**Ideas for analysis**

**Keep it very simple** – a couple of maps, charts will be fine. Surely at this stage they are testing your use of Python and critical thinking rather than anything data science-y.

**Be prepared to work backwards – you might need to change the research question based on the data and analysis that is possible.**

**All data is already in the Census notebook – three core datasets at the end – take what you need. Includes Sadia’s private rented data. Join with the AirBnB counts from Paul. Feel free to add more variables (espec by date).**

**I wouldn’t worry too much about following the Batty paper exactly, just simple exploratory analysis, correlations, basic model and findings should be fine.**

**Maps** – Airbnb by borough (as % of all dwellings), private rents by borough – should look very similar.

Worth mentioning very **local variations at LSOA level** (AirBnB / dwellings) – should be quite a range with the highest LSOAs in the North, Inner London boroughs.

**Compare average Airbnb rents by room size and Inner / Outer London against the same rents for the private rented sector (**I’m not sure how robust doing that by borough would be)**.** They should be higher – if not, how can we claim they are driving up rents?

Do a histogram of Airbnb and the Private rents – if not normal you might need to transform and standardise for further analysis.

I would do a series of scatterplots:

**AirBnB against (use the core housing dataset at the bottom of the notebook) – you can also check against demographic data:**

* **Vacant properties**
* **House Prices (all variables)**
* **Private rents**
* **Homelessness** – I have included those with a preventative duty – ie those at high risk of becoming homeless without actually being accepted as such. I think this is more interesting to work with.
* **Accommodation type (**especially flats**), cars** (should correlate negatively – Airbnb should go up as cars go down assuming a link with central areas). **Addresses** is an interesting one – I would hazard a guess that AirBnB goes up, the number of people who have lived in the same place for a year goes down – kind of sucks for the local community. **PopDensity** – will positively correlate – an Inner London thing. **NsSEC** – self determined social class, will be interesting – I guess there’ll be lots of senior managers etc – Airbnb in already high-income places.

**Regression** – check this model out:

1. **Private rents (possibly transformed - log) ~ Airbnb, House Prices data (all), Homelessness (espec end of AST, landlords wishing to sell), Vacant Properties, Inner / Outer London flag.**

Some variables will correlate with each other – in the regression see if you can find a thing called **Variance Inflation Factors** – Andy touched on them this week – any variable with a VIF of over 10 should be removed – if you still have a lot of variables go down to a VIF of 5.

Or more simply, you can also look at the significance of each variable and remove those that are not significant.

Remember **parsimony** – smaller models are always preferred to larger models if the accuracy is roughly similar – so don’t include 10 extra variables for the sake of an increase in 1% on the R-squared.

Try and keep AirBnB in the model, and determine its significance.

1. **Use the same but including the demographic variables to predict the AirBnB prevalence.**

See what you come up with - what sort of areas have high prevalence of AirBnBs?

Main point – which you will need to discuss (simply) - **Airbnb is adding complexity to an already complex market** – my guess given Pauls map that there are concentrations of AirBnB around K&C, Camden, Westminster, TH, H&F – areas with the highest house prices (especially for first time buyers and newly built properties), already have the high private rents, probably quite a few long term empty properties (except H&F – that might be my fault), Westminster has astronomic homelessness, all high income, all very densely populated, all central to the key tourist sites (and transport hubs be it train stations or airports).

How can these boroughs have “high” everything?

Could make a case for licensing within Inner London? AirBnB licencing (or banning) will help the housing crisis, but not solve it.

**Other technical stuff :**

Remove the old dataframes as they’ll be taking up memory.

**Issues with data other than Census**

Gaps in London in the homelessness data – a few boroughs did not provide data, but should still get some interesting correlations.

**Census 2021** – Paul will know this – first time it was done online (bias), done in Summer, done in the middle of COVID. I don’t know what Camden and Tower Hamlets think, but there is no way in hell the population of Hammersmith and Fulham is rising so slowly.

**Rents** – I don’t know how complete the picture is, there might be some issues to discover here.

**Useful links for data**

These are already in but for information :

[Tables on homelessness - GOV.UK](https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness)

[UK House Price Index: data downloads September 2024 - GOV.UK](https://www.gov.uk/government/statistical-data-sets/uk-house-price-index-data-downloads-september-2024#download-the-data)