IronHack Payments - Cash Request & Fees Datasets

Exploratory Data Analysis

1. Introduction

a. Source

We are given two datasets (cash request and fees) from the company and a lexique explaining the columns in each dataset. The collection of its data most probably comes from the software used to provide the lending service.

b. Scope

Cash requests can be traced back from November 2019 to November 2020.

We can know if the CR was instant (incurring a fee) or regular.

With the status money back we can know if the loan was paid back or not.
Loans get paid back in their full amount at a set date and incur fees if postponed or requested as instant.

Fees can be traced back from May 2020 to November 2020.
Every fee is associated with its corresponding cash request id and amount charged (always 5€), there can be multiple fees per cash request.

We can know the type of fee and if it was accepted or not (generated revenue or not).
The type split payment is not yet in the database (no fee is of type split_payment), so we can conclude that it is not yet an option for users to pay in multiple instalments.

c. Purpose

The purpose of this exploratory data analysis is to gain insights about user behaviour by cohort overtime and the performance of the financial services with these metrics :

0. **Creating a Cohort metric :** Putting users in cohorts based on the first month of their first CR.

1. **Frequency of Service Usage:** Understand how often users from each cohort utilise IronHack Payments' cash advance services over time.

2. **Incident Rate:** Determine the incident rate, specifically focusing on payment incidents, for each cohort. Identify if there are variations in incident rates among different cohorts.

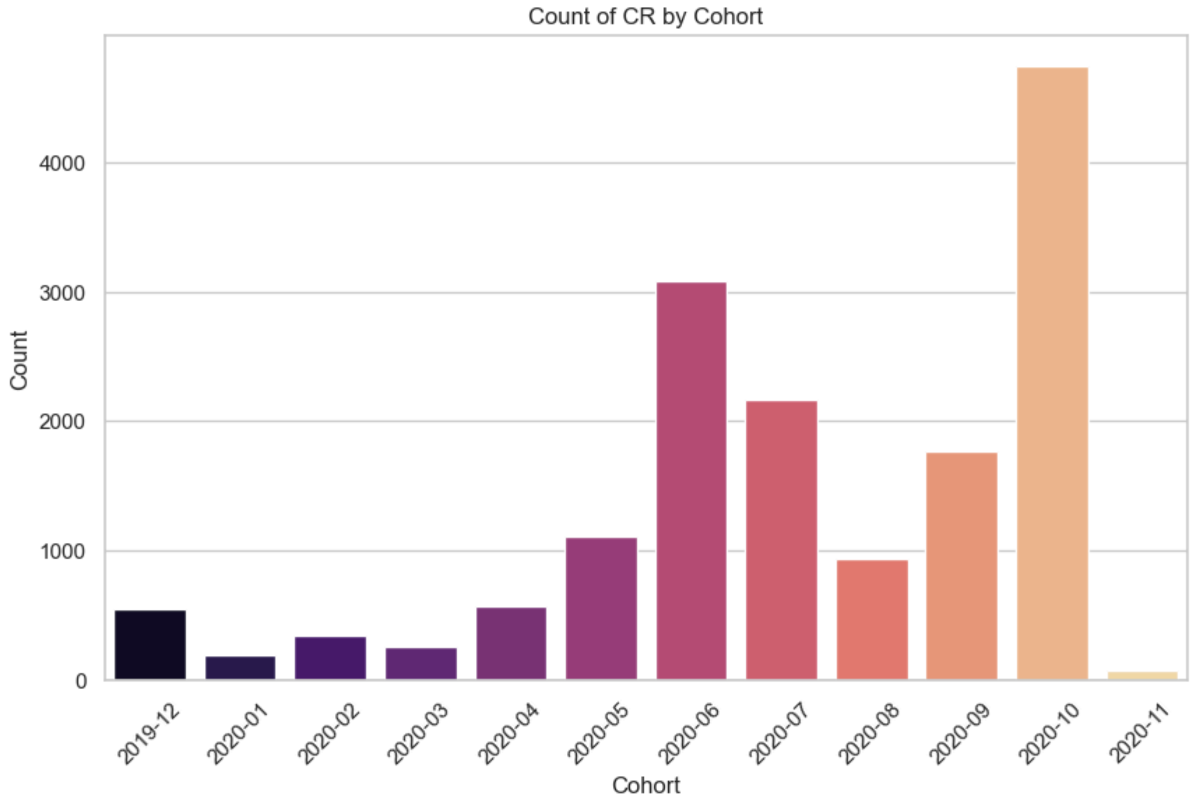
3. **Revenue Generated by the Cohort:** Calculate the total revenue generated by each cohort over months to assess the financial impact of user behaviour.

4. **New Relevant Metric:** Propose and calculate a new relevant metric that provides additional insights into user behaviour or the performance of IronHack Payments' services.

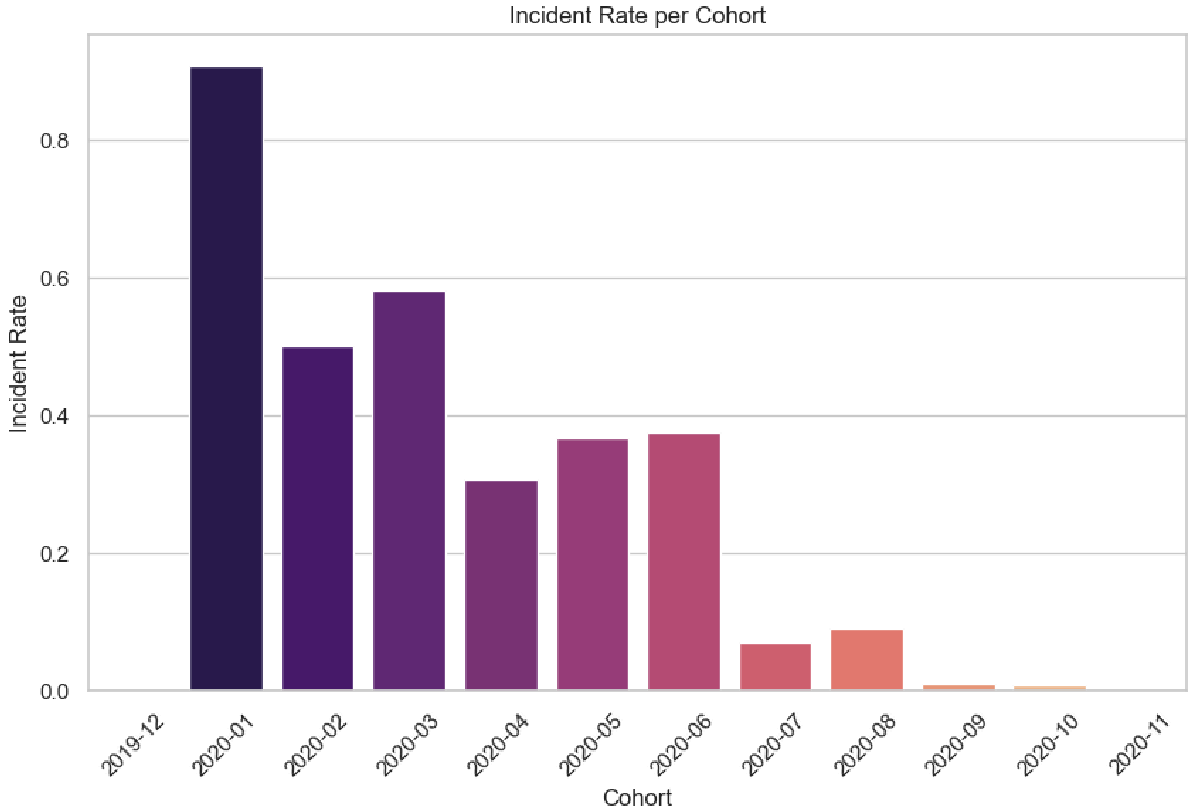
2. Comprehensive understanding

Key statistics

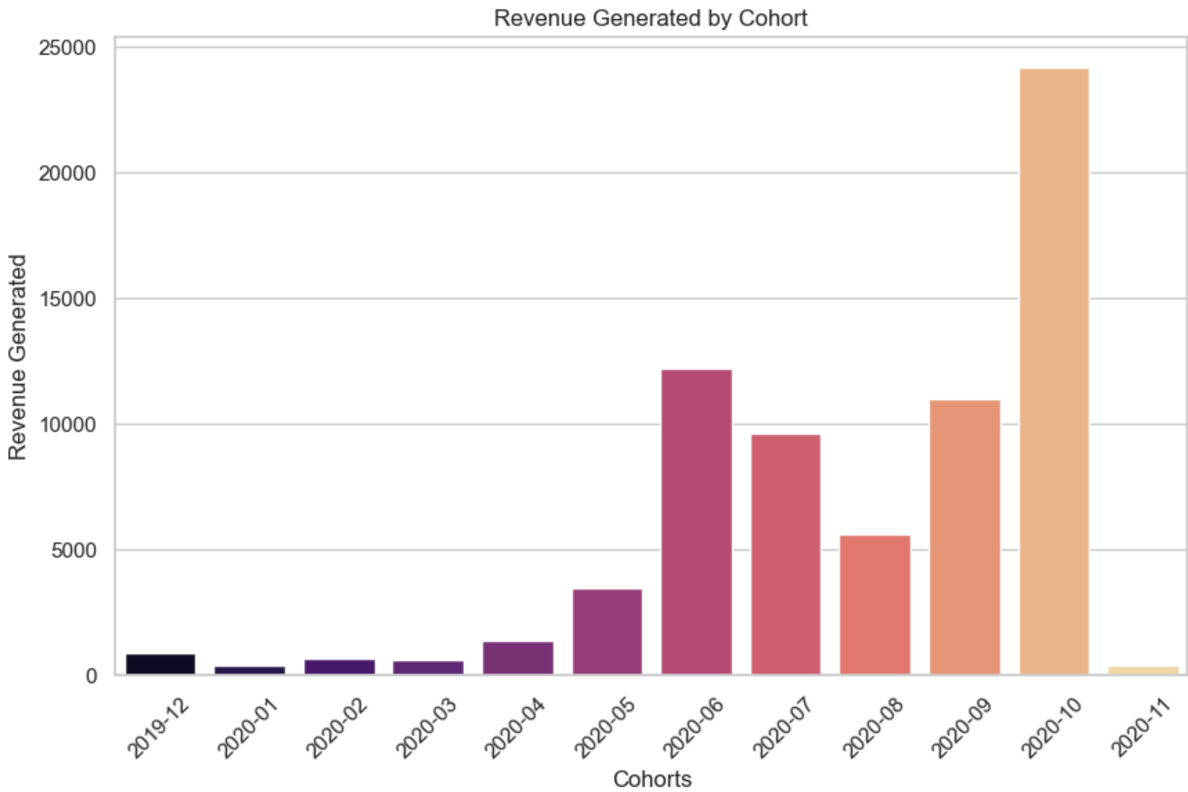
1. **Frequency of service usage per cohort**



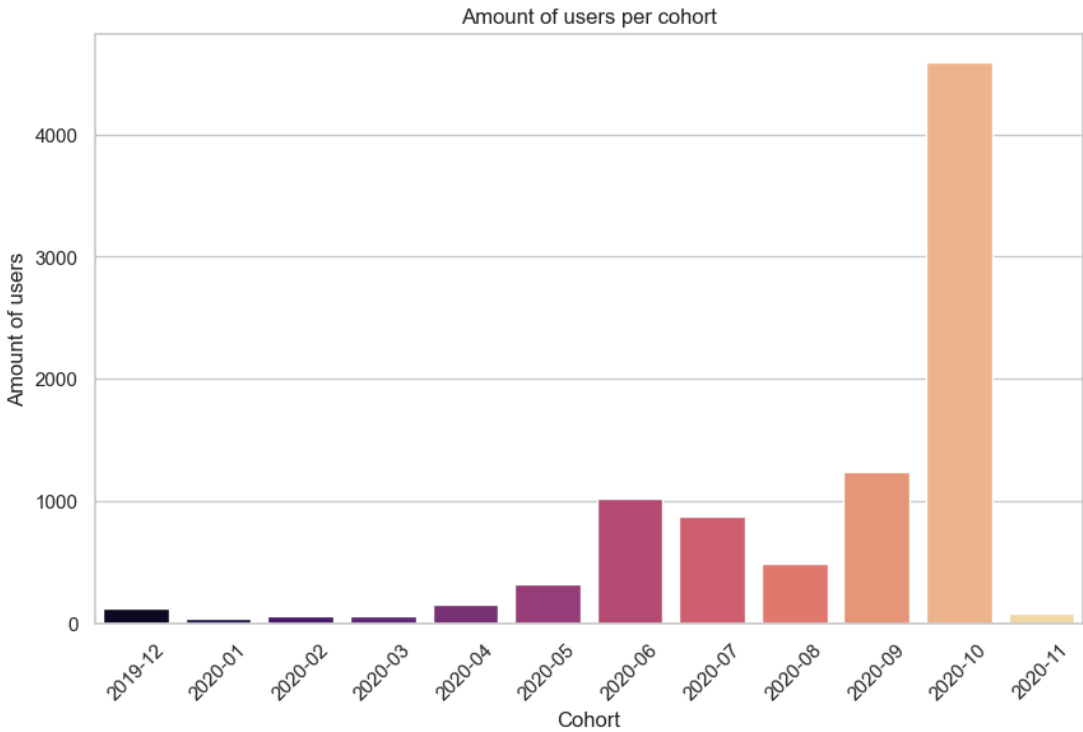
2. Payment incident rate for each cohort



3. Revenue generated by each cohort



4. New relevant metrics



I chose to include the average % return on loans as my main metric which is **8.12%**.

Total capital lended to users from November 2019 to November 2020 is : **1’132’673€**

Total revenue generated rom May 2020 to November 2020: **70’095€**

17290 Cash requests were made in total from November 2019 to November 2020

Suggestions for the business' growth :

- Fees are always 5€, implementing a system where fees are a percentage of the amount lended at the start and a set cost if the user misses a moneyback date or requests an instant loan could benefit the business, especially for borrowers of low amounts.

- Offering to pay in multiple instalments can motivate users to borrow larger amounts, an increasing % fee can be interesting as we offer more flexibility to the user and take on more lending risk.

- There is a large amount of cases where a cash request is rejected, adapting the lending process to be more automated and with less friction can benefit the business (eliminating human manual reviews if possible, including more payment accounts from different apps for the user to receive and pay back the loan)

- Instant payment and postponing the repayment date are the two main reasons for the generation of revenue through fees, we can explore increasing one or both of those fees costs or offer different plans with their own fees based on those options.

Data Quality Analysis and Cleaning Report

1. Data Quality

The only information provided about the users is regarding their lending activity in the platform, we do not know their age, location or other categorical metrics about them making the analysis potential purely quantitative.

Getting additional categorical tags on users could expand our analysis on many aspects (Ex. Marketing insight).

We can know how many cash requests each user id has requested, the date at which it was requested and if it was accepted or not.

Fees are missing data from november 2019 to may 2020 compared to the dates range in cash requests. Meaning we can only know the revenue and its growth for a six month period (from may 2020 to november 2020).

2. Data Cleaning / Filtering / Joins

Filtering / cleaning the cash request dataset :

- To dive into the actual financial activity of the company I have decided to exclude the transactions with a status that indicates the cash request was not accepted or cancelled from the cash request dataset. That way we ensure we are looking at active users and cash requests that have gone through

Adding columns :

- A cohort column was created based on the month of the first user’s cash request since user’s do not get assigned one. This allows us to perform the analysis by user antiquity.
- A ‘Num, of CR’ column was added to pair each user id with the total amount of cash requests his account has done. This allows us to see if customers are returning to our lending business and compute the returning customer rate.
- A ‘Num. of Incidents’ columns was added to match every cash request id with the number of incidents that generated fees. This allows us to get the incident rate and also the revenue generated by incidents of type fees.

Merging the cash request and fees dataset by the cash request id:

- Before merging both datasets I renamed columns with similar names to ensure we can keep track of what dataset their values correspond to.
- Status and created at saw a CR or Fees appended to retain readability after the merge.
- I then merged both datasets by the cash request id, generating a lot of NaN values but retaining as much information as possible.