**READ ME:**

**Introduction**

This is a repository containing all the code for our project for the Code First Girls NanoDegree Data Pathway, *‘Exploratory Analysis of the Social Impact of Airbnb in London Boroughs: Does Airbnb Have a Measurable Influence on Recorded Crime?’* The aim of this study is to evaluate the impact that the growth of Airbnb has on London communities, with the specific objectives of:

* To determine the prevalence of whole properties and rooms listed on Airbnb in London boroughs
* To identify correlations in crime rates and locations of Airbnbs over a period of 4 months (11 December 2022 - 14 March 2023)

**Documentation**

A detailed explanation of the project, in terms of steps taken, challenges faced, and our own personal conclusions, can be found in the ‘Project Document Submission’.

**Archived Data**

We have documented the variety of different avenues we took to solve coding problems. A brief overview of each issue can be found in the ‘*Implementation Challenges’ of the Project Document Submission*, matched with files in the *Archived Data* folder. Within the archived data, we have explained what we were trying to do, and what issues we came up against.

**Data sets**

The data sets for the Airbnb and Crime data, both raw and cleansed, can be found in the *data* file, within *csv,* in *airbnb\_datasets* and *police\_datasets* respectively*.*

*airbnb\_files\_raw\_uncleansed* are the original datasets downloaded from <http://insideairbnb.com/get-the-data/>. The cleansed dataset can be found in the same folder, under the name *airbnbfiles\_final\_cleansed*. The code for the cleaning is in *airbnb\_dataset\_cleansing\_general.ipynb*.

The crime datasets (*police\_datasets\_raw\_uncleansed*) were downloaded from <https://data.police.uk/data/> . Five datasets were initially downloaded on WHAT DATE (referred to as met\_police\_street\_MONTHYEAR). We then discovered that the London City crime statistics were kept in a different dataset, so those were downloaded on WHAT DATE (referred to as city\_police\_street\_MONTHYEAR). The cleansed dataset can be found in the same folder, under the name *police\_datasets\_final\_cleansed*). The code for the cleaning is in *police\_street\_dataset\_cleansing\_general.ipynb*.

**Data cleaning**

The first code to be run is in the Data Cleaning file. Within *airbnb\_dataset\_cleansing\_general.ipynb* it is detailed how the Airbnb data from different time frames were joined and the data was then cleansed keeping only the relevant columns.

*police\_street\_dataset\_cleansing\_general.ipynb* details the cleansing of five crime datasets, ranging from June 2022 to March 2023. This included joining the crime statistics of London with the separate file for the City of London, removing columns we would not be using in in-depth analysis, and matching the LSOA names to the Airbnb boroughs (the many attempts to do this which failed are documented in Archived Data).

**API Connection**

Using two APIS, we have mapped the boroughs we were working with on to an interactive map, accessible through *api\_call\_boroughs\_visualisation\_map.ipynb.*

**Results**

The results from our project can be viewed in **RESULTS FILE.**