1. Who collected the data? *2 points; 25 words; Consider the source of the data and its relation to the underlying data generating process.*

Data was collected by Geocomputation and Spatial Analysis within King’s College London. Forked it from (https://github.com/kingsgeocomp/geocomputation) as indicated from his GitHub account as shown (~jreades/data/2022-09-10-listings.csv.gz).

1. Why did they collect it? *4 points; 50 words; Consider the purposes for which the data was collected and how this might shape its structure or content.*

Data was collected to enable students access the spatial data which is hard to get.

To collected to enable students have better understanding of the spatial data.

To enable students, utilize, visualize and communicate data in geographical context.

To enable students, develop logic, programming and analytical skills tackle complex real-world problems creatively.

1. How was it collected? *7 points; 75 words; What was the method by which the data was collected and how might this shape its structure, content, or completeness.*

Since data was collected by combination of dearth of planners and geographers in within King’s College London, the use of global position system receivers was used as this is the most commonly tool used in spatial data collection in specific. This is because spatial data has to have the latitude and longitude features which indicates where the other features are located at. However, data might not be complete because some values of other features might not be captured.

1. What useful information does it contain? *12 points; 100 words; Discuss how the data might support a range of analyses and note any limitations encountered so far in class or in your own investigations.*

The spatial data collected contains information about geographical context. For example, we have features like accommodations, property type, bathrooms, host total listings , and many more which shares location and longitude. Using this kind of information, one can analyze this data to find best/most used bathrooms and in which latitude and longitude is the bathroom located. One can also wish to know how many accommodations do we have in a given latitude and longitude. As well the different property types in different latitudes and longitudes. The challenge so far is that, the dataset has got many columns needs more attention to have in mind.

1. To what extent is the data ‘complete’? *25 points; 200 words; Reflecting on your earlier answers, and drawing on what you’ve learned about the data so far in class, to what extent is this data a ‘complete’ picture of Airbnb’s operations in London?*

Based on the rows and columns Nan values count , I can say that this spatial dataset is at 90 percent complete. However, this doesn’t mean that we can work with this data the way it is. We must apply data cleaning techniques like working on the missing values (nan) in the dataset entries and columns. The 90 percent is achieved because it contains fewer missing values, the structure is fine to work with and the columns which represents data of specific are in required data format.

1. What kinds of analysis would this support? *15 points; 200 words; Given the issues identified above, what****kinds****of analysis would this data support? You do not need to propose a specific analysis and should instead focus on generic****classes****of analysis.*

Following the issues outlined above, this data could support descriptive analysis. Descriptive analysis will work fine since the data variables in the whole spatial dataset are numerical in data type. Using this analysis, the analysist will be in a position to find the maximum value in each column, the minimum value in each column the, count number of values in the column as well as find the mean value. Using this information especially mean, maximum and minimum values, outliers in the dataset can be easily be notified since data distribution will be clear to the analysts (Reddy, 2018 ).

The spatial dataset could also allow the analysts perform the predictive analysts. The analyst is in a position to making prediction as follows; analyst could want to know, the availability of an accommodation in the next week using the given data. The analyst may also want to know the most liked(used) amenities in a specific latitude and longitude in the coming days. The end of these two analytics the analyst will be a position of first understanding the whole dataset information that is how it is. The analysts will also gain insights using the data on what he wishes to know in future when he applies predictive analytics (Cressie, 2015).

1. Which of the analyses outlined above are ethical? *35 points; 350 words; Discuss the ethics of these classes of analysis with reference to your earlier answers and to the assigned readings.*

The ethics of applying descriptive analytics is that, this class of analytics does not allow the analyst to move with the data which is not in normal distribution state. This kind of analysts as well helps in finding outliers in each dataset column. This creates a room for the analysts to easily see and be in a position to correct before working on the dataset as it was before (Lloyd, 2010). In simple terms, it ensures that the analysts, first has got the information about how the dataset is distributed, if there are any outlier values in the dataset in specific columns, the number of numerical columns found in the dataset. The analysts are feed with this dataset information thus, moving on with a clear indicator on what should be done and where in order to have accuracy in the later analysis and modelling techniques.

On the other hand, the ethics of predictive analytics is that, the accuracy is left feed with what is expected to happen in the coming. This kind of analytics gives the analyst the power to power forecasting on the available data in order to gain insights of the same data in the later times ahead (Anselin, Syabri, & & Kho, 2010). Thus, in our case, the analysts might choose to know which amenities are more likely to be less priced in the coming days due to the changing environment and features around, or the analysts might want to know which amenity will receive more customers in the coming days by analyzing the reviews made per month.

# References

Anselin, L., Syabri, I., & & Kho, Y. (2010). GeoDa: an introduction to spatial data analysis. In *In Handbook of applied spatial analysis* (pp. 73-89). Springer, Berlin, Heidelberg.

Cressie, N. (2015). *Statistics for spatial data. .* John Wiley & Sons.

Lloyd, C. (2010). Spatial data analysis: an introduction for GIS users. . Oxford university press.

Reddy, G. P. (2018 ). *Spatial data management, analysis, and modeling in GIS: principles and applications. In Geospatial Technologies in Land Resources Mapping, Monitoring and Management.* Springer, Cham.