Beer Recommendation System

CJ Johanson

Capstone Project 2

Milestone Report 1

What are recommendation systems?

- Recommendation systems are machine learning programs designed to suggest new products/content to users
- Widely used across websites
- Users are likely to encounter recommendation systems a few times each day without knowing
- Some of the of the most famous sites using them include Netflix and Amazon

Why use a recommendation system?

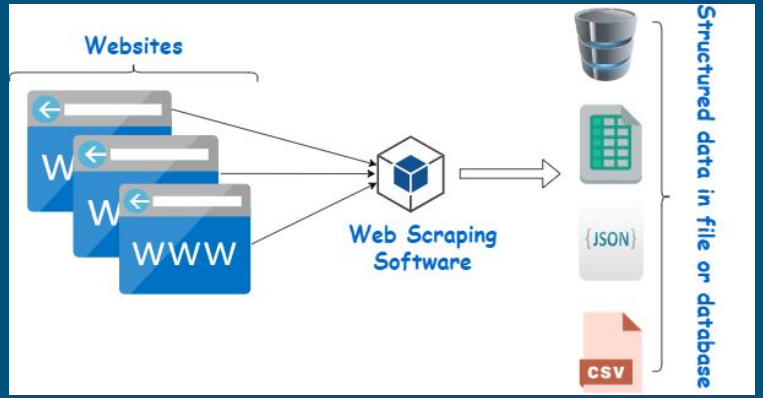
- Using a recommendation system can have various benefits to a business, particularly a small or medium-sized brewery/distributor with nationwide reach.
- Benefits:
  - Increased sales = more profit.
  - Info gathered about customers can lead to valuable insights about their behaviors.
  - Sufficient amount of info could potentially or organized, packaged, and monetized.

Why beer?

- Given the current global coronavirus pandemic, alcohol sales have increased, even though bars and breweries have been temporarily forced to shut down
- Businesses should take advantage of the increase in sales by finding ways to introduce new products to customers
- By creating a beer recommendation system, it can help smaller breweries and distributor to introduce new products to users with the end goal of users purchasing the recommended beers

The data

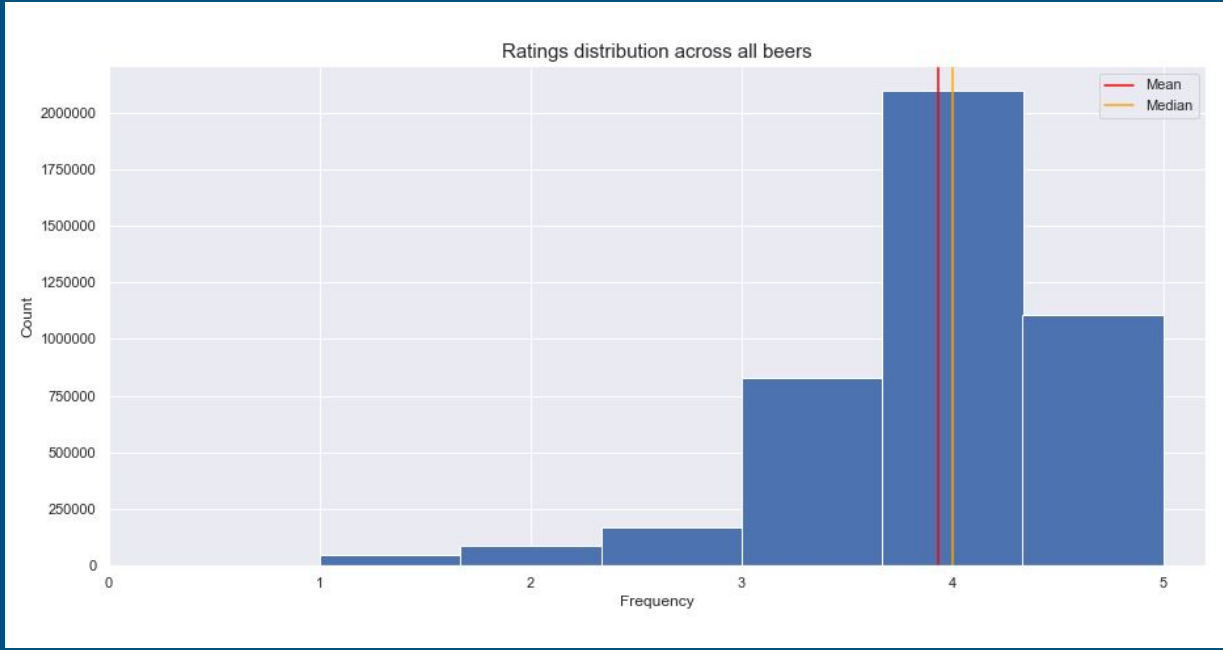
- No existing database of beer reviews
- Web scraping system created using Python, cloud computing, and cloud database storage
- Resulting database included four tables:
  - Beer Styles (111 rows)
  - Beers (~5,500 rows)
  - Breweries (~1,500 rows)
  - Ratings (~4,300,000 rows)



Exploratory Data Analysis

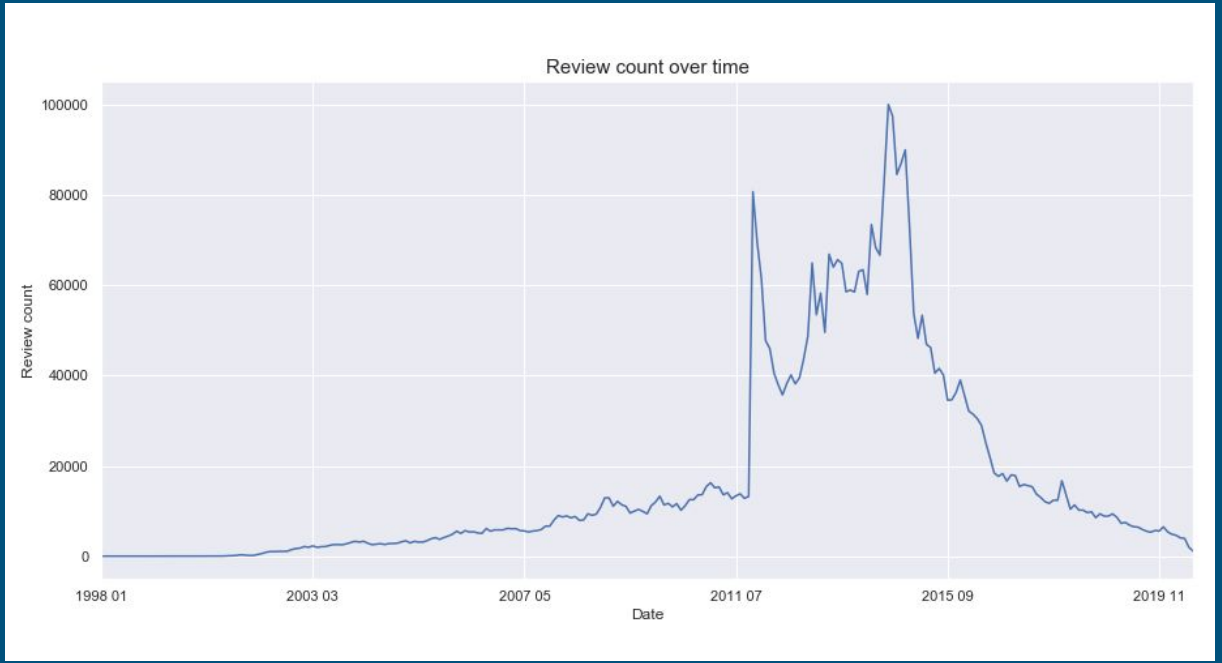
Distribution of ratings

- Ratings distribution is left tailed/skewed, where the mean is less than the median
- Possible explanation: Users of a beer review website are more likely to positively review beers
    - If they didn't like beer or rating beer, they probably would not go to website in the first place



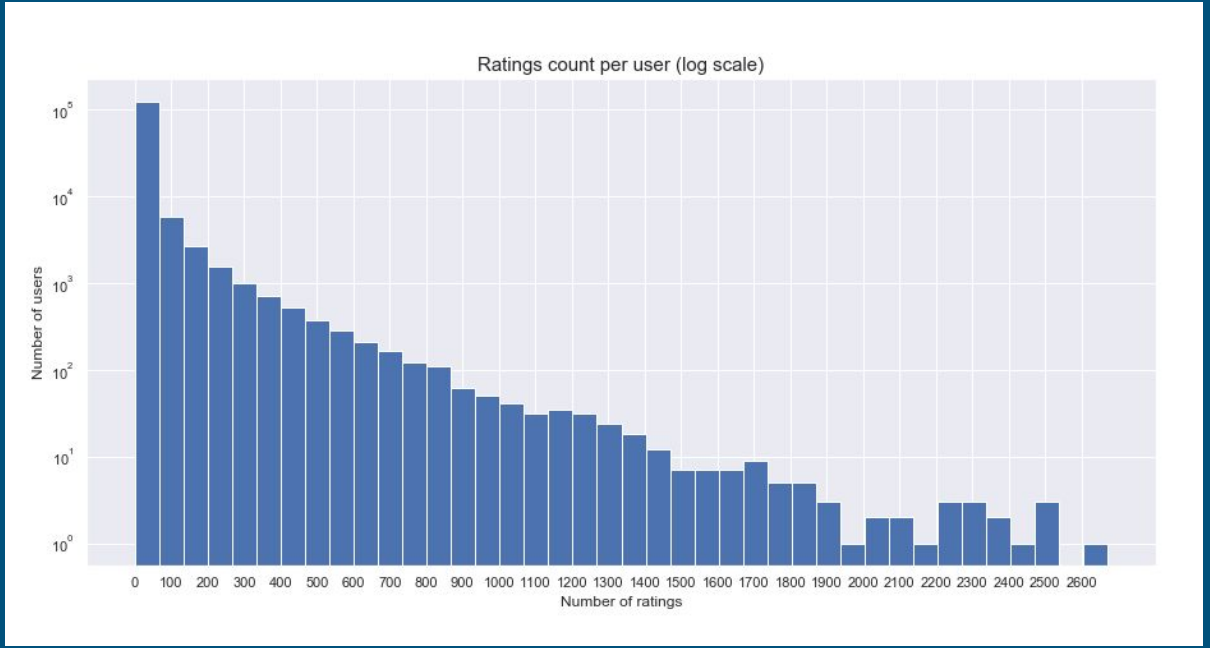
Number of ratings over time

- Number of reviews begins to increase around 2003
- Increase continues and experiences sharp increase around 2011
- 2011 to 2014 experiences overall growth, albeit with ups and downs
- Peak around 2014, steady decrease afterwards



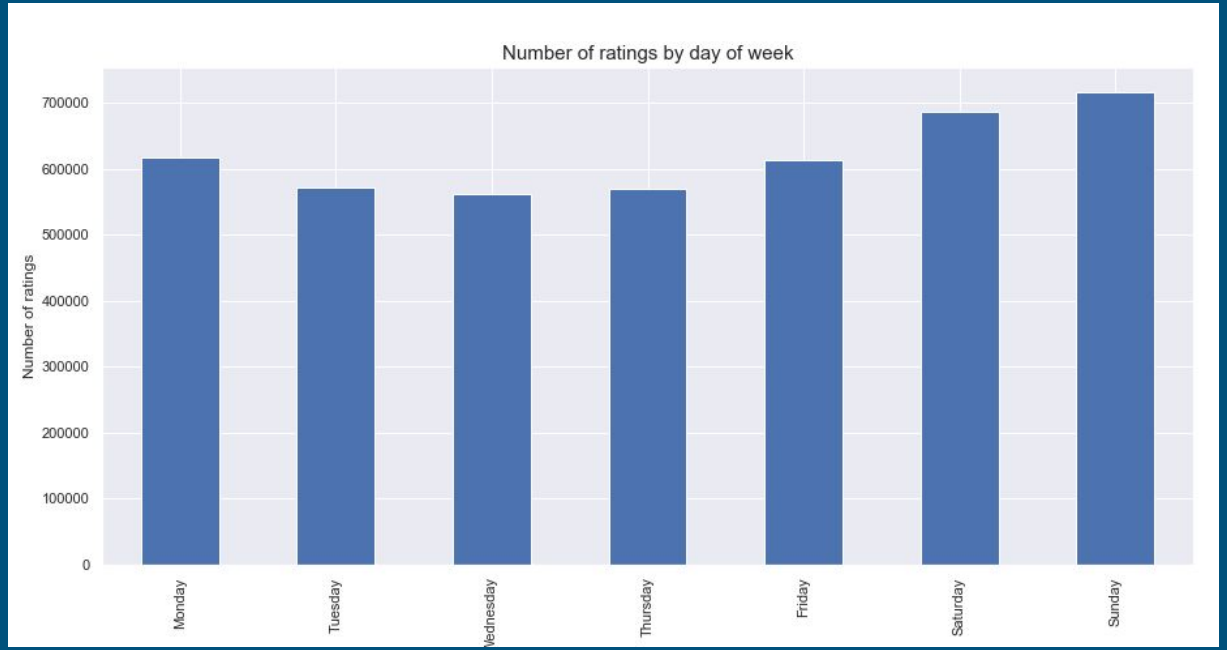
Number of ratings by user

- Vast majority of users write a relatively small number of ratings (less than 100)
- Opposite side of the chart is interesting
- A small group of users (about 300) write over 2,000 reviews per user!



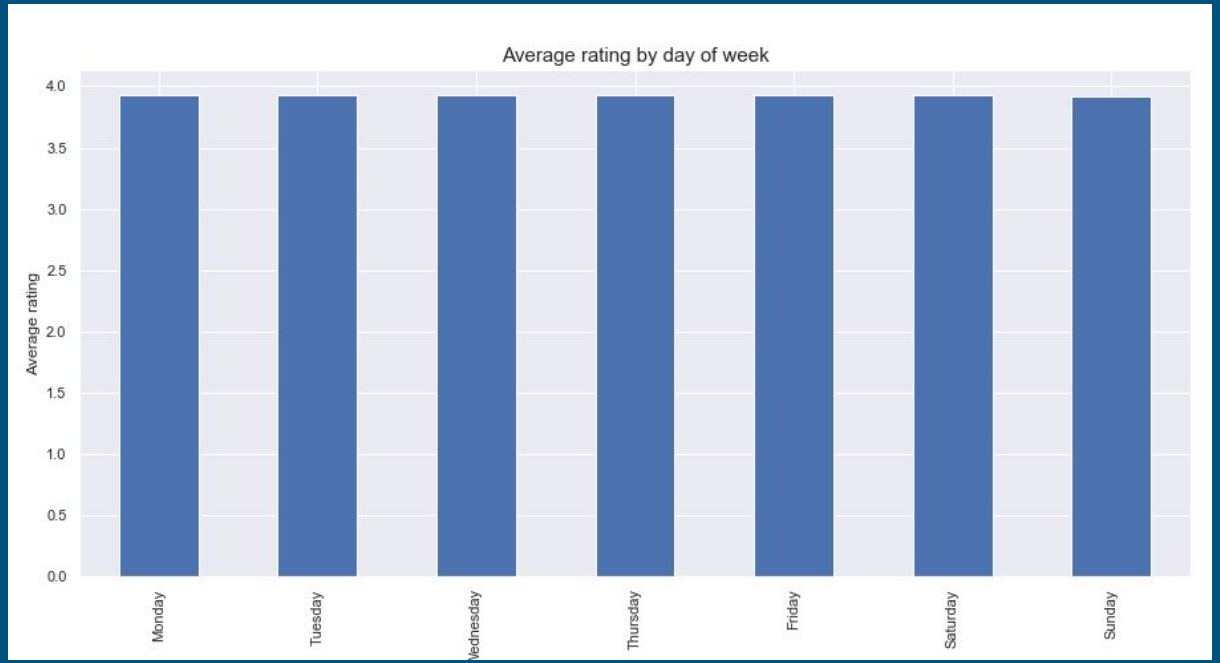
Rating count by day of week

- Initial hypothesis was that most reviews are written during the weekend, when the majority of people have time off from work
- Hypothesis turned out to be true, more or less. One key, interesting difference: Mondays and Fridays are similar in terms of how many ratings are written



Average rating by day of week

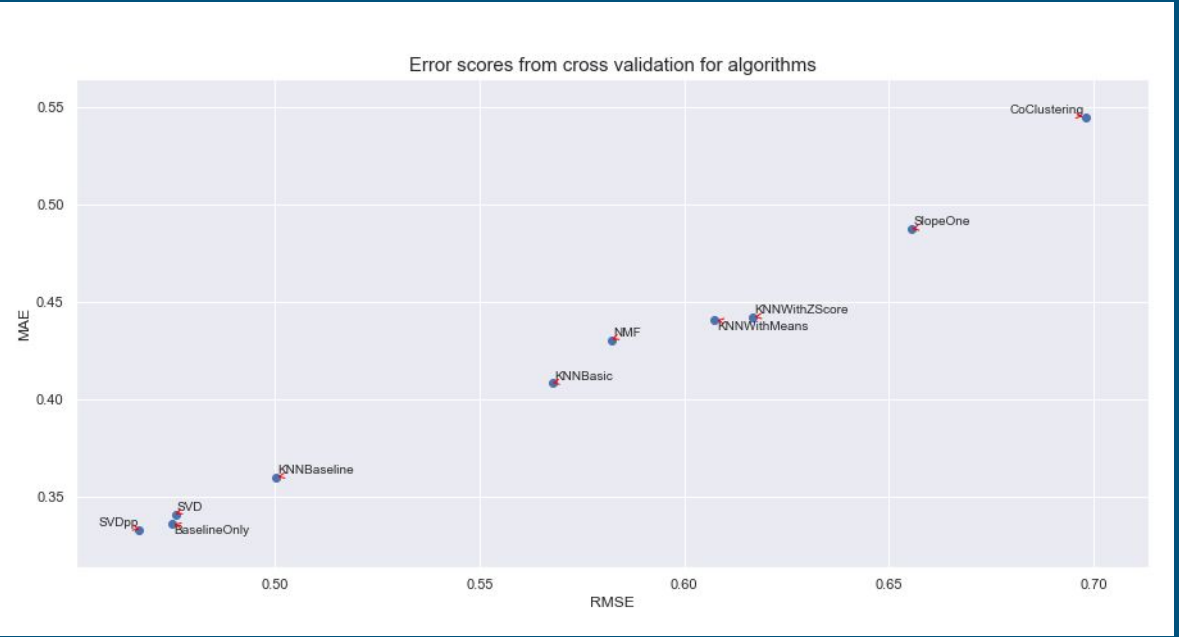
- Initial hypothesis was that weekends could have higher average ratings. Perhaps people drink beer on weekends and increased beer intake means highers ratings?
- Hypothesis was incorrect
- Average ratings are uniform across all days of the week



Machine Learning Analysis

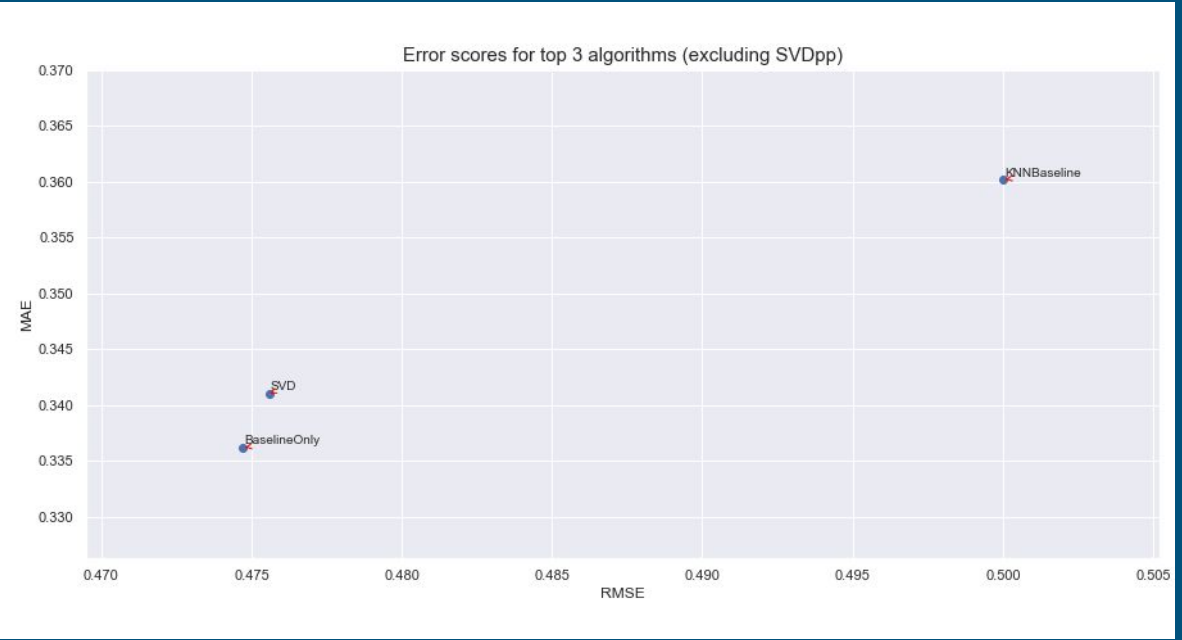
Benchmark model comparison

- Out of the 11 available algorithms, 9 were tested
- The two that were not were NormalizedPredictor and BaselineOnly which are not predictive but are used for comparative purposes
- BaselineOnly is included throughout the analysis for a baseline comparison
- Top performer is SVDpp



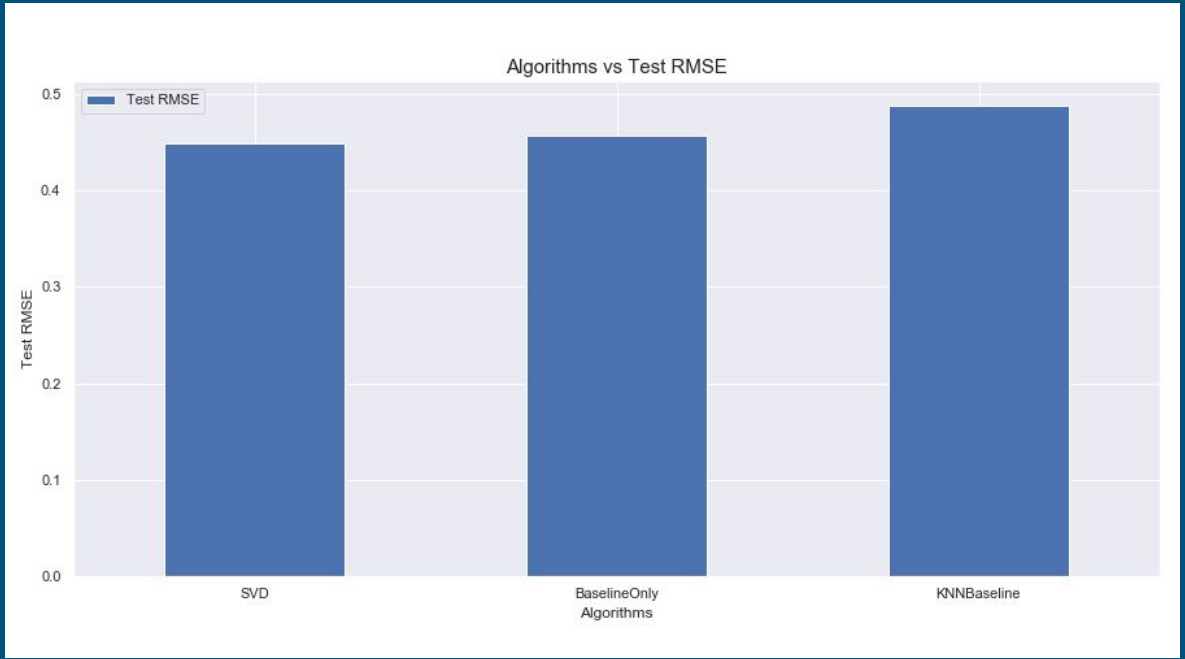
Top 3 Algorithms

- SVDpp excluded due to computational limits
- Remember: BaselineOnly is strictly for comparison
- SVD vs KNNBaseline
- SVD shows better performance but both will be tested



Top 3 Algorithms Compared

- SVD has the best performance by a significant margin
- Result was expected based off the initial comparison of 9 algorithms
- SVD will be used to make predictions



How is similarity determined?

- Cosine distance
  - Has range of 0 to 2 where 0 represents complete similarity
- Compared Budweiser to Budweiser
  - Cosine distance = 0 (they’re the same)
- Budweiser to Bud Light
  - Cosine distance = 0.25 (they’re similar, but not the same)
- Budweiser to Guiness
  - Cosine distance = 0.92 (they’re not similar)

What do the recommendations look like?

Example

- User has a preference for India Pale Ales (IPAs)
    - 5 out of 8 rated beers are IPAs
- Other ratings have some variety in terms of beer styles
- What kinds of beers will be recommended for this user?

	rating	beer_name	brewery_name	beer_style	beer_substyle	abv	user_id
787131	4.75	Fou' Foune	Brasserie Cantillon	Wild/Sour Beers	Belgian Fruit Lambic	5.50	10995
2180183	4.75	Heady Topper	The Alchemist	India Pale Ales	New England IPA	8.00	10995
2190470	4.50	Julius	Tree House Brewing Company	India Pale Ales	New England IPA	6.80	10995
1602982	4.25	Enjoy By IPA	Stone Brewing	India Pale Ales	American Imperial IPA	9.40	10995
3857968	4.25	Breakfast Stout	Founders Brewing Company	Stouts	English Oatmeal Stout	8.30	10995
233200	4.00	La Fin Du Monde	Unibroue	Strong Ales	Belgian Tripel	9.00	10995
1612429	4.00	Lagunitas Sucks	Lagunitas Brewing Company	India Pale Ales	American Imperial IPA	7.85	10995
1818110	4.00	Sculpin	Ballast Point Brewing Company	India Pale Ales	American IPA	7.00	10995

Example recommendations

- India Pale Ales and Pales appear in results, which is understandable given the user's ratings
- Strong Ales
  - Appear once in ratings but 3 times in recommendations
  - SVD algorithm picked up on preference of high ABV and suggested more Strong Ales
  - Power of latent variables

	beer_name	brewery_name	beer_style	beer_substyle	abv
0	Consecrator	Bell's Brewery - Eccentric Café & General Store	Bocks	German Doppelbock	8.0
1	Melcher Street	Trillium Brewing Company	India Pale Ales	New England IPA	7.2
2	Zoe	Maine Beer Company	Pale Ales	American Amber / Red Ale	7.2
3	3 MONTS (Originale)	Brasserie 3 MONTS	Pale Ales	French Bière de Garde	8.5
4	Trial Of Dmitri	Benchtop Brewing Company	Specialty Beers	Russian Kvass	4.5
5	La Clef Des Champs	Brasserie Dieu du Ciel!	Specialty Beers	Scottish Gruit / Ancient Herbed Ale	5.0
6	Double Or Nothing	Otter Creek Brewing	Strong Ales	American Barleywine	10.5
7	Three Philosophers	Brewery Ommegang	Strong Ales	Belgian Quadrupel (Quad)	9.7
8	Gouden Carolus Classic	Brouwerij Het Anker	Strong Ales	Belgian Strong Dark Ale	8.5
9	Blanche De Chambly	Unibroue	Wheat Beers	Belgian Witbier	5.0

Conclusion

- Recommendation systems are popular because they are useful
- Great for driving sales, growth, and increasing quantity/quality of customer interactions
- Amount of data not used to recommendation engine
  - Ideas for the future
    - Create a recommendation engine based on breweries
      - Suggestions breweries that both compliment the user's rating history and are close to visit
- Recommendations engines will likely continue playing a significant role in the future

Questions?

Thank you!