THIS ALE'S FOR YOU:

An exploration of the DMV Beer Scene and Beer recommender system

Will Karnasiewicz

May 2019

What are we trying to learn?

- Explore the DC/MD/VA beer landscape (via RateBeer)
- Build beer recommender models for DMV beer universe:
  - *Content: based on styles and other attributes*
  - *Content: based on review text data*
  - *Collaborative: based on RateBeer user data*
  - *Content: Doc2Vec? (refreshing beer please)*
- Can we classify good beers? Anything else interesting?

What am I going to tell you?

- Quick introduction of data gathering process
- Explore the DC/MD/VA beer landscape (via RateBeer)
- Recommender system methodology
- Demo of recommenders
- Future improvements
- Challenges and key learnings

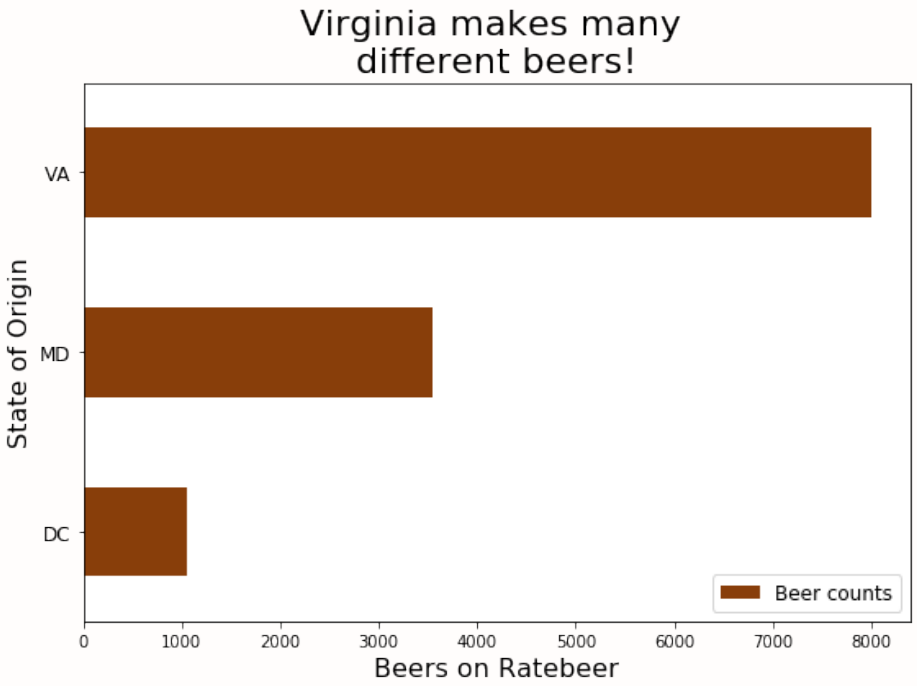
Data Gathering:
Process and Lessons

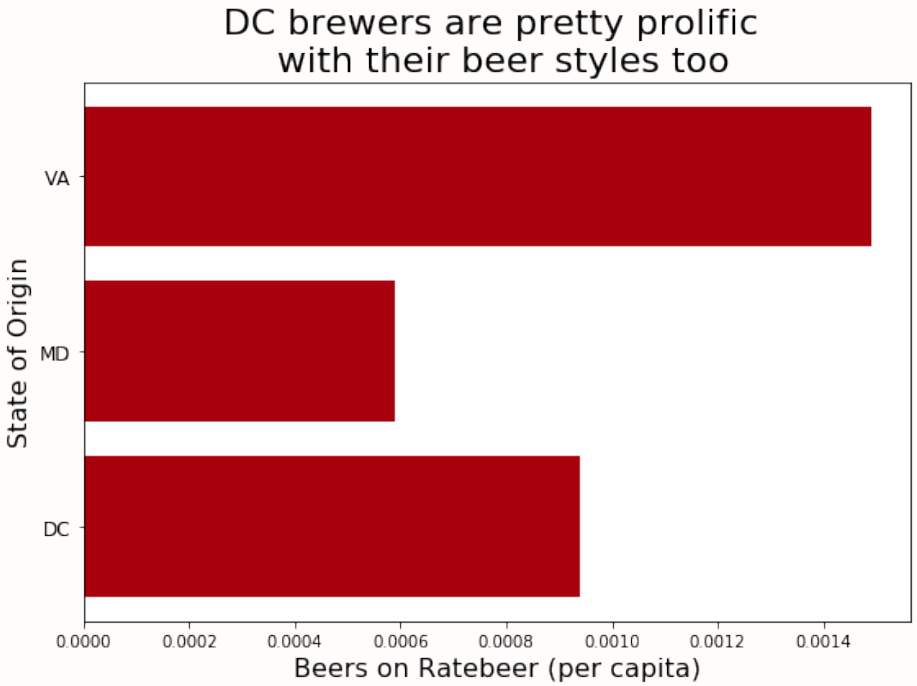
- Check on the API access rules
  - *RateBeer says 7-10 business days for approval (still waiting)*
  - *Untappd strict approval regulations (still waiting)*
  - *BeerAdvocate – none*
- Scraping is fun but infinite scrolling is really annoying
- Selenium works but there may be better ways??
- Key lesson: <u>Set a clear scope and give yourself enough time</u>

BEER-EDA

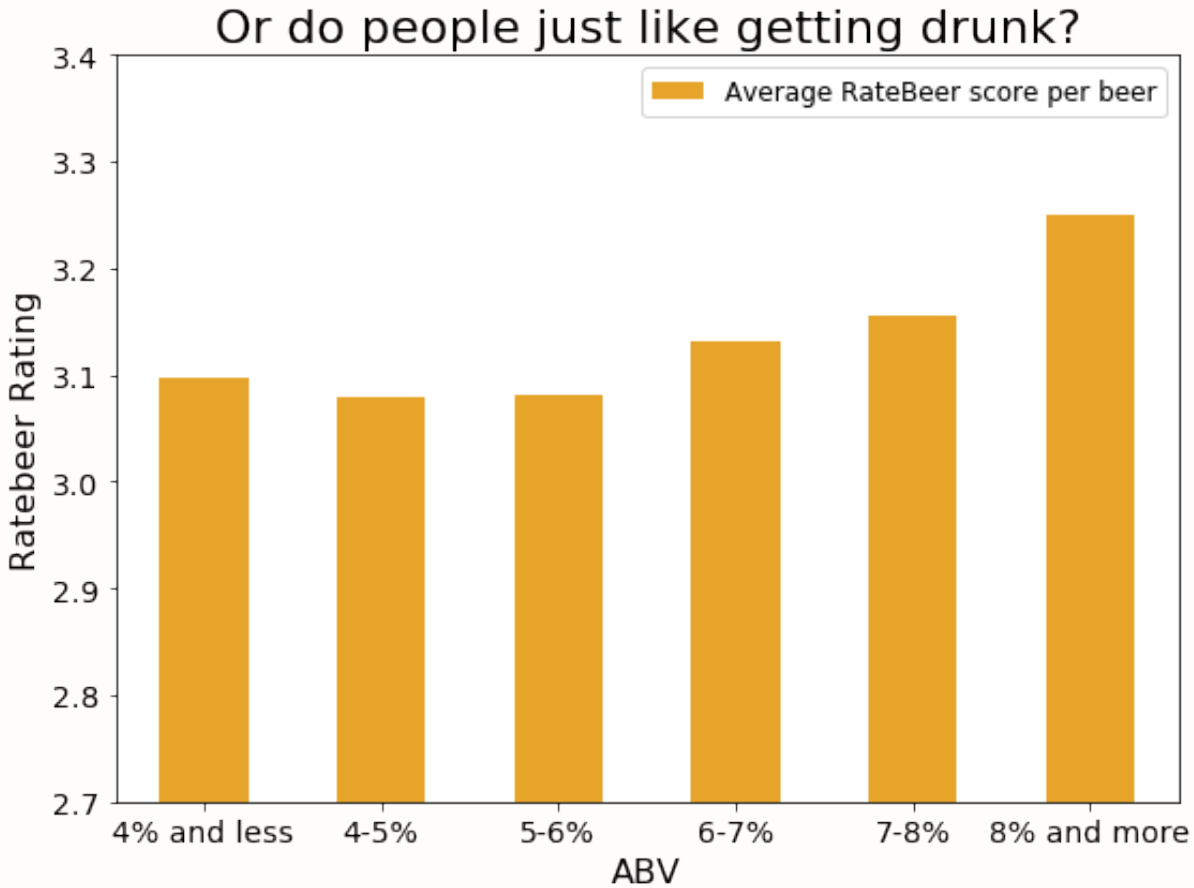
aka Cool Graphs

Beer Stats – Virginia Brewers are Busy

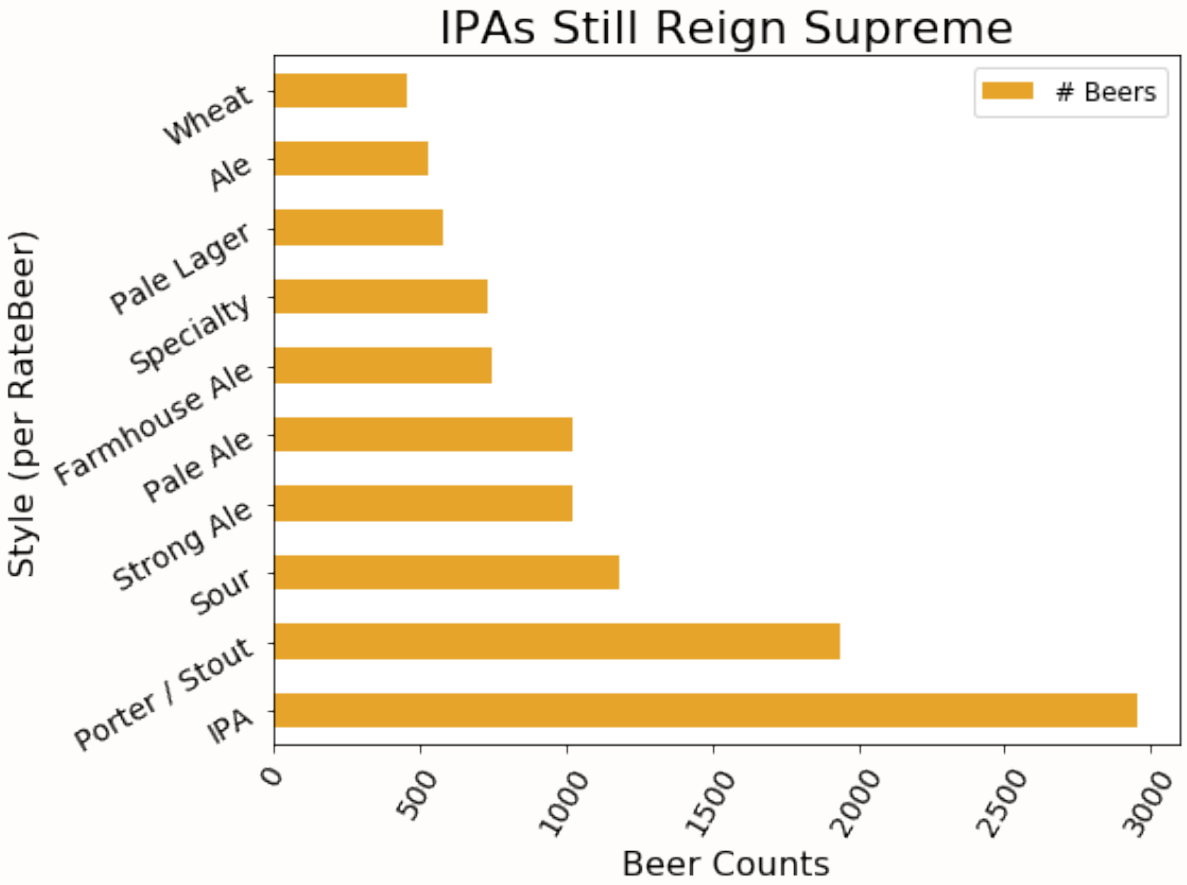




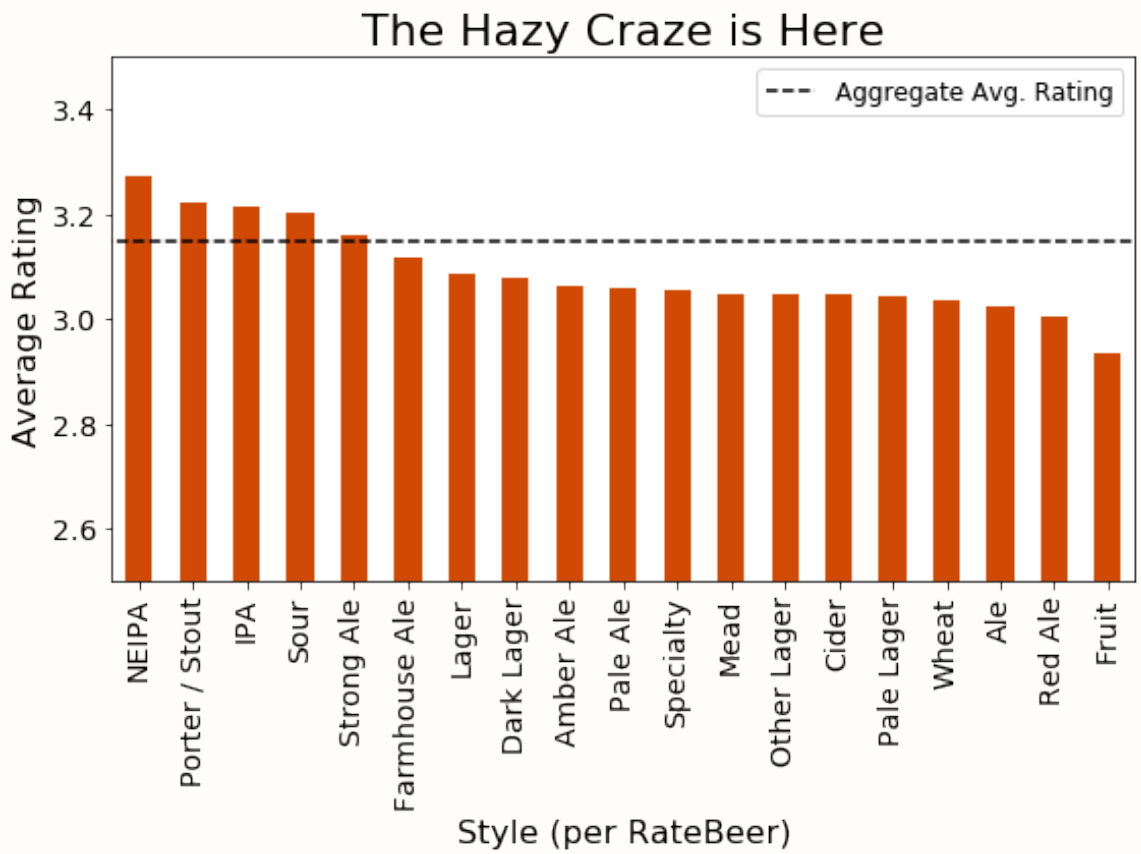
Beer Stats – Bigger is Better!



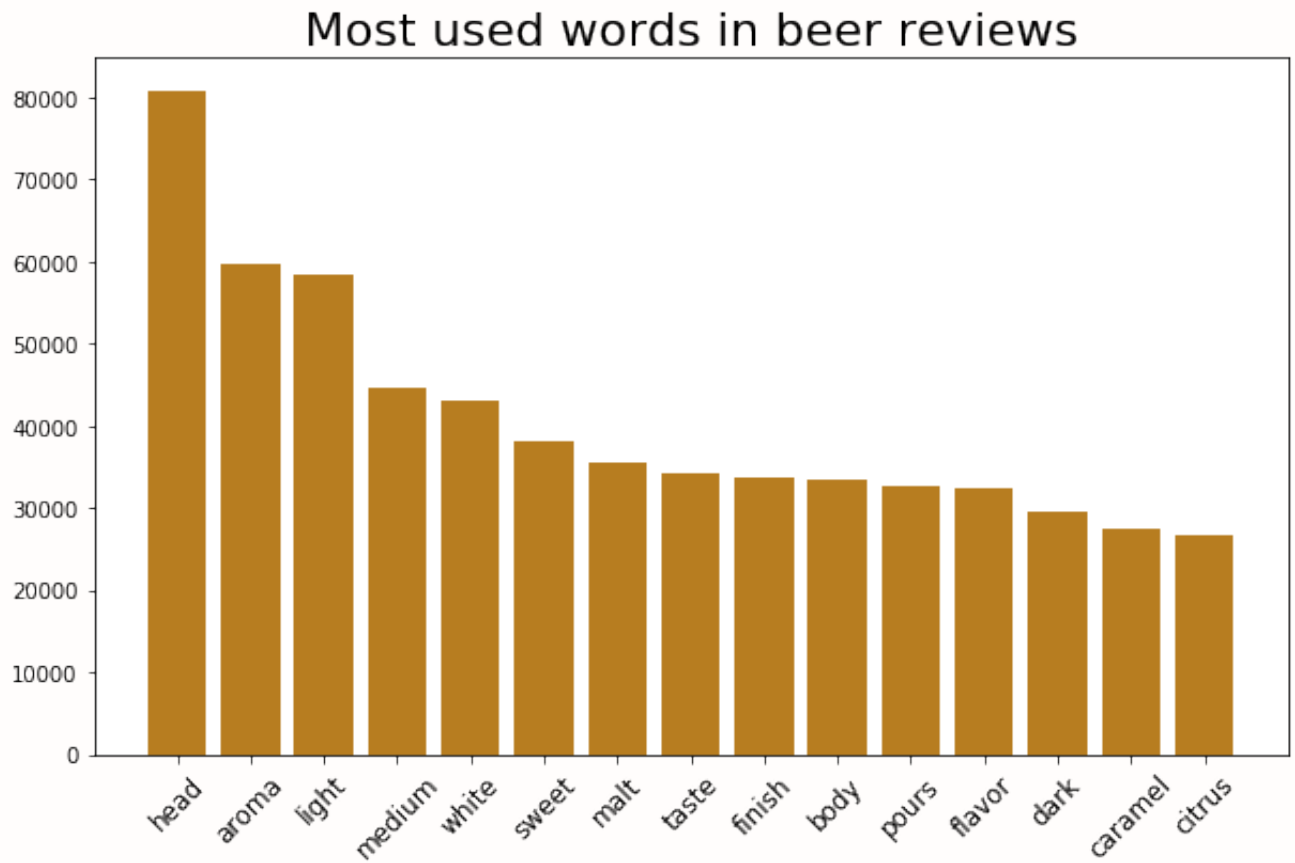
Beer Stats – Can there be too many IPAs?



Beer Stats – Embrace the Haze



Beer Stats – Don’t fear the foam



RECOMMENDATIONS

aka the beauty of cosine similarity

A sample recommendation …

If you like ….
3 Stars Above the Clouds

TRY:

- Stillwater / Mikkeller Two Gypsies Our Side

AVOID:

- Stillwater / Mikkeller Two Gypsies Our Side

How does Poland fit into the DMV beer scene?

Kept noticing The Answer Brewpub beers being listed as extremely different.....

- Apparently they brought their beers to the One More Beer Festival in Krakow in 2018
- Large portion of reviews for those beers are in Polish!
- Lesson: Decrease your max_features on vectorizer!

Na Zdrowie!

Recommender Systems:
Key Takeaways

- Battle of Relevance vs. Diversity
  - You want *similar* but *interesting* as well
  - Hard especially with content recommenders
- Should improve with user data (behavior/demographics)
  - Are the reviewers on RateBeer actually your target?
  - Different recommendations for different audiences?
- Endless tinkering possible …

Project Roadmap

- Large Body of Data ✓
- Working Demo with widgets (semi-✓ …)
- Adding in Doc2Vec so it can accept more general queries
- Finish building Flask app
  - *Add in output for additional details of the recommended beers (directions to brewery, etc.)*
- Deploy with Heroku
- RDWHAHB (relax don't worry have a homebrew)

TIME FOR BEER!

and questions?