README document – Crime project

Mohammed Aminoor Rhaman

What each file refers to

- **Datasets** (*folder):* This folder contains the '**Police_data1'** folder which contains all the crime data in csv format. '**Local_Authority_Districts_May_2024_Boundaries_UK_BGC_- 6307115499537197728.geojson'** is a geojson file which contains the map boundaries for all the districts in the counties we are looking at. **'regionalgrossdisposablehouseholdincomelocalauthorities2023'** is an excel file which contains data on GDHI and population for every district.

- **Project_report**: A pdf file of the written report.

- **Project_Statement_of_Work**: A pdf file that outlines the plan of the project.

- **Pre-processing_documentation:** A PDF file that describes the steps taken to pre-process the Police dataset and GDHI dataset. Also includes flowcharts that further simplify the steps taken.

- **Gnatt_chart**: An excel sheet which contains a gnatt chart for the project.

- **Trello_Timeline**: A PowerPoint file which shows a timeline of when certain tasks were completed.

- **Preprocessing_Police_Data**: Jupyter Notebook file which carries out data cleaning of the police data and then combines the processed data into one dataset. Ensure file paths are set correctly when running the code. After running this code, a new csv file containing

the joined and cleaned dataset will be saved to your current directory.

- **Preprocessing_GDHI:** Jupyter Notebook file which carries filtering of the GDHI excel file. Ensure file paths are set correctly when running the code. After running this code, a filtered population excel file and a filtered GDHI file will be saved to your current directory.

- **Preprocessing_test**: A Jupyter notebook file which was used to first test the data cleaning steps before implementing it on to the entire police dataset.

Github repository

Public repository which contains all the above files can be found in this link:

[rahmanaminoor25/Aminoor_Crime_Report: This repository includes all documentation, the report itself and the raw data used for this project.](#)