Import opportunities in the European second-hand car market for premium brands

Objective

The objective of this project is to search for import opportunities in the European second-hand car market for premium brands. (Mercedes-Benz, Audi, BMW, Porsche and Land Rover)

Summary

It begins with the acquisition of data, second-hand car deals, both from the Spanish and European markets. The data is then manipulated to be able to work with it.

Once the data is ready to be used, the offers from the Spanish market are taken and a Machine Learning algorithm is trained to be able to predict the price of a second-hand car in the Spanish market. Finally, the Machine Learning algorithm predicts the price of European second-hand car deals in the Spanish market.

Those offers that present a greater difference between their real price in Europe and their predicted price for Spain will be the best import opportunities.

Stages

Data acquisition

The second-hand car offers have been obtained by doing Web Scraping of the portal https://www.autoscout24.es/. It is one of the ones that has the most offers both at a national and European level. Notebooks that perform this task can be found in the (Web_Scraping) folder. The data is stored in CSV files in the Raw_Data folder.

Data manipulation

The offers contained in 94 CSV files are manipulated and transformed to be able to efficiently exploit the data in Machine Learning algorithms. Each car brand is treated separately in order to recategorize the models correctly. Notebooks that perform this task can be found in the (Data Manipulation) folder. The data is stored in CSV files in the Clean_Data folder.

Data exploration and visualization

This stage serves to become familiar with the data and to see what relationship the different variables extracted for each offer have with the target variable to predict, the price. The notebook that performs this task can be found in the (Data Exploration and Visualization) folder.

Machine Learning Models

The best Machine Learning algorithm is being sought to predict the prices of second-hand cars in the Spanish market. Various pretreatments to variables are also tested. The chosen algorithm is Sklearn's GradientBoostingRegresor. Notebooks that perform this task can be found in the (Machine_Learning_Models) folder.

Import opportunities

Sklearn's GradientBoostingRegresor is used to predict the price of European offers in the Spanish market. These are ordered according to the difference between their real European value and their prediction in the Spanish market. In addition, these are segmented based on their real price to be able to adjust to the desired investment. The notebook that performs this task can be found in the (Import_Oportunities) folder.

Results

Each of the offers is reviewed by hand, analyzing whether it is real offers or if it was some type of scam or similar. Finally, four CSV files are delivered with the best offers already reviewed.

Conclusions

There are currently import opportunities in the European second-hand car market. The best investment opportunities are found in a group of cars characterized by being large cars, few kilometers, great power and between 2 and 6 years old. This is because in Europe they suffer a faster depreciation.