

Data Mining

Lab sheet N°3: Exploratory Data Analysis (EDA)

ENSIA 2023-2024

Part 2: Exercises on the Chapter Data - Part 1

Exercise 1:

Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio).

Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

Example: Age in years. Answer: discrete, quantitative, ratio

- A. Time in terms of AM or PM. (binary, qualitative, ordinal)
- B. Brightness as measured by a light meter. (continuous, quantitative, ratio)
- C. Brightness as measured by people's judgments. (discrete, qualitative, ordinal)
- D. Angles as measured in degrees between 0 and 360. (continuous, quantitative, ratio)
- E. Bronze, Silver, and Gold medals as awarded at the Olympics. (discrete, qualitative, ordinal)
- F. Height above sea level. (continuous, quantitative, interval/ratio)
- G. Number of patients in a hospital. (discrete, quantitative, ratio)
- H. ISBN numbers for books. (Look up the format on the Web.) (discrete, qualitative, nominal)
- I. Ability to pass light (opaque, translucent, transparent) (discrete, qualitative, ordinal)
- J. Military rank. (discrete, qualitative, ordinal)
- K. Distance from the center of campus. (continuous, quantitative, interval/ratio)

Exercise 2:

For the following vectors, X and Y, calculate the indicated similarity or distance measures.

- A. $x = (1, 1, 1, 1)$, $y = (2, 2, 2, 2)$ cosine, correlation, Euclidean
- B. $x = (0, 1, 0, 1)$, $y = (1, 0, 1, 0)$ cosine, correlation, Euclidean, Jaccard
- C. $x = (0, -1, 0, 1)$, $y = (1, 0, -1, 0)$ cosine, correlation, Euclidean
- D. $x = (1, 1, 0, 1, 0, 1)$, $y = (1, 1, 1, 0, 0, 1)$ cosine, correlation, Jaccard
- E. $x = (2, -1, 0, 2, 0, -3)$, $y = (-1, 1, -1, 0, 0, -1)$ cosine, correlation

	Cosine	Correlation	Euclidean	Jaccard
A	1	0/0 (undefined)	2	/
B	0	-1	2	0
C	0	0	2	/
D	0.75	0.25	/	0.6
E	0	0	/	/