

# Business Challenge: Cohort Analysis for Ironhack Payments (Project 1)

## Introduction

IronHack Payments, a forward-thinking financial services company, has been offering innovative cash advance solutions since its inception in 2020. With a commitment to providing money advancements for free and transparent pricing, IronHack Payments has garnered a substantial user base. As part of their continuous effort to enhance their services and understand user behavior, IronHack Payments has commissioned a cohort analysis project.

## Project Overview

In this project, you will conduct a comprehensive cohort analysis based on data provided by IronHack Payments. The main objective is to analyze user cohorts defined by the month of creation of their first cash advance. You will track the monthly evolution of key metrics for these cohorts, enabling IronHack Payments to gain valuable insights into user behavior and the performance of their financial services.

### Metrics to Analyze

You will calculate and analyze the following metrics for each cohort:

1. **Frequency of Service Usage:** Understand how often users from each cohort utilize 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 behavior.
4. **New Relevant Metric:** Propose and calculate a new relevant metric that provides additional insights into user behavior or the performance of IronHack Payments’ services.

### Data Analysis Tools

You are expected to perform the cohort analysis using Python, primarily leveraging the Pandas library for data manipulation and analysis. However, the main analysis should be conducted using Python.

### Exploratory Data Analysis (EDA)

Before diving into cohort analysis, conduct an exploratory data analysis to gain a comprehensive understanding of the dataset. Explore key statistics, distributions, and visualizations to identify patterns and outliers. EDA will help you make informed decisions on data preprocessing and analysis strategies.

### Data Quality Analysis

Assess the quality of the dataset by identifying missing values, data inconsistencies, and potential errors. Implement data cleaning and preprocessing steps to ensure the reliability of your analysis. Document any data quality issues encountered and the steps taken to address them.

### Deliverables

1. **Python Code:** Provide well-documented Python code that conducts the cohort analysis, including data loading, preprocessing, cohort creation, metric calculation, and visualization.
2. **Tableau Dashboard**: Publish a dashboard in Tableau Public regarding your analysis.
3. **Exploratory Data Analysis Report:** Prepare a report summarizing the findings from your exploratory data analysis. Include visualizations and insights that help understand the dataset.
4. **Data Quality Analysis Report:** Document the results of your data quality analysis, highlighting any issues and the steps taken to resolve them.
5. **Short Presentation:** Create a concise presentation (maximum of 4 slides) summarizing your findings from the cohort analysis and key insights gained from EDA and data quality analysis. This presentation should be suitable for sharing with the IronHack Payments team.

### Bonus:

1. **Operationalize your analysis**: Make sure all the code is in a .py that can be called from the Terminal and whose execution makes sense (if in doubt, ask the Teacher for clarification on this)
2. **StreamLit**: Read about the StreamLit package and create a StreamLit app about this data (you can leverage on ideias from your dashboard)
3. **OPP vs Function**: Take your code and replicate it using an oposite strategy than you have done.