An anonymous salary survey has been conducted annually since 2015 among European IT specialists with a stronger focus on Germany. The data has been made publicly available by the authors. The dataset contains rich information about the salary patterns among the IT professionals in the EU region and offers some great insights.

Source: <https://www.kaggle.com/parulpandey/2020-it-salary-survey-for-eu-region>

Please consider the following tasks.

* Based on 2020 data (IT\_Salary\_Survey\_EU \_2020.csv) build a multiple linear regression model.
* Select a dependent variable which represents the amount of IT specialist’s salary. Select not less than 10 predictors. Out of them at least 3 predictors should be categorical.
* Formulate all possible research hypotheses which will be verified during the data analysis.
* Specify the regression equation.
* Assess the goodness-of-fit of the model.
* Describe the relationships between each predictor and dependent variable in detail. This is a very important point of the task. We should understand clearly from the analysis which predictors are related to the salary and how.
* Perform the model diagnostics. You should interpret all the diagnostic statistics discussed at the lecture.
* Make final conclusions about the results of analysis.

Please send the Jupyter Notebook with comments and answers to amelikyan@hse. The task should be done **individually**.The deadline for submitting the task is **15 April 14:00**.