Goal: predict the number of rented bikes aggregated by hour.

Data: extracted information from bike rental service aggregated by hour here hour.csv with description Readme.txt. Also available in *HOMEWORK_DATASETS['bike_sharing_hour']* library *bds_courseware.*

Target: `cnt` - Number of bikes rented per hour (unless explicitly stated otherwise). It is a sum of 'casual' + 'registered' so make sure the latter does not appear in train sample.

Test sample: random subsample of the data (5-15% size) unless explicitly stated otherwise.

Metrics: MAE

To-do:

- EDA (2 points)
- Preprocessing (1 point)
- Use different advanced regression algorithms and select the best one:
    - KNN Regressor (1 point)
    - Decision Tree Regressor (1 point)
    - Random Forest (1 point)
    - SVM Regressor (1 point)
    - Boosting Regressor - any Boosting algorithm with Regression (1 point)

  Warning! The task is considered to be done for each model if a hyperparameter tuning was performed and the quality is measured on a test sample.
- Considering the best model, train it on the first year and validate on the second year. The results are better or worse? Why? (1 point)
- Considering the best model, train it on the first year with multitarget (**casual, registered**) and validate on the second year. Analyze both targets independently. Then use a sum of both targets to predict **cnt** on the second year. Estimate the quality. Are the results better or worse in comparison with the previous step? Why? (1 point)
- During your research, please answer the following questions:
    - What are the most influential features? (1 point)
    - How the temperature and humidity affect rental count rates? (1 point)
    - What is more important: current season or current hour, temperature or feeling temperature? Why? (1 point)

Total points 13 + 3 for additional research which is not included in the criteria.