# AirBnB Price Prediction Challenge

### Suggest the price of AirBnB listings in major U.S. cities

## Introduction

The aim of this competition is to predict the price of AirBnB listings in major U.S. cities.

## Problem Statement

Given a dataset with 24 variables such as number of bedrooms and a log-price indicator (greater than 0) for each observation in the training data, the objective is to suggest the log-price of a particular listing using the 24 features provided for the test observations. The suggested prices that are closest to the true prices, as calculated by the Root Mean Squared Error, scores highest on the leaderboard.

## Data

The training data for the competition consist of the following columns:

1. id : Record identifier
2. log\_price : log(price)
3. …. : 24 features for the listing

There are three files provided for this competition:

1. train.csv : Training data
2. reviews.csv : Reviews for all of the training and test listings
3. test.csv : Test data
4. sample\_submission.csv : Format for submission of the predictions on the test dataset

## Submission

The submission file for the competition should consist of two columns:

1. id : Record identifier for the test observation
2. log\_price : Prediction score for the record belonging to class ‘1’. Type: Float (between 0 to 1)

Submissions not meeting the requirement stated above will be disqualified. The training, test data and sample submission files can be downloaded here.

## Evaluation

The evaluation metric for the competition is the Root Mean Squared Error (RMSE) score.

During the competition, a public leaderboard will be updated based on the submissions and will be visible to all participants. The public leaderboard will be calculated based on 50% of the test data. The final results will be based on the other 50% of the data, so the final standings may be different.