The "open" in "open data" means that it is provided to you at no cost, you are free to use it in any way you wish, and (usually) there is a "data dictionary" associated with the data that provides you with meta-data about the dataset.

Find datasets that your team finds interesting. Find datasets that would answer **business or management** questions. For example, the City of Kamloops has a lot of management questions that can be answered with data; it is in the "business" of providing civic services to those that live here. Do not re-use the Mobi Bikes data we used for Case #1.

→ While there is lots of sports analytics data out there, **DO NOT** do a sports analytics project! Where the Maimi Heat and Oklahoma City Thunder hit their shots from is interesting and all, but doesn't answer business or management questions…
https://archive.nytimes.com/www.nytimes.com/interactive/2012/06/11/sports/basketball/nba-shot-analysis.html

Questions about data / data sources? Post 'em to the Moodle Discussion Forum.

Step #1B — Get Messy with the Data: Explore the data. Get messy with it. Summarize it. Visualize it. Use applicable tools from this course (database analytics, forecasting, regression, optimization, simulation, statistics, data visualization, etc.) to explore the data and develop some insight for people who may be interested in the data. E.g., if you're looking at economic data for the City of Kamloops, perhaps a homeless shelter or foodbank might be your intended audience. Or a real-estate developer. Or another city of comparable size. Likewise, when we looked at "Mobi Bikes" — audiences could be the operators of the service, related services like tourism services, etc. Thinking about an audience will help with your analysis.

➔ Can you **combine your dataset**(s) with other data? Weather data? Stock market data? Geographic data? GDP data? It needs to make sense of course. Combing Mobi Bikes data with weather creates valuable information & insight. Combining Mobi Bikes data with currency market exchange rates of the Canadian and American dollar is tenuous at best ☺

Questions about what sorts of analysis to do? Post 'em to the Moodle Discussion Forum.

Your team can use any tools you wish — provided we've done the *technique* in class. If you want to do stuff in Excel or Tableau — that's fine and what I expect most will do. If you want to do it all in R Markdown and have it knit to a PDF and PowerPoint — that's clever and will save time. Python? Hey — fill your boots! If you want to do machine learning or k-means clustering — don't. Stick to the techniques we do in class.

Step #1C — develop a presentation of your analysis. Do it in the form of a presentation to a particular audience (see above). For example, think of your audience as **your client** — so give us a suitable title — Executive Director of Kamloops Out of the Cold shelter or Operations Manager for Santander Cycles (the official name of Boris Bikes). You are a team of consultants. Use professional business language. Show up in professional (business) attire. State the objective of your analysis in the opening of your presentation!

Provide your client with details on your approach. For example: “Our team looked at economic data for the 2017-2020 timeframe from the City of Abbotsford open data portal” vs “we copied the data into Excel. Cells A2:B912 were selected and we used the Data Analysis TookPak to complete our regression…”

Provide us with the information you created as well as your interpretation. Don’t just paste some Excel output into your slides — format it, extract it, explain it.

For example, rather than paste something like this into your presentation and leave it to your audience to interpret…

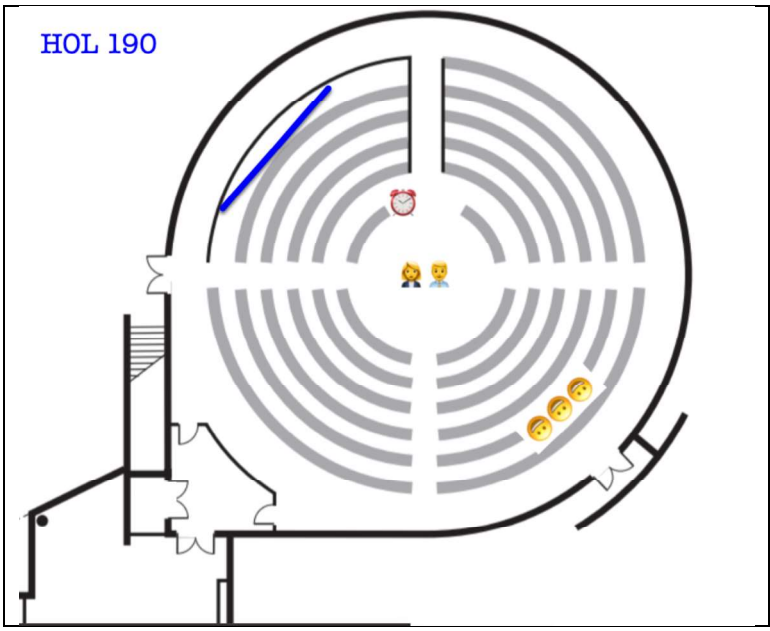
	A	B	C	D	E	F	G
1	**Linear Regression**						
2							
3	**Regression Statistics**						
4	*R*	0.8574					
5	*R Square*	0.7352					
6	*Adjusted R Square*	0.7057					
7	*S*	4.2176					
8	*Total number of observations*	11					
9							
10							
11	**ANOVA**						
12		*d.f.*	*SS*	*MS*	*F*	*p-level*	
13	*Regression*	1.0000	444.4110	444.4110	24.9831	0.0007	
14	*Residual*	9.0000	160.0963	17.7885			
15	*Total*	10.0000	604.5073				
16							
17		*Coefficients*	*Standard Error*	*LCL*	*UCL*	*t Stat*	*p-level*
18	**Intercept**	-3,911.8355	803.8710	-5,730.3179	-2,093.3530	-4.8662	0.0009
19	**Year**	2.0100	0.4021	1.1003	2.9197	4.9983	0.0007

… ***discuss*** your findings (“our regression analysis indicated a positive relationship between year and housing index; more importantly — it’s explanative power was high: 85% of the variation…”) and so on.

Questions? Post ‘em to the Moodle Discussion Forum.

(ii) presented to the audience in a clear and cognizant way…

The Presentation: This will happen in the House of Learning (HOL) room 190 — a large, round, 300-seat theatre space. There are four screens in the room far above you — you won’t be able to stand next to the screen and point to stuff like you could in our classroom. You will have exactly five minutes on a timer that you will be able to see. Your presentation will end at the 5:00 mark whether you are done or not. The five-minute constraint is so we can get through all of the presentations that evening.



The two business people emojis represent where you / your team would stand 👩‍💼👨‍💼

The clock ⏰ is at the console computer where the timer will be.

The blue line represents one of the four screens — your audience will be across the theatre from the screen (😀😀😀) — make it readable for them!

Your team can decide how you want to present — do all five (or six) of you present? Just one person from the team? Maybe a couple of people? You decide.

There are formats for short presentations. One is “3MT” (three-minute thesis) — it is a graduate student competition (usually PhD students) that have 3 minutes and maybe one slide to talk about their research. There are all sorts of tips & tricks videos on YouTube and elsewhere about this format. Another is Ignite! — 20 slides / 15 seconds/slide auto advancing in the background. Here is an Ignite! Talk about the format: https://youtu.be/rRa1IPkBFbg

➔ check these approaches out — not that you must use one of them, but to learn more about how to think about creating a short presentation.

There are a LOT of excellent resources available to help you up your presentation game. At Moodle, I’ll post a PDF document called "Zenify Your Presentation" — this is the handout from a seminar I have run multiple times in the past about applying the principles of Garr Reynolds' "Presentation Zen" to your work. I recorded the seminar and it is split up into three YouTube videos:

Zenify Your Presentation videos
#1 https://youtu.be/8HgnNSStaxk
#2 https://youtu.be/QVEO1i27Jhg
#3 https://youtu.be/heh0kEB4MGM

Note: at the time, I would use two projectors for the presentation, that's why you see slides side-by-side. Easily done with a one screen / one projector set-up.

You will have to think about what you want to put on your slides — what words (if any), what visualizations. You will have to think about your audience — people will be sitting a good distance away from the screen so your design needs to be clean to be readable from across the room. You will have to craft your message — how do you want to guide the audience through your open data journey? If animating a series of builds will help — do that!

Present your data! DO NOT READ your slides. DO NOT READ notes from a set of cards. Practice your presentation! ← you'll have to do this to make sure it fits in the 5-minute limit!
DO NOT "divide & conquer" — i.e., all of the slides should be consistent vs feeling like a few mini-presentations mushed together.

➔ this is a presentation worth 14% of your grade — make it a professional presentation!

Questions about slide design? Post 'em to the Moodle Discussion Forum!

I will get teams to upload their presentations **the day before** the big day (i.e., your presentation must be in the Moodle dropbox by 23:59 on Dec. 1st). This is so I can get them all organized on the console computer in HOL 190 in advance. Do not show up on Dec. 2nd with a presentation on a USB key or e-mail me some last-minute changes! Whatever version is in the dropbox on the 1st is the version that gets presented.

Questions? Post 'em to the Moodle Discussion Forum.

(iii) utilizes the ideas and approaches from this course

Already mentioned above — we've done analysis, statistics, visualization and many other things — demonstrate your mastery of these techniques in your approach to the open data. Finally…

(iv) reveals something interesting from the data!

What interesting things did your team find from your analysis and modelling of the open data? If we think back to Mobi bikes… the proportion of membership types (X% of 365 memberships, Y% of 24-hour, Z% of VIP…) is less interesting than finding out that one type is more likely to use the bikes for commuting while another is more likely to use the bikes on weekends when it is sunny & warm, etc.

So — not "analysis for analysis sake" but analysis for some sort of insight! Remember — think of your audience as a client, you are presenting to them, so they can make a decision and take some action in light of your analysis & presentation of findings.

Last Task —Reflection: Write a 500-word ***personal reflection*** on the project. What worked / didn't work? Why? What did *you* learn — about data analytics, about working in a team, about yourself, etc.? What can you take from this project into the post-business school world?

There are many resources available on the Internet that describe how to write a reflective paper — I'll leave it to you to research those. I found one source at Southampton Solent University a few years ago. Alas — that link no longer works. The big ideas:

To think and write reflectively you have to:

- Experience something
- Think about what happened
- Learn from the experience

You think reflectively all the time, you probably just don't realise you're doing it.

This notion is expressed in the teaching & learning literature as "Kolb's Experiential Learning Cycle." You have an experience — the research, analysis, writing, and presentation of an open data project. Next step — THINK — about what happened. How do you think your team did? Was everyone contributing? Did the team set goals for the project? If so — were those goals met? If not — why was that the case? What did your team do well? Why do you think that was the case?

Again from the Southampton Solent University website:

As a student, and in the workplace, you will be asked to be reflective. Thinking or reflecting on the world around you, your experiences and actions will help you to develop and improve your skills. Reflection is:

- Self awareness: thinking of yourself, your experiences and your view of the world
- Self improvement: learning from experiences and wanting to improve some area of your life
- Empowerment: putting you in control of making changes and behaving in a different way

So — you have all of the "raw materials" to write this paper — what happened over the past few weeks on this project in this course. Think about it. Put those ideas into words. Words into paragraphs and paragraphs into a proper paper.

Questions? Post 'em to the Moodle Discussion Forum.

Inspiration — here are some ideas that students have done in the past…

- National Collision Database — Accidents in Canada; lots of slicing and dicing data. Are men "better" drivers than women? Impact (no pun) of weather conditions, etc…
- Macro Model of Canadian Real Estate (combined building permits data with employment data with GDP by region data)
- Government funding for research (Tri-Agency — Canadian Institutes of Health Research (CIHR) Natural Sciences and Engineering Research Council (NSERC) Social Sciences and Humanities Research Council (SSHRC)) and university enrollment numbers … the idea being more $$ → more students
- Comparison of GDPs of emerging economies (BRIC nations) vs established (G7) nations (econometrics)
- Canadian household debt (mortgage + non-mortgage) combined with interest rate data; predictor of debt service ratio?
- Precarious employment in Canada — educational attainment / type of work (full time / part time) over time.
- City of Vancouver Bike Racks — vs SkyTrain locations, vs crime stats, vs urban planning (Arbutus corridor), vs public buildings, vs parks, etc.
- New graduates employment opportunities … salary discrepancy by region, gender wage gap, college vs university, combinations of those data (college / university | men / women), etc.
- Liquor sales in Ontario … by beverage type, by channel, vs consumption in other provinces, impact of the craft brewing business on LCBO.
- Vancouver crime stats by neighbourhood; moderated by social infrastructure (e.g., community centre, etc.)
- Animated data visualization of Ontario production vs yields in agriculture as a basis for Community Supported Agriculture.
- Commuting times in Ontario and the impact of population of the community on those times.
- Housing values in California by geography with associated demographic data (race: African Americans, Asian, Latino, etc).
- Police activity in the City of Guelph — by geography, incident call-out, etc. Calls for action for mental health crisis management
- AirBnB in NYC — where to locate; home prices vs rental prices, etc.
- Apple crops by counties in Ontario — soil characteristics, harvest values by apple type, downstream apple use (juice processing, cider, grocery store sales, etc)
- Canada / US cross border travel trends tied into CAD/USD exchange rates…