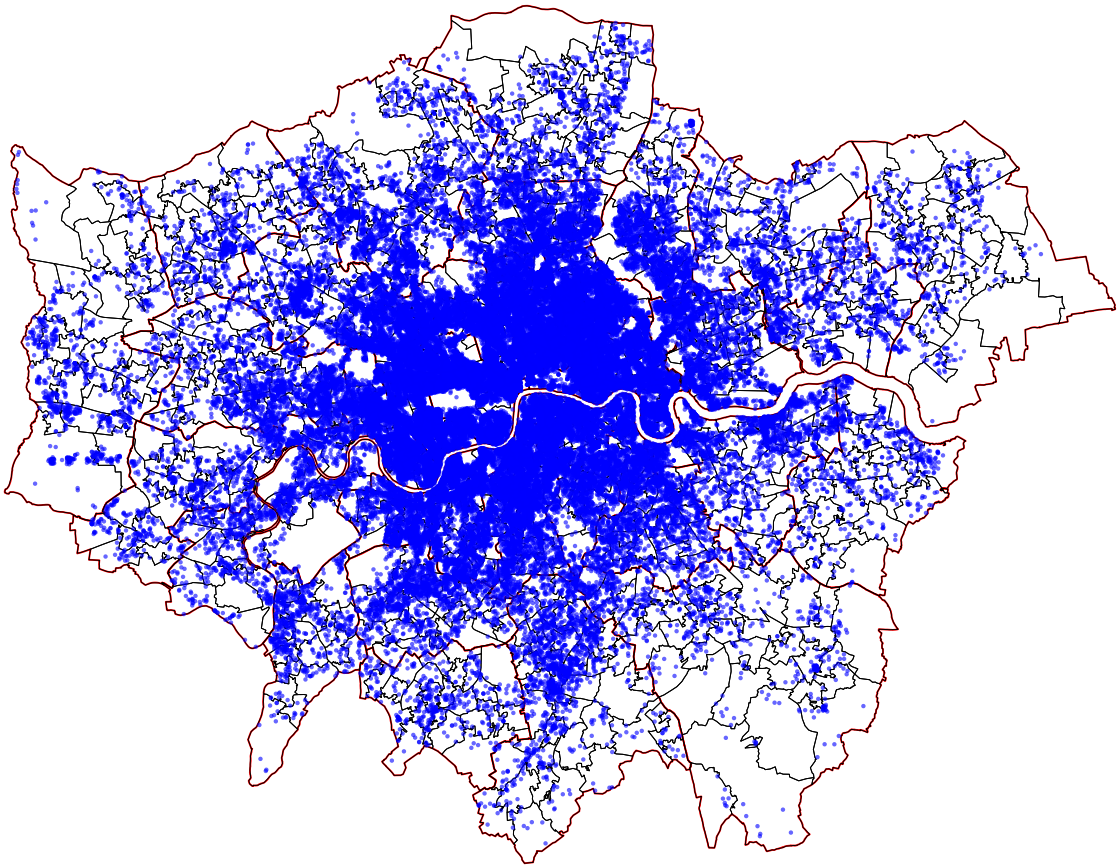
Spatial Analysis of Multi-Host Airbnb Activity in London

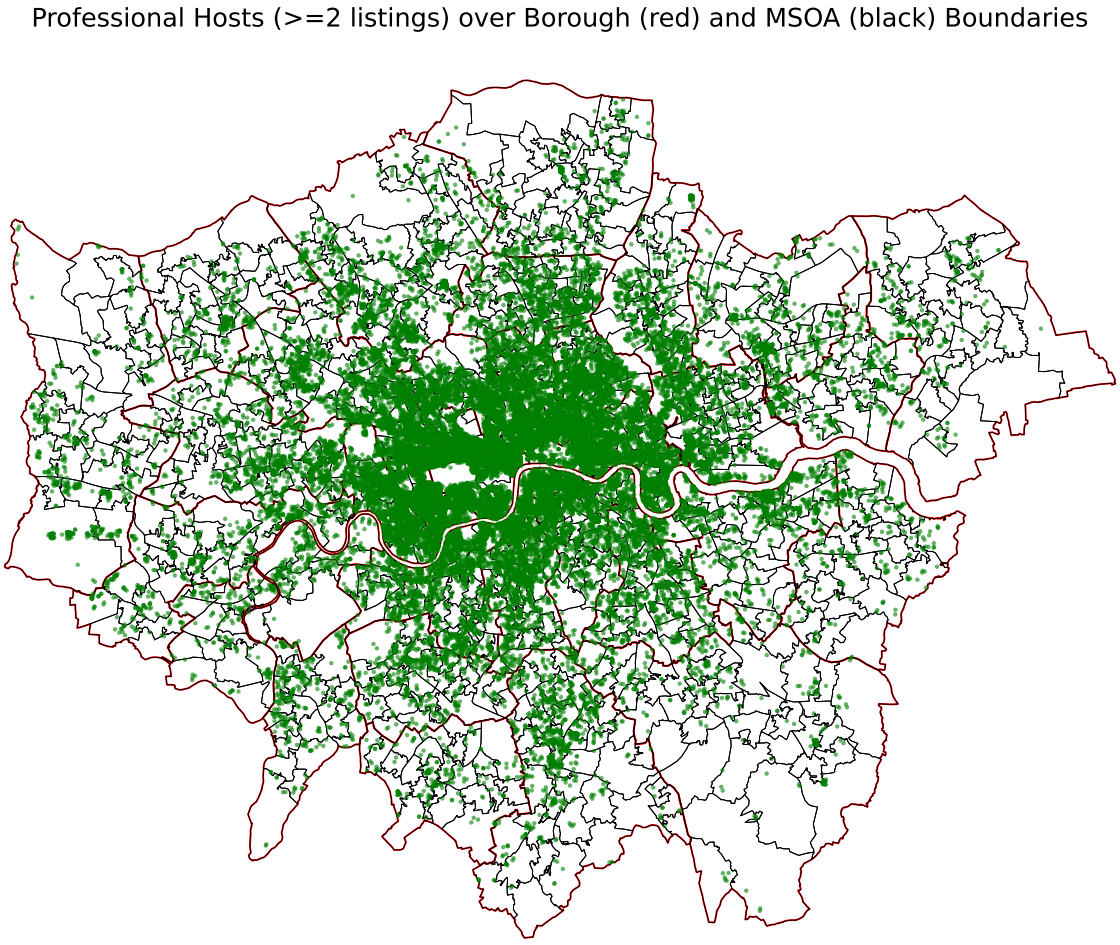
Christian Mulder

2025-11-01

To begin with, I will start by reading in all the necessary files, which are the listings CSV, the borough boundary data, and then the MSOA data.

All Listings over Borough (red) and MSOA (black) Boundaries

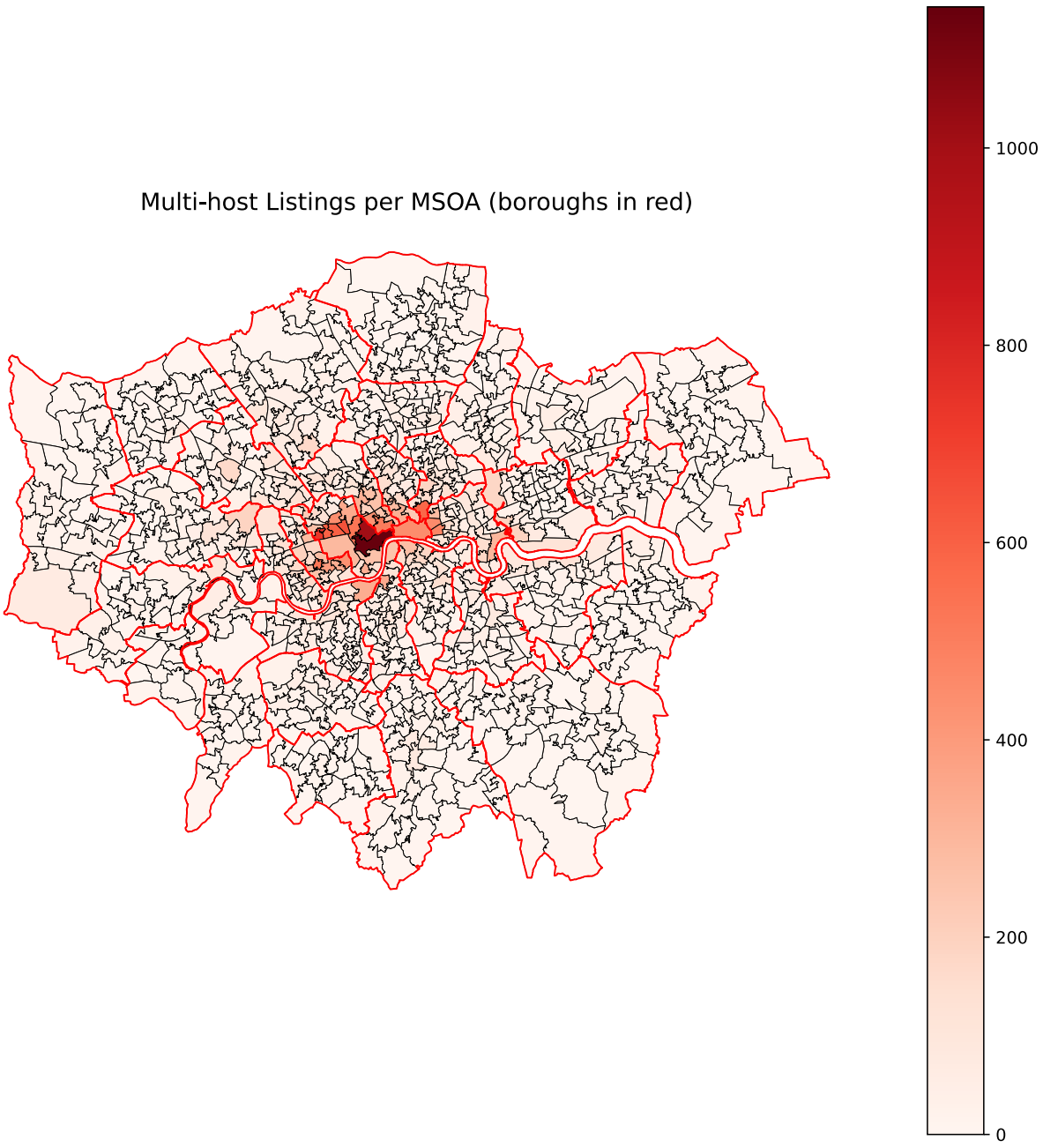


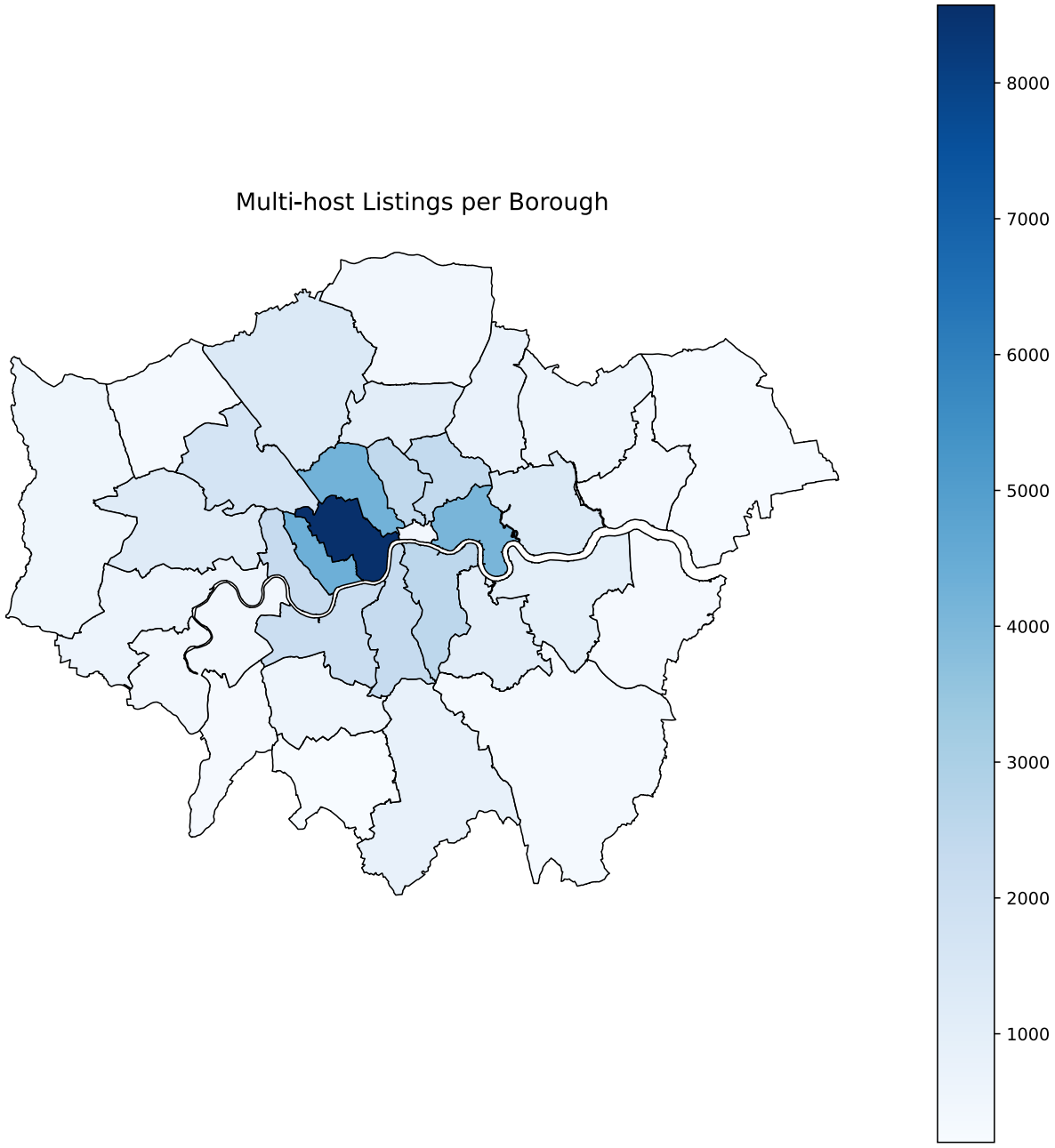


In order to further analyse, I will take these following steps: - Counts and distributions (MSOA/borough, normal/log). - Normality checks (Shapiro–Wilk). - Gini coefficient (MSOA and borough) → here it acts as a descriptive inequality measure alongside distributions. - Density visualization (KDE attempt, maps). - Global spatial autocorrelation (Moran's I). - Neighbor definition comparison (KNN vs Distance Band). - Spatial Error Model. - *Local clustering (Getis–Ord Gi)**. - Return to inequality in discussion → you can revisit the Gini in the conclusion, tying it to spatial clustering.

Counts and Distributions

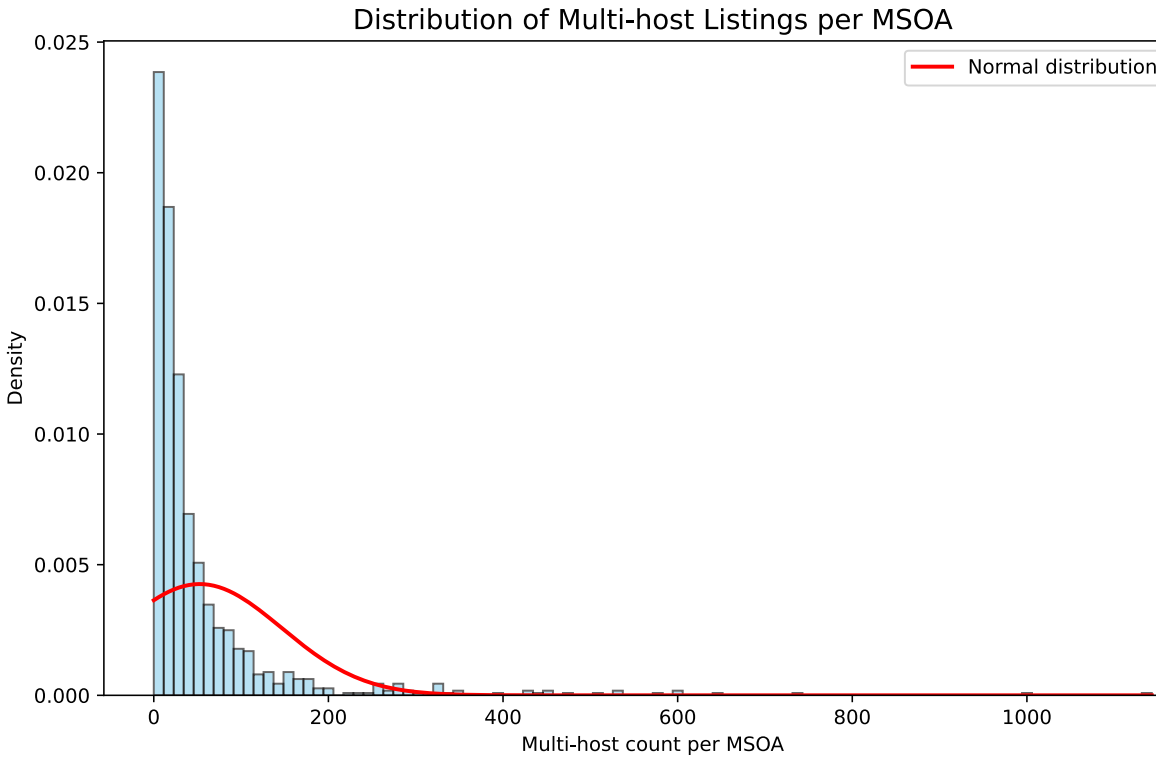
First, the following two maps will be the distribution of counts (tallied per area) for the Borough and MSOA data. ##### MSOA:


Multi-host Listings per MSOA (boroughs in red)
1000
800
600
400
200
0



Borough:

Now that I have plotted these, I want to create a basic distribution analysis.

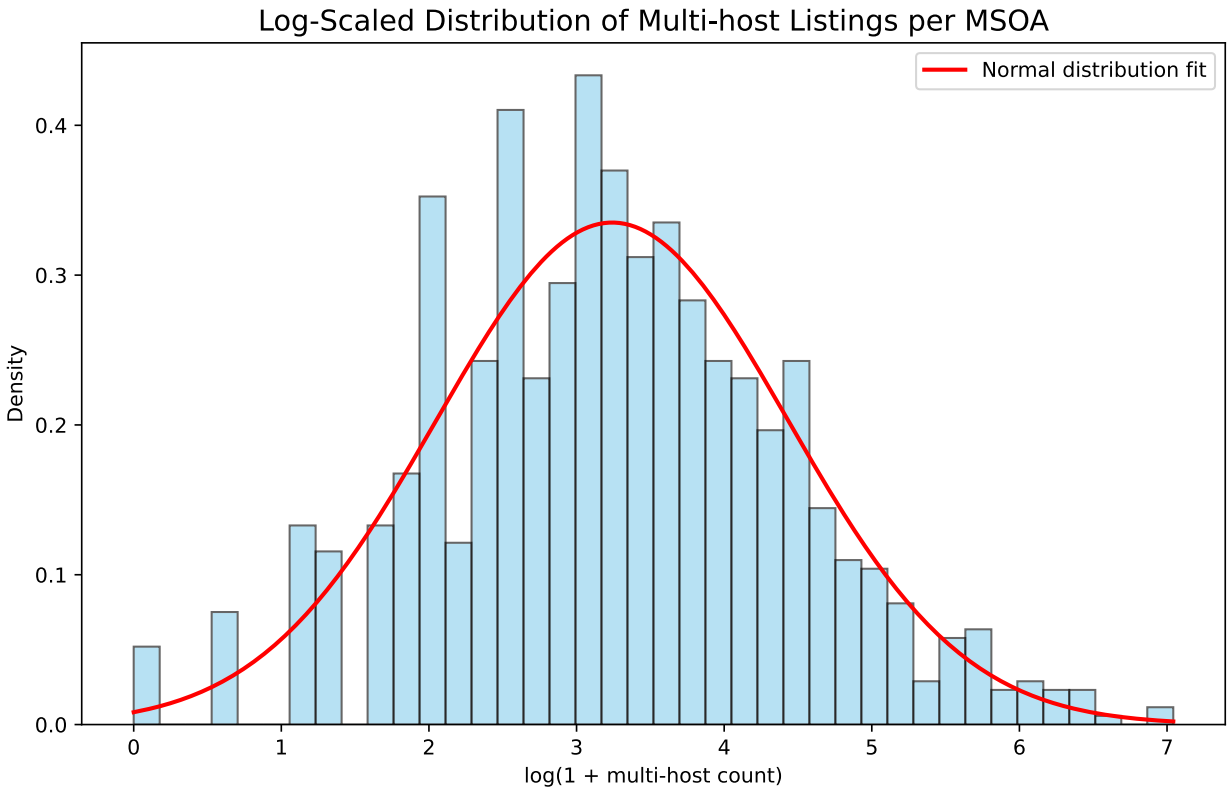


Distribution table:

Just to test whetehr this is a bad skew (although I would say the diagram is pretty obvious…), I will do a Shapiro-Wilk p-value test:

```
Shapiro-Wilk p-value: 2.327478492219949e-46
```

Because from this bar chart of the normal distribution, which is very extremely positively skewed, I want to try creating a log-scaled histogram to see if that changes anything.

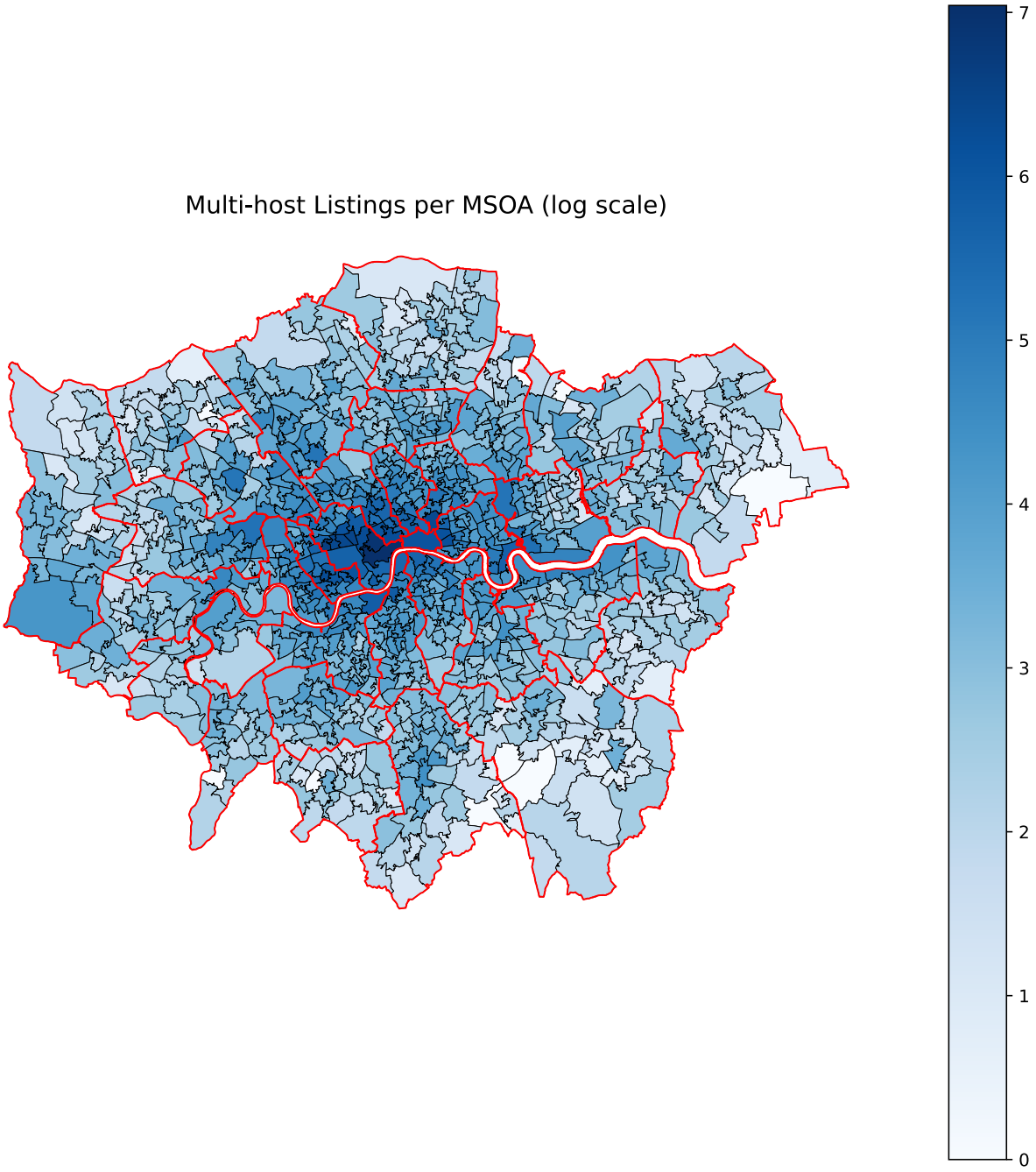


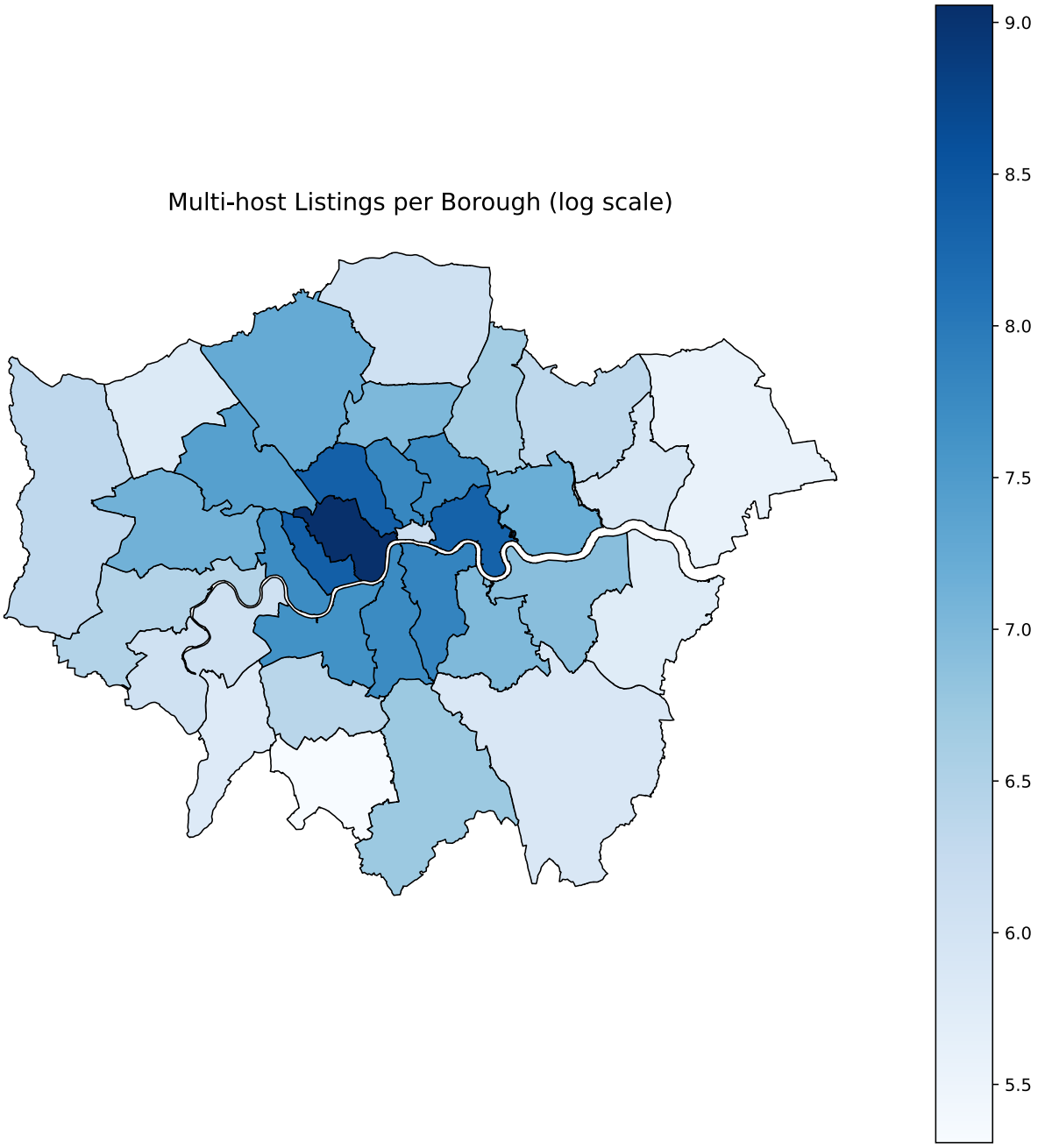
And again let's test the Shapiro-Wilk p-value:

```
Shapiro-Wilk p-value (log counts): 0.0300
```

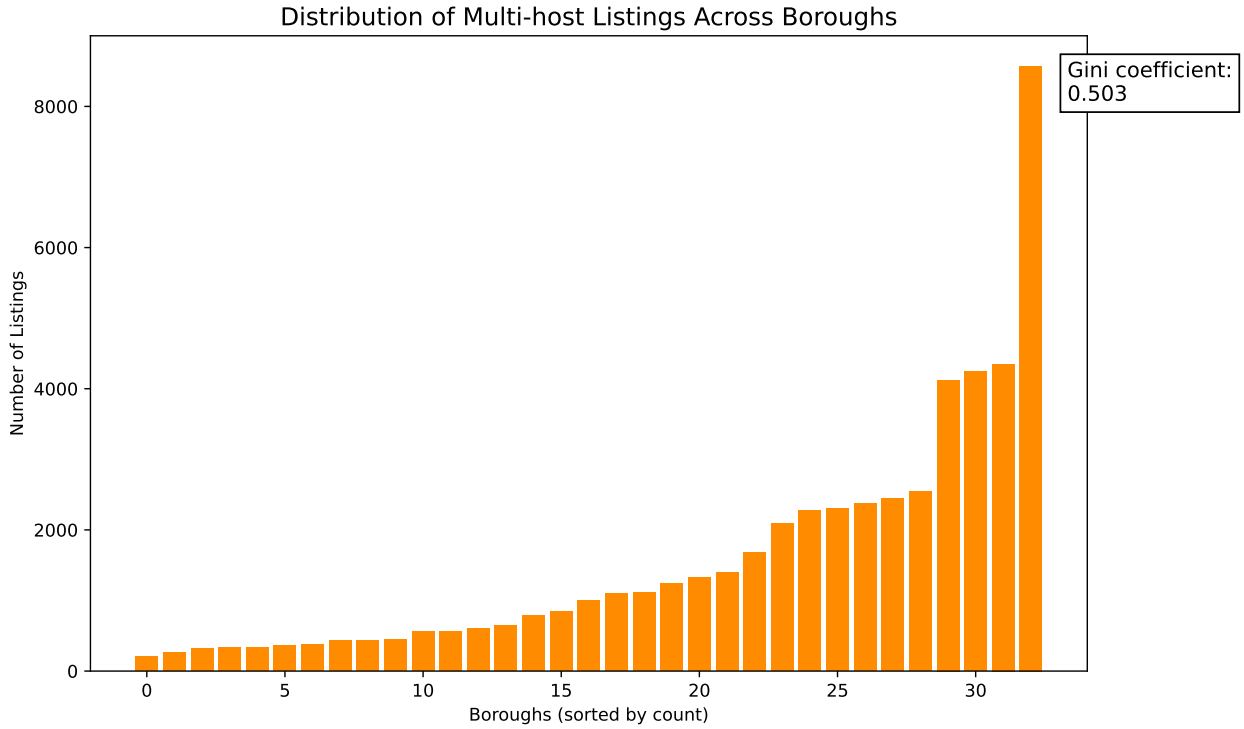
The Shapiro–Wilk test on the raw MSOA counts returned a p-value of 2.3×10^{-46}, providing overwhelming evidence against normality. This confirms what the histogram suggested: the distribution of multi-host listings is extremely skewed. After applying a log transformation, the p-value increased to 0.0300. Although this is still below the conventional 0.05 threshold (meaning the log-transformed data cannot be considered perfectly normal), the improvement is substantial. The log scale reduces skewness and brings the distribution closer to normality, making it more suitable for visualization and interpretation, especially when comparing areas with smaller counts.

To better visualise the distribution, the following map applies a log scale transformation. This adjustment reduces the dominance of areas with exceptionally high counts (such as Westminster), allowing spatial patterns in other boroughs to remain visible and interpretable.

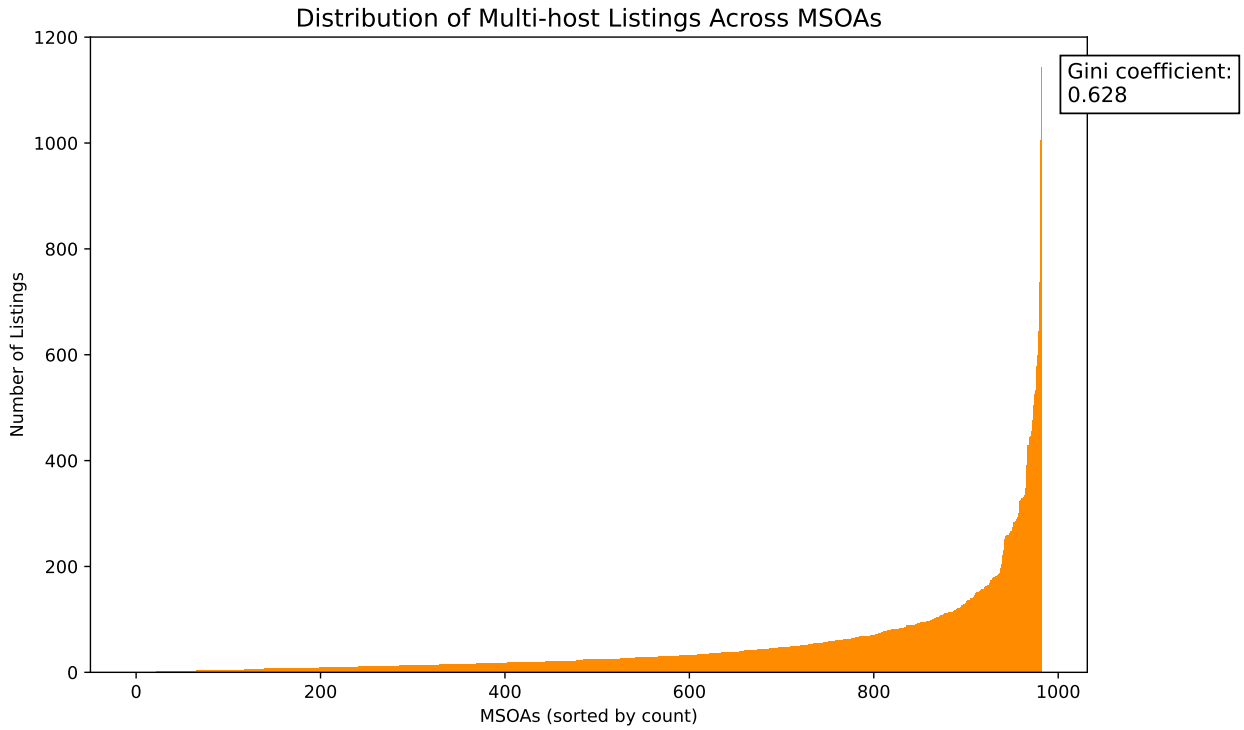

Multi-host Listings per MSOA (log scale)
7
6
5
4
3
2
1
0



Gini I will now do a quick Gini test, which …



The Gini coefficient of 0.503 at the borough scale indicates a moderate degree of inequality in the distribution of multi-host listings across London. Some boroughs host substantially more multi-host activity than others, but the overall pattern is less extreme than at finer spatial resolutions.



At the MSOA scale, the Gini coefficient rises to 0.627, revealing a much higher degree of spatial inequality. This shows that inequality is not only present between boroughs, but also concentrated within them: a relatively small number of MSOAs account for a disproportionately large share of multi-host listings, while the majority contain very few or none. In other words, hotspots of multi-host activity emerge at the neighbourhood level, highlighting that regulation or policy responses cannot rely solely on borough-wide averages but must account for highly localized concentrations.