Doris Suzuki Esmerio

Midterm Exam

Results: I was still not able to get good results, so I am planning to resubmit….

How the parameters were chosen:

ValueError: could not convert string to float: 'Brooklyn'

raise ValueError("Classification metrics can't handle a mix of {0} "

ValueError: Classification metrics can't handle a mix of multiclass and continuous targets

How the data was split/subset:

Limitations of the model:

Latitude and longitude could have been helpful to calculate the distance from important locations, such as most famous tourist places, convention centers, train stations. This way, another variable “distance from reference point” could have been generated and would possibly add relevant information. It would require additional work of identifying the coordinates of main locations of interest (could use GIS to assist in that as well).

This could also be used to select a subset of the data, limiting the observations that are within a certain distance from the location of interest.

The description of the locations (second column) could also have been used, to see whether using certain words are correlated with the price of the airbnb.

I choose to use kFold to split the Data. It is a reasonably huge dataset, so we could try using just subsets of the dataset.

Shuffling the Dataset is a way to

Getting a romdomly

So initially, I tried using only neighborhood and room type to predict, as they seemed to be the top more relevant. My results were

To try to improve it, I then tried

\_

How to run the code:

Write on command :

Python midterm.py

\_\_