CIS 8045 Term Project

Project Team and Due Date

You can do this project in team. **A team should contain no more than 5 people.**

Project Data

The project data is a set of real customer review data collected from Skytrax (www.airlinequality.com), a major website for customers to evaluate various airline companies. This dataset contains 41,396 review entries, with each entry capturing the rating of a particular airline company by a particular author. In this context, **an author is also a rater as well as a customer**. The data is stored in a MongoDB database. A sample document (corresponding to a review entry) is as follows:


```
{
        "authorcountry" : "Israel",
        "rating_valuemoney" : 5,
        "recommended" : 1,
        "airlinename" : "el-al-israel-airlines",
        "travellertype" : "FamilyLeisure",
        "cabin" : "Economy",
        "aircraft" : "Boeing 747-400",
        "rating_overall" : 9,
        "reviewcontent" : "Flight was half-full. I had the whole row to myself. ...",
        "rating_inflightEnt" : 4,
        "reviewdate" : "7/4/2015",
        "rating_cabinstaff" : 5,
        "route" : "TLV to JFK",
        "rating_seatcomfort" : 4,
        "rating_foodbeverage" : 5,
        "authorname" : " Moam Ben-Shalom"
}
```


The overall rating is on a scale of 1-10, with a higher level indicating more satisfaction. The ratings of individual aspects (e.g., cabin staff, seat comfort, food & beverage, etc.) are on a scale of 1-5. The property **"recommended"** indicates whether or not the author would like to recommend this airline to others. Other property names are mostly self-explanatory.

You can download the data file **AirlineReview.json** from iCollege, and use ***mongoimport*** to import it to MongoDB Atlas. Then you can use it.

Project Tasks

Three are three main tasks in this project:

First, you need to develop text-analytics solutions for the following problems:

a. Once you store the MongoDB database, use Python to access. This part can be done using PyMongo.

b. Retrieve customer reviews from the database, and develop sentiment analysis of reviews.
c. Construct a rating network between authors and the airline companies that they rated. Such rating network captures who rated which airline companies. Construct this network in Neo4j.
    - There are multiple ways to do it. You can use any way that you prefer. For example, you are encouraged to explore using Neo4j Python Driver to access Neo4j (A brief tutorial can be found here: https://towardsdatascience.com/neo4j-cypher-python-7a919a372be7). Alternatively, you can use Python to write to a CSV file and import the CSV file to Neo4j.

What to Turn In:

Turn in two documents:

- A MS Word document explaining what you do. When necessary, attach the Python scripts or screenshots for demonstration.
- The complete Python script file.

You can make your report as professional as you can (e.g., using tables, plots or other visualization approaches for results presentation), but you do not have to worry too much about "decorating" your report.

Second, you need to turn in your complete Python code. Use sufficient in-code comments to make your code readable. Make sure that your code is executable, so that others should be able to replicate your solution results using your code.

Grading:

Since this is a project, the solutions are fairly open. Our objective is to solve problems, and therefore there is no absolutely correct answers. The grading will be based on the extent to which your proposed and developed solutions demonstrate that you understand the course materials and are able to apply them.