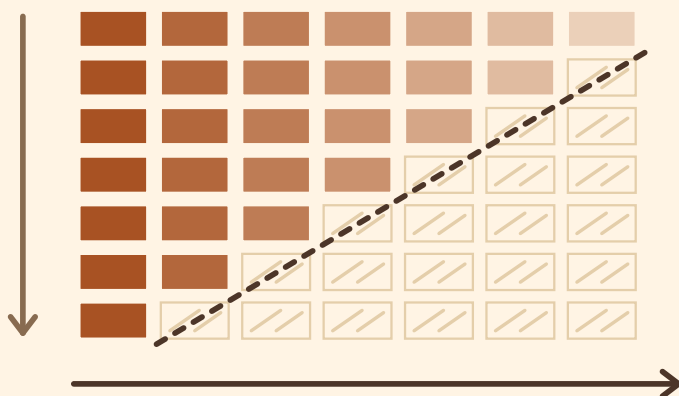




BUSINESS AND CUSTOMER ANALYTICS

Measure Retention with Cohort Analysis

Keep membership fixed, distinguish incomplete periods, and explain what return means



Sanjaay Gowda T S

FREE SAMPLE / INTRODUCTION + CHAPTER 1 / EDITION 1.0.0

sanjaaygowdatai.com

About this free sample

Measure Retention with Cohort Analysis

Sanjaay Gowda T S

First edition 1.0.0, 2026-10-13

Copyright 2026 Sanjaay Gowda T S. This is a labelled free sample, not the complete ebook. You may share this sample unchanged. It contains the cover, sample permissions, contents, introduction, Chapter 1 and a closing product link. The remaining chapters, worked solutions and companion files are not included. The full ebook text, diagrams and artwork have separate single-reader terms; this sample permission does not authorize sharing the full purchased PDF or EPUB.

The companion source code, tests and fictional data are separately MIT licensed. You may adapt or share that project with its included licence notice. MIT does not apply to the ebook text or book artwork. The books and companion projects have no technical activation requirement or buyer-specific watermark.

This first edition was prepared with AI assistance. The case studies are fictional teaching examples. Any measured outputs come from the supplied examples, not production customer data. The teaching scope, assumptions and limitations are explained in the chapters. Independent human technical, editorial and assistive-technology reader review is pending. Automated tests and accessibility checks do not establish certification or suitability for every use.

Run examples in a separate working folder using the supplied synthetic data. Review inputs and calculations before adapting the techniques to real operations. Model software, service access, subscriptions, ongoing updates and consulting are not included.

Author website: sanjaaygowdats.ai

Ebook terms: sanjaaygowdats.ai/ebooks/license

Inside this free sample

Start here.....	4
1. Define membership and return.....	5
Continue with the full guide	6

This free sample contains the introduction and Chapter 1 only. The closing link opens the complete guide. The companion project and worked solutions are supplied with the complete ebook.

Start here

Retention becomes meaningful only after you define who enters a group and what counts as a return. A colorful table can conceal those decisions. It can also turn an incomplete month into an apparent loss, count repeated events as repeated people or compare groups that have had different opportunities to return. This book makes the definitions visible before drawing conclusions.

The worked case is Cedar Workspace, a fictional collaboration service. Four synthetic identities join during January and February 2026. Their fictional events continue through March. Every identity, organization, date, quantity and proposed decision in this case is teaching material. The small dataset is not a benchmark or evidence about an actual business. Its purpose is to make fixed denominators, calendar age and incomplete periods easy to inspect.

You will build a calendar-month activity retention table. Membership is fixed by join month. A return means at least one qualifying activity in a complete calendar month. The tool counts a user once per month even if the event list contains several activities. Future cohort ages are null, not zero. This is one useful definition of retention, not the definition for every product.

Read the nine chapters in order, completing each exercise before reading its solution. The reasoning needs only basic arithmetic. The optional companion requires Python 3.10 or later, uses the standard library and makes no network requests. Keep the original fixture unchanged and experiment on copies. The expected output and behavioral tests help distinguish a changed definition from a programming mistake.

Your deliverable is a retention specification, a reproducible cohort table and a short readout that states what the table can and cannot establish. The table describes activity among supplied identities. It does not prove why activity changed or whether a product intervention caused improvement. Independent human technical and assistive-technology review is pending. For real datasets, use approved identity and event handling arrangements rather than treating these synthetic IDs as a model for personal-data storage.

1. Define membership and return

Ask a complete retention question

"Are customers retained?" omits both the group and the behavior. Cedar Workspace asks a narrower question: "Among identities joining in each calendar month, what share records a qualifying activity in each later complete calendar month?" The group is defined by join month, the counted entity is a supplied user identity and the return event is an activity. Each clause is part of the metric.

The official Google Analytics cohort documentation distinguishes inclusion and return criteria and describes several calculation styles [R1]. Our companion implements its own explicit calendar-month definition. Do not assume that its table matches a platform report merely because both are called cohort retention. Platforms can use different identity, event, timezone and inclusion rules.

Keep membership fixed

January contains A and B. February contains C and D. A returning in February stays in the January cohort; the return does not move A into February's membership. If the denominator changed to include only users still active, a decline could disappear because absent users were removed. Fixed membership keeps the original question intact.

An identity should enter this teaching analysis once. A reactivation question needs a different design if you want users to enter several cohorts after separate episodes. That can be useful, but repeated entries are no longer independent people or one permanent acquisition cohort. State the event that starts each episode and the entity being counted before writing code.

Choose a useful return behavior

Any tracked event may be too broad for the decision. An automatic background event can make an account look active even when nobody completed useful work. For a real workspace, a meaningful return might be creating or editing a shared document. The supplied teaching events have already been selected as qualifying activities. The calculator does not infer whether an event represents value.

Write a metric contract with membership event, return event, identity unit, time basis, observation cutoff and exclusions. Include the purpose: for example, identifying cohorts that deserve an onboarding investigation. A clear purpose helps reject a convenient metric that answers the wrong question. A login count might be easy to extract yet weak evidence that the workspace helped someone finish a task.

Exercise 1: complete the definition

Write a one-sentence retention question for the fictional service. Then list membership, return behavior and counted entity separately. Explain why a background notification event should not automatically qualify as a useful return, even if it appears in every user's activity log.

Continue with the full guide

This free sample ends after the introduction and Chapter 1. The complete guide contains nine teaching chapters, worked exercise solutions, primary references and a separate Python companion project. The full PDF, reflowable EPUB and project files are supplied together.

Use the complete guide to work through the remaining exercises and inspect the supplied examples. The companion code and fictional data are separately MIT licensed; the complete ebook has the single-reader terms described on its product page.

The examples are teaching fixtures, not production customer results. Independent human technical, editorial and assistive-technology reader review remains pending. Software access, subscriptions, consulting and guaranteed results are not included.

[View the complete guide and purchase options.](#)

You may share this labelled free sample unchanged. Keep the full purchased PDF and EPUB private. Thank you for reading.

Primary references

1. [R1] [GA4] Cohort exploration

<https://support.google.com/analytics/answer/9670133?hl=en>

Cohort inclusion and return criteria, time granularity and the distinction among standard, rolling and cumulative calculations.