

Workbook

Analytics for Decision-Making

Instructor: Patrick Boily

Description: Data analytics improves the use of evidence-based solutions to solve difficult governance problems. Organizations are increasingly turning to data to help facilitate transformational changes.

This course provides participants with a baseline knowledge of analytics to support evidence-based decision making. Participants will improve their understanding of the processes used in decision making, and the tools and techniques for the application of analytics to these processes. Discussions include the most common and useful analytics methods, techniques, and software. This course will provide participants with a knowledge base that will equip them to make improved allocations of resources in the most effective way.

This is a survey course. Programming and statistical competencies are not required.

Instructor/Facilitator: Patrick Boily is a graduate from the University of Ottawa. He obtained his Ph.D. in Mathematics in 2006. He has taught over 55 courses at Universities in the Ottawa area since 1999, and has worked on numerous projects as a federal public servant from 2008 to 2012 (including the award-winning Canadian Vehicle Use Study). He started and managed Carleton University's Centre for Quantitative Analysis and Decision Support from 2012 to 2019.

He is now a professor in the University of Ottawa's Department of Mathematics and Statistics. He has provided consulting services to multiple entities over the years, including the Canadian Air Transport Security Authority and the Nuclear Waste Management Organization. He has extensive experience in operations research and data science. He leads various workshops on data science and statistical analysis.

Course materials have been prepared in collaboration with Jen Schellinck, Stephen Davies, John Stroud, and Marwan Kashef.

Contents

Module 1

Decision-Making

Importance of decision-making; decision-making strategies; decision biases; superforecasters and crowdsourcing; augmented intelligence and decision-making; decision making tips.

Module 2

Reasoning, Evidence, Information, Data

Analysis; borrowing strategies from the sciences; types of reasoning.

Module 3

People, Data Ethics, and Laws

Analytics roles and teams; ethics in the data context; data and the law.

Module 4

Business Intelligence and Analytics

Business intelligence history and terms; description, explanation, prediction, prescription; exploring "what if" scenarios; value of presentation and persuasion.

Module 5

Analytics for Decision Support

Decisions about analytics; options, methods, and tools; reasonable expectations; pitfalls; practical advice; communicating results through visualizations and dashboards.

Module 6

Data Engineering and Data Governance

Data pipeline overview; data pipeline components; data governance roles; data governance practices.

Exercises : Module 1 – Data Insight Fundamentals

1. With 26 seconds left in the Super Bowl, the Seattle Seahawks were trailing the New England Patriots by 4 points. At 2nd & Goal, the Seahawks had the ball at the Pats' 1 yard line. The common wisdom in this situation is to hand the ball to the running back and let them try to punch through the defensive line. The Seahawks had two options:
 - A. Run the ball (1 play). **Risk:** Fails to score and time runs out.
 - B. Throw the ball instead, then run if necessary (2 plays). **Risk:** 2% chance of interception.What play should the coach call? Why?
2. In the ***Vanity Fair*** article "You Could Fit All the Voters Who Cost Clinton the Election in a Mid-Size Football Stadium", Tina Nguyen writes:

While nearly 138 million Americans voted in the presidential election, the stunning electoral victory of Donald Trump came down to upsets in just a handful of states that Hillary Clinton was expected to win. It has been cold comfort for Democrats that Clinton won the popular vote—at the last count, she was up by about 2.5 million votes, and climbing, as ballots continue to be counted. Even more distressing is the tiny margin by which Clinton lost Wisconsin, Michigan, and Pennsylvania—three states that were supposed to be her firewall in the Rust Belt, but that ultimately tipped the electoral college map decisively in Trump's favor.

Trump's margin of victory in those three states? Just 79,316 votes.

This latest number comes from Decision Desk's final tally of Pennsylvania's votes, where Trump won 2,961,875 votes to Clinton's 2,915,440, a difference of 46,435 votes. Add that to the official results out of Wisconsin, where Clinton lost by 22,177 votes, and Michigan, which she lost by 10,704 votes, and there you have it: 0.057 percent of total voters cost Clinton the presidency.

It is not entirely unusual for the electoral college to be lost by such a slim margin. In 2000, Al Gore lost Florida (and therefore the election) by 1,754 votes, triggering a painfully drawn out recount drama that only ended with a Supreme Court ruling. And in 2004, John Kerry lost to George W. Bush by losing Ohio by a little over 118,000 votes. But it is worth considering just how few voters ultimately set the country on its current, arguably terrifying course. The 79,316 people who voted for Trump in Wisconsin, Michigan, and Pennsylvania—all states that Democrats carried since 1992—is less than the entire student body of Penn State (97,494 students), or only slightly more than the number of people who attended Desert Trip, the Baby Boomer-friendly music festival colloquially known as "Oldchella." If you put all these voters in the Rose Bowl, there would be slightly over 13,000 seats left over.

There are more people living in Nampa, Idaho, a city you have never heard of.

To put things in even more painful perspective, Green Party candidate Jill Stein won about 130,000 votes in those three states. Libertarian candidate Gary Johnson won about 422,000.

But perhaps the most painful data point for Clinton is this: the Democratic nominee for president never made a single campaign stop during the general election, and largely neglected Pennsylvania and Michigan, too, while Trump canvassed all three states relentlessly. His furious, last-minute blitz throughout the Rust Belt to win white, working-class voters, combined with the lack of resources Clinton invested, essentially handed their combined 46 electoral votes to Trump. Instead, Clinton spent the last few weeks of her campaign expending resources in places like Arizona and Texas—states which went for Trump by huge margins.

So was it bad luck, or a mistake? Why?

3. Revisit the last two questions in light of the ***Luck and Information*** slide.
4. Do the Medical Treatment Exercise (see slide deck).
5. Over the next week or so, pay attention to your decision-making process (or the decision-making process of your organization). What decision biases are you most susceptible to? Least?

Exercises : Module 2 – Reasoning, Evidence, Information, Data

1. Are the following arguments strong? If they are weak, what are their flaws? How would you improve them?
 - a. COVID vaccinations lead to increased hospitalizations as half of the hospitalizations were vaccinated.
 - b. Turning the Large Hadron Collider on was a mistake because either it destroys the Earth or it does not; a 50% chance is way too risky.
 - c. We know that the Earth is not flat because none of the other planets we know are flat.
 - d. You should not vote in the next election because one vote never makes a difference.
 - e. The solution to reduce congestion is to reduce the number of lanes because with fewer lanes, people will seek alternative modes of transportation.
 - f. Airport security measures are proportionate to the risk because it's ok to wait a few hours if it means that my plane won't be hijacked.
2. Consider the items found in a briefing note relating to building a pipeline through caribou territory:
 - a. The last 7 times pipelines were constructed in caribou territories, populations decreased in the territory.
 - b. Biologists created a map showing the caribou migration paths. Based on this map, we conclude that placing the pipeline over the territory will not interfere with caribou migration.
 - c. Pipelines have not affected geese populations; as they and caribous are both social animals, the pipeline will not affect the caribou population.
 - d. Biologists have shown that caribous are not scared of large objects. If caribous are not scared, their breeding habits will not be affected. As pipelines are large objects, constructing this pipeline will not affect the breeding habits of the caribous on the territory.

Identify the reasoning strategies (generalizing from examples, making a deduction, reasoning by analogy, reasoning to the best explanation) being used in each of these arguments. Applying a plausible reasoning lens to this, what would you conclude? What additional information would you need/want, before drawing a conclusion?

3. In Tom Stoppard's 1966 play *Rosencrantz and Guildenstern are Dead*, Rosencrantz flips a coin and observes heads 92 times in a row. Is this possible? Is this plausible? If this happened to you, what would you conclude?

Exercises : Module 3 – People, Data Ethics, and Laws

1. Conduct the generalist/specialist exercise for yourself or your team.
2. A bank has an obligation to increase its shareholder value. It is considering a new AI-driven decision-making process for loan applications, with the objectives being to free staff from having to complete tedious tasks and to reduce the default rate. The bank has access to proprietary and public information about loan applicants. Not knowing what position you may hold in the future, what issues (if any) do you have about this proposal? What ethical principles would you want the bank to keep in mind?
3. Discuss the Amazon hiring case study presented in class.

4. Meltwater offers software that scrapes news information based on specific keywords. Clients order summaries on topics containing excerpts of news articles. The Associated Press (AP) claims that their content was stolen and that Meltwater needed a license before distributing the information that was scraped. The judge found in favour of AP arguing that Meltwater is a competitor. What is your verdict?
5. In 2014, six Italian geologists were cleared of manslaughter charges in the deaths of 29 people, who perished in a major earthquake in 2009. They had initially been convicted of manslaughter for failing to predict the earthquake that ravaged the Italian town of L'Aquila. But in the wake of their successful appeal, some questions remain, including whether scientists will be willing to continue dispensing advice to laypeople, now that they know they can be threatened with prosecution for doing so. What does such a story suggest to you about using data/evidence to make decisions?

Module 4 – Business Intelligence

1. Conduct a premortem/backcasting exercise for a new youth mental health initiative. Assume that your department has created an app which aims to improve the mental health of Canadian teenagers, post-pandemic. Assume that it is now two years from today and you are looking back on the app's launch.
 - a. **Premortem:** Give 3 reasons within your control and 3 outside of your control why the launch failed.
 - b. **Backcasting:** Same, but for why the launch succeeded.
2. Construct a story spine relating to a successful decision that has been made by your organization (such as budget allocation, introduction of a new program, etc.).

Module 5 – Analytics for Decision Support

1. How could you use analytical methods to improve the quality of recommendations in the briefing note exercise of Module 2.
2. What are some analytics use cases in your department?
3. Consider Minard's March to Moscow? In what way(s) is it a good visualization? A poor one? What story does it tell? What concepts are represented in the chart? What about the other charts in the slide deck?
4. What stories do the data visualization posters support? How much context is needed to make sense of them?

Module 6 – Data Engineering and Data Governance

1. Draft your dream data team (finite resources, FTE boxes, cost, priority).
2. What do you think the top 5 data governance priorities should be for:
 - a. your organization?
 - b. the Government of Canada?
3. Have you have had issues related to data availability, usability, consistency, integrity, quality, security, and/or trustworthiness? Discuss with the class.
4. Does your work group create or generate data? What type of data? Do you use data from sources outside of your group? How many? Which ones? Do you publish analysis of data internally to your group, externally or both?