Raciocínio do Projeto

* Primeiramente eu procurei um dataset com um conteúdo mais fácil de se compreender no site kaggle.
* Importei o dataset e analisei os seus dados, utilizando a função setwd() e read.csv().
* Renomeei os nomes das colunas para portugues, para ficar mais compreensível
* Analisando o data frame, utilizando a função view(), notei que apareciam muitas vezes o termo “used” na coluna certificado, então decidi verificar se ele seria um outlier.
* Para isso, criei uma frequência absoluta da coluna certificado, a deixei ordenada usando sort() e depois criei um gráfico de boxplot, o qual aponta se existe algum outlier.
* Após verificar que ele realmente é um outlier, decidi trabalhar apenas com ele, já que tinha muitos dados apenas para carros usados, então criei um novo data frame para ser trabalhado.
* Analisei qual seriam as avaliações mais comuns entre esses carros usando novamente o método para calcular a frequência absoluta de cada valor, também o organizando em um valor crescente, para ficar mais visível.
* Quando criei o gráfico de barras, ele ficou grande demais e aparece um erro, porém com o meu conhecimento, eu não consegui fazer uma limpa para que ele fique mais visível e entendível. Porém decidi deixá-lo para ao menos ficar aparente que as avaliações maiores são mais frequentes.
* Pesquisei alguns métodos que dessem para retirar certas linhas das colunas e fazer uma limpa no conteúdo de certas colunas, para que eu conseguisse tratar elas como inteiros e fazer análises, porém não consegui e as analisei como strings.
* Ao analisar os carros com mais reviews, notei que não bateu com a realidade, pois ele está olhando para os valores como strings e não inteiros.
* Por fim, exportei o arquivo csv do novo data frame modificado.