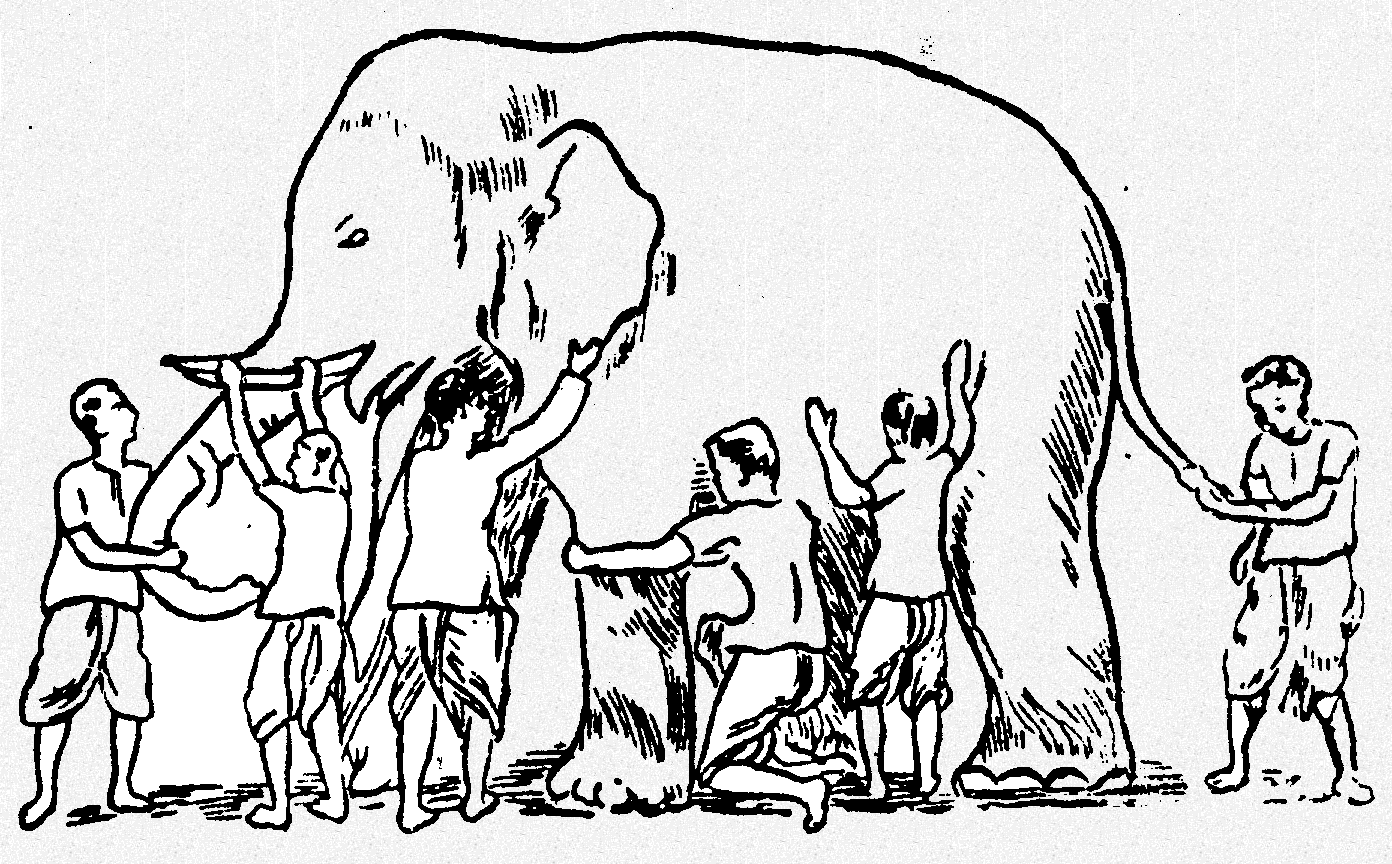
My First Data Science Project



**What are your motivations to do a Data Science Project?**

Improved job satisfaction – actuary.

Become more efficient at –

* find better questions to ask data
* determining the ROI on business actions
* Improve accuracy on current processes
* Respond quickly to market changes
* Find correlations between events.
* Knowing “about” data science and “knowing” datascience.
* Improve technical chops.
* Curiosity.
  + Tell the difference between cats and dogs

**How to identify a problem that can be solved by a Data Science project?**

* Has a lot of diverse data.
* Problems that were left unsolved/solved unsatisfactorily for the lack of computing power/tools.
* Problem has to lead to a solution or a course of action/decision/product.
* Identifying data variables that have influence on decision making.
* Talk to people. Investigate/query the people “Ask the right questions to the right people”.
* Start from the data and identify interesting questions from that. . “what questions can we answer, given this data”.
* Convert qualitative values to quantitative values. (eg: whiskey reviews to star ratings”
* Tell a Compelling Story from data
* Validate the “truth” of existing BI/reporting systems.

**A Data Science project may have one of more this:**

1. Integrate data sources that were ignored before
2. Will use one or more statistical methods.

**What is not a Data Science Project?**

1. If there is no quantifiable aspect to the data.
   1. There is not enough data to have enough confidence in your results
2. A (set of) SQL query alone.
3. There is no statistics beyond aggregation. (BI)
4. Does not question the credibility and generalizability of the results. Concerns solely with reporting existing data.
5. Uses ONLY prepackaged, single-purpose package.

**Who are going to be involved in your Data Science Project?**

**What will you need to start a data science project?**

**What does your toolkit look like?**

**What will you have to show at the end of your project?**

Related topics to discuss in future meetups:

* Trends – why DS?
* Applications in various industries
* Using Python vs R? why not together?
* What is a good data scientist?
* What is the minimum level of competence required in:
  + Mathematics
  + Statistics
  + Programming
  + Visualizations
  + Project management – setting expectations
  + Product lifecycle
* Resources
  + Classes
  + Datasets
  + Tools
  + Books
  + Competitions