

Capstone Project 2 – Datasets

1. [Pitchfork Reviews](#) – dataset holding 18,393 reviews and assorted metadata (including full review text). Pitchfork, to certain people including me, is the best purveyor of music reviews around. Could be interesting to do some text processing and see what sorts of music the site gravitates towards.
2. [Austin Crime Statistics](#) – dataset holding 159,000 crime reports. Would be useful for time series and spatial analysis, both of which I don't have a ton of experience with.
3. [Spotify Song Attributes](#) – I could pull my own profile's data and create a recommender that knows what kind of songs I like. I could then test different songs and see if the model recommends me songs I actually end up liking.
4. [League of Legends Dataset](#) – What appears to be complete data from the last 2 years of professional games. Potentially interesting areas for analysis could be how gold levels fluctuate over a game, and how exactly a team can make a comeback (which stats are most likely to lead to a comeback).