# Predicting the Location of Crime in New York City Based on Historical Data

**Scenario:** You are an aspiring data analyst at the New York Police Department. Your manager has asked you to read the provided documents and gain context for the problem. Following this, you are to analyze the provided data set on shooting incident data since 2006, and subsequently create a model to predict future incidents.

**Deliverable:** The deliverable for this project is a created, working model in either Python or R and an adjunct overview presentation. The model will not be graded on accuracy, simply that it works and the outputs make sense. The presentation should focus primarily on the high level of what was found through context analysis, what model was used, alongside the types of metrics used for the model. The presentation does not need to delve into the intricacies of the model, or how you arrived at the model you chose.

**Due Date TBD**
Submission Format: Github link uploaded repo to Canvas
*Individual Assignment*

**General Description:** Submit to Canvas a link to your case study repository that contains the source code for your model in Python or R, alongside your presentation.

**Preparatory Assignments –** Everything in the course up to this point.

**Why am I doing this?** Case studies provide an opportunity to apply what has been learned throughout your time in the data science minor. This case study provides a dataset with a medium amount of data that is relatively clean and will allow you the chance to focus on your abilities with model creation, rather than cleaning and preparing data.

· **Course Learning Objective:** Prepare findings for presentation to your peers
· **Course Learning Objective:** Create a time-series analysis model
· **Course Learning Objective:** Understand a real-world usage for data science

**What am I going to do?** In this assignment, you will begin by becoming familiar with the project domain using the link provided. This previous presentation will begin to give context to the data and give some potential biases that will arise in the model. You will then read the other provided article, that describes one way to create a location-based prediction model. Following this, you become familiar with the data itself and brainstorm different possible models outside of the one provided in the article, along with choosing one. Finally, you will create your model and the presentation.

**How will I know I have succeeded?** You will follow the rubric below and complete the assignment to its standards.

| Category | Details |
| --- | --- |
| Formatting | *Repository*: Online Github repo containing all project information <br> *Contents*: <br> - README.md <br> - SRC Folder <br> - Figures <br> - License |
| README.md | Overview of the repository. Gives context to the project for anyone who visits. <br> Should contain: <br> - goal of the project <br> - data table and explanation <br> - model used and analysis conducted |

| | |
|---|---|
| |   - figures and table of contents for other folders<br>  - References (minimum the two provided) |
| Figures | Create a minimum of one visualization that gives context to the data. This will help the audience in becoming familiar with the data quickly. It is recommended to create 3+ (with at *least* one being a geo-visualization atop a map of New York) |
| SRC | This folder should contain any source code used for your project. At minimum the file you used to create your model, but also any scripts used to clean or manipulate data.<br>**NOTE:** you will not be graded on the accuracy of your model, but rather that it gives reasonable outputs to the data it is provided. Additionally, there was a genuine effort to apply data science fundamentals in the project. |
| License | Explain the terms and conditions of the repository. MIT license is standard. |
| Presentation | Using the information from the README.md file, along with the figures create a presentation to give to the class. As mentioned on the first page, this presentation *is not* about describing the details of your model, but rather providing a high-level overview. You have a lot of freedom with this presentation, make it interesting! |