

# Data Analyst Case

💡 Notebook on data explorations, findings, assumptions, errors and cleaning

## Introduction

## Company description

Ironhack payments, is a financial services company, offering a cash advance solution service

## Problem statement

Ironhack does not understand what is their monthly cohort evolution usage for the cash advances service

## Key metrics

1. Frequency service usage
2. Incident Rate
3. Revenue generated by the cohort
4. Total amount lend
5. Total amount reimbursed

## Notes

---

[Cash request exploration](#)

[Cohort](#)

[Fees](#)

# Cash\_request exploration

Context : CR = cash request

Status:

Possibles values are :

- approved : CR is a 'regular' one (= without fees) and was approved either automatically or manually. Funds will be sent approx. 7 days after the creation
- money\_sent : We transferred the fund to the customer account. Will change to active once we detect that the user received the funds (using user's bank history)
- rejected : The CR needed a manual review and was rejected
- pending : The CR is pending a manual review from an analyst
- transaction\_declined : We failed to send the funds to the customer
- waiting\_user\_confirmation : The user needs to confirm in-app that he want the CR (for legal reasons)
- direct\_debit\_rejected : Our last attempt of SEPA direct debit to charge the customer was rejected
- canceled : The user didn't confirm the cash request in-app, we automatically canceled it
- direct\_debit\_sent : We sent/scheduled a SEPA direct debit to charge the customer account. The result of this debit is not yet confirmed
- waiting\_reimbursement : We were not able to estimate a date of reimbursement, the user needs to choose one in the app.
- active : Funds were received on the customer account.
- money\_back : The CR was successfully reimbursed.

Shape

23970 rows & 16 columns

Columns names

```
Index(['id', 'amount', 'status', 'created_at', 'updated_at', 'user_id', 'moderated_at',  
'deleted_account_id', 'reimbursement_date', 'cash_request_received_date', 'money_back_date',  
'transfer_type', 'send_at', 'recovery_status', 'reco_creation', 'reco_last_update'], dtype='object')
```

Info:

Need to change dates from 'object' to 'datetime'

Unique values are:

id

created\_at

updated\_at

Columns that will be used to analyze:

id, amount, status, created at, user\_id, deleted\_account\_id, transfer type

Need to convert:

created\_at → object to datetime

reimbursed\_date → object to datetime

Clean up:

user\_id which were blank not taken into consideration in this study as there are deleted users

Using only columns = id, amount, status, created\_date, reimbursement\_date, userd\_id ,  
recovery\_status, first\_request

# Fees

## Cleanup

- To not lose 4 rows where the cash\_request\_id = NaN, I replace them by the id appearing in the reason column
- Convert columns (id, cash\_request\_id) into int

To be able to sum\_up all the fees, I need to create a new table where:

- Group by cash\_id
- Sum amount of fees

# Cohort

Cohort refers to a group of people which share the same characteristics, for instance, the students who join an ironhack course in june. It will be refers to the june cohort since there is a unique characteristic which bring this sample together

Ironhack Payments case, the user will group by month created date of their first cash request (first date)

Creation

Create a new table where:

- Search for the first cash\_request
- Pair user\_id and first\_request date

Merge tables and extract month and year

First\_request == cohort\_date