

PRO CYCLING STATS

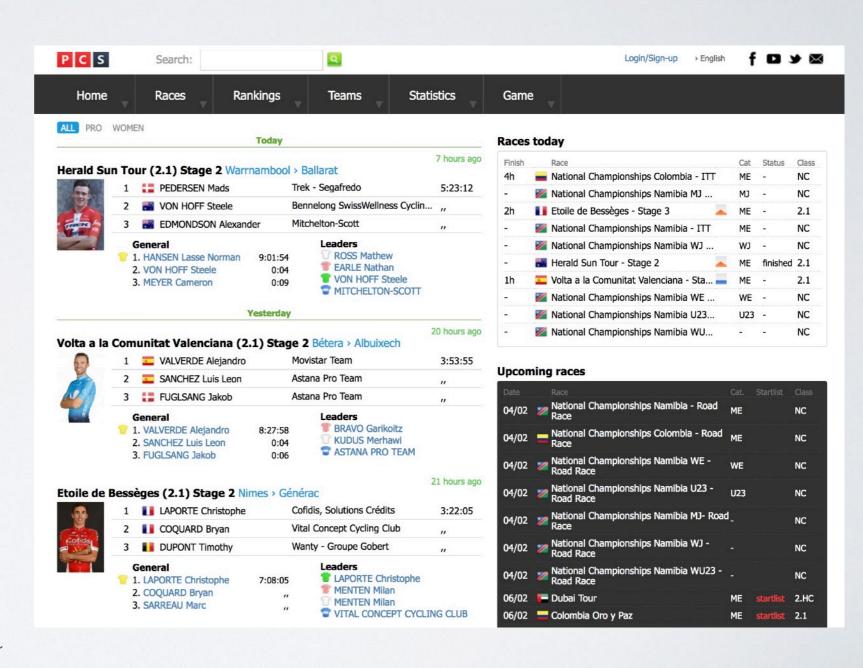
The 'single most useful cycling database'

Why use PCS?

- 140M pageviews/month,
 50k users/day
- cycling writers, TV
 commentators, pro teams,
 riders, fans
- Team Sky: 'marginal gains'

Sports Analytics Innovation Summit, San Francisco, Nov.

 comparing physiological performance and sensor data

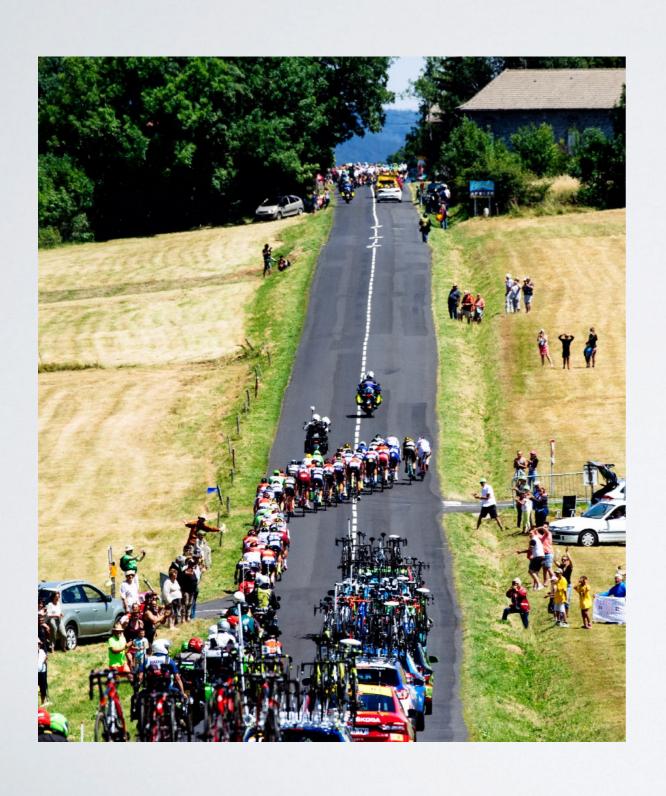


The Challenge: PCS Points



14 races, 15 years → 17,800 pages → 1,730 riders / 10 seasons each

Methodology & Results



Linear Regression

- OLS with 1, 7, and 74 features
 - points per season, career points by category, career length, age by season, nationality
- Lasso selected 21 features
 - Training R-squared (adj):

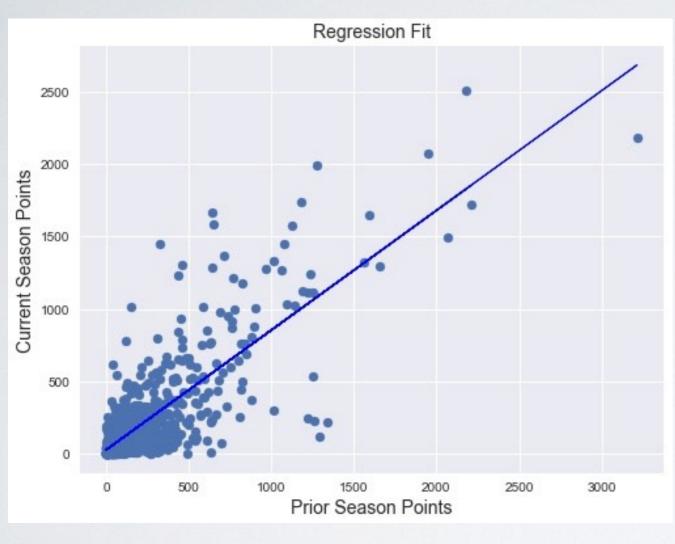
 $0.64 \rightarrow 0.70 \rightarrow 0.75$

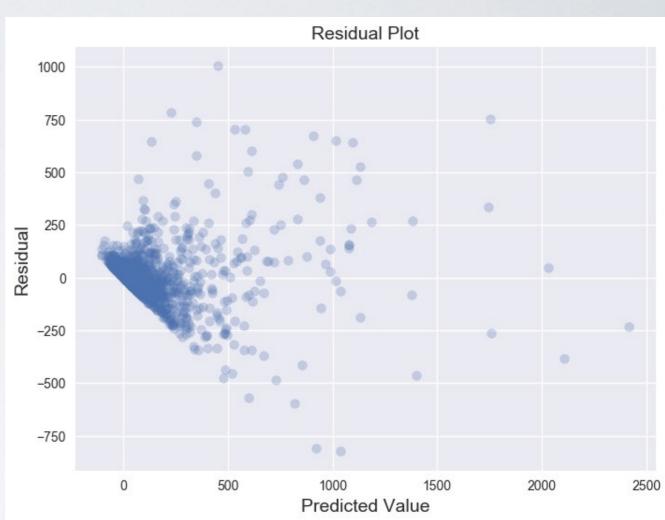
• Test R-squared: 0.68

But is it useful?

• Test RMSE: 137 points

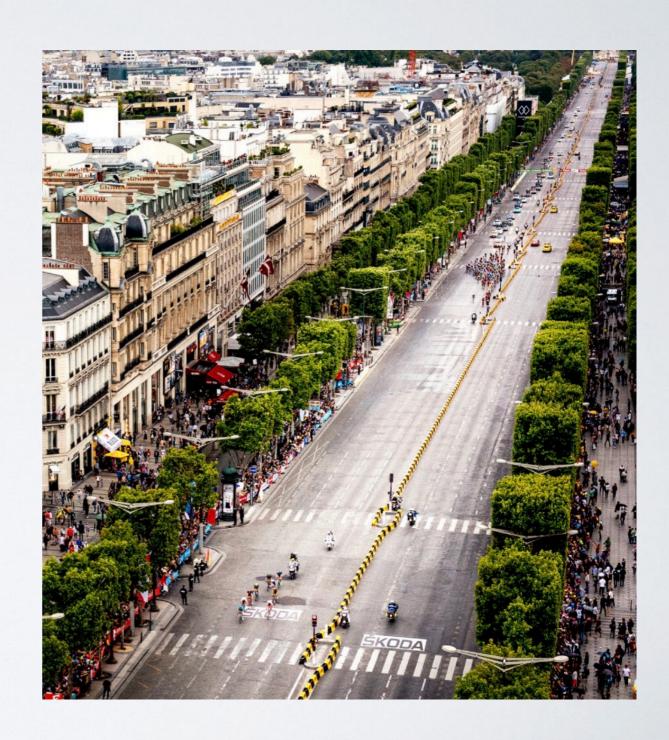
What could be missing?





Opportunities Ahead

- Classification vs. Regression: domestique, sprinter, General Classification, climber, and time trial
- Watts/kilogram
- Dopeology: people, teams, incidents and products
- Team performance and teammate interaction
- UCI vs. PCS points





Time to get moving!

Credits

Number Crunchers: Pro Cycling Stats, Rouleur, https://rouleur.cc/editorial/ number-crunchers-pro-cycling-stats/

The Most Powerful Man In Cycling Data?, Innovation Enterprise, https://channels.theinnovationenterprise.com/articles/how-data-has-changed-cycling-performance-sports

Best of the 2017 Tour de France, Strava, https://blog.strava.com/galleries/ tour-de-france-2017/

Photos by Jered & Ashley Gruber, https://www.gruberimages.pro/