# Prelim project ideas 

Nisheeth

## Trains analysis

- I've shown you how to grab the past year's delay data for any train from irctclive.in
- A nice tutorial for web scraping lives here http://web.stanford.edu/~zlotnick/TextAsData/ Web_Scraping_with_Beautiful_Soup.html
- What more could you do?



## Possible extensions

- Find best train between two stations (***)
- Quantify the confidence in your judgment
- Incorporate time of day into your calculation
- Find the geographical distribution of delay bias across trains (*****)
- Is it topological?
- Is it structural?
- Does it correlate with accidents?
- Rank order trains by their timeliness, compare with fares (*** $1 / 2$ )
- Identify outliers
- Suggest quantitative corrections


## Cricinfo analysis

- I've shown you how to
- Grab data for all Test matches played in the past 25 years
- Grab data for all batting scores from these matches
- Run a basic analysis figuring out how exceptional a particular batting performance was, statistically speaking
- Run a basic analysis measuring how exceptional batting performances connect with matches won


## Possible extensions

- Quantify exceptional bowling performances
- Run an analysis connecting exceptional bowling with match outcomes
- Compare with batting; which is more decisive
- Find which player's performance lines up best with match outcomes
- Treat this as a leave-one-out classification problem?
- Which player's failure influences a team's performance the most over their career
- Build a predictive real-time model of match outcome
- Analyze matches at the session-level
- What constitutes winning a session?
- Can you quantify this winning?

