NYC Taxi Data Challenge

Data sources:

- NYC taxi data (described and available here: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page , also available either through BiqQuery https://bigquery.cloud.google.com/table/imjasonh-storage:nyctaxi.trip_data, or from http://www.andresmh.com/nyctaxitrips/)
- NYC borough polygons: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Questions:

1. Propose a metric or algorithm to assess the potential efficiency of aggregating rides from many vehicles into one, given the available data sources. Make realistic assumptions and necessary simplifications and state them.

2. Implement your proposed metric, and evaluate Manhattan's overall efficiency using **yellow taxi data from the first full week (Monday-Sunday) in June 2016**. Discuss the complexity of your implementation.

3. Based on the implementation in Question 2, use visualizations to show how the efficiency metric varies across different time of day, and day of the week. Discuss potential business implications based on your findings.

4. **Answer the following question theoretically (no need to implement)**. Given the metric you proposed, how would you determine whether a specific zone is more efficient than the rest of the city? How would you determine why it's more efficient?

You will be assessed on:

- The readability of your presentation
- The creativity and validity of your approach
- The quality of your code