

Location, Location, Location

Problem Statement: How far does a property's value depend on its location?



£11,000,000



£110,000

Guide Price ?

2 bedroom apartment for sale

Grosvenor Square, London, W1K

BRAND NEW INSTRUCTION - SOLE AGENT - A rare lateral apartment on the 4th floor with 5 west facing windows overlooking the square gardens.

Added on 03/11/2020 by Wetherell, London



020 7529 5566

Local call rate



2 bedroom maisonette for sale

Oakways, Eltham, SE9

A Self-Contained Two Bedroom Upper Maisonette with accommodation arranged on Two Floors.

Added on 06/02/2021 by McHugh & Co, London



020 8012 8911

Local call rate



We rely on estate agents to know the ‘right’ price for a property



They do this by benchmarking similar houses from the neighbourhood

This is less scientific than they would like you to believe



The closest houses may have different characteristics

A similar house a few streets away may have better amenities

A limited number of very local sales produces a sparse matrix

A lack of good benchmark properties leaves us relying on the ‘expertise’ of estate agents to fill in the gaps

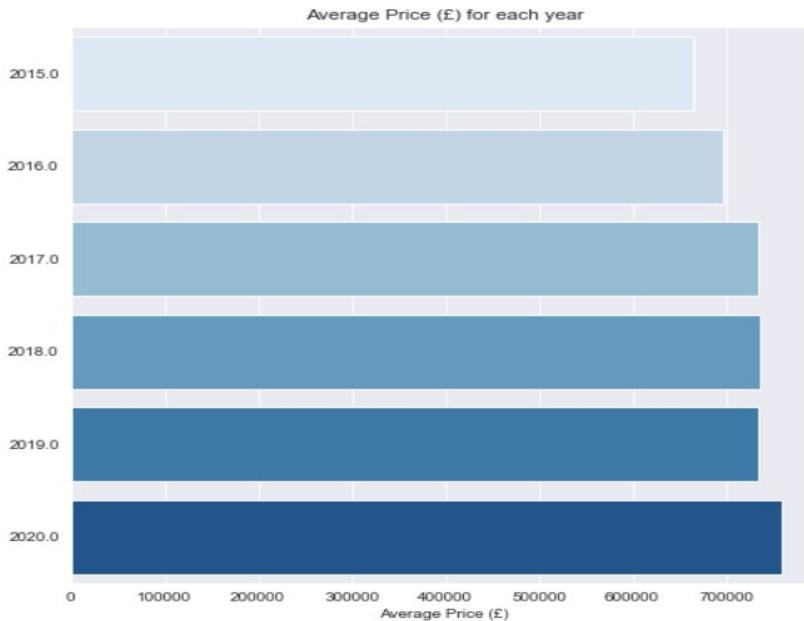


If we can establish the relative value of each postcode, then we can solve this problem



HM Land Registry Open Data

[UK House Price Index](#) [Price Paid Data](#) [Standard-reports](#) [SPARQL query](#)



- Over 320k transactions in five years
- Stable prices (no time series)
- Cutoff March 2020 (Covid19)
- Includes less than 70k postcodes, out of a total of about 200k unique London postcodes
- Data for some features not available across the whole time period

8 Primrose Road, London, E10 5ED

Transaction history

A 2019-07-12 £475,000 

Detailed address

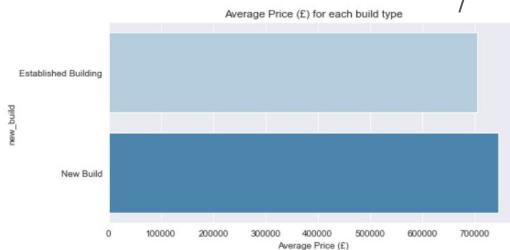
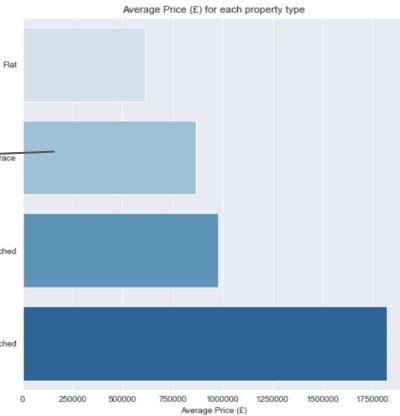
building name 8 
or number
street Primrose Road 
town London
district Waltham Forest 
county Greater London 
postcode E10 5ED 

Attributes

property type terraced

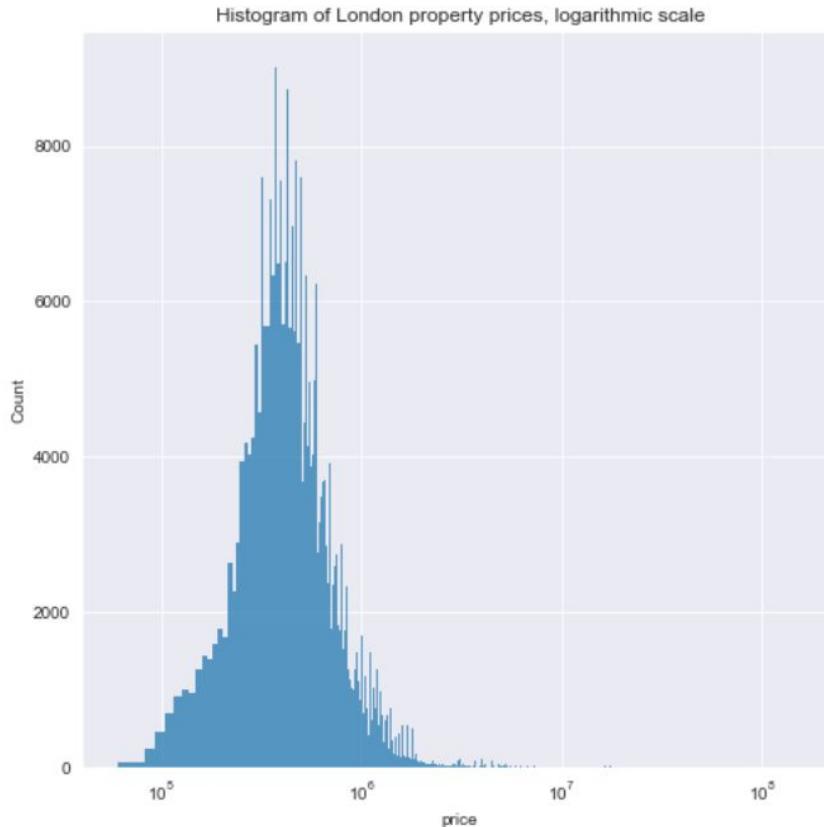
estate type freehold

new build? no



At this stage the plan is to work only with features engineered from the postcode data

The target variable (price) is highly skewed

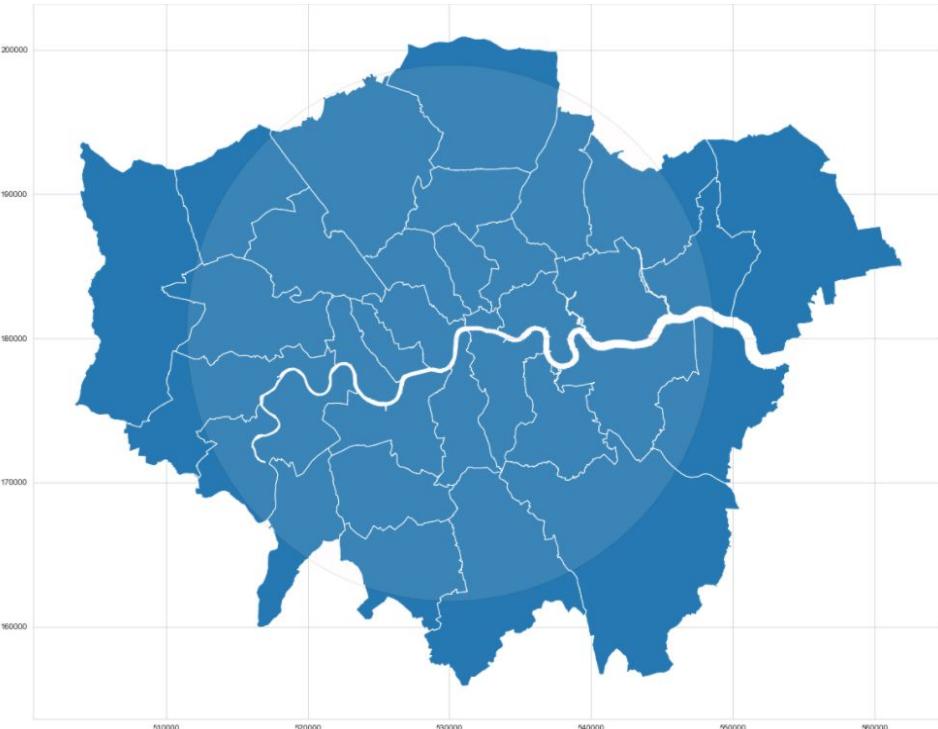


Prices at the upper end seem to be legitimate

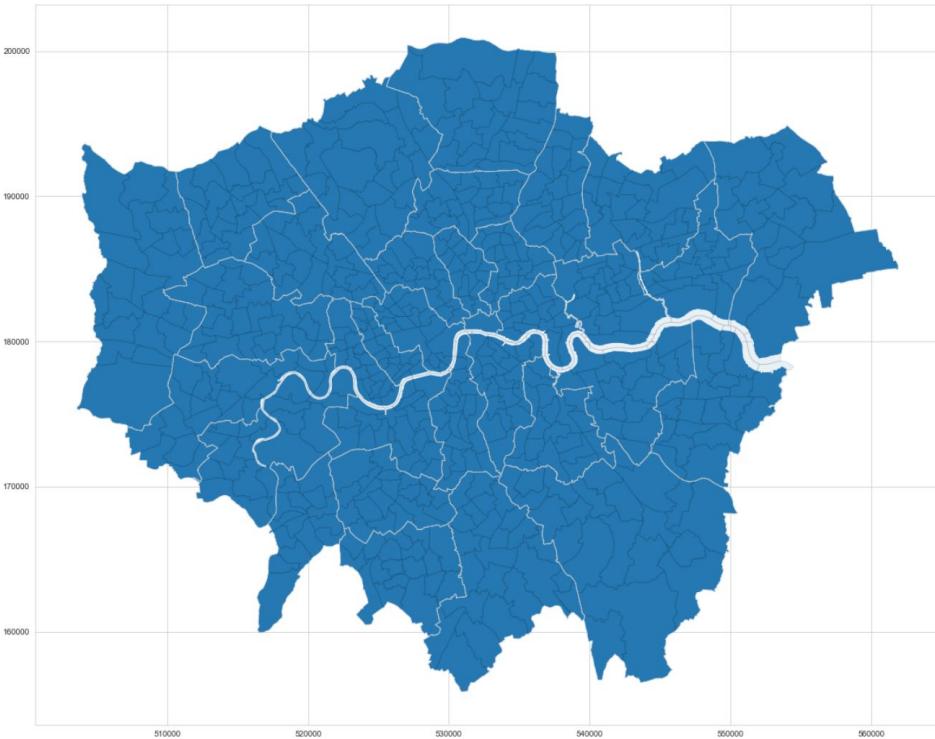
At the bottom end of the scale, many of the lower values look like they might be shared ownerships, short leases etc

Set a lower cutoff of £60k (slightly arbitrary, but it's unlikely that lots of misleading transactions will have got through.

How localised is a postcode?

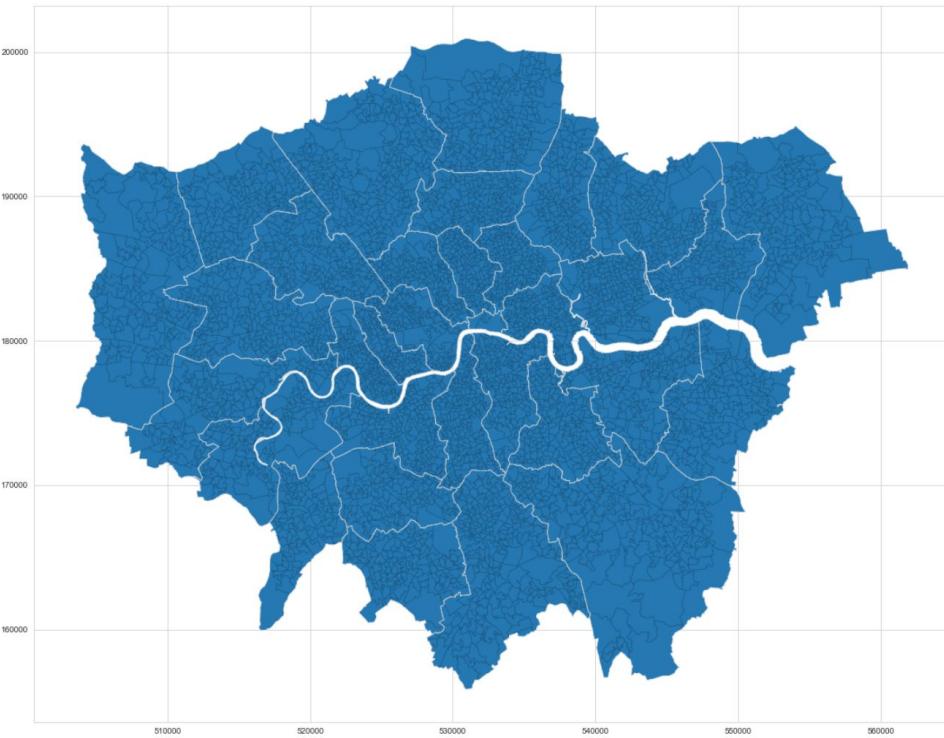


London is split into 32 boroughs, with an average population of roughly 250k



Each borough is divided into
20-30 wards

Average population about 10k



Each ward is divided into LSOA areas,
average population about 1500

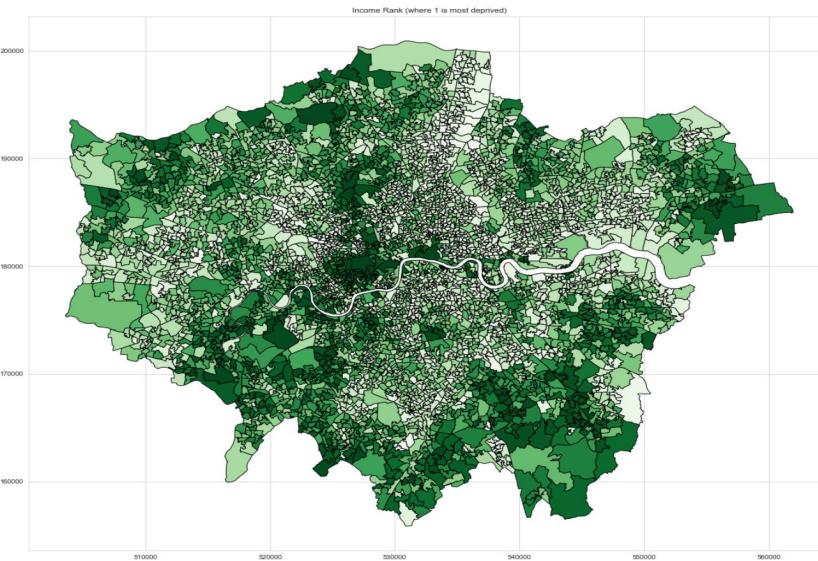
LSOAs are divided into unique
postcodes, which typically cover about
15 properties

Normally 30-50 postcodes per LSOA

FEATURES TO ENGINEER

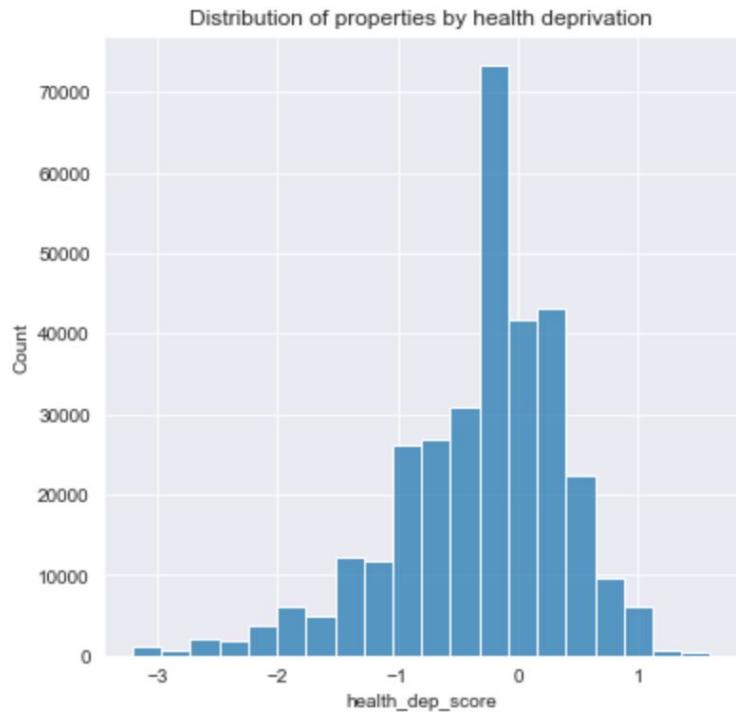
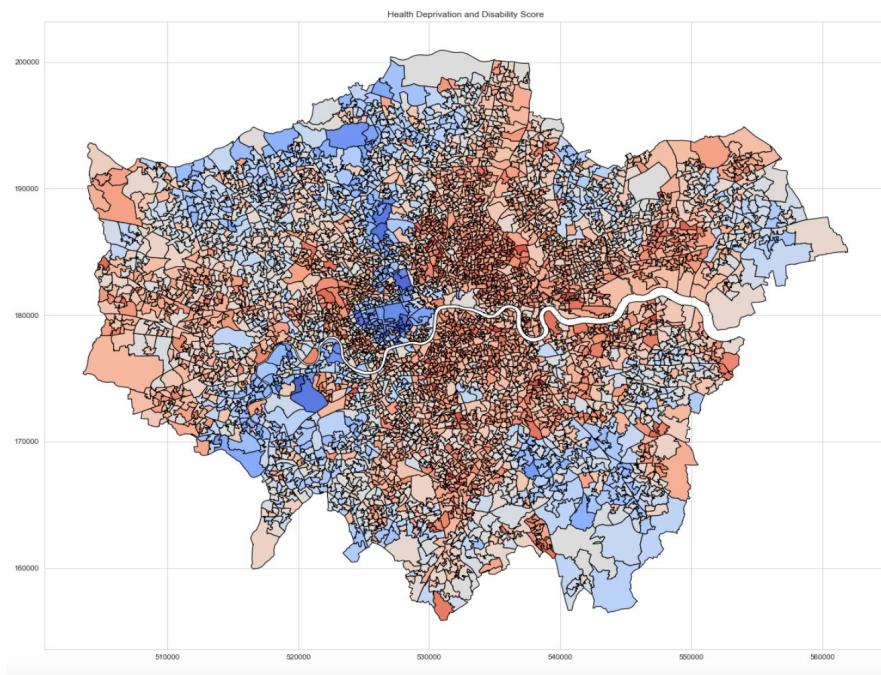
1. Economy
2. Crime
3. Education
4. Transport

1. Economic Factors

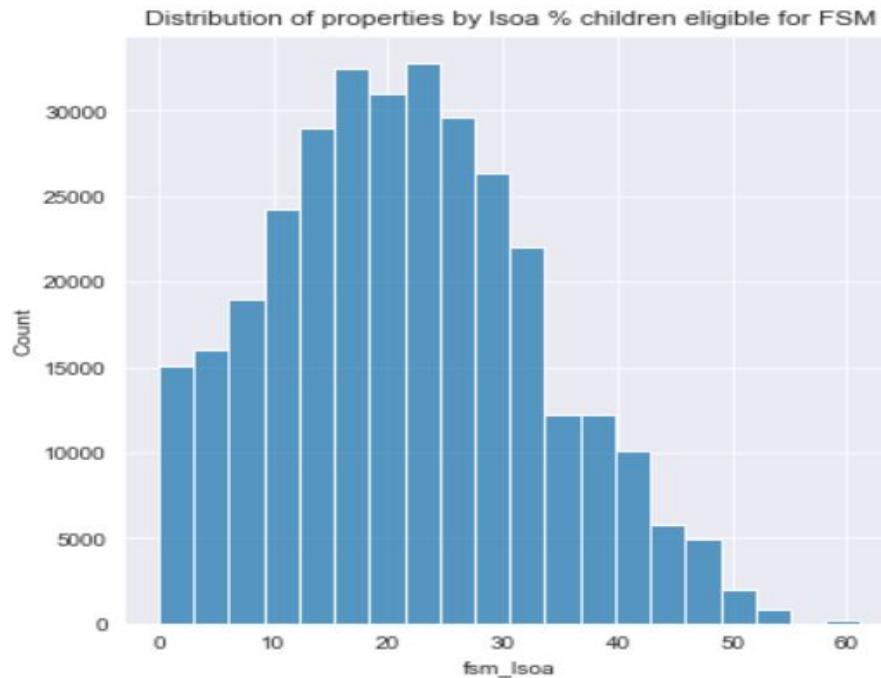


The Ministry of Housing released figures in 2019 for 'Indices of Deprivation'

These include Health, Employment, Education, and Living Environment



Greater London Authority 2015 figures on free school meals



Areas can become gentrified, it's possible that traditional metrics might lag behind



2k postcodes from allinlondon.com

Estate agents are likely to cluster where the property market is heating up



Scraped 5000 + estate agent postcodes



MENU



THE GOOD FOOD GUIDE

WAITROSE
& PARTNERS

LOG IN

JOIN US



MAYFAIR, CENTRAL LONDON

Jamavar



MAYFAIR, CENTRAL LONDON

Galvin at Windows



SPITALFIELDS, EAST LONDON

Hawksmoor Spitalfields



MARYLEBONE, CENTRAL LONDON

Fischer's



COVENT GARDEN, CENTRAL LONDON

Frenchie



HOLBORN, CENTRAL LONDON

The Holborn Dining Room



KNIGHTSBRIDGE, WEST LONDON

Pétrus by Gordon Ramsay

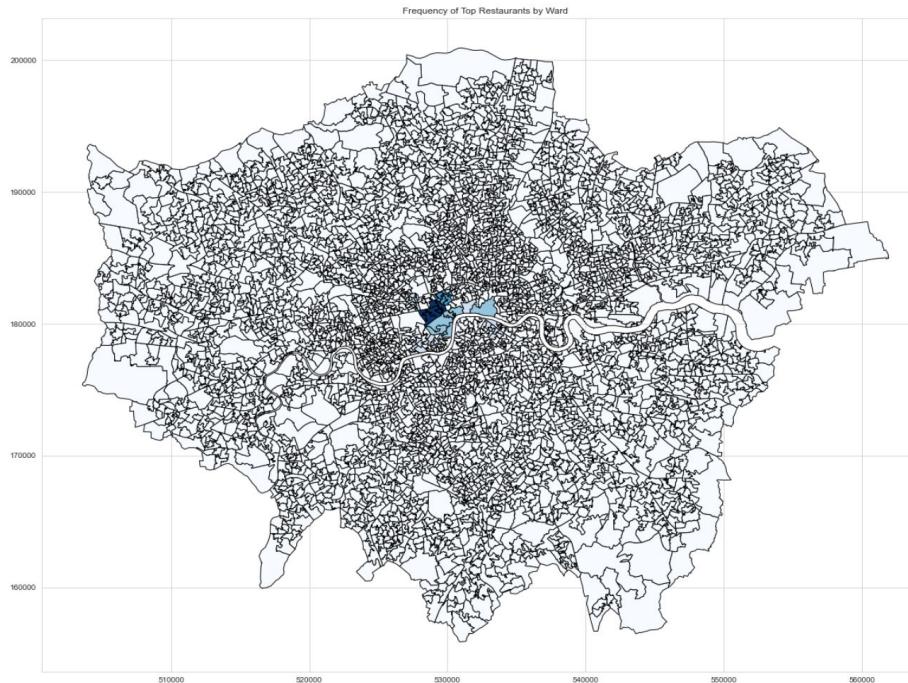


NOTTING HILL, WEST LONDON

Core by Clare Smyth

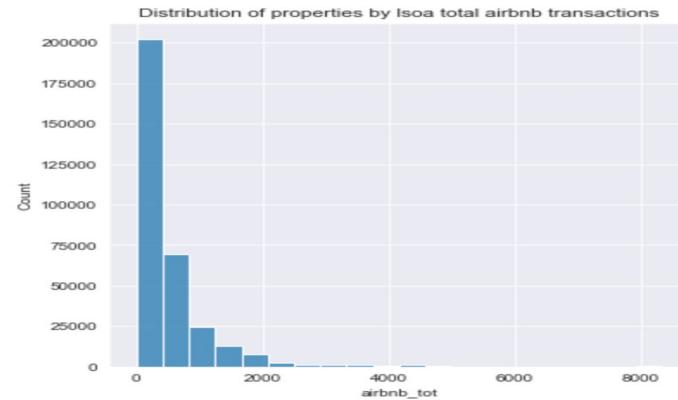
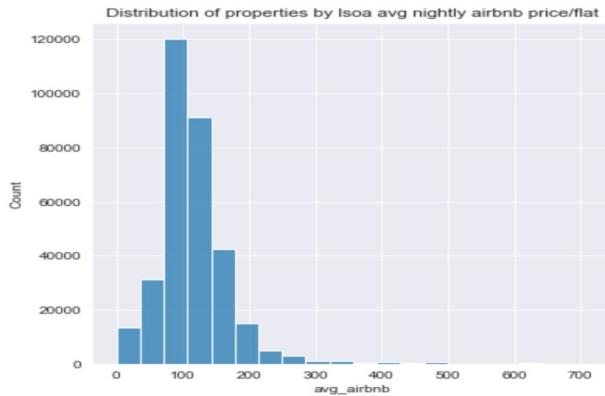


It turns out that top restaurants are almost all clustered in the West End



Not a measure of gentrification, but does have predictive value

How is Airbnb really being used in and affecting the neighbourhoods of your city?



It's possible that the rise of Airbnb has disrupted balance of supply and demand in some areas of the housing market

2. CRIME

Monthly LSOA crime data available from police back to January 2018

DATA.POLICE.UK Cymraeg

Home Data API Changelog Contact About

[Home](#) >

Data downloads

Custom download Archive Boundaries Open data Statistical data

These CSV files provide street-level crime, outcome, and stop and search information, broken down by police force and [2011 lower layer super output area \(LSOA\)](#).

The Police Service of Northern Ireland does not currently provide stop and search data.

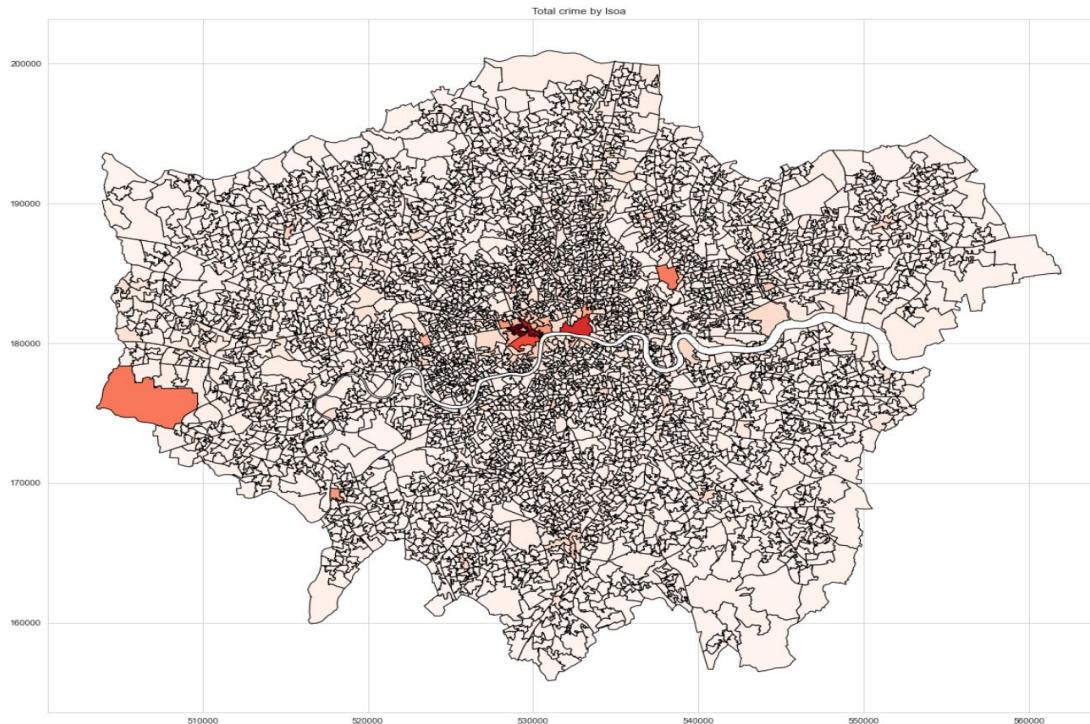
See the [changelog](#) for known data issues, and the [about page](#) for a description of each column in the CSV files.

Date range: to

Forces: All forces

Difficult to get a useful predictor from crime, we find a positive correlation between crime and property prices...

Police data comes in broad categories (robbery, antisocial behaviour etc), not always easy to disentangle, data isn't broken out offences like murder, shootings etc.

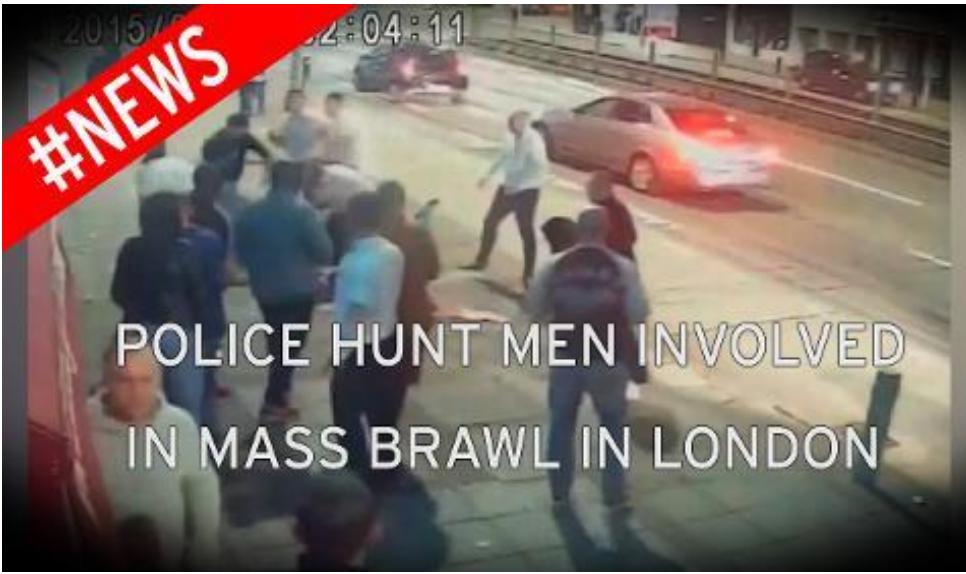


It looks like the West End is to blame for the positive correlation

People visit the West End for many reasons...



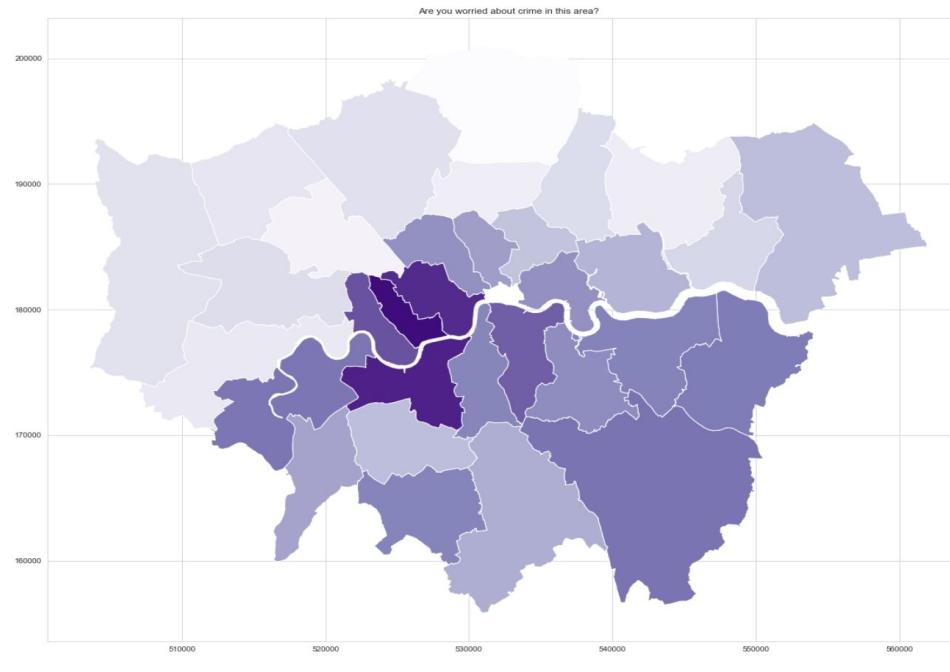
#NEWS





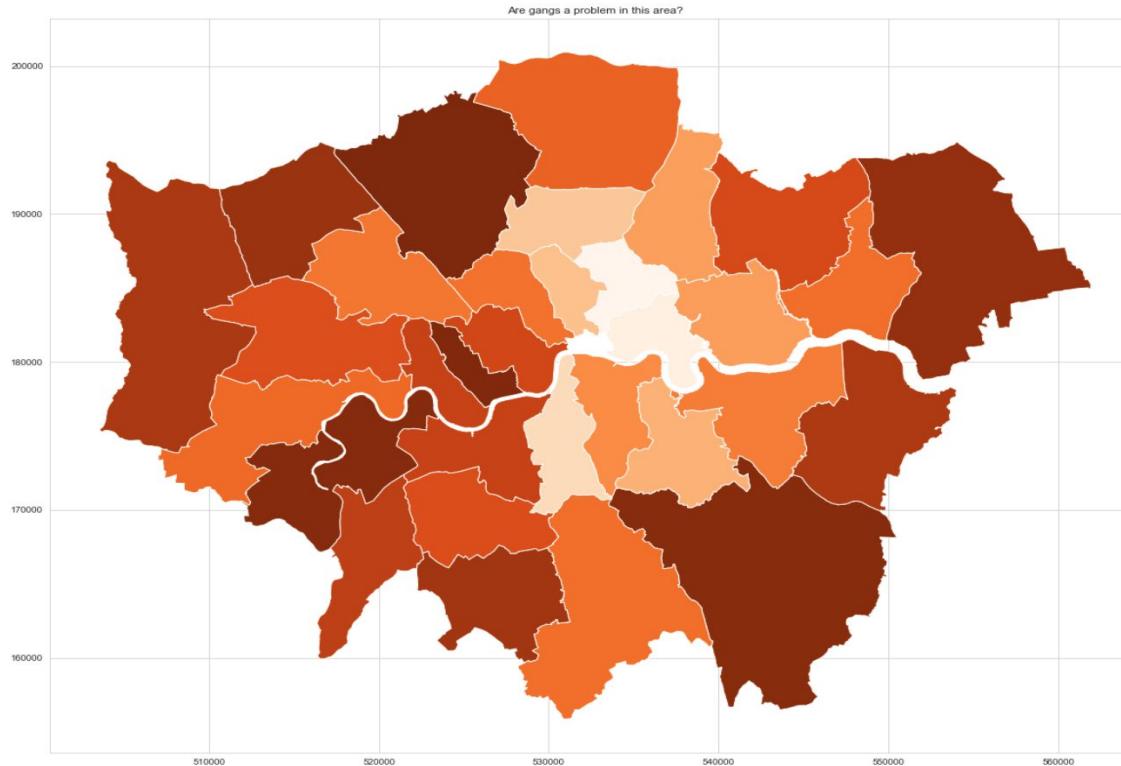


Since late 2016, the Mayor of London's Office for Policing and Crime has commissioned quarterly surveys into the public's attitudes towards crime



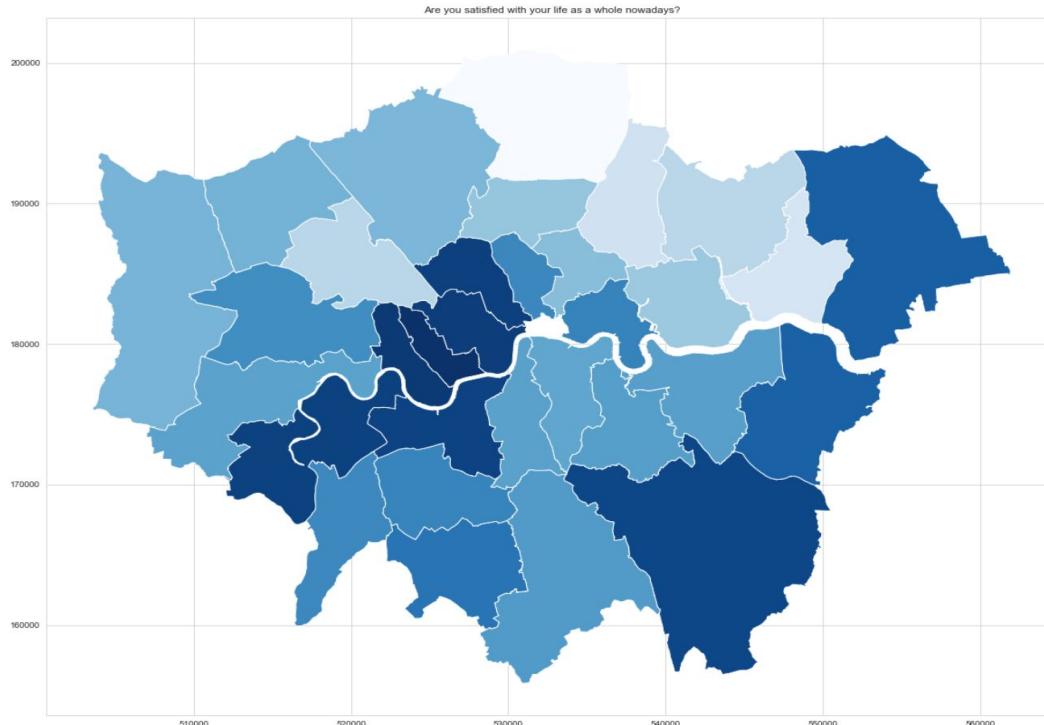
Residents of Westminster don't seem too worried...

Survey responses have more predictive value than overall crime figures



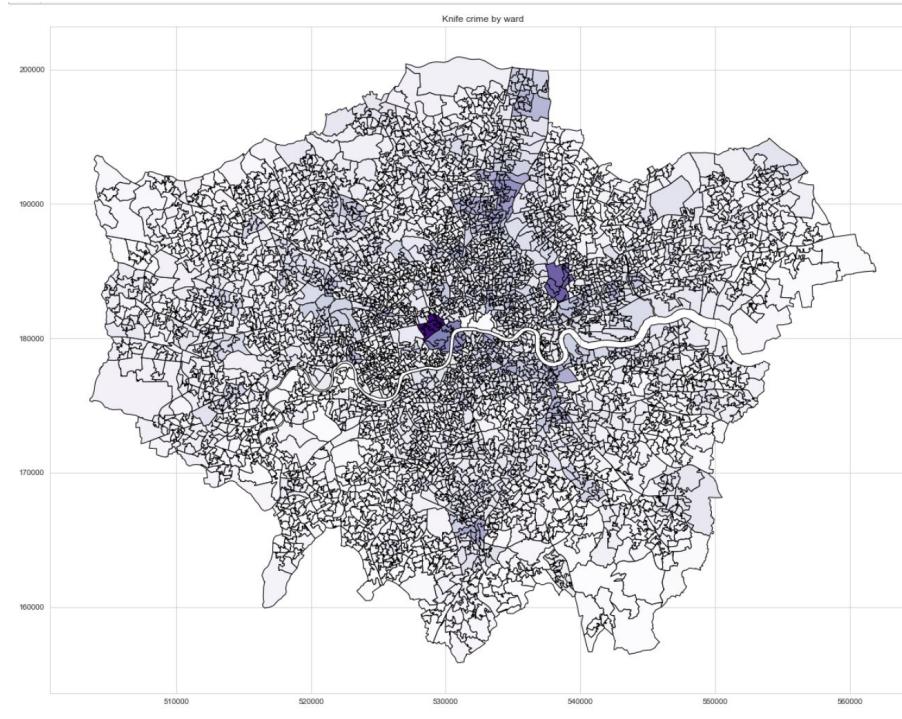
Residents of Hackney and Lambeth seem understandably concerned

Unfortunately this data is less granular than other sources



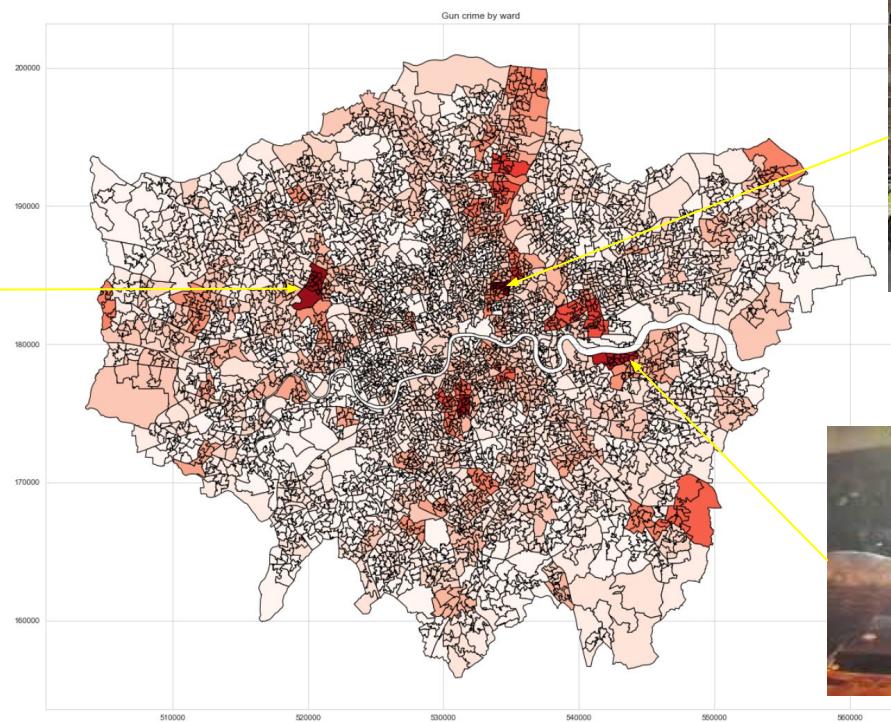
Residents of Camden, Westminster, Kensington & Chelsea are all very satisfied

The London Assembly provides rolling twelve month averages for knife crime by ward



Again the West End is unhelpful...

Gun crime is a more useful predictor



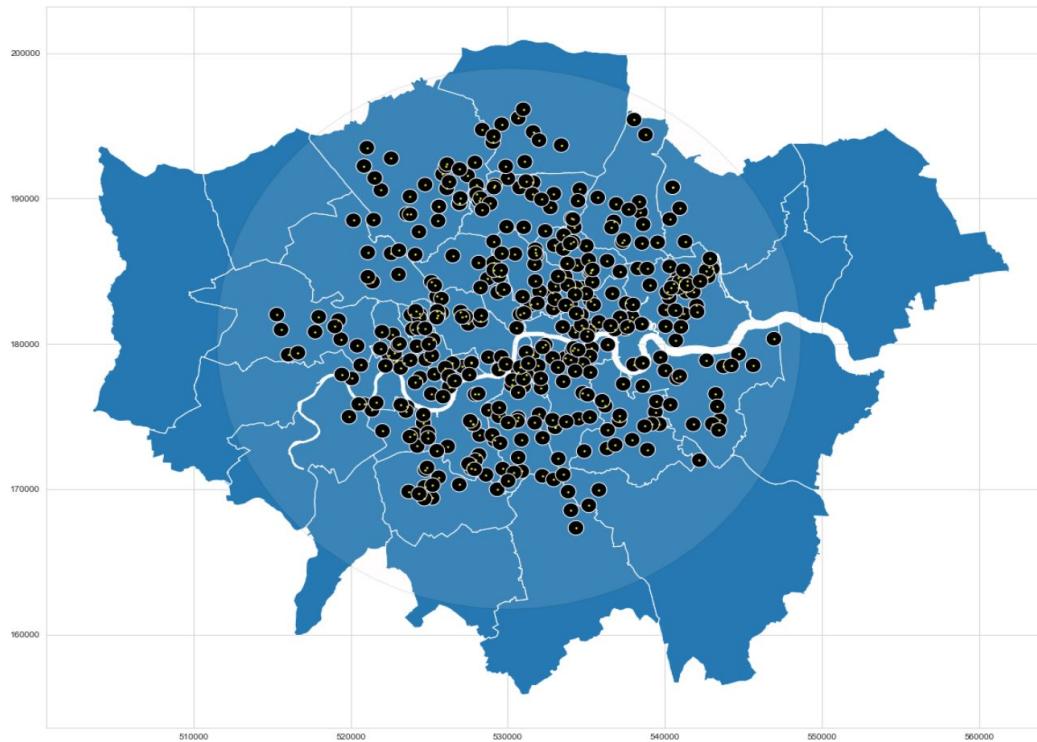
Likely that gun crime is usually linked to turf wars over drugs

3. EDUCATION

Ofsted provides data on ratings for state-funded schools

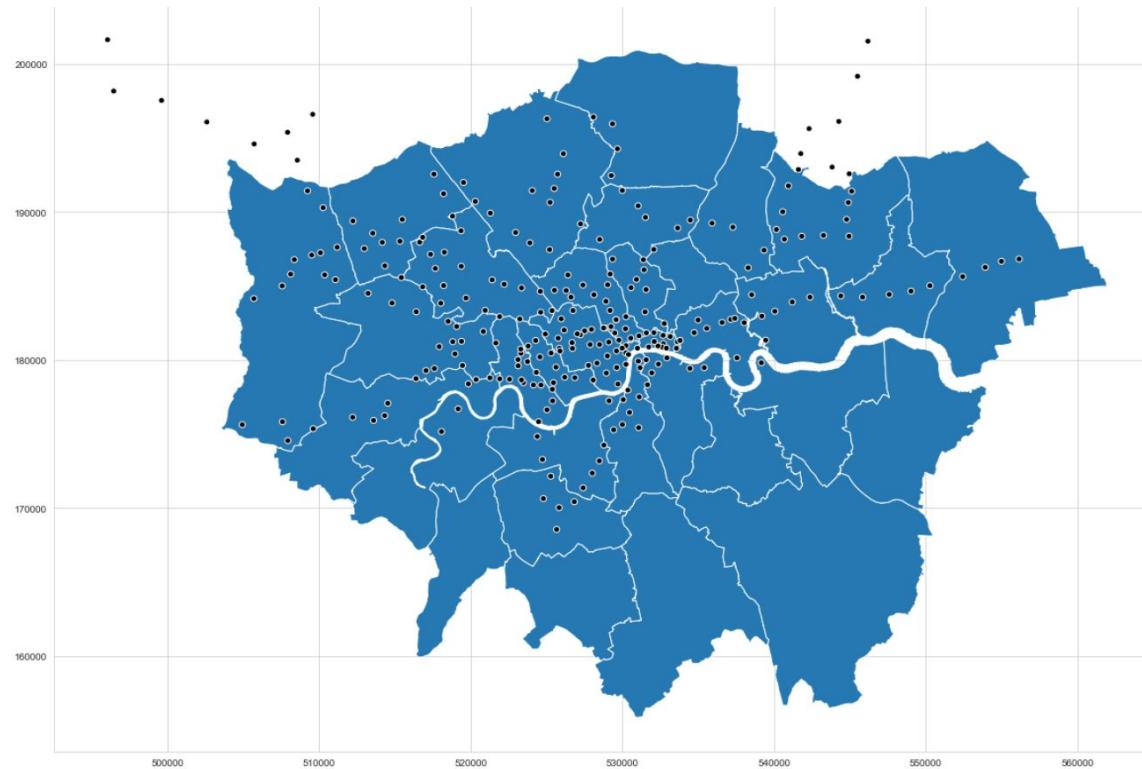


Data as of December 2019, 500m radius around 'outstanding' primary and secondary schools

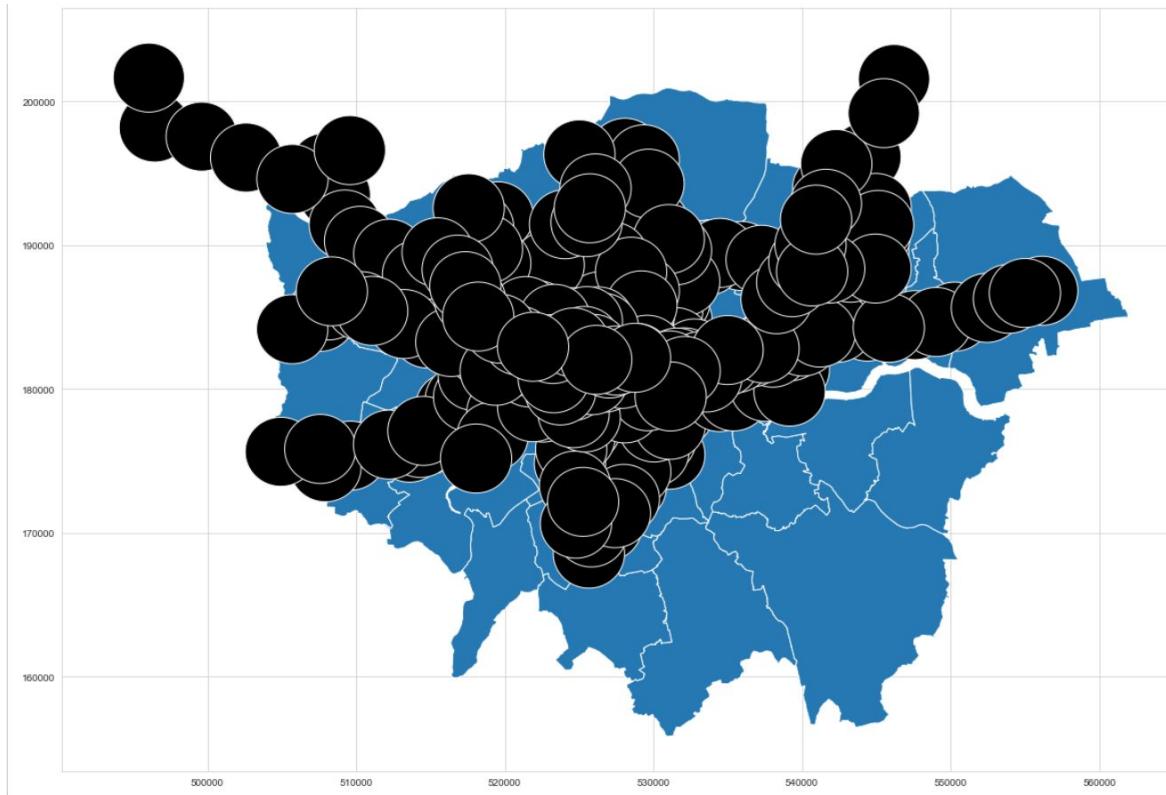


4. TRANSPORT

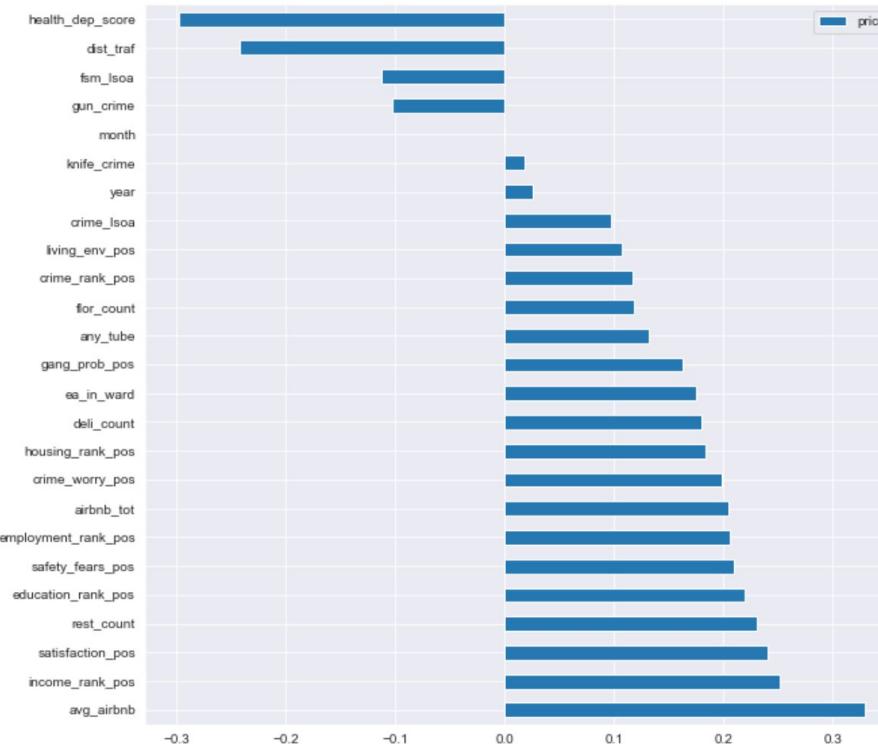
Geopandas lets us check if any property is within 50m of a tube stop, or 200m, 400m etc and create categorical features



Nobody is likely to walk more than 2400 meters



Correlations the numeric features and target variable



Linear Regression, Ridge, Lasso, ElasticNet - R2 scores around 0.21

| | | | | |
|-----|----------------------|---|-------|------|
| 65 | Kingston upon Thames | F | 38 | 0.26 |
| 5 | Bexley | F | 236 | 0.27 |
| 89 | Richmond upon Thames | F | 854 | 0.36 |
| 32 | Enfield | F | 3073 | 0.38 |
| 101 | Waltham Forest | F | 7455 | 0.46 |
| 77 | Merton | F | 3872 | 0.47 |
| 85 | Redbridge | F | 1402 | 0.50 |
| 24 | Croydon | F | 2663 | 0.52 |
| 53 | Hounslow | F | 1381 | 0.56 |
| 28 | Ealing | F | 6469 | 0.59 |
| 48 | Haringey | F | 7538 | 0.59 |
| 1 | Barnet | F | 11142 | 0.60 |
| 36 | Greenwich | F | 11990 | 0.60 |

It turns out that property types aren't evenly distributed, so property type will need to be included in the model.

(94% of properties in Tower Hamlets are flats, only 26% in Kingston upon Thames)

'How much does a property change in value when moved between postcodes?' - Four answers, one for each property type.

Include ‘property type’ as a feature along with purely geographical features

Linear Regression - 0.26

Lasso - 0.26

Ridge - 0.26

ElasticNet - 0.26

AdaBoost - 0.31

Gradient Boosting Regressor - 0.43

KNN - 0.50

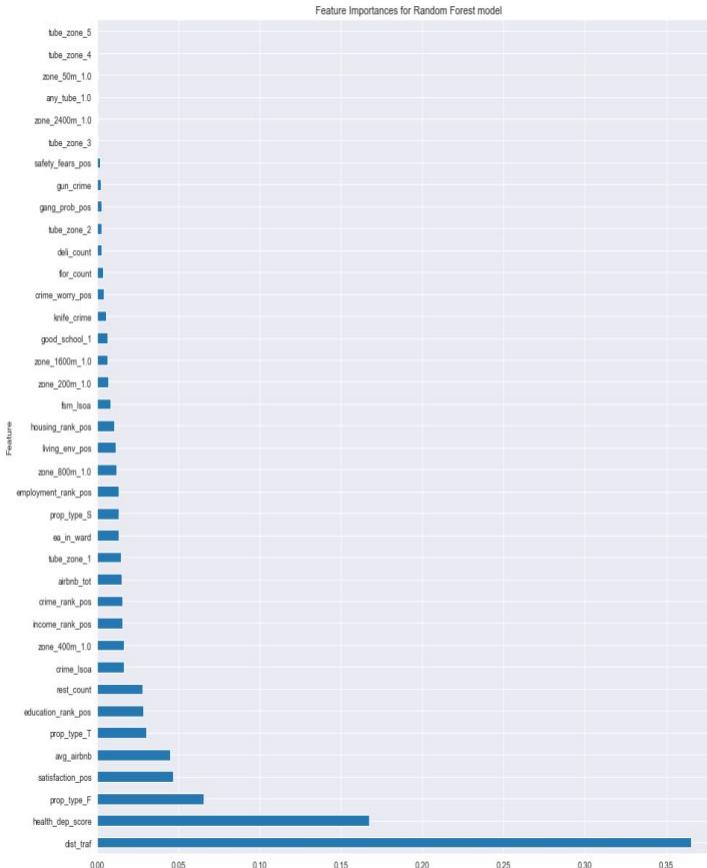
Random Forest - 0.52

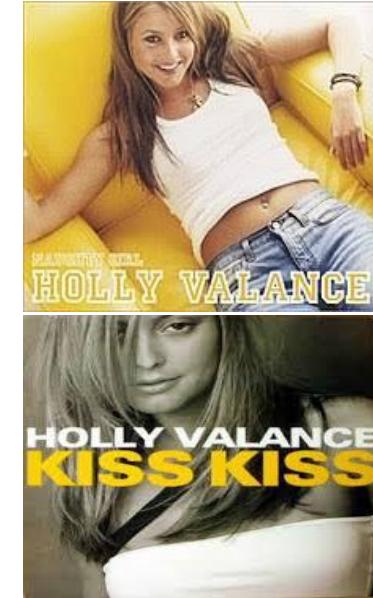
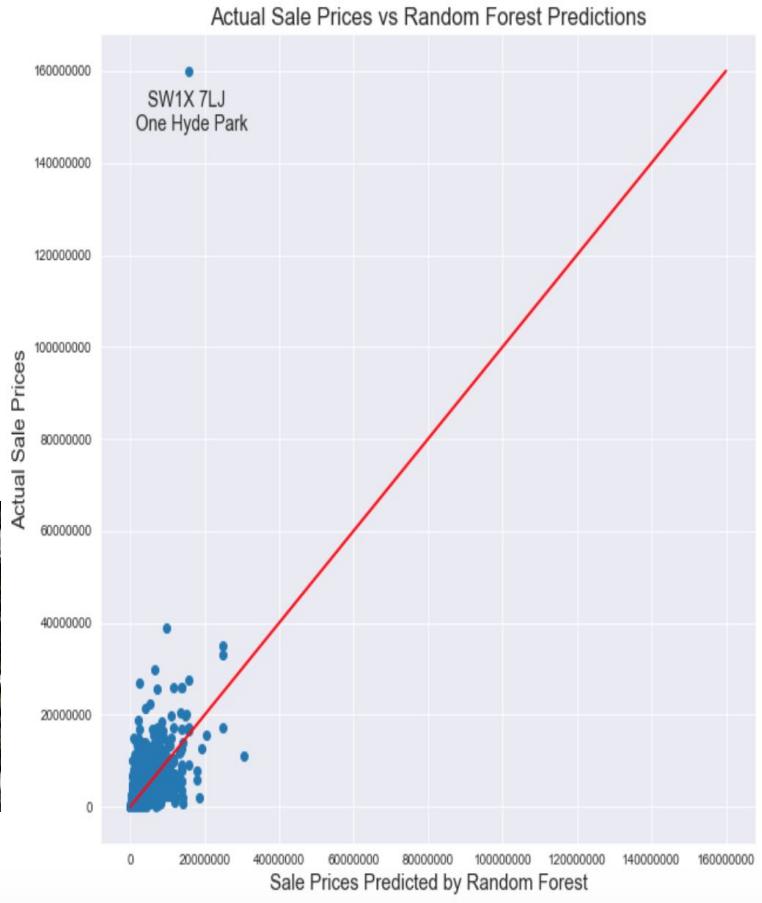
Things that didn't help:



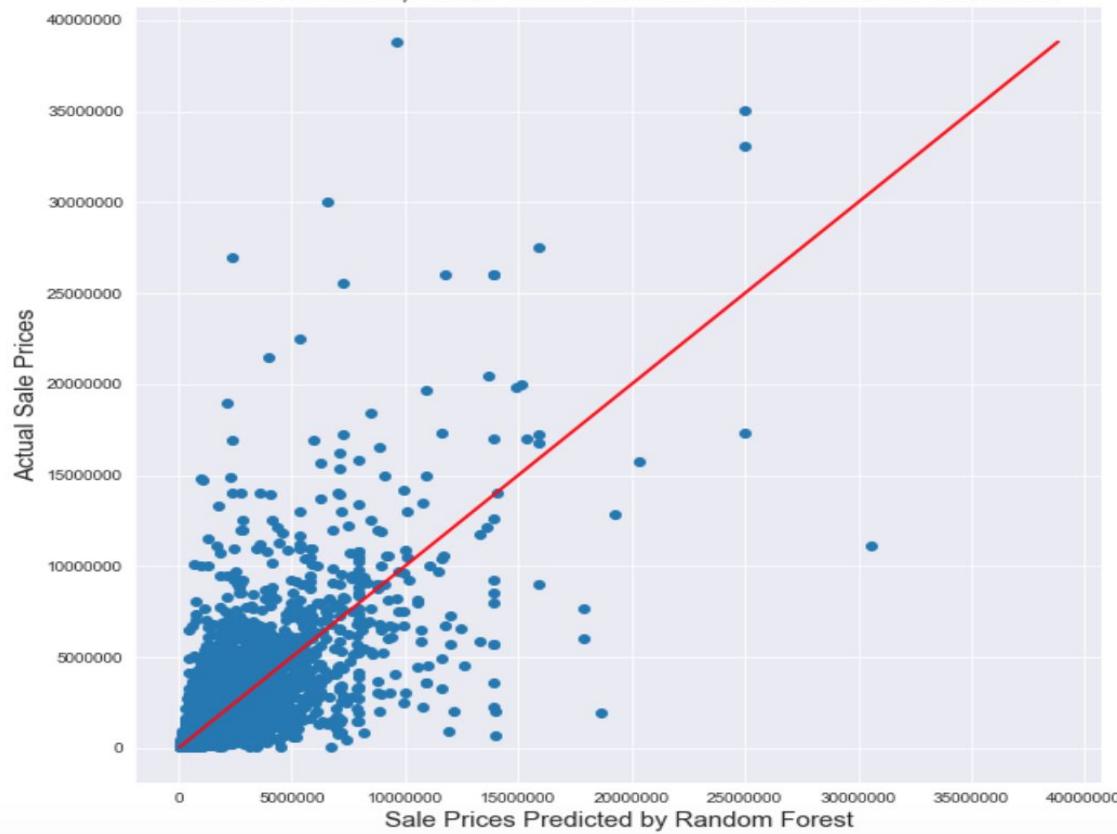
- PowerTransformer on target variable.
- Polynomial Features (all features and interaction features only).
- Power Transformer on predictor variables.
- Dropping 2015-2016.
- Various combinations of the above.

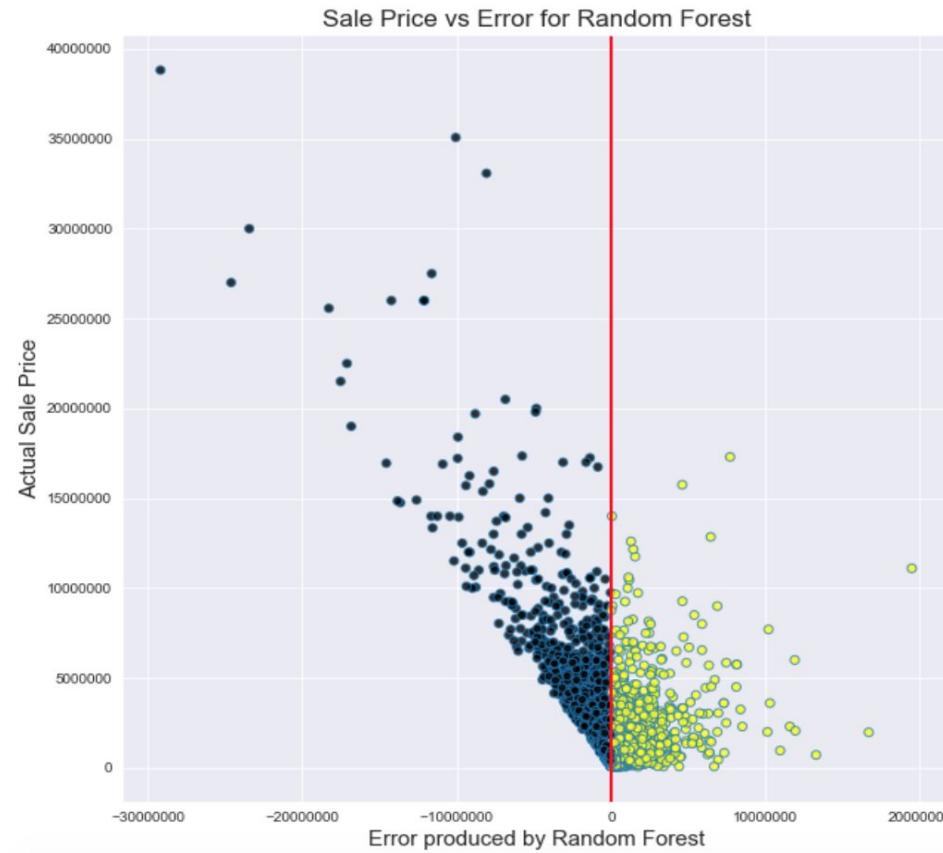
Winning model was Random Forest, r2_score = 0.526





Outlier Removed, Actual Sale Prices vs Random Forest Predictions





Now we can try to answer the original question - how much value is destroyed by moving a flat from Mayfair to Eltham?



£11,000,000



2 bedroom apartment for sale

Grosvenor Square, London, W1K

BRAND NEW INSTRUCTION - SOLE AGENT - A rare lateral apartment on the 4th floor with 5 west facing windows overlooking the square gardens.

Added on 03/11/2020 by Wetherell, London



020 7529 5566

Local call rate



£110,000

Guide Price ?



2 bedroom maisonette for sale

Oakways, Eltham, SE9

A Self-Contained Two Bedroom Upper Maisonette with accommodation arranged on Two Floors.

Added on 06/02/2021 by McHugh & Co, London



020 8012 8911

Local call rate



Create a new dataframe with all 200k London postcodes

| postcode | latitude | longitude | lsoa | fsm_lsoa | ward_code | ward_name | borough | gang_prob_ | crime_worr | ... | zone_200m | zone_400m | zone_800m | |
|----------|----------|-----------|-----------|-----------|-----------|-----------|---------------------------|---------------|------------|-----------|-----------|-----------|-----------|-----|
| 0 | E1 1BY | 51.519554 | -0.059467 | E01004208 | 39.16 | E05009317 | Bethnal Green | Tower Hamlets | 69.360995 | 71.860724 | ... | 0.0 | 0.0 | 0.0 |
| 1 | E2 0ET | 51.527207 | -0.055218 | E01004209 | 41.06 | E05009317 | Bethnal Green | Tower Hamlets | 69.360995 | 71.860724 | ... | 0.0 | 0.0 | 0.0 |
| 2 | E1 7AN | 51.515408 | -0.072613 | E01004309 | 52.18 | E05009333 | Spitalfields & Banglatown | Tower Hamlets | 69.360995 | 71.860724 | ... | 0.0 | 0.0 | 0.0 |
| 3 | E1 7PS | 51.515023 | -0.073249 | E01004310 | 38.95 | E05009333 | Spitalfields & Banglatown | Tower Hamlets | 69.360995 | 71.860724 | ... | 0.0 | 0.0 | 0.0 |
| 4 | E1 7PT | 51.515372 | -0.072487 | E01004309 | 52.18 | E05009333 | Spitalfields & Banglatown | Tower Hamlets | 69.360995 | 71.860724 | ... | 0.0 | 0.0 | 0.0 |

Multiply by four property types to create 800k fictional properties, and use RF model to predict prices for each

Write a function to ‘move’ a property from one postcode to another

```
In [198]: 1 convert_postcode('W1K 2HR', 'SE9 2NZ', 'flat')
Moving the flat from W1K 2HR in Westminster to SE9 2NZ in Greenwich is a journey of 16481 meters.
Its value would change by -96%
```



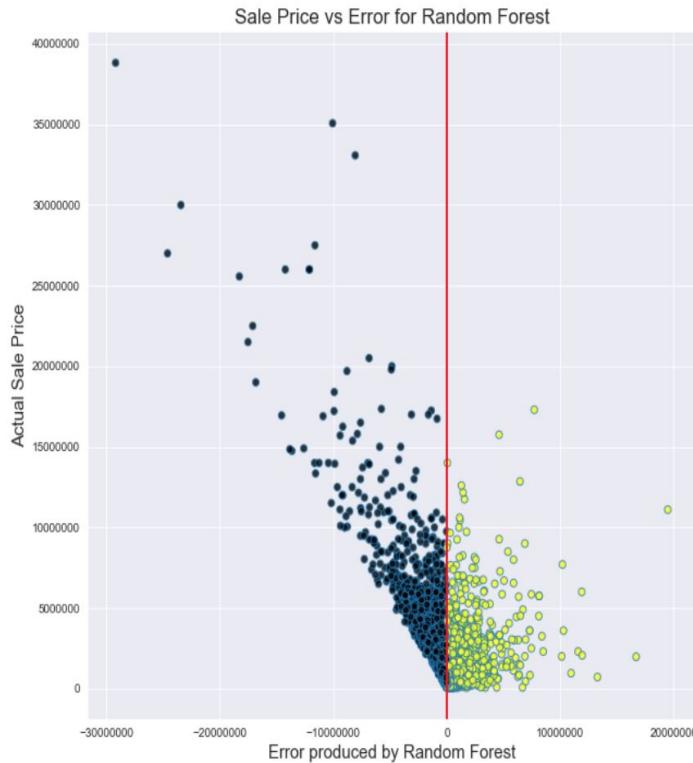
The model suggests the Mayfair flat would be worth £440k if moved to Eltham (down from £11m)
This seems in the right ballpark, but slightly harsh - high end two bed flats in Eltham advertised for > £500k.

Next steps and improvements



The features generally work quite well, but we could definitely add some more - parks, river, overground trains etc.

The model also needs a more fundamental improvement before it could be trusted with real money



Even when we log transform the target variable, the residuals still follow a similar pattern

It's likely that the postcode data is interacting with other missing data (for example overall quality)

Model might also be struggling to cope with irrational behaviour at higher price points

Does the model solve the problem with the estate agents and the sparse matrix?



- The postcode features are a reasonable start, but need to be layered on top of more traditional data (square footage, num bedrooms, overall quality) to get the most value from them
- We may also need to find a way to filter out the most expensive properties (probably on a combination of size, overall quality and location)

THANK YOU!



QUESTIONS?