**ISCS 539 – Data Analytics**

**Presentation, April 25, Monday**

Work on one of the following two tasks and present your work to the class.

1. Choose one of the following built-in data sets in the R package arules: Epub, SunBai, and Mushroom.

Create a histogram for each attribute.

Create a pie plot for each attribute.

Find the frequent itemsets for 3 different values of the support threshold.

Find the association rules for 3 different values of the support threshold and 3 different values of the confidence threshold.

1. Choose one of the following built-in data sets in the R package dplyr: band\_member, starwars, and storms.

Calculate min, max, average, and mode of each attribute.

Create a box-plot for the attributes of the data set.

Place the objects in the data set in clusters using K-means for the number of clusters K = 2, 3, 4 and 5, respectively. Present the result of cluster analysis. Find the optimal number of clusters.

Turn in your presentation slides in Blackboard by 11:59 pm on April 25, Monday.

Each student will present his/her work to the class in five minutes in the order given in the following table.

|  |  |
| --- | --- |
| 1 | Samson O. Ayeni |
| 2 | Dylan J. Craven |
| 3 | Orianna J. Gordon |
| 4 | Ozioma A. Aguegboh |
| 5 | Ileecea Askew |
| 6 | Sumaiya Mahmud |
| 7 | Samiha B. Rafiq |
| 8 | Gowri S. Rajavelu |
| 9 | Bruce E. Taylor, Jr. |
| 10 | Eniola Webster-Esho |
| 11 | Yue Zhang |
| 12 | Kenny W. Fox III |

You should request for my approval asap if you would like to present your work with other data set.

If any student would like to give a presentation on April 22, Friday, feel free to let me know.

**Epub The Epub Transactions Data Set**

Description

The Epub data set contains the download history of documents from the electronic publication platform

of the Vienna University of Economics and Business Administration. The data was recorded

between Jan 2003 and Dec 2008.

Format

Object of class transactions with 15729 transactions and 936 items. Item labels are document IDs of

the form "doc\_11d". Session IDs and time stamps for transactions are also provided as transaction

information.

Source

Provided by Michael Hahsler from the custom information system ePub-WU at https://epub.

wu-wien.ac.at (which has been replaced by eprint).

Examples

data(Epub)

inspect(head(Epub))

**SunBai The SunBai Weighted Transactions Data Set**

Description

A small example database for weighted association rule mining provided as an object of class transactions.

Format

Object of class transactions with 6 transactions and 8 items. Weights are stored as transaction

information.

Details

The data set contains the example database described in the paper by K. Sun and F.Bai for illustration

of the concepts of weighted association rule mining. weight stored as transaction information

denotes the transaction weights obtained using the HITS algorithm.

Source

K. Sun and F. Bai (2008). Mining Weighted Association Rules without Preassigned Weights. IEEE

Transactions on Knowledge and Data Engineering, 4 (30), 489–495.

Examples

data(SunBai)

summary(SunBai)

inspect(SunBai)

**Mushroom The Mushroom Data Set as Transactions**

Description

The Mushroom transactions data set includes descriptions of hypothetical samples corresponding to

23 species of gilled mushrooms in the Agaricus and Lepiota Family.

Format

Object of class transactions with 8124 transactions and 114 items.

Details

The transaction set contains information about 8124 mushrooms (transactions). 4208 (51.8%) are

edible and 3916 (48.2%) are poisonous. The data contains 22 nominal features plus the class attribute

(edible or not). These features were translated into 114 items.

Source

The data set was obtained from the UCI Machine Learning Repository at https://archive.ics.

uci.edu/ml/datasets/Mushroom.

References

Alfred A. Knopf (1981). Mushroom records drawn from The Audubon Society Field Guide to

North American Mushrooms. G. H. Lincoff (Pres.), New York.

**band\_members Band membership**

Description

These data sets describe band members of the Beatles and Rolling Stones. They are toy data sets

that can be displayed in their entirety on a slide (e.g. to demonstrate a join).

Usage

band\_members

band\_instruments

band\_instruments2

Format

Each is a tibble with two variables and three observations

Details

band\_instruments and band\_instruments2 contain the same data but use different column names

for the first column of the data set. band\_instruments uses name, which matches the

**starwars Starwars characters**

Description

The original data, from SWAPI, the Star Wars API, https://swapi.dev/, has been revised to

reflect additional research into gender and sex determinations of characters.

Usage

starwars

Format

A tibble with 87 rows and 14 variables:

name Name of the character

height Height (cm)

mass Weight (kg)

hair\_color,skin\_color,eye\_color Hair, skin, and eye colors

birth\_year Year born (BBY = Before Battle of Yavin)

sex The biological sex of the character, namely male, female, hermaphroditic, or none (as in the

case for Droids).

gender The gender role or gender identity of the character as determined by their personality or

the way they were programmed (as in the case for Droids).

homeworld Name of homeworld

species Name of species

films List of films the character appeared in

vehicles List of vehicles the character has piloted

starships List of starships the character has piloted

**storms Storm tracks data**

Description

This data is a subset of the NOAA Atlantic hurricane database best track data, https://www.nhc.

noaa.gov/data/#hurdat. The data includes the positions and attributes of storms from 1975-2020,

measured every six hours during the lifetime of a storm.

Usage

storms

Format

A tibble with 11,859 observations and 13 variables:

name Storm Name

year,month,day Date of report

hour Hour of report (in UTC)

lat,long Location of storm center

status Storm classification (Tropical Depression, Tropical Storm, or Hurricane)

category Saffir-Simpson storm category (estimated from wind speed. -1 = Tropical Depression, 0

= Tropical Storm)

wind storm’s maximum sustained wind speed (in knots)

pressure Air pressure at the storm’s center (in millibars)

tropicalstorm\_force\_diameter Diameter (in nautical miles) of the area experiencing tropical storm

strength winds (34 knots or above)

hurricane\_force\_diameter Diameter (in nautical miles) of the area experiencing hurricane strength

winds (64 knots or above)

See Also

The script to create the storms data set: https://github.com/tidyverse/dplyr/blob/main/data-raw/storms.R