Task and Logistics

You will have 7 days to complete this task, but we expect you to spend no more than 3 to 4 hours on it. Walk us through your assumptions and thought process, as well as your modeling and tool selection decisions.

Data

For this exercise, the following datasets are available:

1) 311 Service Requests from 2010: https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9 (~10GB as of November 2017 so prepare for a large data transfer)
2) DOHMN NYC Restaurant Grades Results: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

To get started, feel free to download your data using `curl`, as follows:

311 service requests:

```
#!/bin/bash

curl \
    "https://nycopendata.socrata.com/api/views/erm2-nwe9/rows.csv?accessType=DOWNLOAD"
\
    --output 311_service_requests.csv
```

Restaurant grades:

```
#!/bin/bash

curl \
    "https://data.cityofnewyork.us/api/views/43nn-pn8j/rows.csv?accessType=DOWNLOAD" \
    --output restaurant_inspections.csv
```

Questions

1. Answer the following questions. For each question, please provide the code that you used to analyze the data and generate your results. We're really interested in understanding your thought process; please explain your motivation for using any specific algorithm/tool/library.

    a. What are the top features that affect the time to resolve a complaint? Do they change over time?
    b. Are rat sightings more common in areas that have a higher density of restaurants with low inspection grades?

2. Additional Questions:
    a. Ask 2 additional questions you would like to answer with this data.
    b. For each question, briefly discuss which additional data would be necessary, as well as how you would tackle it.
    c. Select one of your questions and dive deeper!