Clustering credit card customers using KMeans

Data analysis, interpretation and modelling

by Amy Birdee

Introduction

- The data consist of details of credit card customers at a bank

- There are 8,950 rows of customer data. All monetary data are assumed to be in pounds sterling (£)

- Included in the data are variables such as card balance, the amount of purchases made, the number of one-off purchases made and proportion of balance repaid

- This project aims to segment the customer data so the bank is better able to target these customers with specific marketing

- The machine learning model used in the project is KMeans clustering which segmented the credit card customers into three distinct groups by using the Elbow method

What does the credit balance and credit limit look like for bank customers?

- Both charts have a right skew indicating that most customers have a balance and credit limit on the lower end of the scale while a few customers have very high credit balances and credit limits

Distribution of Balance for Bank Customers

Distribution of Credit Limit for Bank Customers
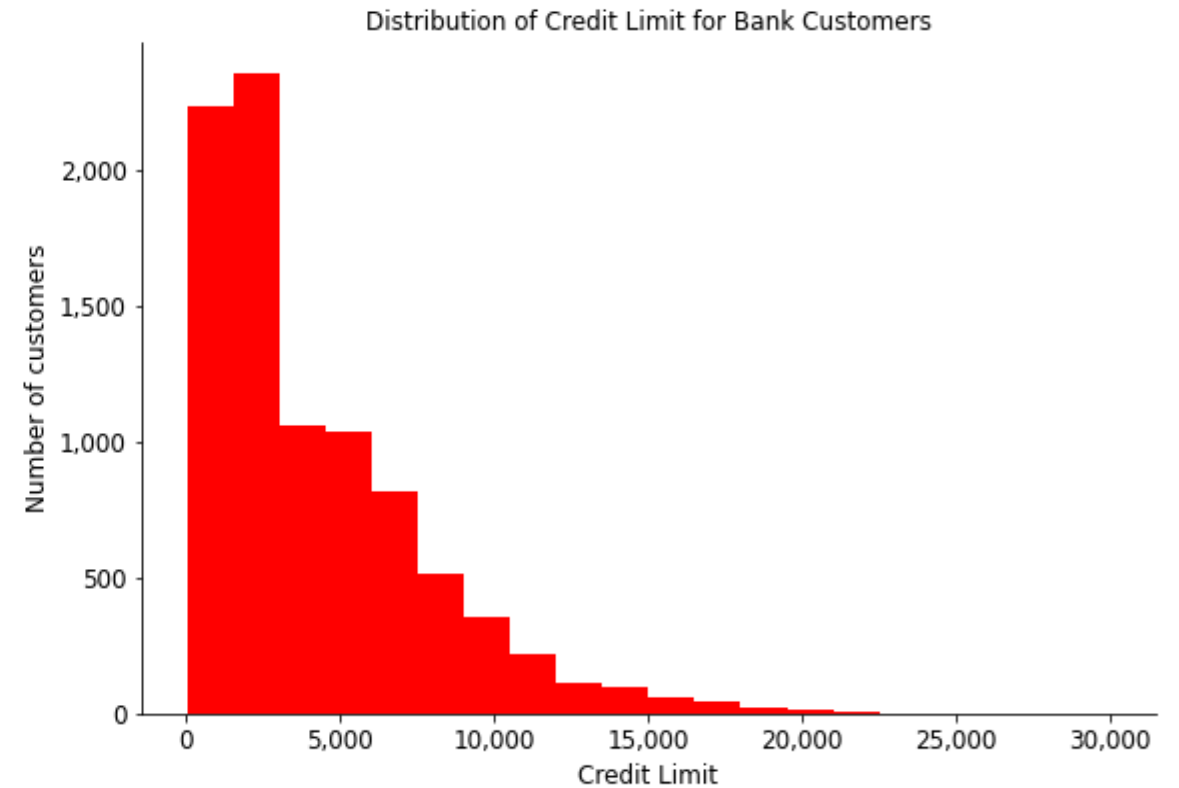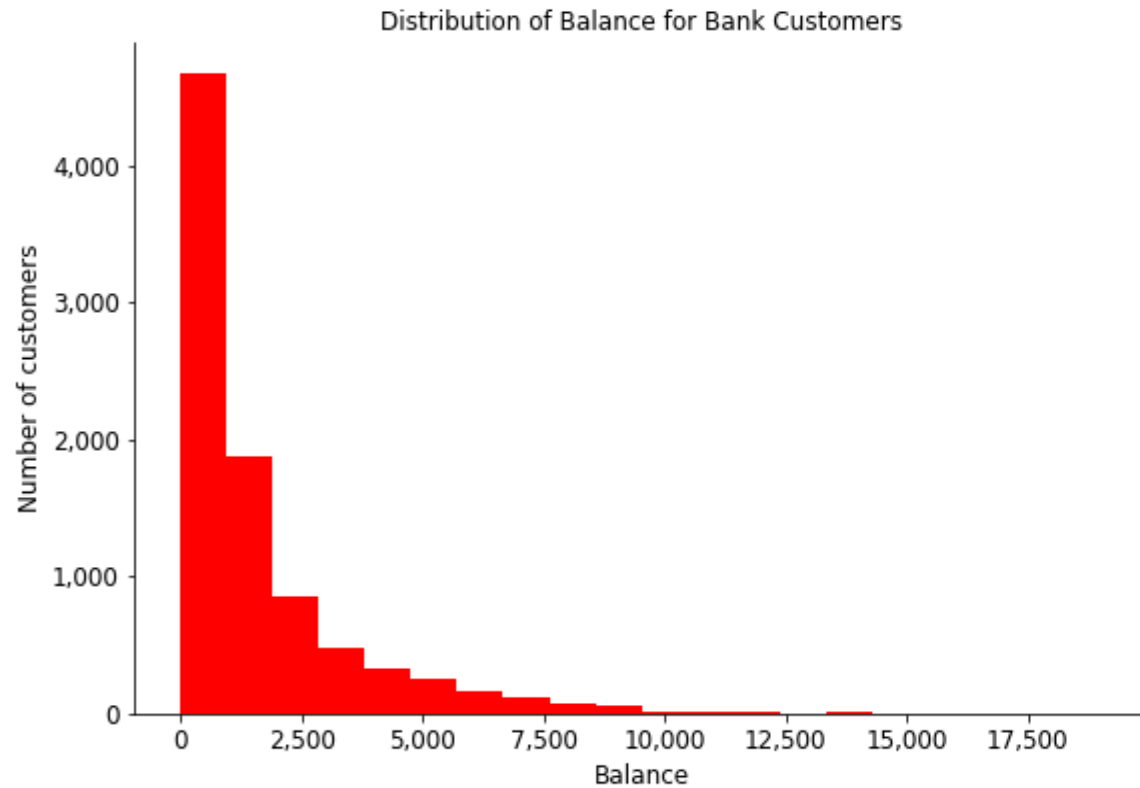
# What do purchases look like for bank customers?

- Again, all charts have a right skew with most customers limiting purchases to less than £5,000
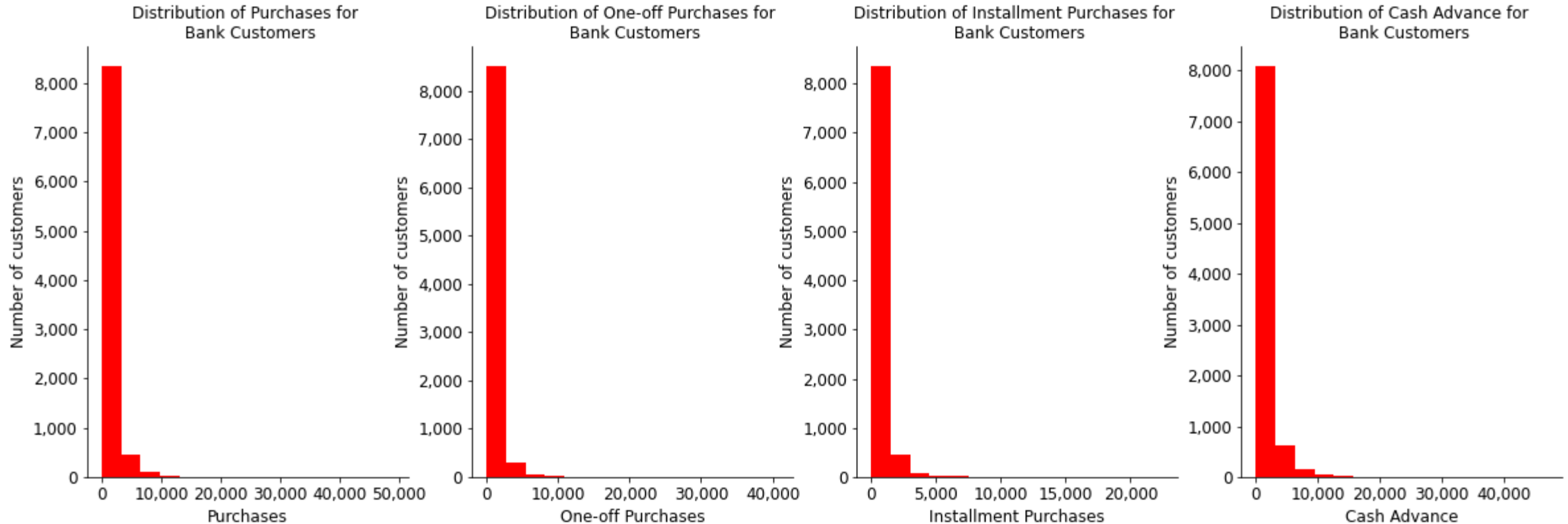

- However, some customers have made very large purchases. There are 21 customers in the dataset who have made purchases of over £20,000 and these are the customers who are causing the data to appear skewed

What do repayments look like for bank customers?

- The repayments data show a right skew indicating that most customers repay the balance in smaller chunks while other customers repay huge sums – these are likely the same customers who borrow huge sums
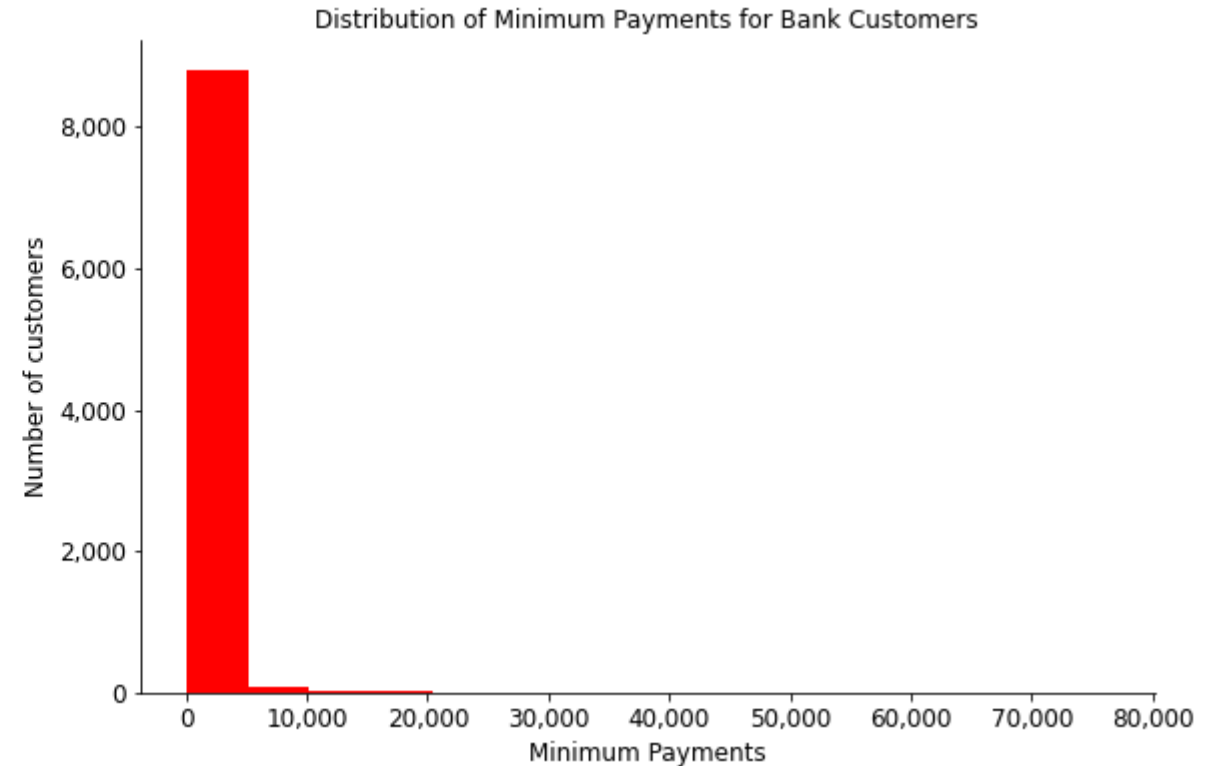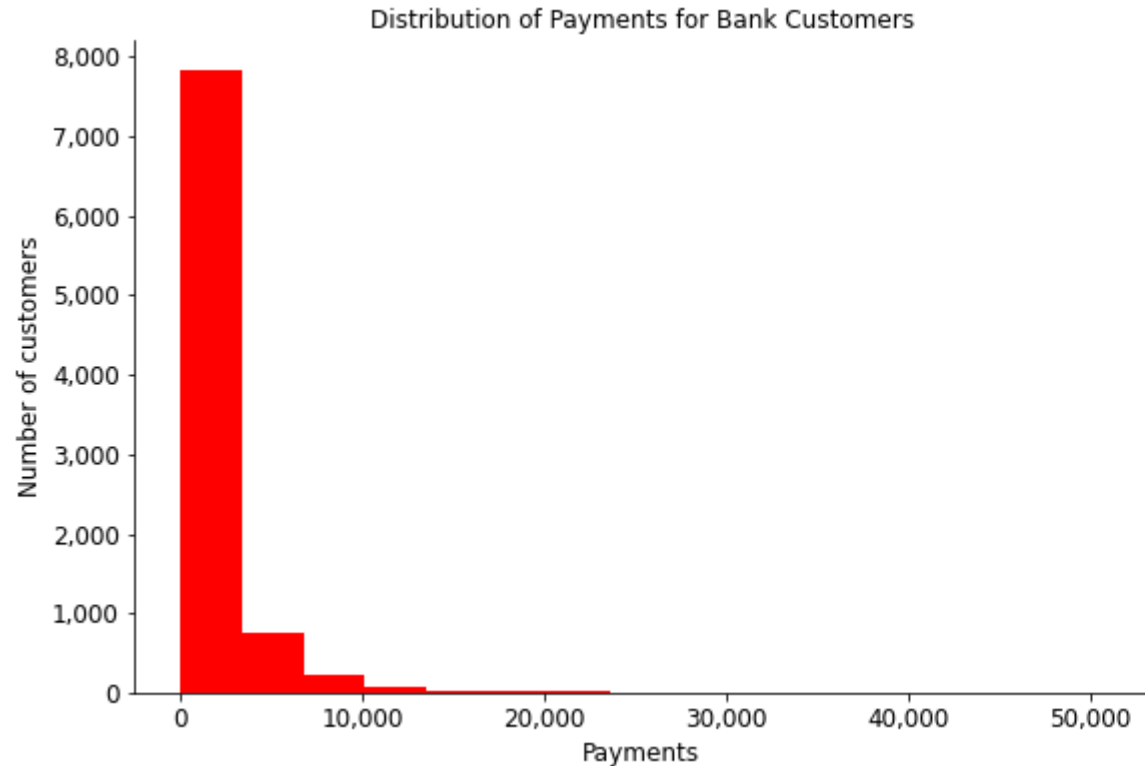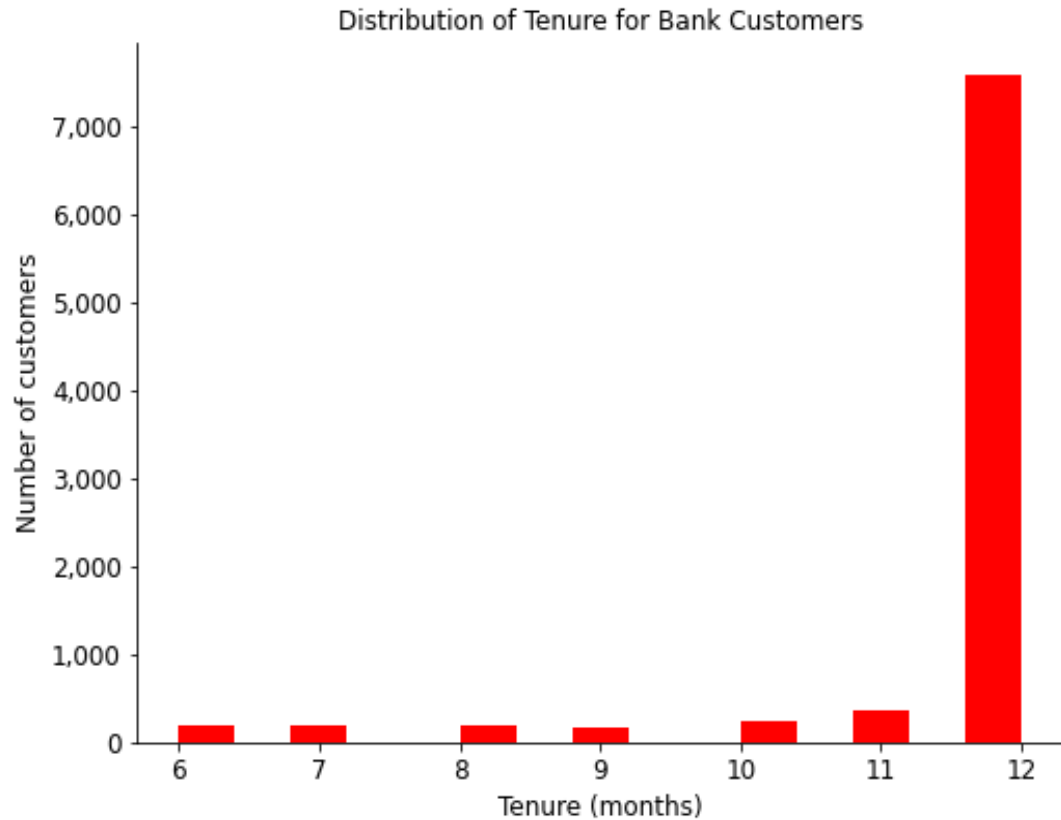
What proportion of the balance do customers repay?

Distribution of Repayment Percentage for Bank Customers
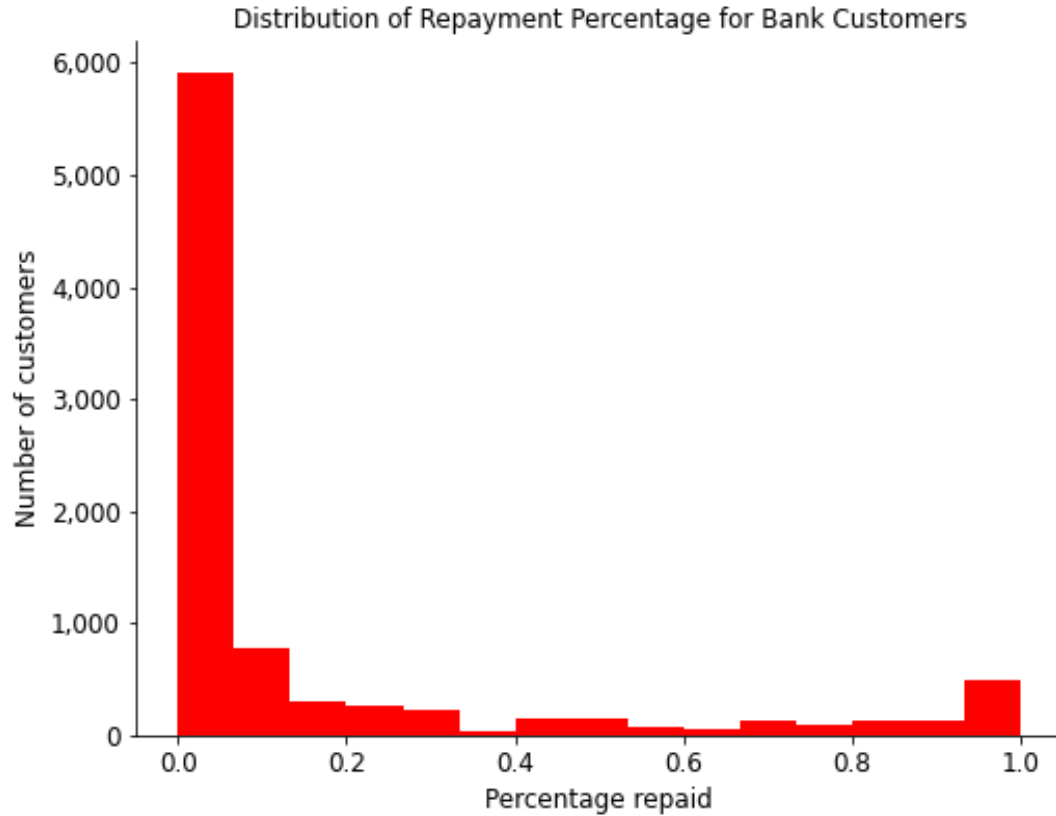
- The majority of customers have repaid 5-10% of their balance

- There are around 500 customers who have repaid their whole balance

# How long have customers been using the credit card service?

Distribution of Tenure for Bank Customers

- The majority of customers have been with the bank for 12 months

KMeans clustering: the Elbow curve

The Elbow Method Showing the Optimal Number of Clusters (k)
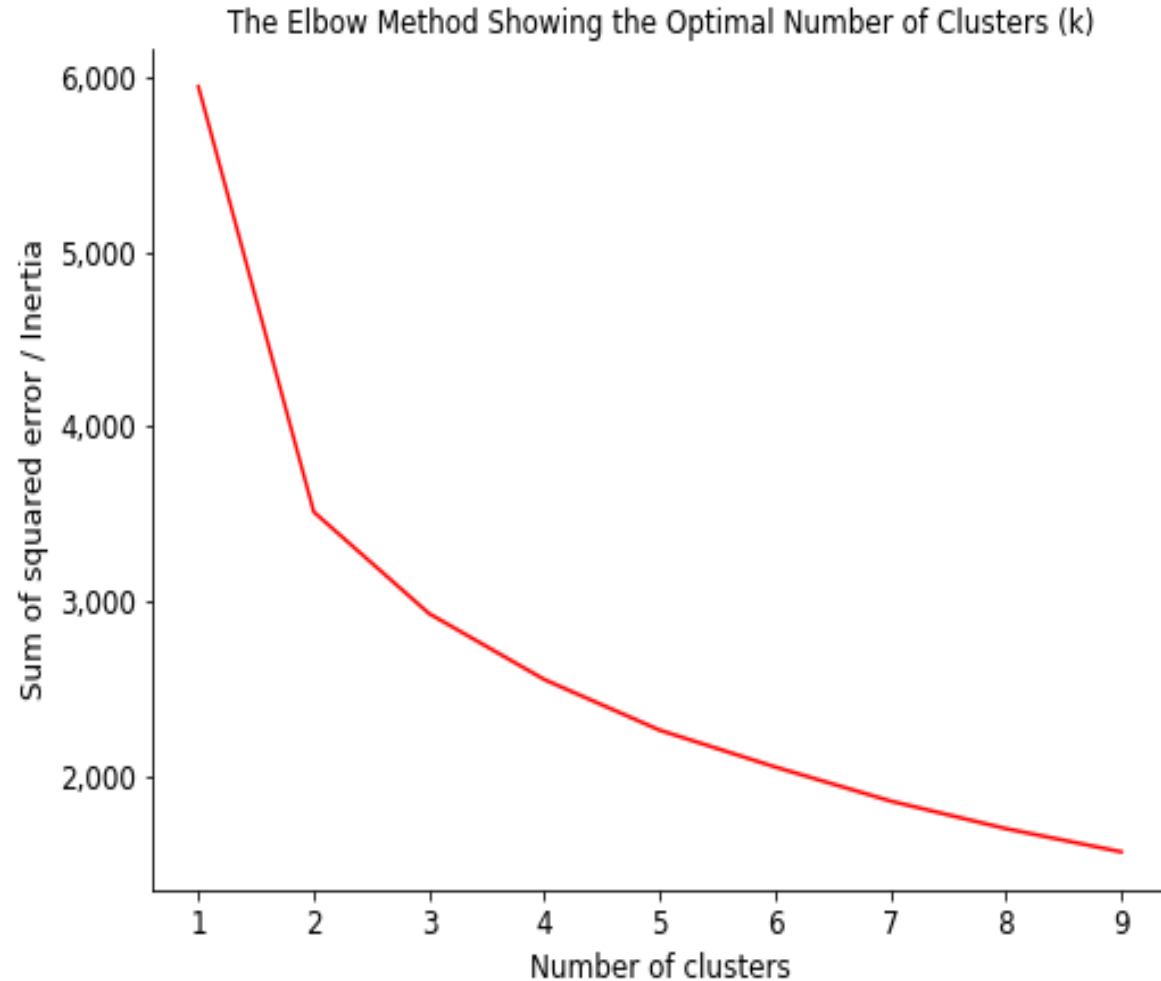
- The Elbow method is one of the most common ways to find the optimal number of clusters, K

- The Elbow curve plots the number of potential clusters on the X axis and the sum of squared error or inertia on the Y axis

- The smaller the inertia value, the denser the clusters (the closer together all the points within the cluster)

- To find the optimal number of clusters, we use the point at which the **elbow chart bends**

- In this chart we could use **2 or 3 clusters**. We will use three clusters

KMeans clustering: the clusters
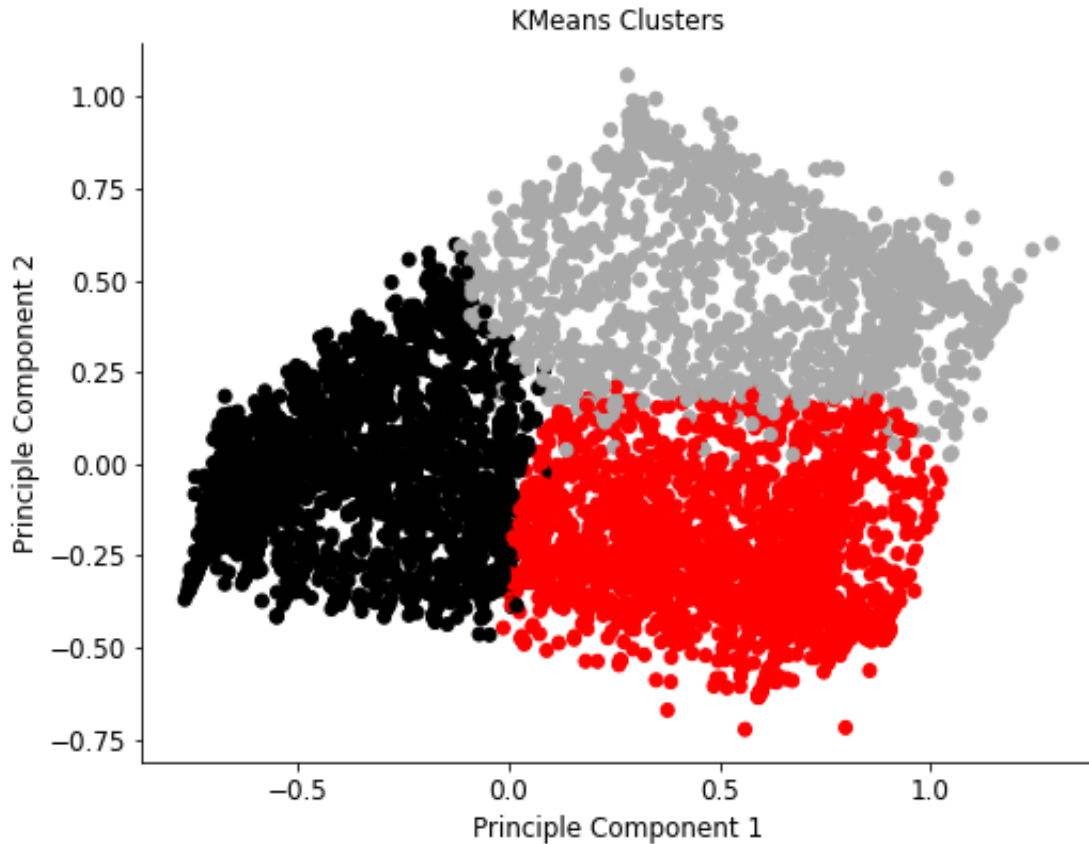
KMeans Clusters

- After implementing the KMeans model with K = 3, we have produced **three distinct clusters**

- The bank can now distinguish between different types of customers and **target them with specific media** communications depending on which cluster they fall in

How do the different clusters compare for each feature of the dataset?
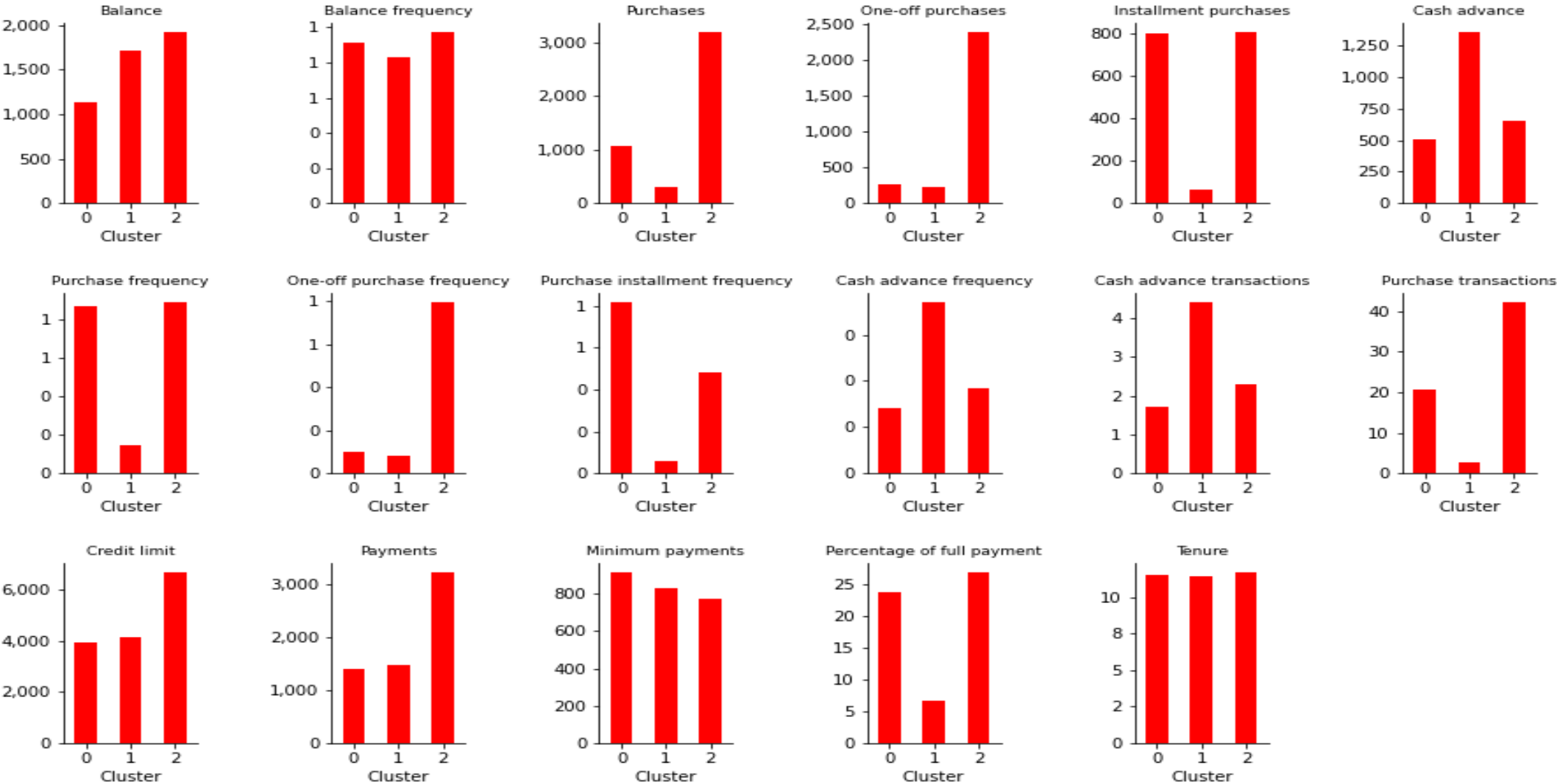
Conclusion

- The data showed that many of the charts had a right skew indicating that the majority of customers have lower balances, credit limits and spend behaviour while a few customers were at the higher end

- The **Elbow curve** showed that customers could be grouped into two or three distinct clusters. Since the number of clusters was small, **three clusters** were chosen in this project

- The bank will now be able to **analyse the behaviours** of the customers within these clusters and **determine the best marketing strategy** for each one

- Targeting the groups with three different types of communication is likely to result in **more meaningful targeting** which will in turn result in **more engagement with the bank's products**

Thank you

Contact details:

Amy Birdee

amybirdee@gmail.com