Assignment II

Data Analysis 3

MS in Business Analytics,
2023/2024 Winter

Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

Goal

- 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.
- 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 the 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 (may check archive data)
- Everybody will have a different city - Please sign up for a city HERE.

Task

- You may use other variables we used in class.
- You may do different feature engineering depending on the selected environment.
- You may make other sample design decisions!
- In each case, document your steps!
- Have at least **3 different models** and compare performance
- Argue for your choice of models
  - One model must be theoretically profound linear regression via OLS.
  - One model must be Random Forest or any boosting algorithm

Extra Task - 5p

- 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 not an easy task, and worth relatively few points, do it only if everything else is already done.

Documents to submit on moodle

- A summary report (pdf), **max 3 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 / quarto 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

Overall you can get 30p from this task plus 5p for the extra (non mandatory) part.

- Data prep, label and feature engineering (8p)
- Model building, prediction and model selection (12p)
- Discussion of steps, decisions and results (7p)
- Quality of the write-up (3p)
- Extra task (5p)

Submission deadline: 11th of February, 2024 23.59 CET.

Good luck!