

ETH Zürich
363-0403-00L
Introduction to Marketing
Fall Semester 2023/24

Continuous Performance Assessment (CPA)

General Information

The Continuous Performance Assessment (CPA) aims to enhance your knowledge and comprehension of (relationship) marketing. You will need to analyze data, apply concepts, frameworks, and methods such as customer lifetime value (CLV), and assist in marketing decision-making. The CPA project will account for 50 percent of your final grade.

The assignment is based on Netflix user data that were collected from users who had allowed the third-party application from [VOD Clickstream](#) to track their anonymized browsing behavior. The dataset includes the desktop and laptop activity of users in the United Kingdom (UK) in the time between 2017 and 2019. According to Netflix ([Kafka 2018](#)), 40 percent of its sign-ups come from desktop and laptop computers. The original data, published in 2021 under a CC-BY-NC-SA 3.0 IGO license, were modified for the purpose of this assignment.

Guidelines

This assignment is to be group work. There are two deliverables — a final project report containing the results of your analyses, and a short video of you presenting selected findings, with a focus on managerial implications. The final project report accounts for 67 percent of the total points for the CPA project and the video presentation accounts for the remaining 33 percent. Both deliverables are due on December 4, 2023, at 11:59 p.m. (23:59). Any submissions past the deadline might be counted as failed attempts.

The assignment is divided into different parts (i.e., questions) that guide you through the process of understanding, analyzing, and interpreting the data. The answers to the questions marked with an asterisk (*) are required for the final project report (see Table 1). We highly recommend that you also attempt to answer the remaining questions. This will help you prepare your presentations, allow you to gain experience applying marketing methods, and potentially benefit you in preparing for the end-of-semester exam.

The final project report provides detailed results of some of your analyses. A template will be provided for you, and you should enter your results in the appropriate fields, as indicated in the questions and instructions below. It is important that you follow these instructions and do not modify the template in any way, including the formatting of the file. Do not use the template for any calculations. If you copy and paste results from a different file into the designated fields of the template, remember to paste the values only (i.e., no formulas or formatting). Ensure that decimals are separated by a dot and not a comma and use at least four digits after the decimal point. Do not include any units, such as "\$". Note that questions that are more difficult carry a higher weight and that your answers for the final project

report may be graded for correctness and consistency. Table 1 in the Appendix shows the weighting of the different parts.

The second deliverable for the CPA project is a short video where you present selected results from your project. This video accounts for 33% of the total points. While the final project report focuses on quantitative results, the presentation requires you to translate those results into actionable recommendations for managers. A good way to do this is to (1) identify a managerially relevant problem, for example, based on an important or interesting finding that you encountered over the course of the project, (2) describe an appropriate method and any data required for analyzing the problem, (3) present results, (4) discuss business implications, and (5) make recommendations for improvement. You can choose any problem that you encountered while working on the project, including mandatory with an asterisk (*) and non-mandatory questions without an asterisk, or even identify a new problem on your own. Additionally, you can run analyses beyond the questions covered in the assignment.

Table 2 in the Appendix provides an example of the grading criteria used for video presentations. The video must not exceed six (6) minutes and should show you presenting, and any props used, such as slide decks. For example, you can record yourself using a Zoom meeting. It is crucial that you are clearly audible in the video. You can decide whether one group member or multiple members will present. Any content in the video exceeding six (6) minutes will not be graded. Videos played at a speed faster than normal will result in a grade deduction.

The deliverables must be submitted through Moodle, where you will be required to sign a declaration of originality. For group submissions, each member's contribution must be specified. Excerpts from your video presentations may be used during the class where we review your projects. By participating in the CPA project, you grant us permission to use the videos of your presentations in this class.

There will an initial check-in for the CPA for the whole class on October 31, 2023. During the project, we will provide a forum on Moodle where you can ask clarifying questions. Please consider that it might take some time for us to respond. Furthermore, we will be available on request on designated dates (see Syllabus) to discuss questions online or at the office WEV G 228 (by appointment). Please reach out to Patrick (pbachma@ethz.ch), or Sandro (sarnet@ethz.ch) with any problems or questions regarding group registration.

All times refer to local time in Zurich, Switzerland.

Assignments

Group*

[Enter the group number, student id, last name, and first name of each member of your group in the rows of the sheet “group”. Group numbers should be entered in numeric format, e.g., “1” or “2” not “group 1” or “group 2”, respectively. The student id should be entered in the regular format with dashes, e.g., “12-345-678”.]

Question 1 Understanding the Context

Different types of customer relationships require different approaches for estimating CLV. One distinction is between contractual and non-contractual relationships. In this assignment, we assume that Netflix has a contractual relationship with its customers. Based on what you know about the Subscription-Video-On-Demand (SVOD) industry and the design of streaming products, briefly discuss (i.e., in 3–4 sentences) whether this assumption is reasonable.

Question 2 Exploring the Data

The file has two sheets (i.e., “clv” and “filter_user”). The data in the sheet “clv” contains rows for different user-months combinations and have a time window of 12 months. We recommend that you number the months, with the earliest starting at zero (0). The last month of the time window is month eleven (11), which is also the current period, i.e., the present, for the most part of this assignment.

- *cohort* is the group of customers that was acquired in each month. For example, a user belongs to cohort 3 if s/he was acquired in month 3 (i.e., the 4th month).
- *user* identifies the different users.
- *time_year* is the year and *time_month* is the month
- *subscription* is binary (1–0) variable that takes on the value one (1) if the user had an (active) subscription at the time. This variable is always one (1) because we cannot observe the behavior of users after they cancel their subscription or after they switch to a device other than their desktop and laptop computers. For the assignment, we define that a user churns if s/he stops watching Netflix online, i.e., because s/he cancels the subscription or switches to a different device.
- *content* is the number of different titles (e.g., shows and movies) that a user consumed in a month.
- *genres* is the number of different genres that are associated with a user’s monthly Netflix consumption. This variable has a missing value if the user did not use Netflix in a particular month.
- *recency* is the share of recent content that a user watched in a month. The streaming of a title is recent if the time between the title’s release date and the date of the stream is less than two years.

- *bounce* is the bounce rate. It is a ratio that is calculated by dividing the number of times that a user clicked on a video and (almost) immediately stopped watching it by the user's total number of clicks in a month.

We would like to get an idea of the number of users in the dataset and how the size of the customer base changes over time.

- 2-a* How many different users did Netflix acquire during the one-year time window, i.e., what is the total size of the customer base in the dataset? [Enter your result in cell A2 in sheet "q2-a" of the template.] When does the time window of your data start? [Enter the month and year in cells B2 and C2 in sheet "q2-a" of the template, respectively.]
- 2-b* How many (new) users did Netflix acquire each month—i.e., what were the initial sizes of the 12 different cohorts? [Enter your results in cells B2 to B13 in sheet "q2-bd" of the template.]
- 2-c Please compare the number of customers acquired in each month and discuss.
- 2-d* How many users were there in each month of the observation window (i.e., size of customer base over time)? [Enter your results in cells C3 to C13 in sheet "q2-bd" of the template.]
- 2-e What can explain the differences in the number of users over time? Please comment.

Question 3 Calculating Retention Rates from Historical Data

You will now look at customer churn and retention using historical data.

- 3-a* Calculate the retention rate of customers acquired in month zero (0) (i.e., cohort 0) at each month of the time window. [Enter your results in cells B2 to B13 in sheet "q3-acegh" of the template.]
- 3-b Plot the retention rate (y-axis) against time (x-axis) and describe your results. Please provide an interpretation of your findings.
- 3-c* Compute the average of the monthly retention rates of cohort 0 from part 3-a*. [Enter your result in cell B14 in sheet "q3-acegh" of the template.]
- 3-d Compare the result with industry benchmarks, e.g., the (monthly) retention rates that Netflix or other major SVOD platforms report. Explain any difference between your result and these benchmarks.
- 3-e* Repeat the calculation of the monthly and average retention rates from parts 3-a* and 3-c* for the remaining cohorts, i.e., cohorts 1,..., 11. [Enter your results in cells C2 to M14 in sheet "q3-acegh" of the template.]
- 3-f Do your observations from parts 3-b and 3-d hold across all cohorts? Please comment.
- 3-g* Calculate the average retention rate for each month across cohorts, i.e., the average retention rate of users in their 1st, 2nd, ... 12th month after acquisition. [Enter your results in cells N2 to N13 in sheet "q3-acegh" of the template.]
- 3-h* Average the monthly retention rates from part 3-g* across months. [Enter your result in cell N14 in sheet "q3-acegh" of the template.]

- 3-i Discuss which approach for calculating average retention rates (i.e., 3-a*, 3-c*, and 3-e* vs. 3-g* and 3-h*) is appropriate.

Question 4 Gauging Expected Lifetime and Survival Probabilities

We would like to get an idea of the duration of the relationship between Netflix and its users.

- 4-a* Compute the expected lifetime of users using the (constant) average retention probability from part 3-h*. [Enter your result in cell A2 in sheet “q4-ab” of the template.]
- 4-b Please provide an interpretation of the expected lifetime and discuss implications for the customer acquisition strategy of Netflix.
- 4-c* Calculate the expected lifetime of customers by using the (changing) monthly retention rate data that you obtained in part 3-g*. [Enter your result in cell B2 in sheet “q4-ab” of the template.]
- 4-d Compare the results from part 4-c* with those from part 4-a* and discuss the (managerial) consequences of using the estimate from part 4-a* vs. part 4-c*.

Question 5 Estimating Survival Probabilities

In this part of the assignment, you will look at the probability that users subscribe to Netflix until a given point in time during the observation window.

- 5-a* Compute the survival probability (i.e., active probability) of the users who were acquired in the first month (i.e., month 0) until the end of the observation window (i.e., month 11). Use the (constant) average retention probability that you have obtained in part 3-c*. [Enter your results in cells B2 to B13 of sheet “q5-ab” of the template.]
- 5-b* Re-compute the survival probability of users in cohort 0 with the (changing) monthly retention probabilities from part 3-a*. [Enter your results in cells C2 to C13 of sheet “q5-ab” of the template.]
- 5-c Plot the survival probabilities obtained in part 5-a* and 5-b* against time since acquisition. The y-axis shows the survival probability and the x-axis shows time (in months). Additionally, draw two vertical lines that crosses the x-axis at the expected lifetime of customers in cohort 0, which you computed in question 4. Does the survival probability based on constant retention probabilities (i.e., part 5-a*) provide a good approximation of the survival probability that you calculated using the (changing) monthly retention probabilities (i.e., part 5-b*)? Please comment.
- 5-d Based on your results in part 5-c, do you expect to overestimate or underestimate CLV using the survival probabilities based on constant vs. changing retention? Please explain.

Question 6 Estimating CLV

We will estimate the CLV of new customers that Netflix is considering adding at the beginning of month 12 (i.e., one month after the end of the observation window or one month into the future).

Before we can estimate CLV, we need to obtain estimates of (future) revenues and costs. Assume that the annual discount rate is 15 percent.

- 6-a* Take a reasonable guess of the revenues per customer and month and explain. [Enter your results in cells B2 to B13 in sheet “q6-a” of the template.]
- 6-b Netflix uses activity-based costing and assumes that the marginal costs of marketing, selling, production, and servicing one additional customer are zero (0). Please discuss briefly (in 2–3 sentences) whether the assumption of zero (0) incremental costs is justified or not.
- 6-c* Compute the CLV of acquiring a new customer in month 12 by using an infinite time horizon and the (constant) average retention rate of cohort 0 from part 3-a*. [Enter your result in cell A2 in sheet “q6-cd” of the template.]
- 6-d* Use the expected lifetime method and redo the CLV calculation. Decide whether you want to use the expected lifetime based on constant or changing retention (parts 3-a* vs. 3-c*, respectively). Compare the result with that from part 6-c*. [Enter your result in cell B2 in sheet “q6_cd” of the template.]
- 6-e Explain your findings, i.e., the difference in CLV you obtain from part 6-c* vs. 6-d*.

Question 7 Using CLV for Targeting Customer Acquisition

We are interested in using CLV for targeting customer acquisition. Specifically, we would like to gauge the value of acquiring 10,000 new Netflix users at the beginning of month 12 (i.e., one month after the end of the observation window or one month into the future). The annual discount rate is 15 percent.

- 7-a* Compute the CLV of a new customer with a time horizon of one year. You can use direct observation (i.e., historical retention data) and keep your estimates of revenue and costs, from before. Calculate the sum of the CLVs of the 10,000 new customers (i.e., customer equity). [Enter your result in cell A2 in sheet “q7-a” of the template.]
- 7-b Please explain the meaning behind the estimate from part 7-a*.
- 7-c Netflix wants to increase the time horizon from one to two years. Discuss briefly, using 2–3 sentences why historical data cannot be used in this case and why using a (constant) average retention rate might not be realistic.
- 7-d* Forecast future retention rates with a model that allows for the retention rates to increase over time, albeit at a decreasing rate. You may want to use that (monthly) retention rates in SVOD are usually 0.9, on average. The rate by which the increase in retention probabilities decays over time is 0.3. [Enter your results in cells B2 to B25 in sheet “q7-d” of the template.]
- 7-e Plot the projected retention rate (y-axis) against time in months since acquisition for a period of two years (x-axis). Assess the quality of your forecast by adding the realized retention rate of cohort 0 to the plot and discuss.
- 7-f* Compute the CLV of a newly acquired customer at the beginning of month 12 using a time horizon of two years. [Enter your result in cell A2 of sheet “q7-f” of the template.]

Question 8 Long-term Value Creation Using Conjoint Analysis

Conjoint analysis can be used to create products whose attributes offer value to customers. It is frequently used prior to the launch of a product but can also be a tool for developing the product after its launch. In this part of the assignment, you will use the Netflix streaming data in the sheet “filter_user” to estimate a user’s preferences for different genres and recency using conjoint analysis. The streaming platform could use the results of such an analysis when deciding which titles to produce or license in the future.

The data in the sheet “filter_user” contain a row for each time that the specific user clicked on a Netflix title. The variables *genre_...* are binary 0—1 indicator variables (dummies) showing the genres of the title that the user was watching. There are 5 different genres (e.g., “family and comedy”) and each *genre_* variable takes on the value one (1) if the title was of that genre and zero (0) otherwise. For this assignment, a title can only be of one genre (i.e., genres are mutually exclusive). The variable *recency_new* is a binary 0—1 indicator variable (dummy) that takes on the value one (1) if the user watched a title that was released less than two years ago and is zero (0) otherwise. *duration_long* is a 0—1 indicator variable (dummy) that is one (1) if the duration of the content is over one hour and zero (0) if it is less than or equal to one hour.

- 8-a* Identify the attributes and attribute levels of this conjoint analysis. Compute the total number of product versions that would be needed for a fully factorial design of the conjoint analysis. [Enter your results in cells A2 to E2 in sheet “q8-a” of the template.]
- 8-b Calculate the number of genres under a fully factorial design if genres were not mutually exclusive. Compare your result to that obtained in part 8-a* and discuss.
- 8-c Count how many times the user watched a particular combination of attribute levels. Rank each combination by their streaming frequency.
- 8-d* Using the ranked preference data from part 8-c, estimate the part-worth utilities for all attribute levels. What conclusions do you draw from the results? [Enter your results in cell B2 to B8 in sheet “q8-d” of the template.]
- 8-e* Compute the relative importance of the different attributes. Interpret your findings. [Enter your results in cell B2 to B4 in sheet “q8-e” of the template.]

Appendix

Table 1
Weighting of Questions in Final Project Report

2a*	1
2b*	2
2d*	2
3a*	2
3c*	1
3e*	2
3g*	1
3h*	1
4a*	1
4c*	2
5a*	1
5b*	2
6a*	1
6c*	3
6d*	2
7a*	3
7d*	3
7f*	2
8a*	1
8d*	3
8e*	2

Note. The maximum number of points for the project report is 38 points (i.e., 67 percent of the 57 total points for the CPA)

Table 2
Criteria and Weighting for Evaluation of Video Presentation

Problem definition	3
Method and data	2
Results	2
Business implications	5
Recommendations for management	4
Quality of presentation	3

Note. The maximum number of points for the video presentation is 19 points (i.e., 33 percent of the 57 total points for the CPA).