CODE BOOK Bài 2a

Thông tin	Nội dung
Tên bộ dữ liệu	Iris Plants Database
Nguồn thu thập và cách thức thu thập	(a) Creator: R.A. Fisher (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov) (c) Date: July, 1988
Số thuộc tính	5
Thông tin tên các thuộc tính	1. sepal length in cm 2. sepal width in cm 3. petal length in cm 4. petal width in cm 5. class: -- Iris Setosa -- Iris Versicolour -- Iris Virginica
Thông tin tác giả	(a) Creator: R.A. Fisher (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)

Raw data: file iris.data

Tidy data: được lưu lại thành file: iris.csv

Instruction list:

CODE R

CODE BOOK Bài 2b

Thông tin	Nội dung
Tên bộ dữ liệu	Bank Marketing
Nguồn thu thập và cách thức thu thập	Paulo Cortez (Univ. Minho) and SÃ©rgio Moro (ISCTE-IUL) @ 2012
Số thuộc tính	16 + thuộc tính đầu ra
Thông tin tên các thuộc tính	*# Dữ liệu khách hàng của ngân hàng:* 1 - age (numeric) 2 - job : loại công việc (phân loại: "admin.","unknown", "unemployed", "management","housemaid","entrepreneur", "student","bluecollar","selfemployed", "retired", "technician", "services") 3 - marital : Tình trạng hôn nhân (phân loại: "married", "divorced","single"; note: "divorced" means divorced or widowed) 4 - education (phân loại: "unknown","secondary", "primary","tertiary") 5 - default: Có tính dụng trong tình trạng vỡ nợ? (nhị phân: "yes","no") 6 - balance: số dư trung bình hàng năm (numeric) 7 - housing: có cho vay mua nhà không? (nhị phân: "yes","no") 8 - loan: có khoản vay cá nhân không? (nhị phân: "yes","no") 9 - contact: phương thức liên lạc (phân loại: "unknown","telephone","cellular") 10 - day: ngày liên lạc cuối cùng của tháng (numeric) 11 - month: tháng liên lạc cuối cùng của năm (phân loại: "jan", "feb", "mar", ..., "nov", "dec") 12 - duration: thời lượng liên lạc cuối cùng, tính bằng giây (numeric) *# Thuộc tính khác:* 13 - campaign: số liên hệ được thực hiện trong chiến dịch này và cho khách hàng này (số, bao gồm liên hệ cuối cùng) 14 - pdays: số ngày trôi qua sau khi khách hàng được liên hệ lần cuối từ một chiến dịch trước đó (numeric, -1 có nghĩa là khách hàng chưa được liên hệ trước đó) 15 - previous: số lượng địa chỉ liên hệ được thực hiện trước chiến dịch này và cho khách hàng này (numeric)

	16 - poutcome: kết quả chiến dịch tiếp thị trước đó (phân loại: "unknown","other","failure","success") Biến đầu ra (mục tiêu mong muốn): 17 – y: khách hàng đã đăng kí kì hạn chưa? (nhị phân: "yes","no")
Thông tin tác giả	[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

Raw data: file bank-full.csv

Tidy data: được lưu lại thành file: bankfull.csv

Instruction list:

CODE R

CODE BOOK Bài 2c

Thông tin	Nội dung
Tên bộ dữ liệu	Car Evaluation Database
Nguồn thu thập và cách thức thu thập	(a) Creator: Marko Bohanec (b) Donors: Marko Bohanec (marko.bohanec@ijs.si) Blaz Zupan (blaz.zupan@ijs.si) (c) Date: June, 1997
Số thuộc tính	6
Thông tin tên các thuộc tính	1. buying v-high, high, med, low 2. maint v-high, high, med, low 3. doors 2, 3, 4, 5-more

	4. persons 2, 4, more 5. lug_boot small, med, big 6. safety low, med, high
Thông tin tác giả	(a) Creator: Marko Bohanec (b) Donors: Marko Bohanec (marko.bohanec@ijs.si) Blaz Zupan (blaz.zupan@ijs.si)

Raw data: file car.data

Tidy data: được lưu lại thành file: car.csv

Instruction list:

CODE R

CODE BOOK Bài 2d

Thông tin	Nội dung
Tên bộ dữ liệu	Wine recognition data
Nguồn thu thập và cách thức thu thập	(a) Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy. (b) Stefan Aeberhard, email: stefan@coral.cs.jcu.edu.au (c) July 1991
Số thuộc tính	13
Thông tin tên các thuộc tính	1) Alcohol 2) Malic acid 3) Ash 4) Alcalinity of ash

	5) Magnesium 6) Total phenols 7) Flavanoids 8) Nonflavanoid phenols 9) Proanthocyanins 10)Color intensity 11)Hue 12)OD280/OD315 of diluted wines 13)Proline
Thông tin tác giả	(a) Creator: Marko Bohanec (b) Donors: Marko Bohanec (marko.bohanec@ijs.si) Blaz Zupan (blaz.zupan@ijs.si)

Raw data: file wine.data

Tidy data: được lưu lại thành file: wine.csv

Instruction list:

CODE R