**Proposal: Spatial Analysis of Airbnb and Housing Market Dynamics in the UK**

The focus here is a GIS task - to demonstrate the spatial relationship between the Airbnb data with other factors such as housing prices, and the short to let market. Using the datasets that I have shared here and using your own research, can you answer a series of research questions using spatial data science principles.

This codebook outlines the variables, data sources, and spatial data science methods that will be used in the analysis of Airbnb listings and their impact on housing prices and the short-let market in the UK. I require this in python.

**Data Sources:** For the UK context, the following data sources are recommended:

**Airbnb Data:**

**Source**: Airbnb data scraped from public sources or platforms like Inside Airbnb.

1. **Variables**:
   * listing\_id: Unique identifier for each Airbnb listing.
   * latitude: Latitude coordinate of the listing.
   * longitude: Longitude coordinate of the listing.
   * price\_per\_night: Price per night for the listing.
   * room\_type: Type of room (e.g., entire home/apt, private room, shared room).
   * availability\_365: Number of days the listing is available in a year.
   * number\_of\_reviews: Total number of reviews for the listing.
   * host\_id: Unique identifier for the host.
   * neighbourhood: Name of the neighborhood where the listing is located.
2. **Housing Prices Data**

**Source**: UK House Price Index (UKHPI) from HM Land Registry, Office for National Statistics (ONS).

**Variables**:

* + postcode: Postal code of the property.
  + price: Sale price of the property.
  + date\_of\_sale: Date when the property was sold.
  + property\_type: Type of property (e.g., detached, semi-detached, terraced, flat).
  + latitude: Latitude coordinate of the property (if available).
  + longitude: Longitude coordinate of the property (if available).

1. **Short-let Market Data**

**Source**: Aggregated from platforms such as Rightmove, Zoopla, or other property listing services.

**Variables**:

* + postcode: Postal code of the property.
  + rental\_price: Monthly rental price.
  + availability\_status: Current availability status (e.g., available, let agreed).
  + latitude: Latitude coordinate of the property (if available).
  + longitude: Longitude coordinate of the property (if available).

1. **Demographic and Socioeconomic Data**

**Source**: Office for National Statistics (ONS), UK Census Data.

**Variables**:

* + postcode: Postal code of the area.
  + median\_income: Median household income.
  + population\_density: Number of people per square kilometre.
  + unemployment\_rate: Percentage of unemployed individuals.
  + latitude: Latitude coordinate of the area (if available).
  + longitude: Longitude coordinate of the area (if available).

1. **Geospatial Boundaries**

**Source**: Ordnance Survey, UK Government Open Data.

**Variables**:

* + LSOA\_code: Lower Layer Super Output Area code.
  + boundary\_polygon: Polygon representing the geographical boundaries of an LSOA or other administrative units.
  + latitude: Centroid latitude of the area.
  + longitude: Centroid longitude of the area.

**2. Variables Overview**

The following variables will be critical in the analysis:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Source** |
| listing\_id | Unique identifier for Airbnb listing | Airbnb Data |
| latitude | Latitude coordinate of listing/property/area | All Sources |
| longitude | Longitude coordinate of listing/property/area | All Sources |
| price\_per\_night | Price per night for Airbnb listing | Airbnb Data |
| availability\_365 | Number of days listing is available in a year | Airbnb Data |
| number\_of\_reviews | Total number of reviews for Airbnb listing | Airbnb Data |
| neighbourhood | Name of the neighborhood of Airbnb listing | Airbnb Data |
| postcode | Postal code of property/area | Housing Prices, Short-let, ONS Data |
| price | Sale price of property | UK House Price Index (UKHPI) |
| date\_of\_sale | Date when the property was sold | UK House Price Index (UKHPI) |
| property\_type | Type of property (detached, semi-detached, etc.) | UK House Price Index (UKHPI) |
| rental\_price | Monthly rental price of short-let property | Short-let Market Data |
| availability\_status | Current availability status of short-let property | Short-let Market Data |
| median\_income | Median household income for area | ONS Data |
| population\_density | Population density of area | ONS Data |
| unemployment\_rate | Unemployment rate for area | ONS Data |
| boundary\_polygon | Polygon representing geographical boundaries | Ordnance Survey Data |

**Spatial Data Science Methods**

The following spatial data science methods will be employed to explore the spatial relationships between Airbnb listings, housing prices, and the short-let market:

* **Geocoding:** Converting addresses into latitude and longitude coordinates to plot properties and listings on a map.
* **Hotspot Analysis (Getis-Ord Gi):** Identifying statistically significant clusters of high or low Airbnb activity relative to the surrounding area.
* **Spatial Autocorrelation (Moran's I):** Measuring the degree to which similar values (e.g., high Airbnb density) cluster spatially, indicating potential spatial dependency.
* **Proximity Analysis:** Analysing the distance between Airbnb listings and key amenities or tourist attractions to assess their impact on prices.
* **Kernel Density Estimation (KDE)**: Estimating the density of Airbnb listings across space to identify areas of concentration and potential market saturation.
* **Regression Analysis: Spatial Lag Model**: Incorporating spatial dependencies in the regression model to account for the influence of neighbouring areas.
* **Geographically Weighted Regression (GWR)**: Assessing how the relationship between Airbnb listings and housing prices varies spatially across different regions.
* **Choropleth Mapping:** Creating maps that visualize the distribution of variables such as median income, housing prices, and Airbnb density across different regions.
* **Impact Analysis:** Examining how the density of Airbnb listings correlates with changes in local housing prices and the availability of long-term rentals.
* **Buffer Analysis;** Creating buffers around key areas (e.g., transport hubs, city centres) to study the impact of proximity on Airbnb pricing and availability
* **Network Analysis;** Analysing the connectivity of Airbnb listings to transport networks, which could influence pricing and availability.

**Jupyter Notebook Structure -** The analysis will be conducted in a Jupyter Notebook, organized as follows:

* **Introduction:** Overview of the study and objectives.
* **Data Import and Pre-processing:** Importing datasets (Airbnb, housing prices, short-let market data, etc.). Cleaning and pre-processing data, including geocoding.
* **Exploratory Data Analysis (EDA):** Descriptive statistics and initial visualizations of key variables.
* **Spatial Analysis:** Geocoding and plotting data on maps. Performing hotspot analysis, spatial autocorrelation, and kernel density estimation.
* **Regression and Impact Analysis**: Running spatial regression models to assess the impact of Airbnb listings on housing prices.
* Conducting impact analysis on the short-let market.
* **Visualization**: Creating choropleth maps and other visualizations to illustrate findings.
* **Conclusions and Recommendations**: Summarizing key findings and providing policy recommendations.
* **Appendix**: Including any additional analyses or data pre-processing steps.