Michael Price
Dr. Guha
COSC 5610
10/8/2020

<u>Dataset</u>

The dataset I will be using looks at citizen complaints filed against police officers in New York City spanning from September of 1985 to January of 2020. This is a public dataset that was released July of 2020 on ProPublica's data repository and provided by New York City's Civilian Complaint Review Board (CCRB). According to its page on ProPublica, every complaint listed in the dataset was fully investigated by the CCRB and was confirmed to have occurred, including cases where the conduct in question was permitted by the NYPD. Each recorded incident contains information regarding the officer in question, the civilian who filed the complaint, the month and year the complaint was opened and closed, as well as details describing the alleged misconduct and whether the CCRB believed the officer violated NYPD rules [1].

<u>Research Questions</u>

Working with data on a sensitive and complex issue like police misconduct comes with serious moral and ethical implications. Because of this, it is important to note that the questions, answers, and results I find in my research are only observations that pertain solely to New York City. There have been countless studies and articles examining issues revolving around police brutality and racism across the United States as well as specific major cities [2][3][4]. With the advent of the internet and social media, this issue has been brough to for forefront in our society [5]. Using this dataset focused on New York City, there are several research questions I plan to explore in hopes of furthering our understanding of the issues surrounding police misconduct:

- Are there identifiable regressors with significant predictors? For instance, could we identify predictors for the type of misconduct using information on the officer and citizen in question?

- Knowing that this dataset is comprised of complaints from the span of over thirty-four years, how much does the state of police misconduct change overtime? If we train an algorithm with the data from 2000 to 2010, how would it preform when testing it with the data from 2010 to 2019?

- How do our findings compare to past research conducted on this topic? Does it reflect common views of police brutality and racism? Have other researchers reached similar findings, or did they come to different conclusions. How is their process different from mine?

References

[1] ProPublica. 2020. Civilian Complaints Against New York City Police Officers. Retrieved from https://www.propublica.org/datastore/dataset/civilian-complaints-against-new-york-city-police-officers.

[2] Chaney, C., Robertson, R.V. 2013. Racism and Police Brutality in America. J Afr Am St 17, 480–505. DOI: https://doi.org/10.1007/s12111-013-9246-5.

[3] Krieger N, Chen JT, Waterman PD, Kiang MV, Feldman J (2015) Police Killings and Police Deaths Are Public Health Data and Can Be Counted. PLOS Medicine 12(12): e1001915. DOI: https://doi.org/10.1371/journal.pmed.1001915.

[4] Currie, M., Paris, B. S., Pasquetto, I., & Pierre, J. (2016). The conundrum of police officer-involved homicides: Counter-data in Los Angeles County. Big Data & Society. DOI: https://doi.org/10.1177/2053951716663566.

[5] Vanderford, J. T. 2020. Wardlow Revisited: How Media Coverage of Police Brutality Makes Empirical Data More Relevant Than Ever. *University of Pennsylvania Journal of Constitutional Law*, *22*(5), 1523.