

Answers for task 5.4: Intro to data mining

By: Alexandra Lindsay

Question 2: Data cleaning

- Some last names have “?” in them. They have been removed.
- Some genders have abbreviations put them all as either female or male or null.
- The countries had had mix data. Some countries abbreviations corrected them all to full names.
- No duplicates found.
- 1 entry empty in the last name changed to null.
- 3 empty credit scores changed to null.
- 1 null gender.
- 1 null age and 11 with age 2. Didn't remove the age 2 due to possibility of parents opening accounts for their children.
- 1 null and 1 blank estimated salary changed to null.
- Didn't remove any null data due to it not impacting the integrity of the analysis.
- removed the row number column since it was not relevant and the last name column for customer privacy.

Question 3: identify the top 3 to 4 factors that lead to clients leaving

Data gathering:

Left the bank:

Column	Average	Max	Min
Credit score	643	850	376
Age	45	69	2
Tenure	4.7	10	0
Balance	\$ 90,239.22	\$ 213,146.20	\$ -
Number of products	1.4	4	1
estimated salary	\$ 19,819,661.74	\$ 199,725.39	\$ 417.41

Column	Criteria	count
Country	France	77
	Germany	75
	Spain	52
Gender	Female	121
	Male	83
Number of products	1	142
	2	32
	3	28
	4	2
customer count		204
has credit card	yes	144

	no	60
is an active member	yes	143
	no	61

To note:

- average age: 45 max: 69
- Tenure average: 4.7 max: 10
- Average balance: \$90,239.22
- Average number of products: 1.4 max: 4
- Most have a credit card
- Most are active members

Stayed with the bank:

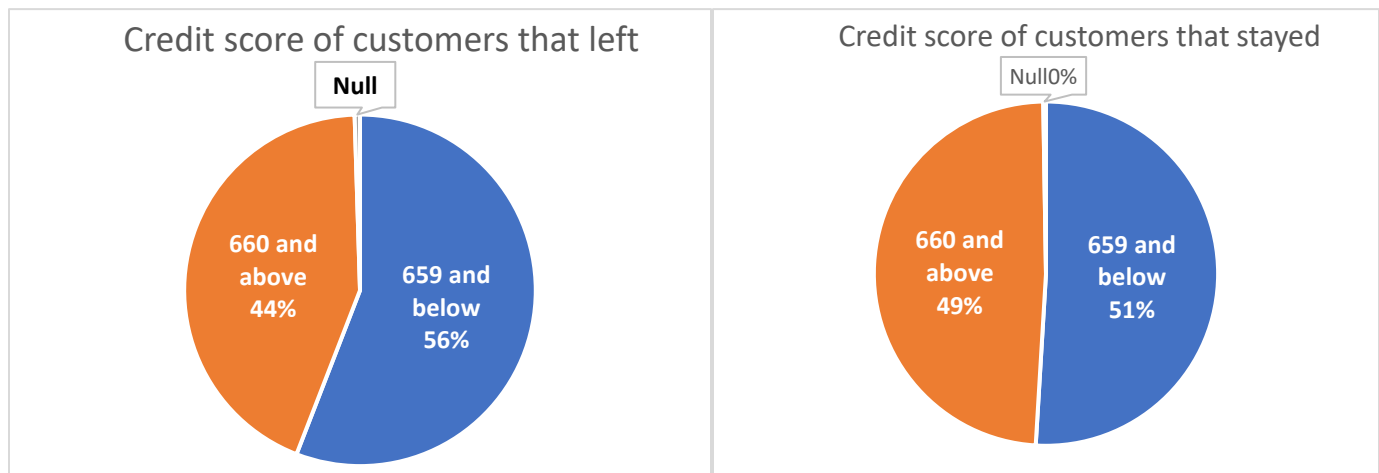
Column	Average	Max	Min
Credit score	652	850	411
Age	37	82	2
Tenure	5	10	0
Balance	\$ 74,830.87	\$ 197,041.80	\$ -
Number of products	2	3	1
estimated salary	\$ 98,943.39	\$ 199,661.50	\$ 371.05

Column	Criteria	count
Country	France	403
	Germany	182
	Spain	202
Gender	Female	341
	Male	445
	Null	1
Number of products	1	368
	2	414
	3	5
customer count		787
has credit card	yes	556
	no	231
is an active member	yes	442
	no	345

To note:

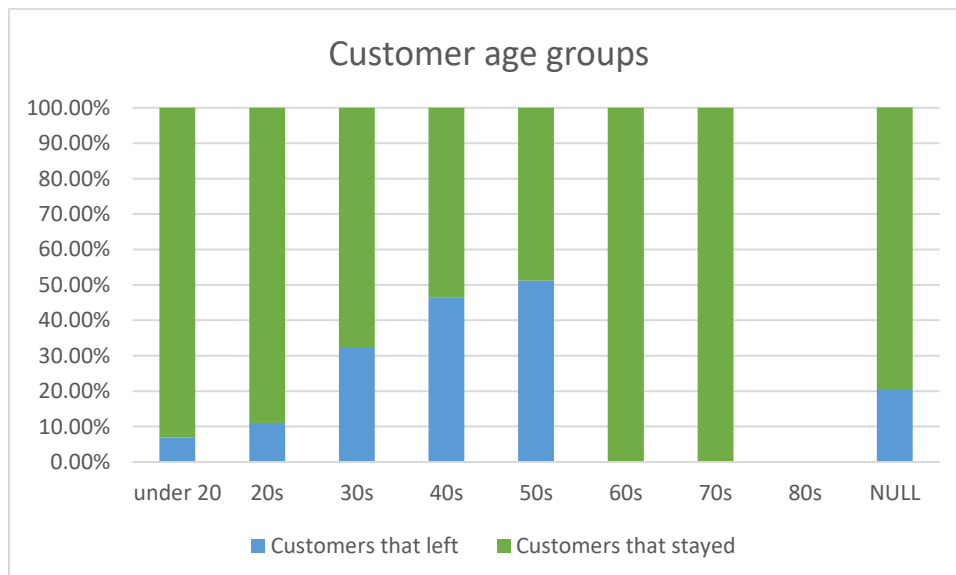
- average age: 37 max: 82
- Tenure average: 5 max: 10
- Average balance: \$74,830.87
- Average number of products: 2 max: 3
- Most have a credit card
- Most are active members

Credit score:

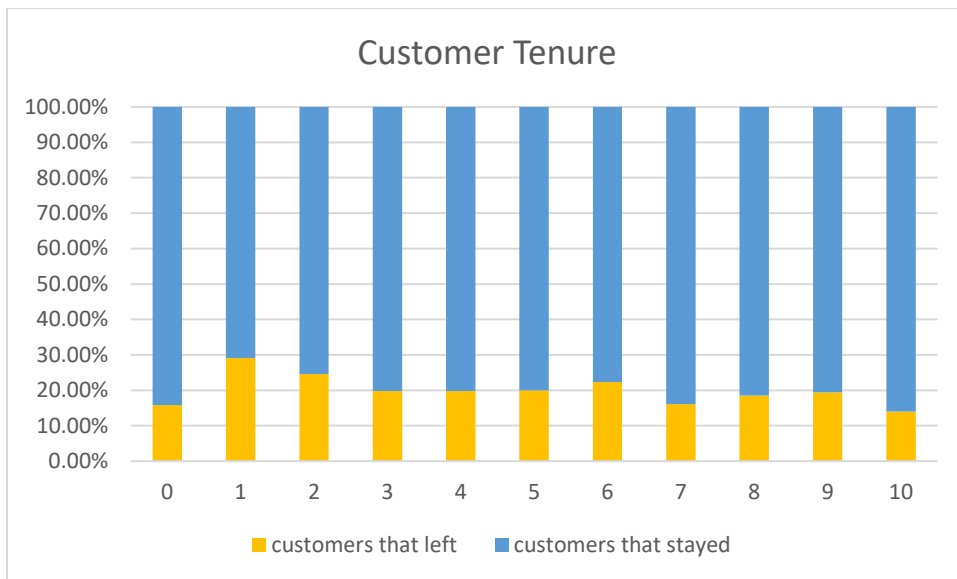


Both customers that left and stayed have the same credit level: below 660 (good credit level establish by Equifax in Canada)

Age Vs Tenure:

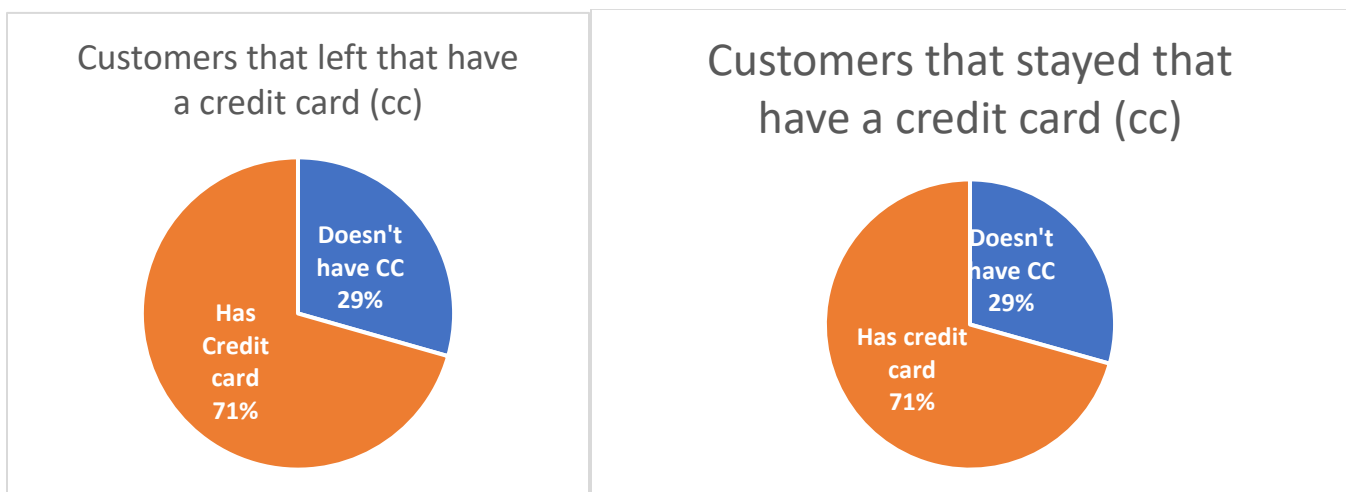
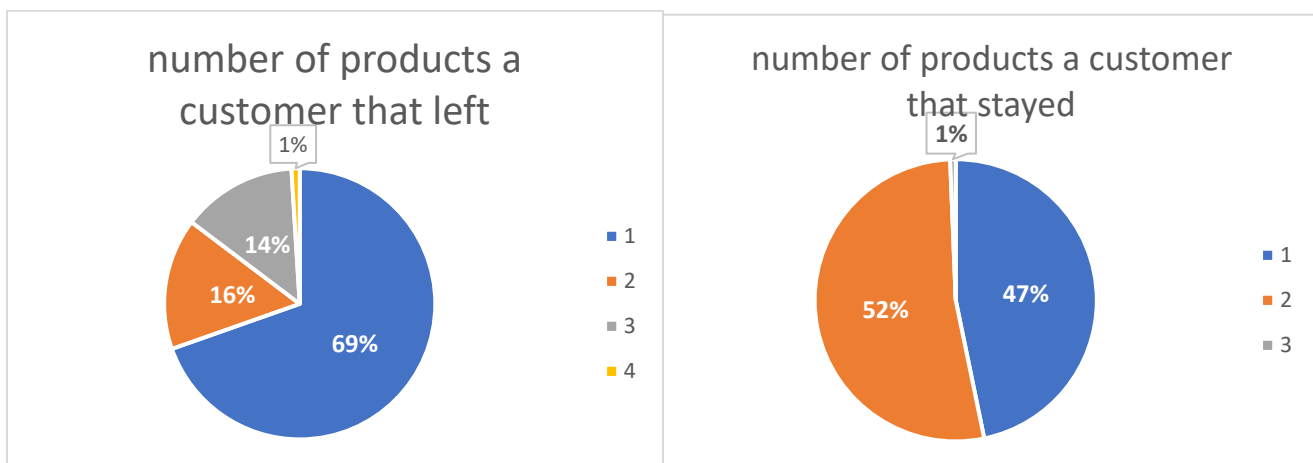


Most customers that left are in their 30s-50s. and those that stayed from 20s to 40s. typically those customers in their 20s and 60s and up do not take investment risks and those between 30s-50s tend to take investment risks being in their prime working years, saving for investments.



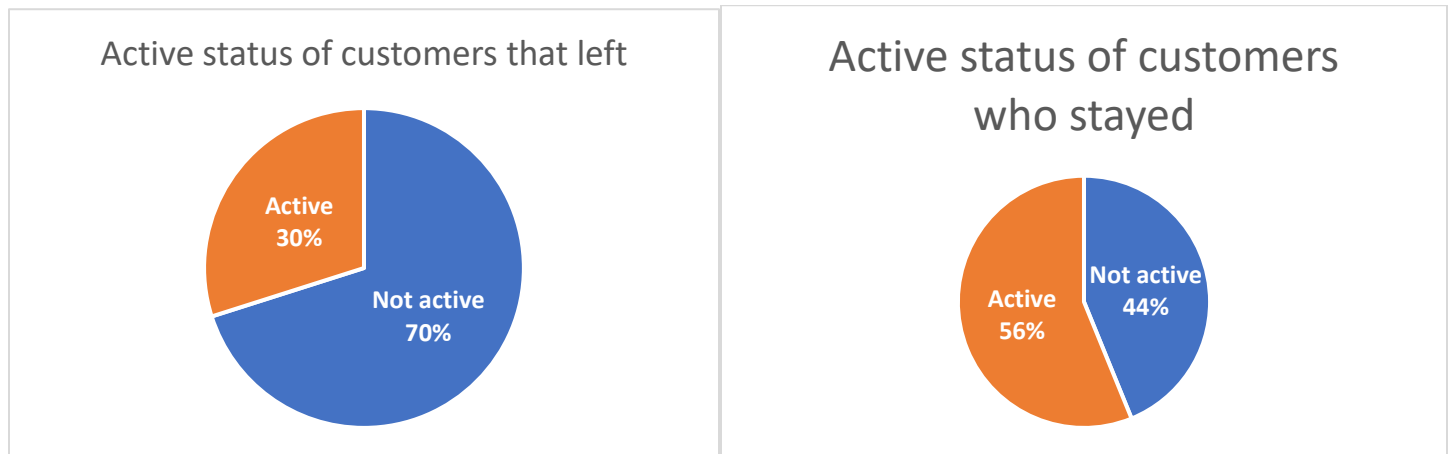
Keeping in mind the investment ages stated before, the top two years for those that left are 1 and 2. And for those that stayed total average out at 72 years. Therefore it is safe to assume that those that left came for investment purposes.

Number of products and Has credit card?:



Most customers now a days have a credit card but if we look at the number of products for those that left, most only took 1 product and those that left most had 2. Meaning those that left only took the product that was most beneficial to them.

Is active member?:



Looking at those that left, since most were not active, it is safe to assume they left a deposit of money leaving in an account like a TFSA or a high-risk investment portfolio to increase their investment and took it out.

Without more information like what products they bought and what were the account balances when they withdrew the money, looking at:

- The average account balance between those that left and stayed only different \$20,000.
- The age groups/ tenure of the customers that left that they came to this bank for short term goals possibly investment.
- Since those that left also only mainly took 1 product this would seem they only wanted the product that benefited them the most in that short amount of time.
- The active status of those that left infers that they only came to get that one product and withdrew their money when they achieved their goals.

In conclusion, the customers that left came with a goal in mind and left when they achieved it. More information would be needed to ascertain why they chose our bank at that specific time (economic situation would help the analysis) and when they left.

Question 4: Decision tree

