Data Analytics and Visualization (Fall 2019)
Rutgers Business School-Newark and New Brunswick
Instructor: Debopriya Ghosh

Assignment 2.
(October 23, 11:59 PM)
Late submissions will be penalized**

Exercise 1. (50 points)
Using the *Olympic Athletes* dataset, answer the following questions:
(a) How many medals have athletes won since the 2000 Games?
(b) How has the number of gold medals changed over time?
(c) Which countries have won the most number of medals in swimming?
(d) How many each of gold, silver and bronze medals has Michael Phelps won?
(e) Create a map view of the total medals won by each country.
(f) Design 3 different views of data of your own design.

Exercise 2. (50 points)
As the new lead analyst for the New York Citi Bike Program, you are responsible for overseeing the largest bike sharing program in the United States. You will be expected to generate regular reports for city officials looking to publicize and improve the city program. Since 2013, the Citi Bike Program has implemented a robust infrastructure for collecting data on the program's utilization. While the data has been regularly updated, the team has yet to implement a dashboard or sophisticated reporting process. City officials have a number of questions on the program, so your task is to build a set of data reports to provide the answers.

Your task in this assignment is to aggregate the data found in the Citi Bike Trip History Logs by merging the datasets from different periods. Next, you will be required to answer the following question using the merged datasets.

(a) How many trips have been recorded total during the chosen period?
(b) By what percentage has total ridership grown?
(c) How has the proportion of short-term customers and annual subscribers changed?
(d) What are the peak hours in which bikes are used during summer months?
(e) What are the peak hours in which bikes are used during winter months?
(f) On a particular date, what are the top 10 stations in the city for starting a journey? (Based on data, why do you hypothesize these are the top locations?)
(g) On a particular date, what are the top 10 stations in the city for ending a journey? (Based on data, why?)
(h) On a particular date, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?)
(i) On a particular date, what are the bottom 10 stations in the city for ending a journey (Based on data, why?)
(j) On a particular date, what is the gender breakdown of active participants (Male v. Female)
(k) How effective has gender outreach been in increasing female ridership over the timespan?
(l) How does the average trip duration change by age?
(m) What is the average distance in miles that a bike is ridden?
(n) Which bikes (by ID) are most likely due for repair or inspection in the timespan?
(o) How variable is the utilization by bike ID?

Your data and analysis needs to be presented in a way that is focused, concise, easy-to-understand, and visually compelling.