Analysis of all rent houses data of Mumbai

# 1. Introduction

## **1.1 Description and Discussion of the Background**

Mumbai is a densely populated city on India’s west coast. A financial canter. It’s India's largest city with over 603.4km area and a population of 1.84 crores. It serves as an economic hub of India, contributing 10% of factory employment, 25% of industrial output.

As I mentioned earlier that Mumbai serves as an economical hub of India, lots of peoples move to Mumbai for doing a job, finding a job, education, labour work, etc. as we know Mumbai is a densely populated city so, it's hard to find the house on rent(in the budget of person) in this type of city.

So, to solve this problem I will be going to analyze all the rents house data of Mumbai and find some reliable house for users.

## **1.2 Description of the data**

After searching a couple of data provider sites, I have found perfect data set to solve this problem on [**kaggle.com**](https://www.kaggle.com/), I’m going to use [**“Flats for Rent in Mumbai”**](https://www.kaggle.com/jedipro/flats-for-rent-in-mumbai/data) data set from kaggle**.**This data set contains data from 12/08/2019 to 14/01/2020 total of 34349 rows.

This dataset contains major key columns like price, latitude, longitude, Locality, City, Description of the house

Followings are a list of columns with small description of each.

1. area : Floor area of the property.

2. bathroom\_num : Number of bathrooms available.

3. bedroom\_num : Number of bedrooms available.

4. city : City in which property is located.

5. desc : Text description of the property.

6. dev\_name : Name of property developer.

7. floor\_count : Total number of floors in building.

8. floor\_num : Floor on which property is located.

9. furnishing : Furnishing status.

10. id : Unique ID.

11. id\_string : Unique ID string used to scrape a particular HTML page element.

12. latitude : Latitude of location.

13. locality : Locality in which property is located.

14. longitude : Longitude of location.

15. post\_date : Date on which property was listed on website.

16. poster\_name : Name of poster.

17. price : Price of the property.

18. project : Name of the residential complex.

19. title : Title of the property ad on the website.

20. trans : Type of property transaction.

21. type : Type of residential complex.

22. URL : URL of the individual property.

23. user\_type : Type of user who posted the ad.

This data set contains 23 columns but, we will filter the data set and remove the un-relevant column and keep only those column which will useful to us. We will also be going to apply data cleaning on this dataset.