# Advanced Regression

## Summary

This assignment is designed to wrap up your knowledges in regression analysis and exploratory data analysis.

Link to the competition [here](https://www.kaggle.com/c/house-prices-advanced-regression-techniques).

This competition is on real world data and helps you to apply your skill in solving simple daily tasks in machine learning area. You’ll learn not only key concepts of nowadays techniques and approaches but also in communication with the data science and machine learning community

## Description

This task’ data contains mix of caveats and pitfalls. Some of them are listed below (but not limited to):

* Missed data
* Outliers
* Skewed data
* Etc.

Those entities can have significant impact on your model’ accuracy, so to achieve your goal you must explore this dataset thoroughly and handle each of them.

In scope of the task you must handle:

* Analyze existing data, clean it out, normalize and extract out list of features meaningful for future model.
* Select best model to solve this task and achieve the accuracy as big as possible.

As a starting point you can look into [this](https://www.kaggle.com/gunesevitan/in-depth-eda-and-stacking-with-house-prices) kernel.

You can apply more than one approach for EDA and / or model selection to select the most effective result. The report can describe all approaches with its pros and cons.

## Report

After the research and implementation parts are completed and you had got the best result you should create report with explanation what has been done during the assignment.

Your work should contain:

* Report with short description of key concepts of your approach and detailed explanation of methods for solving issues with data analysis and model selection.
* Link to the leaderboard with your results