# Week 5 Homework

1. Complete Yelp Data Challenge project preprocessing code: “Yelp_Data_Challenge_Project/Yelp_Dataset_-_Data_Preprocessing.ipynb”
    - Load, visualize, filter data

2. Yelp review classification by user sentiment. Build a document classifier to classify positive and negative reviews. Complete Yelp Data Challenge project NLP code: “Yelp_Data_Challenge_Project/Yelp_Dataset_-_NLP.ipynb”. Tasks:
    - Load, visualize data
    - Define positive/negative reviews
    - Extract Tf-Idf feature vectors from review data
    - Build review classifiers using supervised ML models
    - Use cross-validation and grid search to tune parameters and select models
    - **Question:** Think about the use case for this work, and answer why you want to do this in the first when asked by interviewers.

3. Yelp data clustering. Complete Yelp Data Challenge project code (only clustering part): “Yelp_Data_Challenge_Project/Yelp_Dataset_-_Clustering_and_PCA.ipynb”. Tasks:
    - Load, visualize data
    - Extract Tf-Idf feature vectors from review data
    - Perform K-Means clustering of the reviews, we will be limiting to positive reviews (since we don’t want to cluster good vs bad)
    - **Question:** What does the clustering result tell us, look for answers from the centroid and examples, and how this work can help the business (interviewer may ask this type of questions)?
    - Extra credits: there are 5 optional extra credits at the end of ipynb, can you complete any of them?