datarea

Summer data challenge entry by 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 districs, up and down the postcode hierarchy. Secondly, I fit *ARIMA models* to forecast an area's price trend into the future, 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 real estate insights for data-driven investment decisions.

A full report with additional visualisations is online at **blm.io/datarea** and scripts to reproduce all analyses are available from **github.com/blmoore/summerdatachallenge**.

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 their own limited sample size, but with a large dataset we can address these questions quantitatively and visualise the results.

0.1 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 (Figure 1).

0.2 ARIMA modelling

Given a time series of housing prices, 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) model (Figure 2). Investment returns can be optimised by selecting areas with maximal growth projections.[*]
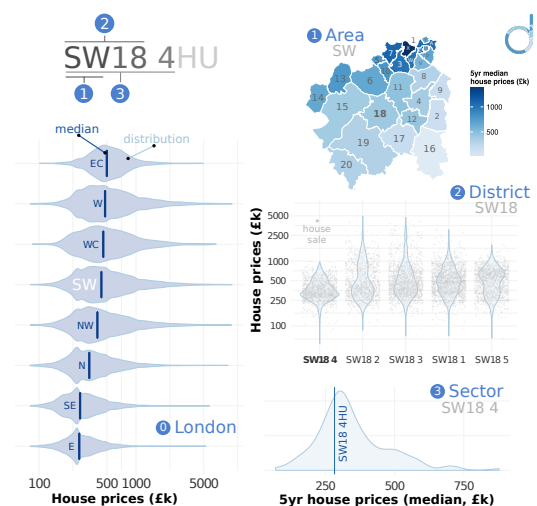

Figure 1: Fractal context of postcode SW18 4HU in Wandsworth, South London.

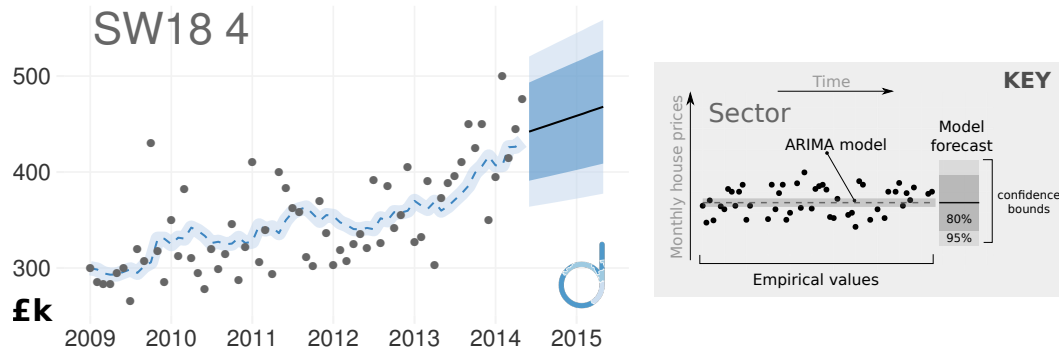[1] *MRC Human Genetics Unit, University of Edinburgh, Scotland, United Kingdom*

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

0.3 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.

1 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 will gradually expand the range of analyses offered, 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.

Additionally 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 first-to-market innovative product positions datarea to succeed as a provider of advanced real estate investment analytics.
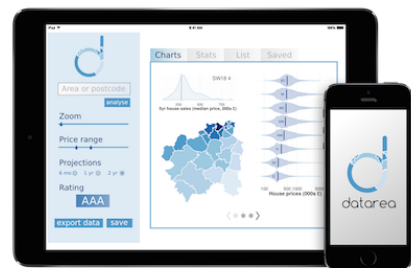

Figure 3: iOS app mock-up.

datarea

Democratising real estate investment

An enhanced overview of this entry is available online at: blm.io/datarea

---

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