Team 5

Technology Review Proposal

My project proposal is a Joint Analysis of Hotel Review & Historical Local Economy Metrics for Causal Topics. As detailed in my project proposal, the tool I plan to build would have at least three major parts:

1. A hotel review text data ingestion agent
2. A economic indicator data agent
3. A main analytic engine to depict insight

The hotel review text data ingestion agent would require the most effort and involve the greatest application from this course. It would need to be able to provide topic detection and/or sentiment analysis.

I plan to build the tool in R. The following are potential software toolkits (each with a short description) that I will evaluate as part of my Technology Review to build this tool.

- mscstexta4r (link to GitHub)
    - Based off of Microsoft Cognitive Services (link)
    - Supports relevant operations:
        - Sentiment analysis
        - Topic detection
    - GitHub info shows last update in June of 2016
- topicmodels (link to CRAN)
    - Can support both LDA (latent Dirichlet allocation) and CTM (correlated topics model)
    - Underlying code written in C and ported to R
    - Most recent version published on April-18-2017
- syuzhet (link to CRAN)
    - Includes tokenizer
    - Built in sentiment function that works at the sentence level
    - Common use is to show sentiment as function of narrative time.
    - Most recent version published on March-6-2017

In addition, to determine causality between topics or sentiments and historical local economy trends, I plan to employ the Granger Causality Test. Below are a couple of toolkits I plan to evaluate to help me:

- lmtest (link to CRAN)
    - includes grangertest function to evaluate causality with a time series.
    - Most recent version published on February 11, 2017
- MSBVAR (link to RDocumentatioin)
    - includes granger.test function to evaluate multiple time series
    - Most recent version published on February 10, 2015

From the evaluation of the toolkits, I plan to use at two or more of the toolkits I described above.