Cloudwalk - Teste Prático

Olá, primeiramente gostaria de agradecer pela oportunidade em participar deste processo seletivo e de poder demonstrar um pouco sobre mim e sobre meus conhecimentos.

Sobre os desafios propostos, notei que a planilha compartilhada leva para uma página do Github que contém um teste de Engenheiro de Software, que particularmente é uma profissão em que eu almejo muito conquistar e por isso eu me dediquei ao máximo na solução do problema proposto lá e espero que meus resultados estejam dentro do esperado pela Cloudwalk.

Neste documento apresento um pouco da minha análise referente ao problema e também apresento a solução desenvolvida através um sistema anti-fraudes capaz de testar todas as transações da planilha compartilhada, gerar gráficos analíticos dos resultados, além de permitir processar novas transações simulando um funcionamento em tempo real.

1 - Identificando comportamento suspeito.

Através da planilha fornecida, podemos realizar diversas análises para identificar os padrões de comportamentos suspeitos que comumente levam a um chargeback de transações.

Padrões de chargeback identificados:

1. Usuário realizando várias transações em um curto espaço de tempo
2. Usuário realizando muitas transações no dia
3. Transações de alto valor em horários atípicos
4. Usuário com vários cartões diferentes
5. Cartão que possui vários usuários utilizando

Possíveis soluções:

1. Usuário realizando várias transações em um curto espaço de tempo.
    - Solução: limite de $1000 ao realizar mais de 1 transação em um período de 4 horas
2. Usuário realizando muitas transações no dia
    - Solução: limite diário de 3 transações
3. Transações de alto valor em horários atípicos
    - Solução: Bloquear transações acima de $3500 entre 21:00 e 04:00
4. Usuário com vários cartões diferentes
    - Solução: Permitir a troca de cartões após 7 dias da última transação, desde que o novo cartão não tenha chargeback registrado
5. Cartão que possui vários usuários utilizando
    - Solução: Permitir apenas 2 usuários utilizando o mesmo cartão em um intervalo de 7 dias, só é possível ser utilizado por novos usuários que não tenham chargeback registrado

Estes foram os padrões mais encontrados sem levar em consideração que os componentes (user, card, device, etc) podem ter histórico de chargebacks anteriormente, porém nada impede que padrões diferentes apareçam e o intuito do sistema anti-fraudes é evitar todas as ameaças possíveis.

Pensando em desenvolver um sistema robusto e eficaz, foi necessário identificar e mapear todas as possíveis ameaças que vão além dos padrões comumente vistos e além disso, para identificar novos padrões, precisamos rastrear se algum componente da transação já esteve presente em alguma transação com chargeback.

Na planilha já temos a informação de qual transação teve chargeback ou não, mas pensando no mundo real a informação de chargeback é recebida dias após a transação ter sido realizada, sendo assim precisamos ter cautela ao tentar identificar padrões levando em consideração os chargebacks.

Como a planilha compartilhada tem poucos dados disponíveis, sendo apenas 30 dias de transações, não é possível simular chargebacks que são informados após muito tempo.

Para tornar o teste mais dinâmico levei em consideração que chargebacks são informados após 3 dias da data de realização da transação, dessa forma é possível identificar quais componentes estiveram presentes em chargebacks anteriores e analisar se ainda apresentam riscos e se o sistema deve ou não autorizar novas transações.

Sendo assim, podemos identificar novos padrões que vão além de uma análise superficial, é possível cruzar informações entre componentes e verificar históricos podendo criar um sistema muito mais seguro e robusto. Pensando nisso foram mapeados 18 casos diferentes para identificar transações suspeitas:

Verificações simples:

Caso 1: Duas ou mais transações que excedam o limite de $1000 valor em um período de 4 horas
Caso 2: Bloqueia transações acima de $3500 entre 21:00 e 04:00

Caso 3: Usuário realizou mais de 3 transações no dia
Caso 4: Cartão realizou mais de 3 transações no dia
Caso 5: Dispositivo realizou mais de 3 transações no dia

Caso 6: Usuário possui mais de 1 chargeback em 7 dias
Caso 7: Cartão possui mais de 1 chargeback em 7 dias
Caso 8: Dispositivo possui mais de 1 chargeback em 7 dias

Caso 9: Usuário possui mais de 5 chargebacks no histórico
Caso 10: Cartão possui mais de 5 chargebacks no histórico
Caso 11: Dispositivo possui mais de 5 chargebacks no histórico

Verificações complexas:

Caso 12: Usuário utilizou 2 cartões em 7 dias
Caso 13: Usuário utilizou mais de 3 dispositivos em 7 dias

Caso 14: Cartão foi usado por mais de 3 dispositivos em 7 dias
Caso 15: Cartão foi usado por mais de 3 usuários em 7 dias

Caso 16: Dispositivo foi usado por mais de 3 usuários em 7 dias
Caso 17: Dispositivo foi usado por mais de 3 cartões em 7 dias

Caso 18: Comerciante possui mais de 1 chargeback em 7 dias

Estes foram os casos que eu identifiquei e analisei e que são interessantes de serem utilizados para aprovar ou negar uma transação. Mas podem existir muitas outras verificações além destas.

Além dos dados presentes na planilha, um outro dado essencial para identificar possíveis fraudes seria o IP do usuário. Golpes e fraudes são comumente realizados via internet com vazamentos de informações ou roubo de contas e credenciais, de tal forma que o usuário fraudador provavelmente fará um acesso através de um IP diferente do usuário legítimo, identificando assim uma transação suspeita.

2 - Sistema Anti-fraudes

Para que fosse possível testar os casos citados anteriormente com mais veracidade, é necessário utilizar uma ferramenta automatizada para isso.

Sendo assim foi desenvolvido um sistema Anti-Fraudes em Python que analisa todo o histórico de transações e faz o processamento de cada transação individualmente levando em consideração que as informações de chargeback somente podem ser interpretadas após 3 dias da realização da transação.

Dessa forma o sistema processa as transações uma a uma e armazena uma cópia em uma nova planilha contendo a recomendação “approved” ou “deny” para cada transação e também qual foi o caso que negou a transação “deny_case”.

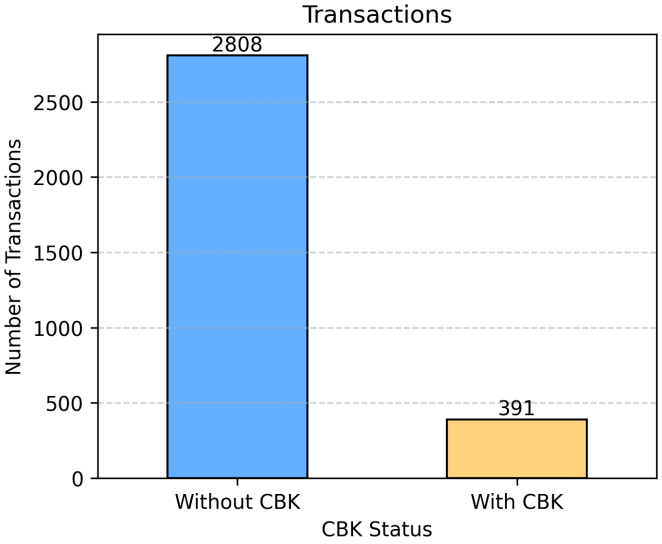
Após ser processado todo o histórico de transações é possível exibir diversos gráficos dos resultados para analisar aspectos como:

- Quantidade de transações aprovadas e negadas
- Quantidade de acertos e erros
- Precisão das operações negadas
- Relação de chargebacks entre novos usuários e antigos

Além disso é possível realizar o processamento de uma nova transação e simular o funcionamento “em tempo real” onde com base no histórico processado o sistema irá decidir se aprova ou não a nova transação.

3 - Resultados do sistema

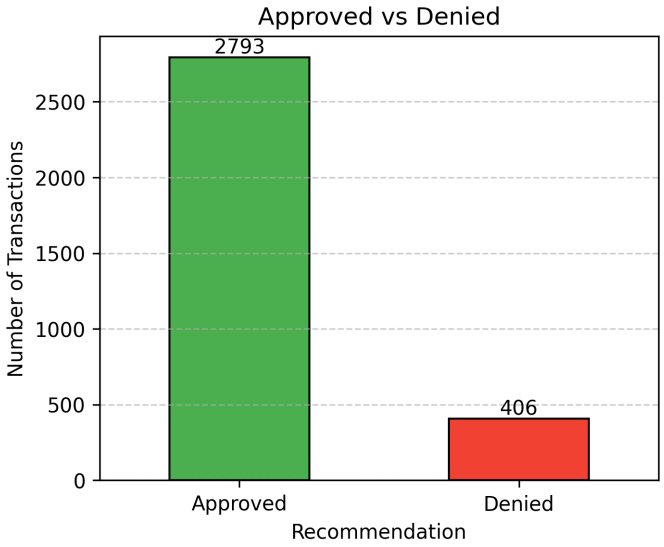
O histórico possui 3199 transações que foram processadas pelo sistema e classificadas como aprovada ou negada. Para analisar os resultados do sistema de diferentes aspectos mais facilmente irei apresentar os resultados graficamente.



No histórico original sem a aplicação de técnicas para evitar fraudes de chargeback.

De um total de 3199 transações no histórico original, 2808 representam transações que não obtiveram chargeback.

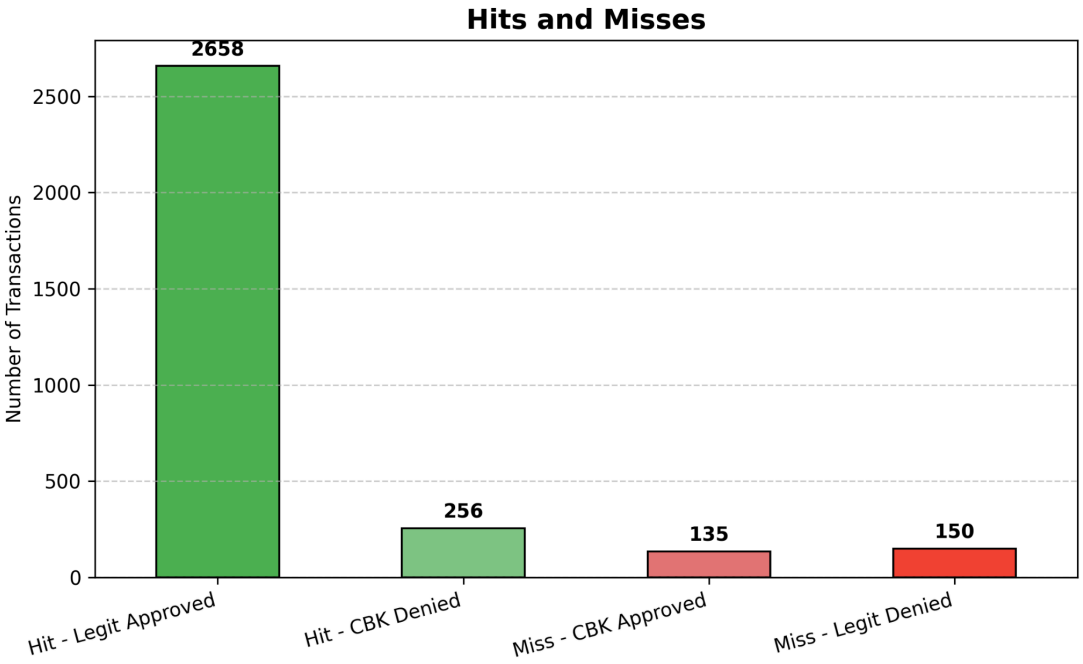
São 391 transações com chargeback confirmado, representando aproximadamente 12,23% do total.



Após processar o histórico original obtemos a quantidade de transações que foram aprovadas e negadas

De um total de 3199 transações, 2793 foram autorizadas, representando 87,3%.

O restante somam 406 transações que foram negadas pelo sistema por serem consideradas suspeitas por algum dos casos de verificação.



Podemos calcular algumas métricas muito importantes para analisar o sistema.

1. **Acurácia - Proporção total das decisões corretas**

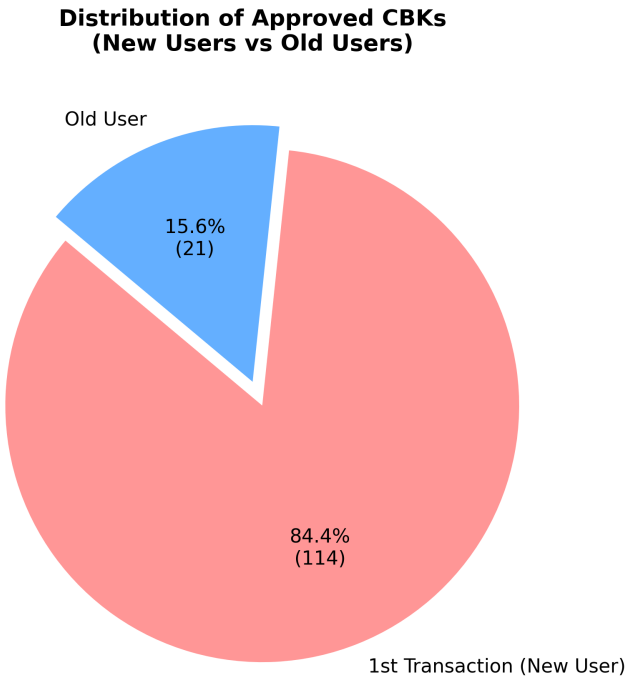
   Acurácia = (2658 + 256) / 3199 = 0.912 ~ 91.2%
   Muito bom - O sistema toma decisões corretas na maioria dos casos

2. **Precisão - Probabilidade de uma transação negada realmente ser fraudulenta:**

   Precisão = 256 / (256 + 150) = 0.63 ~ 63%
   Moderado - O sistema nega alguns clientes legítimos

3. **Recall - Proporção de fraudes negadas:**

   Recall = 256 / (256 + 135) = 0.655 ~ 65.5%
   Moderado - Cerca de 2 de 3 fraudes são evitadas



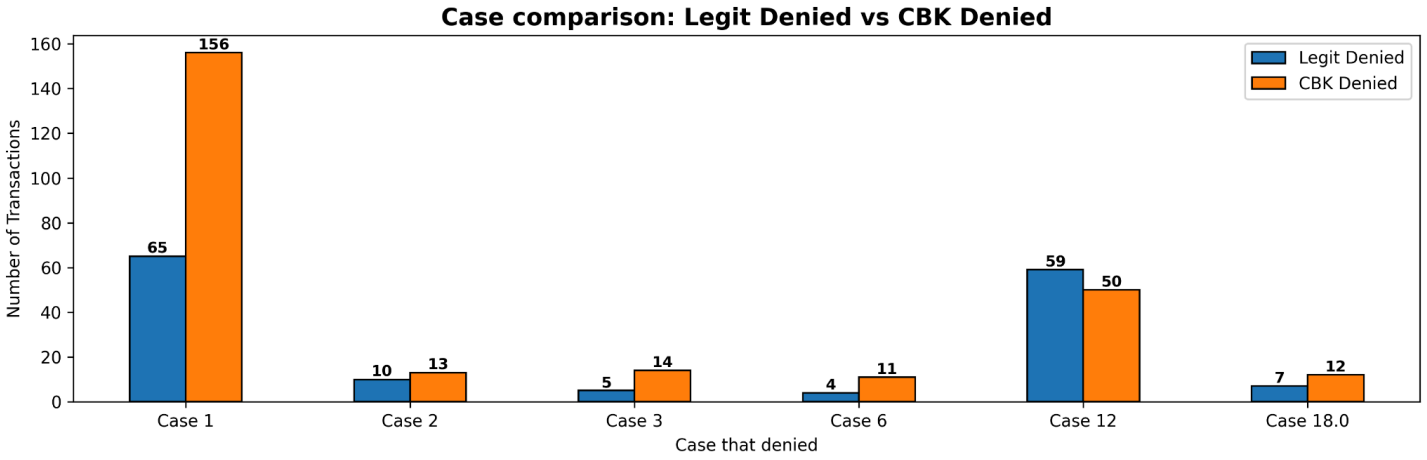
Para um sistema de Anti-fraudes o número de transações fraudulentas que não foram identificadas e impedidas é um fator muito preocupante.

Mas também é preciso analisar os casos em que não é possível evitar um chargeback de nenhuma forma.

Nos casos em que os chargebacks são realizados por novos usuários no sistema é impossível rastrear possíveis riscos para evitar estas transações.

Este gráfico representa que das 135 transações com chargeback, 114 foram feitas por novos usuários e apenas 21 foram feitas por usuários já existentes.

Olhando de outro ângulo, das 391 transações com chargebacks, 256 foram negadas, 114 não podem ser evitadas e apenas 21 transações não foram impedidas pelas regras de verificação de suspeitas.



Também é possível analisar quais dos casos de verificação mais recusaram transações e analisar como a regra está impactando em transações fraudulentas e legítimas.

É possível notar que o Caso 1 teve um alto desempenho, negando mais que o dobro de transações com CBK do que transações legítimas. Mas é importante ressaltar que isso se deve ao fato de que as verificações são feitas por etapas onde o Caso 1 é verificado primeiro que o Caso 2 e assim sucessivamente, agindo como um funil, sendo assim é normal que os primeiros casos de cada tipo de verificação tenham um desempenho um pouco maior.

Já no Caso 12 podemos notar que ele impactou mais nos usuários legítimos que nos fraudulentos, este caso é a verificação de “Usuário com vários cartões”, nesse caso o sistema precisou ser mais rígido com clientes legítimos para que pudesse barrar mais clientes fraudulentos que realizam várias trocas de cartão.

Na prática é como se o cliente legítimo não pudesse realizar 2 transações com cartões diferentes na mesma semana, porque se enquadra como uma operação comumente realizada por fraudadores de cartões. Sendo assim um cliente legítimo só pode trocar de cartão se já tiver passado 7 dias desde sua última transação.

4 - Teste de novas transações

Para conferir se o sistema realmente funciona de maneira correta é possível simular novas transações e verificar se o sistema aprova ou recusa com base no histórico já processado anteriormente, sendo assim o sistema irá analisar todos os casos para verificar se a transação é suspeita ou não.

Exemplo: Transação de teste armazenada no código

```
transaction = {
    "transaction_id" : 2342357,
    "merchant_id" : 29744,
    "user_id" : 97051,
    "card_number" : "434505******9116",
    "transaction_date" : "2019-11-30T23:16:32.812632",
    "transaction_amount" : 373,
    "device_id" : 285475
}
```

```
transaction_id                          2342357
merchant_id                               29744
user_id                                   97051
card_number                  434505******9116
transaction_date      2019-11-30 23:16:32.812632
transaction_amount                          373
device_id                                285475
recommendation                          approve
dtype: object
```

Exemplo: Transação inserida manualmente

```
transaction_id: 0002
merchant_id: 12345
user_id: 81152
card_number: 650516******9201
transaction_date: 2019-12-02 22:00
transaction_amount: 898
device_id: 486
```

```
transaction_id                          2
merchant_id                         12345
user_id                             81152
card_number            650516******9201
transaction_date      2019-12-02 22:00:00
transaction_amount                898.0
device_id                           486
recommendation                     deny
dtype: object
```

5 - Understand the Industry

1. Explain briefly the money flow, the information flow and the role of the main players in the payment industry.

Money flow: The money moves from the cardholder's bank (Issuer) to the merchant's bank (Acquirer) through the card network like Visa, Mastercard, etc.
After validation, the acquirer deposits the merchant's funds into the merchant's account.

I**nformation flow:** The issuer checks if the card is valid, the funds are available, and whether the transaction seems legitimate, then sends an "approved" or "declined" response back

Merchant => Acquire => Card Network => Issuer => Authorize

Main players:

- Cardholder: The consumer making the purchase.
- Merchant: The business selling goods or services.
- Acquirer: The financial institution processing payments for merchants.
- Issuer: The bank that issued the card to the cardholder.
- Card network: The intermediary (Visa, Mastercard, etc.) connecting acquirers and issuers.

2. Explain the main differences between acquirer, sub-acquirer and payment gateway, and how the flow explained in the previous question changes for these players.

Acquirer: A licensed financial institution that connects directly to the card networks (Visa, Mastercard) and manages merchant accounts.
Handles authorization, settlement, and chargebacks.

Sub-acquirer: Acts as an intermediary between merchants and the acquirer. Often used by online platforms and marketplaces.
Aggregates multiple merchants under its own merchant account, simplifying onboarding for small businesses.

Payment Gateway: A service that transmits encrypted transaction data between the merchant and the acquirer.
Does not hold funds; only facilitates the information flow (not the money flow).

Flow difference:

With a gateway, the merchant connects directly to an acquirer, only the data path changes.

With a sub-acquirer, both data and money flow through an extra layer where the sub-acquirer receives funds first, then pays the merchant.

3. Explain what chargebacks are, how they differ from a cancellation and what is their connection with fraud in the acquiring world.

Chargeback: A dispute initiated by the cardholder through their issuing bank after a transaction has already been processed and settled. The issuer investigates, and if the claim is valid, the transaction is reversed, and the merchant loses the funds.
Common causes: fraud (stolen card), product not received, or product not as described.

Cancellation: A reversal initiated by the merchant that occurs before settlement (when an order is canceled immediately after purchase). There is no dispute process involved.

Connection to fraud: Many chargebacks result from fraudulent transactions for example, when a stolen card is used without the cardholder's consent. For acquirers, a high chargeback rate indicates potential fraud risk.

4. What is an anti-fraud and how an acquirer uses it.

Anti-fraud system: These are complex systems that use algorithms, rules, and machine learning models to analyze transactions in real time to detect suspicious or potentially fraudulent activity before authorization.

How acquirers use it: When a merchant submits a transaction, the acquirer's anti-fraud system evaluates it using factors such as user behavior, device ID, transaction history, geolocation, and merchant profile.

If the transaction appears suspicious, it may be flagged, declined, or sent for manual review. This helps reduce chargebacks and protect both merchants and consumers.