You work at a startup that dynamically sets Airbnb pricing for clients that have listings on the site. Your company scrapes data from the Airbnb website and uses the data, plus other data such as local events and time of year, to build models that will predict a reasonable price to set the client's listing.

For this project, you will be creating a baseline model for clients in the Austin, TX region that uses only the data from Airbnb[1]. This data is read in easily[2] with a package such as `pandas`, but the values of the features are not all in a useful format to present to your models. It may be useful to reference this blog post[3] to provide you with an initial direction in your data exploration. **The goal of this project is to predict price for a given property. You must use a neural network to make your prediction.** Note that `scikit-learn` has neural network capabilities, however, `TensorFlow` and `PyTorch` have more flexibility. You are free to use whatever package you want and are not limited to one hidden layer.

Note: I am relaxing the collaboration rules for this project due to the shortened time frame. "Vegas" office hours rules are always in effect. This means that you can talk in any level of detail about the project with others from the class and look at each others code. The only rule is that you can not share (i.e. send each other) your project code.

The Write-Up

Your primary deliverable for this assignment is a Jupyter notebook that promotes your model to potential clients. We will assume that our potential client is an Airbnb owner who is interested in the modeling, but finds it more profitable to outsource the actual work (i.e. we can provide sufficient detail without compromising our company).

You need to be sufficiently precise with your writing and include enough detail that a competent reader could reproduce your results. Here are some specific things to address in your report, in no particular order. This is *not* meant to be an exhaustive list.

- Did you perform any preprocessing on the data? If so, describe these steps (and cite sources if appropriate).
- What architectures did you try for your neural network? How do they compare in terms of performance? What was the best performing model, and how did it do?

[1] This data is provided by http://insideairbnb.com, who scrapes the Airbnb data for us.
[2] I had to add the attribute `low_memory=False` to `read_csv` to see all of the columns.
[3] https://towardsdatascience.com/predicting-airbnb-prices-with-machine-learning-and-deep-learning-f46d44afb8a6

- Has your approach been used for this problem before? If so, how do your numbers compare to the published results? If not, how do your results compare to the most similar/related approaches?
- What was your model-building and tuning regime? How did you address overfitting? How did you make hyperparameter choices?

Recommended Timetable

Here's a recommendation for how to budget your time over the next couple of weeks as you work on this assignment.

- **Oct. 3–6:** Explore the dataset, do feature selection, build your first models.
- **Oct. 7–10:** Run more thorough experiments (hyperparameter tuning, further feature engineering, etc.), analyze your results and iterate, search the literature for related work on the problem, begin writing about your data exploration procedure.
- **Oct. 10–12:** Complete experiments, take a step back and think about your notebook's narrative, write about your experiments and models.
- **Oct. 13–15:** Wrap-up any pending experiments, document code, conclude your writing, revise and proof-read the notebook and submit your final version.