UK Housing Investment Analysis based on Socioeconomic Indicators

Project Overview

The UK housing market is one of the most significant investment sectors in the country, but identifying high-potential locations requires more than simply analysing property prices. This project explores housing investment opportunities across the United Kingdom by combining multiple public datasets and developing a data-driven Investment Score. **The aim is to identify regions that offer strong investment potential by analysing economic conditions, deprivation levels, crime rates.**

By integrating housing transaction data with socioeconomic indicators, this analysis provides a structured approach to identifying areas that may offer better long-term property investment opportunities.

Data Sources

The analysis combines several publicly available datasets to create a comprehensive view of the UK housing market.

Property transaction (PPD) data was obtained from HM Land Registry. The dataset contains the complete record of residential property sales across England and Wales and includes transaction price, postcode, property type, and date of sale.
https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads#using-or-publishing-our-price-paid-data

Geospatial postcode data was sourced from the Office for National Statistics through the ONS Postcode Directory (ONSPD). This dataset maps postcodes to Lower Layer Super Output Areas (LSOAs) and provides latitude and longitude coordinates required for geospatial analysis.
https://geoportal.statistics.gov.uk/datasets/3be72478d8454b59bb86ba97b4ee325b/about

Crime statistics were collected through the public API provided by UK Police. These records contain borough-level crime counts that allow the analysis to incorporate safety and crime conditions into the investment scoring framework.
https://data.police.uk/

Finally, socioeconomic indicators were obtained from the Indices of Multiple Deprivation (IMD) 2019, which provide measures of income deprivation, employment deprivation, and overall socioeconomic disadvantage at the LSOA level.
https://opendatacommunities.org/resource?uri=http%3A%2F%2Fopendatacommunities.org%2Fdata%2Fsocietal-wellbeing%2Fimd2019%2Findices

Data Pipeline and Analysis

The first step in the project involved loading the datasets into a PostgreSQL database and performing data cleaning and transformation. Columns were standardised and unnecessary attributes were removed to create consistent tables containing key variables such as postcode, location coordinates, property price, and transaction date. Because property data is organised by postcode while socioeconomic indicators are recorded at the LSOA level, the next step

involved mapping postcodes to geographic regions. A geospatial nearest-neighbour algorithm using cKDTree was implemented to assign each postcode to the nearest borough centroid. This allowed the datasets to be integrated into a common geographic structure. A weighted Investment Score was then calculated to measure the relative attractiveness of different regions for housing investment. The score was designed to reflect desirable socioeconomic conditions by weighing and combining three key indicators:

Investment Score = Low Deprivation + High Income - High Crime

Higher scores therefore represent areas with stronger economic conditions and lower crime levels, which are typically associated with more stable property markets and long-term growth potential. The investment scores were aggregated at borough and postcode levels and then merged with historical house price data to analyse investment opportunities across the housing market.

Visualisation and Insights

An interactive dashboard was developed using Python, Streamlit, and Plotly to explore the results. The dashboard includes several visualisations designed to reveal patterns in housing investment potential. Three of the visualizations are included here:

A UK investment heatmap displays postcode locations across the country, ranging from red (lower investment potential) to green (higher investment potential). This allows users to quickly identify regions that may offer favourable investment conditions.

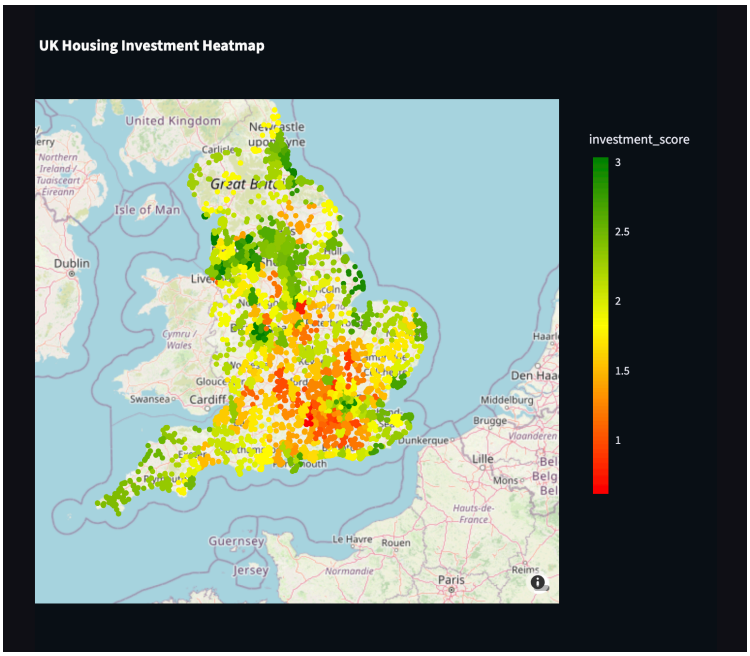

Image 1: Investment Score heatmap

The next is a trend chart covering the period from 1995 to 2023, illustrating how investment scores have evolved over time. These trend can help us predict future investment scores/ prices. Investment score trend of Bristol is shown below:

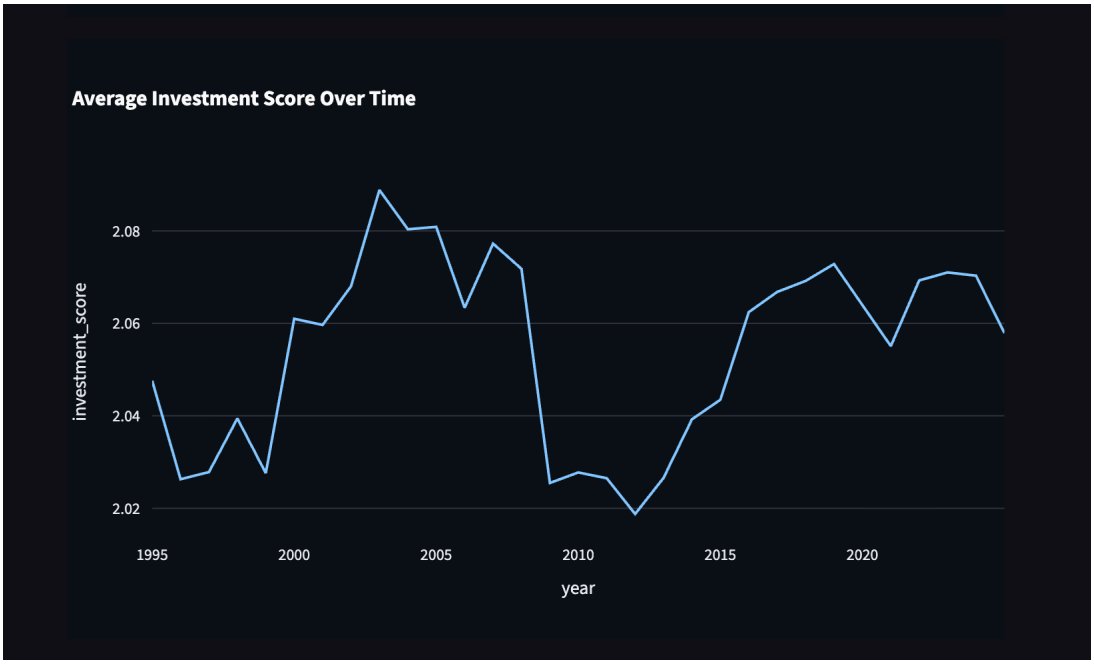

Image 2: Bristol Investment Score trend from 1995 - 2023

Finally, To find undervalued areas, I plotted average house price vs average investment score per borough. This helps you distinguish undervalued (good investment) opportunities from overpriced ones:

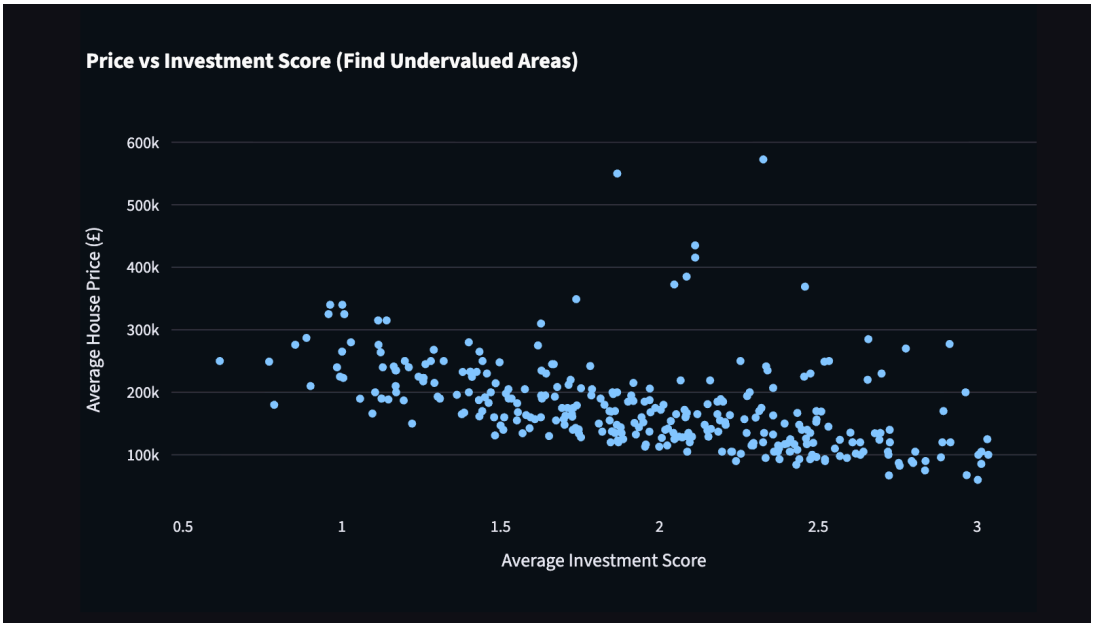

Image 3: Price vs Investment Score (the actual chart is interactive, you get ONS code of each borough along with price and investment score)

Conclusion

While property prices alone provide limited insight into long-term investment potential, combining them with broader contextual factors such as crime levels, income deprivation, and regional economic conditions allows for a more comprehensive assessment of local housing markets. By integrating housing transaction data with socioeconomic datasets, the analysis demonstrates how patterns in deprivation, income, and safety can influence both current property values and future growth potential.

The resulting investment score provides a structured framework for comparing regions and highlighting areas that may offer stronger long-term prospects. More broadly, this approach illustrates how data science can transform publicly available socioeconomic data into meaningful insights for property investment. By moving beyond price data and incorporating the wider social and economic context, investors and analysts can make more informed decisions about where housing markets may present future opportunities.

Tech Stack

Python: pandas, sqlalchemy, plotly, scipy, streamlit
Database: PostgreSQL