Scenario

You and your friends have started a company for business data analytics. Your first customer is Mr. Smith, who is an entrepreneur and runs a restaurant. You offer Mr. Smith the following data-driven consulting service: You will analyze the online reviews of the restaurant and its major competitors. You will then be able to offer advice on how to improve the services and future business plan based on your results.

The review data is provided in an Excel file: review.xlsx. Assume the rid of Mr. Smith's restaurant is 1. The dataset contains more than 50,000 reviews for more than 900 restaurants. The data can be downloaded from my.wbs.

The dataset was collected from yelp.com. You may refer to any review on the website (for example, https://www.yelp.co.uk/biz/the-pig-and-tail-birmingham) for the definition of the data fields. Briefly,

user_name: the user name of the reviewer, it is not a unique user ID.

user_numberoffriends: the number of friends the reviewer has at the time the data was collected.

user_numberofreviews: the total number of reviews the reviewer had written on yelp.com.

rating: the rating given by the reviewer.

date: the date of the review.

numberofuseful: the number of votes given by the people considering the review helpful.

numberoffunny: the number of votes given by the people considering the review funny.

numberofcool: the number of votes given by the people considering the review cool.

reviewcontent: the text content of the review. Requirements:

1. All codes must be implemented using Python.

2. Codes must be well documented with comments. It should also have a separate readme.txt file describing what each file (script, data file and etc.) does, how to run the codes, how to organize your files, as well as a brief summary (no more than one page double spaced) based on your analysis.

3. You need to specify the functionalities you try to accomplish in the documentation. 4. You may search online or discuss with other students, but each group must work independently. Notes: Additional Python packages (not covered in class) are welcomed to use. But they should be well documented.

A simple user interface is recommended. Do not overcomplicate the problem. Your primary goal is to demonstrate your understanding of python programming.

Be creative.

Clarification:

1. You don’t have to use all the data fields for your analysis.

2. You should have at least two files in the submission: one script file and one readme.txt file. You may have more than one script file, depending on your program structure. The readme file should contain the technical instruction/explanation as well as the nontechnical summary based on your analysis. You may have the summary as a separate file.

The summary itself should be no more than one page double spaced.

3. You may make any reasonable assumptions and definitions for your analysis. 4. The recommended user interface should be graphic based, such as webpages or windows. 5. For this project, the primary goal is to demonstrate programming skills. Therefore, you should aim to use all the following techniques in solving the problem, listed roughly in the order of importance:

a. Conditional flow

b. Loop flow

c. Functional design

d. Complex data types

e. File operation

f. Exception handling

g. Installation and use of third-party modules.

h. Graphic user interface i. Commenting/documentation