**Group Assignment 1: Maru Batting Center**

**Objective:** The objective of this group assignment is to understand the value of different customer segments through customer lifetime value (CLV) analysis. You will calculate CLV for various segments and make business recommendations. You will also conduct sensitivity analysis to understand how variations in modeling assumptions affect conclusions.

**Zero tolerance policy:** Violation of honor codes will result in a zero on the assignment.

**Instructions:** Read the case “Maru Batting Center: Customer Lifetime Value”. Download “maru\_data.csv” from Connect and answer the following questions. Prepare a group report for each of the questions. Insert tables or figures in the report rather than attach it at the end. There is no page limit on the reports. However, evaluation is strictly based on the depth of the discussion rather than the length. Single space your document and use Times New Roman font of size 12. All analyses need to be done in R.

You will see your group formation information on Connect. Each group should submit (i) your group report and (ii) all the R codes at connect.ubc.ca by March 5 2018 1pm PST.

**Part I: Answer the questions on pages 5-6 of the case**

Notes: when answering questions, ignore the specification that: “assuming that 100 percent of a customer segment will have experienced attrition once the net present value of annual profits per customer falls below ¥100.” Calculate CLV over an infinite time horizon.

1. Does this assumption of infinite time horizon change results much? Why or why not? Reflect on the conditions under which it makes sense to use an infinite time horizon. Will this assumption ever lead to serious problems?
2. In class we introduced three different formulas that you can use to calculate CLV. You will need to choose one. Choose the one that you think makes the most sense for the business problem faced by management of Maru Batting Center, and be prepared to defend your choice in class. In the report please note your choice and reasoning process behind it.

**Part II: Sensitivity Analysis**

The case asks you to do CLV calculations based on numbers that are merely educated guesses, such as acquisition costs (based on estimated response rates), retention rates and annual margins. You will perform a sensitivity analysis to evaluate how estimated CLV changes when you vary these assumptions.

For this part of the assignment we will concentrate on the “Elite-Ballplayers (Party)” segment. The baseline assumptions for this segment are acquisition costs (ac) of 50,000, annual margins (am) of 30,000 and a retention rate (rr) of 0.6. (Note: for all analyses use an interest rate of 0.1 for discounting).

For each of the three variables (ac, am, ar) generate a vector of alternative values that differ from the assumed values in a reasonable way. The number of “scenarios” is up to you. Start with a small number, say four, to make the coding and interpretation easy, but write your code so it is easy to generate and analyze more scenarios later. So for instance you could test scenarios where the actual retention rate is 20% lower than expected, 10% lower than expect, 10% higher than expected, or 20% higher than expected.

Next, calculate CLV for every possible combination of scenario values. You should also include the original assumed values as one of the scenarios you calculate. If you have 4 possible new values for each variable, and the original assumed values, this should yield 5 by 5 by 5 =125 scenarios.

Now do some analyses and visualization of your results. First calculate whether and how often the CLV is negative. Is it negative for plausible values of the variables? How often does the analysis yield a CLV that is lower than the one you calculated for Little Leaguers in part 1. Reflect on whether any of your conclusions from part 1 are different in light of the analysis. Do you have the same level of confidence in your conclusions? Why or why not?

Next plot how variations in each of the three variables changes CLV, holding the other two variables constant (e.g. boxplot in Tableau or R). One way to do this would be to create three line graphs, one for each variable. Each graph would show how varying the variable affects CLV, using the original assumed values for the other two variables. What do you notice about the shapes of these three graphs? Do all three variables behave in similar ways? Do some variables seem to be more important than others? What broad recommendations do these results imply for Maru Batting Center. What aspects of CLV would be most beneficial to focus on?

**Part III: Monte Carlo Simulation**

Now that you have calculated CLV for various segments and performed sensitivity analyses to assess how variation in assumptions affects conclusions about CLV at the aggregate level, you will be asked to take an individual-level simulation approach to try to estimate CLV and compare the analysis to the conclusions you reached with respect to the aggregate level analysis.

Again, we will use the “maru\_data.csv” and concentrate on the elite ballplayers (party) segment. Your objective is to run a simulation that randomly samples values for the annual margin and retention rate from a probability distribution of your choosing and computes CLV for a large number of “virtual customers.” You will analyze and visualize the CLV distribution of virtual customers and compare it to your conclusions based on the aggregate analysis.

You will simulate annual margins and retention rates for your virtual customers. For all analyses use .1 for the interest rate and 50,000 for the acquisition cost. Note that in reality, both of these numbers have uncertainty, but for simplicity use the fixed values.

**Annual Margin:** To help you determine the right distribution to sample from, look at the file “customers.rdata” This shows annual margins for a subset of Maru’s past customers in the Elite Ballplayer segment. Load the data into R, visualize the data in a histogram or other format and calculate summary statistics. Based on this past data, decide on an appropriate distribution to sample annual margins for your virtual customer.

**Retention Rate:** The only thing you know about retention rate is that the average is 60%. You have no additional data about retention so you will have to use your intuitions and common sense to decide what kind of distribution makes sense. In order to decide on parameters for the distributions, play around and try out a bunch of different parameterizations. Pay attention to how variations in the distribution type and parameters you choose affect your conclusions.

**Simulation:** Write R code to generate a sample of virtual customers. You can choose how large of a sample to generate and you may try different sample sizes to see how this choice affects your results. Two notes: 1. Vector-based calculations are much more computationally efficient in R than For Loops. When possible, use vector computations. 2. If you want to be able to recreate your results exactly generate a random seed at the beginning of your code using the set.seed() function.

Analyses and Visualization: Create graphs and compute summary statistics based on your simulations. Your analyses should focus on two issues:

1. Value concentration. Value concentration refers to how much of your aggregate CLV is concentrated in a subset of customers. If value concentration is high, it means a small number of customers is responsible for a large percentage of your margins. Assess how value concentration depends on the distribution types and parameters you choose for the margin and retention rate distributions. Does value concentration depend on assumptions about how these variables are distributed? If so, how?
2. Comparison to aggregate level analysis. You calculated an average CLV for customers in the elite ballplayer segment. Does the average CLV based on the simulation agree with the number you calculated in part 1? Do assumptions about the distribution of margins appear to affect the extent of agreement or disagreement? How about assumptions about retention rates? Speculate on why the individual analysis does or does not agree with the aggregate level analysis from part 1.