Instructions

In this assignment you are asked to apply your textual analysis skills to the data your group collected in the course of Assignment 1. In particular, you are asked to conduct topic modeling on the textual reviews of your Yelp data set. For the topic modeling analysis, you are asked to use the service of structural topic modeling (stm) package in R.

In your assignment, you can focus on a single topic, a few topics, or many different topics. You can follow a Top-Down or Bottom-Up approach. The choice is up to you. As a suggestion, you could focus on how the topics of "not recommended" reviews evolve over time, for example. As in Assignment 1, each group can choose their own focus of the Assignment. Groups are, however, asked to position their work in the body of literature on online reviews as discussed in class.

After the topic modelling, please link the topic(s) you identified to other variables such as online review metrics (e.g., rating valence) or business attributes (e.g., restaurant type). Please use regressions to investigate these relationships. Please make sure to justify the number of topics for your final model depending on the strategy you pursue (top down or bottom up). If possible, verify the validity of your topic model user intercoder reliability.

The assignment is divided into three steps:

Step 1: Decide Between Top-Down or Bottom-Up Approach

Decide whether you will follow a Top-Down or Bottom-Up approach. After deciding this, please follow the steps for these approaches as shown in the lecture.

Step 2: Conduct Topic Modeling

If you have not done so far, please clean your textual data and prepare them for using them in R. Afterwards, conduct topic modelling on our data. Try to get the best topic model for your purpose.

Step 3: Linking Topics to Other Variables

Present analyses of your data with a focus on the topic(s) you identified. Start out with very descriptive statistics on the characteristics of the topics, e.g., the number of topics in your final specification, the most frequent words in your focal topic(s), your labels (“names”) of the topics, sample reviews from the focal topics etc.

Afterwards, link your focal topic(s) to other variables and investigate how they are related. For instance, you can investigate the rating valence of reviews of a certain topic, how many reviews of a given topic certain restaurant types have received, or how review topics evolve over time. Present more detailed analyses of your data in the form of figures and charts, allowing a more nuanced insight into the industry and its online ratings/reviews. Get creative and think about which insights could be valuable for which audience for which purpose!

Finally, you are asked to run regression models (e.g., linear regressions, logistic regressions) to investigate the relationship between your focal topic(s) and other variables. Linear or logit models are sufficient for that. If you have learned different models during your program that you would like to apply, you are free to do so.