Project Progress

- Did you find one or more libraries/resources… what are they?

  In general, we found these libraries very helpful for our project, including Sklearn, Numpy, Pandas and Tensorflow.

- Have you successfully installed/run the software… ? How did it go?

  We obtained the data from Yelp dataset challenge in tar format. Each file of the dataset is composed of a single object type, one json-object per-line. Every object contains a 'type' field, which tells you whether it is a business, a user, or a review.

- What is your plan for getting to know the library/resources?

  From the course work of CS35, we have some experience of using Sklearn, Numpy, Pandas. We are also provided with some examples to get started from github of Yelp https://github.com/Yelp/dataset-examples.

  In the first week, we explores the tutorial of TensorFlow from https://www.tensorflow.org/get_started/get_started. Also, we followed the examples from github and ran some provided demos on our laptop to get familiar with them. We extracted the dataset and obtained following json objects:

  - Business objects: contains basic information about local businesses.
  - Review objects: contains the review text, the star rating, and information on votes Yelp users have cast on the review.
  - User objects: contains aggregate information about a single user across all of Yelp
  - Check-in objects: contains information about check ins of businesses.
  - Tip objects: contains the texts, compliment counts and date from tips.