

## ĐỀ1

### Câu 1: (2 điểm)

Sử dụng hàm trong R để chuyển đổi dữ liệu trong file DATA1.txt từ dạng cột sang dạng dòng, khóa được sử dụng là MADL.

### Câu 2: (2 điểm)

Sử dụng gói quantmod để lấy dữ liệu về cổ phiếu của mã AAPL, vẽ đồ thị bằng hàm chartSeries cho dữ liệu được lấy trong khoảng thời gian như sau: bắt đầu từ 1/1/2021 đến ngày và tháng là ngày sinh và tháng sinh của sinh viên, năm là 2021.

### Câu 3: (3 điểm)

Cho file dữ liệu DATA3.txt chứa thông tin về các nhóm khách hàng và tour du lịch mà nhóm khách hàng đã đăng ký. Giả sử các tour du lịch được đánh số từ 1-n, các nhóm khách hàng được đánh số từ 1-m. Sử dụng các công cụ trong R để:

- Tìm n, m
- Sử dụng các công cụ cần thiết và vẽ đồ thị để đưa ra phân tích các thông tin về mối quan hệ giữa tour du lịch và khách hàng, nhóm khách hàng.

### Câu 4: (2 điểm)

Xác định hàm hồi quy cho dữ liệu trong file DATA4.txt (Cột 1 là biến phụ thuộc, các cột sau là biến độc lập). Xác định mô hình hồi quy tối ưu và giải thích.

### Câu 5: (1 điểm)

Tìm số cụm tối ưu, sử dụng thuật toán để phân cụm k-means và vẽ phân cụm cho bộ dữ liệu DATA5.

## đề 2

### Câu 1: (2 điểm)

Sử dụng hàm trong R để chuyển đổi dữ liệu trong file **DATA16.txt** từ dạng cột sang dạng dòng, khóa được sử dụng là **MAHS**.

---

### Câu 2: (2 điểm)

Sử dụng gói **quantmod** để lấy dữ liệu về cổ phiếu của mã **CSCO**, vẽ đồ thị bằng hàm **chartSeries** cho dữ liệu được lấy trong tháng là tháng sinh của sinh viên, năm là năm 2022. Nếu tháng sinh là 11 hoặc 12 thì lấy năm là 2021. Ví dụ, sinh viên sinh ngày 21/8/2002 thì sẽ lấy dữ liệu từ 1/8/2022 đến 31/8/2022.

---

### Câu 3: (3 điểm)

Cho file dữ liệu **DATA36.txt** chứa thông tin về số tiền bán hàng của 2 nhóm đại lý.

- Sử dụng R để:
    - Tóm tắt thống kê cho từng nhóm.
    - Trực quan hóa dữ liệu bằng boxplot, đưa ra các nhận xét đặc điểm cho từng nhóm và so sánh giữa hai nhóm.
- 

### Câu 4: (2 điểm)

Xác định hàm hồi quy cho dữ liệu trong file **DATA46.txt** (Cột 1 là biến phụ thuộc, các cột sau là biến độc lập). Xác định mô hình hồi quy tối ưu và giải thích.

---

### Câu 5: (1 điểm)

Tìm số cụm tối ưu, sử dụng thuật toán để phân cụm **k-means** và vẽ phân cụm cho bộ dữ liệu **DATA56**.

