

Môn học: Thu thập và tiền xử lý dữ liệu

MSSV: 20521196

Họ tên: Nguyễn Mạnh Đức

Code Book:

Dữ liệu thu thập tại nguồn:

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-cong-nghe-tphcm/>

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-cong-nghiep-thuc-pham-tp-hcm/>

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-kinh-te-tai-chinh-tphcm>

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-kinh-te-quoc-dan>

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-su-pham-ha-noi>

<https://thituyensinh.ican.vn/diem-truong/dai-hoc-viet-nhat-dh-quoc-gia-ha-noi>

Thu thập bằng việc Ctrl + c và Ctrl + v vào excel, sau đó chỉnh sửa, sắp xếp, xóa bớt

Thông tin mục tiêu bộ dữ liệu:

- Dự đoán điểm chuẩn đầu ngành CNTT ở TpHCM và Hà Nội
- Kiểu giá trị dự đoán: float

Có 8 thuộc tính: STT, Tên trường, Địa điểm, Tên Ngành, Mã ngành, Tổ hợp môn, Điểm chuẩn, Ghi chú.

- STT: Số thứ tự sắp xếp. Giá trị từ 1 → 73. Kiểu dữ liệu integer
- Tên trường: Đại học Kinh tế - Tài chính TP HCM, Đại học Công nghiệp thực phẩm TP HCM, Đại học Công nghệ TP HCM, Đại học Sư phạm Hà Nội, Đại học Kinh tế quốc dân, Đại học Việt Nhật - ĐH Quốc gia Hà Nội. Kiểu dữ liệu string
- Địa điểm: TP HCM, Hà Nội. Kiểu dữ liệu string
- Tên ngành: Công nghệ truyền thông, Thương mại điện tử, Công nghệ thông tin, Khoa học dữ liệu, An toàn thông tin, Công nghệ kỹ thuật điện, điện tử, Hệ thống thông tin quản lý, Robot và trí tuệ nhân tạo, Khoa học và Kỹ thuật máy tính. Kiểu dữ liệu string.
- Mã ngành: 7320106, 7340122, 7480201, 7480109, 7480202, 7510301, 7340405, 7480207, 7480201A, 7480201B, EP03, 7480204

- Tổ hợp môn: Tổ hợp xét tuyển, bao gồm A00, A01, D01, C01, D07, C00, XDHB.
- Điểm chuẩn: 1-70. Kiểu dữ liệu float
- Ghi chú: Lưu ý. Kiểu dữ liệu string.