# ML Engineer homework

Greenscreen - a platform for freight brokers in US. The core product of the platform is Rate Engine (RE). RE provides the price prediction for loads (order for cargo delivered from origin to destination by truck). The price depends on the origin/destination region and many market conditions. It is common for the industry to calculate the rate per mile as price divided by miles and use as a pricing indicator for the load (for example rate = 2,2 $/mi).

We have divided all US territory into KMA (Key Market Regions). These regions grouped by similar market conditions that are inside each market. Try to enhance the current Rate Engine by pushing knowledge about origin and destination KMA into model.

### Input

You have code prototyping of RE model training for rate prediction ([github](https://github.com/mipo47/rate-prediction-test)). RE can predict rates based on several basic parameters. Rate quality prediction is measured by MARE (mean average relative error):

There are two files with data: train.csv and validation.csv (validation set infront of train in term of date), file kma\_description.csv explaining kma codes.

### Output

Enhance the existing code of RE to be able to push KMA knowledge for origin and destination into the model. How much does it help?

Please provide a result as:

1. Enhanced py script that could take csv file to make a prediction for future loads. Program should provide an output csv file with predictions
2. Estimate your RE quality