DA3
Assignment 2
Airbnb prediction models

35 points
Individual or in Pairs

Tasks 1

- Business: Your task will be to help a company operating small and mid-size **apartments** hosting **2-6 guests**. The company is set to price their new apartments not on the market.
- Pick a city you will operate in.
- Build a price prediction model similarly to how we did in our case study for London.
- Discuss your modeling decisions and compare your results to those of the case study.

Getting data

- Visit the website we got the Airbnb dataset from: http://insideairbnb.com/get-the-data.html and download data from **another city**
- *Any city with at least 10k rows (not London) any date*
- *Everybody will have a different city - **Please sign up for a city**.*
- May use any date

Regarding tasks

- You may use other variables we used in class,
- You may do different feature engineering
- You may make other sample design decisions
- But document your steps

- Have 3 different models and compare performance
  - Argue for your choice of models
  - One model must be OLS another must Random Forest or *any* boosting algorithm, the third could be either or stg different

Tasks 2

- Please pick another date. (For instance December vs June)
- Carry out the exercise for the second one
- Take your selected model and use it on the other date
- Compare the predictive power of the model. Discuss briefly in the report

Tasks 3

- Consider your ML model (RF or Boosting)
- Explain features' contribution to predicted values *on average*, using Shapley values
- You shall use a SHAP method
- You may use graphs, but the emphasis is on the discussion part.
- *This is worth relatively few points, do it only if everything else is already done*

Submit two documents to moodle

- A summary report (pdf), min 3, max 5 pages including tables and graphs discussing your work. It is targeted to data science team leaders
  - Can use technical language
  - But need to be the point
  - Focus on key decision points, results, interpretation, decision
- Technical report -- a markdown / notebook in pdf/html with more technical discussion.
  - May include code snitppets
  - May include additional tables and graphs
  - Detail all decisions you made
- Reports should link to code in github

Scoring weights

- Task 1: Data prep, label, and feature engineering (20%)
- Task 1: Model building, prediction, and model selection (20%)
- Task 1: Discussion of steps, decisions, and results (20%)
- Task 2: Technical execution and discussion (15%)
- Task 3: Technical execution and discussion (5%)
- Overall quality of the write-up, prettiness of graphs, etc (20%)