MSIN0094 Marketing Analytics

Dr Wei Miao

October 3, 2023

Preface: How to Use This Guide

Arrangements each week

We will have a 3-hour session each week and I will aim to cover a new marketing analytics model. Whenever we learn a new technique (e.g., a new statistical model or a new analytics tool), the subsequent week will often start with some warm-up exercise and a seminar workshop (with a case study) for you to review and practice the new technique learned in the previous week. This way, you would have time to digest what you've learned and can further reflect on your understanding of the technique by practicing your skills with a real-life application.

For instance, in week 1, I will first introduce the concepts of marketing and marketing process, and then will cover the concept of customer lifetime value (CLV) and how to compute CLV with R. In week 2, we will therefore begin with a case study that helps you practice your knowledge of CLV, so you can understand how to use CLV for better marketing decisions in your future projects/jobs.

In the remaining time of week 2, I will then introduce a new technique: the `dplyr` package in R, which helps us clean and manipulate datasets in R. Following a similar logic, then in week 3, we will start with a case study for you to practice the `dplyr` package. So on and so forth.

About the labels

- Essential: contents core to this week's materials. All pre-class preparations should be completed before class.
- Optional: supplemental readings for those interested in learning more

All materials, including the lecture slides and before-class readings will be released a few days before each Thursday's class.

Module Outline

<div align="center">Table 1: Module Outline</div>

Week	Analytics/Methodology Topic	Substantive Topic	Case Study[1]	Important dates
Induction	Introduction to R			
1	Computation with R	Customer lifetime value	Customer Lifetime Value	
2	Data wrangling with R	Preliminary customer analysis	Preliminary Customer Analysis	
3	Unsupervised learning	Segmentation	Improving Marketing Efficiency for Tesco with Predictive Analytics I (Unsupervised Learning)	
4	Supervised learning	Targeting	Improving Marketing Efficiency with Predictive Analytics II (Supervised Learning)	Friday, 30 October 2023 1st assignment due
5	Rubin Causal Model and Potential Outcome Framework	Causal Inference		
6	RCT	Promotion Analytics	Improve User Engagement on Social Media Platforms using A/B Testings	
7	Linear regression	Pricing analytics	Optimizing Marketing Mixes Using Regression Models	
8	Instrumental variable	Platform Design	Evaluating the Impact of COVID-19 on Ride-sharing Market	Friday, 24 November 2023 2nd assignment due

Week	Analytics/Methodology Topic	Substantive Topic	Case Study	Important dates
9	Quasi-experiments	Difference-in-Differences Regression Discontinuity Design		
10	Causal forest	Causal machine learning		
				Friday, 15 December 2023 3rd assignment due

Induction Week: R Basics

> **!** Pre-class preparation
>
> - Finish reading "An introduction to R" (can be assessed in this link) Chapters 1, 2, and 3.
>
>   – Please try to practice the codes in R along your reading.
>   – Take a note of any questions you may have during your self-study. I will cover R basics in greater details in the induction week.

- What you will learn

  – An introduction to R basics

- After-class exercise

  – (essential) Finish data camp "Introduction to R" tutorial before Week 1 class. We will learn how to use R to compute customer lifetime value next week, so it's very important that you are familiar with R basics before class.

---

[1]Case studies to be discussed in the subsequent week. For example, CLV is introduced in Week 1 and its associated case study "CSLV" will be discussed in Week 2.

Week 1: Module Introduction and Customer Lifetime Value

> ❗ Pre-class Preparation
>
> - Remember to bring your laptop to class **each week**, as we will be practicing R programming in class every week; make sure you have installed R and RStudio following the installation guide
> - Join the Microsoft Teams for the module; all communications will be on this Teams channel
> - Review R basics lecture notes in the induction week, we will compute customer lifetime value using R in class. If your schedule is tight, only focus on the Wednesday tutorial until Vector.
> - Complete the R exercise on the `Case-BreakEvenAnalysis-Stu.qmd`. You can find the file on Moodle.
> - Prepare for Case study: "Break-Even Analyses for PineApple Inc"
>
>   - Please carefully read Section 1, Situation Analysis, before class
>   - Go through the remaining case, highlight the numbers for calculations, and we will be solving the case in class.

Module Introduction

- What you will learn

  - An overview of the course topics and requirements
  - Concept of marketing and marketing process
  - How marketing analytics can empower marketers in the digital era

- After-class reading

  - (optional) The Definitive Guide to Strategic Marketing Planning. Highly recommended if you didn't take marketing undergrad courses and would like to know more about the conventional marketing process.

Customer Profitability and Lifetime Value

- What you will learn

  - How to conduct break-even analyses (static and dynamic)
  - The concept of customer life cycle
  - How to compute customer acquisition cost (CAC)
  - How to compute customer lifetime value (CLV)

- After-class reading

  - (optional) "Hubspot: How to compute CLV". This article introduces alternative ways to compute CLV, which are used in many companies.
  - (optional) "Important lessons for embracing customer lifetime value"
  - (essential) "Common R Programming Errors Faced by Beginners". This guide will be updated constantly and summarizes the common R errors you may encounter as beginners. For any questions, come here first for solutions.

Week 2: Data Wrangling with R

Case Study: Customer Lifetime Social Value

! Pre-class preparation

- HBS 9-518-077: Customer Lifetime Social Value (CLSV).

  - In the first class this week, we will go through how to compute the customer lifetime value for i-basket, a US grocery supermarket, in this case study. We will only focus on the CLV calculation part while leaving CLSV as after-class optional learning contents. **Please read pages 1-8. Pages 1-5 introduce the case background; pages 5-8 introduce key information to compute CLV for the company.**

  - When reading the case, please use the mind map below to find out all key information, including the variable, the value, and where it shows up in the case study. Highlight them on the PDF. This can help you quickly find that information in class.

  - Think about the following questions, which we will discuss in class.

    1. What would be the time unit of analyses? monthly or yearly? How many years of customer's lifetime to consider for CLV calculation? Is this reasonable?

    2. What is the information needed to calculate the net cash flows of a customer in each period? Where can you find the M and c in the CLV formula? Highlight these key numbers so that we can create them in R directly.

    3. How do we incorporate customer churn in CLV calculation?

    4. What are the costs i-basket needs to incur to acquire a customer? Based on this information, how to compute the customer acquisition costs (CAC)?

    5. Can you figure out how to compute CLV with R even before Wednesday class? You can try your best to use what we have learned in Week 1 to

> compute the CLV for i-basket on your own. The exercise Quarto file is already given below.
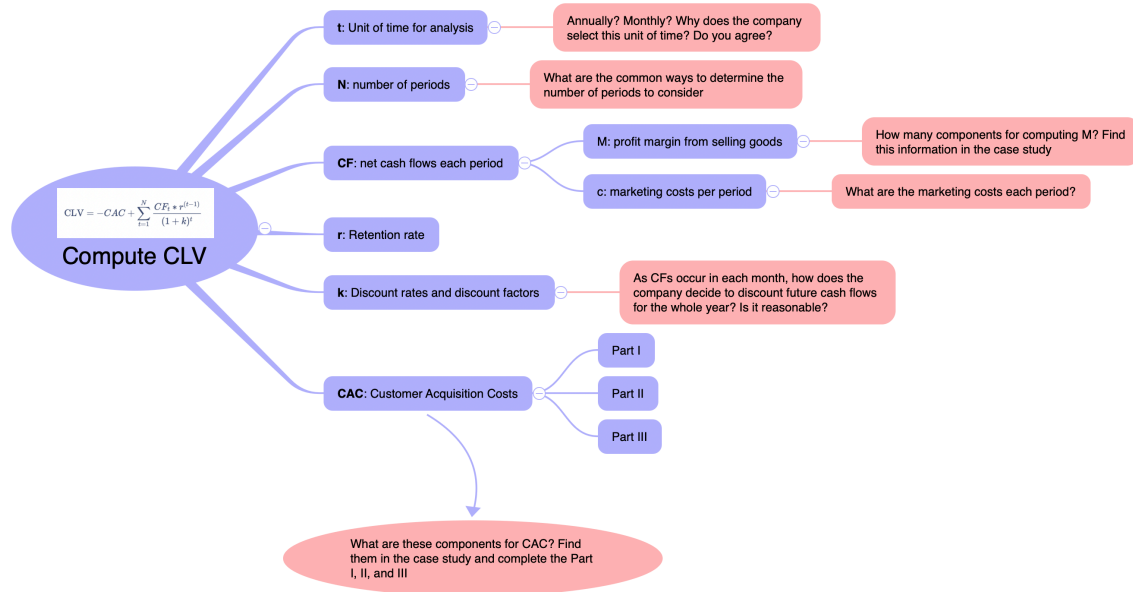

Figure 1: Mindmap for Computing the CLV for i-basket

- What you will learn

  - How to apply CLV calculation in a real-life case study
  - Discuss how CLV can be used by marketers to guide marketing decisions

- After-class reading

  - (optional) The Dangerous Seduction of the Lifetime Value (LTV) Formula

Data Wrangling with R: Part I

- What you will learn

  - Process of a typical data analytics task
  - How to use `filter`, `mutate`, and `arrange` for data manipulation with `dplyr` package in R

- After-class reading

  - (essential) Finish the practice exercise on data camp: Data Manipulation with dplyr: 1.Transforming Data with dplyr

- (essential) [Cheatsheet for `dplyr`](#)
- (optional) [Python Pandas vs. R Dplyr](#). In David's class, you will learn how to use Python Pandas for data wrangling in the last two weeks. This cheat sheet provides a one-to-one comparison of the two libraries for your reference.

Week 3: Predictive Analytics for Customer Segmentation

Data Wrangling with R: Part II

> **❗ Pre-class preparation**
>
> - Before Wednesday class, please finish the following
>   - read the 2 case studies; please familiarize yourself with the variable definitions. **This is very important** as we will use the two datasets to learn data wrangling in R using the dplyr package.
>   - open the two datasets (demographics.csv and purchase.csv) in Excel and inspect the data sets, especially the data structure (what each row means and what each column means)
> - There is no pre-class coding exercise required for this week. We will be doing data wrangling together in class to solve case questions.

- What you will learn
  - The usage of `group_by` and `join` for data aggregation and merge with `dplyr` package
  - How to use `dplyr` to conduct preliminary customer analyses
- After-class reading
  - (essential) Finish the practice exercise on data camp: [Data Manipulation with dplyr: 2.Aggregating Data](#)
  - (optional) Go through the [tutorial](#) of `modelsummary` package to learn more features

Predictive Analytics: Unsupervised Learning

- What you will learn
  - Important concepts in predictive analytics
  - Concept of unsupervised learning
  - How to run K-means clustering in R

- After-class reading

    - (recommended) K-means Cluster Analysis, which provides more details on the maths behind the K-means clustering

Week 4: Predictive Analytics for Customer Targeting

Supervised Learning with Tree-based Models

- What you will learn

    - Important concepts of supervised learning
    - Intuition behind decision tree and random forest models
    - How to build random forest models in R

- After-class reading

    - (optional) Varian, Hal R. "Big data: New tricks for econometrics." Journal of Economic Perspectives 28, no. 2 (2014): 3-28
    - (recommended) Decision tree in R and Random forest in R. Both tutorials introduce the detailed maths behind the two models if you would like to learn more
    - (optional) Available machine learning model packages in R. In class, we have learned how to use R packages to run random forest models. This link lists all other machine learning packages that can be used in R. You can learn how to use these models following their manuals.

Workshop: Targeted Marketing with Predictive Analytics

> ❗ Pre-class preparation
>
> - Case Study: Improving Marketing Efficiency Using Targeted Marketing
>
>   - Please carefully read the case background before class

- What you will learn

    - The application predictive analytics (supervised learning) in targeted marketing
    - How targeted marketing improves ROI of marketing campaigns

Week 5: Causal Inference and RCT

Causal Inference

- What you will learn
    - Concept of causal inference
    - Concept of Rubin's potential outcome framework and various treatment effects
    - Concept of fundamental problem of causal inference and basic identity of causal inference

- After-class reading
    - (optional) Varian, Hal R. 'Causal Inference in Economics and Marketing'. Proceedings of the National Academy of Sciences 113, no. 27 (5 July 2016): 7310–15.
    - (optional) Angrist, Joshua & Pischke, Jörn-Steffen. (2009). Mostly Harmless Econometrics: An Empiricist's Companion.

Randomized Experiments

> **!** Pre-class preparation
>
> - This and next week, we will be conducting t-tests with R in class to estimate treatment effects from randomized experiments. If you're unfamiliar with t-tests, please go through this Review of Statistics with R before the next week's class.

- What you will learn
    - Why RCT is the gold standard of causal inference
    - Steps to design and conduct an RCT

- After-class reading
    - (optional) Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press, 2020., by Ronny Kohavi (Airbnb, formerly Microsoft and Amazon), Diane Tang (Google), and Ya Xu (LinkedIn). This is a very practically oriented guide to experimentation, with many examples relevant to marketing and product management.

Week 6: Linear Regression

Workshop: Improve User Engagement on Social Media Platforms Using A/B Testings

> **!** Pre-class preparation
>
> - Case study: Improve User Engagement on Social Media Platforms Using A/B Testings
>   - Please carefully read the case background before class; we will be discussing the case in class

- What you will learn
    - The business model of the advertising industry and mobile marketing industry
    - Design and analyze A-B testing to solve real-life marketing problems

- After-class reading
    - (recommended) Test and learn: How a culture of experimentation can help grow your business
    - (optional) A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook

Linear Regression Model: Basics

- What you will learn
    - Review of concept for Data Generating Process (DGP) and a model
    - The intuition behind coefficient estimation of linear regression models
    - Run linear regression models in R
    - How to interpret the regression coefficients and statistics

- After-class reading
    - (highly recommended) Introduction to Econometrics with R, Chapters 4-7. These 4 chapters cover very detailed applied knowledge of linear regressions. Due to limited time, we cannot cover all contents in class, so it would be great if you can take time to go through these chapters thoroughly.

Week 7: Marketing Mix Modeling

Linear Regression Model: Advanced Topics

- What you will learn
  - How to model non-linear relationship using linear regression
  - How to interpret the coefficients of categorical variables

Workshop: Marketing Mix Modelling with Linear Regression

> **!** Pre-class preparation
>
> - Case study: Optimizing Marketing Mixes Using Regression Models
>   - Please carefully read the case background before class; we will be discussing the case in class

- What you will learn
  - Build regression-based marketing mix models to solve pricing and promotion optimization problems
- After-class reading
  - A complete guide to Marketing Mix Modeling

Week 8: Endogeneity and Instrumental Variables

> **!** Pre-class preparation
>
> - Case Study: The Causal Impact of COVID-19 on Ridesharing
>   - Please carefully read the case background before class and we will be discussing the case in class. This framework of IV analyses would be very helpful for your dissertation projects. You can work on similar topics such as the causal impact of COVID-19 on offline shopping, etc.
>   - Finish the coding exercise in the Quarto document before Section 4 IV regression (similar questions will be in the last assignment).

Endogeneity

- What you will learn

    - Endogeneity and its consequence in causal inference
    - Omitted variable bias
    - Reverse causality

Instrumental Variables

- What you will learn

    - Intuition of why instrumental variables solve endogeneity problems
    - The two requirements of a valid instrumental variable and how to find good instruments
    - Apply two-stage least square method to estimate the causal effects using instrumental variables
    - Application of instrumental variables in the marketing and business field

- After-class reading

    - (optional) Econometrics with R: Instrumental Variables Regression

Week 9: Quasi-Experimental Methods

Difference-in-Differences Design

- What you will learn

    - Concept of difference-in-differences (DiD) design
    - Estimation of causal effects using the DiD design
    - Application of DiD design in the marketing and business field

Regression Discontinuity Design

- What you will learn

    - Concept of regression discontinuity design (RDD)
    - Estimation of causal effects using the RDD design
    - Application of RDD designs in the marketing and business field

- After-class reading

– (optional) Varian, Hal R. "Causal inference in economics and marketing." Proceedings of the National Academy of Sciences 113, no. 27 (2016): 7310-7315.
– (optional) Econometrics with R: Quasi-experiments

Week 10: Frontiers of Marketing Analytics

Frontiers of Marketing Analytics

- What you will learn

  – Causal machine learning with causal forest
  – Heterogeneous treatment effect estimation with causal forest in R

- After-class reading

  – Estimate causal effects using ML by Microsoft Research
  – Athey, Susan, and Stefan Wager. 'Estimating Treatment Effects with Causal Forests: An Application'. ArXiv:1902.07409 [Stat], 20 February 2019. http://arxiv.org/abs/1902.07409.