**Data Science Challenge**

Dear candidate,

You have made a good impression in the first steps in our recruitment process and we would like to progress with you to the next part of our assessment.

The purpose of the data science challenge is to let you demonstrate the way you think and work.

This exercise should take around 2 to 4 hours to be completed but you have 48 hours to return your solution.

Please take a look at the dataset in the file “Auto1-DS-TestData.csv” (see <https://archive.ics.uci.edu/ml/datasets/Automobile> for information on the features and other attributes) and answer the following questions:

Question 1 (10 Points)

List as many use cases for the dataset as possible.

Question 2 (10 Points)

Auto1 has a similar dataset (yet much larger...)

Pick one of the use cases you listed in question 1 and describe how building a statistical model based on the dataset could best be used to improve Auto1’s business.

Question 3 (20 Points)

Implement the model you described in question 2 in R or Python. The code has to retrieve the data, train and test a statistical model, and report relevant performance criteria.

When submitting the challenge, send us the link for a Git repository containing the code for analysis and the potential pre-processing steps you needed to apply to the dataset. You can use your own account at github.com or create a new one specifically for this challenge if you feel more comfortable.

Ideally, we should be able to replicate your analysis from your submitted source-code, so please explicit the versions of the tools and packages you are using (R, Python, etc).

Question 4 (60 Points)

A. Explain each and every of your design choices (e.g., preprocessing, model selection, hyper parameters, evaluation criteria). Compare and contrast your choices with alternative methodologies.

B. Describe how you would improve the model in Question 3 if you had more time.

**We are looking forward to receiving your solution!**