

KIỂM TRA GIỮA KỲ

Môn: Tương Tác Dữ Liệu Trực Quan

Thời gian: 75 phút

ĐỀ 1

Bài 1: Tính hệ số moment.

Dữ liệu sau đây là thông tin mua bán bất động sản ở ngoại ô Abbotsford, bang Viectoria, Úc lấy từ [Kaggle](https://cloud.vinhthanh.net/s/3MpCNPWDiH7reec). Tải và nạp dữ liệu bằng Pandas từ link dưới đây:

<https://cloud.vinhthanh.net/s/3MpCNPWDiH7reec>

Chỉ bằng numpy, tính và in ra ba hệ số moment sau của giá tài sản: Trung bình, moment trung tâm thứ 3 và độ xiên. Giữ nguyên giá trị thập phân, không làm tròn.

Chú thích:

- Trung bình: *Mean*
- Moment trung tâm thứ 3: *Third central Moment*
- Độ xiên: *Skewness*

Bài 2: Lọc dữ liệu, kết quả lọc bao gồm các bất động sản thỏa điều kiện:

- Đã bán thông qua đấu giá
- Đầy đủ dữ liệu tọa độ
- Tổng diện tích trên 450m²

Bài 3: Thể hiện tương đối lượng phòng ngủ, phòng tắm, xe hơi có thể chứa của nhà thứ 105 và 106 (xét cột chỉ mục có sẵn trong dữ liệu) trên cùng biểu đồ lưới/mạng nhện bằng matplotlib hoặc plotly.

Chú thích:

- Biểu đồ lưới/mạng nhện: Radar chart/web chart/spider web chart

Bài 4: Dựa trên công thức Sturge, chỉ bằng numpy, tính ba giá trị min_x , w , y và vẽ histogram khoảng cách (đơn vị km) từ bất động sản đến CBD theo các số vừa tính. Có thể dùng matplotlib hoặc Plotly để vẽ.

Bài 5: Vẽ biểu đồ điểm với x là diện tích nhà, y là diện tích đất trống. Chỉ bằng numpy, thực hiện hồi quy đơn thức, tính β và hệ số chặn của đường xu hướng. Vẽ đường xu hướng đó trên biểu đồ vừa vẽ.

Lưu ý:

- Có thể tách thành hai biểu đồ: Trước và sau khi vẽ đường xu hướng, hoặc chỉ vẽ biểu đồ có đường xu hướng.
- In kết quả theo đúng thứ tự: β và hệ số chặn.

STT	Mục tiêu	Bài
1	Biết cách dùng Pandas để đọc dữ liệu.	Bài 1
2	Hiểu và tính toán các hệ số moment thống kê.	Bài 1
3	Biết cách lọc dữ liệu với nhiều điều kiện cùng lúc.	Bài 2
4	Hiểu dữ liệu, thông qua đọc hiểu mô tả và tìm kiếm thông tin.	Bài 2
5	Có thể vẽ đồ thị dạng cơ bản.	Bài 5
6	Biết cách vẽ đồ thị nâng cao.	Bài 3, bài 4
7	Tính toán hồi quy và thêm các đường thống kê (đường xu hướng, đường trung bình,...).	Bài 5

Bài	Mức độ	Điểm
Bài 1	Dễ	2.5
Bài 2	Dễ	2.5
Bài 3	Trung bình	2
Bài 4	Trung bình	1.5
Bài 5	Trung bình	1.5
<i>Tổng</i>		<i>10</i>