October 17, 2023

**PROJECT PROPOSAL**

**Team 3**

**CRIME ANALYSIS USING GIS FOR DATA SCIENCE**

**OBJECTIVES:**

The objective of this project is threefold:

**a.** To visualize crime data spanning five years (2018-2022) to delineate the crime pattern within the city of Washington DC using ArcGIS Pro.

**b.** To conduct predictive modelling on the consolidated data from the five years to anticipate future crimes using machine learning techniques.

**c.** To spatially represent the forecasted crimes based on the predictions derived in part b using ArcGIS Pro and Python.

**SCOPE OF WORK:**

The tasks to be undertaken are as follows:

• Collection and preprocessing of crime data from the years 2019 to 2023 for Washington DC.

• Visualization of historical crime patterns within the city using ArcGIS Pro.

• Combining the data for the specified years and conducting machine learning techniques for crime forecasting and trying to understand the nature of the crime.

• Spatial representation and mapping of the forecasted crime using Python and ArcGIS Pro.

• Evaluation of the forecasting model's performance and accuracy.

• Documentation and presentation of the findings to relevant stakeholders.

*Hypothetical Questions*

*Q1. Does the density or sparsity of a location's geographical attributes, such as the number of unique properties nearby, affect the frequency of crimes?*

*Q2. In terms of crime features like types, regions, and numerical difficulty of events, are there time series patterns?*

**OUTPUTS AND PRODUCTS:**

The deliverables under this project include:

• An interactive crime pattern visualization for Washington DC spanning the years 2019-2023.

• A machine learning model detailing anticipated crime hotspots, features and trends.

• A spatial representation of the forecasted crimes for stakeholder reference.

• A final report detailing the methodologies employed, results obtained, and areas of potential improvement.

**METHODOLOGY**

Crime patterns, though seemingly sporadic, can exhibit discernible trends when analysed with the right tools. Utilizing ArcGIS Pro allows for intricate visualization of crime hotspots and temporal changes. Combining this with Python and its robust libraries for data analysis, machine learning, and geospatial operations, the project aims to not only depict the current state of crime in Washington DC but also to provide a forward-looking perspective.

The project will commence with data collection and preprocessing to ensure data integrity. Visualization in ArcGIS Pro will highlight key patterns, aiding in feature selection for the machine learning phase. The model will then be trained, validated, and tested on the historical data to ensure robustness. The resultant predictions will be spatially mapped and shared with stakeholders.  
  
**DATASET DESCRIPTION:**

The dataset titled "Crime Incidents in 2019" is sourced from Open Data DC, provided by the City of Washington, DC. It underwent a methodology change on February 1, 2019, enhancing geographical assignment accuracy. The dataset contains subsets of incident locations and attributes from the ASAP crime report database by the District of Columbia Metropolitan Police Department (MPD).  
  
We will be working on the same dataset updated for years 2019 to 2023. All 5 dataset have same set of features.

Key Attributes:

CCN: Character string

REPORT\_DATE: Date

SHIFT: Character string depicting the time of the crime (EVENING, DAY, MIDNIGHT)

METHOD: Character string indicating the method of crime (OTHERS, GUN, KNIFE)

OFFENSE: Character string detailing the type of offense

BLOCK: Character string representing the crime's location

XBLOCK & YBLOCK: Double-precision decimal numbers for spatial data

WARD & ANC: Character strings detailing administrative regions.

The project will prioritize the timing (SHIFT), method (METHOD), and location (BLOCK, XBLOCK, YBLOCK) among other attributes for visualization and analysis.

Also, Economic characteristics survey data of DC Census Tracts (specific clusters of regions) includes features of Employment, Commuting, Occupation, Income, Health Insurance, Poverty, etc. (U.S Census Bureau). It is also sourced from Open Data DC and could be used for this analysis by joining to Crime Incidents dataset.

**TIMELINE:**

Oct 17 **Proposal & Draft Pseudocode** **Due**

Oct 22 Gather data, Set small goals of project objectives.

Oct 30 **Data Due**, Make Pipeline between ArcGIS and Python

Nov 5 Define GIS features, set potential targets to forecast.

Nov 13 **Revised Pseudocode & Progress Report Due**: Visualize Crime D.C,

Nov 25 Preparation and Performing the Model development

\*The modeling schedule will be specified (Define methodology, etc)

Dec 19 **Final Project Due**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Week  1  of October | Week  2  of October | Week  3  of October | Week  4  of Oct | Week  5  of Nov | Week  6  of Nov | | Week  7  of Nov | Week  8  of Nov | | Week  9  of Dec | Week  10  of Dec |
| Review Literature | Min & Abhi | |  |  |  |  | |  |  | |  |  |
| Initial Data Collection | Min | |  |  |  |  | |  |  | |  |  |
| Develop Data pipeline between ArcGIS and Python |  |  | Murthy | Sasank |  |  | |  |  | |  |  |
| Modelling |  |  |  |  | Min and Murthy | | | | | | | |
| Visualization |  |  |  |  | Sasank and Abhi | | | | | | | |
| Finalize and comment code |  |  |  |  |  | | Sasank and Murthy | | |  | | |
| Presentation |  |  |  |  |  |  | |  |  | | Abhi | |

**Reference:**

DC Crime data

https://opendata.dc.gov/datasets/f9cc541fc8c04106a05a1a4f1e7e813c\_4/explore

Metropolitan Police Department https://crimecards.dc.gov/all:violent%20crimes/with%20a%20gun/2:years/citywide:heat

ACS Economic Characteristics DC Census Tract : <https://opendata.dc.gov/datasets/DCGIS::acs-economic-characteristics-dc-census-tract/explore>

Yue Deng, Rixing He, Yang Liu,Crime risk prediction incorporating geographical spatiotemporal dependency into machine learning models, *Information Sciences,Volume 646, 2023, 119414,ISSN 0020-0255*, <https://doi.org/10.1016/j.ins.2023.119414>.

Anneleen Rummens, Wim Hardyns, Lieven Pauwels, the use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban *context, Applied Geography,Volume 86, 2017,Pages 255-261, ISSN 0143-6228*, https://doi.org/10.1016/j.apgeog.2017.06.011.