

HOXHUNT SUMMER TRAINEE 2021

BANK CHURNERS ANALYSIS

Prepared by Nhu Nguyen



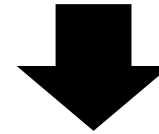
PROBLEMS

- Find factors that have the highest linkage to churning customers
- Find the most likely customers at highest risk of churning



SOLUTIONS

Exploratory Data Analysis



Build a prediction model

Summary of the data

21
columns

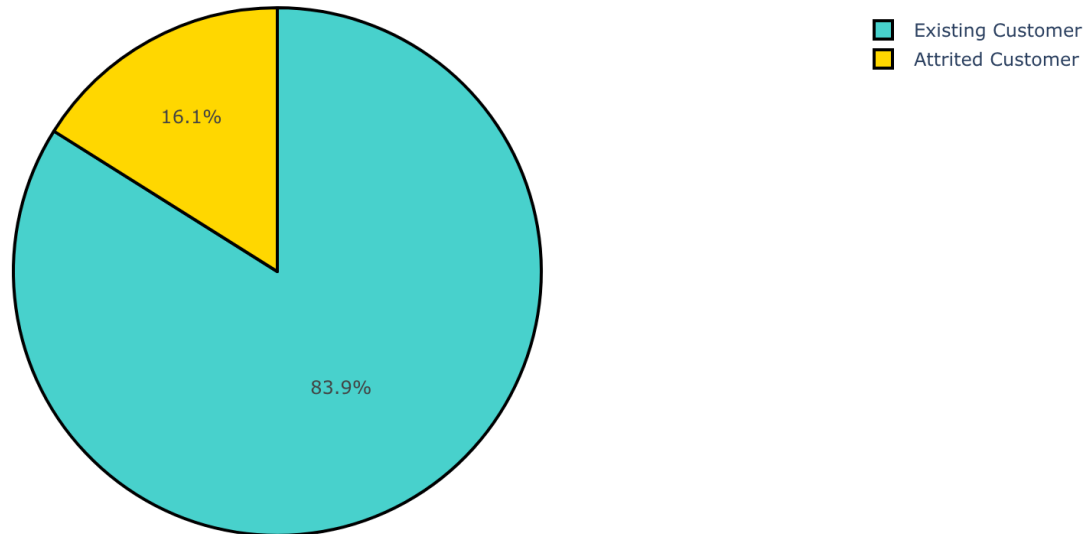
10127
Rows

0
Null values

0
Duplicated
values

Target variable

Attrited customer vs Existing customer



1627

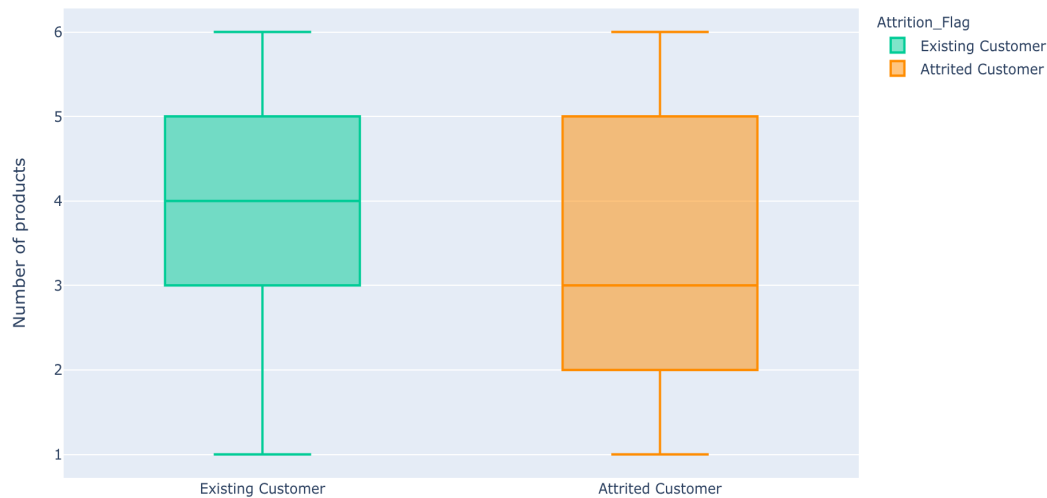
Attrited Customers

8500

Existing Customers

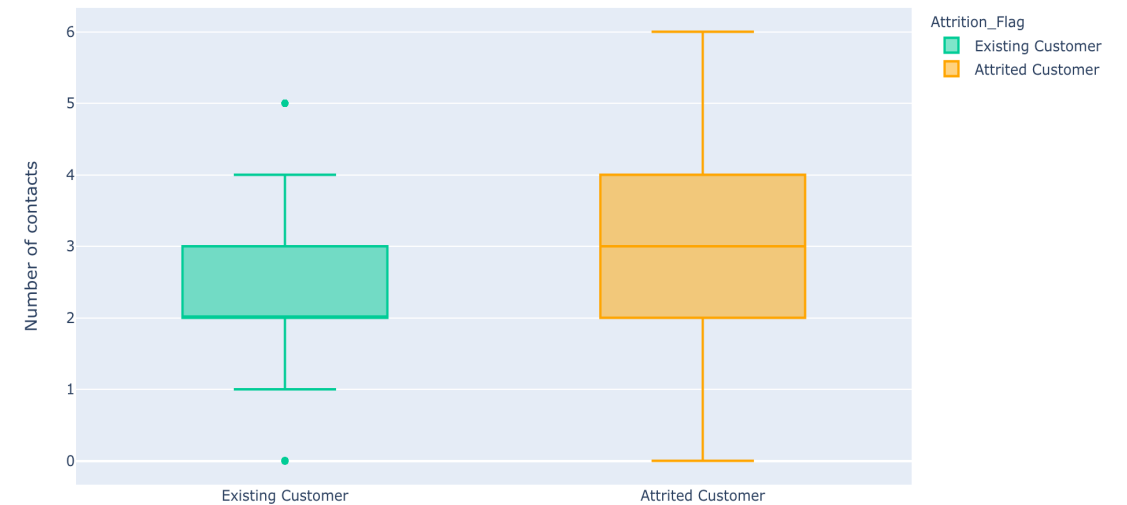
PRODUCT VARIABLES

Number of products held by customers

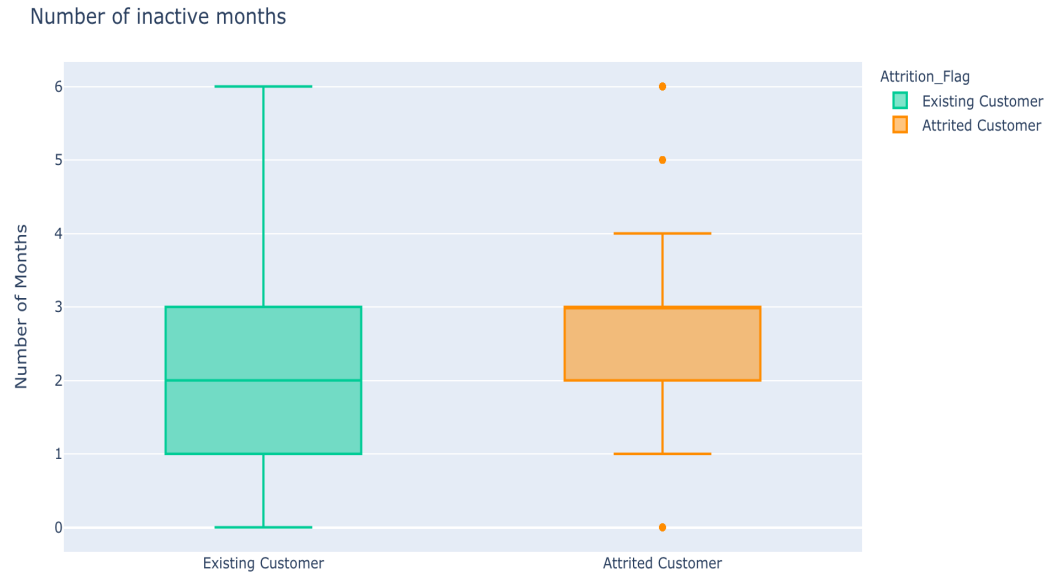


- Existing customers own 4 products on average
- Attrited customers own 3 products on average.

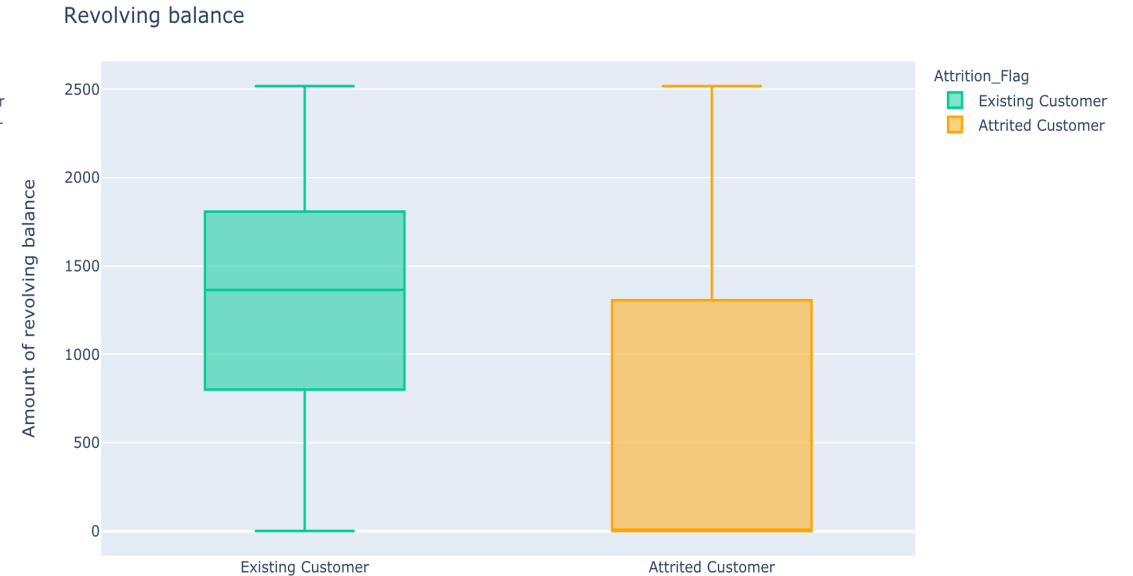
Number of contacts in last 12 months



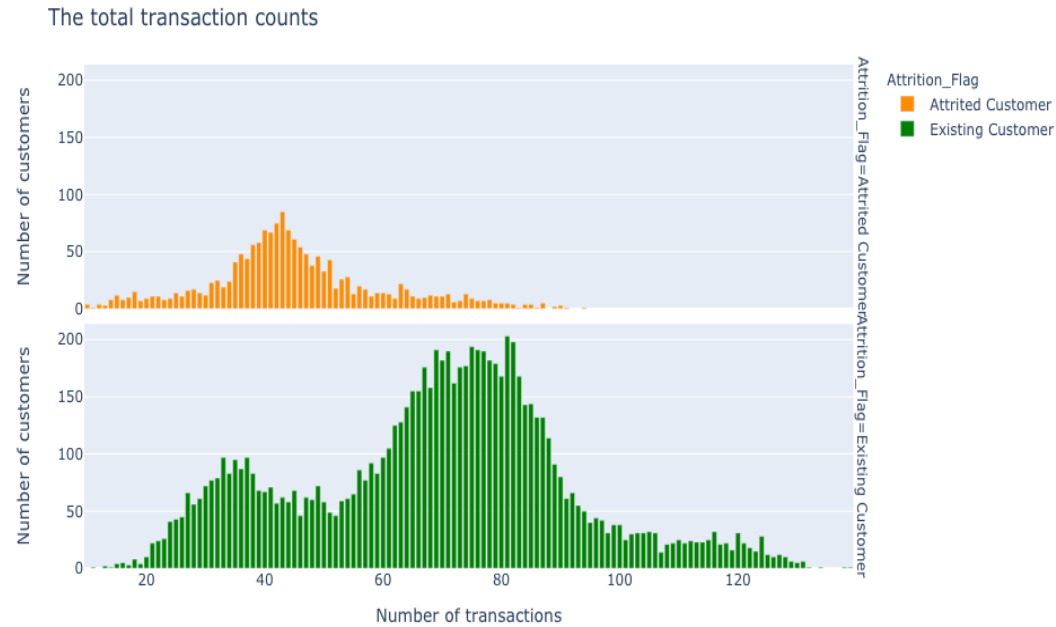
- Existing customers have 2 contacts on average with two outliers
- Attrited customers have 3 contacts on average



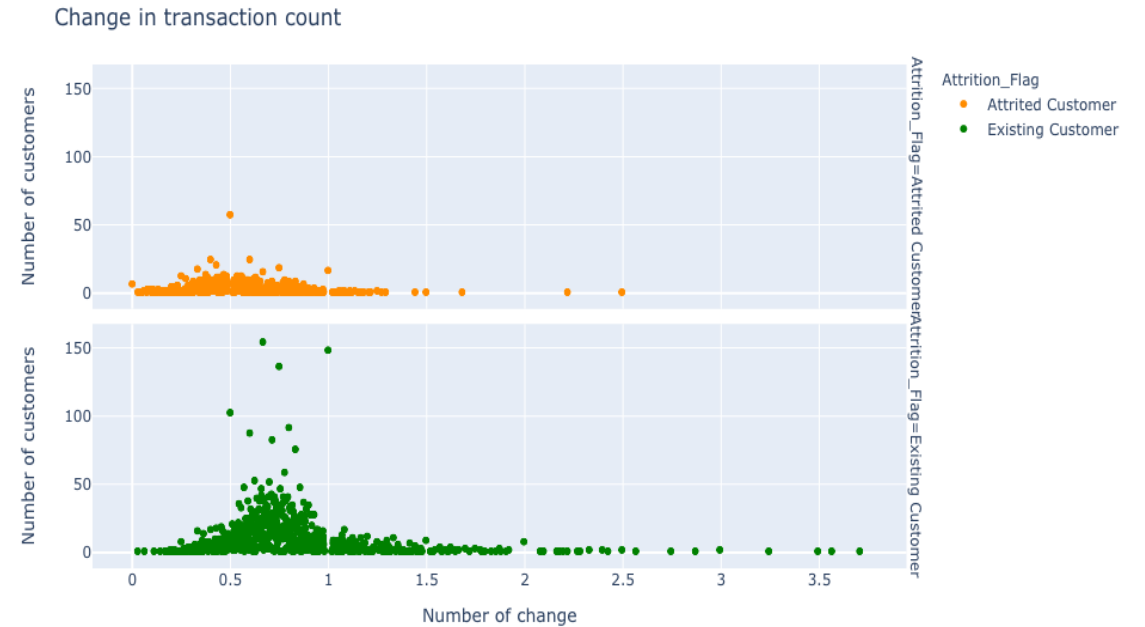
- The average inactive months of existing customers is 2
- The average of attrited customers is 3.
- Outliers show in attrited group



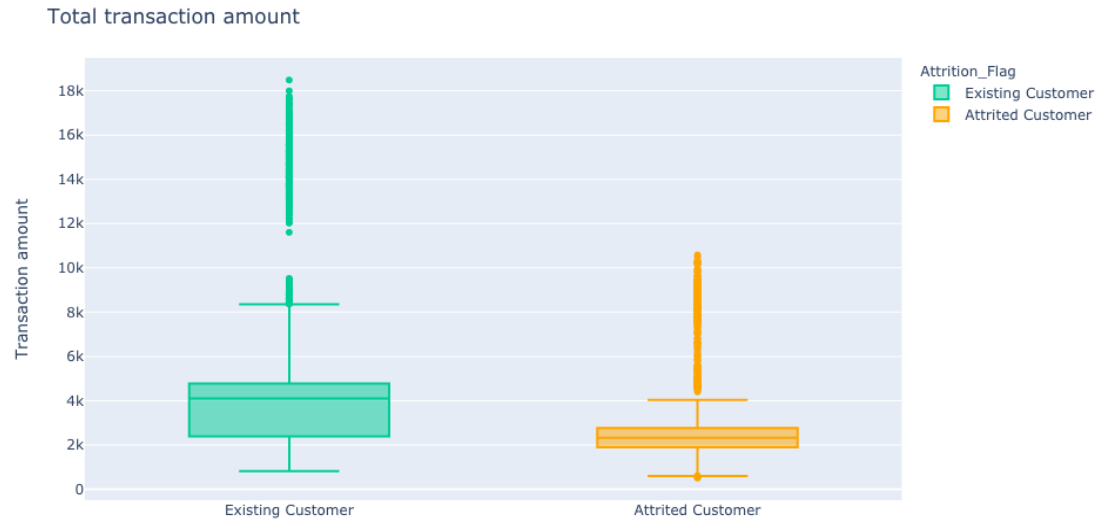
- Existing customers have high revolving balance (around 1400)
- Attrited customers have reached the lowest point (near 0)



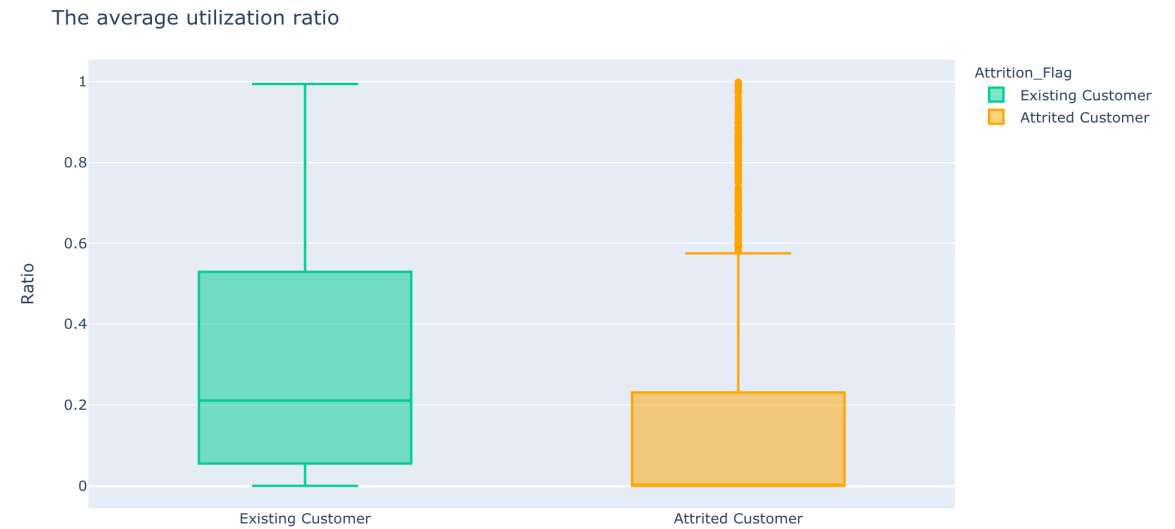
- Existing customers have around 70-90 total number of transaction
- Attrited customers have around 30-50



- The number of changes in transaction of existing customers is around 0.5-1 as the highest.
- 0.3-0.7 for attrited group



- Total transaction amount of existing customers is around 4000.
- Its of attrited customers is around 2000
- Outliers detected in both groups



- The utilization ratio of existing group is 0.2 on average.
- Attrited customers group reaches the lowest point 0.
- Outliers lie in attrited customers group



The factors that have the highest linkage to churning customers (with the summarized average values)

	Attrited Customer
Number of products	3
Inactive months	3
Number of contacts	3
Revolving balance	0
Total transaction amount	2229
Total transaction count	43
Total count change	0.53
Utilization ratio	0

For example: A customer who has 3 number of products, 3 inactive months and the total transaction amount around euros, etc will have the high risk of churning.

PREDICTION MODEL

Confusion matrix and classification report of Random Forest prediction model

```
[[1657   28]
 [   50 1665]]
```

Accuracy on test set: 0.98

	precision	recall	f1-score	support
0	0.98	0.97	0.98	1707
1	0.97	0.98	0.98	1693
accuracy			0.98	3400
macro avg	0.98	0.98	0.98	3400
weighted avg	0.98	0.98	0.98	3400

98%

Correct prediction

RECOMMENDATIONS

It is necessary to pay attention in changing in product variables. The customer who most likely intend to leave will have:

- Low number of products (2-5 average products)
- High inactive months (average 2-3 months)
- High number of contacts (2-4 contacts on average)
- Low revolving balance amount (under 600)
- Low in total transaction amount (under 2300)
- The total transaction range from 30-50 times
- Change in transaction range from 0.25-0.75 times.
- Low utilization ratio (almost 0)

Use the prediction model to predict the customer at highest risk of churning (the accuracy reached 98%).

THANK YOU!