

# STAT 418 – Final Project Proposal

## Steam Player Reviews

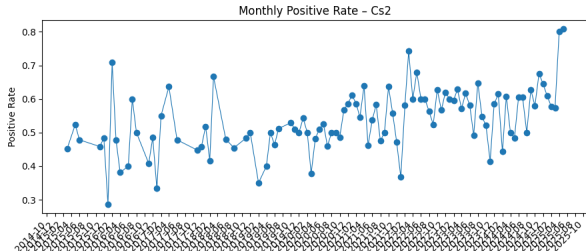
Yanze Guo

Week 6

- **Source:** Steam /appreviews JSON endpoint  
(no API key, cursor paging, 100 reviews/request)
- **Collected:** 10k reviews    Language: english    Period: last 12 months
- **Games:** Dota2, CS2, PUBG, and Stardew Valley
- **Column names:** *review\_id, Appid, voted\_up, timestamp, language, playtime, review*
- **Storage:** each game is stored in its own SQL database



- Overall positive rate: **72.96 %**
- Top tokens:  
POS: best, love, fun, team  
NEG: cheater, hacker, toxic, lag



- **Problem:** Steam “Recommended / Not Recommended” votes often disagree with the actual review text, so the official positive rate can be misleading.
- **Solution:** Train an NLP sentiment model, compute a positive text-based rate, and show it next to the vote-based rate in the Shiny app.