Instructions

The aim of the project is to write a **functional and structured code in python using a jupyter notebook**.

The code should be able to make a prediction on the dataset Auto Insurance based on the Kaggle competition https://www.kaggle.com/c/auto-insurance-fall-2017 .

The target for this project is **TARGET_FLAG**.

The code should explicitly:

- return the performance of the algorithms tested using the **appropriate metric**
- explain with markdown **comment important modelling decisions** including metric selection
- generate a csv **file with the predictions**

Performance is not the main goal. The first objective of the project is to write the **main steps of a data-science project in 2 to 4 hours** with a proper style and well written comments. The algorithms used do not have to be hyper-optimized on all parameters. However, the project should test different algorithms adapted to the type of problem.

The project should be shared via a github, gitlab or any other public git repository kept available until the end of the technical interview.

We will discuss about the project for about 20 min during the interview.

FAQ:

The dataset contains insurance claim statics on the customers. The names of the keys and categories are nearly all transparent. We can answer to specific questions on the variables.

The files SHELL_AUTO and MEAN_AUTO can be ignored for this project.