Dataset: Airbnb listing Dataset Washinton DC

Problem:
We want to find the reasonable nightly rent so that we don't set the rent too high and lose customers to cheaper options or set the rent too low to not earn any profit.

What we are going to do:

Use KNN to check for similar listings and set our price equal to the average of the price of the listings similar to us.

Shape of Dataset :

3723,92

Steps we are going to follow:

1. Taking each entry in the test set.
2. Finding its euclidean distance from each entry in the training set.
3. Appending the calculated distance to a new column 'distance' in the training set.
4. Randomly shuffling the resulting set.
5. Sorting the set in ascending order of distance.
6. Choosing the first 5 entries(let K=5) i.e. the five nearest neighbors.
7. Calculating the mean of their 'price' column which is the predicted price.
8. Appending the resulting predicted price to a new column 'predicted_price' in the test set.

Functions we created:

Euclidean_distance: to calculate the euclidean distance btw two rows.

Predicted_price: To calculate the predicted price for a particular test row.