

Kiểm tra Môn: Lập trình Python

Thời gian: 75 phút

SV thực hiện các thao tác với dữ liệu trên dữ liệu IMDB -Movie - Data.csv về thông tin của các bộ phim

Với các yêu cầu sau:

A. Tìm hiểu về dữ liệu

- a. Hiển thị 5 hàng đầu tiên của bảng dữ liệu
- b. Hiển thị thông tin đầy đủ của các kiểu dữ liệu của các field
- c. Thống kê đầy đủ về dữ liệu (describe)
- d. Lấy các cột Title, Genre , Actors, Director, Rating
- e. Kiểm tra số lượng các giá trị null có trong từng cột của bảng dữ liệu
- f. Với phương án thế giá trị mới vào các ô bị mất dữ liệu, ta có thể sử dụng các giá trị mean, median... của cột dữ liệu tương ứng để thay thế (việc chọn giá trị để thay thế còn tùy thuộc vào tính chất của bộ dữ liệu, bài toán đang giải quyết). Thay thế các giá trị trung bình trên các dòng của Revenue mang giá trị null
- g. Kiểm tra những dữ liệu trùng, xóa dữ liệu trùng

B. Thống kê dữ liệu

- a. Lấy các bộ phim từ 2010 tới 2015, với rating nhỏ hơn 6.0 nhưng lại có doanh thu đứng đầu.
- b. Lấy ra các bộ phim thuộc thể loại “Action”
- c. Lấy ra 5 bộ phim được votes nhiều nhất.
- d. Tìm số rating trung bình mà các đạo diễn đạt được
- e. Tạo cột phân loại Rating “Good”, “Average” và Bad theo tiêu chí sau: Với Rating \geq 7.5 thuộc loại: Good;
Rating \geq 6 và <7.5 thuộc loại: Average,
Rating < 6 thuộc loại Bad
- f. Tính tổng doanh thu của các bộ phim theo Director.

Sau khi làm xong SV nộp bài trên: <https://forms.gle/FVV2TUbjh2diUDxp8>