IronHack Payments Cohort Analysis Report

1. Exploratory Data Analysis (EDA) Report

Overview of the Dataset:

- **Total Records:** 26,598
- **Columns:** 28, including transaction amounts, statuses, user IDs, dates, and more.

Key Findings:

- **Distribution of Transaction Amounts:**
  - The distribution is right-skewed, with the majority of transactions being smaller amounts.
  - Most common transaction amounts are concentrated in the lower range (e.g., around $50 to $100).
  - Indicates a preference for smaller cash advances, possibly due to user cautiousness or short-term financial needs.
- **Status of Cash Requests:**
  - Several categories exist for cash request statuses, including `rejected`, `accepted`, and others.
  - The frequency of these statuses varies, with certain statuses being more prevalent, such as `money_back`.
- **Time-Based Patterns:**
  - Users' activity is spread across different months, with some months showing higher user engagement.

2. Data Quality Analysis Report

Data Quality Issues Identified:

- **Missing Values:**
  - Several columns have missing values. Notably:
    - `deleted_account_id` has 24,494 missing values.
    - `moderated_at`, `cash_request_received_date`, `money_back_date` have a significant number of missing entries.
  - **Resolution:** For critical columns like `amount`, rows with missing values were dropped. For categorical columns like `status_cash_request`, missing values were filled with 'unknown' to maintain data integrity.
- **Data Consistency:**
  - The date fields contained time zone information, which was dropped to ensure consistency in date processing.

Actions Taken:

- Rows with missing `amount` were dropped to ensure the reliability of revenue-related analysis.
- Missing values in `status_cash_request` were filled with 'unknown'.
- Time zone information was removed from datetime fields to simplify date analysis.

3. Cohort Analysis Insights

Cohort Definition:

- Users are grouped into cohorts based on the month of their first cash advance request (`cohort_month`).

Key Metrics and Insights:

1. **Average Usage per User Over Time:**
   - Measures the average number of times users in each cohort used the service.
   - **Observation:** The average usage per user fluctuated slightly over time. Some cohorts show higher engagement, indicating that certain periods may attract more frequent users.
2. **Incident Rates by Cohort:**
   - The proportion of transactions that were `rejected` for each cohort.
   - **Observation:** Incident rates vary across cohorts. Some cohorts have higher rates of rejected requests, suggesting either a change in user behavior or modifications in service policies.
3. **Total Revenue by Cohort:**
   - The total revenue generated by each cohort over time.
   - **Observation:** Revenue contributions vary by cohort, with some months generating significantly higher revenue. This may correlate with user engagement levels or the overall economic climate during those periods.
4. **Retention Rates by Cohort:**
   - The percentage of users making additional transactions after their first request.
   - **Observation:** Retention rates varied across cohorts, with some cohorts showing higher rates of repeat usage. This could indicate the effectiveness of the service in retaining users or differences in user needs.

Overall Insights:

- **User Engagement:** Certain cohorts show higher average usage and retention rates, suggesting potential periods of high user satisfaction or effective service marketing.
- **Risk and Incident Management:** Variations in incident rates highlight the need for further investigation into factors influencing request rejections.
- **Revenue Trends:** Monitoring revenue by cohort helps identify the most profitable user groups and optimize marketing strategies accordingly.

Recommendations:

- **Further Analysis:** Investigate cohorts with high incident rates to identify potential causes and improve user experience.
- **Targeted Marketing:** Focus on cohorts with high average usage and retention rates for targeted marketing campaigns.
- **User Feedback:** Collect feedback from cohorts with lower engagement to understand barriers and improve service offerings.