Project 3: Classification

Week 4 - Week 6.5

Back story:

Using data from the web (API or scraped) or one of the optional supplied data sets (possibly in conjunction with your own data), create models using supervised learning techniques. OPTIONAL: Extend your findings with a flask website and/or Tableau visualization.

Note you can work as a 'group' (with other folks working on the same data source as you) for brainstorming, design, additional data, etc. However, the final projects will be individual.

Data:

- **acquisition**: download, api's, scraping, etc.
- **storage**: PostgreSQL or do SQL challenge 1, 2 and 3 (check with your instructors)
- **example sources**: (see sql_data_sets for descriptions and links to data.)

- NYC Transportation
- Traffic Fatalities
- US Health Insurance Marketplace
- Militarized Interstate Disputes
- Sports, Sports, and More Sports!
- The Simpson's
- Climate Change
- Python StackOverflow Questions

Required Skills & Tools

- supervised learning
- SQL: either storage of your data or SQL challenge 1, 2 and 3
- Optional: flask and/or Tableau

Deliverable/communication:

- organized project repository
- slide presentation
- visual and oral communication in presentations
- write-up of process and results
- 4-min presentation
- Project Logistics

Design:

- iterative design process
- "MVP"s and building outward

More information:

Data sources for this project are all about options. We can choose from a number of pre-selected data sets. We can also use our own data (either scraped from the web or pulled from api's) or supplement the pre-selected data with some of our own. Either way, we will be honing our database skills by storing data in PostgreSQL and doing some of our analysis there.[^1]

Example Projects:

- Predicting Telecom Churn
- Baseball Pitch Recommendation
- Major or Minor? Classifying the Mode of a Song
- Ads That Click