# **Assignment 2**

Use the hotels-europe dataset and pick a city (other than Vienna). Each student/pair should pick a different city with at least 250 hotels in the dataset after appropriate filtering of the data. Sign up with the city chosen in this spreadsheet.

Once the city is chosen, first combine the prices dataset with the hotel features dataset. Then, use hotel user ratings to create a binary variable: highly_rated=1 if rating ≥ 4, 0 otherwise. Examine how high rating is related to the other hotel features in the data. Make sure to document and argue for the data cleaning/filtering decisions you make.

Estimate linear probability, logit, and probit models with distance and stars as explanatory variables. You may add other variables if you wish.
Compare coefficients, marginal differences, and predicted probabilities, and discuss your results.

This is an assignment that you may do alone or in a pair within your group. You shall write a short (max. 1 page, excluding exhibits) report that describes your major decisions, your estimated models, interpretations and summary. You may have a descriptive data table (optional), a regression table and up to two graphs. All codes should be pushed to your version control repo, with an appropriate commit policy.

You need to upload your report in pdf format. Add the link of your repo to the online text field.