Project Report(Task 1)

Project name: Yelp User Review Visualization
Group members: Yilai Yan(yy97), Weijian Zeng(wz48), Wanrong Cai(wc52)
Github link: https://github.com/YilaiYan/comp631-yelp-project
Start date: 2/1/2022
Expected Task 1 end date: 2/22/2022

Introduction/Motivation/Problem Definition (What is it that you are trying to solve/achieve and why does it matter)
Introduction:
We are creating a website that allows users to search Houston restaurant names and the website shows the trend of customer comments in five years in the form of cloud words and line charts and the website will give the prediction of next year toward restaurant reputation.
In the cloud world, users will see positive words in bright color and negative words in gray color.
In the line chart, users will see years on the x-axis and positive comment levels on the y-axis. Besides, it also shows some tags of the restaurants. All the tags come from the review of the previous. And the tags that we show are the most common tags from users.

Motivation**:**
We are smart computer, science students. We are disdain to use regular ways to review comments on yelp since it has thousands of comments. As we are computer science students with a strong mathematical background, we are eager to use a way to simplify the decision process. This thought motivate us on creating this project.

Prior Work
(How does your project relate to previous works. Please give a short summary on each paper you cite and include how it is relevant)

Using sentiment analysis to analyze the yelp data, with the following tutorial material (https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk)

Improving the accuracy of sentiment analysis:
(https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.660.5055&rep=rep1&type=pdf)

Model/Algorithm/Method
(This is where you give a detailed description of your primary contribution. It is especially important that this part be clear and well written so that we can fully

understand what you did)

We will use term-document incidence matrix, inverted index, permuterm index, spelling correction, Solr, Django, Lucene, sentiment analysis model
After task 2, we could know the implementation details of our model and method.

Results and findings
Expected result: the prediction of the future good reviews and bad reviews of the restaurants.
Showing the top 5 tags of the restaurant to users.