**1. Initial Setup**

* Confirm understanding of project objectives:
  + Analyze user behavior and performance of financial services.
  + Focus on cohorts based on the month of their first cash advance.
* Gather required data from Ironhack Payments.
* Review the data-sharing protocols and permissions.

**2. Data Quality and Preprocessing**

* Conduct a **Data Quality Analysis**:
  + Check for missing values, inconsistencies, and duplicates.
  + Document data quality issues and the steps to resolve them.
* Clean the dataset:
  + Remove or address inconsistencies.
  + Handle missing values appropriately.
  + Verify accuracy of date fields and transaction values.
* Create derived columns for cohort analysis:
  + First cash advance month.
  + Service usage frequency metrics.
* Ensure data format is compatible with analysis tools (e.g., Python, Tableau).

**3. Exploratory Data Analysis (EDA)**

* Perform descriptive statistics on key metrics (e.g., usage frequency, revenue).
* Visualize data:
  + Monthly trends for each cohort.
  + Distribution of incident rates.
  + Revenue patterns over time.
* Identify outliers and trends that need further exploration.
* Document findings in an EDA report, including key patterns and any hypotheses.

**4. Cohort Analysis Execution**

* Define cohorts by their first cash advance month.
* Calculate:
  + **Frequency of Service Usage**: Monthly activity rates per cohort.
  + **Incident Rates**: Variability in payment incidents.
  + **Cohort Revenue**: Total revenue generated per cohort over months.
  + **Proposed New Metric**: Brainstorm, implement, and calculate (e.g., customer retention rates or average revenue per user).
* Visualize metrics over time for each cohort:
  + Line charts or heatmaps for service usage trends.
  + Revenue graphs by cohort.

**5. Reporting and Presentation**

* Prepare the following deliverables:
  + **Python Code**: Ensure it includes data loading, cleaning, cohort creation, metric calculation, and visualization.
  + **Tableau Dashboard**: Interactive views for the Ironhack Payments team.
  + **EDA Report**: Include visualizations and insights.
  + **Data Quality Report**: Document identified issues and resolutions.
  + **Presentation Slides** (maximum 4 slides):
    - Key findings and insights from cohort analysis.
    - Recommendations based on the findings.

**6. Review and Iteration**

* Share findings and visualizations with stakeholders for feedback.
* Iterate on analysis or visualizations based on feedback.

**7. Bonus (Optional Enhancements)**

* Create a **Streamlit App** for dynamic data interaction.
* Implement an operationalized .py script:
  + Include clear documentation and execution instructions.
  + Test script functionality in a terminal environment.
* Convert code to use an alternate strategy (e.g., object-oriented if the current is functional, or vice versa).

**8. Submission and Handover**

* Verify all deliverables are completed and uploaded.
* Prepare final documentation for the Ironhack Payments team.

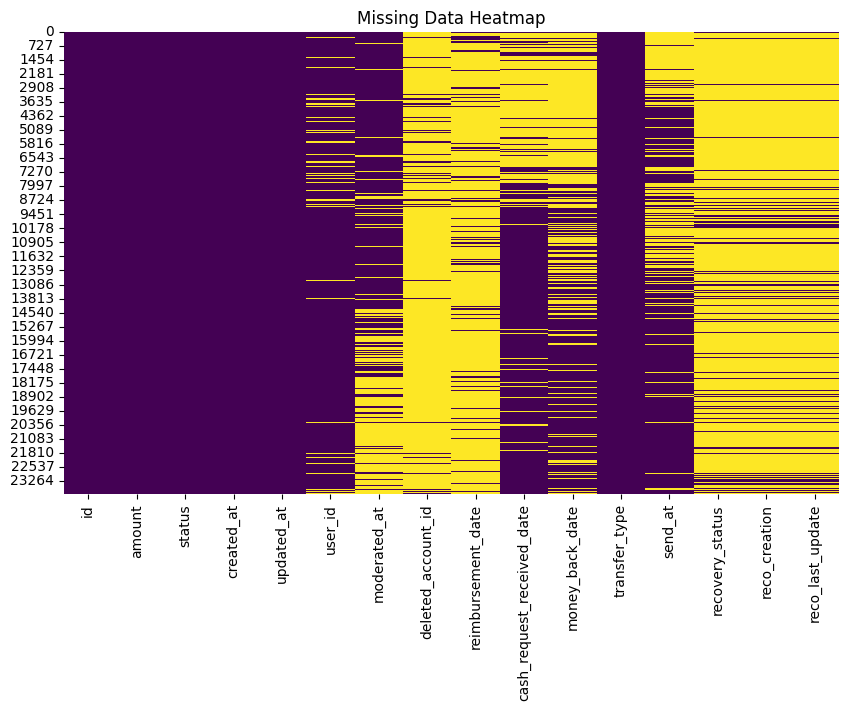
EDA Notes.

Cash-request

* + Missing column ‘reason’ in cash request file.
  + Date format **ISO 8601**

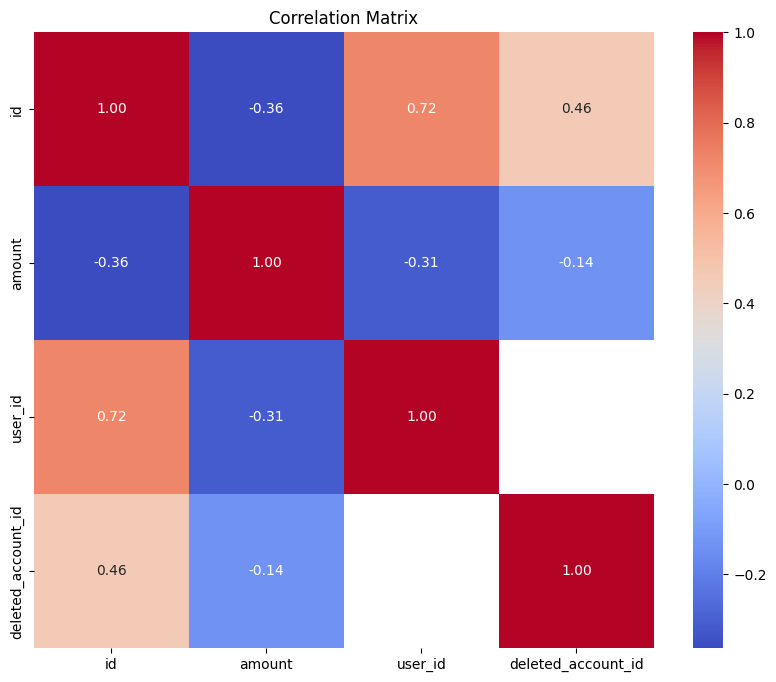
### ****Structure of the Format****

1. **Date**: YYYY-MM-DD (e.g., 2019-12-10)
   * YYYY: Year
   * MM: Month
   * DD: Day
2. **Time**: hh:mm:ss.sssssss+zz:zz
   * hh: Hours (24-hour format)
   * mm: Minutes
   * ss.sssssss: Seconds (with microseconds or milliseconds precision)
   * +zz:zz: Timezone offset from UTC (e.g., +00:00 means UTC)



**user\_id, reimbursement\_date, deleted\_account\_id have significant missing data**

id and amount are almost entirely complete



**id and user\_id**:

* Correlation = **0.72**: A strong positive correlation, suggesting that the id and user\_id are likely related or follow a sequential pattern.

