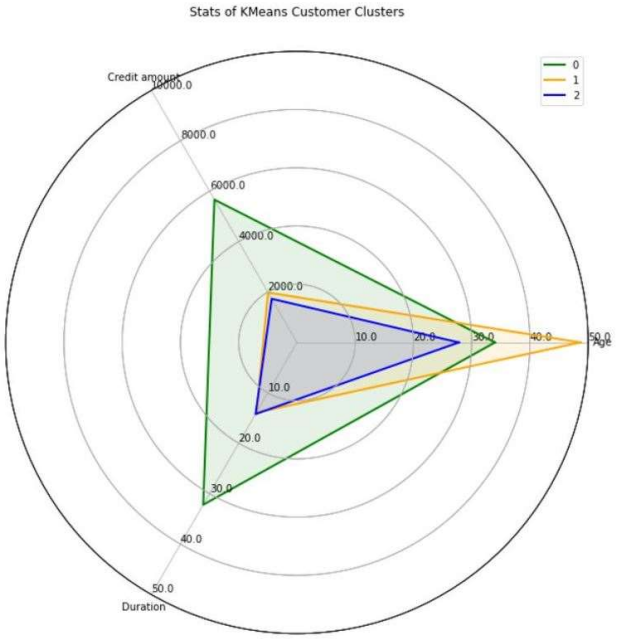
German Bank Credit Risk Analysis



Cluster	Age	Credit amount	Duration
0	34.0	5665.4	32.2
1	48.7	1979.7	13.9
2	27.8	1732.2	14.2

Cluster 0 – lower mean of credit amount, short duration, older customers

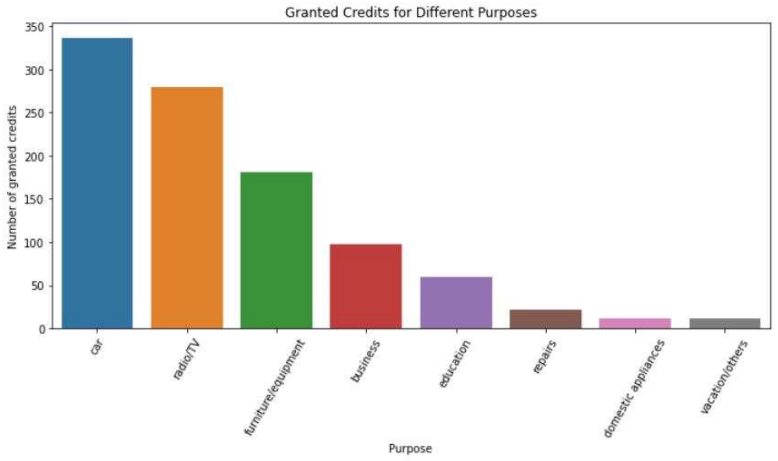
Cluster 1 – high mean of credit amount, long duration, middle-aged customers

Cluster 2 - lower mean of credit amount, short duration, young customers

- The German credit dataset (1994), prepared by Professor Hans Hofmann from the University of California Irvine, was used to identify customers with either good or bad credit risks based on their financial and banking details.
- The bar plot below shows how many credits were granted for various purposes. Most of credits were granted for car and radio/TV.
- The k-means algorithm clustered customers into three segments based on their age, credit amount granted, and credit duration.

1) The 1st cluster represents elderly customers over 48 years old who are offered with low credit amount and short credit duration
2) the 2nd cluster entails middle-aged customers granted with higher credit amount as well as long credit duration
3) the 3rd cluster is characterized by young customers in their late 20s who receive low credit amount and short credit duration

For the next step, I would recommend inquiring about reasons behind the k-means algorithm, and whether the k-means clustering segmentation results are aligned with business assumptions as well as how the clustering methodology can facilitate the decision of the bank of giving credit to a client in the future.



Data Source: Statlog (German Credit Data) Dataset provided by Professor Dr. Hans Hofmann from UCI