Strava
Fraida Levilev
SF_DAT_15
8/26/15

Strava is a website/app for people to track athletic activities. It is really popular among cyclists. It can get competitive. An activity is broken into segments and each segment has a leaderboard. The person who records the fastest time on a segment becomes the KOM/QOM (King/Queen of the Mountain). People put a lot of effort into getting PRs on big segments and possibly becoming the KOM. Being an active strava user I thought it would be an interesting data set to work with.

There is a website called roadresults.com where they look at the people registered for upcoming races and they're results in past races and try to predict who will win. I wanted to do something similar with my project where I say this is how someone did on this type of segment and see if I can predict how they will do on a similar segment.

Most of my time was spent collecting the data. Strava has a pretty good API with access to a lot of information. I wrote some functions to pull that data and then put it all into a CSV. Once that was done I was able to look at the data and see what I could find.

I started with a linear regression. The results weren't surprising. It told me if someone used more power on their ride they were faster; if the average grade of the segment was higher, the athlete was likely slower than usual. As I added more features, such as the category of a segment or the actual segments themselves, the model became more accurate. Eventually I got the error down to 13%.

There is still a lot I would like to do with it to make it better. Gender and weight play huge factors in how people perform. A large chunk of the athletes I pulled did not have gender on their account. I want to try to do a tree to see if I can predict whether they are male or female and then work with that. The issue I had with weight was that although the API says it's accessible, it is private on the athlete accounts so my weight is the only one that showed in my data. My friends told me I could work around it by looking at someone's w/kg and getting the data that way, so that is something I could look into.

I'm really excited about this project and definitely want to work further on it. Eventually I will get a bunch of people together for a ride and then see how my predictions held up.