Capstone Project - The Battle of Neighborhoods

1. A description of the problem and a discussion of the background. (**15 marks**)

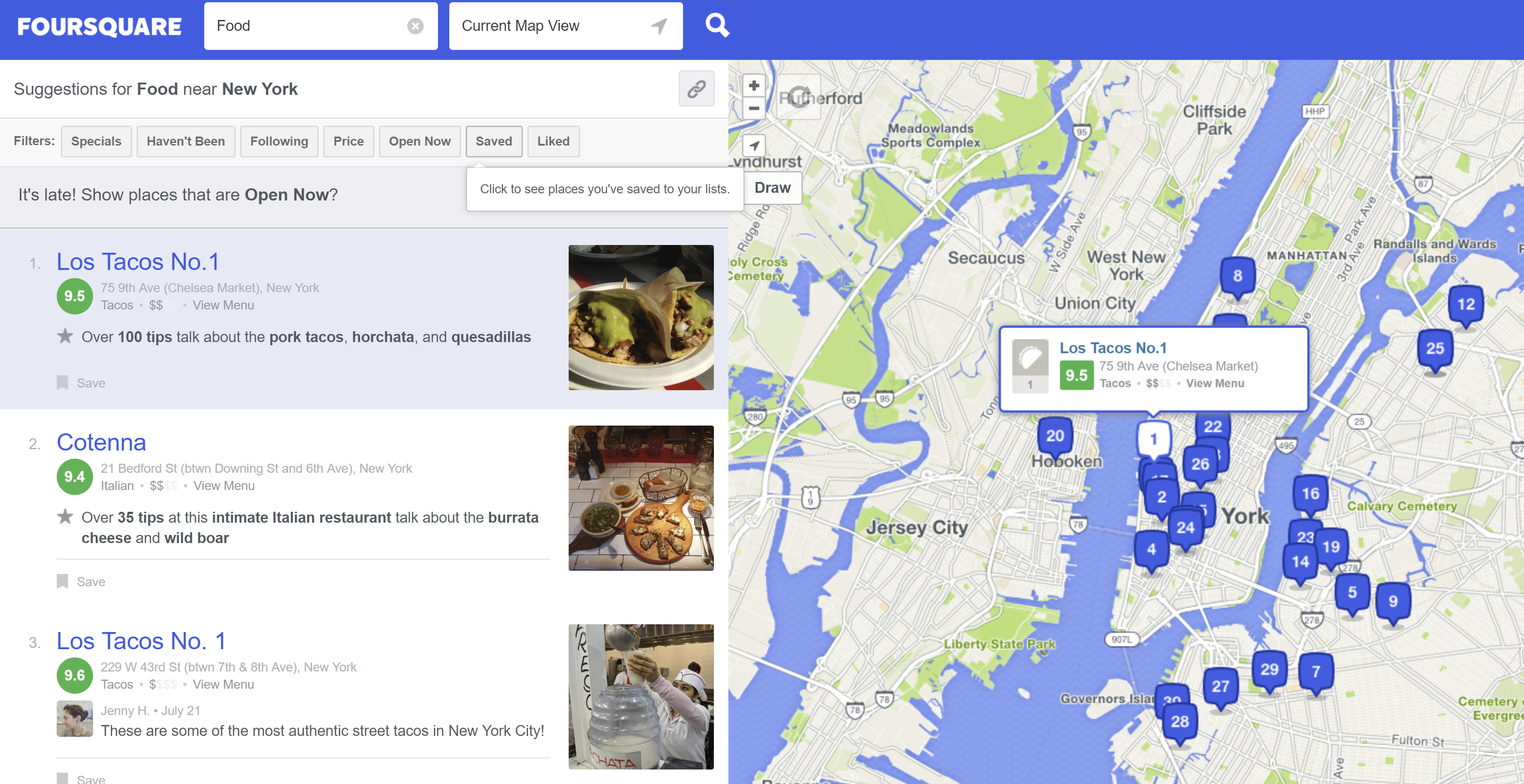
With the help of Four Square, people can learn which venue is the best option for them to eat dinner or have fun. Although venues on Four Square have been categorized into different filtering tags and scored by previous customers, people are still likely to review venues’ tips. In other words, reviewing tips is a common method by which people can learn the venue thoroughly and from various perspectives.

However, reading tips is very consuming because the tips are unstructured. It is not that easy to extract useful information from a bunch of tips quickly and correctly. In this project, a data visualization method – Word Cloud is applied so that the key information in tips can be presented in an effective and apparent way.

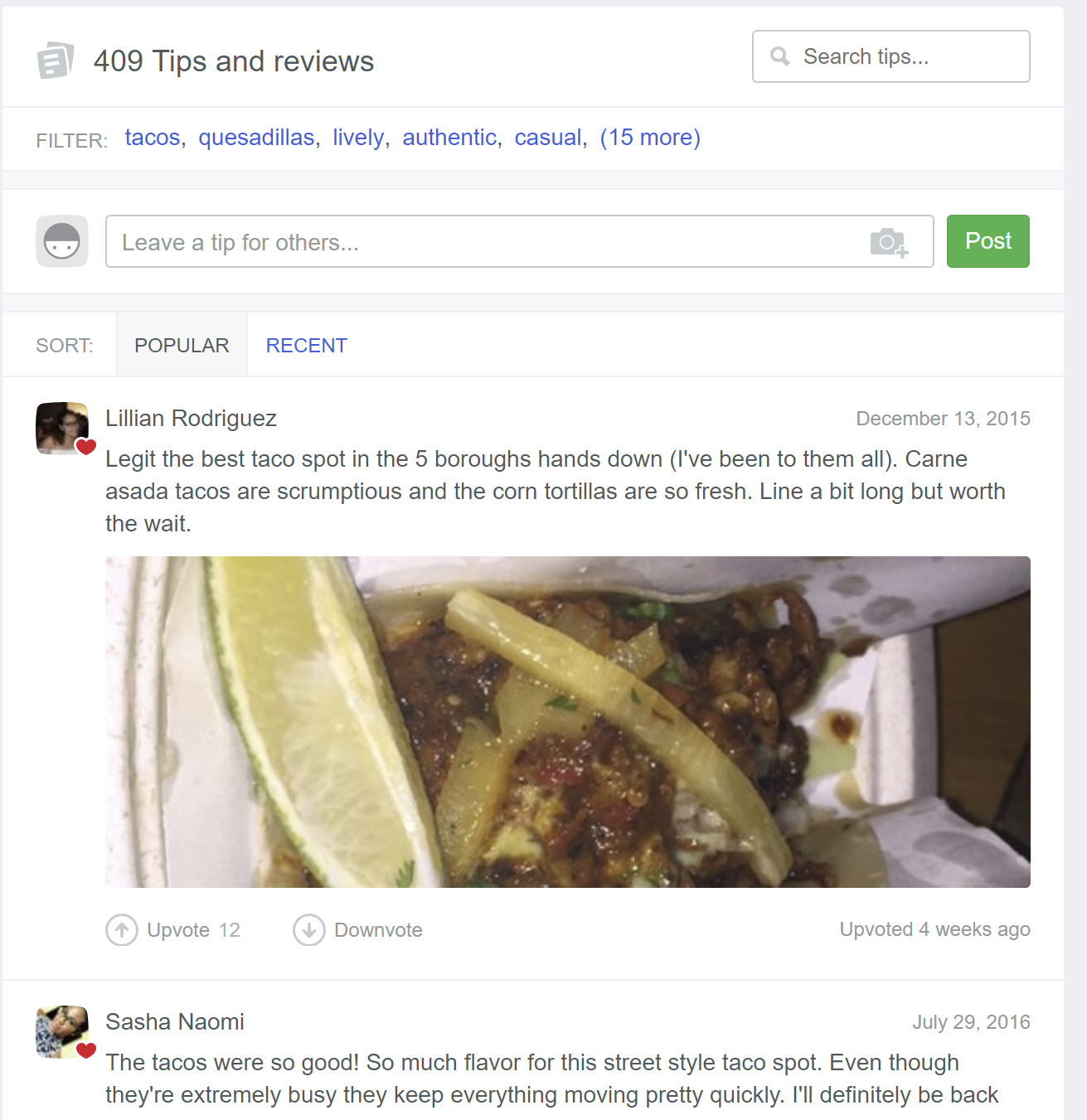
1. A description of the data and how it will be used to solve the problem. (**15 marks)**

The data I will use in this project is the venues’ tips in Manhattan.

First, I will get the top 50 rated restaurants in Manhattan.



Then, I will collect the 10 most popular and 10 most recent tips for each restaurant.



Since there are photos in some tips, I will extract the text only and use the text for the following language processing. The tips collected here will be categorized into two groups: “like” or “neutral/dislike”.

For each group, I will generate the word cloud figure individually. From the two figures, people can learn the advantages and disadvantages of the venue intuitively.