Customer (Banking) Segmentation

March. 4th, 2022

Siyu Mao
Hashim Javed

Project Importance

Source:
https://medium.com/analytics-vidhya/credit-card-customers-segmentation-bc3c5c87ddc

Company has limited resources, and must focus on how to best identify and serve its customers.

(a) an analysis of how products should be sold or developed, based on an analysis of current customer segments

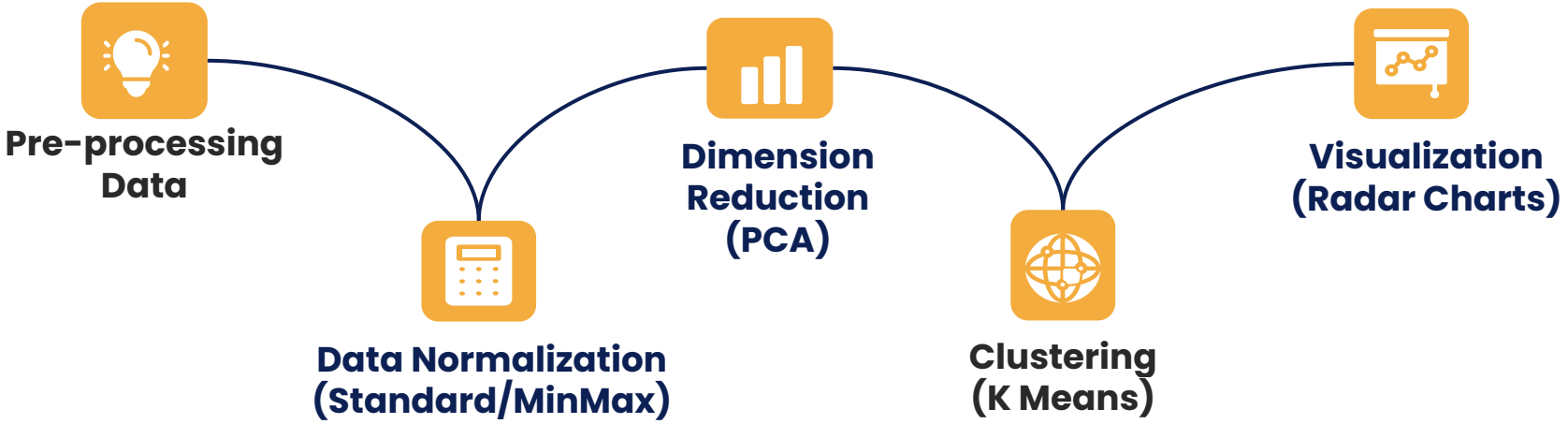
(b) the identification of new segments as targets for existing products or the development of new products.

Table of contents

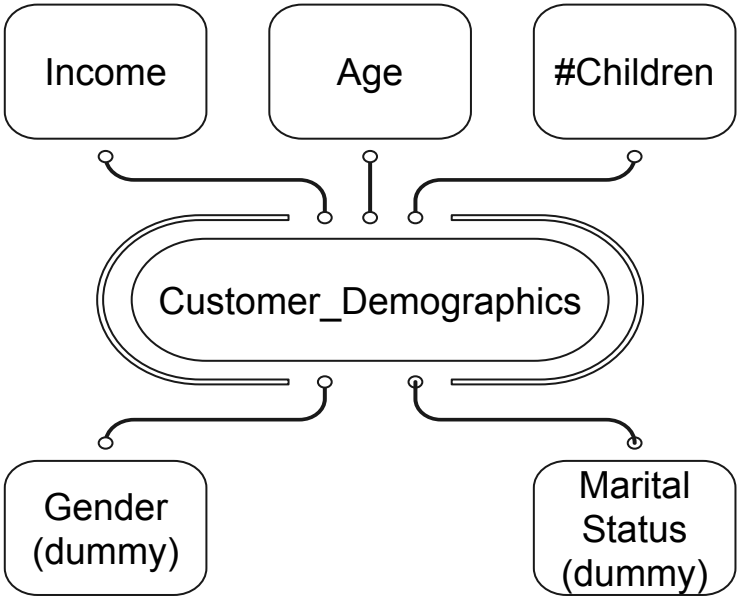
01

Project Workflow

Steps:


Pre-processing Data
Data Normalization (Standard/MinMax)
Dimension Reduction (PCA)
Clustering (K Means)
Visualization (Radar Charts)

Demographics



income	float64
age	float64
nbr_children	float64
F	float64
M	float64
marital_status_1	float64
marital_status_2	float64
marital_status_3	float64
marital_status_4	float64

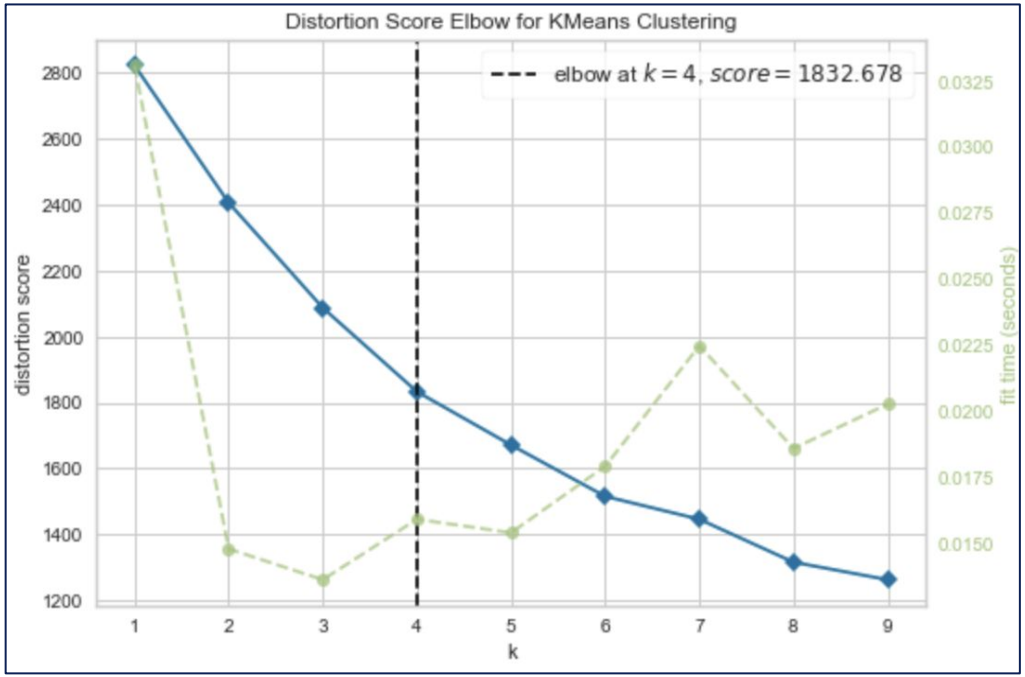
Banking Behavior

acct_nbr	tran_id	tran_amt	principal_amt	interest_amt	new_balance	tran_date	tran_time	channel	tran_code
13624802					34.47	2.1.1995	0	P	WD
					34.32	2.1.1995	0		FK
					28.86	21.1.1995	134752	A	WD
					49.01	21.1.1995	0	P	WD
					161.88	20.1.1995	191749	K	TR
45611432					-7100.00	13.5.1995	154408	E	CG
					-6464.42	4.5.1995	0	M	PM
					-7093.01	29.4.1995	81523	E	CG
					-6466.65	9.4.1995	83112	A	IQ
					-6869.00	22.12.1995	182026	E	CG

credit_limit	int64
change_in_debts	float64
change_in_savings	float64
income	int64
years_with_bank	int64
sav_trans_count	int64
sav_trans_sum	float64
sav_trans_avg	float64
cre_trans_count	int64
cre_trans_sum	float64
cre_trans_avg	float64

KElbow visualizer

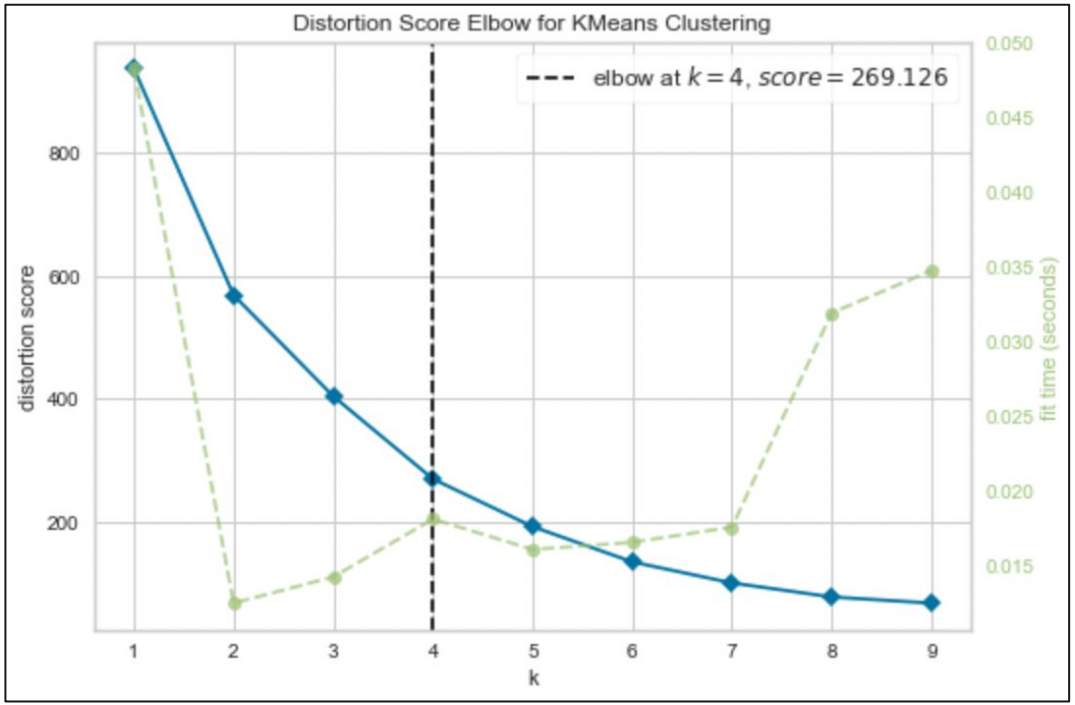
(from yellowbrick.cluster import KElbowVisualizer)



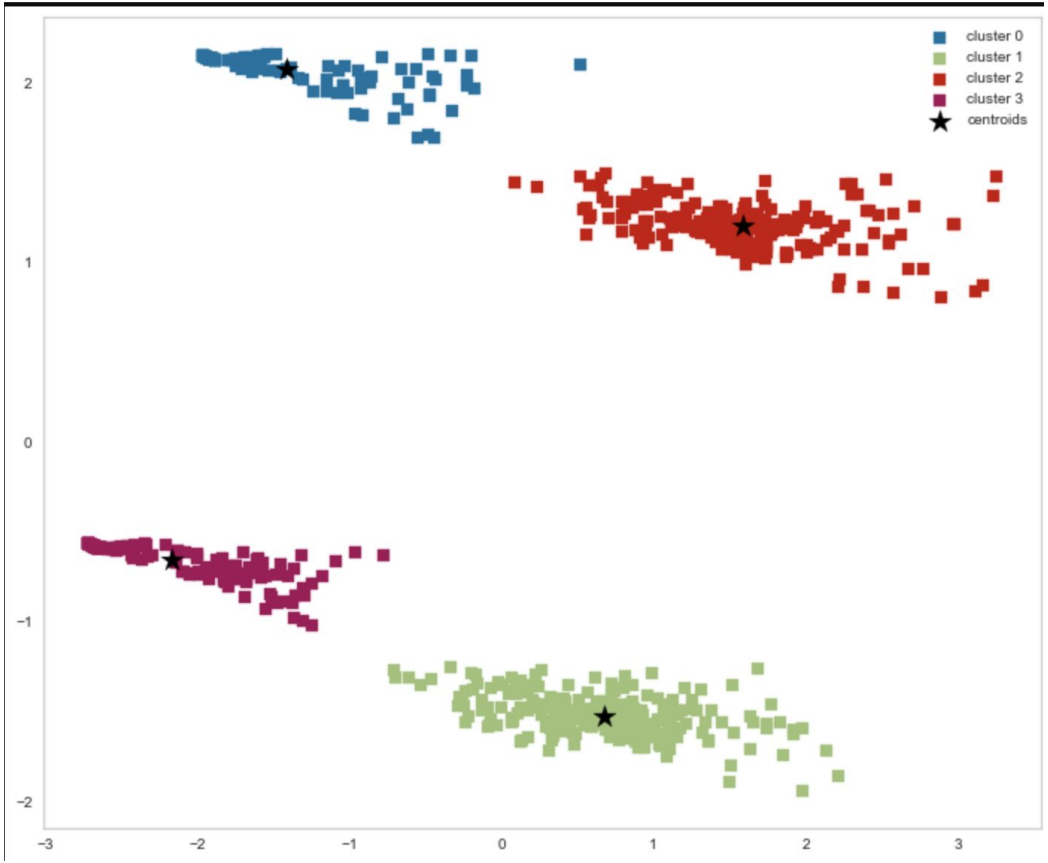
02

Results

Demographics - Kmeans

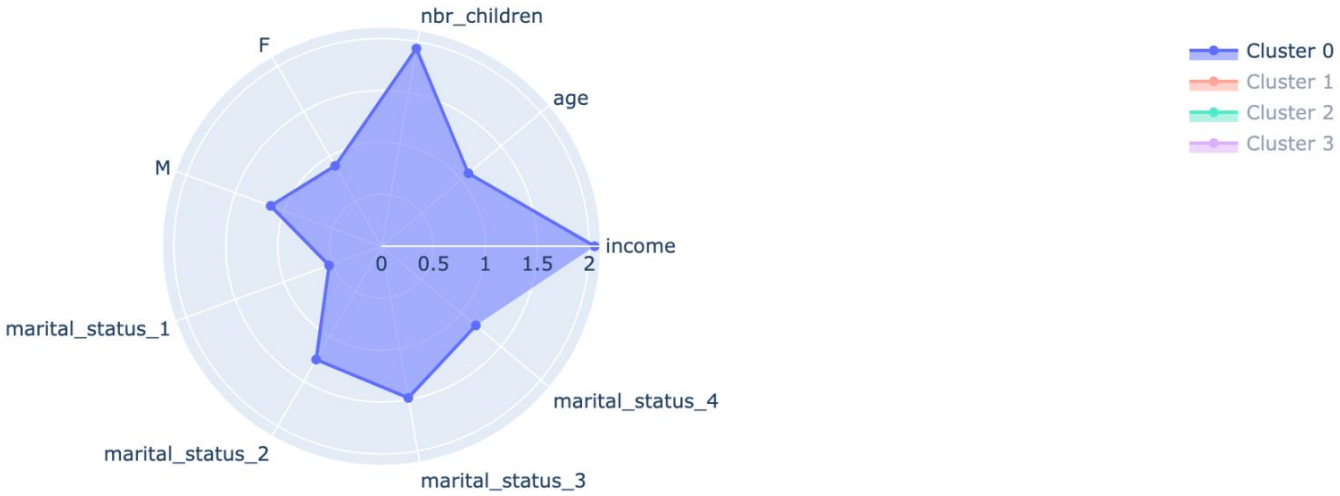


Demographics



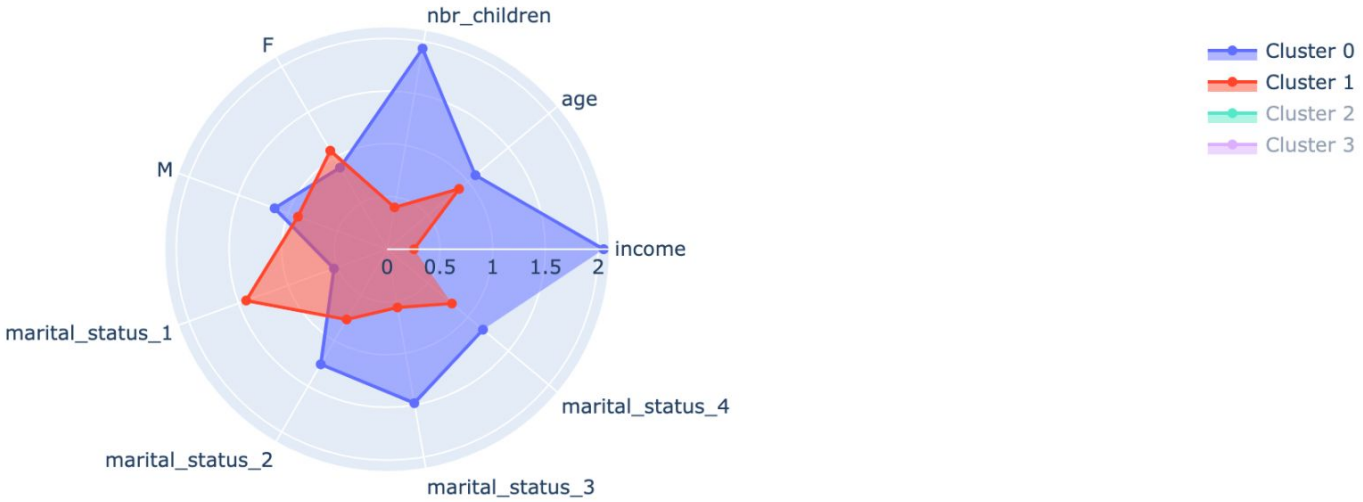
Demographics - Cluster 0

Customer Demographics Radar



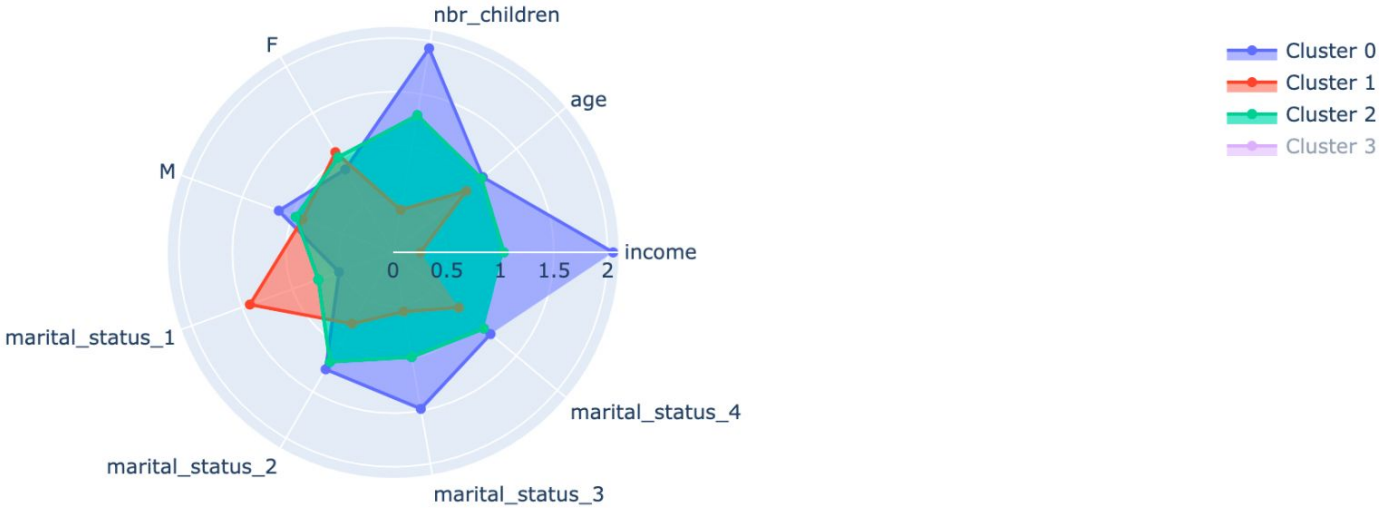
Demographics - Cluster 1

Customer Demographics Radar



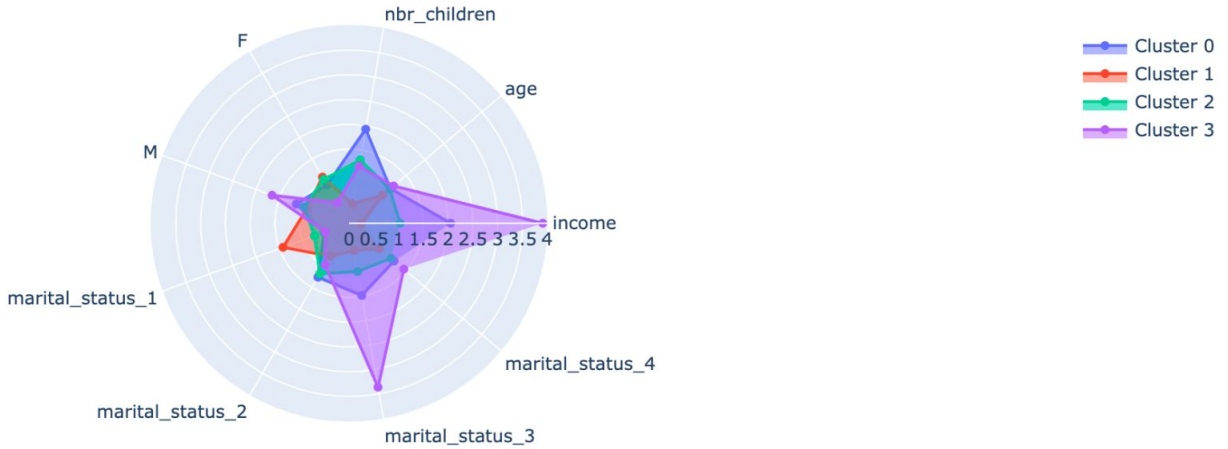
Demographics - Cluster 2

Customer Demographics Radar



Demographics

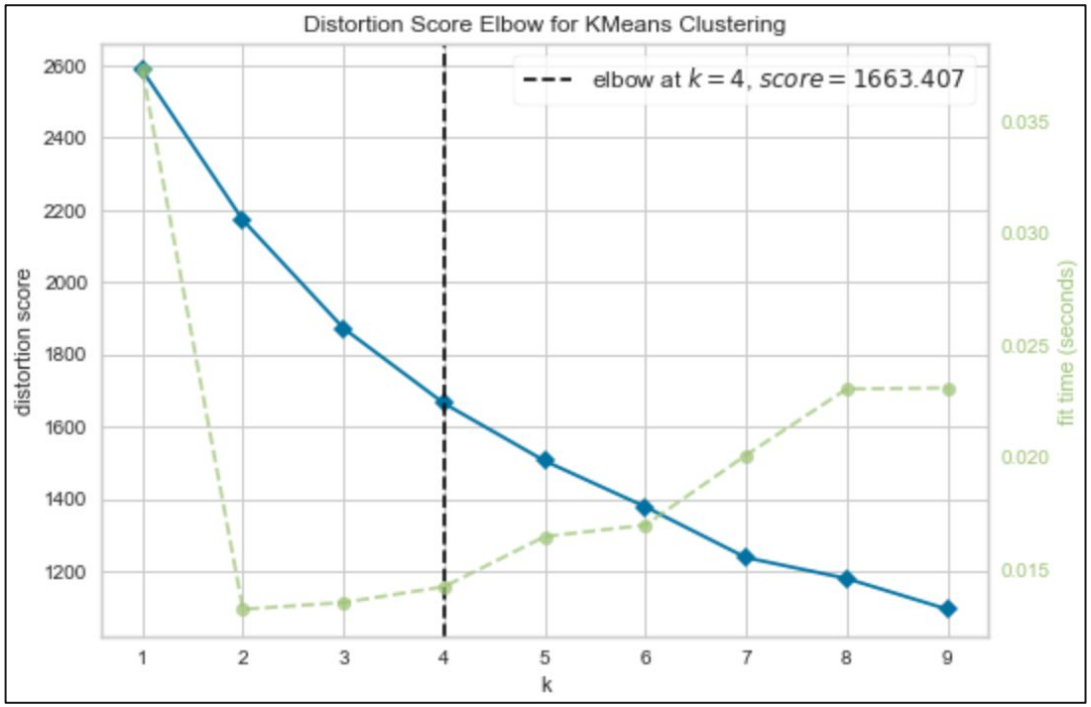
Customer Demographics Radar



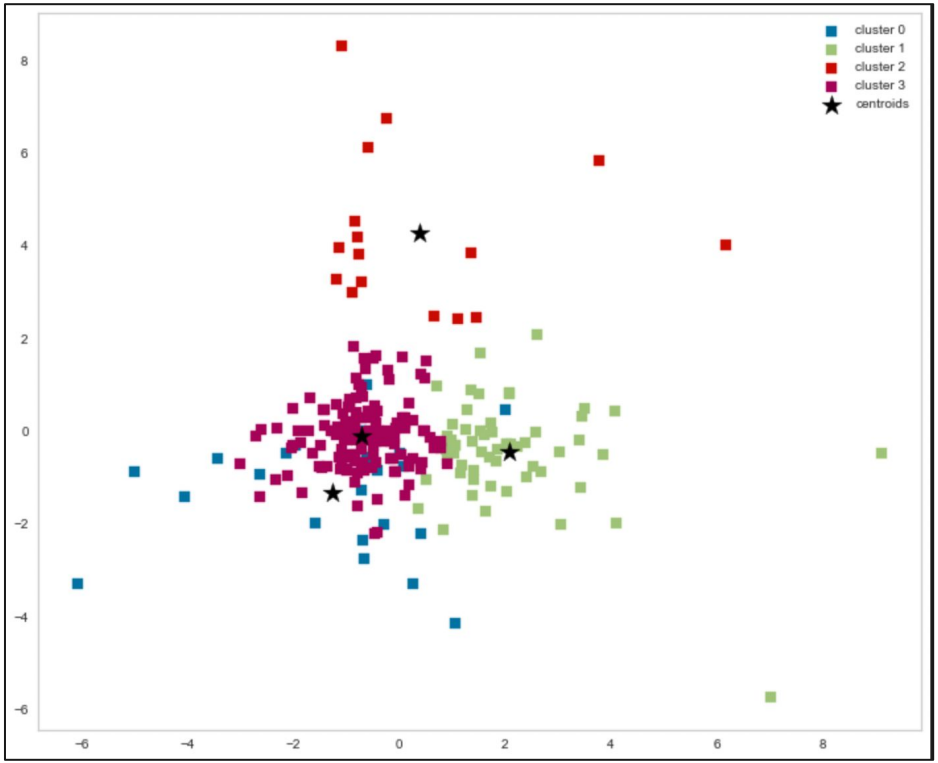
- Cluster 0 :
  - Customers with lowest income,
  - Youngest
  - Lowest number of children
  - Gender not distinct
  - Marital status 1
- Cluster 1 :
  - Customers with 2nd highest income
  - 2nd oldest
  - Mostly Male
  - Marital status 3
  - Highest number of children
- Cluster 2 :
  - Customers with highest income
  - Oldest
  - Mostly Male
  - Marital status 3
  - Low number of children
- Cluster 3 :
  - Customers with 3rd highest income
  - Small number of children
  - 3rd oldest
  - Gender & Marital status not distinct

*Marital status: 1 = Married, 2 = widowed, 3 = Separated, 4 = Divorced

Banking Behavior - Kmeans

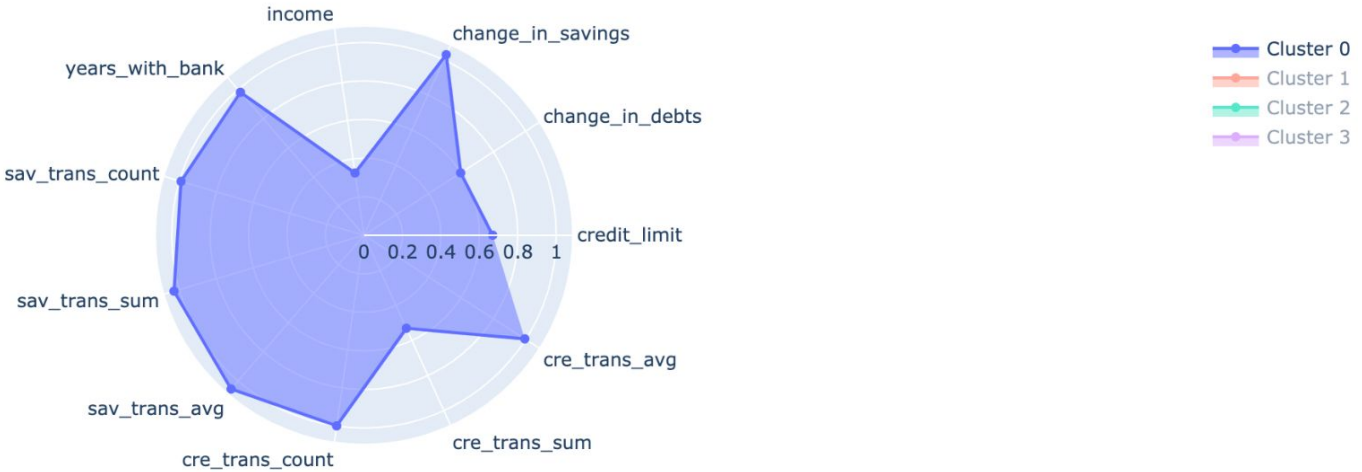


Banking Behavior



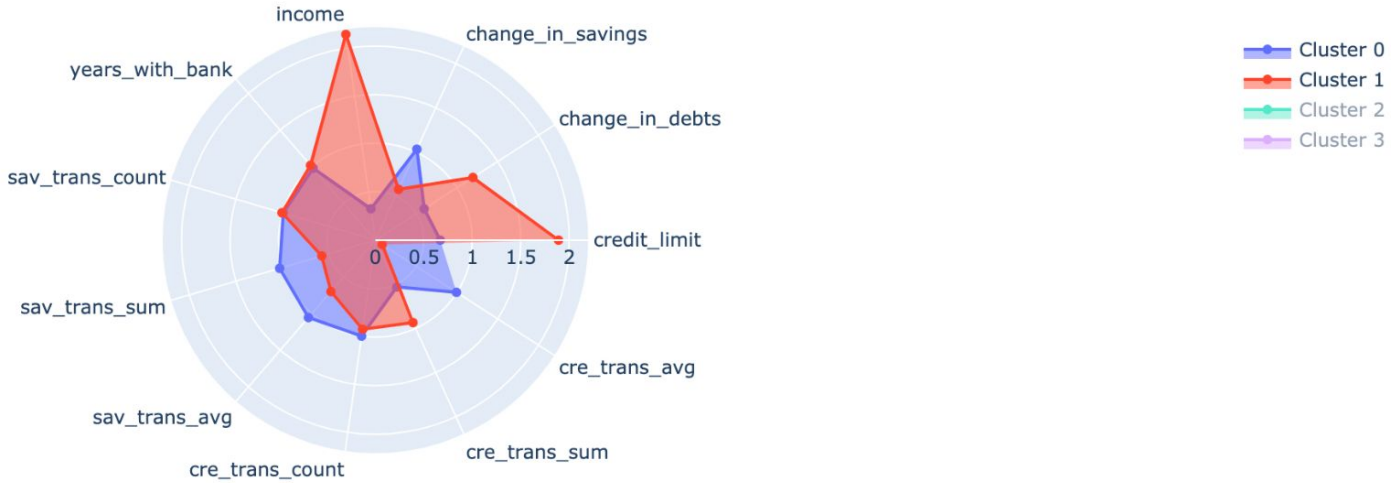
Banking Behavior - Cluster 0

Customer Banking Behavior Radar



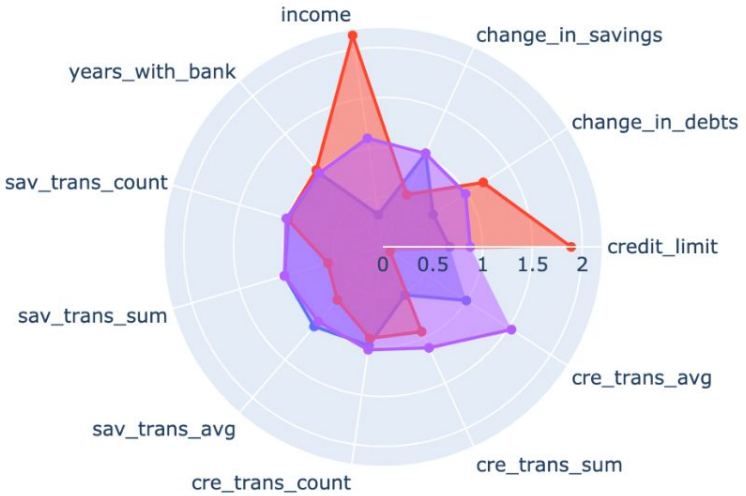
Banking Behavior - Cluster 1

Customer Banking Behavior Radar



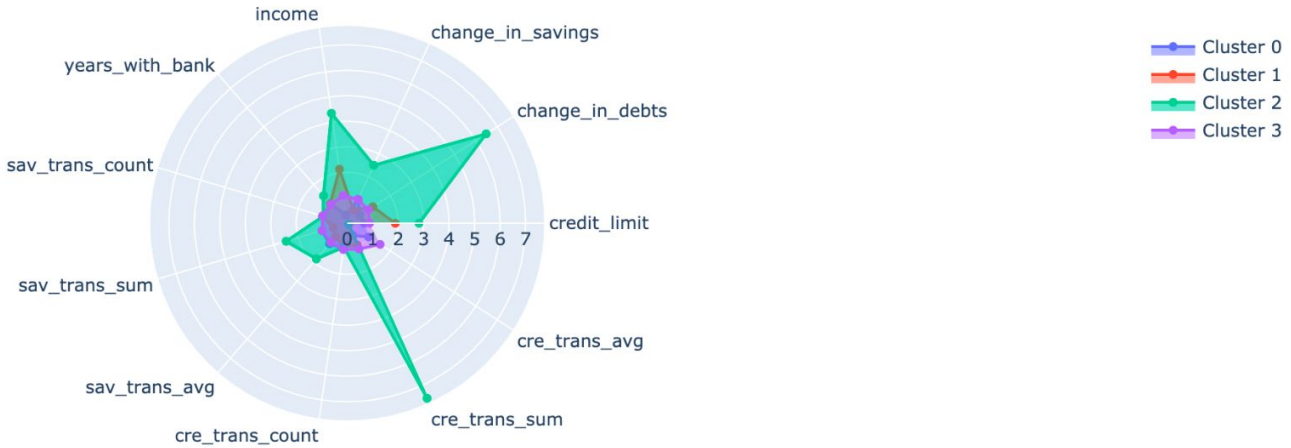
Banking Behavior - Cluster 3

Customer Banking Behavior Radar



Banking Behavior - Cluster 2

Customer Banking Behavior Radar



- Cluster 0 :
  - Customers who have lowest income
  - Medium spenders
  - Lowest credit limit.
- Cluster 1 :
  - Customers who have 2nd highest income,
  - Small spenders
  - Second high credit limit
- Cluster 2 :
  - Customers who have highest income
  - High spenders, highest credit limit
- Cluster 3 :
  - Customers who have 3rd highest income
  - Medium trans count
  - Low credit limit
- We can assume that customer Cluster 2 uses their credit cards the most
  - Highest debts and highest total credit card transaction amount

03

Challenges

Data Selection

Not domain knowledge so sometimes features can be confusing to select

Data Combination

Not enough time to explore more possible combination of features

Data Normalization

Standard or MinMax or Robust or Normalize

Radar Chart

Need to convert and normalize cluster output

04

Improvement

If we have more time..

Thank you!