

## BDIPlus Inc. DS take home challenge Set 1 Level 1

The data set “complaints.csv” consists of all rodent complaints made on September 2016 in New York City.

*Source: NYC Open Data.*

Allowed Resources:

- Attached Data Dictionary
- Guidance online for open source tools

Requirements:

- Use of Python

Task:

- 1) Review the data carefully. And hand in a report with answers from each step and briefly explain your approach.
- 2) Your goal is to create a model that accurately predicts the three types of rodent complaints.\*
  - a. Mouse Sighting
  - b. Rat Sighting
  - c. Signs of Rodent

\*there is a 4<sup>th</sup> one called “condition attracting rodents” you may delete the records that have this descriptor or tag it as a 4<sup>th</sup> prediction.

You may ignore any other types of rodent complaints.

- 3) While creating your report, please try and follow the following format:
  - a. Describing statistics
  - b. Primary Data exploration
  - c. Choice of Analysis
  - d. Variable exploration
  - e. Model exploration and testing cycle
  - f. Choice of model

Your model will be tested against the test data that is with your interviewer.

Enjoy!