AIT722_Data Collection: Collecting the data from the 4 GB yelp dataset. We are only interested in business and user information. So, the file describes the code for extracting reviews from the large yelp dataset.

AIT 722_Data Preprocessing : All the pre-processing steps of textual data is implemented here. Feature selection from top 1000 words and word2vec model is also implemented in this file

AIT722_GeographyTaste: The main modelling part, clustering is performed here.