**Overview**

[**Data**](https://drive.google.com/drive/folders/13zKKNCa5srQSC8pqYWCIn5oRPXMlOGid?usp=sharing)

* Regression analysis of Air Quality dataset
* Target - C6H6(GT)
* **-200** indicates missing values

**Required**

Jupyter notebook file (.ipynb) or link to colab/github attached

**Evaluation Criteria (20 max)**

* EDA: exploration of variables and properties of data with conclusions: 2
* Data preparation: all missing values are filled/dropped with explanation of why some method was chosen: 2
* Data preparation: normalization of data / scaling: 1
* Feature engineering - basic transformations (nonlinear): 2
* Baseline model - linear regression without regularization: 2
* Metrics chosen as well as reasoning behind each metric: 2
* Feature importance, hyperparameters tuning: 2
* Statistics material used (residual analysis, factor-importance hypothesis testing): 2
* Quality of delivered work:
  + Analytical comments provided: 1
  + The experiment is structured (file is readable, pictures have titles): 1
  + Code is clear (reusable code in functions, comments,   
    code is easy readable): 1
* Extra points for improvements not considered in the criteria: 2