**CSYE 7245 - Big-Data Systems and Intelligence Analytics**

**MakeMyTrip**

<https://www.kaggle.com/PromptCloudHQ/hotels-on-makemytrip/>

**Abstract**

This is a pre-crawled dataset, taken as subset of a bigger dataset (more than 615,000 hotels) that was created by extracting data from MakeMyTrip.com, a travel portal in India.

**Introduction**

In this project using the data collected by Kaggle for MakeMyTrip, we are going to determine the hotel rating using

* Facilities provided by each hotel.
* Type of rooms available in each hotel.

Higher the count of the above two column, higher is the rating.

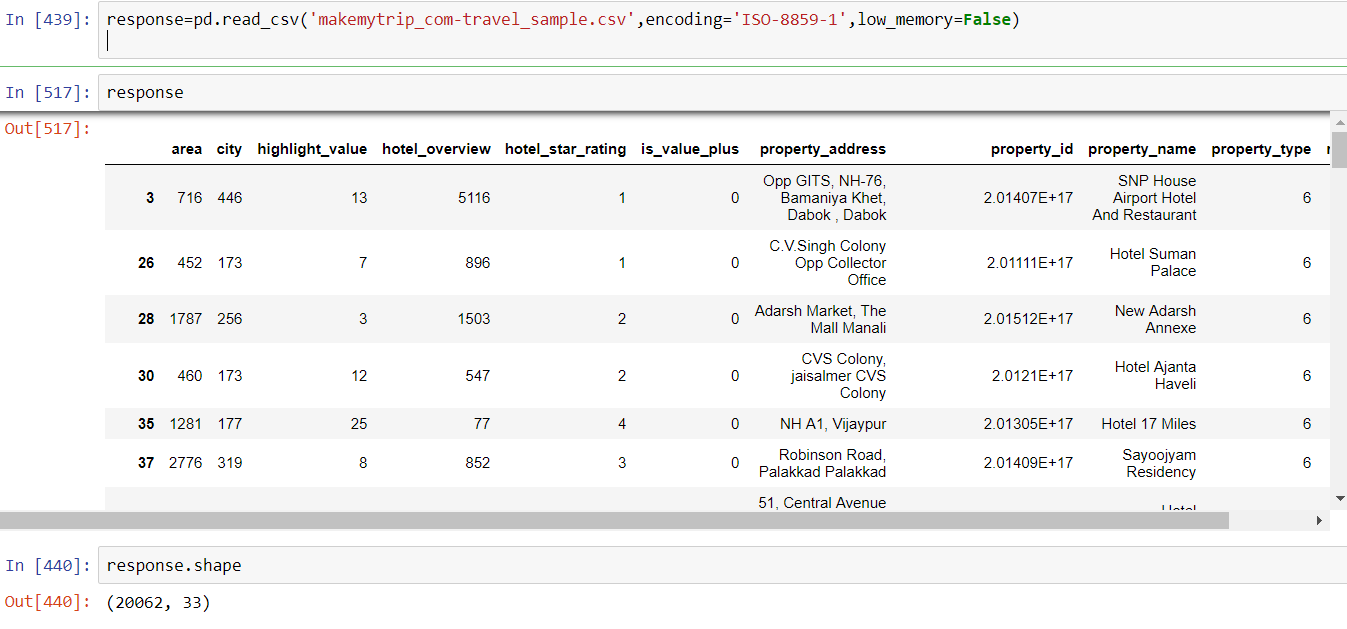
So for example 1 hotel is providing facilities like [ Airport Transfer, Car rental, Conference Hall, Currency Exchange, Doctor on Call, Dry Cleaning, Guide Service, Internet access, Jungle Safari, Laundry Service, Restaurant, Room Service, Wi-Fi Internet] which totals up to 13. Similarly, we will count types of rooms.

There is a more accurate way of doing it, we can split each value of the columns and then plot a ‘Yes’ and ‘No’ table for each row. This method would have given a better prediction but it requires a lot of computation power to execute it.

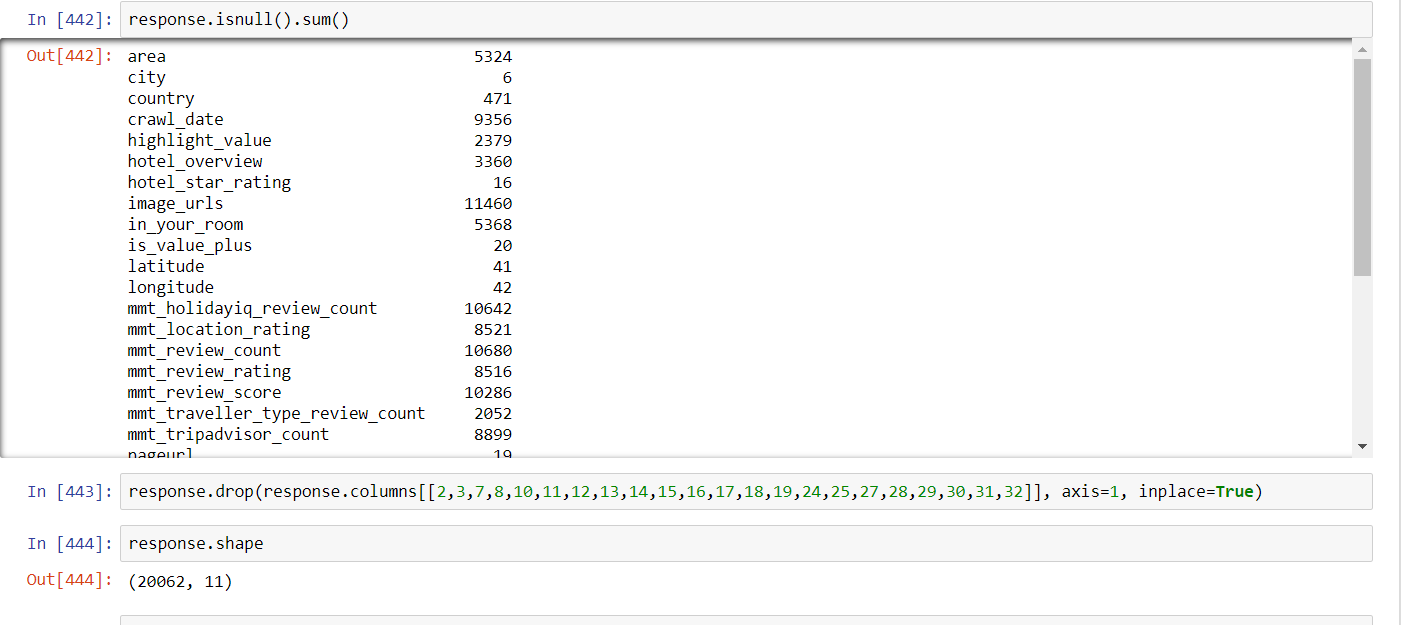
**Code with Documentation**

We need to import the data set in notebook and then start removing the unnecessary data and columns.

Below we are importing dataset and cleaning it

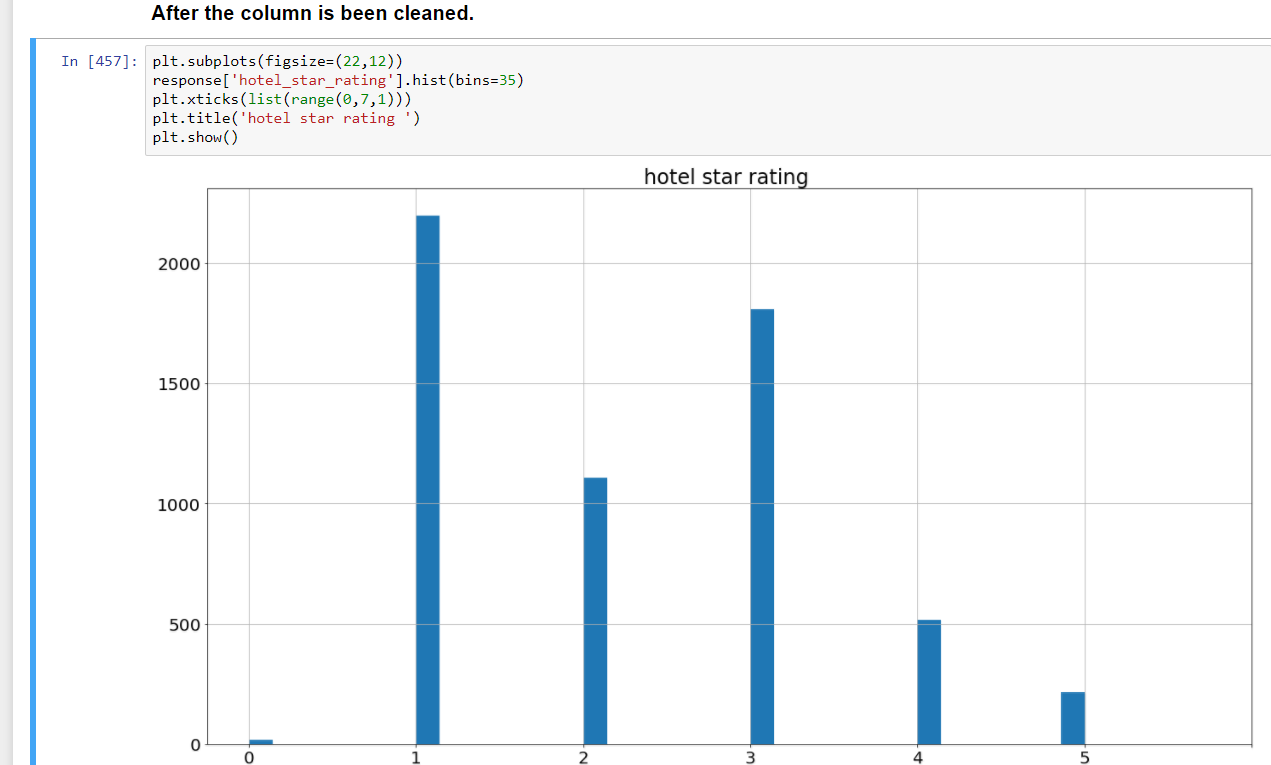
****

Here we check for sum of null values and then delete unrelated rows from our dataset.

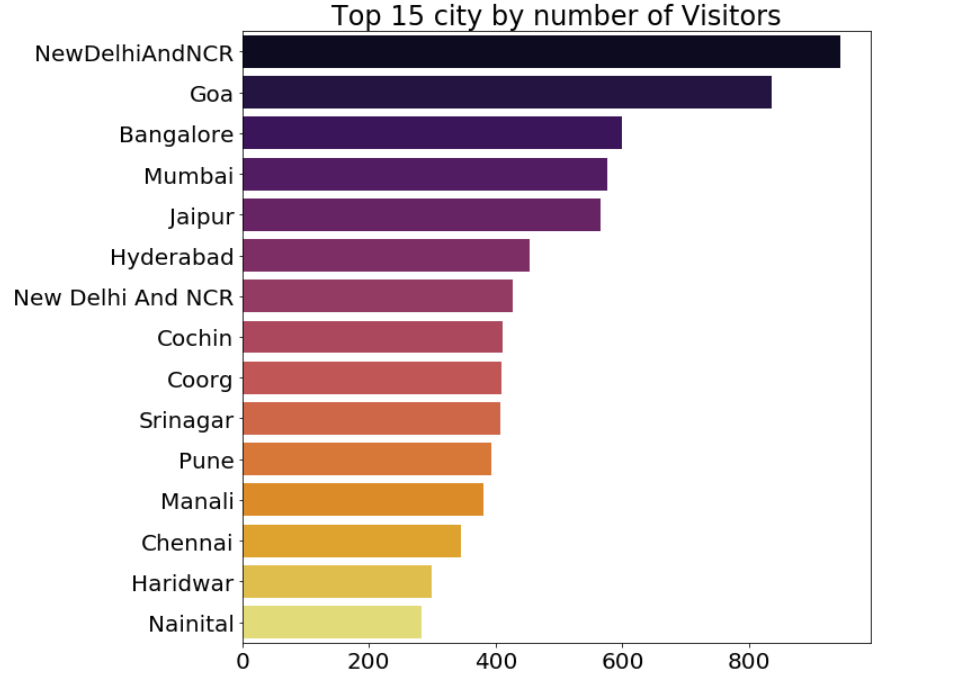


For cleaning a particular column, we need check unique value of that column. For hotel\_star\_rating , some value of row was 1 star, 2 start, 3, 5, 1. So we need to make it follow a standard representation i.e. integer. In the below code we are removing ‘star’ from each cell and the blank spaces.





**Exploring some data that has the highest number of reviews.**



**Results**

**Discussion**

**References**

<http://www.pythonforbeginners.com/loops/for-while-and-nested-loops-in-python>

<https://stackoverflow.com/questions/5278122/checking-if-all-elements-in-a-list-are-unique/5281641>

<https://www.kaggle.com/ash316/novice-to-grandmaster>

<https://stackoverflow.com/questions/28538536/deleting-multiple-columns-based-on-column-names-in-pandas>

<https://www.dataquest.io/blog/jupyter-notebook-tips-tricks-shortcuts/>