

### Retention data and customer intelligence

- Individual assignment.
  - Work due (by 11:55PM on the following dates):
    - November 12<sup>th</sup>: First round of decisions (10 pts).
    - November 26<sup>th</sup>: Second round of decisions (10 pts).
    - December 4<sup>th</sup>: a 3-page executive report on customer intelligence (10 pts).
  - You must provide your R code through ZoneCours.
  - The executive report must be given as PDFs made with Rsweave OR Rmarkdown.
  - You must provide the code for the executive report as well.
  - This work is worth 30% of your final grade.
- 

### Context

*Telc*, a telecom company specialized in mobile communications, is turning 10 years (120 months). After their initial big splash, they have grown steadily, welcoming about 3% more clients every year. Shareholders are satisfied with this performance considering that telecom is a very competitive market where clients are constantly looking for a better deal. To celebrate its 10-year anniversary, *Telc* is planning an innovative marketing campaign to reward their clients and improve their loyalty.

Over the last six months, the Marketing and Business Intelligence department has been planning a campaign where some clients are invited to a 4-course dinner with live music performed by a local talent. The database of *Telc* identifies which clients are in a same family. It has been decided that if a client gets invited to the dinner, all family members from the same household who are *Telc* clients get invited to the same dinner. The main objective is to make the clients more loyal by developing a sense of belonging to the company, with a desire to improve retention of the clients, by inviting to the dinner those who are more likely to leave.

You will be given some data about the current *Telc* customers (a description of the variables is found later in this document). You must now decide what families to invite to a dinner (and how many dinners to put together). You will submit your decision through the online platform at : <http://dsgames.hec.ca/play> that will give you immediate feedback on your performance in terms of profit in the next 24 months in comparison to having invited nobody.

In their planning, the Marketing and Business Intelligence department has put a pilot study where they chose around 5% of their clients at random and invited them to a dinner. These clients are identified with `promo="True"`. We are now twelve months after the pilot, so it is possible to know who has left in that time window.

The pilot study was useful in determining the cost of organizing those dinner events. Even when clients are not able to make it to the show, getting an invitation may have a positive effect. In any case, we cannot force clients to attend the dinner, we can only decide to invite them or not, hence it makes sense to measure the effect of the action we are able to take. The costs have been broken down by invitation, whether the invited client goes to the dinner event or not, and amount to:

- \$10 per invitation for the food and drink,
- \$10,000 per 5,000 invitations for the fixed costs (renting room and sound system, paying the artist, etc.)

This is a business simulation in which clients and their behaviour are simulated. Therefore, we know how each individual will behave in the next 2 years if you invite them, and if you do not. *Telc* constantly prices their plans to keep a margin of profit 50% on the monthly rates for the plan, hence the profit generated by one client amounts to 50% of the fees that they pay for their plans in the next 24 months, or until they leave. Your performance will be based on the total profit in 24 months, minus the total cost of the dinners based on the number of invitations. With a good retention strategy, you should do better than the baseline of not inviting any further clients to those dinners (which leads to additional profits of 0\$).

To clarify, the time frame:

- *Telc* was founded 10 years ago, the first clients arrived at time  $T = 1$  month, we are now at time 120.
- A pilot study was run at time  $T = 108$  month, with the covariates of all clients at that time. The variable "churn\_in\_12" indicates if the client left between time 108 and 120. This file is called `train_student.csv` and contains information on all clients at time 108.
- The file `score_student_withID.csv` contains the covariates of all clients now, at time 120, and you need to decide who to invite. Clients who left between times 108 and 120 are not in the file, but new client are in there.
- The online platform knows how much profit is made by all clients until time 144 (two years later) if you invite them, and if you do not invite them. Your performance is based on the total amount of profit you obtained minus the cost of the dinners.

### Round 1: Targeting churning customers without the pilot

IMPORTANT: For round 1, you need to make a decision without using the information obtained by the pilot. The costs presented remain the same, but you must only use the clients with `promo="False"` in your analysis – do not forget that you need to provide your R code, do not cheat.

We do not give further instructions on what you should predict and how, you are on your own for that. There is no single good answer, and multiple strategies that can support the business problem reasonably well. We expect each student to come up with their own approach.

The platform `dsgames.hec.ca` allows for multiple uploads (a total of 70 per round) and will provide immediate feedback. You need to prepare a one column csv file with the list of `unique_family`, the code that uniquely identifies the families that you want to invite. Before the deadline, you must select one of your uploaded decisions as your final decision. You also need to share a clean version of the R code that produces your final decision from the raw data – it will not be systematically reviewed but could be taken into account for determining your grade.

### Round 2: Using the information obtained from the pilot

In round 2, you may use all the data. Information from the pilot study may help you target clients even more efficiently, possibly leading to a better solution than in round 1. The same upload site will be used, and you will submit your R code again. Note that both rounds are open now, but they have different end times.

### Step 3: Customer intelligence executive report

We ask you to perform a cluster analysis of the data to address a business problem of your choice with the retention data. You may use any other methods, supervised or unsupervised, as well as descriptive statistics, in addition to the clustering. Identify key discoveries about the clients and present them as in a 3-page executive report that has explicit suggestions of actionable items. The report must be produced with RMarkdown or Rswave.

**Additional information**

*Where is the data?*

On ZoneCours, you will find a link to a zip file containing both data files.

*How to submit a decision?*

You must decide which families to invite. Prepare a one column file with the `unique_family` variable of the families to invite. The cost of the invitations will be automatically determined. You will have to upload that file to a Web platform. Codes to unlock the right games on the online platform are on ZoneCours.

*Where do I submit my R code and my report?*

In the assignments tab of ZoneCours where assignments have been created. Feedback will be returned through ZoneCours as well.

*What grade will I get for round 1 and 2?*

The points of round 1 and 2 are uniquely based on performance. When you upload a decision, you will immediately see your performance as well as that of the other students in the course. You are ranked based on the amount of profit gained in the 24 months following the dinner initiative. The *Baseline* of the leaderboard shows how much profit is made if nobody gets invited to the dinner show.

During the course of every round, you see the “interim” leaderboard. To ensure that there is no benefit in overfitting, there is another leaderboard called “real-life”, which will only be revealed at the end of the round. The grades are based on the real-life leaderboard in the following way:

- 0/10 if you make less profit than inviting nobody (if you do not beat the baseline),
- At least 5/10 if you beat the baseline on the interim leaderboard,
- 10/10 if you are among the 3 people with highest scores – this is approximately the best 10%.
- If you got  $x$  profit, let  $B$  be the baseline and  $T$  be the result of the 3<sup>rd</sup> best student, then your grade out of 10 will be:  $5\{1 + (x - B)/(T - B)\}$ . In other words, we use a linear interpolation between the baseline and the top people who got 10/10.

| No. | Variable name                 | Description  |
|-----|-------------------------------|--|
| 1.  | id                            | Client ID  |
| 2.  | family_id                     | Family ID  |
| 3.  | unique_ID                     | Unique ID containing information about the individual ID and his family ID   |
| 4.  | unique_family                 | Unique ID for the family (to be used to identify families to invite)         |
| 5.  | unique_individual             | Unique ID for each individual  |
| 6.  | age                           | Age  |
| 7.  | gender                        | Gender   |
| 8.  | period_id                     | Month of client arrival (1=first month when the company was founded)         |
| 9.  | active                        | The client's contract is active ('TRUE'=yes, 'FALSE'=no)                     |
| 10. | plan_type                     | The type of plan (rent, buy or bring your phone)                             |
| 11. | phone_price                   | Value of the phone   |
| 12. | cashDown                      | Initial cash down on the phone   |
| 13. | phone_balance                 | Balance on the phone   |
| 14. | workphone                     | Work phone ('TRUE'=yes, 'FALSE'=no)  |
| 15. | base_monthly_rate_phone       | Monthly payment rate for the phone   |
| 16. | base_monthly_rate_plan        | Monthly payment rate of the plan   |
| 17. | voice_minutes                 | Monthly number of minutes of voice in the plan                               |
| 18. | unlimited_voice               | Plan has unlimited voice (1=yes)   |
| 19. | unlimited_text                | Plan has unlimited SMS (1=yes)   |
| 20. | data                          | Monthly GB of data in the plan   |
| 21. | voice_consumption             | Total minutes used in the most recent month of the plan.                     |
| 22. | total_voice_consumption       | Total minutes since the beginning of the account                             |
| 23. | text_consumption              | Total SMS's in the most recent month of the plan.                            |
| 24. | total_text_consumption        | Total number of SMS's used since the beginning of the account                |
| 25. | data_consumption              | Data consumption in the most recent month of the plan                        |
| 26. | total_data_consumption        | Total GB used since the beginning of the account                             |
| 27. | technical_problem             | Number of reported technical issues this month                               |
| 28. | time_since_technical_problems | Number of months since the last technical issue                              |
| 29. | total_technical_problems      | Number of all reported technical issues                                      |
| 30. | complaints                    | Number of complaints in the last month                                       |
| 31. | time_since_complaints         | Time since last complaint in months  |
| 32. | total_complaints              | Number of complaints since the client's account was created.                 |
| 33. | time_since_data_ouverage      | Number of months since data limit was last exceeded                          |
| 34. | total_data_ouverage_fees      | Number of times the data limit was exceeded                                  |
| 35. | time_since_voice_ouverage     | Number of months since the voice limit was last exceeded                     |
| 36. | total_voice_ouverage_fees     | Number of times the voice limit was exceeded                                 |
| 37. | time_since_ouverage           | Time since last overage  |
| 38. | total_ouverage_fees           | Total overage fees   |
| 39. | promo                         | Invited to the trial dinner ('TRUE'=yes, 'FALSE'=no) (Used only for Round 2) |
| 40. | churn_in_12                   | Client has left in a 12-month window ('TRUE'=left, 'FALSE'=still client)     |