

- delete columns with more than 50% null data
- delete columns without useful information
- data cleaning
- fillna numeric data by mean
- normalizing numeric data
- delete outlier
- Which hosts are the busiest : Michael 362 , David 356
- delete column that low importance
- Location of rental houses in different cities with scatterplot
- Is there any noticeable difference of traffic among different areas and what could be the reason for it? The reason is the number of hosts
- Build a column called all_year_availability with scatterplot
- Number of active houses on all days of the year
- relation between neighbourhood_group & price
- relation between neighbourhood_group_edit & price
- check distribution of price
- transfer price (to ln price) to make the disk normal
- correlation matrix
- split future and target