Building Recommender Systems to Identify Top Beauty Products

Jonathan Lee

Business Problem

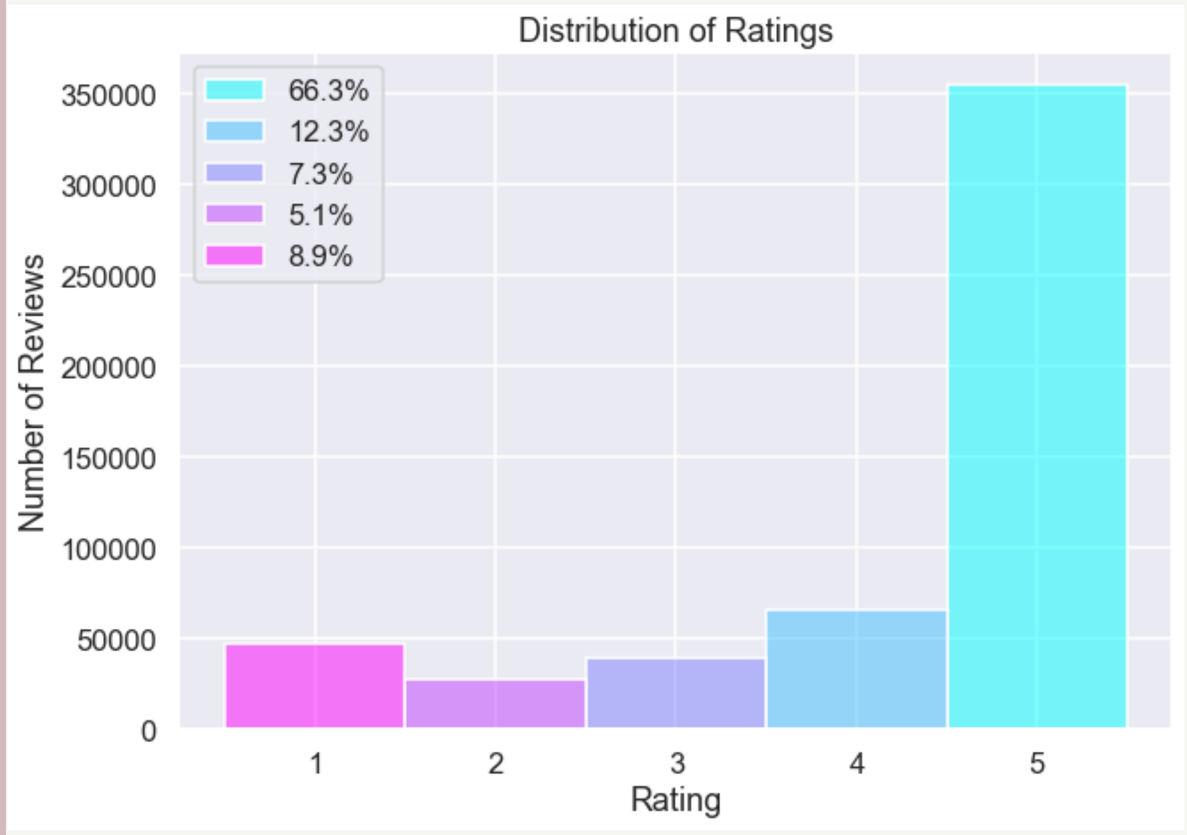
- Relatively new to the beauty product industry
- Want to build a recommender system based on Amazon's data, to determine what model and parameters to use for our own recommender system
    1. Memory-based or model-based?
    2. What specific algorithm?
    3. What hyper parameters?

Data Overview

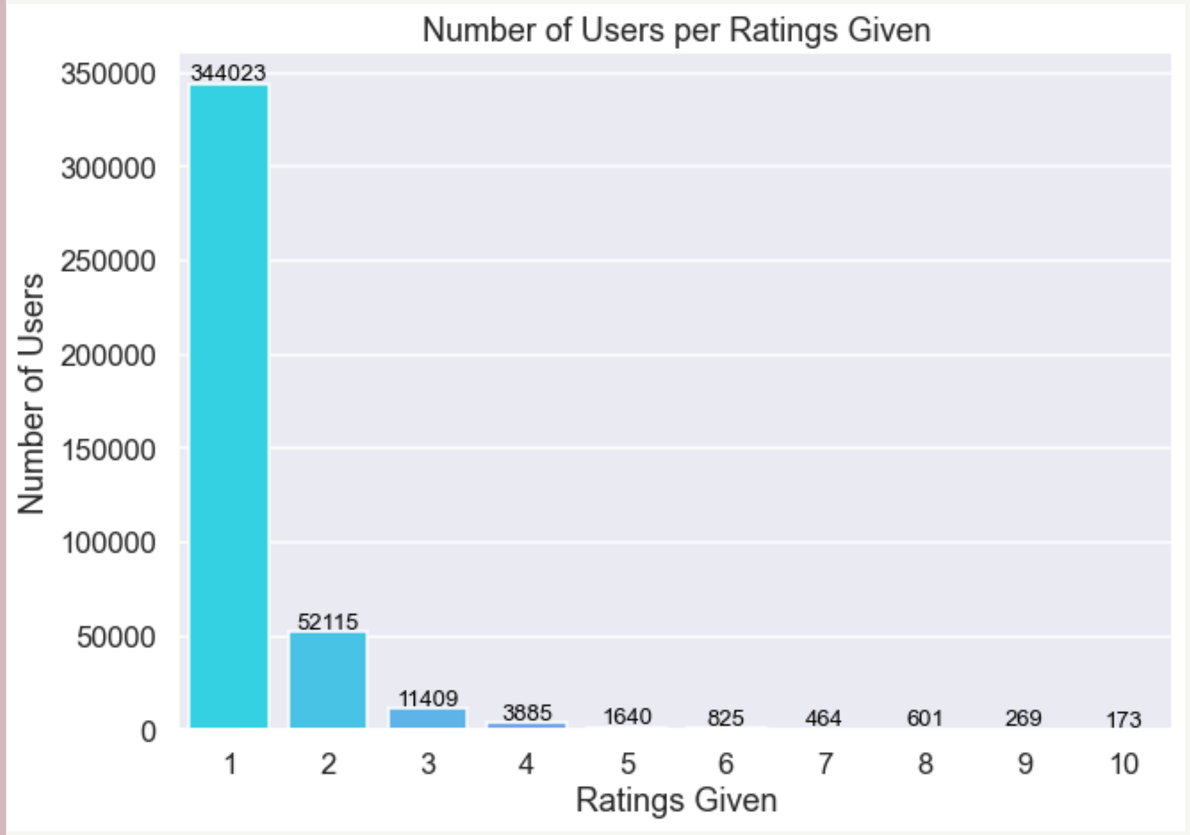
Amazon Ratings

- Contains ratings collected between May 1996 to Oct 2018
- Products under “Luxury Beauty” category
- 536,111 total ratings
- Contains product ASIN codes for 12,111 listings
- Includes ratings from 416,077 unique users
- Ratings are on a scale from 1 - 5

Rating Distribution



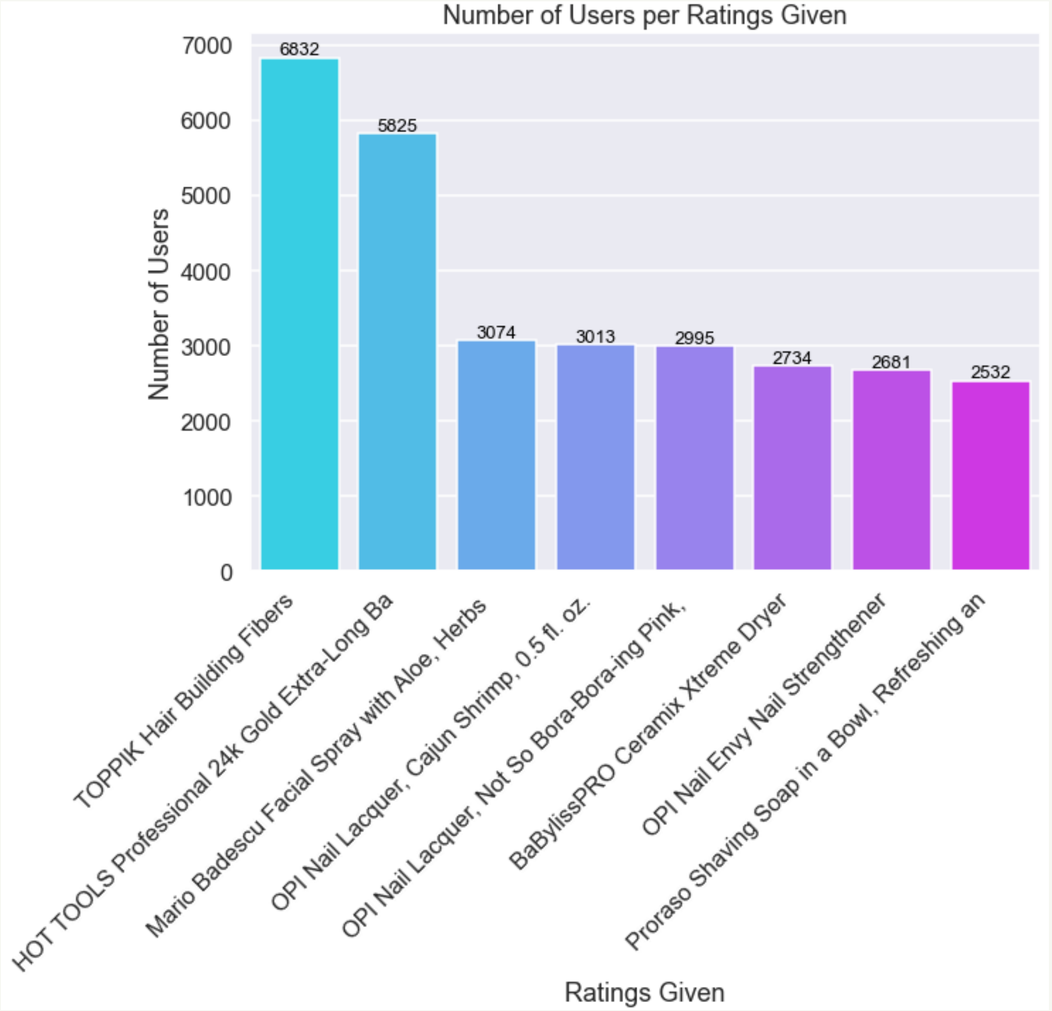
Users per Number of Ratings Given



What Products are Most Popular on Amazon?

Most Popular Products

by Number of Ratings



What Other Products Should We Consider Selling?

How the Model Works

- Model-based Collaborative Filtering Singular Value Decomposition model was most effective in reducing the error between actual ratings and predicted ratings
- Takes existing user ratings of items and breaks them down into item characteristics as well as user preferences that correspond directly to those characteristics.
- Final model breaks down our data into 150 item characteristics and 150 user preferences with respect to those characteristics

Example Customer #1

User Ratings:

- **WEN Sweet Almond Mint Texture Balm - 5 Stars**

Recommended products:

- **Kneipp Lavender Mineral Bath Salt, Relaxing**

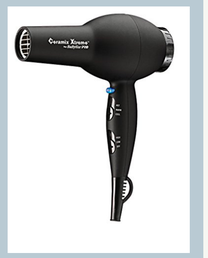
- **Aromatherapy Associates Deep Relax Bath And Shower Oil**

- **L'Occitane Green Tea Eau de Toilette**

Example Customer #2

User Ratings:

- **BaBylissPRO Ceramix Xtreme Dryer - 5 Stars**



- **theBalm INSTAIN Blush - 4 Stars**

Recommended products:

- **Eau Thermale Avène Avène Thermal Spring Water Gel**

- **boscia Clear Complexion Blotting Linens**

- **JAPONESQUE Travel Smudger Brush**

Conclusion

- The best way to build our own performance optimized recommender system for our new eCommerce platform:

  1. Model-based

  2. Singular Value Decomposition Model

  3. 150 Latent Features to optimize accuracy with large number of products

Further Analysis

- Amazon has a category for “All Beauty”, so it might be worthwhile to combine these ratings with the data used in this analysis
- Our model was capable of predicting user preferences with an average error of about 0.92 stars, which might be improved by adding more data or by tuning our model
- Try using alternate modeling algorithms on combined dataset to improve recommendation accuracy

Thank You!