Capstone Proposal
311 Service Requests

Briefing:
I am a data scientist employed by MODA – The Mayor's Office of Data Analytics. Given the recent political climate and global health crisis, the mayor and city council need to ensure the happiness of New Yorkers now more than ever. Thus, I've been tasked with gathering information to find out how the city's various departments can more quickly and efficient resolve the people's woes. This will help the city officials allot funds and make decisions that affect basic infrastructural services.

Problem:
311 is a hotline system that captures non-emergency service requests made by the city's residents, including but not limited to noise complaints, unsafe housing conditions, air quality issues, unsanitary restaurants, etc. 311 has come very far in the past 17 years. Gone are the days where human operators had to manually input a request ticket when a resident phoned in a complaint. Today, New Yorkers can text the hotline, visit the 311 website, or use the mobile app in hopes of having their concerns heard. The requests are then forwarded to the relevant agencies, such as the NYPD, Department of Sanitation, Transportation, Housing Authority etc. The agency responds to the request, then closes it in the 311 system once resolved.

But many residents have never used 311 with the thought that the problem will never get fixed, so there's no point in using the service. This could change if New Yorkers were reassured that their request or complaint would be handled in a reasonable and realistic timeframe. This project will focus on response time of a request – that is, how long does it take for an agency to complete and close a specific type of complaint for a specific area of the city and ultimately answering the question: Should I call 311?

Data
In order to answer this question, I will use 311 data provided by NYC Open Data, a website that provides free public data published by NYC agencies and partners. The 311 data set is updated every day and contains over 20M rows and 41 columns that timestamp, geo-tag, categorize, assign an agency and create a close date for all service requests from 2010 to Present.

Approach
A subset of this data will be used to create a model with the goal of predicting a close date for a given complaint in a specific location based on several input variables taken at the time of the call (i.e., predict the time necessary to complete a particular request). This is a supervised learning regression problem so Classification models will be necessary: logistic regression, decision tree (possibly gradient boosting) and random forest. Before a model can be created the following steps must be taken:

- Subset the data and remove least useful data (e.g. rows where close date < created date)
- Pick target feature

- Answer the following question:
    1. What are the different types of complaints/service requests?
        i. Which is most common?
        ii. Which are least common?
    2. Who is submitting the request?
    3. Where are the requests based?
        i. Which borough?
        ii. Which neighborhood?
        iii. Which zip code?
    4. When are these requests made?
        i. Most/least common day of the week?
        ii. Most/least common time of day?
        iii. Most/least common time of year?
    5. Which agencies are more efficient in closing requests?

Once these questions have been answered, the model(s) can be created and evaluated for accuracy using an error matrix. The final deliverables of this project include: a GitHub repo containing the code, a slide deck and a project report.