The Brave Ducks
Patrick Pei, Numair Ahmed, Danhua Lu
10/01/2021

Project Proposal

Background and Theme

The police violence has become a very heated topic. It is speculated that in different demographics, lower-income neighborhoods have been unproportionately affected by policy violence. The number of officer allegations per capita is a way to comparatively measure the police allegations per capita in a low, middle, and high-income neighborhood. Our definitions for different income neighbors are following:

- a **low-income neighborhood** is a community that has less than $30,000 of median income,
- a **middle-income neighborhood** is a community that has a median income between $30,000 - $75,000,
- a **high-income neighborhood** is a community that has a median income of over $75,000.

We want to use data to observe any correlation between the number of officer allegations filed to the income demographic of the neighborhood in which the misconduct occurred from the beginning to the most recent allegations (with the most recent date as of this writing in 2021). We want to see if there are any trends that are worth noting.

Research Questions

To evaluate the trend of change for police allegations, we will further differentiate the "income demographic" of a neighborhood to be designated as low-income (less than $30,000/yr), middle-income (between $30,000/yr and $75,000/yr), and high-income (greater than $75,000/yr).

Relational Analytics

- What is the number of people living in **low, middle, high** income neighborhoods?
- Using our definition of types of “income neighborhood", what is the total number of officer allegations for all **low, middle, high income** neighbors?
- What is the rate of increase for officer allegations for **low, middle, high** neighborhoods between 2002-2007 and 2007-2012 timeframes, 2007-2012 and 2012-2017 timeframes?

- What is the percentage of misconduct allegations (drug/Alcohol, illegal search, use of force, etc) out of all allegations for these **low, middle, high** neighborhoods?
- Among the officer allegations with complaints filed in the **low, middle, high** neighborhood, what percentage of the cases are dismissed?

Data Exploration

- Symbol Map (or Filled Map): We would like to obtain a clearer relationship between socioeconomic status and officer complaints, misconduct allegations, and the areas raised complaints on geographical factors.
- (Horizontal) Bar chart: We hope to draw a more visual report of socioeconomic status and number of officer allegations in different areas across a certain timespan. The bar chart is capable of representing the total number of officer allegations of an area per year.

Interactive Visualization

For this section, we are interested in finding more answers via interactive approaches for questions 3 and 4.

- To better answer the percentage of misconduct allegations (illegal search, use of force, etc) for the **low, middle, high income** neighborhoods, we can use Zoomable Circle Packing for our theme because it allows us to visualize these misconduct allegations more interactively for each of the neighborhoods. i.e. focus on the portion of majority types of misconduct allegations among different economic status neighborhoods, or vice versa.
- For our last question regarding the percentage of the **dismissed** cases for each neighborhood, we can use interactive Treemap to represent these cases. It will give us a better understanding of how many cases are dismissed.

Graph Analytics

Since we have identified police allegations with regards to three different income neighborhoods, we will analyze the relationship between the mentioned areas.

- is a high-income neighborhood area more likely to have less police misconduct allegations per capita than a low-income neighborhood area across all three time spans?
- is a low-income neighborhood area more likely to have more police dismissed cases per capita than a high-income neighborhood area across all three time spans?

Natural Language Processing

We would like to implement NLP models to [based on cp1 output data table] [task specification]. For this task, we will apply the [model name] to analyze / make classification on text.
(Q: what is our purpose for the trained model?) -> training outcome perception

Future work

Using these conclusions from our questions and answers, is there a way to predict misconduct based on neighborhood median income or any other socioeconomics factor? And is there a way to apply these learnings and predictions to other police departments in the US using transfer learning tools?