Predicting NFL QB Fantasy Performances

DAT 8

Victor Yang

Why we care

- Fantasy football is a multibillion dollar industry and began back in the 1960s
- Recent surge in funding and advertisements due to advent of weekly leagues
- Similar to online poker in its early days, not just luck
- Can we model QB fantasy points using historical data and Las Vegas odds?
  - Bonus question: Can we predict better than Yahoo?

How a QB is scored in fantasy football

- QBs tend to be one of the more consistent positions in football from year to year
- $FP = .04 * PaYd + 4 * PaTd + .1 * RuYd + 6 * RuTd - 2 * FL - 2 * Int$
- Other features or metrics to look out for:
  - Opposing defensive strength
    - Yards allowed, TDs allowed
  - Las Vegas spread
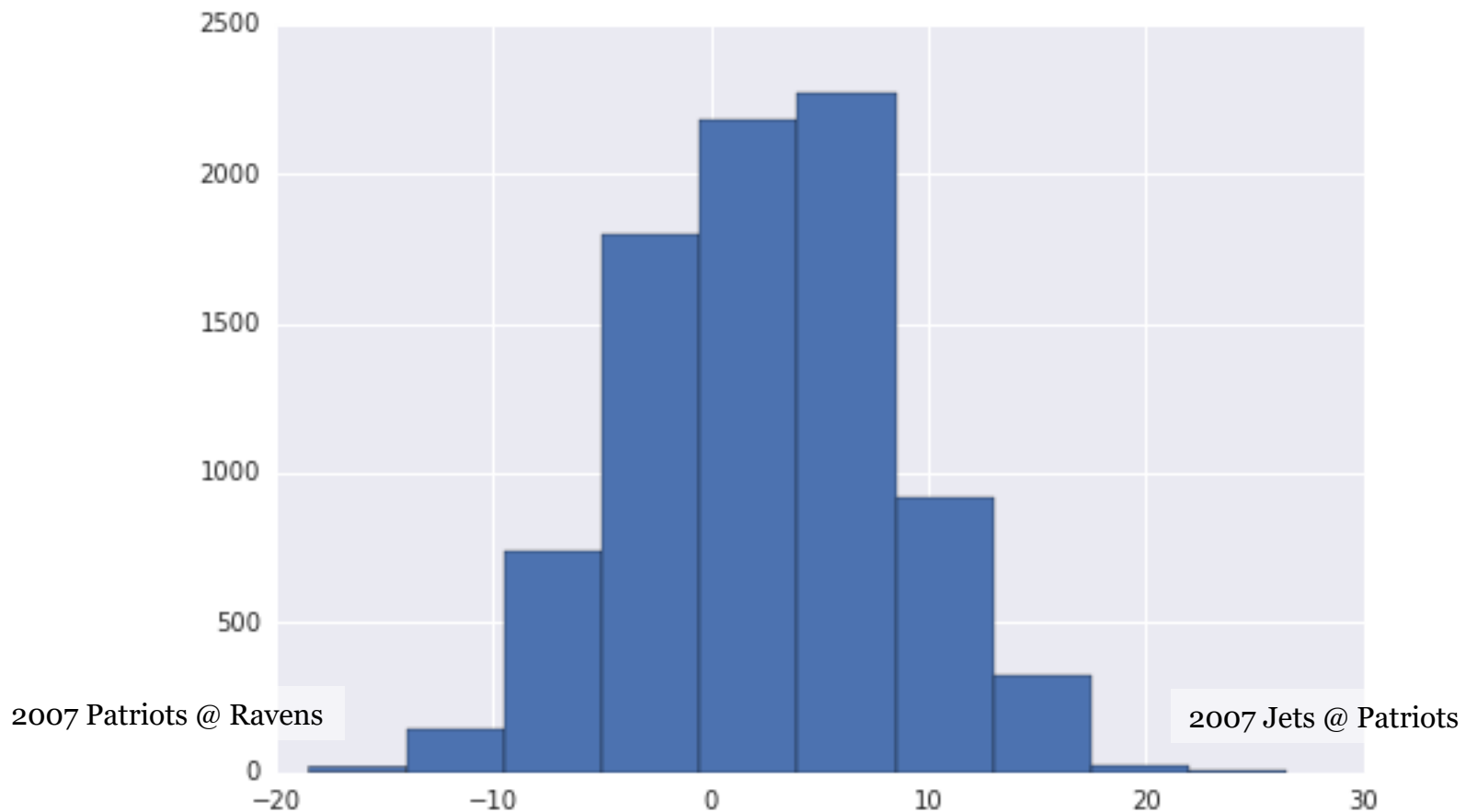 - Extras: Age, season, 538.com's team ELO

Data sources

- Python packages called *nflgame* and *nfldb*
  - APIs that retrieve and parse JSON data from NFL.com's GameCenter (only goes back to 2009)
- FootballLOCKS.com
  - Sells "expert picks" for bettors claiming it can beat Vegas's spread (goes back to 2006)
- Repole.com
  - A professional handicapper's personal website that keeps track of historical lines and other stats (goes back to 1978)
- 16 weeks of 14-16 games between 2 starting QBs over last 5 seasons ~ 2400 observations
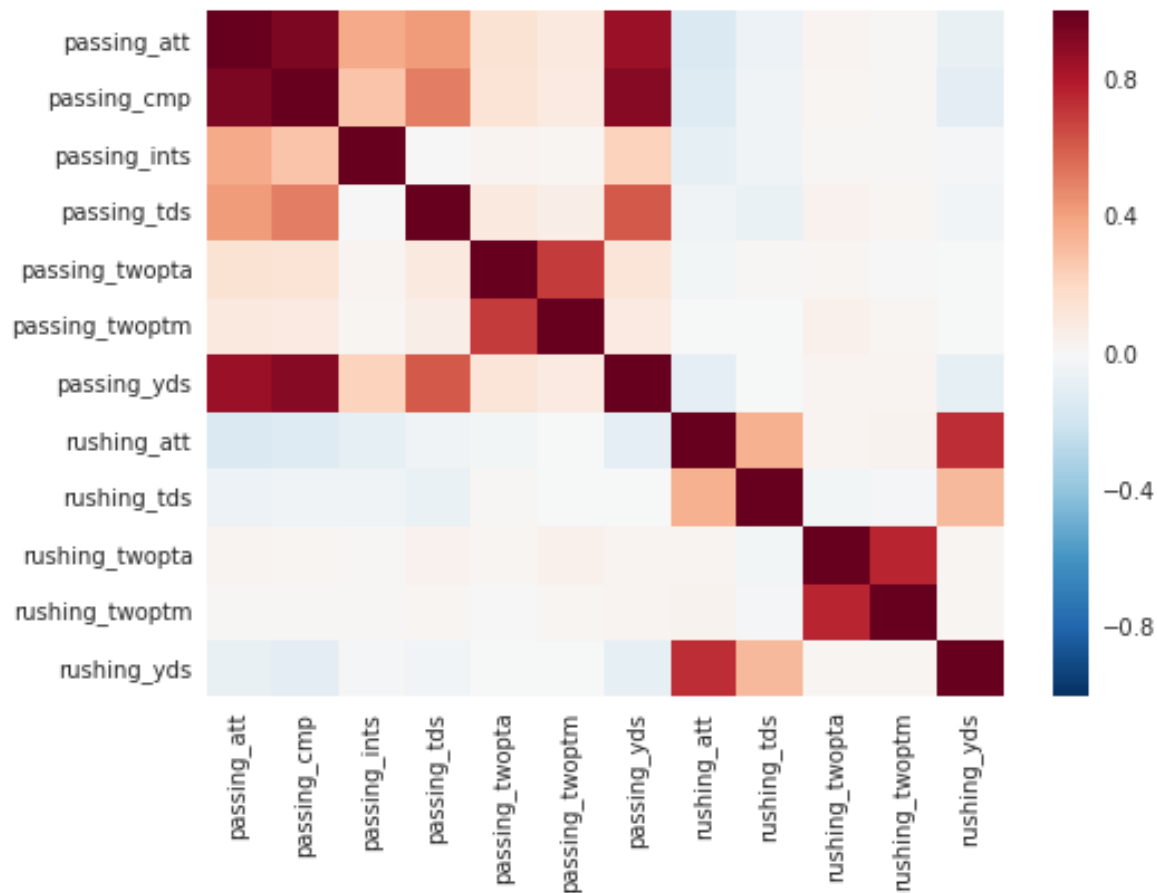 - Would like to stretch this further if possible

The data and bumps so far

- Team stats scraped
  - Formatting changed between seasons
  - Most series/columns need cleaning
- Historic Vegas lines downloaded
  - CSV's were missing comma at end
  - Bash scripting and sed fixed this relatively painlessly
- Game-level player stats collected
  - Documentation lacking in certain areas
  - Reddit and IRC served as excellent resources

Distribution of Vegas spreads since 1978

2007 Patriots @ Ravens

2007 Jets @ Patriots

QB Stat Correlations

Next steps

- Data cleaning
  - Way more "QBs" than expected
  - Team stats not converting to numeric as expected
- Create common key to merge datasets and begin feature engineering
- Model selection
  - Basic model will likely be a rolling average of individual QB stats