The Effective Analyst

Lessons we’ve learned

Deirdre O’Brien
Amir Najmi
2011/06/27
Our qualifications

- Two oldest Metricians (after mmm)
  - A decade between us
  - How many is that in dog years?
- Learned most things the hard way
- Mentored many analysts
The Analyst role in AQ

• Ensure that AQ makes best possible product decisions
  • Support decision under uncertainty
  • Highly leveraged
• Owns the semantics and significance of signals and metrics
  • SWEs build awesome systems
  • Analysts worry about the quality of the data that goes through them
Typical analysis pipeline

- Pull data from one or more sources
  - Sawzall, SQL
- Glue the data together, aggregate
  - Sawzall, python
- Explore, fit and plot
  - R
Data sources

- RASTA query
- Dremel
- Ads DB
- Logs
  - Adquery, Sessions
Look at your data ASAP

• Even if you know it isn’t perfect
  o (It may be totally junk)

• Example: demo-days
  o Two things needed to agree
Sanity checking

- Ads Nav
- RASTA Experiment 42
- Develop intuition about our data!
Know when to punt

- Occasionally, there are problems with logging or external libraries
- Resist the temptation to solve it all
  - Do enough to figure out whom to punt to
- Example: RHS Ads analysis
  - Clear problems with data (likely images)
Workaround

• If possible, avoid known problems by looking at alternative data
  o Else, be sure to note caveats in reports

• Example: RHS Ads Analysis
  o Looked at google.com experiment
Don’t thrash. Really!

- Don’t be stuck for more than 20 minutes
- Do due diligence
  - Don’t overdo
  - Like look for and read documentation
    (search email)
- Seek expert advice
  - You’ll be on the other side soon enough
Ads Metrics is an IM culture

- Faster, less invasive than email
- Anyone online at anytime is fair game
  - They may not have time
- Keep it concise
  - yt, qq, nm
- Reciprocate by having chat on
  - Go red if you have to
Mailing lists

• Ads-metrics-core
  o Please join the discussion
  o Mail filter for code-reviews!
• Ads-quality-team
  o Big list, only important information
• Sawmill-announce
• Rasta-announce
• Many other mailing lists useful for archiving
Expect to iterate, evolve

• You actually don’t know what you are doing until after you have done it
• Start with end-to-end immediately
  o Rapid, minimal, hacky
• Look at the data, iterate, evolve

• Example by Wael
Do not start by launching an amphibious assault on Normandy

...especially if you’ve never been there before
Quality without perfection

- We do high impact, high quality work
  - Timely yet thoughtful
- Quality through iteration
  - Surprising results are usually wrong
  - Peer review is crucial
- We stop when the purpose is fulfilled
- Perfectionism = failure to launch
Credibility is all we have (and all we need)

• We give the unvarnished truth
• Own up to mistakes proactively
  o Feels awful but actually earns you respect
• Admit ignorance
  o Radical for professional statisticians
• Respect levels of sharing
  o Ads-metrics-core vs. one-pagers
Engaging across Ads-Metrics

• Lots of different projects
• Once starter project is complete, you are an expert
  o Share that knowledge!
  o Don't wait to be invited
• Forums:
  o Ads metrics meetings
  o ads-metrics-core mailing list
  o Office chats
  o White board discussions
Distractions

• The data
  • Data itself is infinitely rich, distracting
  • Must approach it with a question
• The system
  • Don’t need to know every system detail
  • Selectively watch deep-dive videos
• Lesser distractions
  o Math, engineering, R … typing speed

• You can find God in all things
We’re all about impact

• Hard to imagine impact on Google
• What are the big questions
  o For my team
  o For Ads Metrics
• How will what I find be used

• Scope vs. impact
Reminder: how you are judged

• Scope
• Impact
• Complexity
• Leadership
Summary

- Analysts care about serving decisions
- We’re about significance and semantics of information
- Keep focus on the marginal value of your effort
- Believe in evolution
- Only do that which only you can do
The End
General principles

• Engineering is a means to an end
  o Temptation to build tools and abstractions
  o Programming is like typing
• Our data is always complex, imperfect
  o Data/code changing constantly
• Keep focus on the marginal value of your effort

• Only do that which only you can do