Questions

1. Please describe the central limit theorem and provide an example.

Given a series of independent and identically distributed random variables with sample mean and sample variance, bound to under infinity, the central limit theorem states that the distribution of the sample means will be approximately normally distributed around the true population mean given that Xi is sufficiently large.

1. Describe a classification algorithm that you have previously put into production and why it was chosen.

1. Describe the difference between bagging and boosting methods, and when to use one or the other.

1. Describe 2 regularization techniques for a random forest model

Data Challenge

Attached you will find the file listings.csv which provides data on around 50,000 AirBnB listings in New York City.  Given only this data, you want to create a model to predict how much you can charge for new listings while keeping vacancy down.

Please put together a brief analysis of the dataset and show how you would go about creating a model to predict a listing price, while taking market demand into account.