Data Glacier

Your Deep Learning Partner

Final Report

Customer Segmentation for XYZ Bank

December 2, 2022

Team Member Details

Data Glacier
Your Deep Learning Partner

Group Name: Elizabeth's Analytics

Name: Elizabeth Banning

Email: estall@hotmail.com

Country: USA

College: Western Governors University

Specialization: Data Science

Agenda

Executive Summary

Problem Statement

Approach

Correlations

EDA Summary

Customer Segments

Recommendations

Executive Summary

- **Purpose:** Segment customers into 2-5 groups for marketing campaign
- **Methods:** Clean data, then use k-means clustering analysis
- **Timeline:** Final results by December 30, 2022
- **Results of EDA:** Dataset cleaned, correlations and distributions explored
- **Results of Clustering:** 4 distinct customer groups identified

Problem Description

In order to develop its promotional campaign, XYZ Bank needs to know the answers to the following questions:

- What is the best number of groups to divide customers into?
- What are the primary characteristics of each group?

To answer these questions, the k-means clustering algorithm will be used to segment the customers, and the inertia metric will be used to determine the optimal number of groups (k). Finally, the characteristics of each group will be summarized so that XYZ Bank can determine which offers to develop and target to each group.

Problem Statement

- XYZ is a bank that wants to do a promotion
- 1,000,000 customers: need to tailor different promotions to different types of customers
- Maximum 5 groups
- **How can customers be grouped?**
- **What are the characteristics of each group?**

Approach

Data Glacier
Your Deep Learning Partner

1 file used

- 1,000,000 customers (rows)
- 48 features (columns)

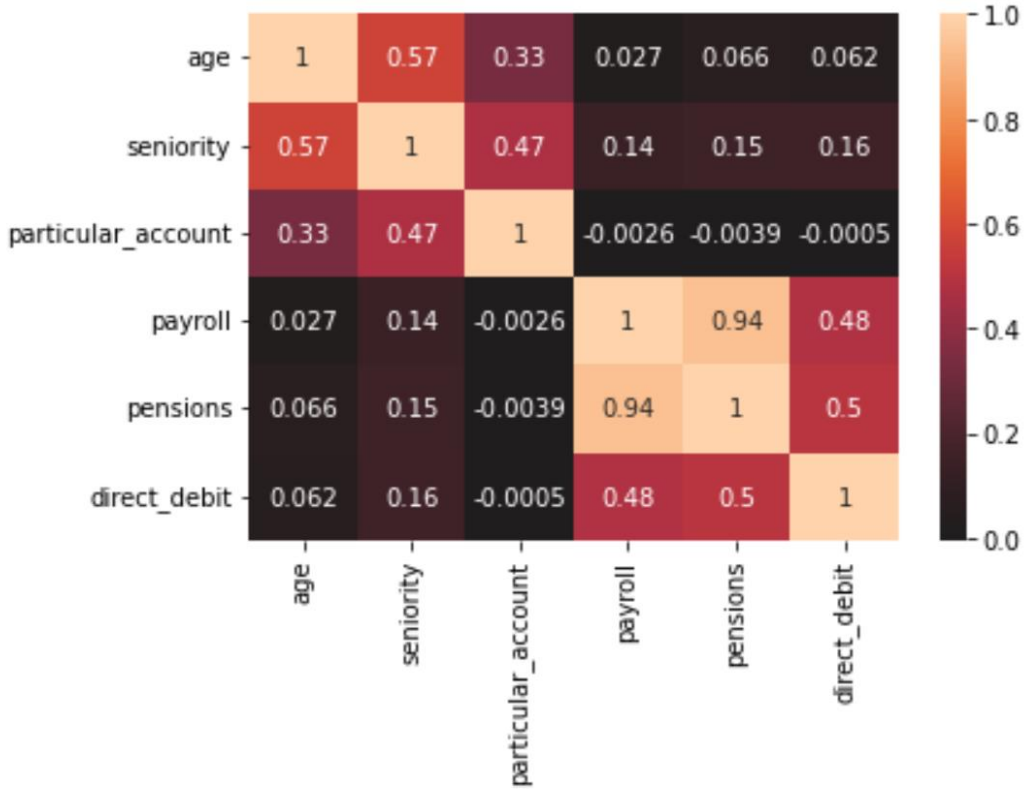
Clean the data:

- Check for duplicates and remove
- Check for missing values (treatment depends on type of data)
- Check for impossible/nonsense data and correct if necessary
- Drop irrelevant features (ID number, etc.)

Cluster:

- Determine ideal number of clusters (up to 5)
- List characteristics of each group

Correlations

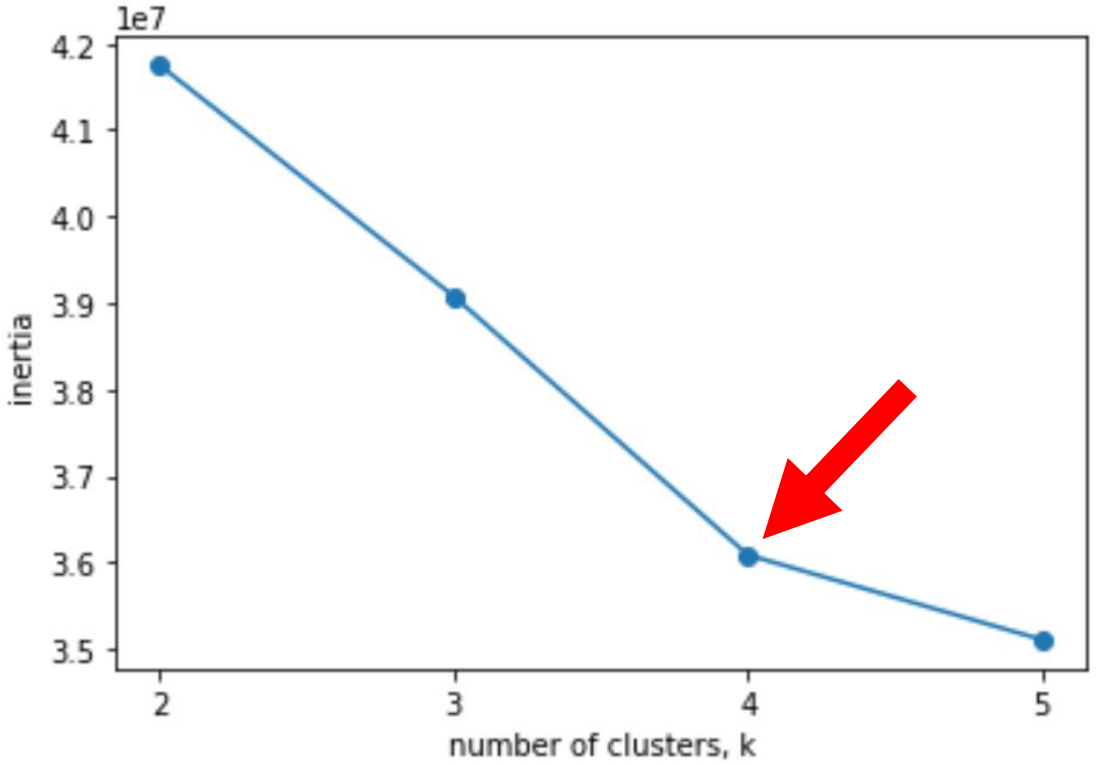


- Age, seniority, and particular account were correlated. These may be important together in grouping customers.
- Payroll, pensions, and direct debit accounts were also correlated.
- These groups may help distinguish different customer groups. For example, older customers are more likely to have a particular account.

EDA Summary

- Dataset was cleaned: no more missing or nonsensical values
- Customer age, seniority, and income were all positively skewed and could help distinguish customer groups
- Some categories could also distinguish customer groups, such as gender, active level, and various accounts held
- Age, seniority, and particular account are correlated with each other
- Payroll, pensions, and direct debit are correlated with each other

Ideal number of customer segments (clusters)



- Ideal number of clusters is where adding more doesn't result in additional information gain
- The "bend" in the graph
- Here ideal number is **4 groups**

Group 1: Current, long-time personal accounts

- 459,759 customers (the largest group)
- Median age: 45 years old
- Median seniority: 108 months (9 years)
- Median income: 106,651.90
- Typical customer in this group:
  - Is active user
  - Has current account
  - Is primary account holder

Group 2: Inactive personal accounts

- 441,071 customers
- Median age: 37 years old
- Median seniority: 63 months (5 years 3 months)
- Median income: 106,651.90
- Typical customer in this group:
  - Is **NOT** active user
  - Has current account
  - Is primary account holder

Group 3: Current, long-time business accounts

- 88,347 customers
- Median age: 46 years old
- Median seniority: 147 months (12 years 3 months: the longest-term customer group)
- Median income: 111,303.20
- Typical customer in this group:
  - Is active user
  - Has payroll, pensions, and direct debit services
  - Is primary account holder

Group 4: Former customers

- 41 customers (the smallest group)
- Median age: 33 years old
- Median seniority: 0 (no longer with company)
- Median income: 125,137.50
- Typical customer in this group:
  - Is not active user
  - Is a former customer
  - Is **NOT** primary account holder

What all 4 groups have in common

- More males than females in all groups
- Resident, not foreigner
- Not employee or spouse of employee
- Typical customer does **NOT** have:
  - Savings account or credit card
  - Guarantees
  - Derivative account or securities account
  - Junior account
  - Other account
  - Particular account
  - Plus account
  - Short/medium/long-term deposit
  - E-account
  - Funds
  - Mortgage, loan, or home account
  - Tax account

Recommendations

- Group 1 (current, long-time personal accounts): Offer promotions on accounts individuals typically don't have but might want to use such as savings, credit card, and loans
- Group 2 (inactive personal accounts): Offer promotions to encourage activity using incentives
- Group 3 (current, long-time business accounts): Offer promotions on business services such as tax account or loans
- Group 4 (former customers): Offer promotions such as incentive for opening account to encourage customers to return

Thank You

GitHub repository link:

https://github.com/ebanning/DataGlacierProject

Data Glacier

Your Deep Learning Partner