



# IronHack Payments



Project II  
Data Science & Machine Learning Bootcamp April  
2024



# Challenge

- Export and clean data for IronHack Payments
- EDA and QA of data set
- Analyze relevant metrics
- Data visualisation in Python
- Create Tableau dashboard

# Create analysis for top KPIs:

- Frequency of Service Usage
  - Incident Rate
  - Revenue Generated by the Cohort
  - Relevant Metric
-

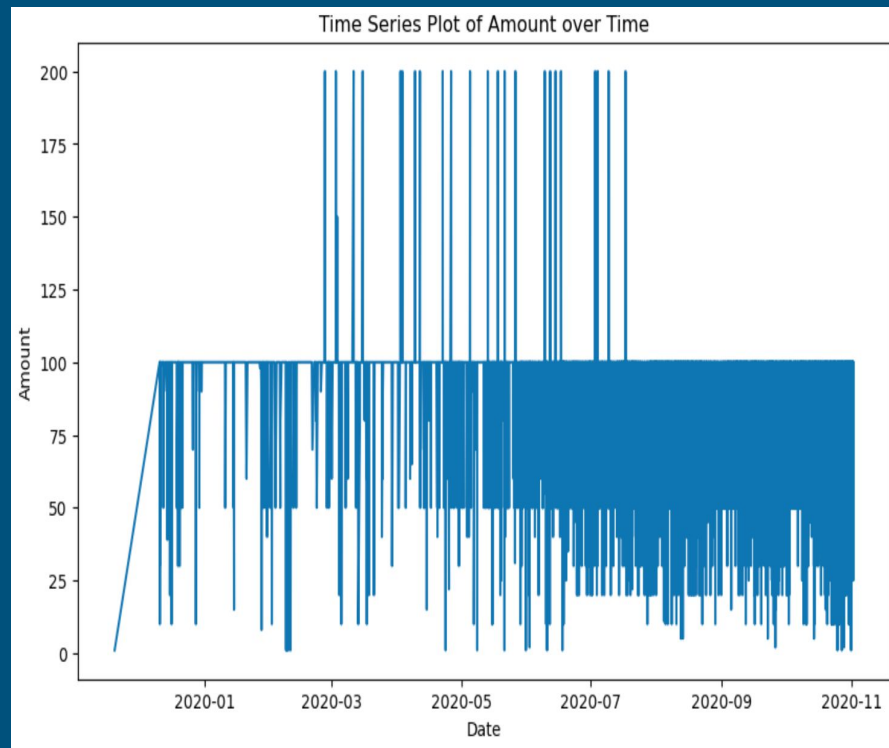
# Cleaning & Quality Analysis

- Date format has been converted from object to DateTime for both files
- Year and Month from created\_at have been concatenated to a separate column 'cohort'
- NaN values have been replaced by 0 for the purpose of calculations
- Missing values in user\_id column have been dropped accounting for approx. 9,6% data loss (deleted user\_id)

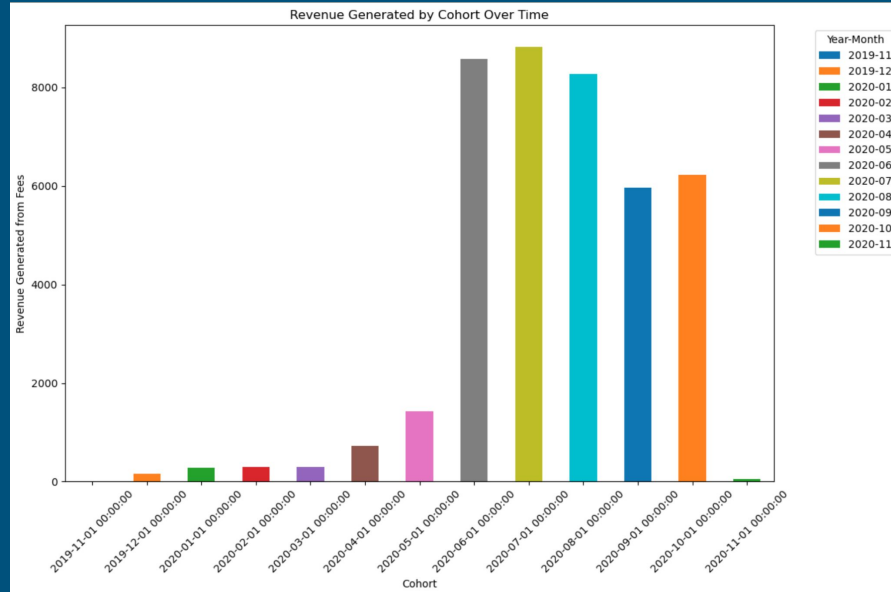
```
Data columns (total 16 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id                                           23970 non-null  int64
1   amount                                       23970 non-null  float64
2   status                                       23970 non-null  object
3   created_at                                   23970 non-null  object
4   updated_at                                   23970 non-null  object
5   user_id                                       21867 non-null  float64
6   moderated_at                                 16035 non-null  object
7   deleted_account_id                          2104 non-null   float64
8   reimbursement_date                         23970 non-null  object
9   cash_request_received_date                 16289 non-null  object
10  money_back_date                             16543 non-null  object
11  transfer_type                               23970 non-null  object
12  send_at                                     16641 non-null  object
13  recovery_status                             3330 non-null   object
14  reco_creation                               3330 non-null   object
15  reco_last_update                           3330 non-null   object
dtypes: float64(3), int64(1), object(12)
memory usage: 2.9+ MB
None
id                                           0
...
reco_last_update                           20640
dtype: int64
Number of Columns with Missing Values: 9
All Columns Have Missing Values: False
```

# Exploratory Data Analysis

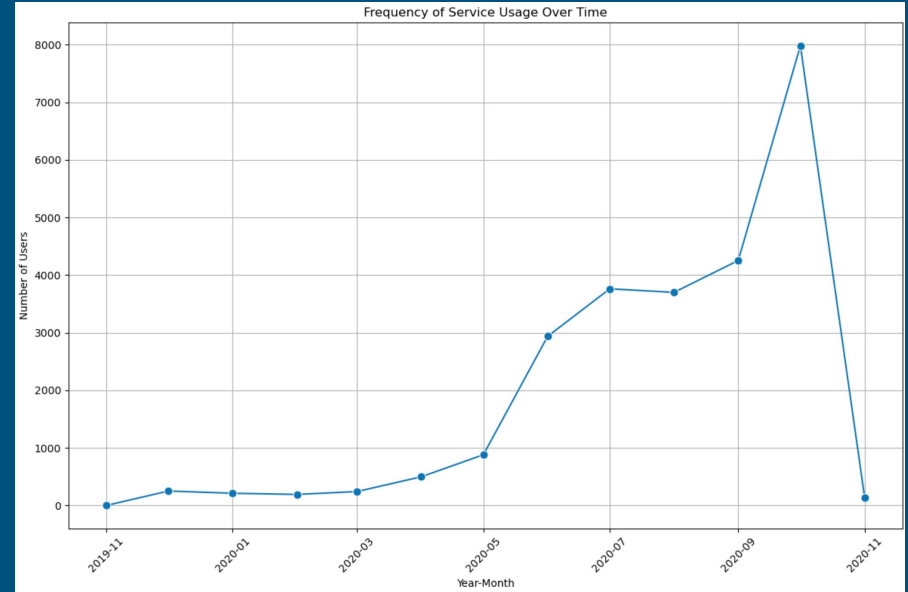
- Amount requests over 125 are outliers
- Most of outlier amount requests happened between 03/2020 - 09/2020
- Decided cohorts are grouped by months
- User\_id used for calculations of fees
- Revenue for instant, postponed and incident



# Revenue Generated by Cohorts

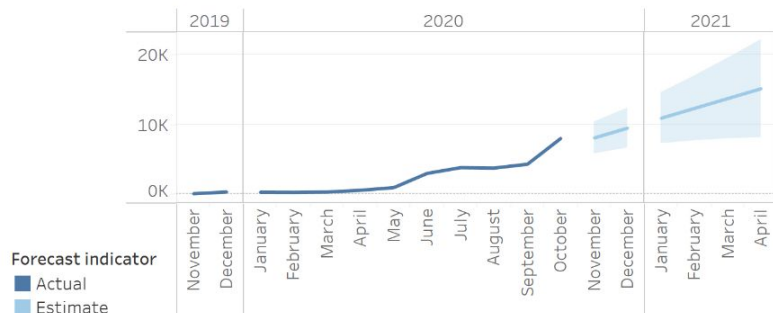


# Frequency of Service Usage

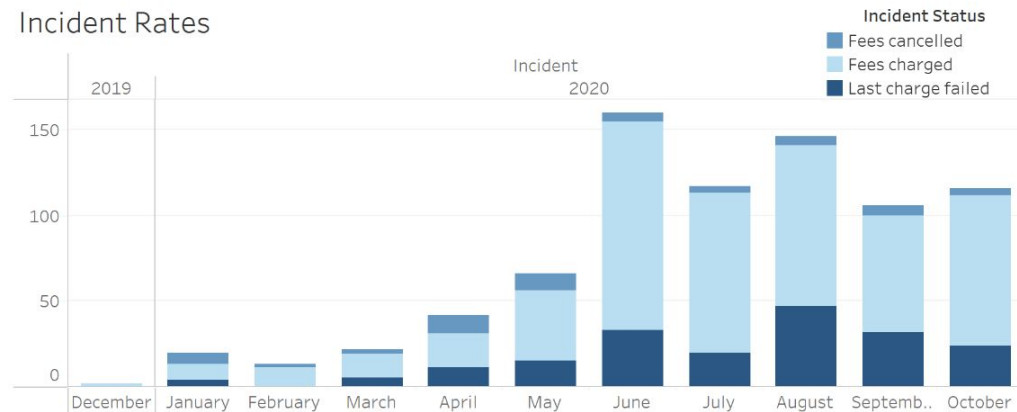


# IronHack Payments Overview

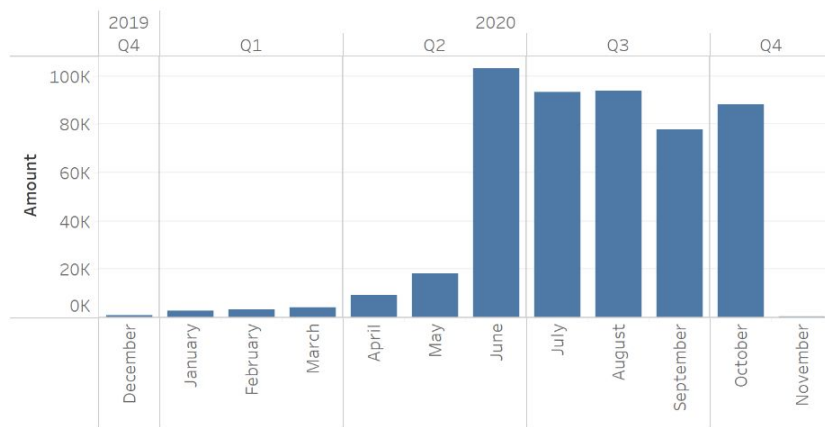
## Frequency of Service Usage



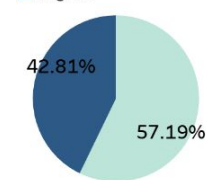
## Incident Rates



## Revenue by Month



## Transfer Type



## Transaction Status

|                        | 2019   | 2020   |
|------------------------|--------|--------|
| Transaction Failed     | 46.79% | 21.26% |
| Transaction Successful | 53.21% | 78.74% |

## Failed Transactions

|                       |        |
|-----------------------|--------|
| canceled              | 0.10%  |
| direct_debit_reject.. | 12.77% |
| rejected              | 86.45% |
| transaction_declined  | 0.68%  |



# Thank you!



Adrianna Domaszewicz

