Data Science Exercise:

The candidate will work on the data science exercise at home using any tools / computing platform of choice. Where necessary, please document your assumptions for the analysis. We mainly look for the thought process and approach used for solving business problems using data sciences skills. You can provide a solution with any format of your choice. You can choose to use any or none of the questions listed below (basic & open) for your analysis.

Data:

This exercise is based on the public NYC taxi data. The data set used in this exercise is a one month snapshot.

- The data is available at the following link:
  https://drive.google.com/file/d/0B_eZ1n8fw-T6SGp4bXJTaUF3Sm8/view?usp=sharing
- The data dictionary is available at the following link:
  https://drive.google.com/file/d/0B_eZ1n8fw-T6VVdmbVRMWmxyUlk/view?usp=sharing

Questions:

Below are two set of questions for the candidate. The candidate is free to work with a sample or use the entire months' data. You can choose to answer any or all of the following questions. If you choose not to answer any of the questions below, please provide the reason. In that case, you can build an analysis of your choice, but it has to be based on the data provided.

Basic Questions:

1. What are the distributions of the number of passengers per trip, payment type, fare amount, tip amount, and total amount?
2. What are top 5 busiest hours of the day, and the top 10 busiest locations of the city?
3. What is the hourly taxi activity for each day of the week?
4. Which trip has the most consistent fares?

Open Questions:

a. Can you predict the fare and tip amount based on the pickup / drop off location, time, and day of the week?
b. Can you predict the pickup / drop off geographical distribution for each hour of a weekday?
c. If you were a taxi owner, how would you maximize your earnings in a day?
d. If you run a taxi company, how would you maximize your earnings?

Notes for the candidate:

- Please document your work and provide the results along with commented code.
- We will not use the candidate's work for any commercial purpose outside the interview process.
- This is not a competition using NYC taxi data.
- You earn bonus points for "Story Telling": well communicated and presented results.