



Customer Segmentation

11/09/2022



Problem Statement

Motivation:

XYZ bank wants to roll out Christmas offers to their customers. But Bank does not want to roll out same offer to all customers instead they want to roll out personalized offer to particular set of customers.

Objective:

Clustering the XYZ Bank customers into 5 groups by finding the pattern which group certain kind of customer in one category.

Data Set

Total number of observations	1000000
Total number of files	1
Total number of features	47
Base format of the file	csv
Size of the data	366.2+MB



Data types:

Float type features:9

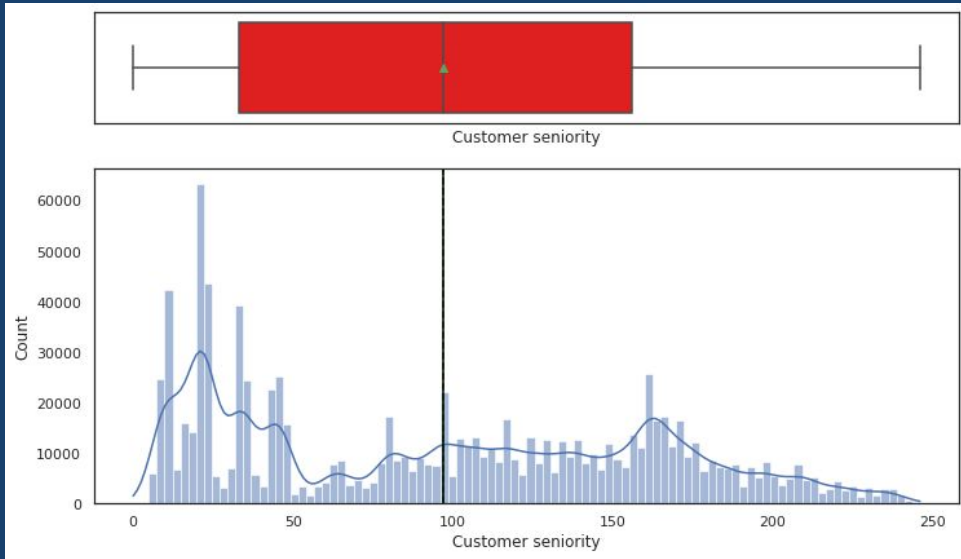
Int type features:24

Object type features:15

Memory usage: 366.2+ MB

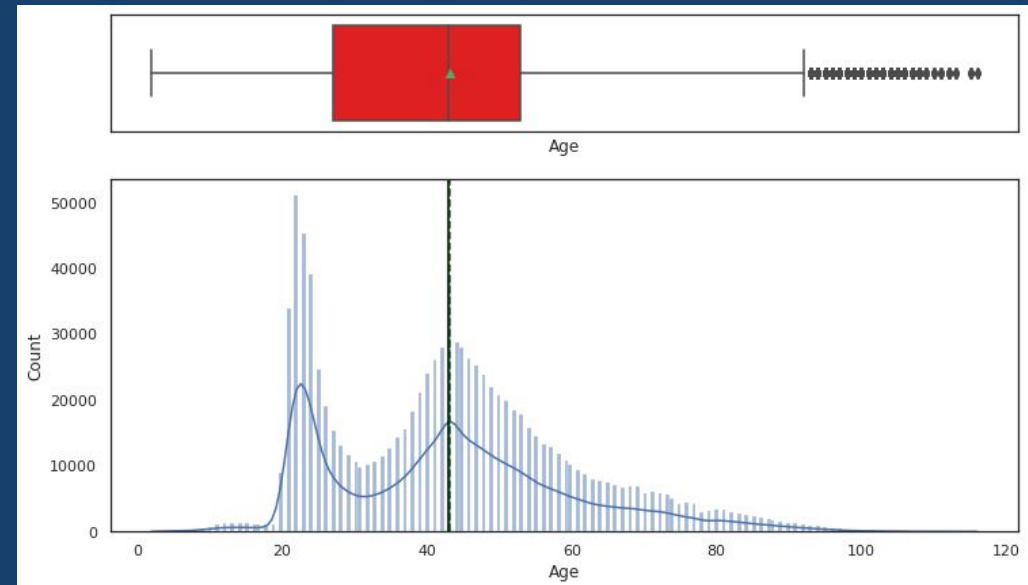
Exploratory Data Analysis

Univariate Analysis



The customer seniority is bimodal . With two peaks at 20 month and 170 months. The mean and median of Customer seniority is 100 month.

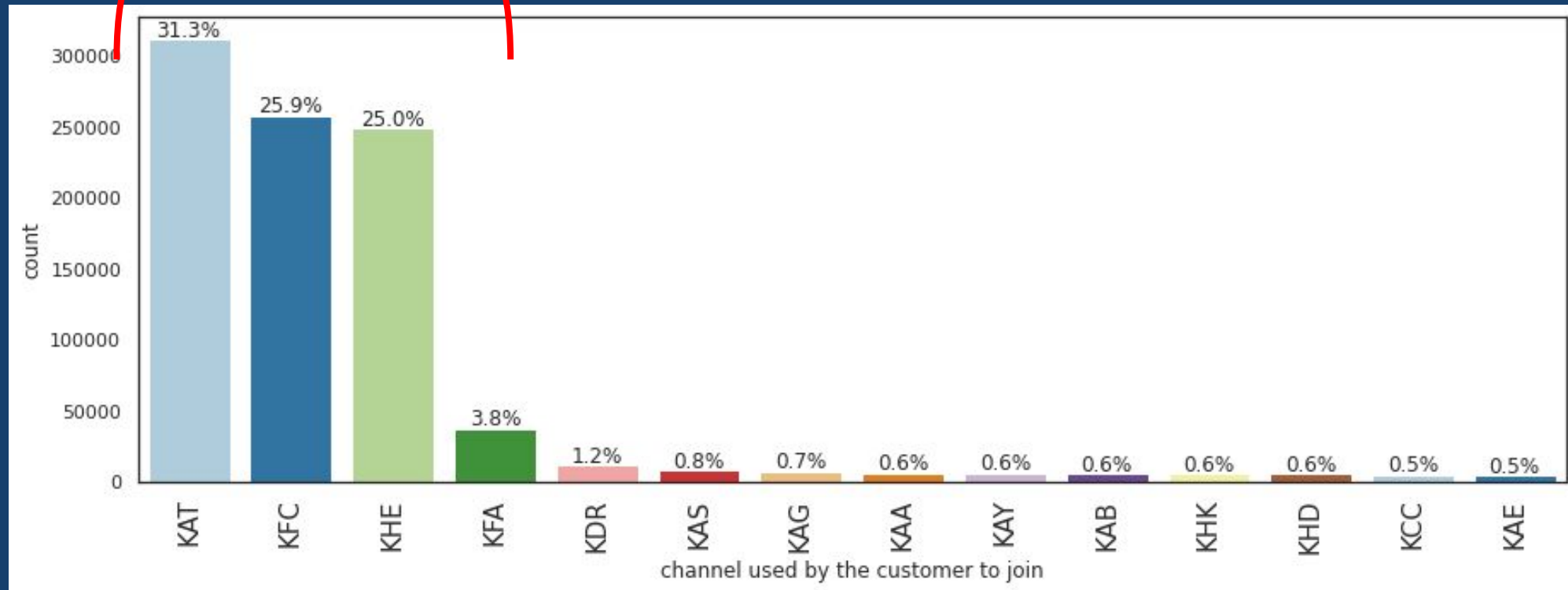
The Age is bimodal . With two peaks at 20 year month and 40 year. The mean and median of Customer seniority is 40 year



Exploratory Data Analysis

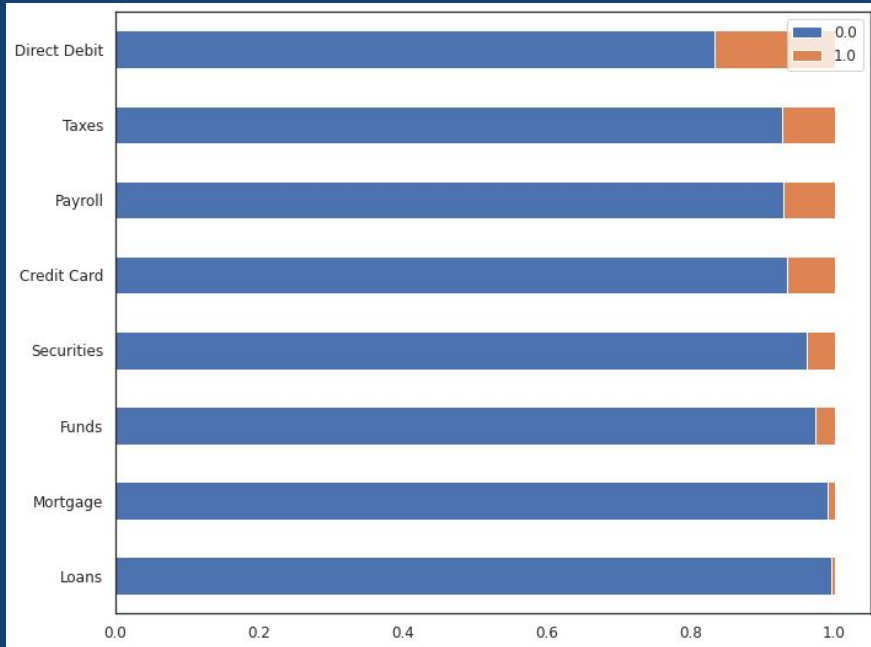
Univariate Analysis

80% of the customers are in these 4 channels



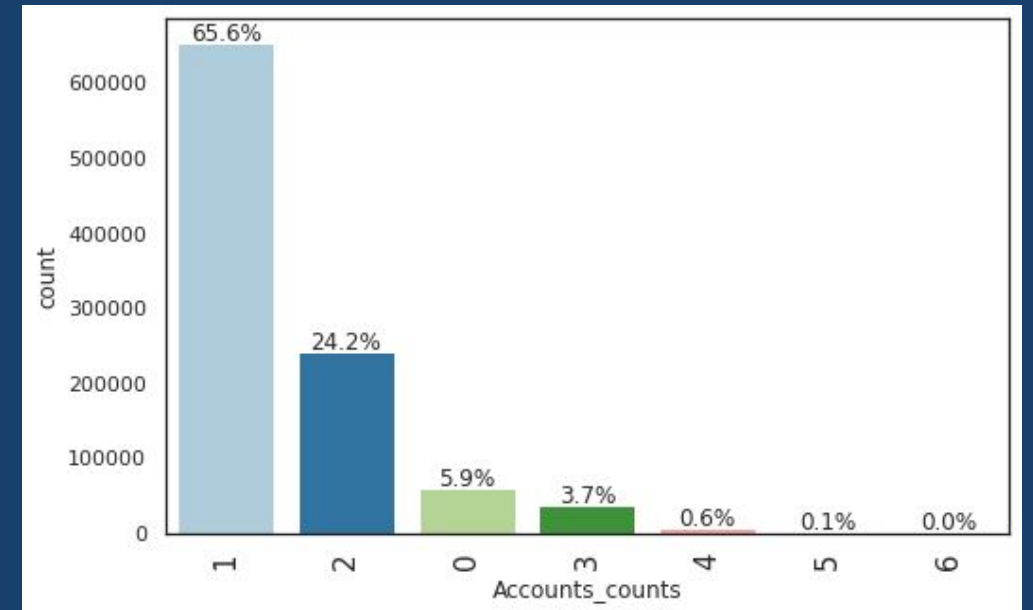
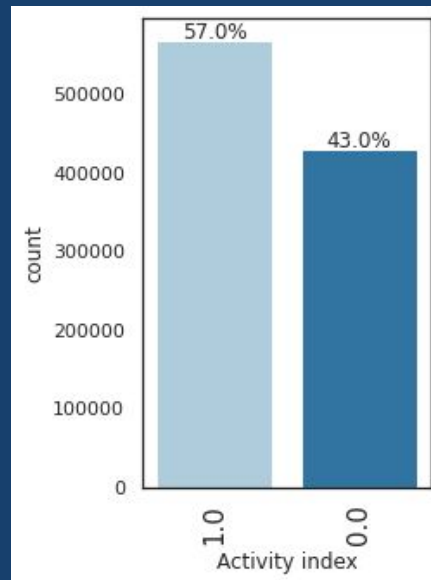
Exploratory Data Analysis

Univariate Analysis



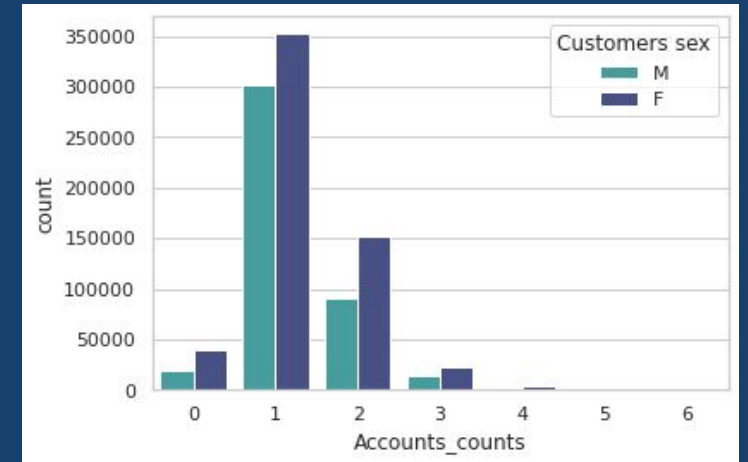
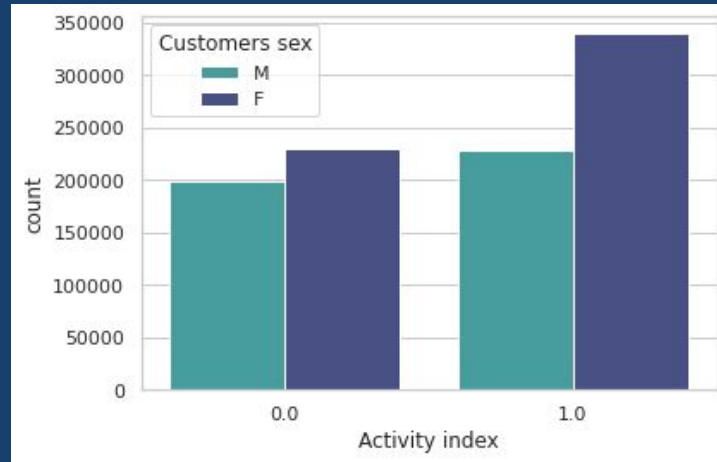
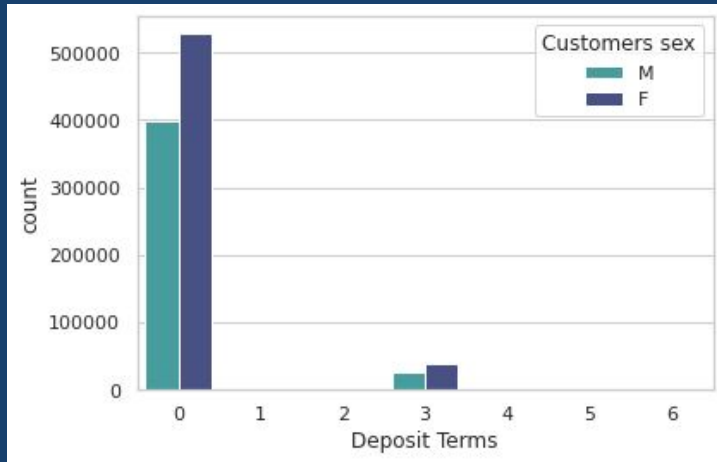
- 56% of accounts is active.
- 92% of accounts are payroll.
- 98% are residence in the same place of account.
- 4.3% are foreigner accounts.
- 2.7% of the account are for funds, 1% for mortgage
- 0.5% for loans, 7.3% for taxes, 6.6 for Credit Card
- 3.9 for securities, 16.7% for direct debit.

- 93% has no term deposit, 6.3% has long term deposit
- 5.9% have no accounts and 65% of customers has one accounts while 24% has 3



Exploratory Data Analysis

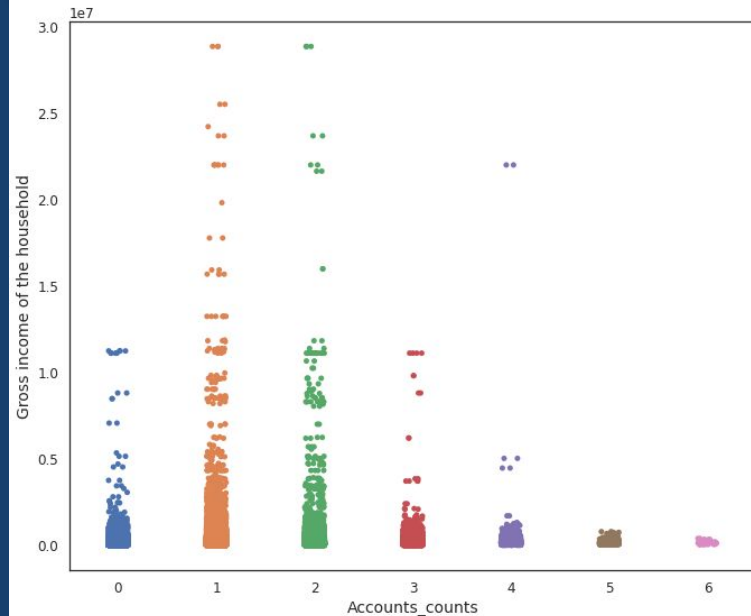
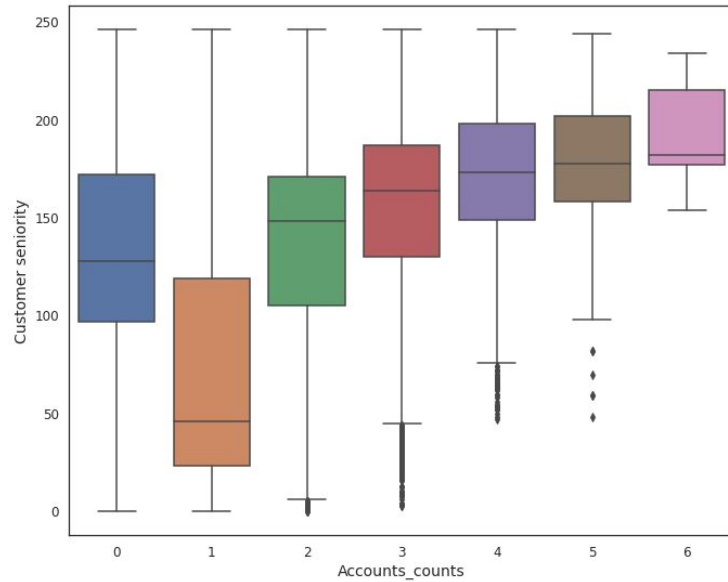
Bivariate Analysis



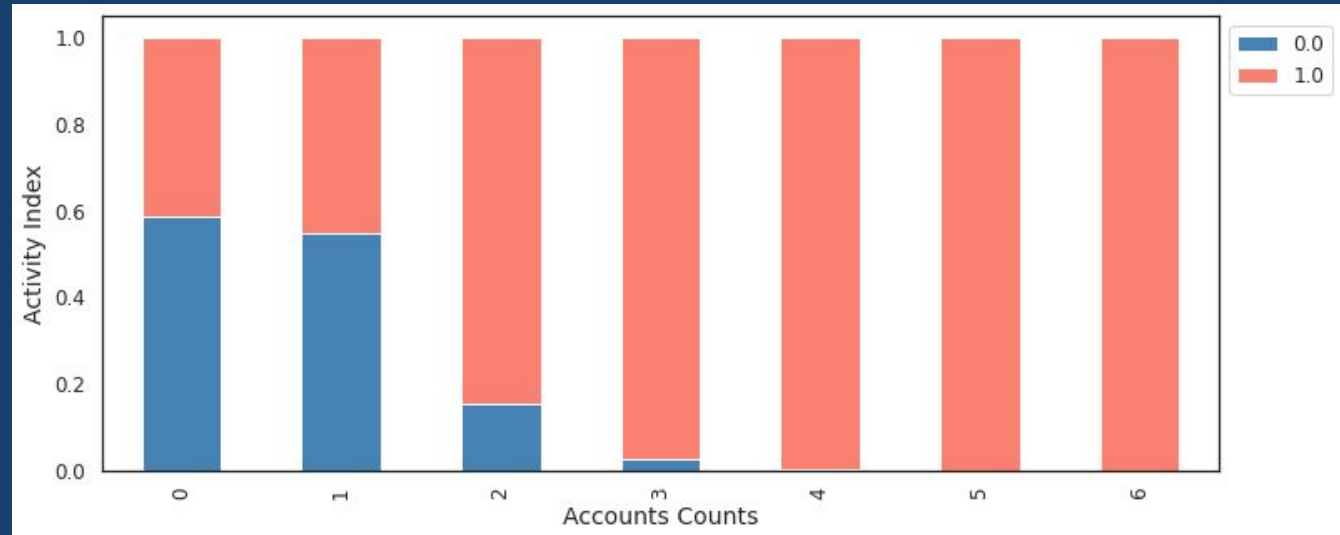
- Female customers to male customers ratio are higher for active accounts than in active accounts.
- Female customers to male customers ratio are higher for two account counts than one account and three accounts.
- For customers without any deposit terms the ratio of female to male is much higher than the ratio of other deposit terms.

Exploratory Data Analysis

Bivariate Analysis

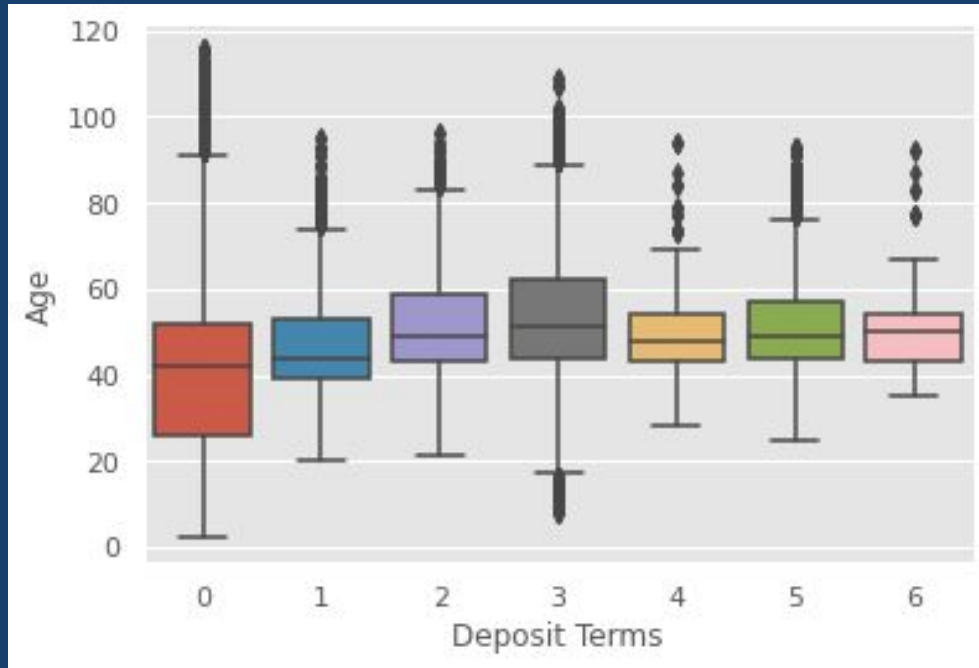


- The more seniority the customer is the more bank accounts they have.
- Customers with at less 3 accounts, they are active customers.
- Customer with 1 and 2 accounts have the high income gross.



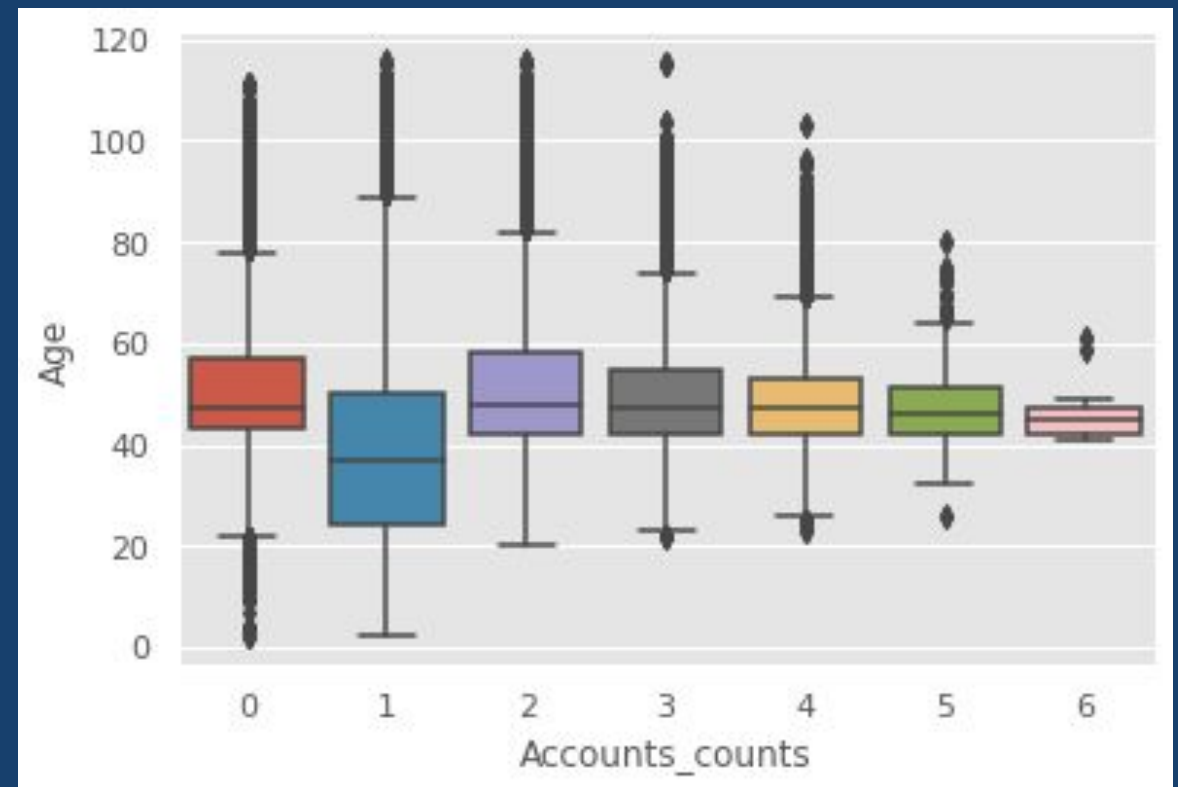
Exploratory Data Analysis

Bivariate Analysis



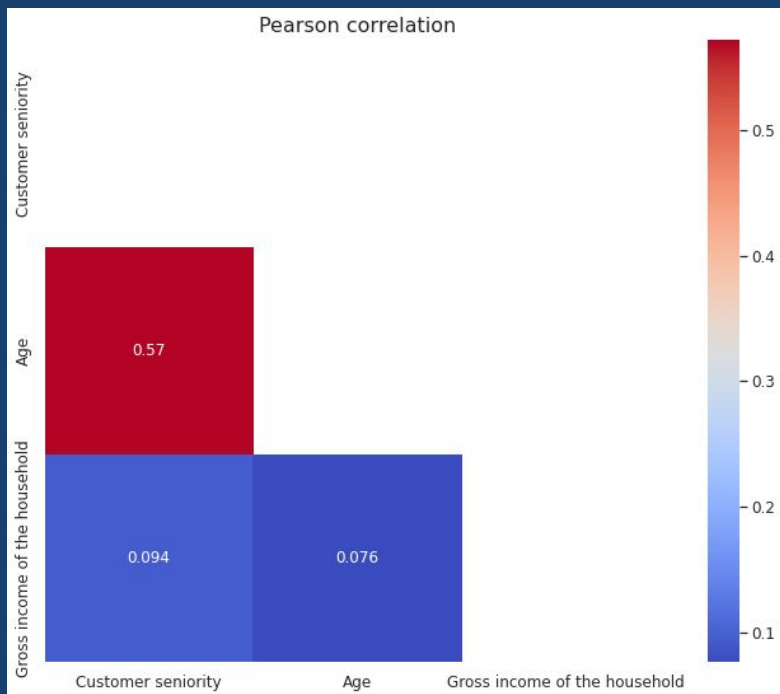
Customers with long term deposit have the higher median age 50 year. While customers age doesn't have relationship with customer with multiple deposits terms.

Age of customers doesn't have relations to how many account you have. However, customers with median age 40 have one account

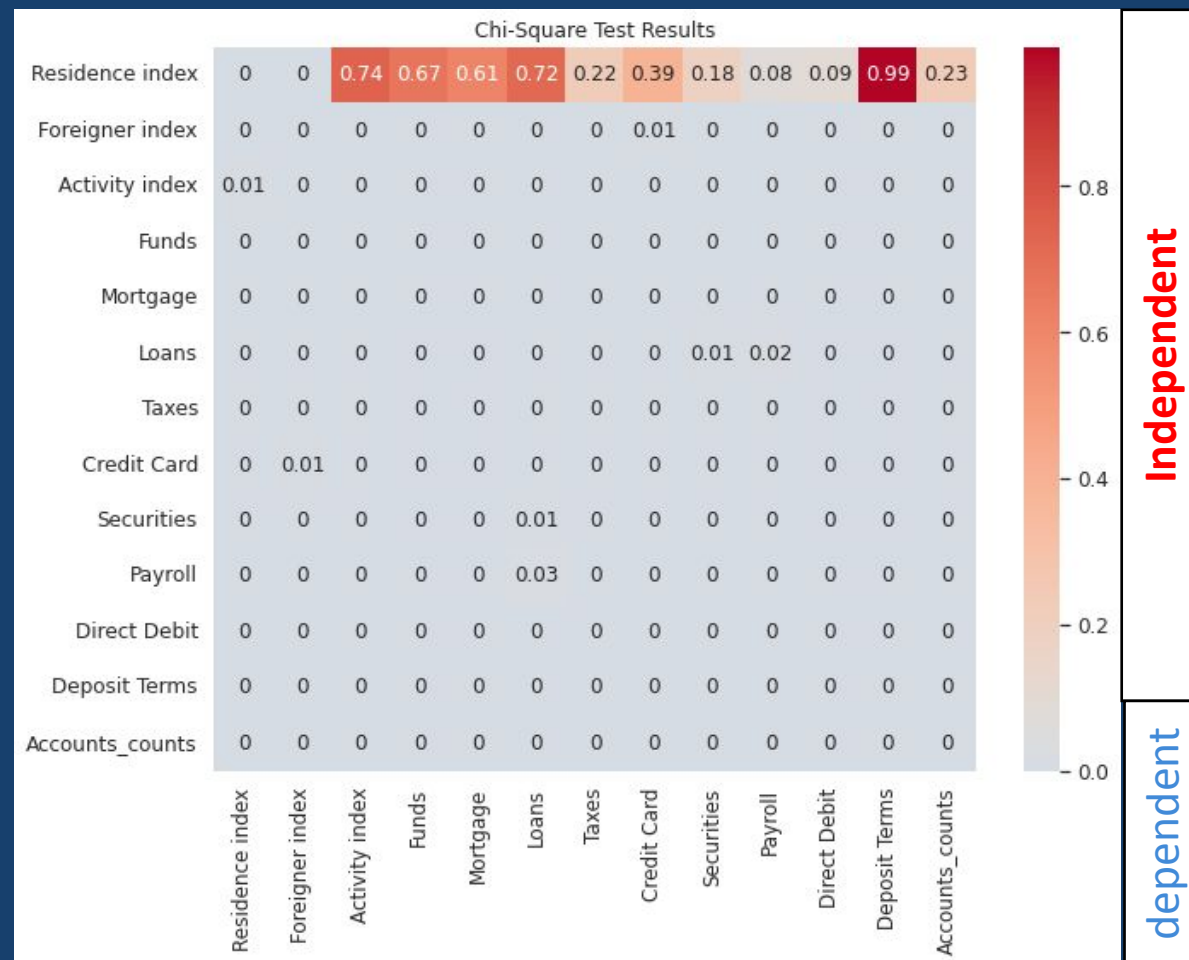


Exploratory Data Analysis

Bivariate Analysis



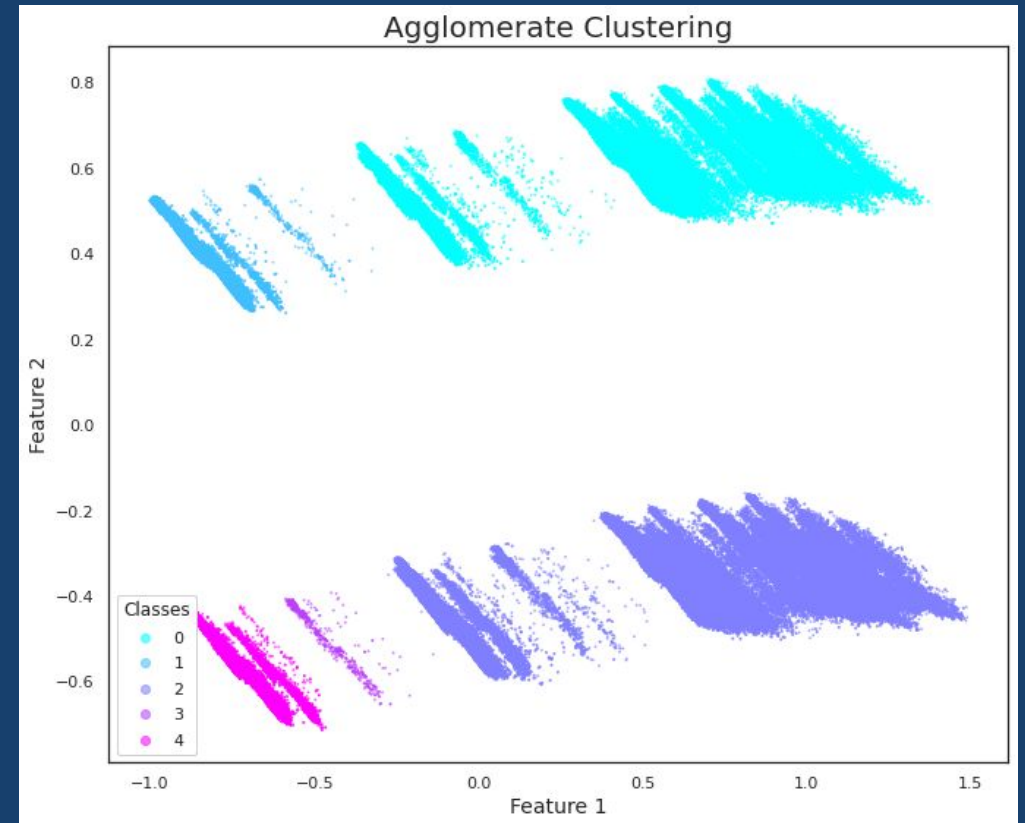
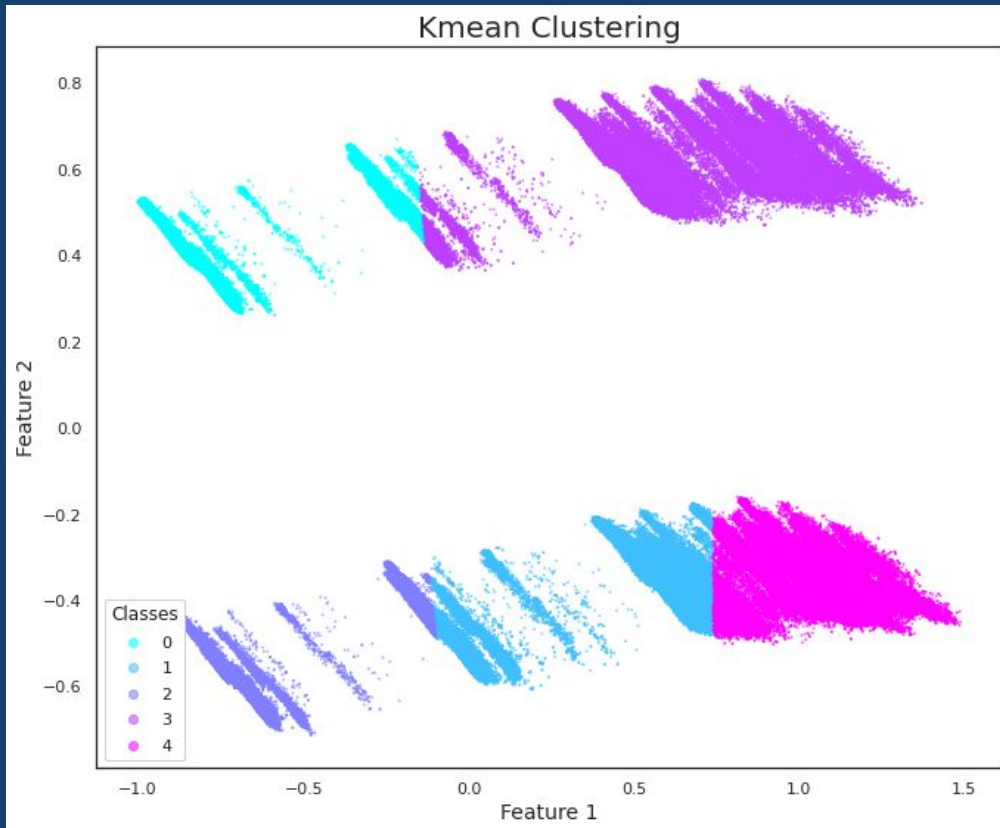
- No correlation between the numerical features.
- The Chi2 p-values < 0.05 represent features depended on each others, while the value > 0.05 are independent features.



Clustering models

Kmeans and AgglomerativeClustering from RAPIDS CuML was applied. The results shows that agglomerate clustering is more efficient in segmenting the data.

Winner



Insight and Recommendation

- Female customers to male customers ratio are higher for active accounts than inactive accounts.
- Customer with 1 and 2 accounts have the high income gross.
- Age of customers doesn't have relations to how many account you have.
However, customers with median age 40 have one account.
- The more seniority the customer is the more bank accounts they have.
- Customers with at less 3 accounts, they are active customers.
- **80% of the customers are in these 4 channels, KAT,KFC,KHE, and KFA**