Preparation week 2 Data science

1. <https://www.safaribooksonline.com/library/view/creating-a-data-driven/9781491916902/ch01.html> - **What do we mean by data-driven?**

* Data-drivenness is about building tools, abilities, and, most crucially, a culture that acts on data
* Prereq: An organization must be collecting data.
  + Data undoubtedly is a key ingredient. Of course, it can’t just be any data; it has to be the right data. The dataset has to be relevant to the question at hand. It also has to be timely, accurate, clean, unbiased; and perhaps most importantly, it has to be trustworthy.
  + This is not always easy. It is easy to bias conclusions and cleaning data is a long and tiresome process.
  + Even if you have quality data, it will only get you so far. It does not make you data driven. A small amount of usefull clean data is more worth that giants amounts of junk data.
* Data must be accessible and queryable
  + Having accurate, timely and relevant data is not sufficient to count yourself as being data driven. The data must be:
    - **Joinable -** The data must be in a form that can be joined to other enterprise data when necessary.
    - **Shareable –** There must be means to share the data within the organization
    - **Queryable -**  there must be appropriate tools to query and search for the right data in a big database.
  + Now the data is accessible, but not sufficient enough yet. You need people with the right skills to use that data. That can mean the mechanics of filtering and aggregating data, such as through a query language or Excel macros, but it also means people who design and choose the appropriate metrics to extract and track.
  + **So, for an organization to be data-driven, there have to be humans in the loop, humans who ask the right questions of the data, humans who have the skills to extract the right data and metrics, and humans who use that data to inform next steps. In short, data alone is not going to save your organization.**
* **Reporting**
  + Data also needs to be presented in the right way. You could have a lot of valid and accurate data, but if you can’t present it in a good way, the information won’t mean anything.
* **Alerting**
  + Alerts are essentially reports about what is happening right now. They typically provide very specific data with well-designed metrics.
* **Hallmarks of data-drivenness**