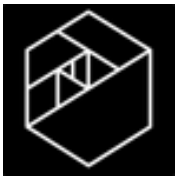

Metis Take Home Practice

We will be using New York City CitiBike data for this analysis. CitiBike is New York City's public bike sharing program which opened in May 2013. Bicycles are rented and then returned to various stations located throughout Manhattan, Brooklyn, and Queens. Memberships and various time-limited passes are available for purchase.

Historical CitiBike usage data can be downloaded here.

Python libraries should be used to answer all questions except the last one, where any tool(s) can be used. Your analysis should be performed and annotated so that it could easily be handed off to another data scientist in the group. This exercise is designed to require 4 – 8 hours to complete.

Question 1

- Programmatically acquire, load, and prepare CitiBike data for the year 2015.
- How many data points are there in the data used for analysis?

Question 2

- Plot the distribution of the trip duration.
- Describe the structure and hypotheses you have about that structure.

Question 3

- Make a plot of the total number of trips per hour of the day.
- What does this plot look like if days are broken into weekdays vs weekends? What does this usage pattern tell you?
- Find another interesting way to segment all or some of the data. Discuss your findings.

Question 4

- There are instances where a bike's next trip originates from a different station than it was returned to for its most recent previous trip. What do you think causes this?
- Explore this phenomenon.

Question 5

- Build a predictive model(s) for the total number of daily CitiBike trips. This model should include at least some features from an external data source, *i.e.* that is not included in the CitiBike data.
- Evaluate your model(s), discuss pros/cons, and alternative approaches.

Question 6

- Using the tool of your preference, develop a visualization that displays something interesting about the data.