15.819 **Marketing Analytics**

MIT Sloan School of Management Spring 2021

Section A: TTh 11:00 AM – 12:30 PM, online and then in-person sessions after March 1st
Section B: TTh 2:30 PM – 4:00 PM, online

Email full teaching team: **15.819-staff@**

Instructor:
Dean Eckles **eckles@**

Teaching assistants:
Madhav Kumar **madhavk@**
Leon Yao **leonyao@**

Please put "15.819" in the subject when emailing course staff. To email all course staff at once, you can use our list 15.819-staff@.

It is increasingly possible to use analysis of quantitative data to inform, make, and even automate marketing decisions. This is possible in the sense that the data is often available or economical to collect, but many firms lack the analytical and managerial expertise and culture to effectively use this data. The Internet industry has led in both the availability and use of behavioral data, so I will draw on experiences there, but we will apply many course ideas to other industries, as well as in politics and public health.

We will cover quantitative approaches to: learning customer lifetime value, customer acquisition and viral growth, measuring preferences and demand, recommendations and personalization, product management, advertising, retention and churn, social influence and network effects, and discovering customer needs.

The goal of this course is to give students the expertise to initiate, participate in, manage, and evaluate marketing analytics efforts with substantial business impact. This requires knowledge of fundamental ideas at the intersection of statistics, machine learning, and human behavior, but also benefits from knowing specific techniques. Impactful analytics can require a challenging balance of rigor, speed, repeatability, and interpretability.

In exercises and assignments, students will work with data sets. The assigned data sets will include retail purchase micro-data, ridesharing data, data from randomized and non-randomized marketing campaigns, voter turnout experiments, and data describing behavior on social media.

Schedule of topics

This course is under continual development. The schedule may be adjusted as we go (except for assignment due dates).

Readings listed with a session should be read *before* that session. Optional readings are shaded according to the likelihood you should actually read them.

Feb 16 Introduction: Marketing decisions as interventions
Types of data relevant to marketing. Purpose-built data and data exhaust. Ways to use data, from generating ideas to automating decisions. Modeling interventions on your customers and potential customers. Gaps between data and marketing decisions.

Feb 18 Metrics, customer lifetime value & ROI
Customer lifetime value, choosing metrics, KPIs, and proxies.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 6.

Gurley, B. (2012). The dangerous seduction of the lifetime value (LTV) formula. *Above the Crowd*. [optional, grumpy]

Feb 23 Randomized experiments I
Designing and deploying randomized experiments (e.g., A/B tests) — online and offline. Statistical inference with minimal assumptions (randomization inference). Long tails in practice.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 1-2.

Gerber, A., & Green, D. (2012) *Field Experiments: Design, Analysis, and Interpretation*. Ch. 2–3. [optional, helpful for reviewing lectures]

Feb 25 Randomized experiments II, lab
Power to detect effects. Improving precision through design and analysis. Typical designs and outcomes for marketing decisions.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 17. [optional]

Mar 2 Advertising and marketing experiments
Measuring ROI of marketing campaigns, especially advertising.

Lewis, R., Rao, J. M., & Reiley, D. H. (2013). Measuring the effects of advertising: The digital frontier. (2015). *Economic Analysis of the Digital Economy*. Goldfarb, A., Greenstein, S. M., & Tucker, C. E. (Eds.) University of Chicago Press. [optional]

D.E. Using covariates to increase the precision of randomized experiments. [optional, some notes]

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 18. [optional]

Lewis, R. & Rao, J. (2015). The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics*. [optional, technical]

Gordon et al. (2016). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. Whitepaper. [optional, technical]

Mar 4 Product experimentation
Automating / routinizing experimentation. Accumulating and creating new knowledge with many experiments. Examples of successful, confusing, and failed experiments.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 3-4 & 7.

Zhuo, J. (2019). The agony and ecstasy of building with data. [optional]

Mar 9 Monday classes held

Mar 11 Learning during and after a product launch
Instrumentation and tracking use of a product. Summarizing consumer behavior. How is my product broken? Finding unexpected uses.

Seward, Z.M. (2013). The first-ever hashtag, @-reply and retweet, as Twitter users invented them. *Quartz*. [optional]

Teehan, G. (2016). Reactions: Everything in life is not Likable. Facebook Design blog. [optional]

Guo, J. (2015). Seriously, here's one amazing math trick to learn what can't be known. Wonkblog. *Washington Post*. [optional, on synthetic control methods]

Mar 16 Predictive modeling with marketing data: Penalized regression
Modeling probability of purchase with generalized linear models. Understanding overfitting and the bias–variance tradeoff. Improving noisy estimates with shrinkage/regularization/priors (e.g., lasso and ridge regression).

James et al. (2013). *An Introduction to Statistical Learning*. Ch. 2.

James et al. (2013). *An Introduction to Statistical Learning*. Ch. 5-6. [optional]

Hastie et al. (2009). *Elements of Statistical Learning*. Ch. 3 & 7. [optional, more advanced]

Mar 18 Predictive modeling with marketing data, with lab
Applying these models with many, possibly related, predictors. Inspecting and checking predictive models.

Mar 23 No classes held

Mar 25 Targeting marketing interventions
Who should we treat? Using predictive models for decision-making. Reusing previous experiments to learn new policies.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 20. [optional]

Mar 30 **Experiments for customer acquisition and product innovation**
Guest: Zach Winston (Stitch Fix)

Modern "Growth" organizations. Perils of measuring success and conflicting incentives. Randomized experiments. **Zach Winston** is a data scientist at Stitch Fix, where he has worked on both their marketing data science team and, now, a team that leads experimentation. He is an expert in assessing effects of targeted advertising. He was previously a data scientist at eBay after leaving a doctoral program at Wharton.

Katrina Lake (2018). Stitch Fix's CEO on Selling Personal Style to the Mass Market. *HBR*. [optional]

Eric Colson, Brian Coffey, Tarek Rached and Liz Cruz. Algorithms tour: How data science is woven into the fabric of Stitch Fix. [optional]

Erin Boyle (2018). Understanding latent style. [optional]

Apr 1 **Recommendation and personalization**
Review of targeting interventions. Recommendation and ranking as high-dimensional targeting. What is the value of recommendations?

Apr 6 **Recommendation and reproducible analytics in practice**
Guest: Hilary Parker (Biden for President)
The value of systematic, repeatable analysis in businesses. Making decisions with experimental and non-experimental data. Metrics, including for recommendation systems. **Hilary Parker** is a data scientist with experience in ecommerce and political campaigns. She spent the 2020 cycle with Biden for America. She was previously a data scientist at Stitch Fix and Etsy. She hosts a popular podcast on statistics and data science and has a PhD in biostatistics from Johns Hopkins.

Roger D. Peng and Hilary Parker, ed. (2016). *Conversations on Data Science*. Leanpub. [optional]

Apr 8 **Political and consulting perspectives on marketing analytics**
Guest: Solomon Messing (Twitter)
How modern political campaigns use data for strategy and decision-making. Sources of data in politics. Analytics teams in consultancies. **Sol Messing** recently started a new data science team at Twitter focused on social science and machine learning. During the 2020 US Elections, he was chief scientist at Acronym, a nonprofit founded by ex-Facebook people to run digital marketing against Trump's reelection. Previously, he founded Pew's Data Labs group between two stints at Facebook as a scientist.

Apr 13 **Including more in your metrics, scaling up surveys & unstructured data**
Quantitative approaches to measuring attitudes and feelings. Survey measures as outcomes in marketing experiments.

D.E. Adjusting biased samples. [optional, notes on adjusting survey data]

Apr 15 **Marketplace analytics**
Guest: Andrey Fradkin (Boston University & MIT)

How do we do analytics in marketplaces and platforms? Market design, transaction costs, and congestion. Reviews, ranking, marketing, and pricing. Challenges with experimenting in marketplaces. **Andrey Fradkin** is faculty at Boston University and a fellow at MIT's Initiative on the Digital Economy. He is a leading expert on the economics and analytics of digital marketplaces. He has a PhD in economics from Stanford University, and he previously worked at Airbnb as a data scientist and consultant.

Gurley, B. (2013). A rake too far: Optimal platform pricing strategy. *Above the Crowd Blog*. [optional]

Interview with Greg Lewis. (2016). Marketplace design and strategy. *Economic Frontiers podcast*. [optional]

Chetan Sharma & Jan Overgoor. (2016). Scaling knowledge at Airbnb. *Airbnb Engineering Blog*. [optional]

Andrei Hagiu and Simon Rothman (2016). Network effects aren't enough. *HBR*. [optional]

Fradkin, A. Digital marketplaces. *The New Palgrave Dictionary of Economics*. [optional, more academic]

Apr 20 **No classes held**

Apr 22 **Network data and multiple units**

Interventions and experiments in networks and marketplaces. Spillovers and interference as a nuisance. Cluster-randomized and time series experimentation. Users, households, cookies, and other units.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Ch. 22. [optional]

Aronow, P. M., Eckles, D., Samii, C., & Zonszein, S. (2020). Spillover effects in experimental data. To appear in *Advances in Experimental Political Science*. [optional, technical]

Taylor, S. J., & Eckles, D. (2018). Randomized experiments to detect and estimate social influence in networks. In *Complex Spreading Phenomena in Social Systems*. [optional, technical]

Coey, D., & Bailey, M. (2016). People and cookies: Imperfect treatment assignment in online experiments. In *Proceedings of WWW*. [optional, technical]

Apr 27 **Network effects, virtuous cycles, and marketing strategy**

Detecting network effects and winner-take-all dynamics. Amplifying network effects in your product. Substitution and competition between products. Using network data to learn otherwise unobserved consumer characteristics.

Apr 29 **Viral marketing and seeding in networks**

Compartmental (e.g., SIRS) models of contagions. Viral incentives. Targeting influencers.

Jackson, M. O. (2019). *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors.* Pantheon Books. [except here & on Canvas: Chapter 3, "Contagion"]

Watts, D. J. (2011). *Everything Is Obvious: Once You Know the Answer*. Crown Business. [excerpt on Canvas: Chapter 4, "Special People"]

May 4 **Dashboarding, illustrated with novel third-party data sources**
Data generated or captured outside our firm. Tracking pixels. Purchase data. Location data. Kicking off Assignment 5.

May 6 **Building analytics products and the digital marketing ecosystem**
Guest: Marco Matos (Pinterest)
The contemporary digital marketing ecosystem. Building analytics products, such as tools for advertisers. The changing landscape of data brokerage and analytics services. **Marco Matos** is head of ad measurement at Pinterest. He was previously a manager at Facebook, Google, and Microsoft, working on advertiser and developer tools. He studied electrical engineering and finance at Princeton.

Deighton, J. (2016). Acxiom. Harvard Business School Case 516-037.

May 11 **Recent developments in data use for marketing**
Guest: Eric Seufert (Heracles & MobileDevMemo)
On-going developments in how customers are tracked and targeted and how outcomes are measured. Recent proposals from Apple, Google, and Facebook, their expected impacts, and the future of digital marketing. **Eric Seufert** is principal at Heracles Media and owns and edits MobileDevMemo and QuantMar. His book *Freemium Economics* was published in 2014. He previously led marketing, advertising, and business intelligence teams at N3TWORK, Rovio, Wooga, and Digital Chocolate. He studied finance and applied economics at University of Texas, Austin, and University College London.

Ben Thompson (2021). An interview with Eric Seufert about Apple, Facebook, and mobile advertising. Stratechery. [optional, most recommended]

Eric Seufert (2017). The coming war between Apple and Facebook. MobileDevMemo. [optional]

Eric Seufert (2021). Facebook may take 7% revenue hit from Apple privacy changes. MobileDevMemo. [optional]

Eric Seufert (2021). How does Facebook's advertising targeting algorithm work?. QuantMar. [optional]

May 13 **CLV revisited: Churn, prediction, and networks**
Putting many pieces from the course together. Modeling customer acquisition, retention, churn. Interventions to prevent churn. Targeting and network effects. Examples from the telecommunications industry.

Gupta, S., et al. (2006). Modeling customer lifetime value. *Journal of Service Research, 9*(2), 139-155. [optional CLV resource]

de Matos, M. G., Ferreira, P., & Belo, R. (2015). Target the ego or target the group: Evidence from a randomized experiment in pro-active churn management. Working paper. [optional; academic paper used as example in class]

May 18 Analytics in B2B marketing and startups
Guest: Bryan Gaertner (DocSend)

What can we borrow from B2C for B2B? Defining product lines through analytics and listening to customers. Growing and democratizing analytics in a growing company. Observing and responding to big societal changes. Hiring in analytics / data science building data-friendly culture. Learning pricing policies. **Bryan Gaertner** is chief strategy officer for DocSend. DocSend, recently acquired by Dropbox, is a document sharing and analytics company whose products are widely used for fundraising, sales, and marketing. Bryan previously worked as an analyst and manager at Google and studied economics at Stanford.

Ron Miller (2021). Dropbox to acquire secure document sharing startup DocSend for $165M. TechCrunch. [optional]

D.E. (2014). It's better for older workers to go a little fast: DocSend in Snow Crash. [optional, mainly an excerpt from a novel]

May 20 Assignment 5 critique & highlights; Wrap-up

Kortina, Andrew. (2017). Metrics, incrementalism, and local maxima. [optional; Kortina is founder of Venmo]

Said, C. (2016). Four pitfalls of hill climbing. [optional]

Books

There are a few books that are particularly relevant for the course.

One is a new book, ***Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing***, 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. This text is **required**, so get a digital or hard copy.

In the past, we have relied more heavily on *Field Experiments: Design, Analysis, and Interpretation* by two political scientists, Alan Gerber (Yale) and Donald Green (Columbia), who helped create and study much of modern, data-driven political campaigning. This is a highly recommended resource on randomized experiments. It uses some of the notation and technical concepts that we employ in the lectures. Some of the suggested chapters are available on Stellar, but the whole book is recommended and inexpensive.

James et al.'s ***An Introduction to Statistical Learning*** and its more advanced predecessor / big brother (Hastie, Tibshirani & Friedman's *Elements of Statistical Learning*) are both available online for free, which makes them even more useful as resources to refer to in analytics teams.

Guests
We will have guests from industry, to be announced. This may require adjustments to the schedule above, which is preliminary.

Requirements
This section provides an overview of the main requirements of the course. More details about the requirements — and the resources, such as data, needed to complete them — will be provided throughout the course.

Class participation (10%)
Attendance (including online attendance) is required except for extenuating circumstances, which we understand students may encounter at some point.

Everyone's learning can benefit from vibrant discussion that reflects being prepared for class and drawing on additional knowledge or experience. The teaching staff will track attendance and contributions to discussion during class. Productive contributions will be rewarded.

Because of the work required for the assignments, all required readings have been carefully selected; read them before class.

Assignments (90%)
There will be five graded assignments given; based on previous iterations, we have gone with more, but smaller assignments with the goal of giving you more rapid feedback.

For Assignments 1 and 4, you can work on the assignments with other students but must turn in your own answers reflecting you actually doing data analysis and writing. Many of the questions will have you inform marketing decisions using a provided data set.

Assignments 2, 3, and 5 can be done as groups, though only Assignment 5 requires this.

Assignments 2 and 3 involve contests, whereby you submit predictions about consumer behavior or decisions about marketing interventions and these are scored; top performers get a bonus to their grade and prizes.

Assignment 0
Ungraded assignment to get started with R & Tableau. Exploratory data analysis.

Assignment 1
Planning and analysis of experiments. Analyzing a marketing experiment.
Due Mar 9.

Assignment 2
Penalized regression and modeling. Contest of predicting purchases.
First submissions due on Mar 19, contest ends Mar 25.

Assignment 3
Who should be treated? Contest of deciding how to target a costly intervention.
First submissions due by April 8, contest ends April 13.

Assignment 4
Analyzing a non-randomized marketing campaign.
Due April 29.

Assignment 5
Analyzing a market with geo-data and creating an interactive dashboard.
Due May 19, noon Eastern.

Late work policy for assignments
To add some flexibility, assignments 1 and 4 can be turned in slightly late with a 10% penalty (i.e., your score is multiplied by 0.9 ± rounding). But in order to allow us to keep the class moving (e.g., by providing solutions to other students), there is a limit to this. We will provide these "last dates" for each assignment.

Tools
Students can use any tools they like for the data analysis assignments, but we will provide some support for learning R and Tableau.

Tableau is commercial software for exploratory data analysis (EDA), dashboards, etc., which is seeing substantial use in industry. It can be used in combination with other tools, such as SQL, R, and Excel. Students can use it without cost via a MIT license.

R will be used for modeling, planning, and statistical inference. You can also use it for EDA and plotting. The free online course that I created with Udacity and Facebook colleagues might be helpful for that: https://www.udacity.com/course/data-analysis-with-r--ud651.

Further guidelines and policies

The teaching staff intends to follow all relevant guidelines and create a suitable environment for learning. Please let us know if you think we are coming up short.

Values and ethics

We expect that your behavior will be guided by the MIT Sloan Values.

Assignments 1 and 4 fall into Type 1 collaboration under the MIT Sloan standards:

> *"The professor states that collaboration is allowed, but the final product must be individual. An example of this might be a problem set.*
>
> - *You are allowed to discuss the assignment with other team members and work through the problems together.*
> - *What you turn in, however, must be your own product, written in your own handwriting [DE: please don't turn in handwritten assignments!], or in a computer file of which you are the sole author.*
> - *Copying another's work or electronic file is not acceptable."*

Assignments 2, 3, and 5 can be completed in groups of up to five students (from either section), under a "Type 3" collaboration:

> "*The professor states that collaboration is expected and that each team member must contribute substantially to the deliverable.*"

Assuming we have a sufficient number, I also intend to require that — for Assignment 5 only — all groups of at least three students not include only Master of Business Analytics students. I would encourage all groups to do so, but we may end up with a few smaller groups.

Communicating with teaching staff

All emails should include "15.819" in the subject or be to 15.819@. Quality of service for all other emails may vary!

For questions about assignments or course content, we encourage using the Discussion feature of Canvas, so that other students can see your query and the answer.

Electronics in class

For the in-person section, please bring your laptop to class and join the Zoom meeting without audio and video.

Student Support Services

"Whether you are struggling with a pset because of something going on in your life, you feel too ill to take an exam, you are considering taking time away from the Institute, or you just don't know who to talk to, we can help." - S^3

Please contact S^3 about any personal or medical issues that will interfere with your ability to complete any course requirements: http://web.mit.edu/uaap/s3/

Student Disability Services

If you may need disability accommodations, contact Student Disability Services (SDS) as early as possible. If some accommodation has already been put in place by SDS, please contact your TA ASAP so that we can implement it as smoothly as possible.