### **Project Proposal: Analyzing Racial Profiling in Police Stops**

#### **Define Project**

Racial profiling in law enforcement has long been a critical issue, raising concerns about fairness, equality, and justice in policing practices. This project aims to analyze police stop data from NJ to identify patterns that may indicate racial disparities in stop outcomes, such as arrests, citations, and warnings. Specifically, it will investigate the relationships between race, stop conditions (e.g., time of day), and police actions to determine whether certain demographic groups are disproportionately impacted.

The strategic goal is to use data science techniques to provide data-driven insights into potential systemic biases. This project relates to class lectures and readings on data analysis, machine learning, and database management by incorporating advanced analytics and ethical considerations into the exploration of societal issues. It also highlights the importance of transparency and accountability in using data for decision-making.

#### **Novelty and Importance**

This project addresses a highly relevant social issue: racial profiling in policing. While existing studies on police stops have highlighted disparities in outcomes, many fail to leverage modern data science tools, such as machine learning models or explainable AI techniques, to analyze these issues deeply. This project is novel in its approach, integrating machine learning, SQL databases, and ethical analysis to identify and explain disparities.

The project is significant because it has the potential to provide actionable insights for policymakers and law enforcement agencies. By identifying patterns of inequality, the findings can help inform reforms aimed at promoting fairness and equity in policing practices. Current challenges in data management, such as handling missing demographic information and ensuring unbiased algorithmic outputs, will also be addressed. These elements make this project both technically challenging and socially impactful.

#### **Plan**

* **Data**:  
  The project uses the police stop dataset, which includes over 195,000 entries with details on subject demographics (e.g., race, age, sex), stop conditions (e.g., time, location), and outcomes (e.g., arrests, warnings, citations). This real-world dataset provides a comprehensive view of policing practices in a specific urban area. The dataset will be cleaned to handle missing values and inconsistencies, and additional features (e.g., nighttime stops) will be derived for further analysis.
* **Models/Techniques**:
  + **Exploratory Data Analysis (EDA)**:
    - Analyze the distribution of stops by race, time, and outcomes.
    - Create geographic heatmaps to identify high-incident areas by outcome and race.
  + **Database Integration**:
    - Store the cleaned dataset in an SQL database for structured querying.
    - Write queries to identify patterns, such as arrest rates by race and time of day.
  + **Machine Learning**:
    - Train a classification model to predict stop outcomes based on demographics and stop conditions such as make or model of car.
    - Use explainable AI tools to make predictions transparent and understandable.
* **Implementation Steps**:
  + **Data Cleaning**: Handle missing values, standardize categorical variables, and derive new features (e.g., stopped\_at\_night).
  + **Exploratory Analysis**: Use visualizations (e.g., bar charts, pie charts, heatmaps) to uncover trends and disparities.
  + **Database Integration**: Store the cleaned data in SQLite, and write SQL queries to analyze trends and relationships.
  + **Machine Learning**: Train and validate classification models to predict outcomes like arrests or citations, ensuring fairness and accuracy.
* **Evaluation**:  
  Success will be measured using:
  + Statistical metrics for machine learning models (e.g., F1 score, precision, recall).
  + Insights gained from exploratory analyses and SQL queries.
  + Ethical assessment of the findings, with a focus on fairness and transparency.

#### **Expected Outcomes**

The project will deliver:

1. **Data-Driven Insights**: Patterns in stop outcomes by race, time, and location, highlighting potential disparities.
2. **Machine Learning Model**: A trained and explainable model to predict stop outcomes.
3. **SQL Database**: A structured database for querying patterns in police stops.
4. **Ethical Recommendations**: A written discussion on findings and their implications, with actionable suggestions for promoting equity in policing.

#### **Conclusion**

This project represents a meaningful application of data science to a pressing societal issue. By combining technical rigor with ethical considerations, it seeks to uncover insights that can inform policy and promote fairness in law enforcement. The findings will demonstrate how data can drive accountability and contribute to positive societal change.