Rent Price Prediction - Steps that are followed

1. Data pre-processing –
   1. Loading the test and train data
   2. Validating the shape of data (rows, columns)
   3. Inspecting high-level stats of train data
   4. Checking for Null values and imputing them
2. Exploratory Data Analysis –
   1. Checking the distribution of rent amount
   2. Univariate analysis for categorical variables like country, property\_type and city
   3. Plotting correlation Map and finding the most correlated variables (60% of correlation is observed between few variables)
   4. Plotting Mean price for different countys(Adams, Allegheny, Armstrong, Beaver are the top most counties as per rent)
   5. Plotting average rent for different property\_types (Quadplex and Triplex are the top most property types )
   6. Creating yearwise bins from 1800-2016 (10-year bins). 1820-1830 interval has largest average rent.
   7. Plotting top-20 cities by rent (Millerton is the expensive city in terms of rent)
3. Feature engineering–
   1. Removing un-necessary columns (Zipcode, address, state, id)
   2. Converting categorical data to numerical data (using pd.getdummies()). The total number of columns for train data yielded to be 836 columns and test data – 616 columns
4. Data Modelling –
   1. Splitting the train data and test data with 80:20 ratio
   2. Trying different models for regression (Random forest, Linear regression and Decision trees).
   3. Evaluation of best model among 3 models (Random forest is the best model which yielded best R^2 value and RMSE)
   4. Performing hyper parameter tuning using RF Regressor.
   5. There is no much difference between RF regressor using default parameters and tuned parameters.
   6. The default R^2 = 0.82 and RMSE = 386 for RF Regressor
   7. Reshaping the test dataset to make sure that train dataset and test data have equal dimensions (836 columns)
   8. Plotting top most important features and their feature important scores. The most dominant features are
      * 1. Bath
        2. Bed
5. The above two features explain 61% of variance in the data

j) making the final predictions using test data and saving as csv file.