Atul Aravind Das

U73275808

AI Ethics

**DETECTING BIAS IN DATASETS**

In the present data-driven world, a solution-based approach is crucial to solve problems. Acting as champions to this movement are the various algorithm-based applications such as AI tools and machine learning predictor models. These algorithm-based applications make use of various types of data to train and then test their models and many of them have been successful. However, in some cases, these solutions become controversial such as the case of the Amazon hiring system. It used to hire only male candidates and very few female candidates were hired in the process. This is because the historical data used to create the model considered the best-performing candidates and most of the candidates were males which stirred a huge controversy. The problem was that the data used for training was gender biased which jeopardized the model and compromised the reliability and fairness of the model. This is an indicator of the importance of the detection of bias in datasets.

Predictive Policing is one major application that uses criminal data and does analysis and predicts when and where crimes will occur to prevent crime before they happen. However, despite its effectiveness, in many cities like Miami, it has come under great scrutiny because the data has been biased based on parameters like race and zip code. This is a concerning problem as it may wrongly classify a person as a criminal despite not committing any crimes. This paper conducts a study and gets inferences to detect potential biases in the Los Angeles Crime Data from the year 2020 to the present day using Python. Through this study, we aim to highlight how potential biases can affect the performance of algorithms and how mitigating such biases play an important role in the development of fair and effective models.

Los Angeles is the largest and the most populous city in the state of California. The Los Angeles Police Department (LAPD), has 21 areas in its precinct. These 21 areas cover the 88 cities present in Los Angeles County. Three datasets are considered here. The first dataset lists the occurrence of crimes in each area [1] and the second dataset contains the area and perimeter of the areas [2]. The third dataset contains the demographic data of all 21 areas. The columns considered for the data are LAPD Area, Total Population, White, Hispanic/Latino, Black /African American, Asian, Other, and Median Income[3]. The initial data being used had details in percentages and the values have been converted to numeric values for easier implementation. The third dataset has been obtained with the help of Claude AI and it referred to the 2020 Census/LA based data.

In the scenario of predictive policing, the data considers parameters such as population, income, race, and number of crimes committed, and using this data, a predictive model to detect the occurrence of a crime in a particular area has been created. This model aims to detect “high-risk areas” and try to monitor and mitigate crimes as soon as possible. However, despite its intention to improve the greater good of society, it faces potential biases. Some areas among the 21 precincts have higher populations than the other areas. This will ensure greater monitoring of people in such areas as they pose a higher risk than areas with a lower population and so because of this, more crimes get reported from those areas. Another possible case of bias will happen in the00 case of parameters such as race and zip code. Consider a densely populated area that is diverse. But, in that area, the predominant race present there are Blacks/African Americans. A lot of people assume the 13/52 assumption, i.e. the 13% African American population causes 52% of the crimes [5], and then arrest people. Therefore, there will be increased monitoring in that area and it will report more crimes thus indicating the Collinearity between race and area name. The model has to ensure that this potential bias is mitigated during creation.

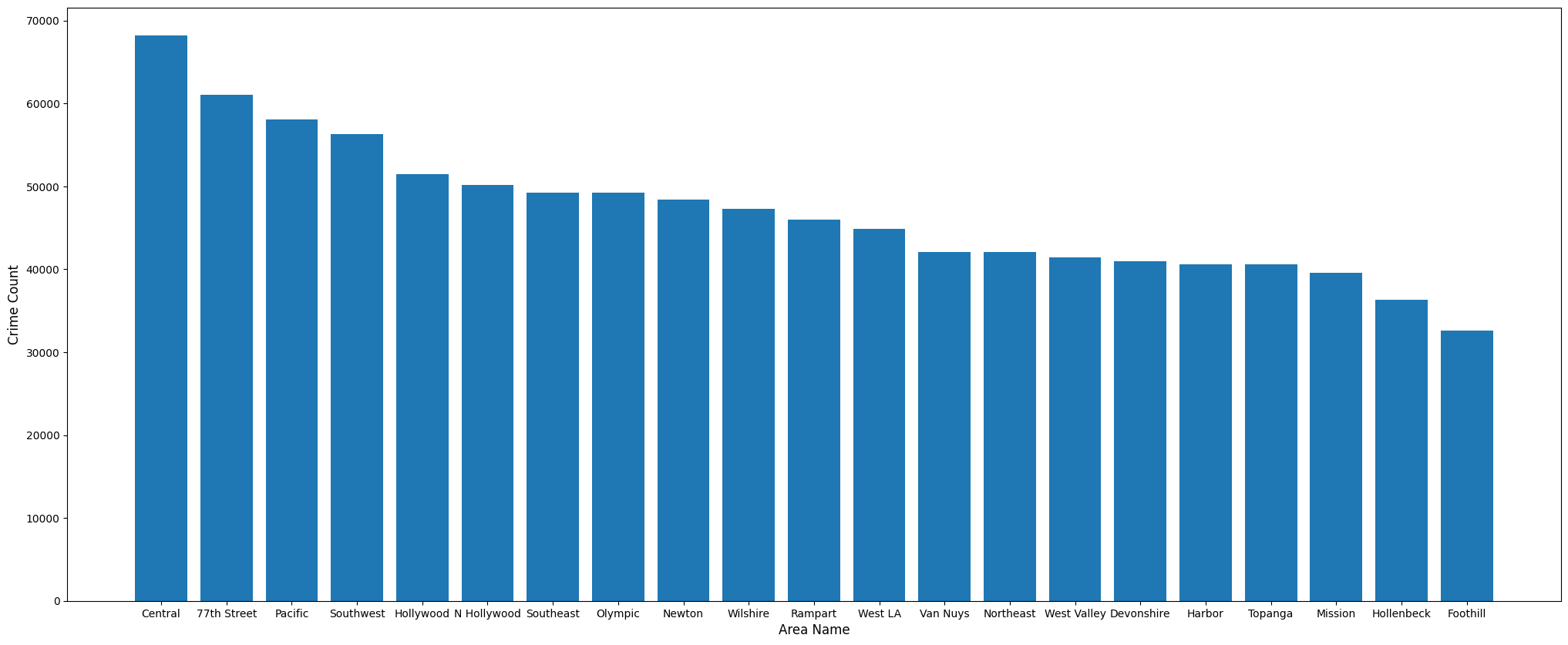
Feature Engineering is an important aspect of creating a predictive policing model. A lot of features such as crime rate can be generated to get better inferences. However, there is a flaw in the Crime Rate. The Crime Rate is defined as the Total Number of crimes per population and this implies that it’s inversely proportional to the population when the areas have the same number of crimes. An area with a smaller population will have a higher crime rate as compared to an area with a larger population. This will lead to an increase in monitoring in the areas with higher crime rates and they will be oblivious to the areas with lower crime rates. They will also be oblivious to the type of crime that is happening in both these areas as there will be greater monitoring of the area with the higher Crime Rate.

Along with the bias due to features, there is another potential source of bias and that is the data obtained from the US census itself. The US census is taken once in 10 years and the only existing data present in the model is from the year 2020. The next census is going to be taken in the year 2030. This does not reflect the number of changes that have happened in the last 10 years i.e., it only measures the population parameter values in the year 2030 and thus it acts as a source of bias and a possible solution to mitigate this bias is to take a yearly census

Biasing problems can be identified and resolved by considering and implementing solutions by keeping an ethical concept as the base of the solution. The ethical concept that can be considered in this case is the Rule Utilitarianism concept. According to Tavani, Rule Utilitarianism can be defined as: “An act, X, is morally permissible if the consequences of following the general rule, Y, of which act X is an instance, would bring about the greatest good for the greatest number.”[4]. In simple terms, the aspect of rule utilitarianism aspires to be fair to everyone by the set of rules and also maximize the happiness of everyone being affected by it. Rule Utilitarianism has been chosen in this case because it will consider the well-being of every citizen and ensure that fair treatment is given to everyone irrespective of their social status. If we consider a biased dataset, then it will jeopardize the algorithm as it will become unfair and it will decrease the happiness of everyone being affected by it.

0

Consider the case of Area-based crime count

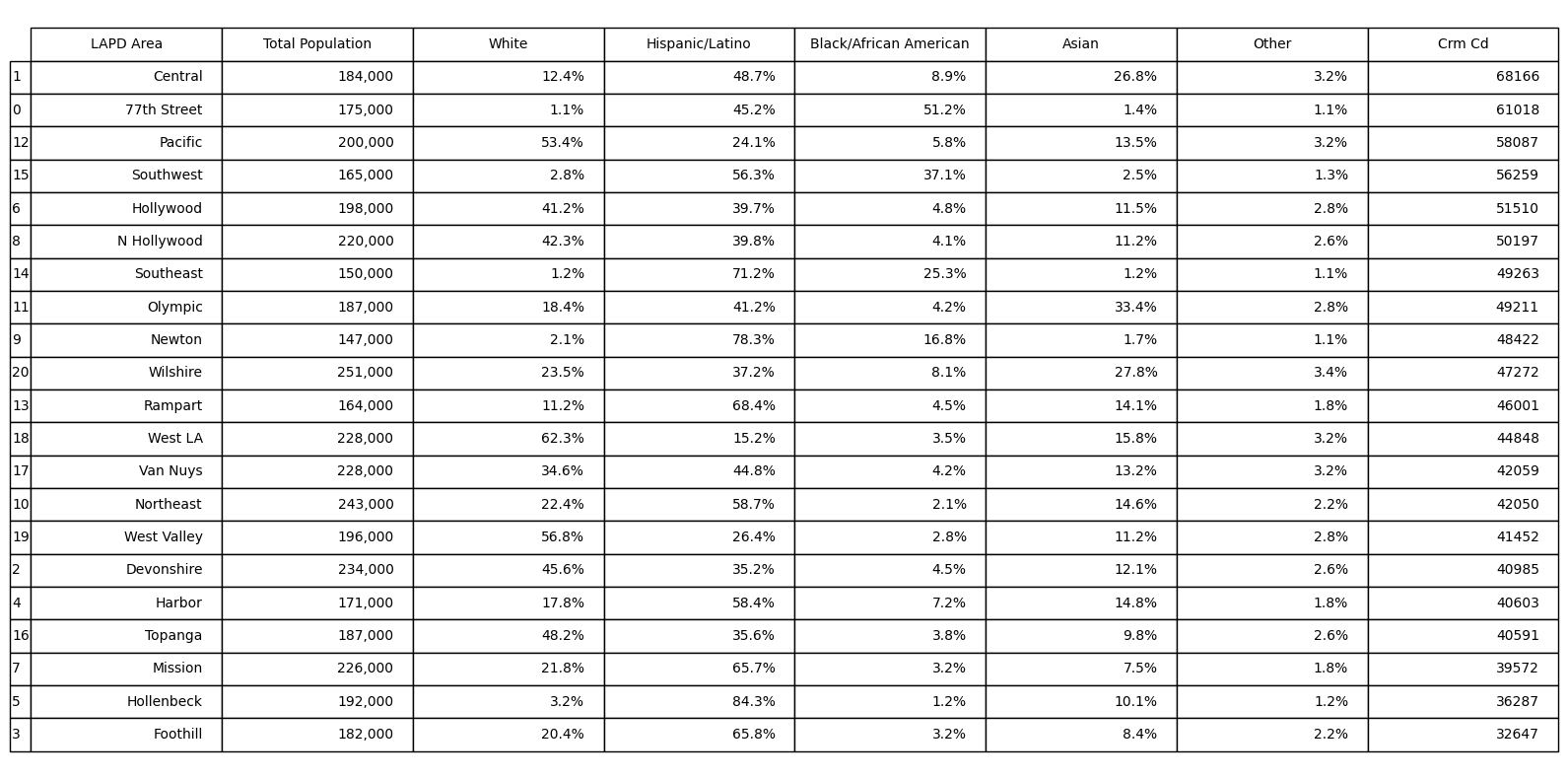


In this scenario, the areas with Top 5 crimes are the following: Central(68166), 77th Street (61018) Pacific (58087), Southwest (56259), and Hollywood(51510). This implies that these five areas are to be avoided as much as possible.

If we consider the five areas with least crime the following areas are considered: Harbor (40603), Topanga (40591), Mission(39572), Hollenbeck(36287), and Foothill(32647).

This implies that the least crime happens in these five areas and people who live in these five areas feel the most secure about their lives.

However, consider the case of demographic values in these areas



If we consider the top 5 areas with the most number of crimes, three of the five areas (Central, 77th Street, and Southwest) have a Hispanic/Latino majority with their proportion being approximately 50% in these areas. In the other areas (Pacific, Hollywood), the Hispanic/Latino people make up the second highest majority. If we consider these five areas to be extremely unsafe because a lot of crime is happening in these areas, it will discriminate against the Hispanic/Latino population in that area. Other than the areas with the most crimes, even in the areas with the least crime, there is a significant Hispanic/Latino population. The usage of this type of data will discriminate against the Hispanic/Latino population and usage of such data to create a predictive policing model will lead to a decrease in happiness and well-being of Hispanic/Latino people in the area. Using this dataset will go against the idea of Rule Utilitarianism as well and this is because the Hispanic/Latino population in the county will be unfairly treated and this will not maximize the happiness of people affected by it. Hispanic/Latino businesses and reputations will also be affected by it as it will bring them down and it may lead to a lot of losses for them. Therefore, this indicates a sign of potential bias.

Detection of bias in a dataset is a crucial step in the initial stages of model creation. Failure to detect bias in a model will result in the entire algorithm being biased. In the case of topics like Predictive Policing, a lot of algorithms fail to detect the inherent patterns that lead to bias and a lot of unfair arrests are made due to the bias being present. The Los Angeles Police department crime dataset inherently did not have any bias. But, if the demographic case is considered, there is a potential bias present in the dataset. Thus, to design such algorithms, a detection of biases and mitigation of biases is necessary for its proper functioning. The results have been published as a code[6].

References

[1] <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

[2] <https://geohub.lacity.org/datasets/lahub::lapd-police-stations/explore>

[3] <https://claude.site/artifacts/f52c1170-9778-4649-9b03-7ad746950d21>

[4] Herman T. Tavani - Ethics and Technology\_ Controversies, Questions, and Strategies for Ethical Computing, 4th Edition-John Wiley and Sons (2012)

[5] <https://www.adl.org/resources/hate-symbol/1352-1390>

[6]<https://colab.research.google.com/drive/1jtISvGChSE9DaSeLz2782XUoSFAdw4fY?usp=sharing>