Final Project Proposal

Zhenkun Fang, Weiqi Liang, Shiyu Zhang, Zeqi Li

2024-11-08

Team members

- Zhenkun Fang (zf2352)
- Weiqi Liang (wl3011)
- Shiyu Zhang (sz3319)
- Zeqi Li (zl3545)

Tentative project title

Barking Up the Right Tree: Analyzing Trends, Geographic Hotspots, and Contributing Factors for Public Safety

Motivation

Dogs have always been known as "people's best friend," playing an important role in American households. However, while they bring us joy and companionship, we cannot overlook the rising number of dog bite incidents. In fact, data from 2022 shows that dog bites are among the top 15 reasons for nonfatal emergency department visits across all age groups, highlighting a serious public health issue.

This project aims to investigate the complex relationships between dog bite incidents and various factors, such as dog characteristics, geographical distribution, and trends over time:

- **Geographical Distribution:** By examining dog bite rates in different zip code areas, we hope to pinpoint potential hotspots and explore whether there is a correlation between the frequency of incidents and the density of dog populations in those areas.
- **Dog Characteristics:** We'll look into whether specific traits of dogs—like breed, age, gender, and whether they are neutered—are linked to how often bites occur. This analysis could help in developing targeted strategies to reduce the risk of dog bites.
- **Trends Over Time:** Additionally, we will investigate how the rates of dog bites have changed over time, particularly from 2015 to 2021.

We hope this analysis could help in developing targeted strategies to reduce the risk of dog bites.

Intended final products

The expected final website will consist of the following parts:

Introduction page: Demonstrate the background of our study, display screen cast, pictures of dog breed in the dataset, a general map illustrating where dog bites occurred.

Statistics analysis: This page will introduce the statistical methodologies used to analyze the relationship between dog bite occurrences and dog demographics. It will also present the results and diagnostic assessments.

Visualizations plot: This section will include charts, graphs, and maps showcasing the geographical distribution of dog bites in New York. It will display occurrences across different dog breeds, sexes, and ages, and use facet plots to show trends over different years and boroughs. It will also compare the distribution of dog bites with the NYC dog licensing data.

Meanwhile, we will prepare a detailed analysis report on the distribution of dog bites in New York, examining contributing factors such as sex, age, breed, and licensing status.

Data sources

Dog bite incidents

A dataset of dog bite cases in New York City for different genders, breeds, and ages. Data were taken from NYC Open Data for 2015-2021.

Dog license

A dataset of dog license records that were active during 2014-2023. Data is sourced from the DOHMH Dog Licensing System, where owners can apply for and renew dog licenses.

Planned analyses, visalizations, and coding challenges

I. Correlation between dog bite incidents and the background of dogs

1.1 Trends in annual number of dog bite incidents

A combined bar and line chart illustrates the fluctuations in dog bite incidents in NYC from 2015 to 2021, with the years on the X-axis and both the frequency and rate of incidents plotted on the Y-axis. This dual-chart approach succinctly captures the yearly variance in the occurrences of dog bites, offering a clear visual representation of trends over the specified period.

1.2 Correlation between the number of dog bite incidents and dogs' background

Aggregate data from 2015 to 2021 will be used to analyze the various background characteristics of dogs, including gender, breed, age, and neutering status, to determine their correlation with the number of dog bite incidents.

II. Correlation between dog bite incidents and the geographical distribution of dogs

2.1 Spatial distribution of dog bite incidents

Display the spatial distribution of new dog licenses and dog bite incidents in NYC from 2015 to 2021. On a map of NYC, use color blocks to show the relative geographic increase in dog populations and mark the total number of dog bite incidents in each zipcode with dots. By comparing the map's background colors with the sizes of the dots, conduct a preliminary analysis of the spatial correlation between the increase in dog numbers and dog bite incidents.

2.2 Spatial characteristics of dog bite incidents

Using the Global Moran's Index, assess whether the spatial distribution of dog bite incidents in New York City exhibits spatial correlation and clustering characteristics.

III. Cumulative impact of various factors on dog bite incidents

Use negative binomial regression (or Poisson regression, depending on test outcomes) to measure the combined effects of year, dog background (i.e., gender, age, and neutering status), and the borough rate of dog

bite incidents. The dependent variable is the incidence rate of dog bites, adjusted for the population in each zipcode. The predictor variables include year, dog background, and the area of the incident. The year is measured as an ordinal variable ranging from 1 (corresponding to 2015) to 7 (corresponding to 2021). To capture potential curvilinear effects of the year on the incidence rate, a multiplicative term created by squaring the year variable is also included in the analysis. The area of the incident is categorized into a variable with values from 1 to 6, representing the five boroughs of NYC plus others. The dog gender is a binary variable, with 0 indicating unknown, 1 for male, and 2 for female. Similarly, the neutering status is a binary variable, with 0 indicating not neutered and 1 indicating neutered. The age variable encompasses several categorical intervals.

Timeline

- Nov 03 - Nov 09:
  - Choose a project topic
  - Familiarize with the datasets
  - Complete proposal *(due 11/08)*
- Nov 10 - Nov 16:
  - Tidy up the datasets
  - Draft the data, motivation, related work, and initial questions sections of the report
  - Get preliminary visualizations
  - Meet with a course staff to review the project
- Nov 17 - Nov 23:
  - Finalize visualizations
  - Draft the exploratory analysis, additional analysis, and discussion sections of the report
- Nov 24 - Nov 30:
  - Draft website
  - Draft screencast outline
- Dec 01 - Dec 07:
  - Record screencast *(due 12/07)*
  - Finalize website and report *(due 12/07)*