

BÀI TẬP TUẦN 3

I. Bài tập tại lớp

Bài 1: Metacritic là trang web nổi tiếng chuyên đánh giá sản phẩm trò chơi điện tử, phim ảnh,... Dữ liệu được Patrick Klein thu thập thông qua công cụ scraper và đăng lên [Kaggle](#).

Trong môi trường Spark của Zeppelin, đọc bảng games.csv bằng Python từ đường dẫn cuối bài. Sau đó, xóa cột summary và title, lấy mẫu có kích thước là 200 và gửi cho R.

Link games.csv:

<https://drive.google.com/uc?export=download&id=1ZR3F9ZhveHmLkhTaK3uxZ2rmTIB0B8Zv>

Bài 2: Dùng R vẽ biểu đồ phù hợp thể hiện tính tương quan điểm số trung bình của Metacritic và người dùng theo hệ máy.

Bài 3: Lựa chọn và vẽ biểu đồ phù hợp để thể hiện miếng bánh thị phần game theo nền tảng trong năm 2022.

II. Bài tập về nhà (nhóm)

Bài 4: Sử dụng R để đọc dữ liệu điểm kì thi Quốc Gia 2022 từ [Kaggle](#). Lọc ra các thí sinh dự thi ở Hà Nội, Đà Nẵng và thành phố Hồ Chí Minh.

Bài 5: Từ dữ liệu đã lọc ở bài 4, chọn và vẽ biểu đồ phù hợp thể hiện mức độ tương quan giữa năm môn toán, lý, hóa, văn và ngoại ngữ bằng R.

Bài 6: Dùng R, khám phá dữ liệu (từ bảng bài 4 và các bảng khác trên Kaggle) và viết báo cáo về những điều tìm được trong notebook. Đưa ra ít nhất hai cơ sở lập luận và một kết luận.

Yêu cầu:

- *Nộp file JSON Zeppelin hoặc IPYNB Jupyter.*
- *Upload dữ liệu từ Kaggle lên các trang đám mây như Google Drive, viết code để tải link trực tiếp.*