1. Audience: Companies Interested in Implementing Big Data technologies for Data Science projects.
2. Central question:

I am trying to address which Big Data Technologies have more users and are performing well.

1. Question breakdown:

* Which Big Data technologies are catching up with the industry
* What technologies are more popular among users
* Which technology/Programming language has more users?

1. Data:   Twitter

Stack over flow

Github

Use twitter API to collect the tweets about the technologies and programming languages like Mongo DB, Cassandra, Python etc. In Big Data Echo system and compare how each technology is doing in terms of number of tweets, does the tweets are more about performance or general tweets

* Number of projects implemented in a programming language.

1. Analysis:   Explore
2. Start with logistic regression to see if the point streak relates to Stanley Cup outcome (1,0). If this is not interesting, do either linear regression or categorical regression with playoff outcome (numerical values for no playoffs, eliminated in first round Stanley cup loss, Stanley cup victory). If this is not interesting, ... not sure.
3. Deliverable:   d3 dashboard.

Challenges:   if nothing comes out statistically significant, I hope people like playing with the dashboard controls and seeing the outcomes change. Maybe I could look at additional statistics like strength of schedule or injuries to chase after significance a bit more, but if it is not close (sample size will be very small) I'll just focus on making the dashboard fun to play with.