**Problem Statement:**

**Description**

In Russia, vehicle registration plates (commonly called “numbers”) consist of three letters (from 'ABEKMHOPCTYX'), three digits (except '000'), and a region code. Our dataset contains car plate sale offers collected from various sources, including aggregators, Telegram channels, and websites. Since the same plate may have been listed multiple times at different prices, we recorded all observed price changes over time. Some plates are affiliated with government agencies or have symbolic meanings, making them significantly more valuable (more on that in the "Data" page). Participants may want to consider these factors when building their models.

**Evaluation**

Submissions will be evaluated using the **Symmetric Mean Absolute Percentage Error (SMAPE)**. This metric measures the accuracy of predicted prices relative to actual prices, ensuring a fair evaluation regardless of price scale. A lower SMAPE score indicates better model performance.

**Submission File**

For each id in the test set, you must predict the price variable. The file should contain a header and follow this format:

id,price  
2,100000  
5,200000  
6,300000  
etc.

**Existing Approach:**

Extract the numbers and alphabets from the plates

Identify patterns in both numbers and alphabets such as Palindrome, repetitiveness, symmetry score, memorability score, sequential & ladder pattern.

Checking if the plate has been forbidden by the government or not, what is the significance that the author mentioned.

Built SMAPE error metric and then built the model using autogluon

**Expected:**

1. Haven’t done any date related analysis to add required features which can significantly improve the SMAPE metric. Try integrating this with advanced time based patterns to add valuable features.
2. Did not use the Region\_code and Government\_code from the supplemental\_english file appropriately.
3. Did not transform the Price values even though there is no distinct patterns in the raw values as is. Try Log or any other transformation to generate valuable patterns.
4. Did not perform k-fold cross validation to handle over-fitting. Apply the best CV technique to generalize the model
5. Try only ensemble models using autogluon and generate only SMAPE evaluation.
6. Generate the Submissions.csv file as mentioned in the problem statement.
7. All the steps must be one-by-one categorized and the results must be generated within the same block before moving on to the next one.