INSEAD
The Business School for the World®

MBA Programme
Period 3 – Jan/Feb 2016

Data Analytics for Businesses

Theos Evgeniou

Some Lessons from Sessions 7-8 (February 20, 2016, INSEAD)

Some of the lessons discussed (and more) are in the readings for session 8 on the course website.

How do I know I have enough data?

- You never have all the data. This implies you should be extra careful with overconfidence "just because some data say something". Beware of "black swans", "blind spots", etc.
- Be ready to spend > 60% of your time on data gathering and data curation/cleansing
- At the end you need to have a **process** to keep gathering, integrating, and improving data quality.
- It is not about volume of data (not "Big Data") but about diversity of data ("Diverse Data")

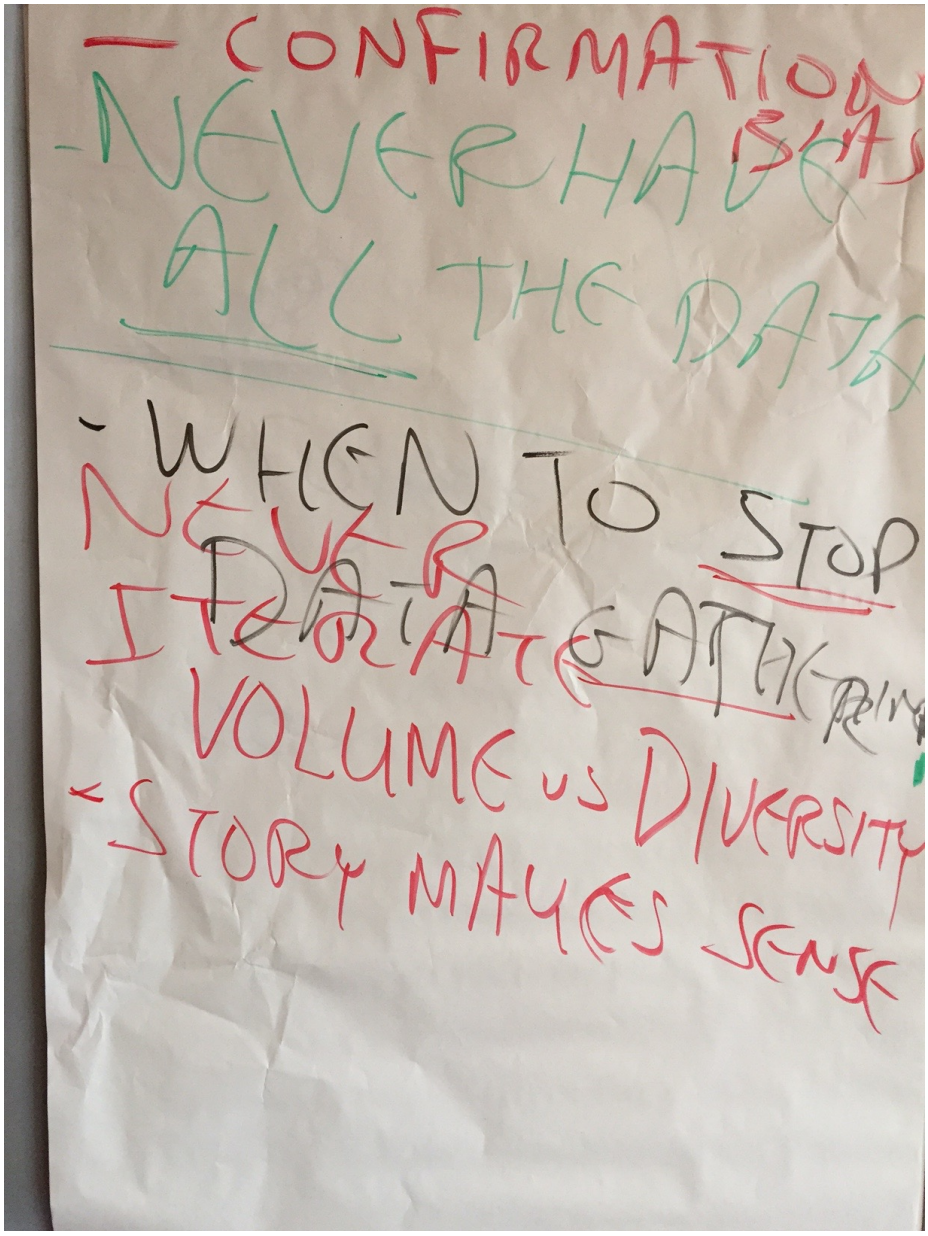
How do I know I am done with my data analytics/science project?

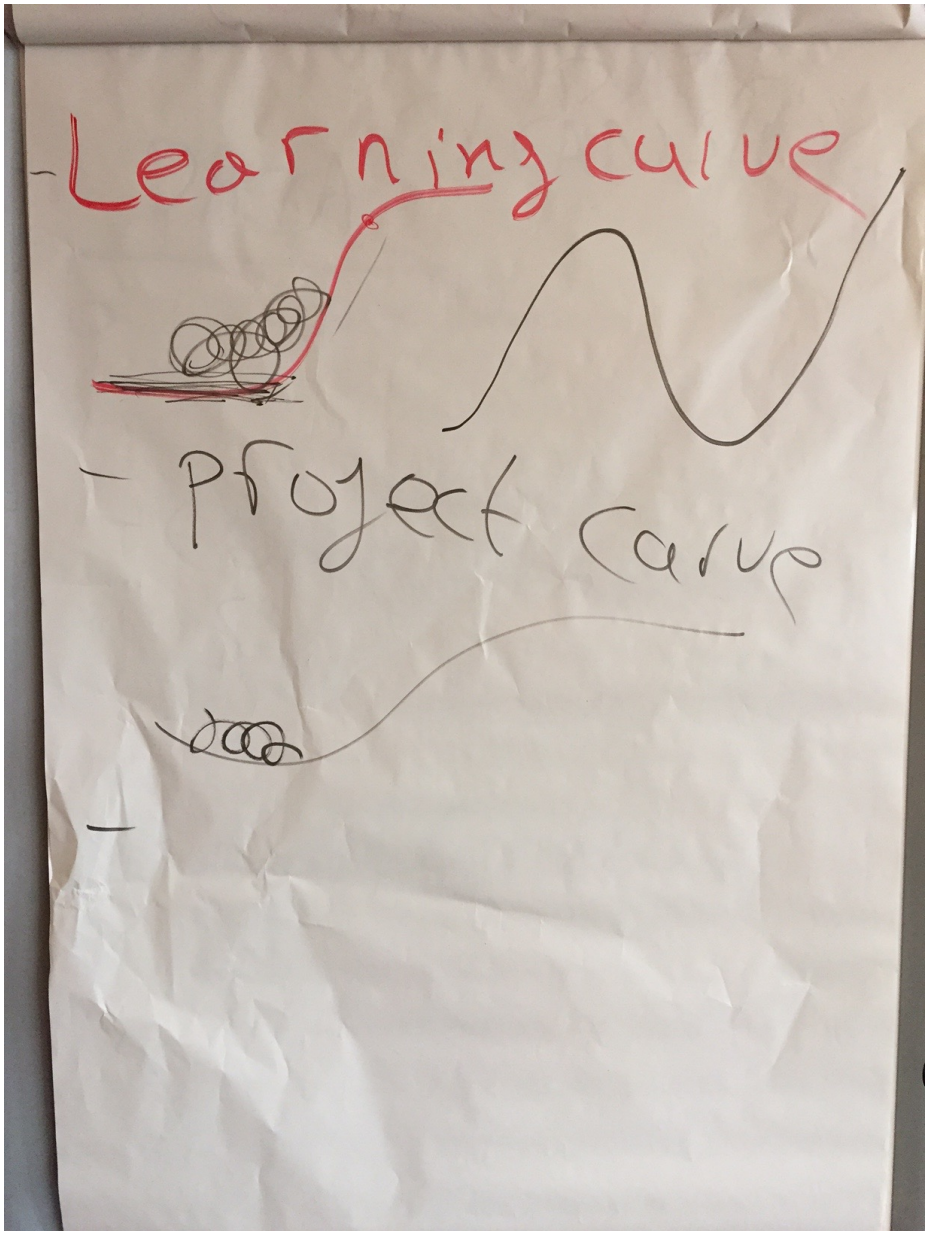
- You are **never** done with a project, as the deliverable it is **not a project but a process**
- You need to end with a process (e.g. how new data will be integrated in the future, how to manage the data quality/update, how to maintain all software, how to reuse components, documentation, etc) that you can use any time in the future
- An indication that you "are done for now" is that you have some clear "story" about your business findings that makes sense and is robust to different analyses/viewpoints/data – and ideally is not the same as what you had at the beginning, although it may be (if the data say so, too).
- It is all about setting up an **iterative learning process**: start from not knowing much, and iteratively learning about the business through the data. The project curve is very slow at the beginning.
- Make sure you "**protect the plant":** It takes time to go past the initial slow "learning/project curve". Be patient with the team, give time. Analysis results some mostly at the very end, the first stage is slow with no findings (e.g. largely focusing on data gathering etc).

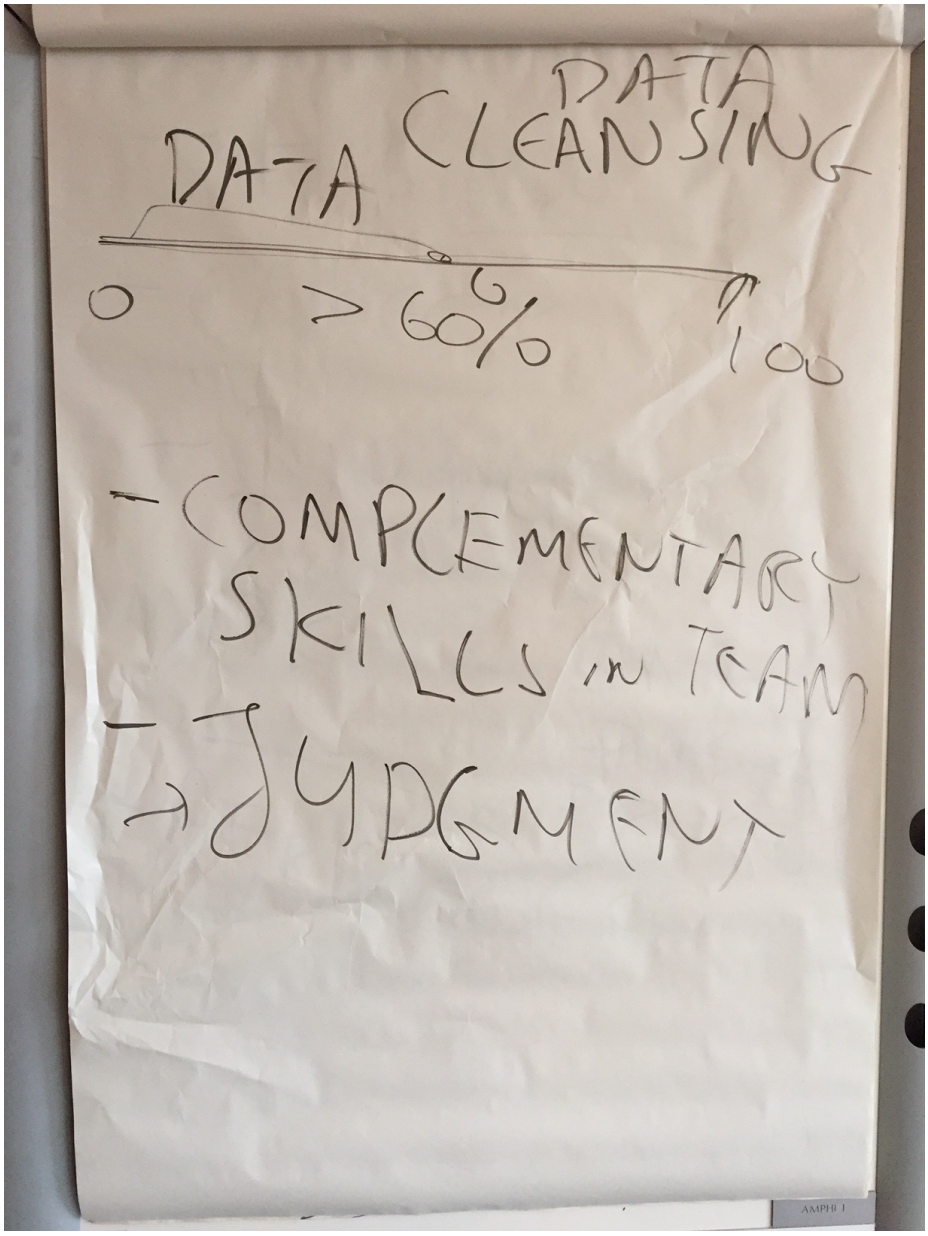
How do I hire data scientists? What do I look for?

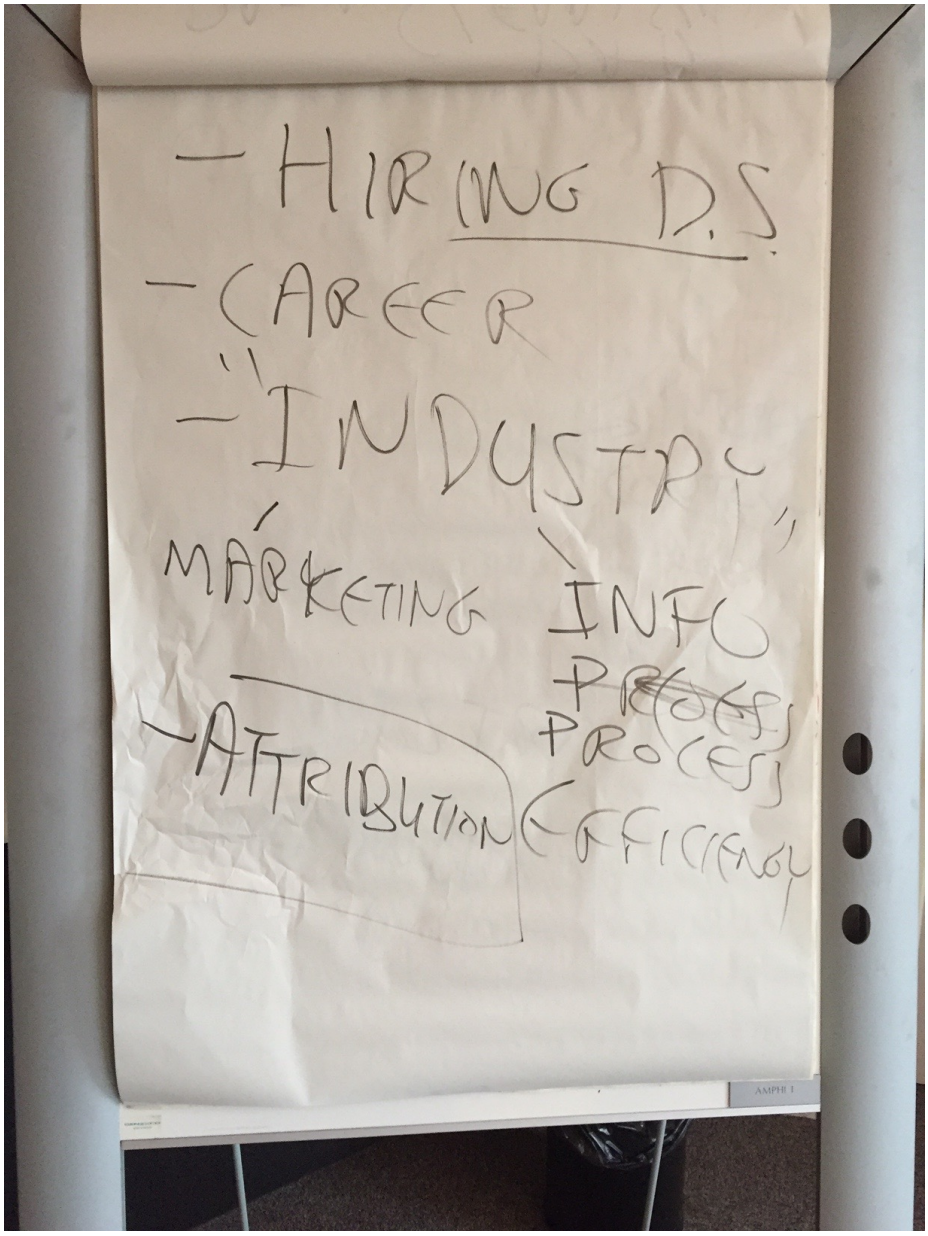
- Technical skills:
    - Coding: ask for sample code, github account, etc; coding case study with business context.

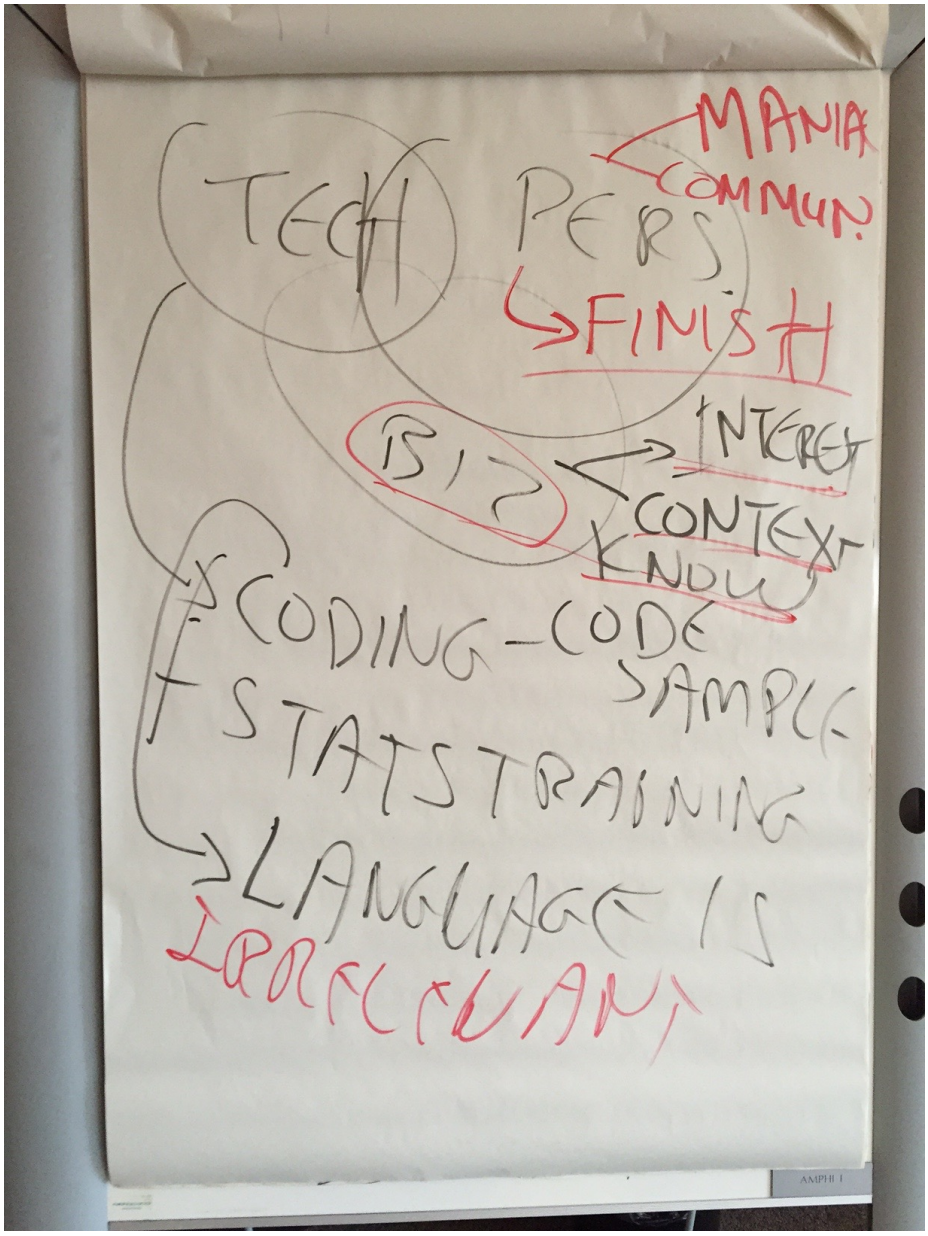
    - Some scientific/engineering thinking (not necessarily training) to be “evidence based” and be familiar and/or able to use basic statistical methods (e.g. from simple regression to the topics we covered in class)
- Business skills:
    - Absolutely necessary to be interested in the business/context. Data science is about creativity and that come only if the people are interested in the context;
    - Some understanding of the industry/business issues already – if possible.
- Personality:
    - Totally obsessed with learning (from the data/business/others) about the problems at hand. “Maniac”.
    - Curiosity to learn (about the business context/problem)
    - Obsessed with quality (e.g. data quality, etc)
    - Team player
    - Communication skills


- CONFIRMATION BIAS
- NEVER HAVE ALL THE DATA
- WHEN TO STOP
NEVER
ITERATE
VOLUME vs DIVERSITY
- STORY MAKES SENSE


- Learning curve
- Project curve


DATA
DATA CLEANSING
0
> 60%
100
- COMPLEMENTARY SKILLS IN TEAM
- → JUDGMENT


- HIRING D.S
- CAREER
- INDUSTRY
MARKETING
INFO
PROCESS
ATTRIBUTION
EFFICIENCY


MANIA
COMMUN
TECH
PERS
FINISH
BIZ
INTERET
CONTEXT
KNOW
CODING - CODE
SAMPLE
STATS TRAINING
LANGUAGE IS
IRRELEVANT