❖**Project Title:** Location Mining and Urban Planning

❖**Team information:** Harish Gitta

❖**Objective and overview of the project:**

* Objective of the project is to find whether location really matter for a business. If location really matter for a business then our projects aims to predict what kind of business is ideal at a given location and also aims to find how amenities like WIFI, Parking helps business.

❖**What are the data mining tasks you will perform on the Yelp dataset?**

* Data wrangling: Removing stop words and stammer the review and tips dataset to convert into a consistent dataset.
* Tokenizing Data: Tokenizing the data using NLTK(Python Library) to generate tokens and categorize them as positive and negative tokens.
* Data Classification: Classify the reviews as positive or negative review.
* Prediction: Support vector Regression to keep the statistics up to date.
* Visual Representation: Visually representing the patterns on Maps.

❖**What do you plan to deliver at the end of the semester? How would you present the outcome of your study? How would you demo your work? What will you place on your website?**

* A detailed report on how location substances a business, and an application that helps you find prime business location for different kinds of business.

**❖What are the challenges in this project?**

* Large Dataset: Working on a large dataset is a challenge as there are close to 1.6M reviews, 500K tips.
* Tokenizing Data: Tokening the data into positive and negative sets is a challenge as we don’t have predefined classes of positive and negative words.
* Data Classification: I am not sure on what classifier best suits for the current task.

**❖How do you plan to address the challenges? How would you design and implement the solution?**

* Large Dataset: Partition the dataset into small sample sets to evaluate accuracy and performance, later use the same on large datasets.
* Tokenizing Data: We derive the tokens from reviews and then categorize data into positive and negative tokens using Wordnet (WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms).
* Data Classification: Evaluate different classification algorithms on a sample set of data to find the best classifier that supports given dataset.

❖**How would you partition the tasks and coordinate among team members?**

* I am the only team member.