

Ayiti Analytics

BI Project II / Fraudulent credit cards

Group_3 Team Members:

- Wedsanley JEAN PHILLIPE
- Djeninnah THIMOTHEE
- Hardiles THERMORIS

I. Understanding the Objective

Business Problem:

The main problem of the bank is the complaint that it receives from the customers about the loss that they suffer, the second problem is the fact that the media reports that the bank does not do any follow-up to solve this difficulty

Since the customers are losing a large amount of money, it is damaging the image of the company.

Currently, the bank is not able to provide better customer service to protect the financial assets of customers. As a result, it receives a lot of complaints from customers about losing their money fraudulently.

As a group of analysts, we will analyze the 6 months of available data and suggest the best solutions to solve this problem.

Background of the problem:

The problem was identified by the bank's customers. The bank did not respond, so customers went to the media to report their losses

Conceptual Model

- The Client:

Sans les clients il n'y a pas de banque. Pour qu'il y ait

Pour pouvoir sauvegarder l'image de la compagnie il va falloir regagner la confiance de ses clients

En attendant que nous proposons un modèle pour les transactions frauduleuses et des suggestions pour combattre le problème.

The Stakeholders:

The main people affected by the problem are the Clients, the Bank and the media.

- 1) **Customers:** They are affected due to the loss of large sums of money.
- 2) **The Bank:** The unreliability within the bank, can lead to a general loss of confidence in the system. From the depositors' point of view, the stability of the financial markets is equally important. The bank risks losing its customers because it has not been able to protect them against active financial risks. Customers may decide to leave the bank and go to a more reliable bank. It is necessary that customers have absolute confidence in the chosen bank. The credibility of the bank is questioned. No investor wants to invest his money in a company that loses its customers, so customers will have to invest to make a profit, otherwise the bank will fail.
- 3) **The media:** Given the impact of the problem, it could also have repercussions on the whole economy. which will not spare the media

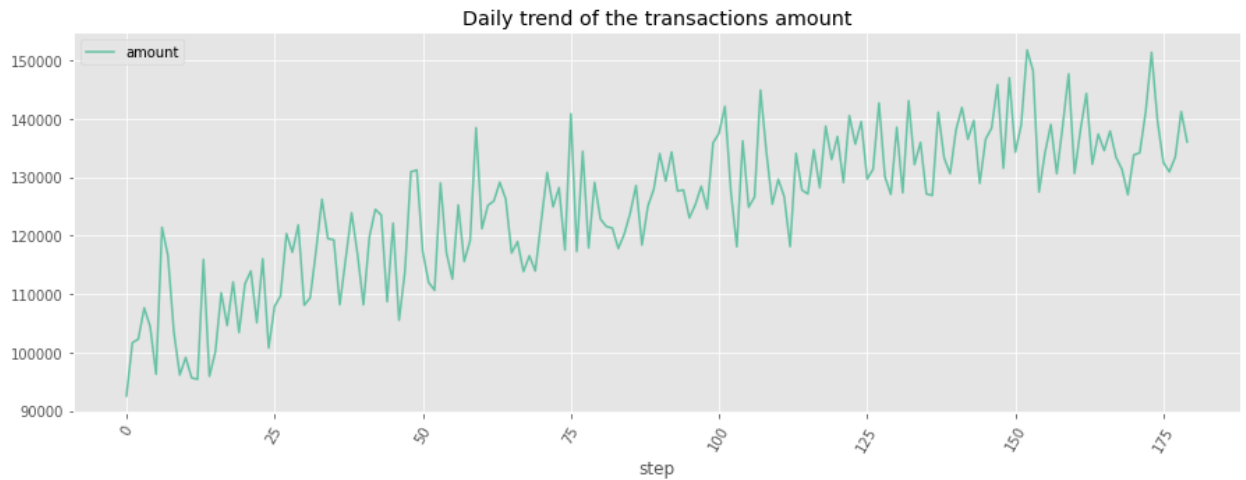
Measurements of performance :

Retention and growth of its customers

The objective of the bank is to maximize its profit. To do this, it must reassure its customers as to the confidentiality of their information and the security of their money. Thus, the customers will be confident and make the disclosure of positive information about the bank, which would motivate other people to choose the bank. As long as there is an increase in customers, the profit will increase.

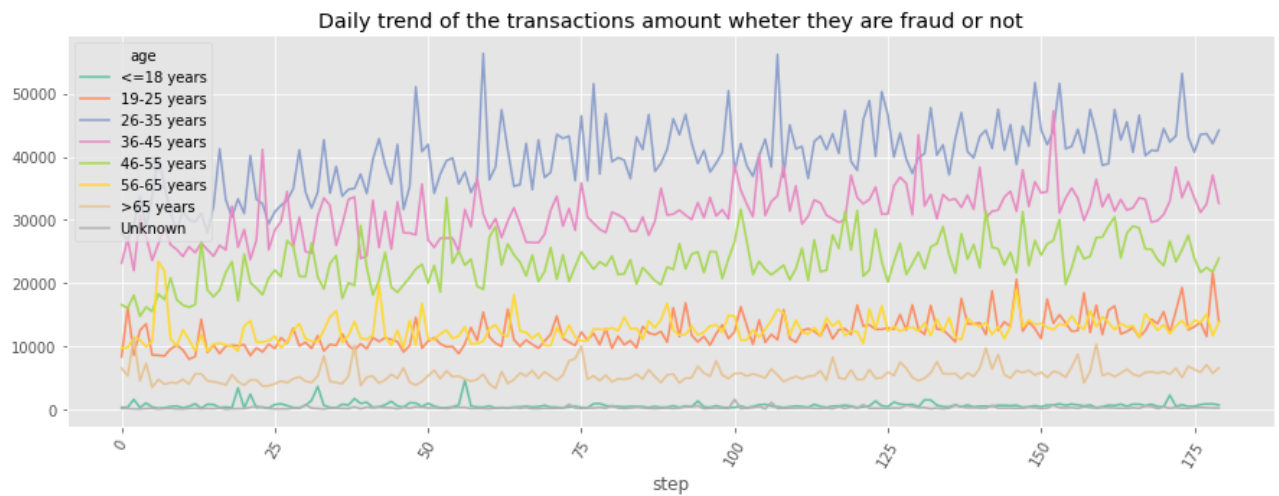
II. Results

The daily trend of the transactions.



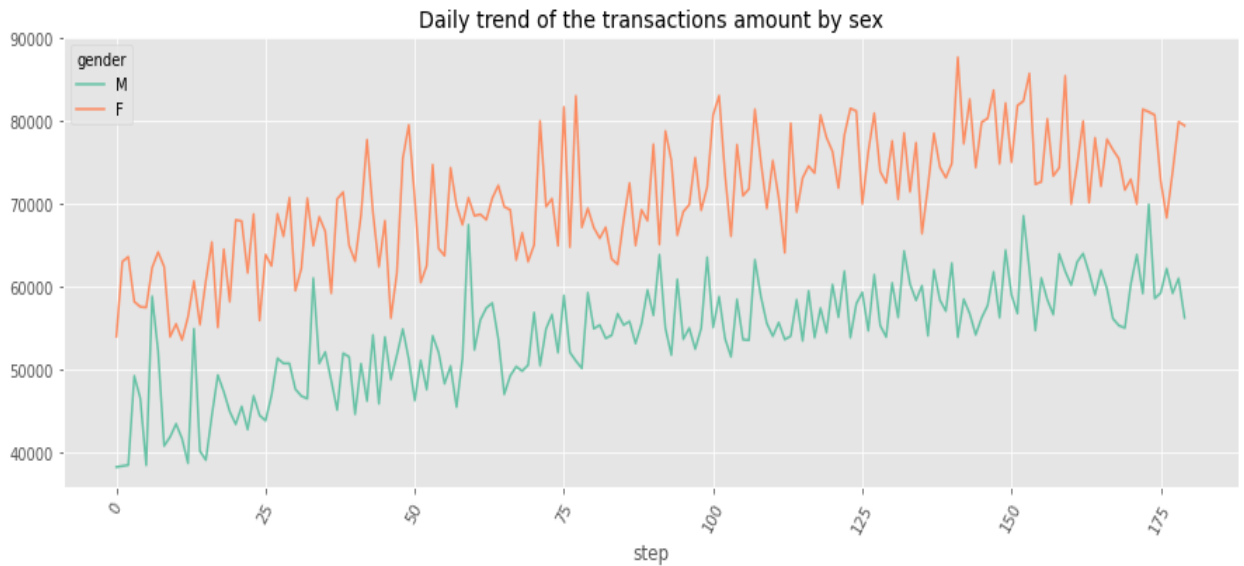
The general trend of the transactions is increasing.

Daily trend of the transactions by age



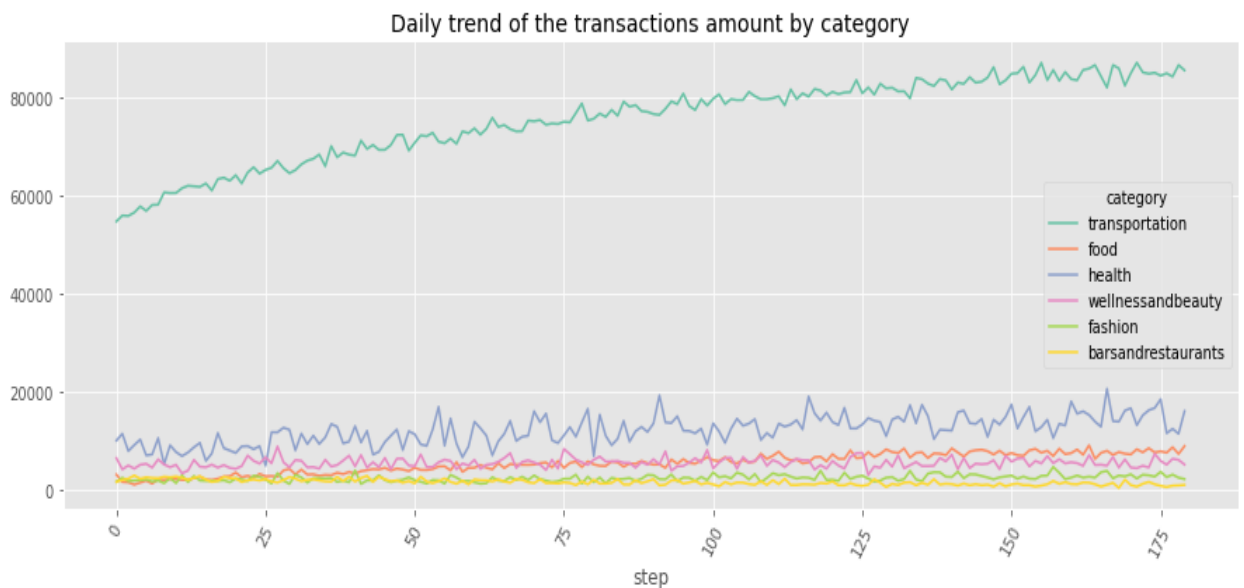
Daily trend of the transactions amount by age, the 26-35 years by more daily than any other class of age and they are the most victimized of fraudulent transactions among class of age.

Daily trend of the transactions by sex.



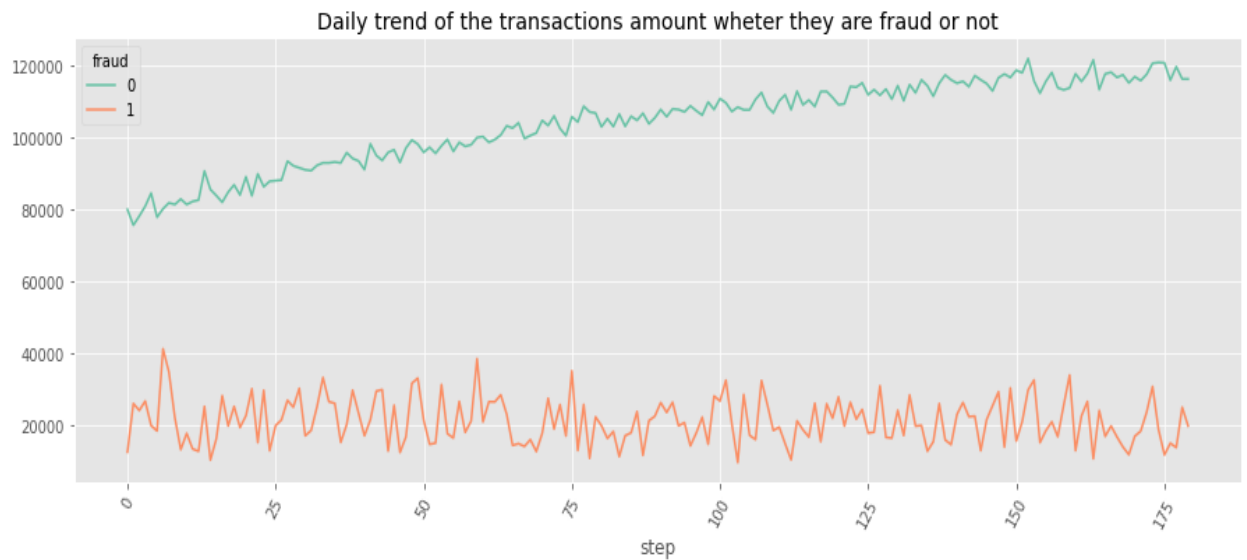
The transaction trend of the women seems to be the same but the amount done by the women is bigger. They grow up at the same time but the total amount of transactions stays bigger.

Daily trend of the transactions by category of the service.



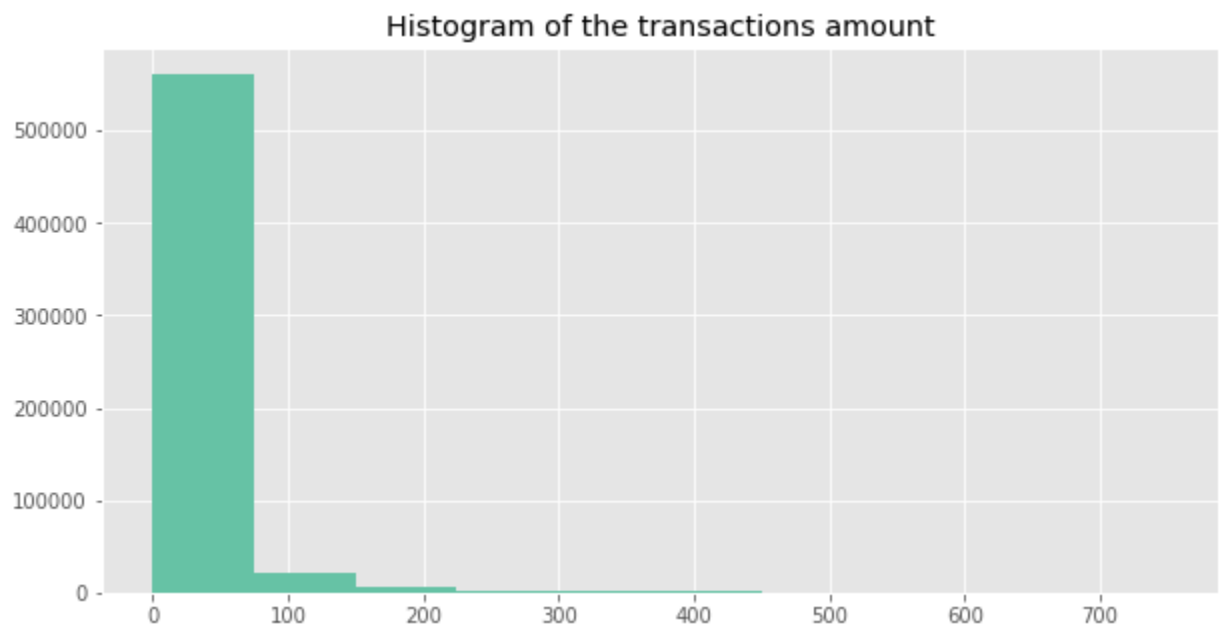
Like we have seen, the amount of money in transportation transactions is bigger than the others added. They represent 84% of the transactions.

Daily trend of the transactions by good and fraudulent transactions



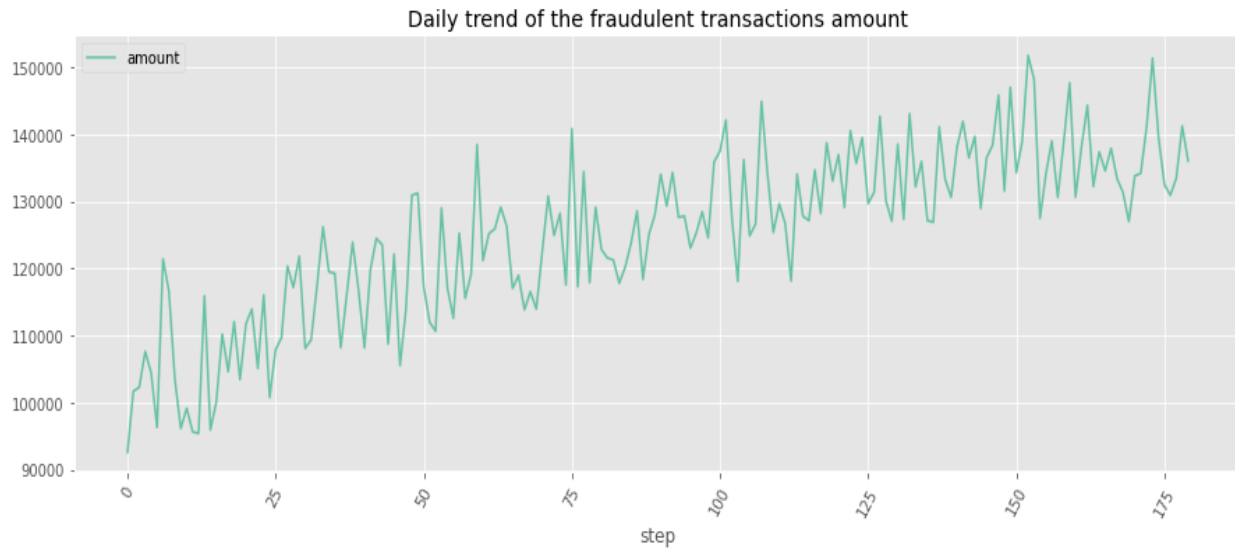
1.21% of the transactions are fraudulent, they never depass 40000 a day. The good transactions represent over 98% , they are over 80000 a day.

Histogram of transactions amount.



The transactions are generally between 0 and 100 dollars. We understand that. The transportation represents 84% of the total transactions and the amount average per transportation is 26.95 dollars.

Daily trend of fraudulent transactions



Fraudulent transactions tend to increase day by day. What is amazing to note is that there are frauds every day, so the control must be pushed continuously. Hence the urgency for the bank to solve the problem as soon as possible otherwise it risks losing many of its customers. The total amount of fraudulent transactions is 7200 which represent 3822671.17 dollars us.

<hr/>		
fraud	0	1
gender		
E	0.994058	0.005942
F	0.985340	0.014660
M	0.990927	0.009073
U	1.000000	0.000000
All	0.987892	0.012108

We performed a chi-square test to see if gender has some dependence on whether one is a victim of fraud or not.

The P-value is inferior to 0.05

We conclude that there is dependency between gender and fraud.

A woman would be more exposed to fraud than a man and those who do not specify their sex are less at risk. In fact, none of them were victims. See the graphic of fraudulent transactions by sex.



fraud	0	1
category		
barsandrestaurants	0.981171	0.018829
contents	1.000000	0.000000
fashion	0.982027	0.017973
food	1.000000	0.000000
health	0.894874	0.105126
home	0.847936	0.152064
hotelservices	0.685780	0.314220
hyper	0.954083	0.045917
leisure	0.050100	0.949900
otherservices	0.750000	0.250000
sportsandtoys	0.504748	0.495252
tech	0.933333	0.066667
transportation	1.000000	0.000000
travel	0.206044	0.793956
wellnessandbeauty	0.952406	0.047594
All	0.987892	0.012108

We performed a chi-square test to see if a category has some dependence on whether one is a victim of fraud or not.

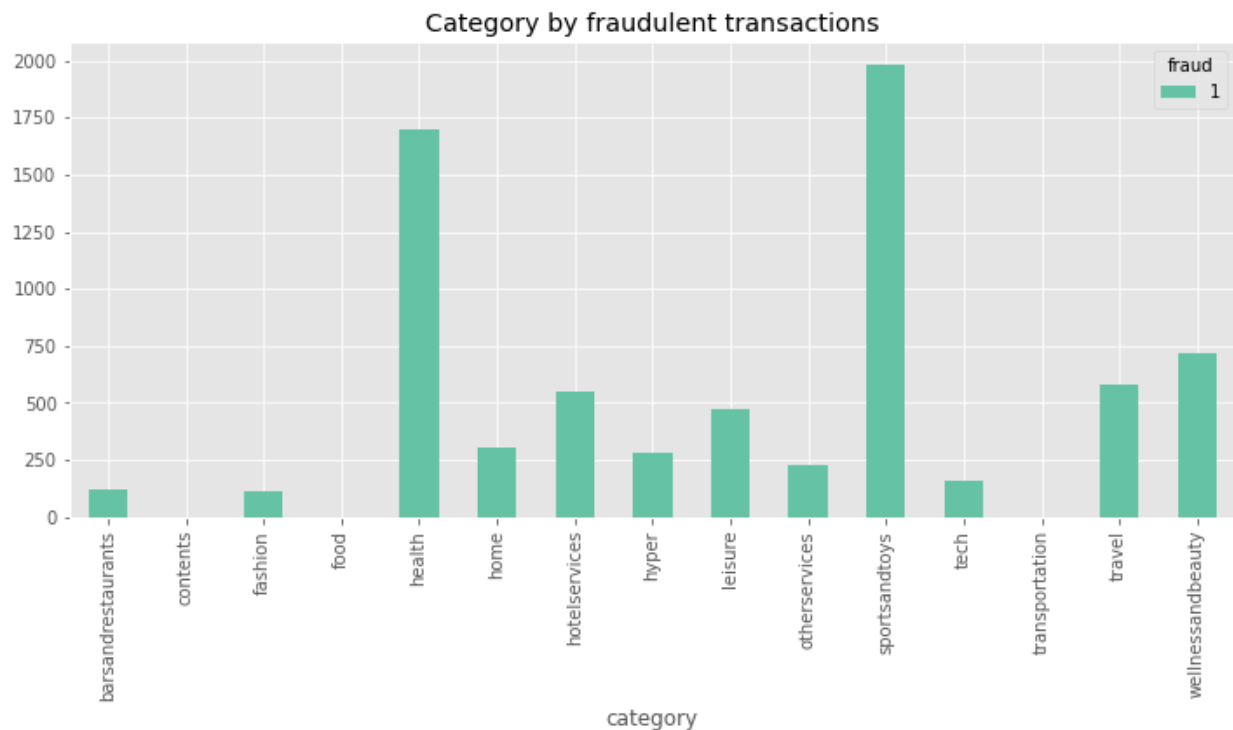
The p-value is inferior to 0.05

We conclude that there is dependency between category and fraud.

The leisure service would be more exposed to fraud, 94.99% of them are fraudulent. Food, contents and transportation are not at risk. In fact, none of them were victims.

Sports, toys and health are the most used service among the fraudulent transactions

Transportation, food and contents are no fraudulent transactions. We are reminded that transportation is the most used service, 84%. See the following graphic:



We classify the amount of transactions in different classes, we found that there is essentially fraud when the amount is over 2000 dollars. See the following table:

fraud	0	1
Class_amount		
Sup to 2000	1	343
inf to 2000	587442	6857

We have shown that there is a dependency between the amount of purchases and whether there is fraud or not. The pivot table above shows that from 2000 dollars for a transaction, there are no more safe transactions. The bank needs to increase control over these transactions.

We can say that 2000 \$ US is the threshold of good transactions.

Modeling

We perform a logistic model to predict whether a transaction is fraudulent or not.

We split the data to understand model performance, dividing the dataset into a training set and a test set is a good strategy.

The dataset is broken into two parts in a ratio of 75,25. It means 75% of the data will be used for model training and 25% for model testing. We leave 25% percent to have enough data to test the model. In fraud detection, the model has to be the most efficient possible.

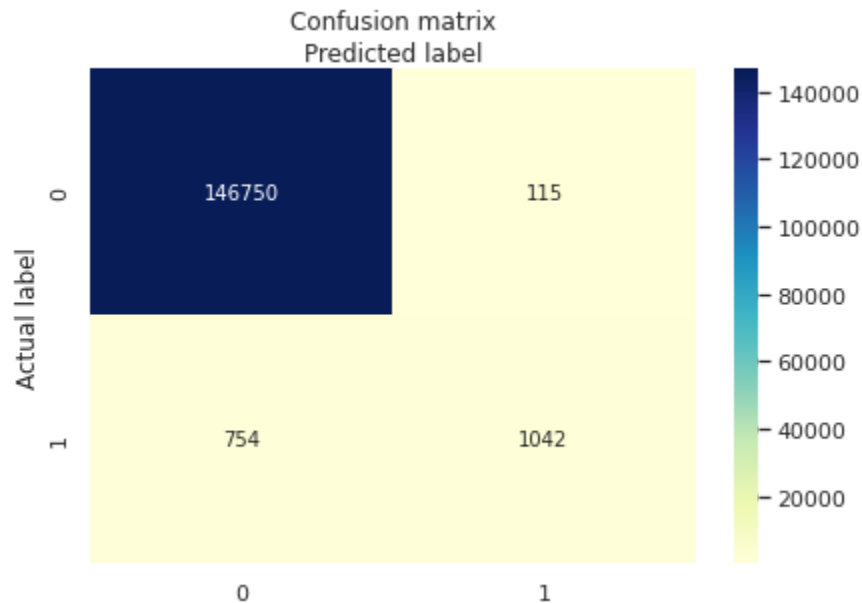
We have chosen the logit model because we think it is the most adapted to predict phenomena like whether a transaction is fraudulent or not. It is used to model a binary dependent variable.

We use the dummies of variables like : Age, category, gender. We transform each of the modalities of these variables into a binary variable. To remove the effect of multicollinearity, we delete a modality from each variable. We use “amount” as an independent variable.

The significant variables are: Enterprise status, bar and restaurants service card usage, travel, health, home, hotelservices, hyper, leisure, other services, sports and toys, tech and amount.

The LLR p value that could be compared to fisher test is 0.000 inferior to zero, we reject the hypothesis that the restricted model performs better. and prefer the full model over the null model.

The confusion matrix



Here, you can see the confusion matrix in the form of the array object. The dimension of this matrix is 2*2 because this model is binary classification. You have two classes 0 and 1. Diagonal values represent accurate predictions, while non-diagonal elements are inaccurate predictions. In the output, 146 750 and 1042 are actual predictions, and 115 and 754 are incorrect predictions.

Accuracy: 0.9941544857090965

Precision: 0.9006050129645635

Recall: 0.5801781737193764

Well, we got a classification rate of 99%, considered as good accuracy.

In the prediction case, when your Logistic Regression model predicts transactions are going to be fraudulent, that transactions have 90% of the time.

Possibly complete quasi-separation: A fraction 0.90 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

III. S.W.O.T

Strengths

- 1) To provide better assistance to protect the financial assets of its customers, to raise awareness and to regain confidence in the company
- 2) To have the confidence of its customers. Despite this, they decide to stay.
- 3) We could see that the problem is not general, it is only a certain number of customers who are affected.

Weaknesses

- 1) The non-authentication of customer transactions.
- 2) Media already aware of the difficulty.
- 3) Lack of active communication between the bank and its customers.

Opportunities

- 1) It is a good opportunity for the bank to analyze its data to identify flaws in its system.
- 2) Closing the accounts would be the first reaction of the affected customers, but since they did not do it that is they still trust, for that it is necessary that the difficulty is solved as quickly as possible. For that a better communication on the issue and the strengthening of the security policy can contribute to the loyalty of the customers once the problem is solved

Threats

- 1) Due to the flaws in the authentication system, if the customer loses his card, and it is another person who has it in his disposal then the bank will not be able to know who is the owner. For this we will need to have other measures much more reliable
- 2) Lack of communication between the bank and its customers, external agents, such as the media, other banks have interfered and this has many more disadvantages to ensure the reliability of the bank and restore its image.

IV. Solutions alternatives immédiates, à court et à long terme

Immediate solution

The bank should suspend all transactions in excess of \$2000 and those that do through leisure.

	fraud	0	1
Class_amount			
Sup to 2000	1	343	
inf to 2000	587442	6857	

The reduction on frauds if we cancel the transactions for the leisure services is 142335.98

The reduction on frauds if we cancel the transactions over 2000 because they are mostly fraudulent is 1 298 859.85

The amount of money the customers could save from fraud if we stop leisure and transactions over 2000 us is 1441195.83. They represent a reduction of 37.7%.

Strength : Fraud is decreasing while waiting for the system to be readjusted.

Weak point: The affected customers cannot make all the desired transactions since there is already a limit. This can provoke the customer's anger.

Challenge: Customers could leave the bank due to the constraints imposed by the bank on its transactions. With this, the bank must manage this situation as quickly as possible.

Short-term solution

1)Trace the last purchases made on the cards of customers experiencing fraud

Strengths: this solution will help find fraudsters

weak points: the bank can increase its vulnerability to the fraudster by leaving him viruses on his traceability

Challenge: Substantial data could be corrupted. To remedy this, the bank will have to recruit specialists in computer security and will use quality antiviruses.

Long-term solution

- 1) Implement an IT awareness system for Customers.

Strength: Customer awareness of security issues is essential, especially in the fight against fraudulent attacks. Effective communication, using all available channels (mail, telephone, Internet sites, etc.), is therefore desirable from all actors in the payment chain in order to draw users' attention to risk factors and good practices to be respected.

Weaknesses: it can happen that customers change their phone numbers or forget their email passwords.

Challenge: Client unavailability. To address this, clients should be given the opportunity to schedule their own training

- 2) Implement an online system to alert customers when their data is misappropriated

To achieve this solution, the system must be able to authenticate card owners during the transaction. to authenticate them, it will need to update and verify (phone numbers and email addresses)

Strengths: It will detect fraud attempts in an automated way. And customers can dispute fraudulent payments and secure their account

Weakness: Customers may not have the necessary resources (computers, internet) at the time of the fraud.

Challenge: The unreachable customer. To overcome this problem, the system must not accept under any circumstances a transaction without verification by phone and valid email address of the customer. To do this, the customer will receive an access code

Other type of data that could explain the fraudulent transactions and help solve the problem:

1. The location of the customers and merchants: This data could help us see if the fraudulent transactions are made in a specific area. To see if a lot of transactions are made in other places than the real address of the customer.
2. The frequency of card usage: To see if the card is often used, the client seems to be less secure. To see if a lot of transactions are made in a short timeframe.
3. Website where customer put his card information previously: Generally some bad website of selling can use the card information to make fraud.

