Instructions

-Complete this exercise in a Jupyter Notebook using either R or Python.

-Include comments every step of the way so we can observe your thought process

-Use an environment such as Virtualenv or Anaconda to isolate your work

-Include the conda and/or pip exported file(s) containing your installed packages in your submission so your work can be replicated

-If you are using both pip and conda to install packages, please make sure you include both files so all the packages you have used can be installed easily on another computer.

The attached CSV contains ~8000 samples of rental listing samples. There are 5 feature columns in the file (listing_type, tenant_rating, bedrooms, bathrooms, monthly_rent).

Some info about the data:
Each sample is recorded when a tenant moves out of a rental. The tenant provides some of this information below (most of the time, sometimes they are forgetful). The other information comes from the owner of the rental.

listing_type: what the owner has told us their type of rental is

tenant_rating: what the most recent tenant has rated that rental at move out

bedrooms: the number of bedrooms the owner has told us exist in the rental

bathrooms: the number of bathrooms the owner has told us exist in the rental

monthly_rent: the monthly rent the tenant has been paying for their contract duration

Task 0 - Get a feel for the data
a) Load the file into your notebook
b) Print the sample size
c) Print the default head and tail of the data
d) Print the distinct listing_type categories

Task 1 - Clean it up
We only care about data that is complete.
a) Remove any samples that are incomplete or missing features
b) Print the number of incomplete samples removed and the new "complete" sample size
c) Make all the characters in the listing_type column lowercase
d) Print the head showing 25 samples

Task 2 - Some numbers
a) Print the average tenant_rating and monthly_rent from

all samples

b) Print the average monthly_rent for listing_type "private Room"

c) Print the average tenant_rating for listing_type "entire home/apt"

d) Print the IQR for monthly_rent from all samples

e) Print the mode for bedrooms and bathrooms

Task 3 - Some Visualization

a) Generate a box and whisker plot for listing_type "private room" where bedrooms are equal to 1. Label the upper extreme, upper quartile, median, lower quartile, and lower extreme (with their values)

b) Are there any outliers? If there is, generate a table displaying some information of your choosing about the outliers.

c) For listing_type "private room" and "entire home/apt", generate a plot with tenant_rating on the y-axis and monthly_rent on the x-axis. Give each plot a name and don't forget to label the axes.

d) Generate a plot for listing_type "entire home/apt" with tenant_rating on the x-axis, bedrooms on the y-axis, and monthly_rent on the z-axis.

Task 4 - Let's separate some data

a) Create three distinct datasets (training, testing, validation) with a 70/20/10 split.

b) Print the total number of samples in each dataset

Task 5 - Let's make a model

Let's say our goal is to help owners try to figure out what their rental property should be listed for (monthly_rent).
a) This part is up to your creativity - design a model using the features of your choosing. Explain your reasoning for choosing that model. Explain what features you used as explanatory variables and if you didn't use features explain why.
b) Train the model using the appropriate dataset (set the verbose to a level where we do not see the progress bar, but we can see the loss & accuracy)
c) Plot the epochs vs accuracy
d) Plot the epochs vs loss

Task 6 - Let's test the model
a) Use the appropriate datasets you made in Task 4 to see how well your model performs. Plot the predicted monthly_rent and actual monthly_rent as a line graph. Make sure to add labels so we know what we are looking at.
b) If it's not performing as well as you had hoped, explain what you could try changing to improve.
c) Repeat part a with the validation set

Task 7 - Some fun stuff
Have any funny jokes? What are your hobbies? Do you like hot sauce? Here is your chance to tell us a little bit about yourself! Maybe something that doesn't fit into a resume but let's us here at REBCA know a bit more about you!

When you are finished the assignment, save your notebook and anything else we might need to replicate your work. Email or share it with   kevinschw@rebca.ca

Good luck!!