Data Analysis for Business Analytics

CIS621

Professor - Micah Melling

Chief Data Scientist at Americo Financial Life and Annuity

Former Senior Director of Data Science at Spring Venture Group

Availability

- Email: Micah.Melling@park.edu
- Before class
  - I will try to always get to class by 5:30
- After class
  - I will stay for 15 minutes after class ends
- Other evenings, including weekends
  - Mostly available after 9:00 pm

Classroom Expectations

We’re all adults

Course Purpose

“This course will teach students to apply statistical analysis and other forms of data analysis to solve business problems. Students will design analytical methodology and use relevant software tools to implement the analysis. Students will also learn to evaluate the results of the analysis, its limitations, and their applicability to support business decisions.”

Course Structure

- Blended course
- Weekly class meetings from 6:00 pm - 9:00 pm
  - Lecture materials
  - Technical examples
  - Discussions
  - Activities
- Weekly online assignments
  - Discussion posts
  - Homework assignments (papers and data analyses)
- Final project

Online vs. Lecture Components

- Online
  - Curriculum and assignments laid out by Park University
  - You can work ahead if you would like
- Lecture
  - Discuss “real world” realities
  - Engage in discussions
  - Gain exposure to a variety of concepts

Software

- Paid options
  - JMP
  - JASP
  - SPSS
- Free options
  - R
  - Python

Lecture Outcomes

1. Understand the *real-world landscape* of using data to make decisions
2. Understand the *limitations* of using data to make decisions
3. Understand the *reasons* we use data to make decisions

Major (General) Areas of Analytics

Descriptive:
What happened?

Predictive:
What will happen?

What about Prescriptive Analytics?

- This is an over-hyped term that, in reality, means one of the following:
  - Optimization (often constrained optimization)
  - Predictive model output mapped to an action / decision
    - Example: If the probability is more than 80%, send an email.
  - Predictive model where the labels are actions / decisions

Group Discussion

What are sources of information that businesses use to make decisions?

Examples: Customer Database, Subject-Matter Expertise

Quantitative Business Decision Spectrum

Magic 8-Ball | Hunches | Expertise | Numbers | Analysis | Terminator

Where Should a Business Try to Operate on the Spectrum and Why?

Magic 8-Ball | Hunches | Expertise | Numbers | Analysis | Terminator

Where Should a Business Try to Operate? (General, Ballpark Answer)

Magic 8-Ball	Hunches	Expertise	Numbers	Analysis	Terminator
5%	0%	20%	0%	75%	0%

Individual Activity

Make a list of at least 10 challenges a business might face when trying to use data to make decisions.

Example: Not having funding to purchase certain datasets

Realities of Real-World Data Decision Making

Imperfect and incomplete data

Ever-changing environments

Time pressures and windows of opportunity

Highly-complex, intertwined systems

“Chaos”

Randomness

Human biases

The goal of business decision making is often not about being “right”. It is about being “less wrong” than your competition or than you were previously.

Why can’t we always be “right”?

We live in a world of scarcity - we don’t have full information.

We live in a world with tradeoffs - we can’t always get the data we want or it might be too costly.

We live in a world of randomness - we can’t always detect the signal clearly.

We live in a world that is non-stationary - what is valid today may not be so tomorrow.

We live in a world with imperfect models - in fact, models are designed to make errors because they search for patterns!

However, we can use math, statistics, and data to cut through the challenges and take steps in the right direction.

“Operationalization is the creation of measurable proxies for rigorously investigating fuzzy concepts.” - Cassie Kozyrkov

https://towardsdatascience.com/operationalization-the-art-and-science-of-making-metrics-31770d94998f

But…a misapplication or miscommunication of math, statistics, and data can move us in the *wrong* direction. What’s worse - in such a scenario - we believe we are closer to “right” when we have actually moved closer to “wrong”.

Know the Rules to *Bend* the Rules

- Statistical methods have rules, which can often turn into pitfalls.
  - LINE assumptions in linear regression
  - (To note, beyond very simple cases or “research”, you will almost never use plain linear regression)
- If you know the rules, you can note the limitations to which your analysis is subject.
  - “We used a technique intended for a normal distribution even though our data wasn’t perfectly normal”.

Know the Rules to *Bend* the Rules

- Sometimes we have *bend* the rules in our pursuit of being “less wrong”.
    - Having the results of a regression model, even when some assumptions are violated within reason, can produce more useful knowledge than a heuristic in many cases.
- We have to be careful: If we don’t understand the limitations and *really break* the rules, we could end up being *more wrong*.
    - We use a linear model to predict an inherently nonlinear process.

Knowing the rules is science. Applying - including breaking - the rules is science *and* art.

Example: Omitted Variable Bias

Intellectual Honesty

Intellectual honesty is seeking the truth regardless of personal beliefs, rejecting assertions that don't have sufficient proof, and always applying rigor to the work.

This is a must in our field - especially since we know there are limits to our knowledge and to our data.

Own Your Mistakes

You’re going to make mistakes. Own them and set the record straight.

Brave Things a Data Analyst Can Say:

“I don’t know”

“We need more data”

“The results are tentative”

Group Discussion

You discovered you made an error in an important report.

How do you tell your boss?

Outline the points you want to cover and important statements you want to make.

Process Transparency

- Always share the following when applicable:
  - Assumptions you made
  - Decisions you made about handling the data that could have changed the outcome (e.g. outliers)
  - Known data quality issues
  - Level of confidence in results (e.g. these results are ‘directional’)
- However, don’t bog a business audience down in technical details.
  - If you applied an ARIMA model to a time series that you couldn’t make exactly stationary, you don’t need to share these technical details. Rather, don’t be overconfident when sharing predictions. Instead, state the predictions will give us a *decent idea* of how a series might move.
- Many business people often think “analytics” is magic: you throw in data and get perfect output. It’s the job of the analyst to properly temper expectations.

Example: Can Data Say Whatever We Want it to Say?

Be a Scientist: Know Your Main Research Question

For every analysis, know the main problem you are trying to solve. Also understand that though your analysis may only be a piece of this puzzle, you should still have the final goal is mind.

Examples:

- We are trying to determine the efficacy of a recent marketing strategy.
- We are attempting to create customer segments.
- We are working to find coaching opportunities for customer service reps.

This Isn’t Freshman English: Be Concise

- The goal of *business* analytics is to help *business* stakeholders make decisions.
- The primary goal of any report is to clearly and concisely articulate the main findings and how they answer the main research question at hand.
- Major supporting evidence should be accompanied by a concise explanation.
- Further technical details and documentation can be put in an appendix.

A Business Report Template

- Introduction (2-4 sentences)
  - What problem you were trying to solve
  - What you did to solve this problem
- Takeaways (2-4 bullet points)
  - What you found, what it means, and how it relates to the main research question
- Important Notes (2-4 sentences)
  - Any key assumptions or limitations that should be communicated up front
- Supporting Evidence (1-2 visuals for each takeaway)
  - Table / non-technical graph and 1-3 sentence summary for each takeaway
- Appendix
  - Documentation
  - Additional supporting evidence and technical analysis

Report Collaboration

- When developing a complex report, don’t show up with a report several weeks or months down the road.
- Rather, share bits and pieces over time with the target audience and get feedback. (The final report can still be comprehensive rather than piecemeal).
- Additionally, getting some structure on paper in short order and getting early feedback is often invaluable.
  - “This is a working draft of what we are seeing thus far. What do you think about this direction?”

Having No Finding is Many Times a Valuable Finding

Be Humble in the Face of Complexity

- Humans crave answers.
- We find patterns where none exist.
- We are “fooled by randomness”.
- Don’t provide a “statistical” answer that doesn’t exist.

Working with Data is Recursive, Not Linear

- We often cannot chart out in advance exactly how to do an analysis.
- Likewise, we often do not know *exactly* the types of analyses will help yield worthwhile answers before diving in.
- Oftentimes, we can sketch out a plan and produce a buffet of ways we *could* analyze the data. But we must remain flexible.
- Plans go out the window when we start interacting with real world data.
- We have to start analyzing, see the results, and adjust.

Considering all these challenges with analyzing data, why should we do it?

Properly Analyzing Data is the Best Way to be *Less Wrong.*

Using statistics and mathematics to make decisions is often treated as a given. However, we should look deeper into *why* these tools are useful for humans.

Math and Statistics are “Pure”

- Statistics and models lay bare certain assumptions about how a process works. They are transparent.
- The math itself is unbiased.
- The rules of a statistical technique are repeatable and auditable.
- The methods will tell you when they cannot find an answer.
- It’s the human - not the math - that messes things up through misapplication and miscommunication.

Math and Statistics Distill Information Efficiently

- We can take reams of data, throw them into a statistical technique, and get an interpretable output in seconds. Humans cannot perform equivalent calculations.

Math and Statistics Answer Challenging Questions

- Mathematical techniques can often solve incredibly challenging problems.
- Big Examples: self-driving cars, Google Translate, FaceID

Math and Statistics Answer Challenging Questions

- More “Down-to-Earth” Example:
- A company wants to know if multiple sales promotions interacted with evergreen marketing efforts to produce synergistic results. This is a difficult problem to solve, one that could never be perfectly scientifically solved. Without data analysis, we are left with opinions and hunches (and the Magic 8 Ball), which won’t suffice for answering such a complex question. With proper data analysis, we may not get a perfect answer, but we could get a *realistic idea of what the answer could be*, which will help us be *less wrong* than before.

Math and Statistics Combat Human Weaknesses

- They don’t always tell us what we want to hear (they only do so when you manipulate them).
- They work to find patterns and ignore noise (though not perfectly).
- They don’t care about preconceived notions (well, unless you’re a Bayesian).
- They let you know how they work and what assumptions they make (even complex machine learning models can be explained with modern techniques…don’t believe people who say ”black boxes” exist).
- They are consistent - the same input will give you the same results (in *many* cases but not *all*).

Notice all the caveats on the last slide? That’s the purpose of this class - you need to know the amazing potential data analytics has but also how it can burn you!