

OVERVIEW

I analysed the **London house prices** dataset in several creative ways. Firstly, the *fractal context* of an area visualises how house prices relate to those of neighbouring areas at different scales 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, growth and other metrics were combined into an *investment grade* for a given postcode area. Together this set of analyses forms the basis of **datarea**, a startup aimed at *democratising real estate investment*, offering individuals and businesses quantitative property area insights to enable data-driven investment decisions.

An enhanced report with interactive visualisations is online at blm.io/datarea and scripts to reproduce all analyses are available from github.com/blmoore/summerdatachallenge.



A map of the house prices dataset (brightness is proportional to price and density).

SECTION A: ANALYSES

When considering a property in a given area, you may want to ask questions like: how do prices here compare to the surrounding areas? How have they changed over time? Where will they be in a year's 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 through statistical analysis and visualise the results.

Fractal context

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

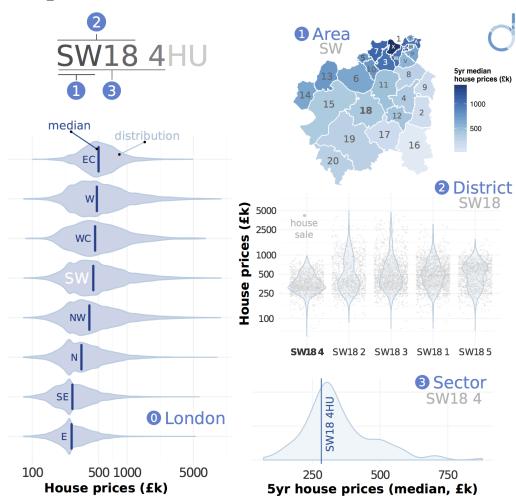


Figure 1: Fractal context of postcode SW18 4HU in Wandsworth, South London.

These questions can be answered with a clear and intuitive data visualisation of price distributions in the postcode hierarchy — what I've called the *fractal context* of a property price (Figure 1).

Any chosen area can be seen in the context of its surroundings, from full postcodes to much wider areas, clearly showing how the pricing of any given property relates to those nearby. The example graphic for postcode SW18 4HU (Figure 1) combines ranked violin density plots and a geo-heatmap for a clear, intuitive view of a location's price context.

ARIMA modelling

Property investors need a method of evaluating the potential investment returns of a given property area. This can be achieved through time series modelling with autoregressive integrated moving average (ARIMA) models (Figure 2). For this application regularised models were fit which allowed a drift term to capture non-stationary trends (i.e. allowing for gain, loss or neither) as well as coefficients

reflecting periodicity or seasonal effects. Hence potential investment returns could then be optimised by selecting areas with maximal growth projections.

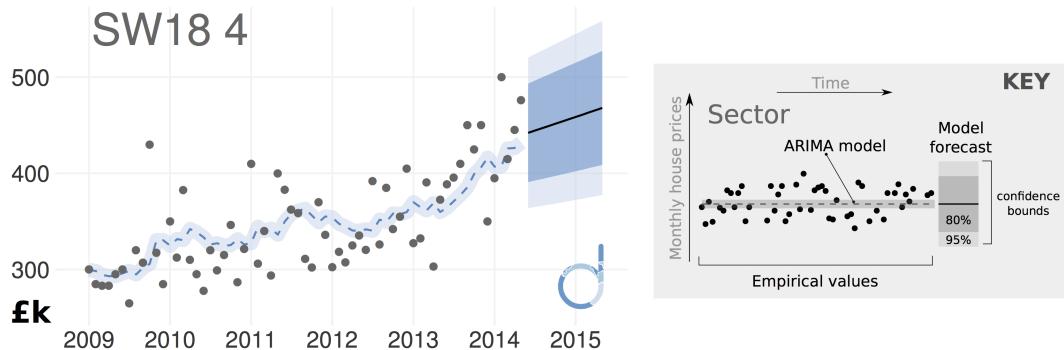


Figure 2: Example ARIMA forecast of median house prices for a specific postcode.

Investment grading

Combining growth forecasts and with an additional measure of variability allows an approximate investment grading of each sector, relative to other postcodes within the area covered by this dataset. This intuitive output metric can help prioritise areas ripe for property investment and combines the theoretical model projections with an empirical metric of past stability.

As an example, Figure 3 shows the top 5 postcode sectors in which to invest, according to their 12-month price growth forecast and historical annualised volatility. These include Brockley, whose train station within SE4 1 was linked with the London Overground network in 2010.

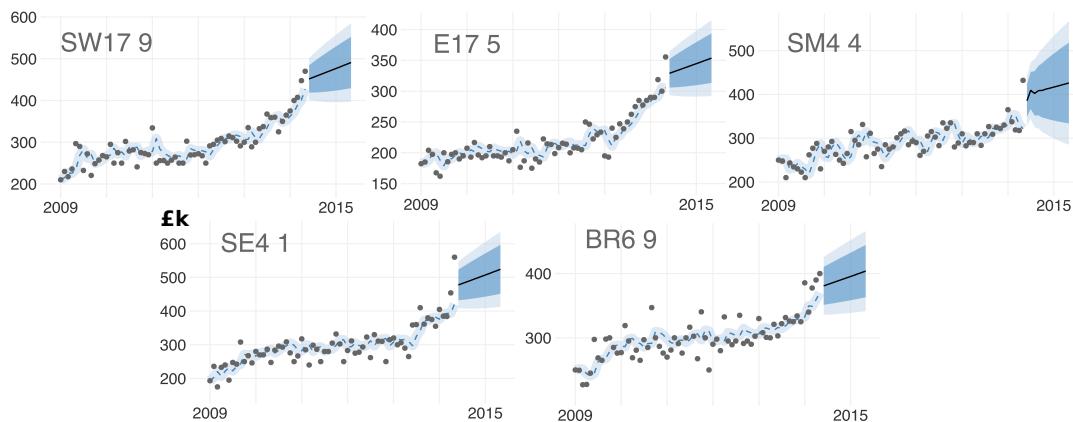


Figure 3: Top 5 property investment sectors by projected returns and historical volatility.

Growth forecasts and volatility were converted to quantiles and equally weighted (Figure 4); those sectors with the highest projected growth and least historical volatility receive the best investment grade (on a scale of AAA, AA, A, BBB...C).

Returning to our original example, SW18 4HU is a promising choice of area for property investment.* It's median-priced for the sector, which is the cheapest of SW18 (in turn part of the desirable SW London area; Figure 1). It has good growth projections of over £28,000 in the next year (Figure 2), and receives the top AAA investment grade for being in the 90th percentile for predicted price growth and low historical volatility compared to all other sectors covered by the dataset.

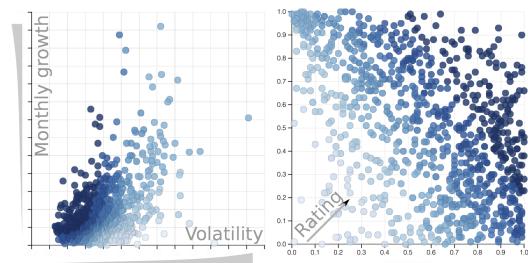


Figure 4: Quantile transformation for investment grading (AAA rated sectors in navy).

* Data intended to assist investors but does not constitute investment advice; independent investment advice should be sought where appropriate. The value of investments can go down as well as up.

SECTION B: VALUE GENERATION

The above-described analyses form the basis of **datarea**, a start-up aimed at “democratising real estate investment” by making deep-dive analytics data available both to a consumer and professional investment market. Initially **datarea** is a provider of information aimed at investors but cannot dispense investment advice until authorised to do so by the FCA. Three initial areas of interest to commercialise and generate social and economic value are listed:

1. Analytics provider for online house sales and lettings agents

datarea can provide **real-time property area analytics** to existing market leaders in **online sales and lettings**.



Potential partners.

There are currently no mainstream property sales or rental sites that attempt to provide any quantitative insights that contextualise a property's list price. Some list recent nearby sales in plain text, but the wealth of historical sales data available allows for much richer displays (as demonstrated with the fractal context view, Section A: Figure 1, and by the interactive data-driven map shown in Figure 5), which offer valuable and practical insights to prospective homebuyers.

This added feature would help differentiate a given sales and letting site as well as increase the brand positioning of **datarea** within the real estate sector. Contextual maps could be drawn as interactive D3.js visualisations and neatly integrated with a typical property search. Additionally, web analytics data from our partnered site could be fed-back to help further enhancement and expansion of the data-driven insights and visualisations we offer, such as how often a given area is searched for on their site and by which demographics.

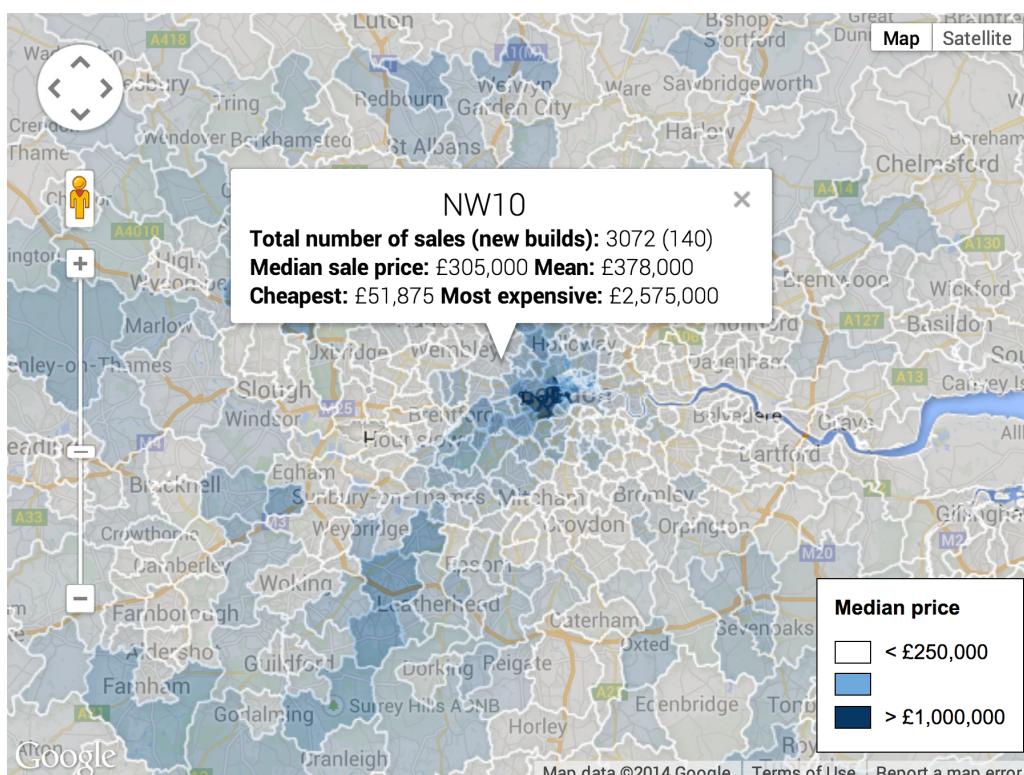


Figure 5: A basic embeddable web app built with the Google Maps API (viewable at blm.io/datarea).

2. Direct-to-consumer demo application

datarea will release an initial product both as a **web service** and associated **mobiles and tablet applications** (Figure 6). These offer limited analysis and visualisation without any fee, with collected user data and feedback helping to validate the business model and further develop our software.

Releasing these products aligns with our aims of democratising real estate investment. Amateur property investors and prospective landlords can make data-driven decisions, the likes of which are already employed by analysts for large investment funds.

A small data science team will gradually expand the range of analyses offered, integrating novel public and privately-acquired datasets and release them to a premium app 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.

3. Real estate data science consultancy

Having established a property area analytics brand, **datarea** will look to develop business relationships with real estate investment funds and high net-worth individuals as a **quantitative property investment consultancy** and independent analyst for the financial services. Our domain expertise and existing partnerships and products would place us as a market leader in this largely undeveloped segment.

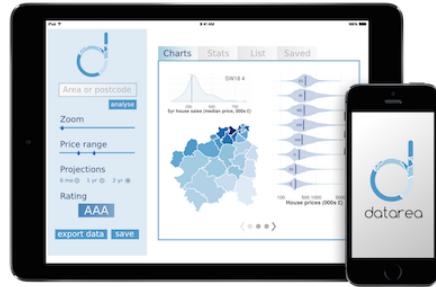


Figure 6: iOS app mock-up.



Figure 7: Branding.

SUMMARY

In this report I've given a brief overview my attempts to elicit maximal insight from a single large dataset through statistical modelling, visualisations and other data science techniques. Together these form a narrative of investigative analysis into a single exemplar postcode, and the assessment of the potential investment returns within.

I have shown that these analyses have the potential to generate valuable insights for a range of customer segments:

- Online sales and letting agents would benefit from plug-in analytics provided by **datarea** — offering their users advanced insights into spatiotemporal property statistics.
- A web and mobile app to make a set of analyses widely-available to amateur property investors, new homebuyers and landlords.
- A consultancy specialising in real estate investment to partner with REITs and other specialist investment funds, as well as charitable trusts.