

# Môn: Trực Quan Hóa Dữ Liệu

## Câu Hỏi Cá Nhân

26/06/2022

### Dữ liệu

Tập dữ liệu [ToyotaCorolla.csv](#) là thông tin liên quan đến giá bán xe Toyota đã qua sử dụng của công ty kinh doanh xe cũ được ghi nhận trong tháng 8 năm 2004.

Biến	Mô tả
<a href="#">Price</a>	Giá xe
<a href="#">Age</a>	Tuổi xe (tính đến 8/2004)
<a href="#">Kilometers</a>	Quãng đường đã đi
<a href="#">Fuel_type</a>	Loại nhiên liệu
<a href="#">HP</a>	Động cơ
<a href="#">Metallic</a>	Màu metallic? (Yes = 1, No = 0)
<a href="#">Color</a>	Màu
<a href="#">Automatic</a>	Xe tự động (Yes = 1, No = 0)
<a href="#">CC</a>	Dung tích
<a href="#">Doors</a>	Số cửa
<a href="#">Quarterly_Tax</a>	Thuế
<a href="#">Weight</a>	Trọng lượng

### Yêu cầu

1. Hãy trực quan hóa các thông tin thống kê mô tả cho các biến.
2. Tìm và trực quan mối quan hệ *tương quan* giữa các cặp biến (nếu có).
3. Hãy trực quan hóa biểu đồ histogram cho [Price](#) theo từng biến biến

- `Fuel_type`
- `Color`

4. Hãy đưa ra mô hình dự báo về giá xe `Price` (có thể sử dụng mô hình hồi quy logistic hoặc mô hình học máy bất kỳ).

## Lưu ý

Sinh viên làm

- Bằng **notebook** (colab hoặc jupyter, ...)
- Lưu thành tập tin `[Student ID].pdf`
- Nộp tập tin `[Student ID].pdf` qua Moodle