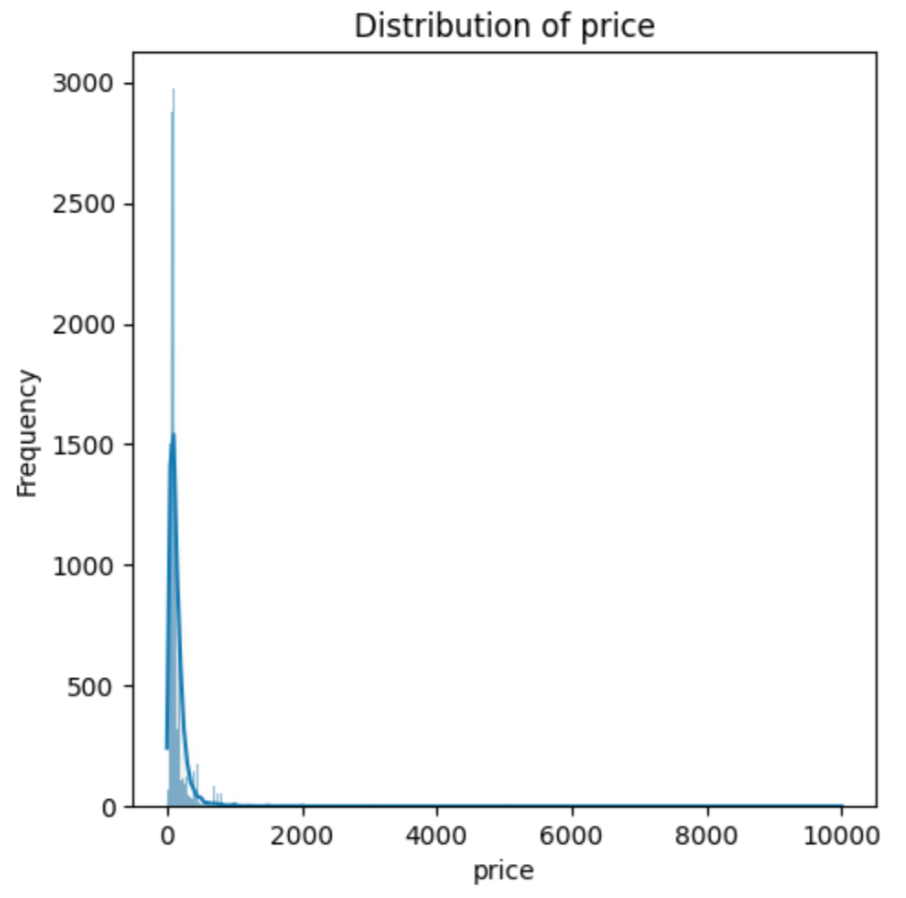
New York City (NYC) has recently introduced new short-term rental regulations, prohibiting rentals of entire housing units for less than 30 days. The new regulations have been established to address NYC’s reduced housing stock and its plummeting affordability (Zaveri, 2023). This report comprehensively analyses the Airbnb New York City 2019 dataset. It aims to identify whether certain NYC neighbourhood groups are more adversely affected by the presence of Airbnb and what impact the new regulations may have on Airbnb’s profitability.

The dataset consists of 48,895 entries across sixteen columns, of which our report primarily focuses on five: room type, price, availability, and location-related variables. The dataset is mostly complete, with missing values commonly appearing only in the 'last\_review' and 'reviews\_per\_month' columns, which will not be used in the research.

Initial exploration highlighted a particularly high variation in the prices of listings, with an average of $153 per night but prices going up to as much as $10,000. As may be expected, our average suggests a higher supply and perhaps customer preference for mid-range accommodation in NYC. Only 0.04% of listings cater to the most opulent customers with prices per night above $1,000. This is confirmed by the right-skewness of the data (Figure 1).

Despite what could be considered highly-priced outliers, no values were removed from the feature. Osborne (2004), highlights that when the data is suspected to be legitimate, it is better to include seemingly ‘extreme’ data points in the analysis as it makes the data more likely to be representative of the true population.

  
Figure 1: Distribution of data of the ‘price’ feature

Following the trace of luxury properties, we explored the distribution of room types across neighbourhood groups (Figure 2). This showed that Manhattan and Brooklyn hold, by far, the highest numbers of entire apartment listings.

Further analysis shows that these types of units are also the highest priced in the two neighbourhoods (Figure 3). Both areas being central for tourists and lucrative for investors may reduce rental affordability for locals to the area.

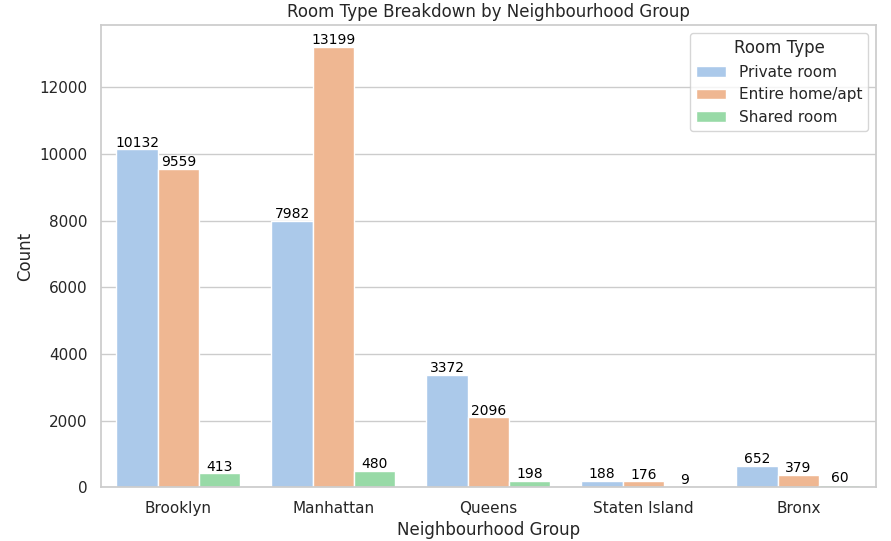


Figure 2: Room type distribution by neighbourhood group

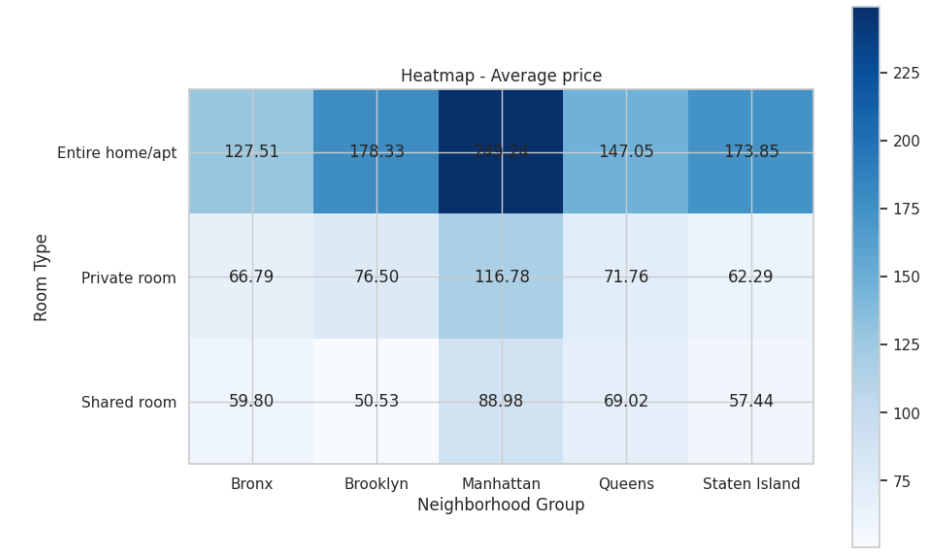


Figure 3: Heatmap of average price by Room Type and Neighborhood Group

Another interesting feature is ‘availability\_365’. Overall, it shows a fairly even distribution with two clear spikes: listings available for less than 10 days per year and those available for over 360 days (Figure 4).

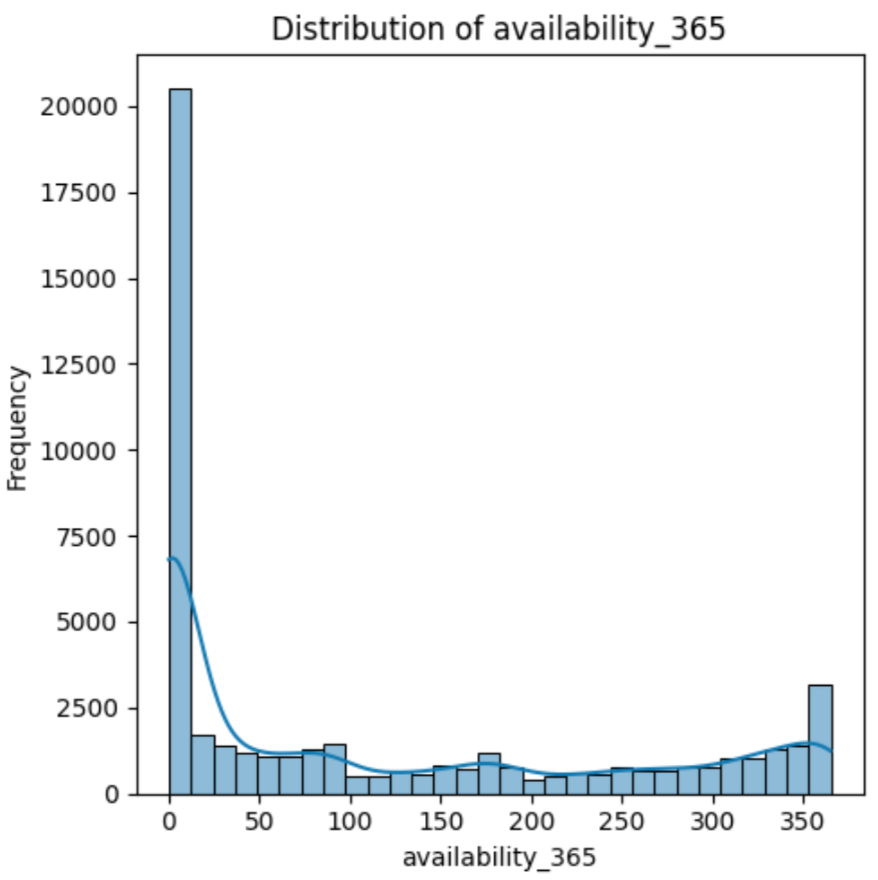


Figure 4: Distribution of data of ‘availiability\_365’ column

46% of total listings are available for less than 30 days per year. Solely based on availability, this would make almost half of Airbnb’s NYC listings not rentable against the new regulations, though the truth may be even harsher. Our analysis shows that actual listing violations are most common in the two highest-priced areas: Manhattan and Brooklyn (Figure 5). Since Airbnb earns its revenue by charging a percentage fee of the booking price, the removal of a large amount of its highest-priced listings could have a long-lasting impact on the company.

K-means clustering was applied to further explore and confirm the above findings. The 'latitude', 'longitude', and 'price' features were used to identify whether clusters of particularly high-priced short-term rentals exist across NYC.

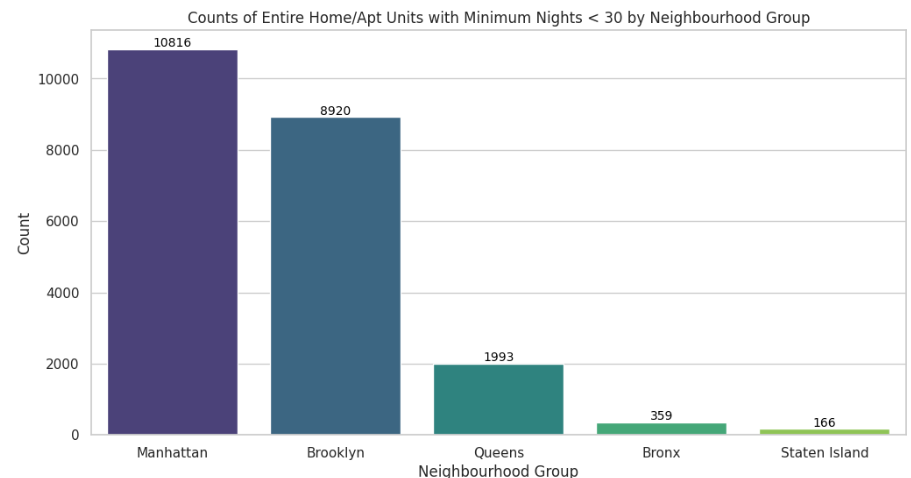


Figure 5: Distribution of units that are violating the new regulations

Clustering algorithms are sensitive to the scale of the data. Since 'price' is on a different scale compared to 'latitude' and 'longitude', we standardised the features, ensuring each feature contributes equally to the analysis (Banks et al., 2011). The Elbow method was applied to determine the optimal number of clusters (Figure 6).

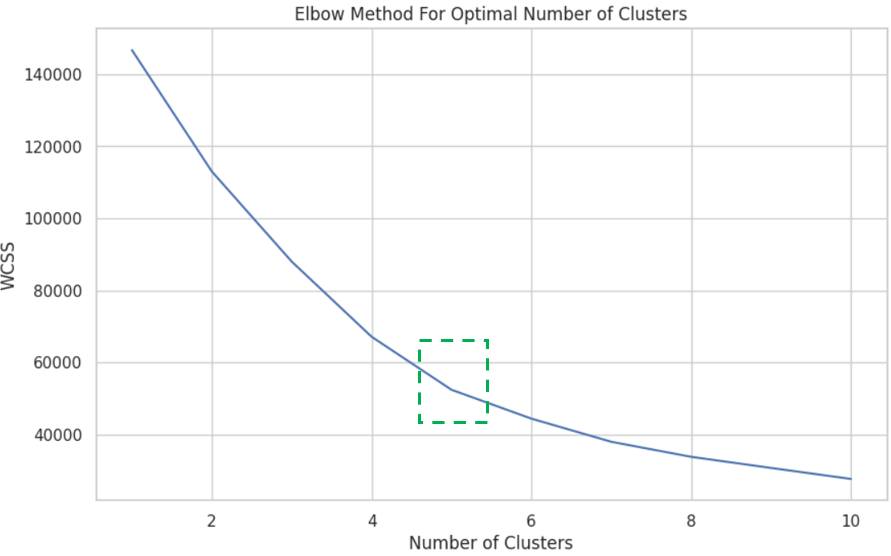


Figure 6: Elbow curve, demonstrating the “elbow”, where k equals five.

Certain clusters, particularly those with higher average prices, confirmed our initial findings as they were located in more central or sought-after areas. Clusters with lower average prices were found in more peripheral areas, with higher ratios of shared accommodation options (Figure 2), such as Queens.

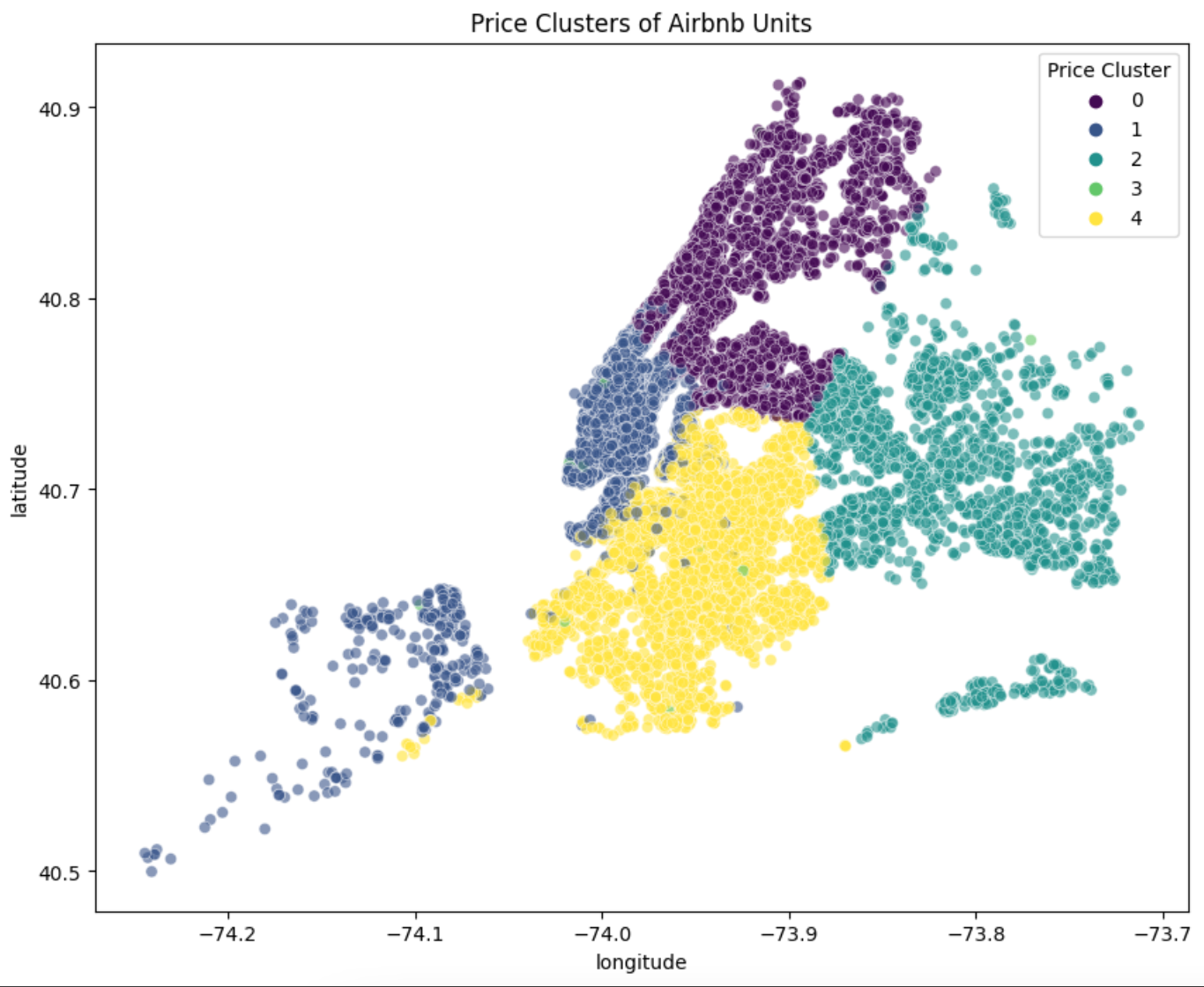


Figure 7: Price Clusters of Airbnb units

The clustering has effectively segmented the market into different categories, ranging from luxury or high-priced listings to more budget-friendly options. Below is a brief analysis of each of the identified clusters (Appendix B).

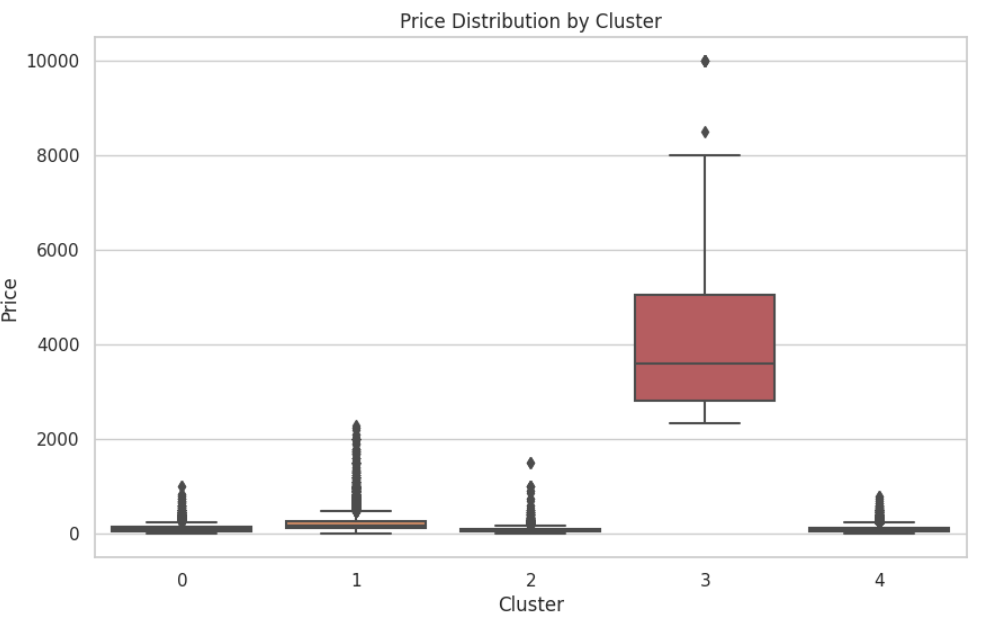


Figure 8: Price distribution by cluster

**Cluster 0**

Moderately priced and quite densely populated area of listings across the Bronx, Upper Manhattan, and Queens. A good option for travellers on a budget seeking to stay central.

**Cluster 1**

This area has a moderate average price that is slightly higher than the cluster 0’s. It also has a few more listings making it possibly less affordable for long-term renting than the previous cluster. It spans the neighbourhood groups of Manhattan and Staten Island and exhibits the highest standard deviation, most likely reflecting the difference in centrality and, thus, desirability between the two neighbourhood groups.

**Cluster 2**

This cluster represents the lowest-priced area and largely covers the Queens neighbourhood. This cluster does not include too many listings, less than 18% of the previous cluster. Based on the price and number of listings, it is likely to be the most affordable for long-term rentals.

**Cluster 3**

Very highly priced units averaging $4,435 per night. This is the most luxurious cluster with probably the least affordable housing for locals. It doesn’t cover a particular neighbourhood but listings belonging to it can most commonly be found within the Manhattan neighbourhood.

**Cluster 4**

This cluster has a mid-price point and the highest number of listings, it covers the neighbourhood of Brooklyn. Locals may find it difficult to find affordable long-term rentals here judging by the number of units listed dedicated to short-term renting.

In conclusion, the New York City government’s regulatory changes targeting short-term rentals would have a material impact on AirBnB’s business in the city. A total of 22,254 units are likely to violate this regulation; this is 45.5% of the total inventory. The loss of revenue due to the regulatory changes may be as high as 31 million dollars per year. Manhattan and Brooklyn appear to be the worst affected neighbourhood groups; these two neighbourhood groups account for 89% of all potential violations. Manhattan is also the neighbourhood group with the highest rental prices, with prices reaching as high as $10,000 per night. Based on our analysis, we recommend that AirBnB management take proactive actions to support the hosts to remain compliant with the regulatory changes to reduce the adverse effect of these changes on AirBnb's business; it is imperative to target Manhattan and Brooklyn hosts — considering the high number of units that are at risk of violating the new regulations.

**References**

Banks, D., House, L., McMorris, F.R., Arabie, P. and Gaul, W.A. (2011). *Classification, Clustering, and Data Mining Applications*. Springer Science & Business Media.

Coles, P.A., Egesdal, M., Ellen, I.G., Li, X. and Sundararajan, A. (2017). *Airbnb Usage Across New York City Neighborhoods: Geographic Patterns and Regulatory Implications*. [online] papers.ssrn.com. Available at: https://ssrn.com/abstract=3048397 [Accessed 15 Dec. 2023].

E. Adams, A. (2023). *The New York City Council - Committee on Housing and Buildings*. [online] legistar.council.nyc.gov. Available at: http://legistar.council.nyc.gov/DepartmentDetail.aspx?ID=6910&GUID=AAAF96D6-8CCD-46E5-9DDB-CFF23FE775DB&Search= [Accessed 15 Dec. 2023].

Mahmoud Parsian (2015). *Data Algorithms: Recipes for Scaling Up with Hadoop and Spark*. Boston, Ma: O’reilly Media, Inc.

Osborne, J.W. (2004). *The Power of Outliers (and Why Researchers Should Always Check for Them)*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/242073851\_The\_Power\_of\_Outliers\_and\_Why\_Researchers\_Should\_Always\_Check\_for\_Them [Accessed 30 Nov. 2023].

Stoddard, A.M. (1979). Standardization of Measures Prior to Cluster Analysis. *Biometrics*, 35(4), p.765. doi:https://doi.org/10.2307/2530108.

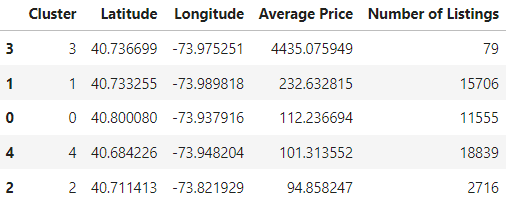
Suresh Kumar Mukhiya and Ahmed, U. (2020). *Hands-on exploratory data analysis with Python: perform EDA techniques to understand, summarize, and investigate your data smartly*. Birmingham: Packt Publishing.

Zaveri, M. (2023). New York City’s Crackdown on Airbnb Is Starting. Here’s What to Expect. *The New York Times*. [online] 5 Sep. Available at: https://www.nytimes.com/2023/09/05/nyregion/airbnb-regulations-nyc-housing.html [Accessed 15 Dec. 2023].

**Appendix**

**Appendix A**: Python code used to perform analysis in this report  
<https://colab.research.google.com/drive/1Y8qcIFEYsS9eZYJF41fNBQNFGClmqO1T?usp=sharing>

**Appendix B:** Summary of cluster statistics

****