Decision making using marketing analytics

MKT 566

Instructor: Davide Proserpio

A little about me

- My 10th year at USC (office HOH 332)
- For the past two years, I have been working full-time for Amazon, first in the Ad Measurement organization and then in the Climate Pledge Friendly organization
  - Worked closely with data scientists, research scientists, sales teams, and product managers
  - I am still working for Amazon one day a week
- I run **Real-Estate Analytics Lab (REAL)** and the associated free Substack with Marco Giacoletti (www.realab.com)
  - Subscribe if you are interested!
- Personal website: https://dadepro.github.io/

A little about me

- Research on online marketplaces & policy
  - Trust & Reputation
  - Platforms manipulation
  - Short-term rentals (Airbnb)/Real estate
  - Advertising
  - Sports betting

- My website has a link to all my papers
- You can also find them on SSRN if you are interested

A little about you

- What’s your background?
- What do you expect from this class?
- Have you worked in marketing yet?
- …

The course

- Structure
- Syllabus
- Office hours
- Semester-long project
- Exams
- Grading
- …

The course

Instructor:

- Davide Proserpio
- Office: HOH 332
- Email: proserpi@marshall.usc.edu

Teaching assistant:

- Ignacio Riveros
- Office: HOH 103
- Email: iriveros@marshall.usc.edu

The course: Organization

Schedule

- Session 16546: 12:30 pm – 13:50 pm (Mon and Wed) in JKP 104
- Session 16547: 2:00 pm – 3:20 pm (Mon and Wed) in JKP 104

Office hours:

- Office: Hoffman Hall (HOH) 332
- Monday 4-6 pm (in person)
- Outside of this time slot, I am available via Zoom or in person by appointment

Class website: https://github.com/dadepro/mkt566

Syllabus: https://raw.githack.com/dadepro/mkt566/main/syllabus/mkt566-syllabus-proserpio.pdf

The course: Organization

- Most lectures will involve examples/exercises, and sometimes will dedicate the whole class to exercises/discussions/guest speakers

- The goal is to have at least three guest speakers from (very likely) tech companies (Amazon, StubHub, Netflix)
  - Potentially on Friday

- This class is very hands-on, so bring your laptop and be prepared to code and present your results

The course: Readings/book

- Slides
- Readings links (open source) will be provided for each topic
- Useful open–source materials :
  - My course on data storytelling
  - Data Visualization: A practical introduction
  - R for data science
  - R markdown
  - R for Marketing Students (similar to R for Marketing Analytics)
- Not open source:
  - R for marketing analytics (slides/exercises/code are open source)
  - Python for marketing analytics

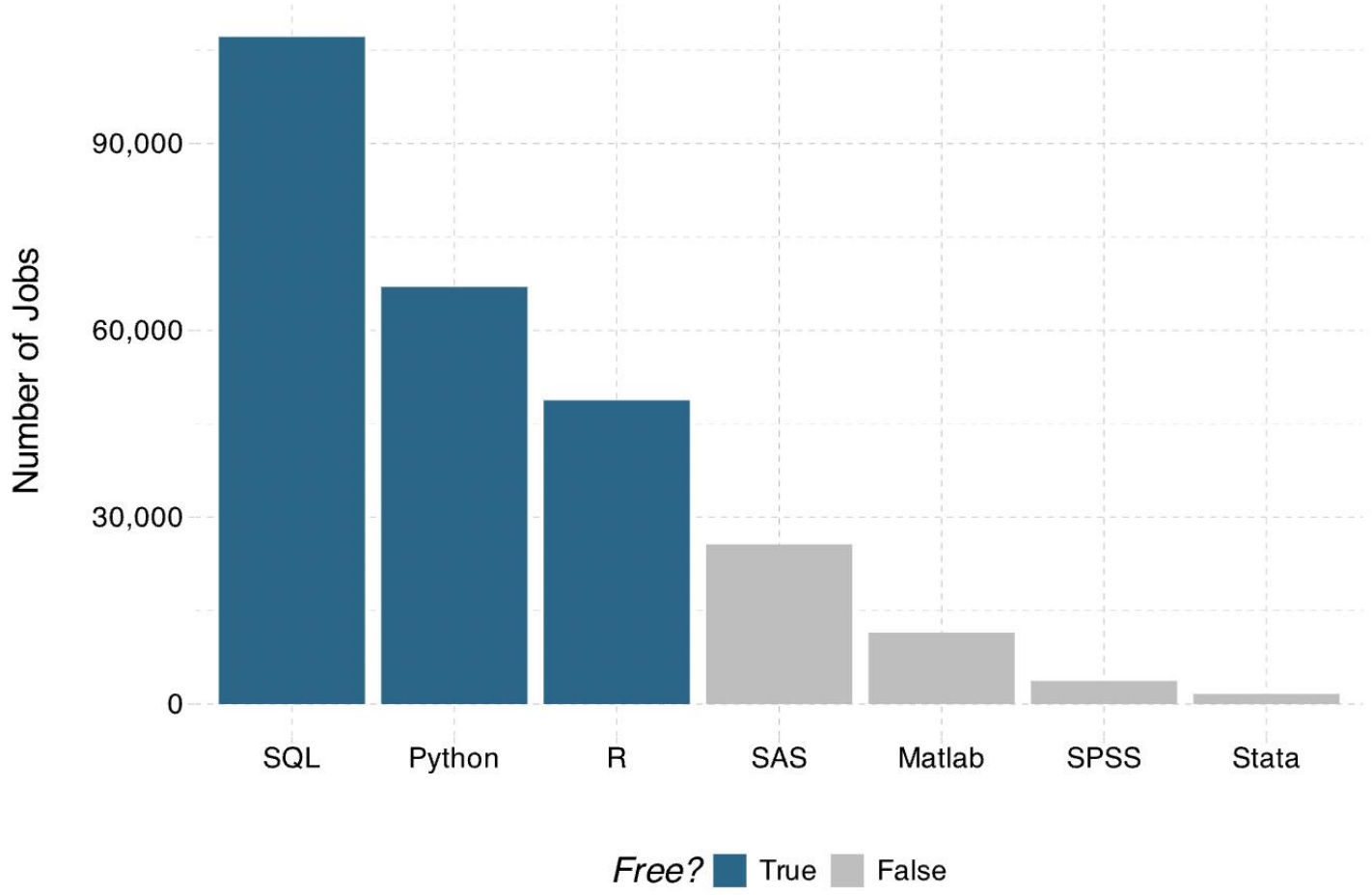
The course: Analyzing data/coding

- I am going to use (mostly) R and Python
  - R is primarily for data viz/statistics
  - Python is primarily for machine learning applications

- For your assignments, you are free to use either, but it is likely better to follow what I am doing (especially if you are not familiar with coding)
  - No, Excel is not allowed

- Suggestions about Integrated Development Environment (IDE):
  - Visual Studio Code (Python/R, and pretty much anything else)
  - R Studio (only for R)
  - Both work with Copilot ☺

The course: Analyzing data/coding



The course: Analyzing data/coding

Khyati Thakur • 2nd
Building Fridayy AI | MERN | AI Agents | Gen AI
1d •

+ Follow ···

AI PMs is the new hot job role. Netflix is paying $900K for AI PMs. Meta is paying $1M+.

Here are 8 skills every AI PM should have and a short course to start each one.

1️⃣ AI & ML Fundamentals
You don't need to train models, but you must understand how they work, their lifecycle, and their limits (hallucinations, bias, cost).
📚 Course: AI Product Management Specialization – Duke University (Coursera)

2️⃣ Applied AI Patterns & Tooling
Know agentic patterns (ReAct, Plan-and-Execute, CodeAct) and AI stacks (LLM APIs, vector DBs, LangChain, CrewAI).
📚 Course: IBM AI Product Manager Professional Certificate

3️⃣ Data Literacy & Evaluation
Be able to query, inspect logs, run A/B tests, and measure model quality (precision, recall, cost per token).
📚 Course: Generative AI for Product Managers Specialization

4️⃣ AI UX & Product Thinking
Design flows that balance creativity and control. Confidence scores, prompt chaining, guardrails.
📚 Course: AI for Product Management Course (Pendo)

5️⃣ Cross-Functional & Documentation Skills
Translate between engineers, data scientists, and business teams. Write PRDs with clear inputs, outputs, guardrails, and metrics.
📚 Course: Artificial Intelligence Product Certification (Product School)

6️⃣ AI Strategy & Business Acumen
Spot high-ROI AI use cases, evaluate competitors, and model trade-offs (latency, accuracy, cost).
📚 Course: Mastering Generative AI for Product Innovation (Stanford Online)

7️⃣ Experimentation & Deployment Mindset
Own offline + online evaluations. Know when to tweak UX, retrain models, or change data inputs.
📚 Course: AI Product Management Bootcamp (Maven)

8️⃣ Emerging Trends Awareness
Stay ahead on new models (Gemini, Claude, Mistral), multimodal AI, and evolving standards like MCP.
📚 Course: Elements of AI – University of Helsinki

LinkedIn link

Use of AI (LLMs)

- I expect you to use AI (e.g., ChatGPT, Gemini) in this class
- Learning to use AI is an emerging skill, but keep in mind the following:
  - AI tools are permitted to help you brainstorm topics or revise work you have already written
  - If you provide minimum-effort prompts, you will get low-quality results. You will need to refine your prompts to get good outcomes. This will take work
  - Proceed with caution when using AI tools , and do not assume the information provided is accurate or trustworthy
- **To foster transparency, I require each student to submit a log file of the LLM prompts used (if any) to help solve the assignments**

The course: Grading

- 4 Individual assignments: 60%
- Semester-long project (in groups, 5-6 students): 30%
- Participation: 10%

The course: Assignments

4 Individual assignments: 60%

- Roughly **three weeks** for each assignment
- For each assignment, you will submit a **R (Markdown/Quarto)/Python notebook (converted to HTML or PDF)** with code properly commented, outputs (e.g., figures, tables), and the answer to the assignment questions
- Log file of the LLM prompts used (if any) to help solve the assignments

- Why do I ask you to submit a notebook-generated file?
  - Reproducibility is very important!

The course: Semester-long project

Analyze a real-world dataset to answer a (or a set of) marketing-related question(s)

- Social media sentiment analysis:
  - To learn brand perceptions and how they changed over time
  - To understand which content performs better
- Customer segmentation: Who are our most valuable customer segments?
- Churn prediction & retention: Which users are likely to churn next month?

The course: Semester-long project

Some examples from a colleague's class

- **ChatGPT and online content:**
  - Using data from Stack Overflow + YouTube, students found a decline in activity on Stack Overflow for coding topics (i.e., python/java) but an increase in AI related videos for some YouTube channels (i.e., tech channels)
  - Link: https://tarikajain.substack.com/p/the-ai-disruption-chatgpts-takeover

The course: Semester-long project

Some examples from a colleague's class

- **Sentiment analysis of Sephora reviews**
  - Using data on customer reviews for Sephora, students aimed to determine whether reviews play a role in helping to identify which Sephora products will be featured as "trending" on the website.
  - Link: https://medium.com/@chaitanyaparachotill_75259/nveiling-sephoras-beauty-secrets-a-data-driven-approach-to-customer-engagement-e0d06a96aa05

The course: Semester-long project

Some examples from a colleague's class

- **Breaking box office**:
  - Using data on movie reviews from Kaggle and revenues from Box Office Mojo, the students documented a null effect i.e., customer reviews generally don't explain box office performance but may still be useful for less popular movies
  - Link: https://sites.google.com/view/imdb-marketing-analytics-grp6/imdb-review-analysis

The course: Semester-long project

Some examples from a colleague's class

- **Gender effects in a dating app**
  - Using review data from Bumble, a female-friendly dating app, my students discovered that, although more Bumble users appear to be male, women tend to rate the app higher than men. Nice application of the gender-guesser package!

The course: Semester-long project

Where can we find data?

- Tons of available datasets online (Google is your friend), e.g.,
  - Kaggle datasets (e.g., customer churn, customer shopping trends, Steam)
  - Yelp open dataset
- Amazon reviews, Google reviews, Steam dataset, and more: https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews
- Marketing Science research papers often come with data and code for replication
- I do have some data that I can share about Airbnb, TripAdvisor, Expedia, and Yelp
  - Let me know if you want more info about this data
- Collect it yourself via API (e.g., Twitter)

The course: Semester-long project

Deliverables and deadlines:

1. Form groups: September 15
2. Slides for mid-term project proposal (~15 mins including Q&A) on October 13,15 (due date October 12)
2. R (Markdown/Quarto)/Python notebook with data cleaning and analysis properly commented by December 3 (HTML or pdf format)
3. Slides for final presentation (~15 mins including Q&A) on December 1,3 (due date November 30)
4. Peer evaluations: due date December 3

The course

Everything I just discussed can be found in the syllabus, so please **read it at least once**.

Let’s begin!

Marketing analytics

Marketing analytics is the practice of measuring, managing, and analyzing data from marketing activities to optimize business performance. It covers:

- **Data Collection & Integration**
- **Visualization & Reporting**
- **Advanced Analysis**
- **Defining, Measuring, and Tracking Performance Measurement (KPIs)**
- **Translating Insights into Decisions**

Data Collection & Integration

- Gathering information from sources such as the web, social media, e-commerce platforms, etc.
- Unifying datasets from different sources (e.g., online clicks, offline transactions) to answer marketing-related questions

Visualization & Reporting

Creating figures, building dashboards and reports to present insights for stakeholders, e.g.,

- Advertising effect on sales
- Eco-labeling program growth
- Visualizing the relationship between variables
- …

Advanced Analysis

- **Attribution Modeling (causality):** Determining which channels or campaigns deserve credit for conversions.
- **Recommendation & segmentation (clustering):**
  - Grouping customers by behavior or demographics to tailor messaging
  - Identify products that consumers will likely buy
  - Product placement
- **Predictive Modeling**
  - Forecast ad campaign performance
  - Predict customer churn
  - Identify fraudulent “players” (e.g., fake reviews, accounts, clicks, etc.)
  - Predict/forecast prices

KPIs

Tracking key metrics:

- **Acquisition** (e.g., cost per click, cost per acquisition)
- **Engagement** (e.g., click-through rates, time on site)
- **Conversion** (e.g., lead-to-customer rates, average order value)
- **Retention** (e.g., churn rate, customer lifetime value)
- **Incrementality** (e.g., how much sales can we **attribute** to the ad campaign)

Translating Insights into Decisions

Turning insights into action:

- **Reallocating budgets** to top-performing channels
- **Optimizing** messaging and creatives,
- **Changing** the landing page
- **Identify** and leverage new trends
- **Optimize** prices

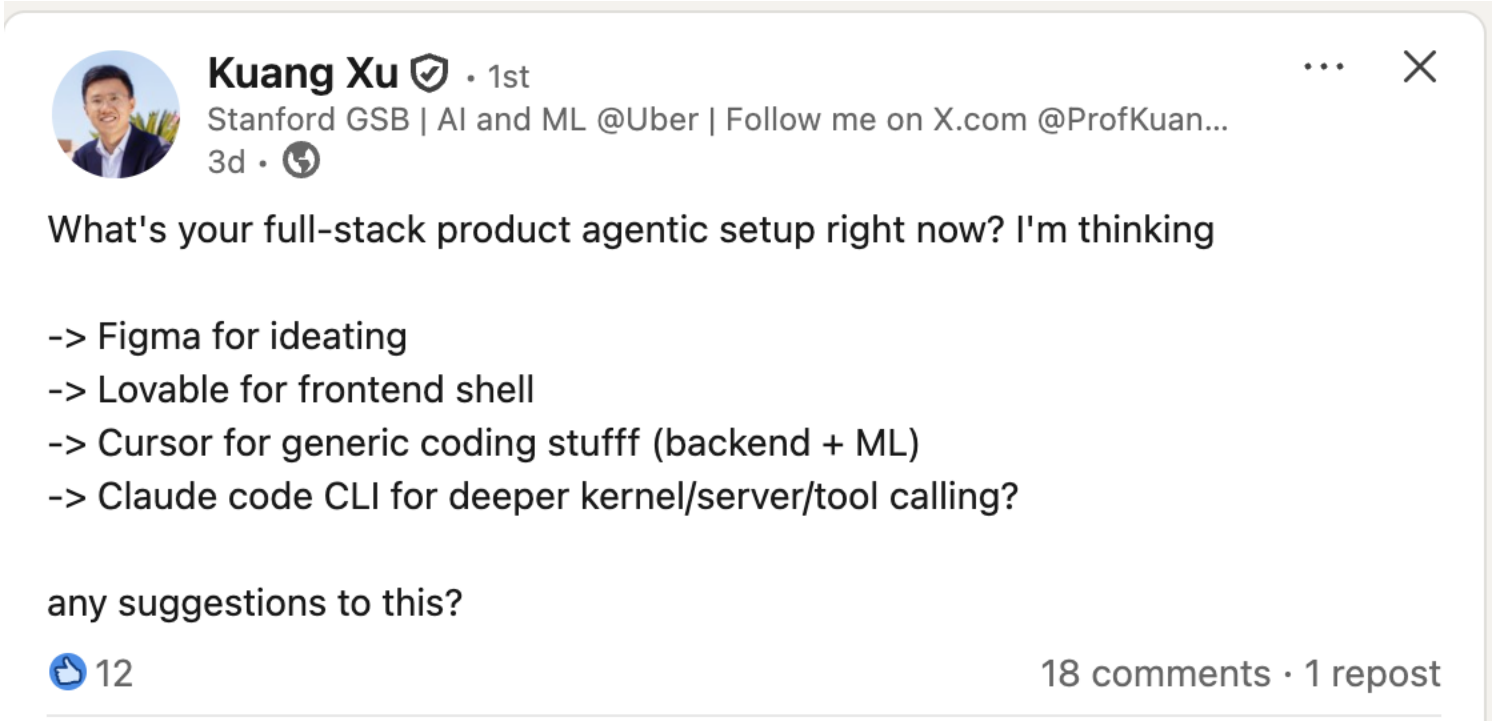
Marketing analytics in a nutshell

By systematically applying statistical methods and data-driven storytelling, **marketing analytics enables organizations to understand what works, identify growth opportunities, and maximize their return on marketing investments.**

What we are going to cover

- Data analysis and visualization
- Regressions
- Clustering and recommendations
- Classifiers
- Causality (experiments and observational data)
- LLMs and agents (new challenges for brands, e.g., search; how firms are leveraging LLMs, etc.)
  - https://x.com/bearlyai/status/1942327568948158780

What we are going to cover



LinkedIn link

What we are NOT going to cover

- This is not a programming language course, so it is (mostly) up to you (and LLMs) to learn how to code

- This is not an advanced statistics/econometrics/ML methods course. The emphasis of this course is on data-driven decision-making. So, we will focus on a high-level understanding of the concept covered
  - I am assuming you **are familiar with basic statistics and probabilities,** but I can post some review slides if needed

Questions?