

Melissa Cho (mccho3)

Iris Kan (iriskan2)

CS 410

15 November 2021

Project Progress Report

1) Which tasks have been completed?

We have completed all of Week 10, Week 11, and Week 12 tasks:

Week 10:

- Web scraping google reviews of apartments and hotels in Chicago
- Building dataframe of google reviews of apartments in Urbana-Champaign
- Classify/give scores to word with positive/negative connotations
- Assigning a sentiment score to each google review

Week 11:

- Build sentiment analysis model
 - data cleaning
 - split data frame
 - create bag of words
 - import logistic regression
 - split target and independent variables
 - fit model on data
 - make predictions

Week 12:

- Testing model's accuracy
 - Producing classification report

2) Which tasks are pending?

The pending tasks are the tasks of Week 13:

- Debugging/testing model
- Cleaning up code
- Analyze results/form conclusions
- Preparing for presentation

3) Are you facing any challenges?

We initially faced a roadblock with web scraping- we could not find a good website of apartment and hotel reviews that had the sufficient information we needed to conduct our sentiment analysis. We needed reviews with a score rating (scale of stars), the text body of the review, and a one-line summary of the review. On top of that, we had trouble finding hotels and apartments that had a sufficient number of reviews to gather. However, after searching, we found the websites and web scraped the information that we needed to conduct our sentiment analysis.