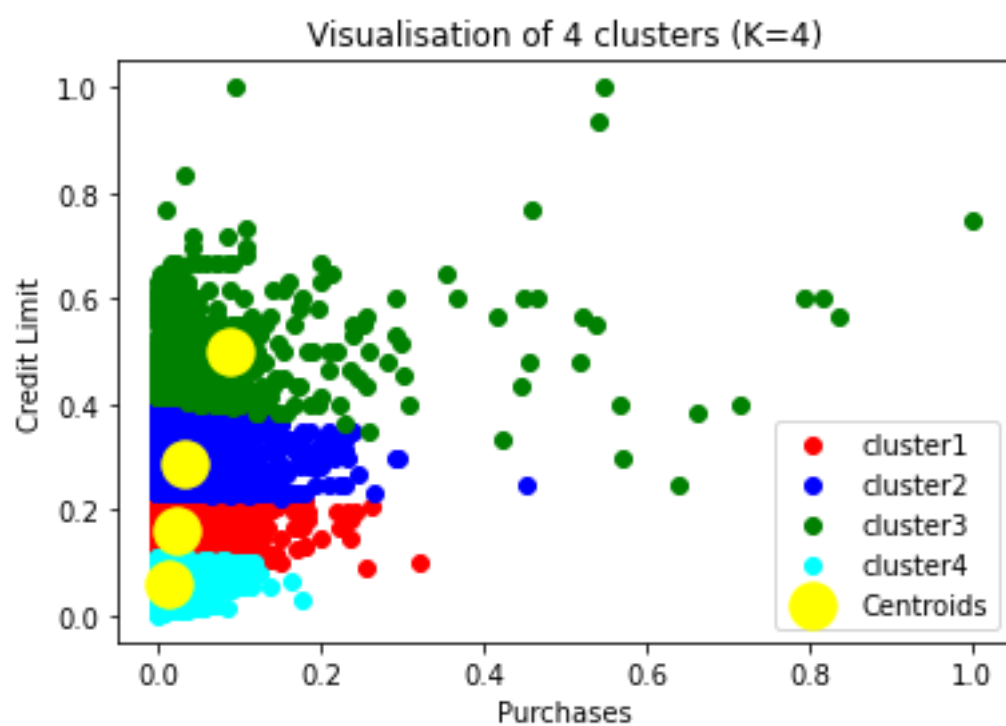
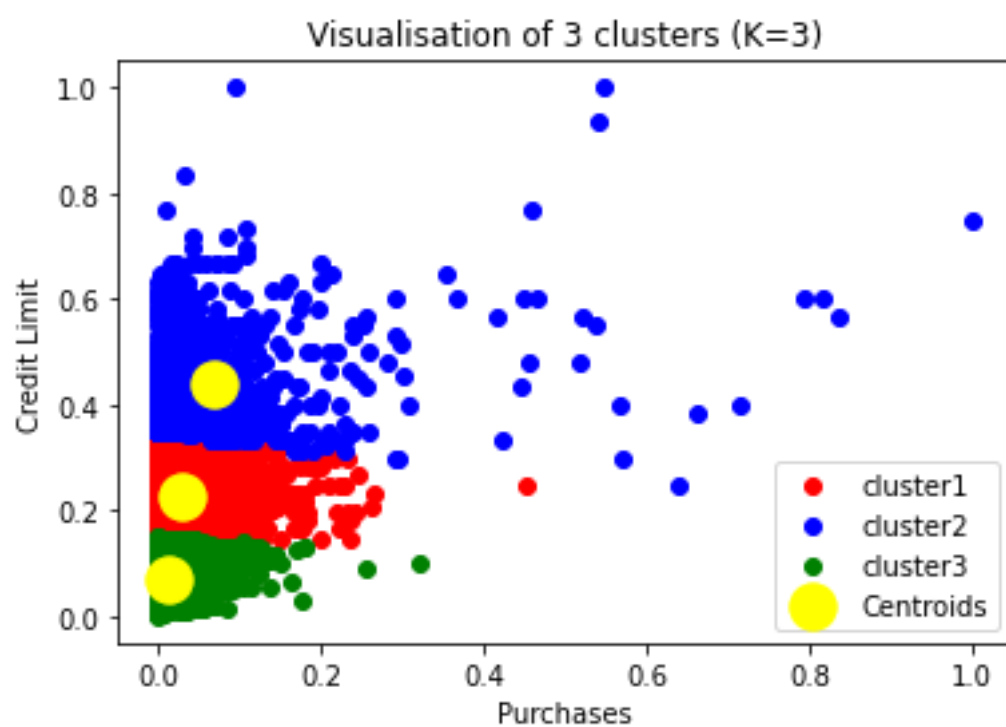
- With K=2 there are data grouped into 2 clusters
 1. Credit card holders with low credit limit and low purchases within range of 0 to 0.2 on a scale of 0 to 1. (Red color)
 2. Credit Card holders with credit limit ranging from greater than 0.2 to mostly less than 0.9 on a 0 to 1 and most purchases in the range of 0 to 0.8 on scale of 0 to 1. (Blue color)
 3. Several outlier points clearly seen on the graph are clustered with red cluster(Data skewness)
- Cluster differentiates well on the credit limit scale, but limited differentiation is seen on the purchases scale with most purchases in the range 0.2 to 0.8 for both clusters.

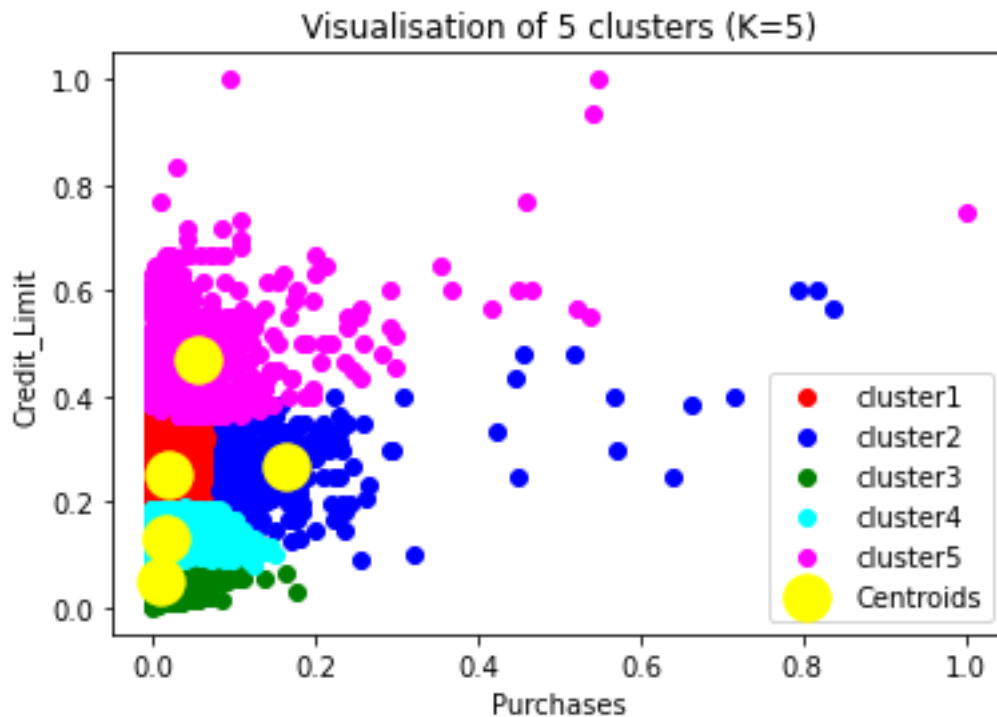
Hence, following could be interpreted from the clusters

1. **Low credit availability, low spending group** (Red cluster)
2. **Elastic credit availability but cautious spending group** (Blue cluster)

To confirm this interpretation, additional data analysis can be conducted using all other available variables.

Q3. Redo question-2 on different values of K = 3,4,5. For each case, draw the plot of clusters as stated above. Visualize these plots, copy the plot diagrams in the word document, and comment on which is better clustering (and reasons) based on visualization only





Interpretation:

- As the credit limit increases, spending also increases with K=3,4, 5 cluster groups.
- With K=4,5 low spending customers are further divided in clusters with 0.2 range of scale
- Moderate spenders are with in 0.2 to 0.4 range are in divide into sub clusters with k=4,5
- With k=5, we see a separate cluster for outliers (small group)
- K=4 Seems to best grouping based on credit limit as the data distribution across clusters and it helps business to more further analysis on spending with variety of group of customers

Q4. Use the code written by you to cluster the data using all the features in the dataset. Take k=5

for this. [1.5 marks]

- With All features in K means with 5 cluster, its multi-dimensional data and can't make any decisions from centroids.
- From Data distribution across with 2 features and all features are 90% similar grouping. We can infer that Credit limit is important features among all other features.
- We can apply PCA to reduce dimensionality and make better clusters when we have more features.

Q5. Write a few lines as comments in the notebook about the interpretation of the best clusters obtained. Also write a few statements about how these clusters can be useful.

- From the data visuals in previous steps, we can make conclusion that spending increased with credit limit. With K=4 best cluster obtained where as K=5 , outliers are formed as new cluster and we can't make any interpretation with outliers group.

- With $K=4$, we can make more inferences on credit limits vs spenders' analysis. It helps business to see how to improve the spending of less spender cluster
- Again, best cluster decision can be varied on what business question we are trying to answer.
- We can attract low spenders by giving offers or increase the credit limit to make more business.

Clustering is useful because

- Customer segmentation is not only important, but vital, in order to optimize your marketing strategies, maximize a customer's value to your business, and improve customer experience and satisfaction.
- Finding an optimal number of unique customer groups will help you understand how your customers differ, and help you give them exactly what they want. Customer segmentation improves customer experience and boosts company revenue. That's why segmentation is a must if you want to surpass your competitors and get more customers. Doing it with machine learning is definitely the right way to go.