# Project Check-in 1

Individual Report

Project Info: Ethical Data Mining

Team: Group 1

I found the datasets that we shall be performing our project upon. It is a good dataset and I feel this is one of the special cases where the absence of certain data values speaks as much as the that is present for analysis. I moved into the exploratory data analysis phase of the project and performed imputation on 1 of the datasets. The second dataset I started has been explored but needs to be imputed for missing/irrelevant values. So far, I have noticed traces of the inhumane nature of the American state apparatus and society. When I say that, it is a feeling mostly arising from the fact that the deaths of K-9 units have an active record that is logged and noted in the files but a lot of details on the victims of the state-police machinery are missing. I will intentionally be provocative here and state that ‘in America, the life of a dog holds more importance than the victims of police violence.’ I read an extensive amount of literature surrounding the state of affairs in American society which includes, but is not limited to, the NY times, Jacobin magazine, Huffpost, Guardian, etc. I have no sources to cite as of now but will have a lot more to share by the next check in. So far, most of the work I have performed can be seen in the master branch of our git repository as I am the only one who has pushed to the master branch as of now. I state this as the status as of 7:35PM on the 6th of April. So far, I have analyzed the dataset containing the data on police deaths and have explored the police killings dataset. The data that has been used so far is solely from a Kaggle grandmaster with the username ‘John M’ [1]. I would like to mention that there are 3 such datasets available from this user as it was a 3-part series of Kaggle datasets posted by him on his page. We shall discuss adjusting the scope of the project as we finish working on the current datasets and shall also mention the scope limits in future reports.

## Project Activities Completed Since Last Check-in

* Found the datasets that we shall be working on for the remainder of the project
* Discussed project scope with the group
* Created the Discord server for the group to discuss the project
* Initialized the git repository
* Downloaded and added the project files to the git repo
* Performed EDA on the police deaths and police killings datasets
* Imputed and got rid of some data in the police deaths dataset

## Project Activities in Progress

Currently, I am working on imputing and analyzing the data in the police killings dataset.

## Project Activities Planned

I shall first perform EDA on all the datasets and wrangle the data. Then I shall perform a deeper analysis on it using data visualization. Finally, I shall try to share all my findings in the report.

## GitHub/Google Drive links

Github repository link: https://github.com/shahbazmansahia/EthicalDataMining

# References

|  |  |
| --- | --- |
| [1] | JohnM, "JohnM | Grandmaster | Kaggle," 2020. [Online]. Available: https://www.kaggle.com/datasets/jpmiller/police-violence-in-the-us. [Accessed 18 03 2022]. |