Recommendations, anyone?

By: Zach Hyde
Flatiron School- Part time Data Science
Cohort 7/6/2020

Business Case

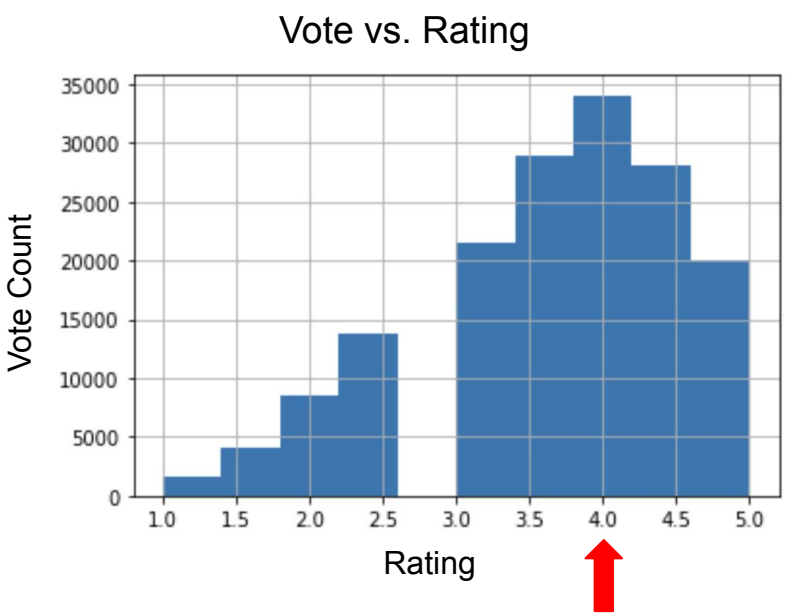
I took a Yelp API database to create a Recommendation System utilizing Surprise. Recommendation systems are extremely important to a corporation due to learning behavioral patterns of a user and their preferences or interests are and effectively advertising new or similar product to them. This in turn, keeps the consumer engaged and interested in product leading to consistent purchasing behaviors and increased loyalty within the brand.

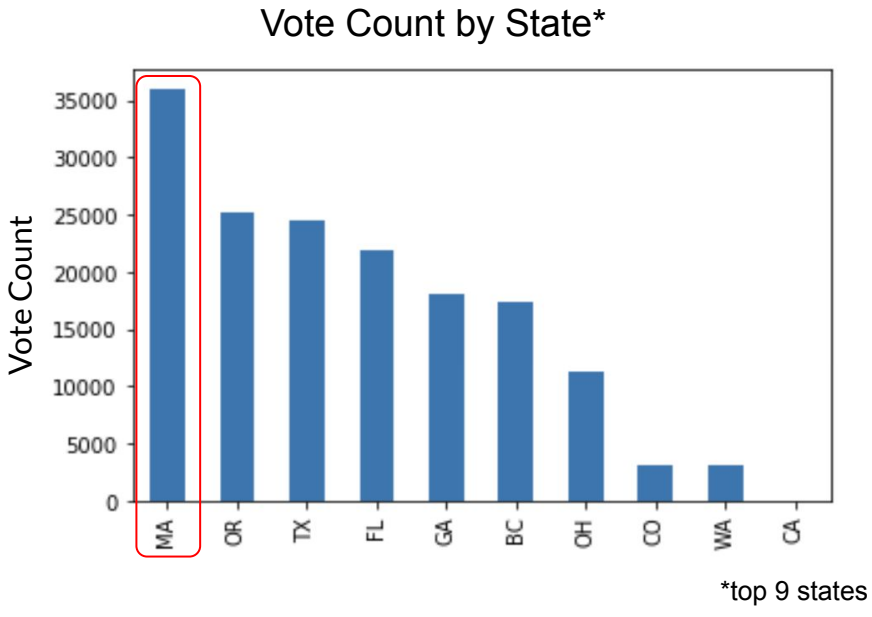
EDA

Data Entries = 8.5 million +

Features = 14

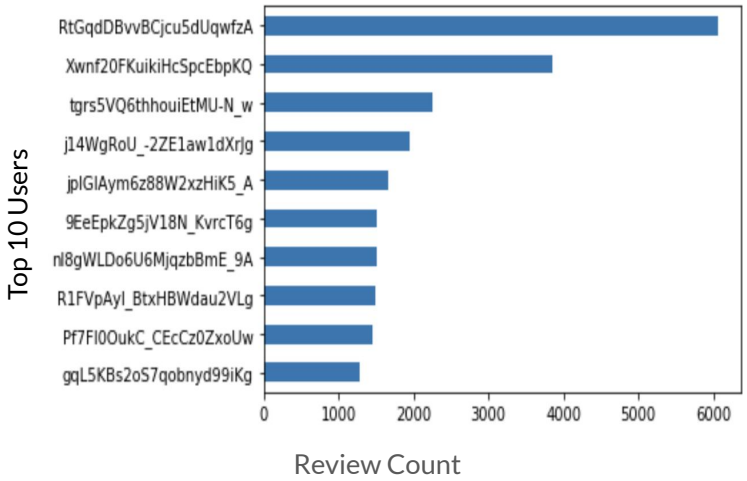
- What are they?





*top 9 states

EDA continued



	unique_values	counts
0	RtGqdDBvvBCjcu5dUqwfzA	6073
1	Xwnf20FKuikiHcSpcEbpKQ	3861
2	tgrs5VQ6thhouiEtMU-N_w	2265
3	j14WgRoU_-2ZE1aw1dXrJg	1958
4	jpIGlAym6z88W2xzHiK5_A	1671
...	...	...
2189452	-YIC6RNgwad72SM48Taf8Q	1
2189453	0tzhQW4TFk0rAimfzBt3fQ	1
2189454	XgXP7GaBmJ5Bfsas-svbnA	1
2189455	GROXfaefwvXir12s2bPRPw	1
2189456	0c12y_Q0qkV40BkQl21uBA	1

2189457 rows × 2 columns

	unique_values	counts
0	RtGqdDBvvBCjcu5dUqwfzA	6073
1	Xwnf20FKuikiHcSpcEbpKQ	3861
2	tgrs5VQ6thhouiEtMU-N_w	2265
3	j14WgRoU_-2ZE1aw1dXrJg	1958
4	jpIGlAym6z88W2xzHiK5_A	1671
...	...	...
5743	cTVJlVQBfEtsrPW0dM3z4A	101
5744	D_liFPO9WgITdQSun24geA	101
5745	JqEOJxG4J5ybEoBH363jwQ	101
5746	UU4zFYWXQKwqWLlt2veXMQ	101
5747	VMuvh0otnR-YOz2abRl2tw	101

5748 rows × 2 columns

Over 2 million low quality user entries redacted

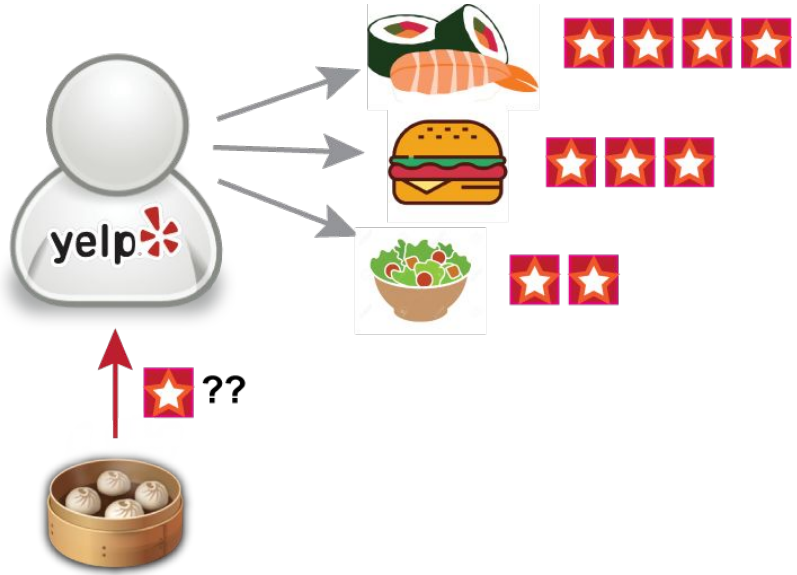
Model's Evaluated

NMF ( )	RMSE accuracy = 1.4348	Base model
SVDpp ( )	RMSE accuracy = 1.2840	Took over an hour to run
SVD ( )	RMSE accuracy = 1.2744	Model of choice being the 2nd lowest initial RMSE and 2nd fastest to run
BaselineOnly ( )	RMSE accuracy = 1.2704	* Even though this had the best RMSE, it shouldn't be utilized for final model

**One can interpret Root Means Squared Error (RMSE) as how far off the prediction is from the actual review score.

Model's Evaluated with subset

NMF ( )	RMSE accuracy = 1.0733
SVD ( )	RMSE accuracy = 0.9827
BaselineOnly ( )	RMSE accuracy = 0.9815



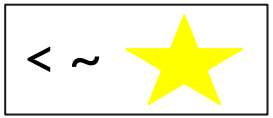
Final Model with Params

SVD (n_factors = 5, reg_all = 0.020)

After implementing parameters ⇒ 0.9804 RMSE

params	param_n_factors	param_reg_all
{'n_factors': 5, 'reg_all': 0.002}	5	0.002
{'n_factors': 5, 'reg_all': 0.02}	5	0.020
{'n_factors': 5, 'reg_all': 0.1}	5	0.100
{'n_factors': 7, 'reg_all': 0.002}	7	0.002
{'n_factors': 7, 'reg_all': 0.02}	7	0.020
{'n_factors': 7, 'reg_all': 0.1}	7	0.100
{'n_factors': 10, 'reg_all': 0.002}	10	0.002
{'n_factors': 10, 'reg_all': 0.02}	10	0.020
{'n_factors': 10, 'reg_all': 0.1}	10	0.100

Results



After implementing parameters ⇒ 0.9804 RMSE

SVD model within Surprise can recommend a restaurant within less than one star accuracy of what a user would rate on their own.

Utilizing this model within the Yelp app or website would effectively recommend a new restaurant to a user that hasn't been to a location similar to those they have previously rated before, gaining new business for the restaurant and expanded options for the user.

Future Recommendations

This modeling specifically isolated the term ‘Restaurants’ within the business category column of the JSON file utilized. This modeling can be expanded to isolate many other types of businesses with some minor adjustments during the EDA portion.

There is also the opportunity to web scrape sites such as Google to build a model off of those reviews of the same businesses. There can possibly be a merger of the two sources with averages of individual businesses to have a well rounded model to effectively recommend a new establishment to a user who may look at both sites for reviews to make a final selection on how to spend their finances.

GitHub: Here

LinkedIn: Here

Email: zacharyhyde14@gmail.com