

ĐỀ KIỂM TRA GIỮA KỲ 02 – THỐNG KÊ MÁY TÍNH

Đề bài: Cho dataset mpg.csv về mức tiêu hao nhiên liệu. Thông tin của dữ liệu như sau:

- mpg: Miles/(US) gallon số gallon xăng trên 1 dặm
- cylinders: số lượng xilanh
- displacement: Displacement/ dung tích xilanh (cu.in / inches³)
- horsepower: công suất theo mã lực
- weight: trọng lượng (pound)
- acceleration: Gia tốc
- model_year: Năm sản xuất
- origin: Xuất xứ
- name: Tên dòng xe

Hãy thực hiện:

Câu 1. (2.0 điểm) Đọc file dữ liệu trên và Kiểm tra có dòng nào không có dữ liệu? Xóa các dòng không có dữ liệu

Câu 2. (2.0 điểm) Tìm trung bình, trung vị của các thuộc tính: ‘cylinders’, ‘horsepower’, ‘weight’. Tìm miền giá trị và miền phân vị của thuộc tính ‘mpg’.

Câu 3 (2.0 điểm) So sánh mức tiêu hao nhiên liệu trung bình của dòng xe dựa trên xuất xứ. Tính tỷ lệ các dòng xe dựa trên xuất xứ.

Câu 4. (2.0 điểm) Vẽ Histogram trọng lượng các xe. Vẽ Boxplot so sánh công suất của các dòng xe thập niên 70 và 80.

Câu 5. (2.0 điểm) Khảo sát mức tiêu hao nhiên liệu bằng Scatter Plot theo các thông số: Số lượng xi lanh; Công suất theo mã lực; Gia tốc, Năm sản xuất