Guided Capstone 2 Project Proposal

1) Problem Identification

a) **Problem Statement:** The goal of this project is to analyze felony crime patterns in New York City using NYPD arrest data to identify trends, hotspots, and demographic correlations

b) **Context:** New York City is one of the largest and most populous cities in the United States. Therefore, it presents a unique and complex landscape for crime and law enforcement. Historically, the city has experienced various phases of crime rates, with significant reductions in major crimes since the peak levels of the 1990s. The diverse urban setting of NYC, encompassing a range of socioeconomic backgrounds and dense neighborhoods, contributes to the varied nature of its crime patterns, making it a critical area for ongoing and detailed criminological study and analysis.

c) **Criteria for Success:** Successfully identify key patterns and trends in felony crime. As well as successfully building a model that accurately predicts future occurrences of crime in the city.

d) **Scope of Solution Space:** Analysis will focus on demographic dimensions of felony crimes.

e) **Constraints:** The data I am working with is mostly textual data as opposed to numerical data making it more difficult to make numerical analyses without specialized techniques such as NLP.

f) **Stakeholders:** NYPD, policy makers public safety analysts, community leaders

g) **Data Sources:** I am using publicly available arrest data from the New York City Open Data. The data includes information such as the type of crime, the location and date of the arrest, and information on the perpetrators. https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc

2) Approach and Methodology

a) After initial analysis, the project will focus on developing predictive models using machine learning techniques. These models aim to forecast the likelihood of felony crimes in different areas or times based on historical data patterns.

b) Various machine-learning models will be tested and will implement the most effective model for the project

3) *Deliverables:*

a) Github repo
b) Slide deck
c) Project report