DATA SCIENCE
SYD DAT 8

Thursday 13th July
Week 8 – *Less* Technical Skills
              (but still *really* important)

1. Non technical skills
        1. Communication
        2. Presentation
        3. Networking
        4. Résumé
        5. Interviews
2. Lab
3. Discussion

COMMUNICATION

COMMUNICATION

"Data science is the practice of turning tools and raw data into something that non-data scientists might care about." - Sandy Ryza

- Who's the audience?

- Clear and concise

- Know the business

- Always ask why

- What's the lever?

- Over-Communication is better than under-communication

- Listen hard

- Break bread

‣ Does it reveal something that couldn't be done in Excel?

‣ No speling errors!

‣ What would happen if your document was forwarded to other people in the organisation?

‣ Are you communicating in the same language?

PRESENTATION

‣ How you present your work will determine if it gets implemented.

‣ If your work isn't implemented then it's worthless

‣ Start with the results and action then dive into how you got there.

‣ Can your audience understand what's in front of them?

‣ What questions do you think your audience will have?

‣ Peer Review

‣ Explain complex concepts in three stages if you have a mixed audience

1. Very general level - something your Mum would understand

2. Slightly technical - something an analyst would understand

3. More technical - something a peer would understand that communicates the key points

NETWORKING

‣ Sydney is an incredibly connected place for Data Science

‣ Meetups are a great way to socialise and find people with common interests.

‣ Conferences are great too (but harder to make lasting connections)

‣ Your workplace may have some great people you may not have realised - talk to people!

RÉSUMÉ

‣ Tailor it to the position you are applying for

‣ Relevant experience

‣ Format

   ‣ Keep it to one page

   ‣ Use past tense

   ‣ Keep any descriptions succinct

   ‣ Avoid colour coding

   ‣ Send it as a pdf (the best way to ensure there are no scaling issues)

‣ Only include a "Interests/Hobbies" section if

‣ (a) you make it specific such that it conveys something about yourself

‣ (b) you have the sense from networking / interactions with the company you're applying to, that they value some of the softer skills.

Proof Points

‣ Relevant thesis

‣ Relevant course(s) taken - and grades if available

‣ Independent projects completed (e.g., Kaggle, self-driven work)

‣ Github profile - especially with the code from the independent projects

‣ Technical blog

‣ Don't lie

‣ Brush up on technology you use before an interview

‣ Highlight the benefits of your analysis (e.g. x% reduction in customer churn which increased revenue by $y)

‣ What were some of the major projects you worked on?

‣ What was the purpose of them?

‣ What did you contribute?

‣ What technologies did you use to complete the project?

INTERVIEWS

- ‣ Problem solving

- ‣ Coding ability

- ‣ Communication
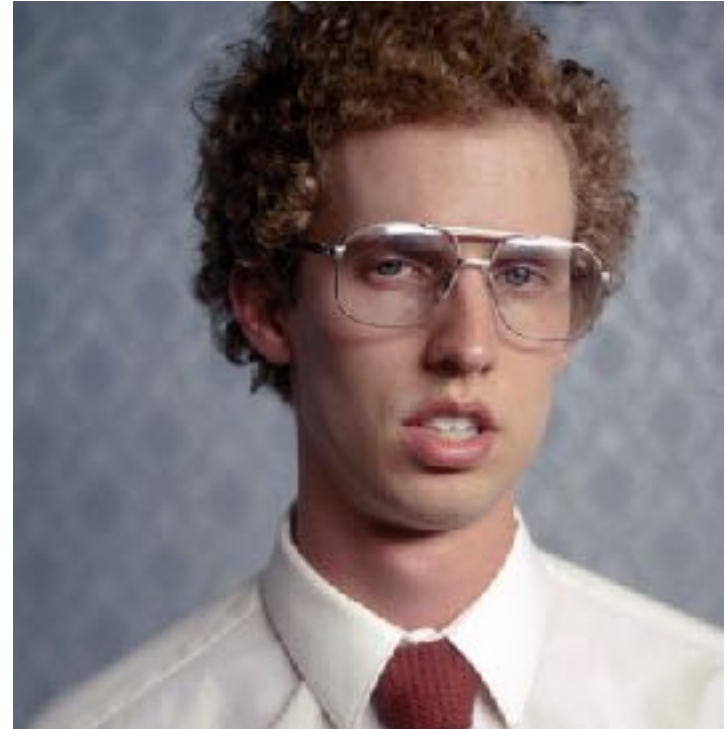
- Technical knowledge
  - Machine Learning
  - Tools (software)
  - Certifications
- Business Knowledge
  - how the world works
- Statistics
- Data Handling

‣ Interest in Technology

‣ Interest in data and what it represents

‣ Good Communicator

‣ Networking

‣ Willing to share ideas

- What is R^2 (for linear regression)

- How do you assess whether to include a variable in a linear regression model?

- How would you assess model accuracy?

- Explain what regularization is and why it is useful.

- Explain what resampling methods are and why they are useful. Also explain their limitations.

- Give an example of how you would use experimental design to answer a question about user behaviour.

‣ What is a recommendation engine? How does it work?

‣ Which tools do you use for visualization? What do you think of Tableau? R? SAS? (for graphs). How to efficiently represent 5 dimension in a chart?

‣ Are you familiar with pricing optimization, price elasticity, inventory management, competitive intelligence? Give examples.

‣ How can you prove that one improvement you've brought to an algorithm is really an improvement over not doing anything?

1. What was the last thing that you made for fun?

2. What's your favourite algorithm? Can you explain it to me?

3. Tell me about a data project you've done that was successful. How did you add unique value?

4. Tell me about something that failed. What would you change if you had to do it over again? ...

5. You clearly know a bit about our data and our work. When you look around, what's the first thing that comes to mind as "why haven't you done X"?! ...

THE
DATA SCIENCE HANDBOOK
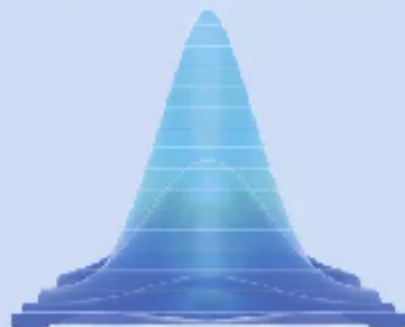
ADVICE AND INSIGHTS FROM
25 AMAZING DATA SCIENTISTS

FOREWORD BY JAKE KLAMKA

DJ Patil, Hilary Mason, Pete Skomoroch, Riley Newman, Jonathan Goldman, Michael Hochster, George Roumeliotis, Kevin Novak, Jace Kohlmeier, Chris Moody, Erich Owens, Luis Sanchez, Eithon Cadag, Sean Gourley, Chris Corthell, Clare Wu, Joe Blitzstein, Josh Wills, Bradley Voytek, Michelangelo D'Agostino, Mike Dewar, Kunal Punera, William Chen, John Foreman, Drew Conway

BY CARL SHAN HENRY WANG WILLIAM CHEN MAX SONG

DATA SCIENCE PART TIME COURSE

LAB

CAUSALITY

‣ **What is Causality?**

‣ **How is it different to Correlation?**

‣ **What are some ways to detect Causation?**

HOMEWORK

Homework

▸ Homework3

Reading

▸ https://linkurio.us/panama-papers-how-linkurious-enables-icij-to-investigate-the-massive-mossack-fonseca-leaks/