Cohort Analysis for User Retention

Name : Imran Bin Salam
Role : Data Analyst Intern
Date : 01/30/2026

1. Introduction & Objective

Cohort analysis is a **behavioral analytics technique** used to group users based on a shared characteristic (most commonly **signup date**) and track their behavior over time.

Objective of This Analysis

- Measure **user retention** over time
- Identify **churn patterns**
- Understand **long-term vs short-term user behavior**
- Provide **actionable retention improvement strategies**

---

2. Dataset Description

Since no real dataset was provided, a **synthetic but realistic dataset** was created.

Dataset Structure

Column Name Description

user_id	Unique user identifier
signup_date	Date user registered
activity_date	Date user was active

Key Assumptions

- Users can have multiple activity records
- Activity spans up to **6 months** after signup
- Dataset simulates a SaaS / application usage model

---

3. Cohort Analysis Theory (Important for Interviews)

What is a Cohort?

A **cohort** is a group of users who share a common event during the same time period.

Example:
Users who signed up in **January 2023** form one cohort.

What is Retention?

Retention measures:

How many users return after their first interaction?

Retention Rate Formula:

$$Retention = \frac{\text{Active users in month N}}{\text{Total users in cohort}}$$

What is Churn?

Churn is the **inverse of retention**:

$$Churn = 1 - Retention$$

---

4. Step-by-Step Methodology

Step 1: Assign Cohorts

- Extract **signup month**
- Extract **activity month**
- Convert both into YYYY-MM format

Step 2: Create Cohort Index

The cohort index represents **number of months since signup**.

$$Cohort\ Index = (Activity\ Month - Signup\ Month) + 1$$

Cohort Index Meaning

1	Signup month
2	First month after signup
3	Second month after signup

Step 3: Build Cohort Table

- Count **unique active users**
- Pivot data to form a **cohort matrix**
- Normalize values to calculate **retention percentages**

5. Cohort Retention Table (Conceptual)

Signup Month	Month 1	Month 2	Month 3	Month 4
Jan-2023	100%	62%	45%	31%
Feb-2023	100%	58%	41%	29%
Mar-2023	100%	65%	49%	34%

6. Retention Heatmap (Visual Analysis)

Why Heatmaps?

- Quickly identify **drop-off intensity**
- Compare **cohort performance visually**
- Darker color = better retention

Visual Insights

- Steep drop between **Month 1 → Month 2**
- Retention stabilizes after Month 3
- Indicates **early churn problem**

7. User Behavior Interpretation

Key Observations

1. **High Initial Drop-off**
   - Most users churn after first month
   - Indicates weak onboarding or unclear value proposition
2. **Retention Stabilization**
   - Users who stay beyond Month 3 tend to remain loyal
3. **Consistent Pattern Across Cohorts**
   - Suggests **systemic product issue**, not seasonal

8. Business Impact of Findings

Insight	Business Meaning
Early churn	Poor onboarding experience
Stable long-term users	Strong core value
Similar cohort trends	Product-driven issue

---

9. Retention Improvement Recommendations

1. Improve Onboarding

- Interactive walkthroughs
- Product tours
- First-day activation goals

2. Early Engagement Campaigns

- Email nudges in first 7 days
- Push notifications for inactive users
- In-app tips

3. Feature Adoption Tracking

- Monitor which features retained users use
- Push those features earlier

4. Loyalty Programs

- Rewards after 30 or 60 days
- Gamification strategies

---

10. Tools & Technologies Used

- **Python**
- **Pandas** – data manipulation
- **Matplotlib** – heatmap visualization
- **Jupyter Notebook** – analysis documentation