Ayiti Analytics Data Processing Bootcamp

Due Friday, July 02th at 2:00pm

AYITI ANALYTICS
Data Science
Training &
Research Center

Introduction

Team: Group 3
Client: The Bank's Senior Management

The business Problem

Description

the bank is unable to provide better customer service to protect customers' financial assets from fraud

Systems Analysis of Problem

- **Stakeholders**

1) The bank
2) The clients
3) The Media

Systems Analysis of Problem

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

Methodology

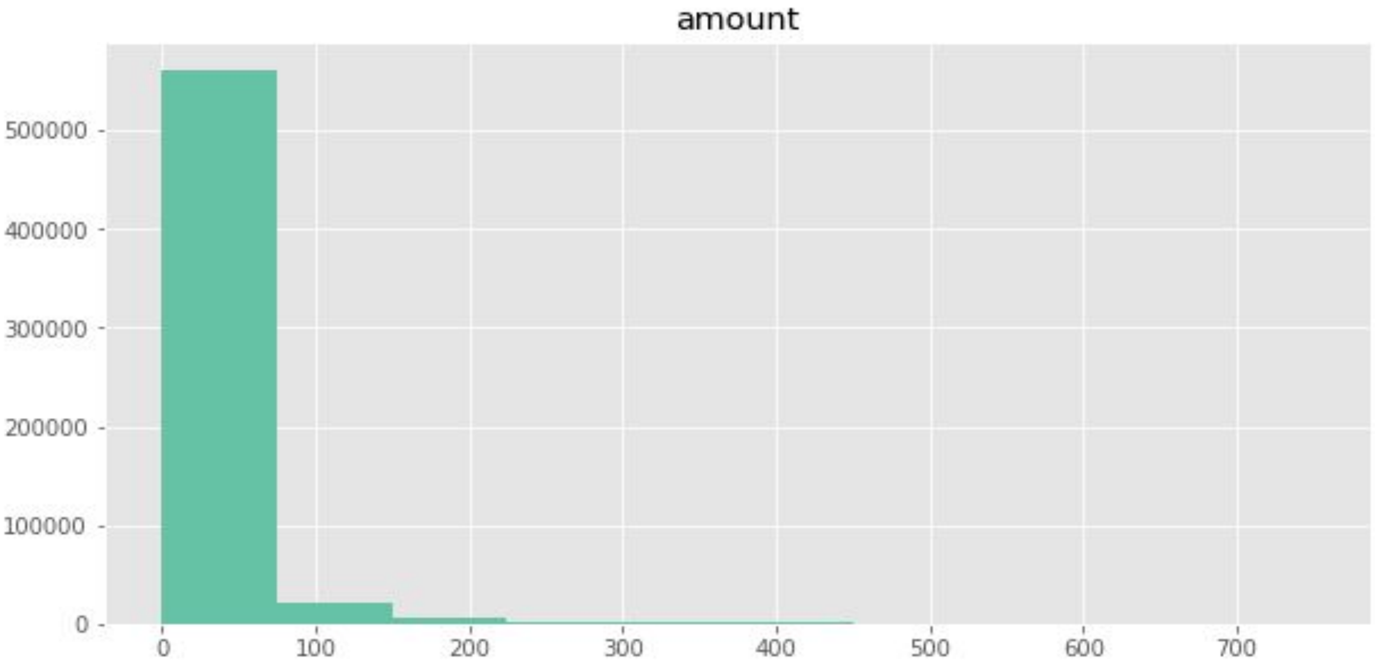
- Bank Dataset :
- Data processing and Analysis using Python.
- Calculating summary statistics to understand the data
- Daily trends to understand the trend
- Finding correlation between different variables and the fraud

Relevant Analytics

What data is important to understanding the situation?

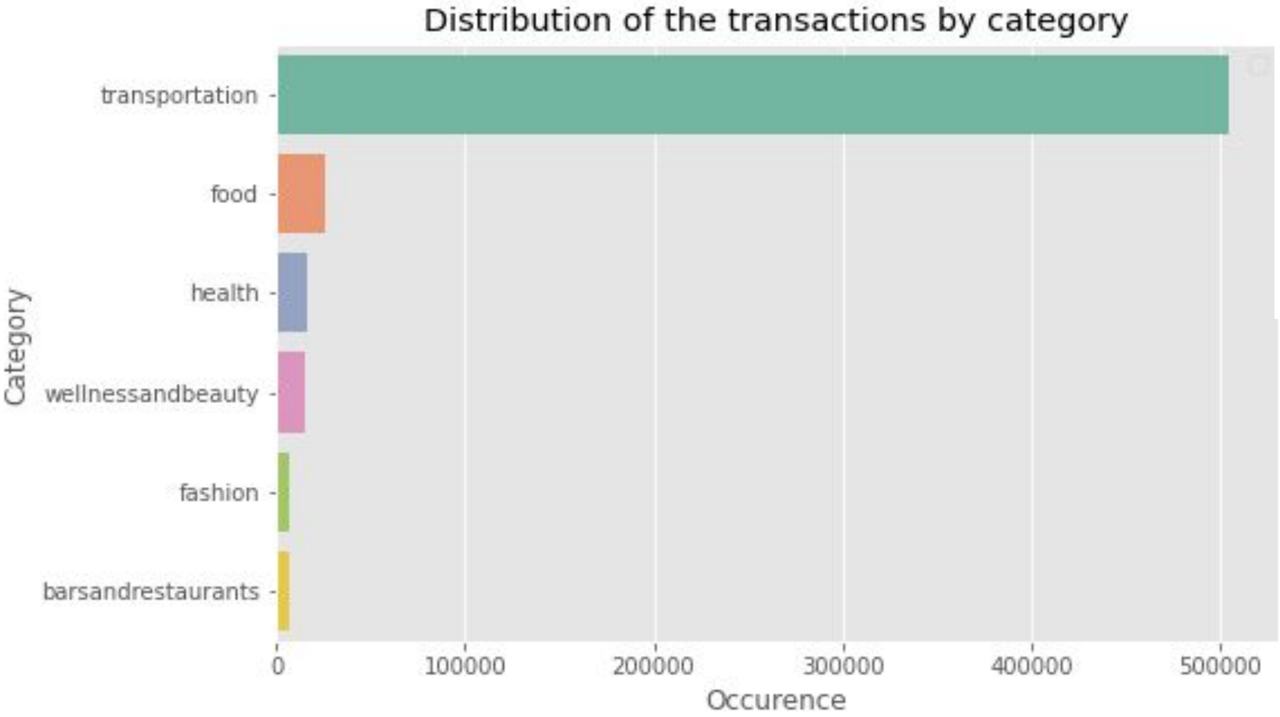
In the DataSet, these data are : age, Customer, gender, merchant, Category, Amount, Fraud

Result



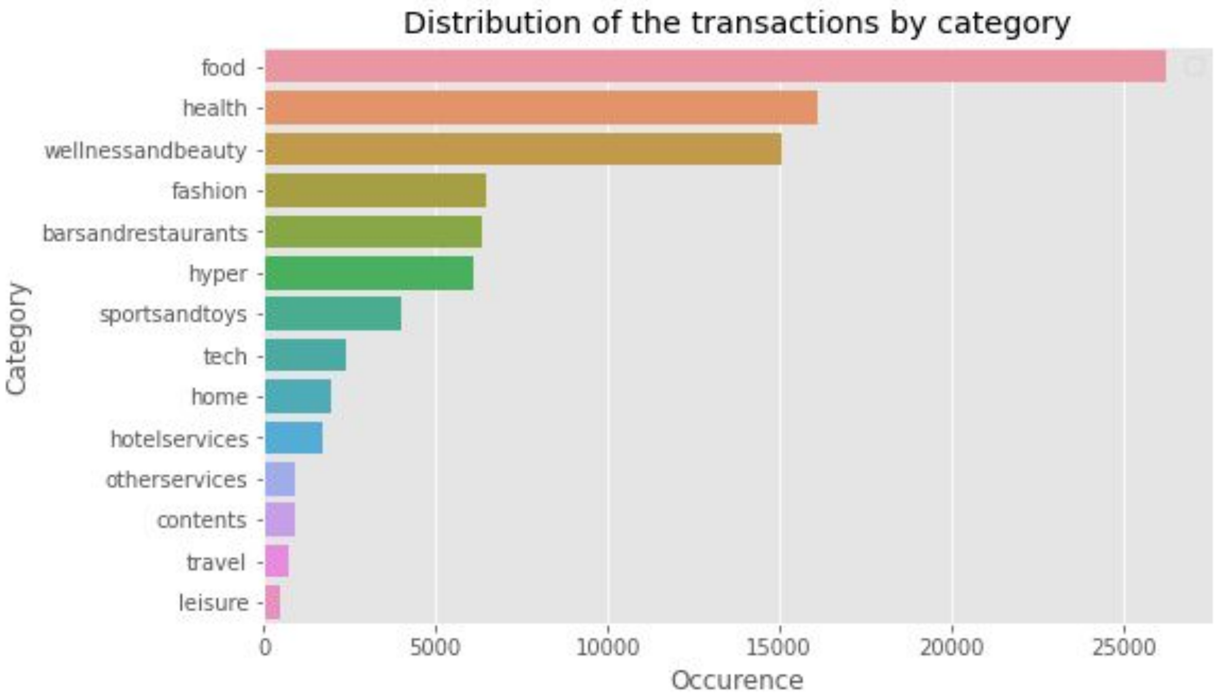
The amount of transactions are for the big part between 0 and 100 dollars.

Result



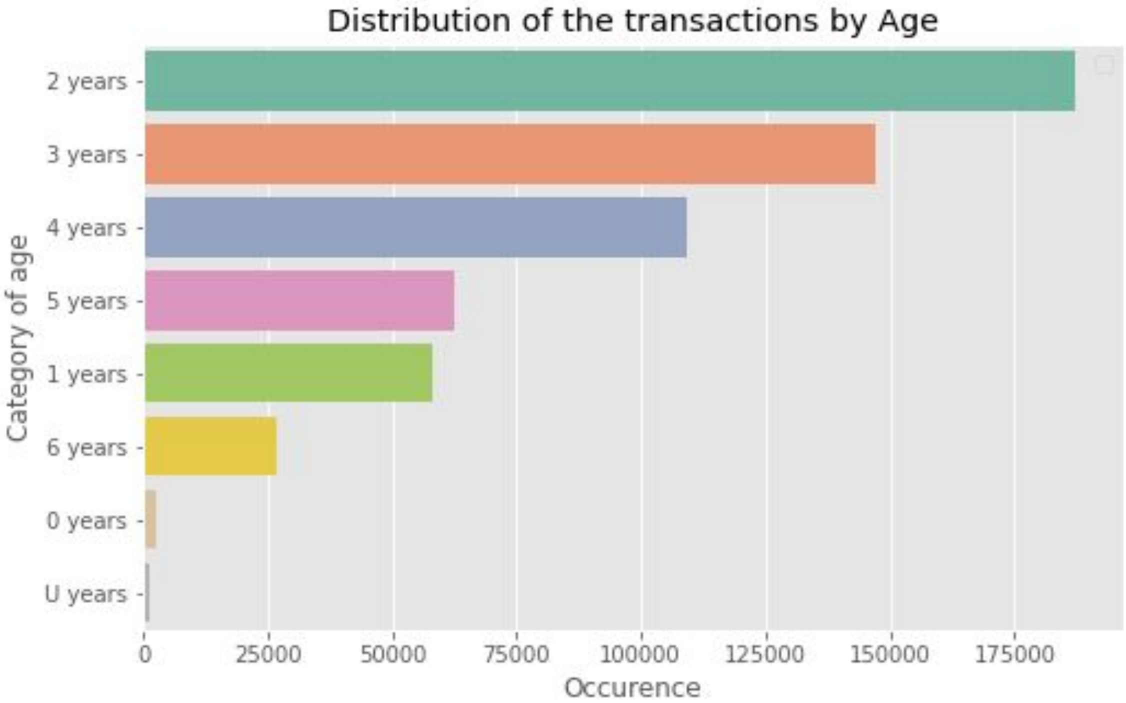
The most used service is transportation, the costumer use the card every day to pay for their transport.

Result



This graphic better present the services used by client. Leisure is the least used service while it is the most fraudulent service.

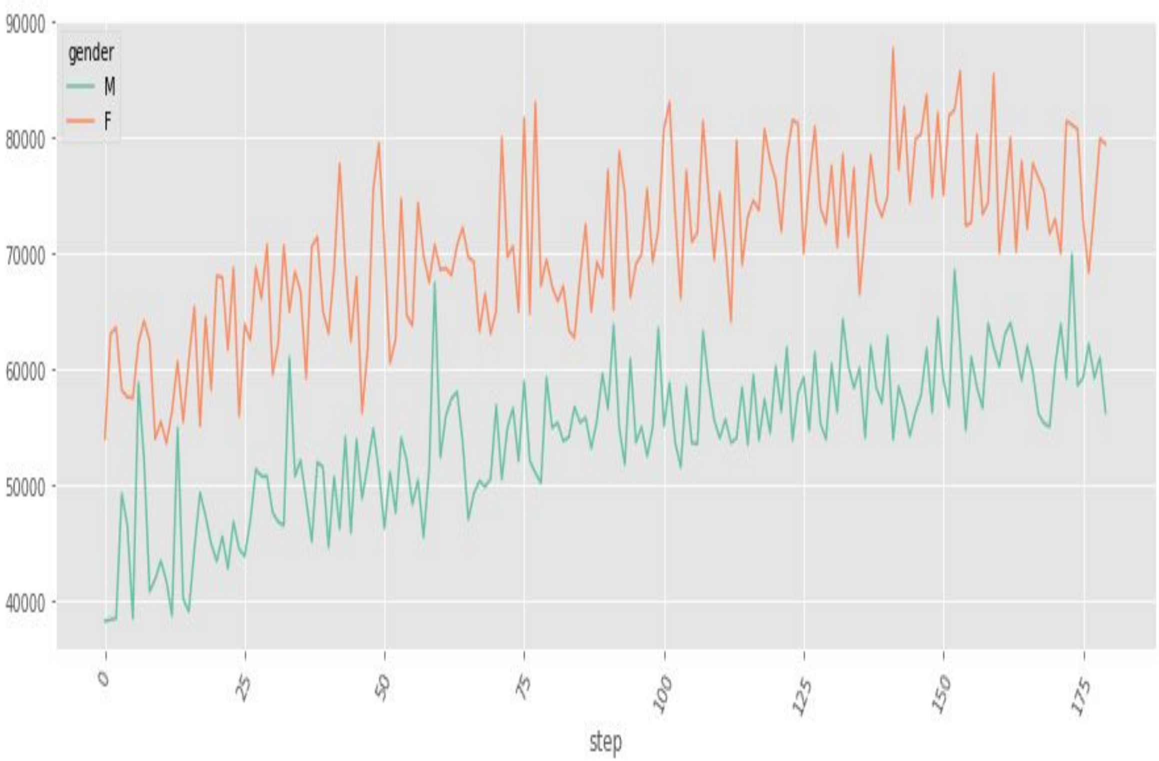
Result



The second class, 18-25 years old use the credit more often than the others.

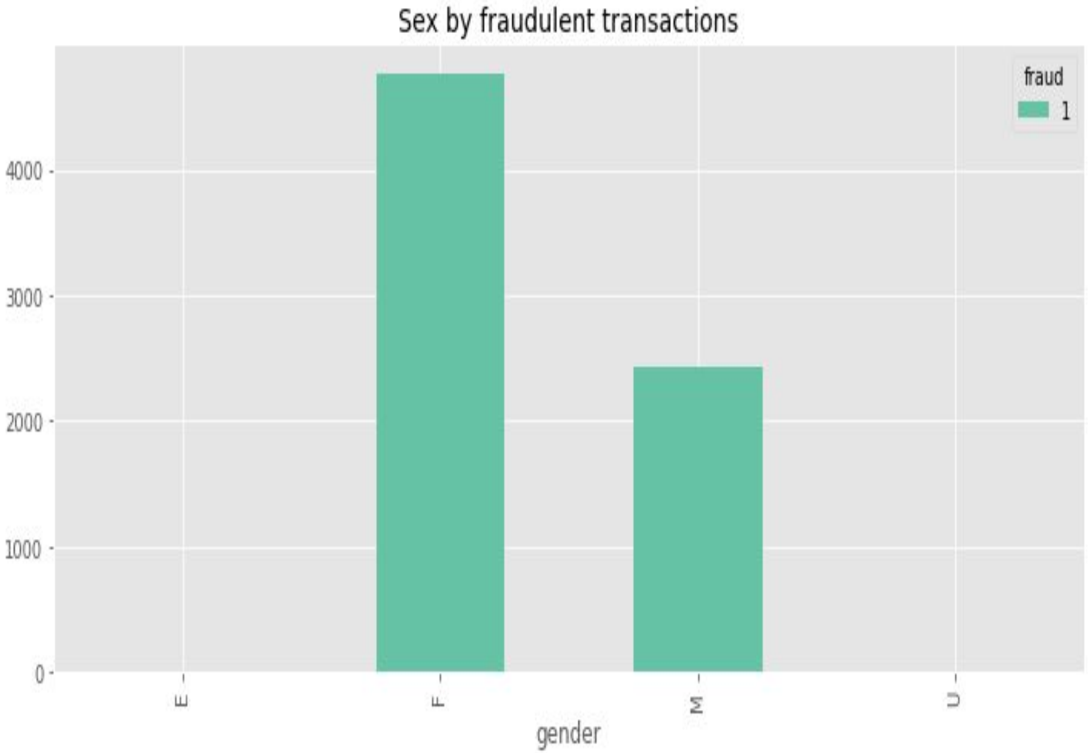
AYITI ANALYTICS

Result



This graphic presents the daily trend of the transactions by sex. The women make more transactions than the men. We'll show later that they are the most risked people to be victims of fraudulent transaction.

Result



This graphic presents the distribution of the sex by fraudulent transactions.

We see that the probability to be a woman and victim of fraud is bigger than when being a man.

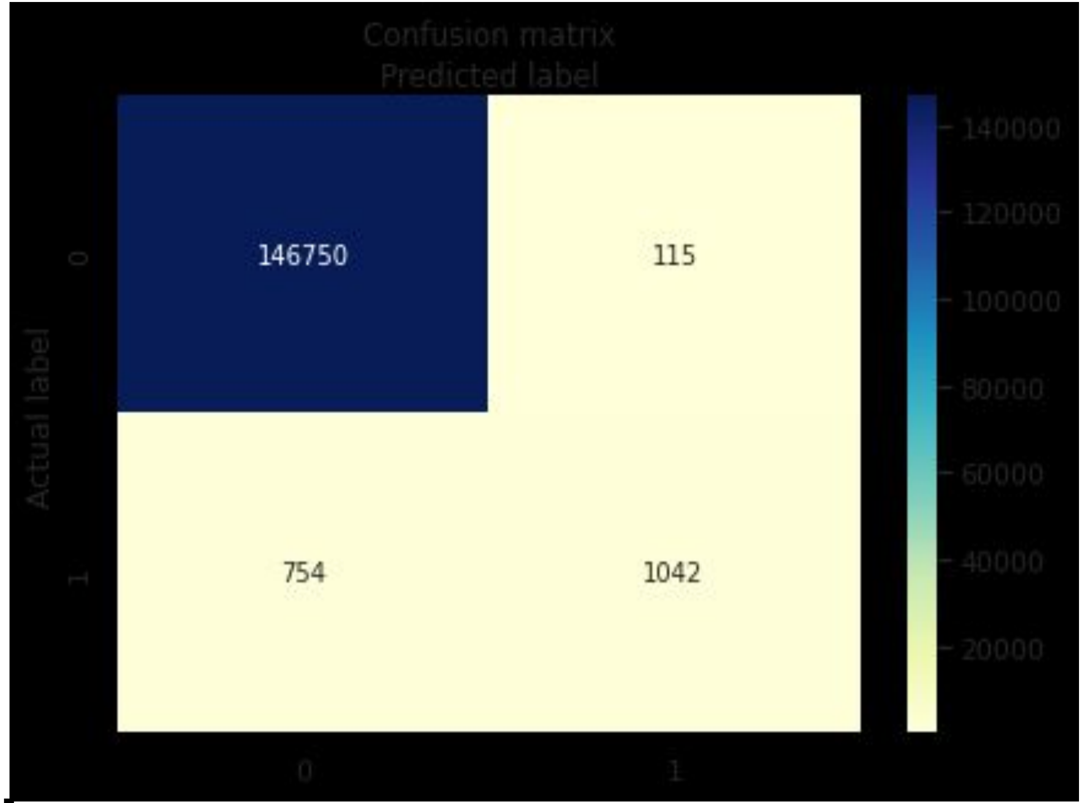
Result

We realize 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.

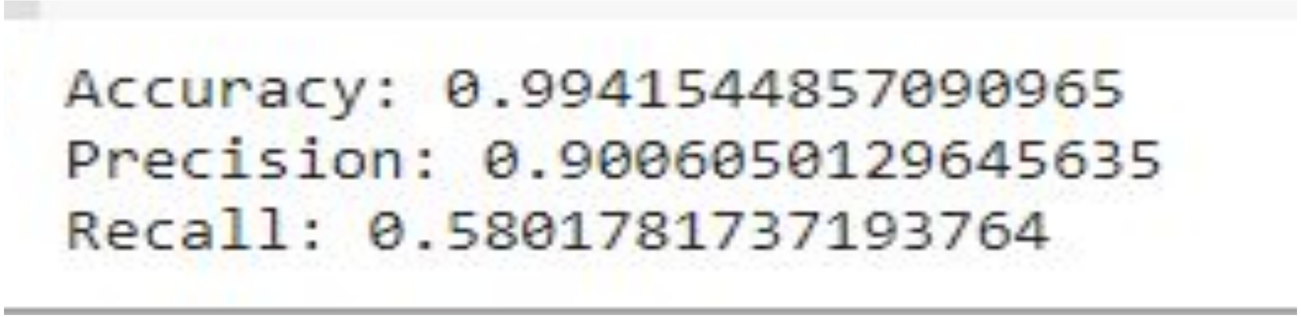
Result



	0	1
0	146750	115
1	754	1042

Here, 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.

Result



```
Accuracy: 0.9941544857090965
Precision: 0.9006050129645635
Recall: 0.5801781737193764
```

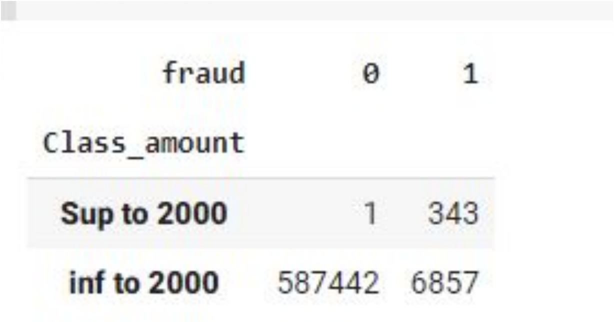
Well, we go a classification rate of 99%, considered as good accuracy

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

DISCUSSION AND SOLUTION

Immediate solution

The bank should suspend all transactions over $2000 and those using leisure and use a decision trees model to predict which variable is most likely to characterize Fraud.



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%

DISCUSSION AND SOLUTION

Immediate solution

The bank should suspend all transactions over $2000 and those using leisure

Strength:	Weakness:	Challenge:
Fraud is reduced while waiting for the system to be readjusted	The customers concerned cannot make all the transactions they want, since there is already a limit. This may cause customer anger	The customers could leave the bank because of the constraints imposed by the bank on its transactions. On that note, the bank has to manage this situation as soon as possible.

DISCUSSION AND SOLUTION

Short-term solution

Track the last purchases made on the cards of customers experiencing fraud

Strength	Weakness:	Challenge:
this solution will allow to find the fraudsters	the bank may increase its vulnerability to the fraudster by leaving viruses on its traceability	A lot of data could be corrupted. To remedy this, the bank will have to recruit specialists in computer security and will use quality antivirus software.

DISCUSSION AND SOLUTION

Long-term solution:

1) Implementing an IT awareness system for customers

Strength:	Weakness:	Challenge:
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	It can happen that customers change their phone numbers or forget the passwords to their email addresses.	Client unavailability. To address this, clients must be given the opportunity to plan their own training schedules

DISCUSSION AND SOLUTION

Long-term solution:

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

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

Strength:	Weakness:	Challenge:
It will detect fraud attempts in an automated way. And customers can dispute fraudulent payments and secure their account	It can happen that customers change their phone numbers or forget the passwords to their email addresses.	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

AYITI
ANALYTICS

DISCUSSION AND SOLUTION

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.

Thank you

Contact Us

If you have any question about this presentation, Feel free to send us a message in the #business-analysis (https://app.slack.com/client/T01V5TV0YMQ/C021PS8GCG2) channel on slack.

Ressources

➢ See the full project
Thttps://github.com/Wedsanley/fraud_detection

Team Members

Wedsanley JEAN PHILIPPE
Djeninah TIMOTHEE
Hardilès THERMORIS