Ironhack Payments
Cohort Data Analysis

By: Debora, Martyna, Zhoubin

Project Overview

Objectives

- Practicing the recent data manipulation skill sets

- Gaining Generic Insights from a Complex Data

- First-hand Experience at Creating Cohorts

- Recognizing the key metrics

- Presenting the Findings in Streamlit

Project Overview

Key Learnings for Future Projects

- ○ Discover Patterns in User Cohort Interactions with Cash
  Advance Services

- ○ Drive Revenue by Quantifying the Financial Impact of User
  Cohorts

- ○ Ensure Data-driven Insights and Improve Decision-making
  for Service Enhancement

Implementation

- Exploratory Data Analysis (EDA) on a pair of pre-processed datasets from Ironhack Repository
  - Cash request
  - Fees
- Performing Cohort Analysis for Data Insight
  - Cohort calculated on the month of first request
- Extracted Metrics
  - Frequency of Service Usage
  - Calculating the Incident Rate
  - Calculating Revenue Generation
- Visualization
  - Line Plots
  - Bar Charts
  - Histograms

Challenges

- A primary encounter with the dataset, in particular finding the connection between the variables, was a bit challenging
- Some terms such as Incident rate appear ambiguous
- Cleaning the datasets and merging them on common key values appeared a bit puzzling
- Converting the cohort month column to a datetime object was not immediate- got us stuck for hours!
- The mix of discrete and continuous numeric values makes visualization a bit challenging
- Deciding on which data to be cleaned/removed (e.g. non-existing elements in user_id)
- Trouble with configuring the local Python environment via console commands on macOS

Tools

- **Pandas** Library for Data Manipulation and Analysis, particularly on DataFrames
- **Matplotlib** Library for basic plotting
- **Seaborn** Library with Advanced Plotting Features such as Mapping and Statistical Aggregation to produce informative plots
- **Numpy** Library for working with Numerical Arrays
- **Streamlit** Library that makes it easy to create and share custom web apps
- **Some knowledge** of descriptive statistics such as aggregate functions (e.g. min, max, mean, std,...) and well-known plots (e.g. line, bar, histogram)
- **Some knowledge** of data story telling

Streamlit demo

Conclusion

Dealing with financial and business data for the purpose of gaining insight, decision-making, and devising strategies can be particularly challenging and requires significant expertise beyond our current knowledge. Through this project, however, we had the opportunity to discover patterns in data and implement well-known plotting and graphical tools for visualization.

Management and Teamwork

- **Debora** mainly responsible for data manipulation and insight
- **Martyna** mainly responsible for data analysis and partly coding
- **Zhoubin** mainly responsible for coding and developing Streamlit interface

Thank You!

Questions and comments are welcome :)