# Geospatial Analysis and Prediction of Rodents in New York City

## Focus:

To discover the hot-spots of rodent incidents and epicenters of spread of vector-borne diseases in New York City and predicting the trend of future incidents to take proactive control measures.

## Aim:

Performing Geo-Spatial Analysis on the spatial data and the occurrence of rodents in that location over past 10 years (2009-2019). Performing ARIMA algorithm over timestamped data to predict the trend of occurrences in the next 2 years (2020-2022).

## Problems:

* The raw data is huge numbered and consisted of individual randomly timed data (no periodicity).
* Locations (Latitude and Longitude) was pointing very specific location making it harder to draw insights about an area.
* Outliers existed, there were a few months with almost no sightings of rodents which is unusual.
* Geo-Spatial mapping of data and time based visualization of incidents.

## Approach:

The Problem is broken down into modules which can be individually considered as deliverable. We had 3 main jobs: Data Cleaning and Preprocessing, Geo–spatial Analysis, Prediction. Problem breakdown is a below.

Processing stage

Deliverable

## Problem Solution:

* Aggregating the data based on time stamp and inducing periodicity (Overall, Yearly and Monthly aggregation based on location).
* Strategic rounding of location points resulting in grouping of nearby location points.
* Removal of outliers post statistical analysis (used box-plot and data description).
* Mapping Geo Spatial data using **folium** and **pandas**.
* Implemented **season decomposition** on data to extract trend and seasonality.
* Used **ARIMA** for prediction of incidents for 2 years (**SARIMAX** from statsmodel API).

## Results:

|  |  |
| --- | --- |
| Geo Spatial Analysis: | Trend and Seasonality: |
| Testing ARIMA fitted model: | Predicting Future Trend of rodent incidents: |

## Analysis Report:

From the above analysis, we can draw following insights

* The rodent occurrences are peaked near the Brooklyn, Empire State and The Bronx regions. The Northern and Central New York City are largely affected by rodents in the last 10 years.
* Although there is a decrease in incidents from 2013 to 2016, a significant increase in incidents after 2017 is observed. This can be inferred from analyzing the trend plot.
* The incidents generally attain peak value in Spring season of the year (First Quarter). If a time frame right before this season is targeted for eradication, their population can be drastically controlled. This can be inferred from the seasonality plot.
* The prediction shows that the trend and seasonality continues and the direction of incidents will increase if no measure is taken.