Kojak MVP

Concept: A recommender system that takes into account words used in reviews. Using data from Yelp, I plan to recommend new restaurant to user not only using traditional recommender methods, but also NLP. For instance, if a particular reviewer uses certain words more frequently in their higher rated reviews (ex. "trendy" , "new") the recommender would suggest place that that also contain similar words in their respective reviews.

Domain: I have used a bit on NLP in my last project and plan to pick up where I left off in this one. I have not worked with recommender systems in the past.

Data: Yelp review data from Kaggle (provided by Yelp for a past competition).

Variable Name	Type	Description
Business ID	ID	Unique ID of Business
Business Name	String	Name of Business
Business Location (state)	String	
Business Location (city)	String	
Stars	Float	Average of all review stars
Review Count	Integer	Number of reviews per business
Category	String	Type of Business
Is Open	Boolean	Check if the place is still in Business
Business Attributes	String	Pick which ones to use for the recommender
Review ID	ID	Unique ID of every review
Text	String	Body of the review
Useful	Integer	Votes on the review
Funny	Integer	Votes on the review
Cool	Integer	Votes on the review

Known Unknowns: Currently the data I plan on using exists in 4 different tables. I will need to figure out how to combine it a way that makes it useful for my project.