

Hướng dẫn Đồ án Quan hóa Dữ liệu

Mục tiêu

Mục tiêu của đồ án này là giúp sinh viên làm quen với các thao tác cơ bản trong quá trình phân tích dữ liệu và đưa ra kết luận dựa trên phân tích đó. Sinh viên cũng sẽ được phát triển kỹ năng lập trình bằng việc tự cài đặt các phân tích dữ liệu sử dụng ngôn ngữ lập trình Python 3.

Thời gian

Đồ án sẽ kéo dài trong vòng 3 tuần. Chi tiết về thời gian cụ thể có thể được tìm thấy trên Moodle.

Yêu cầu nộp đồ án

- Sinh viên hoặc nhóm sinh viên nếu có cần nén thư mục bài làm thành định dạng zip và nộp qua Moodle.
- Tên thư mục bài làm:
 1. Nếu chỉ có một sinh viên: **<MSSV>**
 2. Nếu có 2-3 sinh viên: **<MSSV1>_<MSSV2>_<MSSV3>**
- Thư mục bài làm bao gồm các phần sau:
 1. Báo cáo trả lời các câu hỏi tự luận: Định dạng PDF, trang đầu tiên ghi thông tin nhóm, tỉ lệ thực hiện của mỗi thành viên và các câu hỏi chưa làm được.
 2. Mã nguồn chương trình cài đặt: Đặt trong thư mục "Source", bao gồm các file mã nguồn liên quan trong bài tập lập trình. Ngôn ngữ sử dụng là Python 3.

Dataset

Sinh viên có thể chọn một trong những Dataset sau để thực hiện việc phân tích:

[Dataset](#)

Chi tiết các phần

I. Data Understanding (20%)

1. Đếm số dòng và số cột.
2. Viết bảng mô tả về các cột.
3. Lấy 5 điểm dữ liệu ra làm mẫu.
4. Phân tích tỷ lệ missing rate.
5. Phân tích tỷ lệ duplicate.
6. Fill missing rate.

II. EDA 1D (20%)

1. Chia loại dữ liệu num hoặc cate.
2. Phân tích tỷ lệ đó với biến cate.
3. Phân tích phân phối đối biến num.

III. EDA 2D (20%)

1. Phân tích hệ số tương quan giữa các biến num.
2. Sử dụng Scatter plot để phân tích dữ liệu 2D.
3. Sử dụng bar chart để phân tích dữ liệu num và cate.
4. Tính tỷ trọng đối với hai biến cate.

IV. EDA 3D (20%)

1. Sử dụng Scatter plot để phân tích dữ liệu 3D cho ba biến num.
2. Sử dụng Scatter plot 2D và màu đối với hai biến num và cate.
3. Tính tỷ trọng theo bin chia theo thể loại với hai biến cate.

V. Insight (20%)

Rút ra insight và kết luận từ những phân tích đã thực hiện ở các phần trước.

Chi tiết xem ở file notebook: [MyHotel.ipynb](#)

Lưu ý

- Sinh viên cần tuân thủ đúng các yêu cầu và định dạng nộp bài như đã quy định.
- Mọi vấn đề hoặc thắc mắc có thể được thảo luận trực tiếp trên Moodle hoặc thông qua email: truongthk62014@gmail.com với giảng viên hướng dẫn.