Analytics In Practice (MIS 64038)
Assignment II – Data Visualization

Total Points: 100

Exercise 1:

An Airport manager is receiving complaints from customers regarding the delays in departure and arrival of flights. He thinks that there are hardly any delays and even if there are any delays, it is mostly due to weather. But he is unable to convince his supervisor. He has now hired you, a Data Scientist, a graduate from Kent State University. Your job is to explain to the Airport manager about the overall delays at the Airport. You have been provided with complete air traffic data to analyze. Your task is to analyze the data and provide insights about the delays. Following are the set of questions, he is looking for an answer from you. However, you are not restricted only by these questions (extra bonus money(points) will be provided if you can provide answers beyond these questions).

1. What is the pattern of arrival traffic and departure traffic delays with respect to days and weeks?
2. Can you interpret the traffic delays?
3. Which Airport ('Origin Airport') has highest departure delay?
4. Which Airport has highest Arrival delay?
5. How do you relate the delay pattern to the distance travelled?
6. Is there any correlation between weather delay and carrier delay?
7. What is the delay pattern you can find in respective states?
8. How many delayed flights were cancelled? (approximation)
9. How many delayed flights were diverted? (approximation)
10. What time of the day do you find Arrival delays?
11. What time of the day do you find Departure delays?

Instructions:

Please answer all questions. You should use R (ggplot2() package) and include the screen shots and word/pdf file along with your of your R-notebook in your submission. **NO GITHUB links. You will get ZERO, if only GITHUB link is provided. Your document should have answers for all the 11 questions in the same order.**

For this assignment, you need to use the 'Airport_delays.ZIP' dataset which can be downloaded in ZIP format from the Dataset folder. The data contained in the compressed file has been extracted from the

Marketing Carrier On-Time Performance (Beginning January 2018) data table of the "On-Time" database from the TranStats data library. (Ref: https://www.transtats.bts.gov/)

1. I have sampled only one state, **CALIFORNIA**, in the dataset that I have provided for this assignment. However, the full data set can be downloaded from website. **All your answers are for the only ONE state, CALIFORNIA.**
2. Data description "readme,html" is provided to understand the data variables. Please read the readme.html before you start the assignment.
3. All work to be done using ggplot() library
4. Experiment with the color, shape and size aesthetics.
5. All the plots/graphs should be labelled appropriately
6. There can be multiple plots/graphs for each of the below questions. You can do as many graphs possible. You will be given extra credits for the extra efforts.
7. Your presentation is very important for this course.
8. All your work may be posted on your **github or website for your reference only**.