datarea

#SummerDataChallenge

Benjamin L. Moore[1]

OVERVIEW

I analysed the **London house prices** dataset is several innovative ways. Firstly *fractal context* shows how house sales within an arbitrary area relate to those if neighbouring districts, up and down the postcode hierarchy. Secondly, I employed *ARIMA forecasting* to extrapolate an area's price trend with an eye to suggesting profitable investments. Finally, these and other metrics were combined into an *investment grade* for a given postcode area. Combined this set of analyses forms the basis of datarea, a product aimed at "democratising real estate investment", offering individuals and businesses quantitative insights for data-driven investment decisions.

A fully-featured report is online at **blm.io** and scripts to reproduce all analyses are available from **github**.

1 DESCRIPTION OF ANALYSES

When considering a property in a given area, you may ask a realtor questions like, how do prices here compare to surrounding postcodes? How have prices increased over time?

A real estate professional can give an opinion based on anecdotal evidence and the limited sample size they have experienced, but with a large dataset we can address these questions quantitatively and present them as clear visualisations.

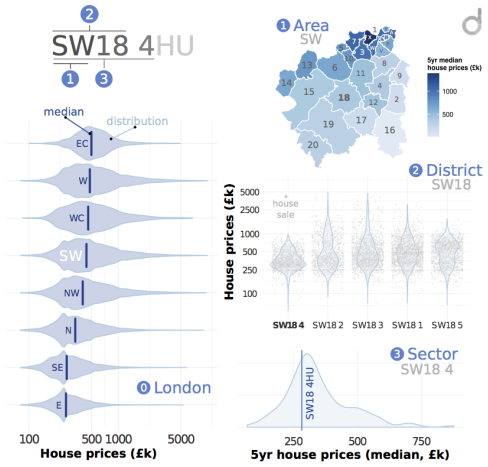

Figure 1: Fractal context of postcode SW18 4HU in Wandsworth, South London.

1.1 Fractal context

Letting and sales sites may currently list some recent sales, but it's not currently possible to see a quantitative overview of property prices in a region, within the context of its sector, district and postcode.

These questions can be answered with a clear and intuitive data visualisation of price distributions in the postcode hierarchy — we call this the fractal context of a property price.

1.2 ARIMA forecasting

Given area price histories, it's of interest to investors to speculate on potential investment returns for a given property area. This can be done through time series analysis using the autoregressive integrated moving average (ARIMA)

[1] *MRC Human Genetics Unit, University of Edinburgh, Scotland, United Kingdom*

model. Investment returns can be maximised by selecting areas with maximal growth projections.*

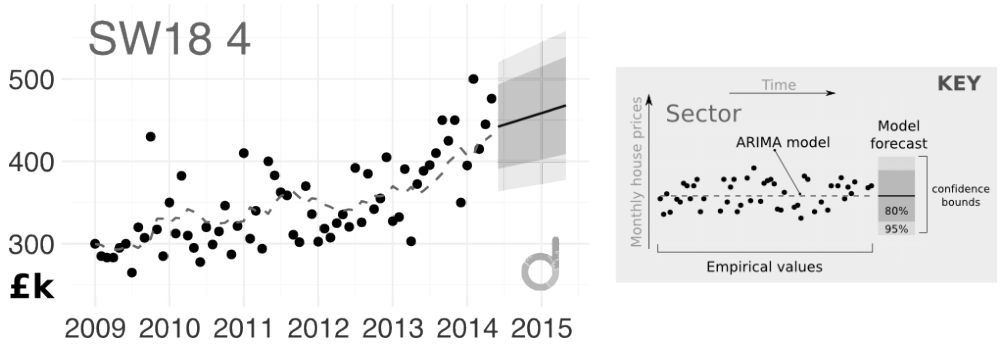

Figure 2: Example ARIMA forecast of median house prices for a specific postcode.

1.3 Investment grading

Combining growth forecasts and contextual information, as well as other derived metrics from this dataset, allow an approximate investment grading, relative to other postcodes within the area covered by this dataset. This intuitive output metric can help prioritise areas ripe for property investment.

2 VALUE GENERATION

datarea would first be released as both a web service and associated apps for mobiles and tablets. These would offer limited analysis and visualisation for free, with user analytics and feedback helping to validate the business model.

A small data science team could expand analyses, integrating novel public and privately-acquired datasets and release them under a rolling subscription model, aimed at private landlords and property speculators, as well as professional real estate investment trusts (REITs). Significant revenue could be generated through tailored partnerships with funds and high net-worth individuals.

Figure 3: iOS app mock-up.

Should the product not succeed in a direct-to-consumer market, the business could pivot to an analytics provider for existing letting and sales sites, offering unique data insights indirectly to the consumer market, or as independent fund consultants for the financial services. Having produced a functional product positions datarea to succeed in innovative real estate analytics.

Full version online at: blm.io/datarea

datarea

Democratising real estate investment

* Data intended to assist investors and does not constitute investment advice; independent advice should be sought where appropriate.