Cohort Retention Analysis (Python)

By: Tushar Kalra

Date: 22 Oct 2025

Objective: Measure month-over-month loyalty by user signup cohorts, and identify retention cliffs.

Executive Summary

We compute cohort-month retention and visualize as a heatmap. This reveals early churn vs sticky cohorts and informs lifecycle interventions.

Definitions

- Cohort Month: user first-order month.

- Cohort Index: months since first order (1,2,3...).

- Retention %: active users in month n / cohort size.

Pandas: Cohort Pivot Steps

```
df['cohort_month']                                                =
df.groupby('user_id')['order_date'].transform('min').dt.to_period('M')
df['order_month']  = df['order_date'].dt.to_period('M')
df['cohort_index'] = (df['order_month'].dt.to_timestamp() -
                      df['cohort_month'].dt.to_timestamp()).dt.days // 30 + 1
cohort = (df.groupby(['cohort_month','cohort_index'])['user_id']
          .nunique().unstack(fill_value=0))
retention = (cohort.divide(cohort.iloc[:,0], axis=0).round(3))
```

Reading the Heatmap

- Row = signup month; Column = months since signup.

- Top-left to right: decay curve. Look for sharp early decay.

Playbook

- If Month-2 drop steep: activate winback campaigns.

- If a cohort outperforms: study acquisition source and replicate.