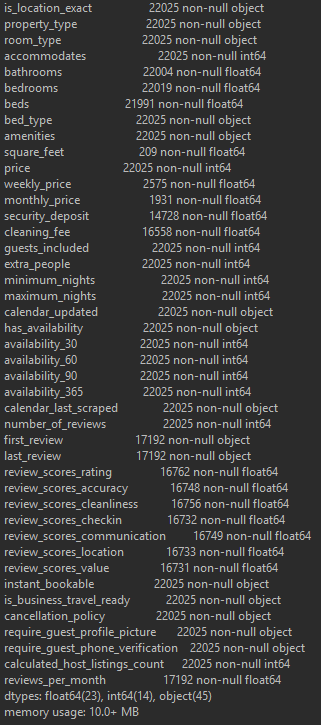
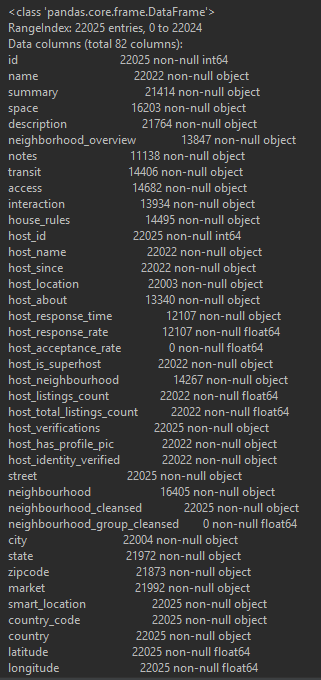
The data was obtained from Inside Airbnb to build prediction models about its prices in Melbourne. Firstly, the data was downloaded in a csv format and loaded into the Jupyter Notebook.

The downloaded data has the following lists of attributes in the data:-

1. As part of this data wrangling exercise, I have to identify the columns that I would like to retain as some columns are not necessarily useful for modelling. In this part, I’ve removed:

* is\_business\_travel\_ready
* country
* country\_code
* neighbourhood\_group\_cleansed
* host\_verifications
* host\_neighbourhood
* host\_acceptance\_rate
* market

1. I’ve also removed any rows that did not have alphabets or numbers in them. A couple row of data contained different languages that couldn’t be converted back to into English. Hence, these rows have been dropped.
2. First review, last review, host since and calendar last scrape date are all recorded as objects and I’ve changed them to datetime
3. One of the messiest column is ‘City’ as some categories are not labelled correctly. I’ve replaced their correct names
4. From data.info() we know that we have null objects in them so I’ve filled them with NaN instead as of removing them or leaving them blank
5. Saving the clean file into a new csv