

# **Analiza asocjacji**

Marzena Nowakowska

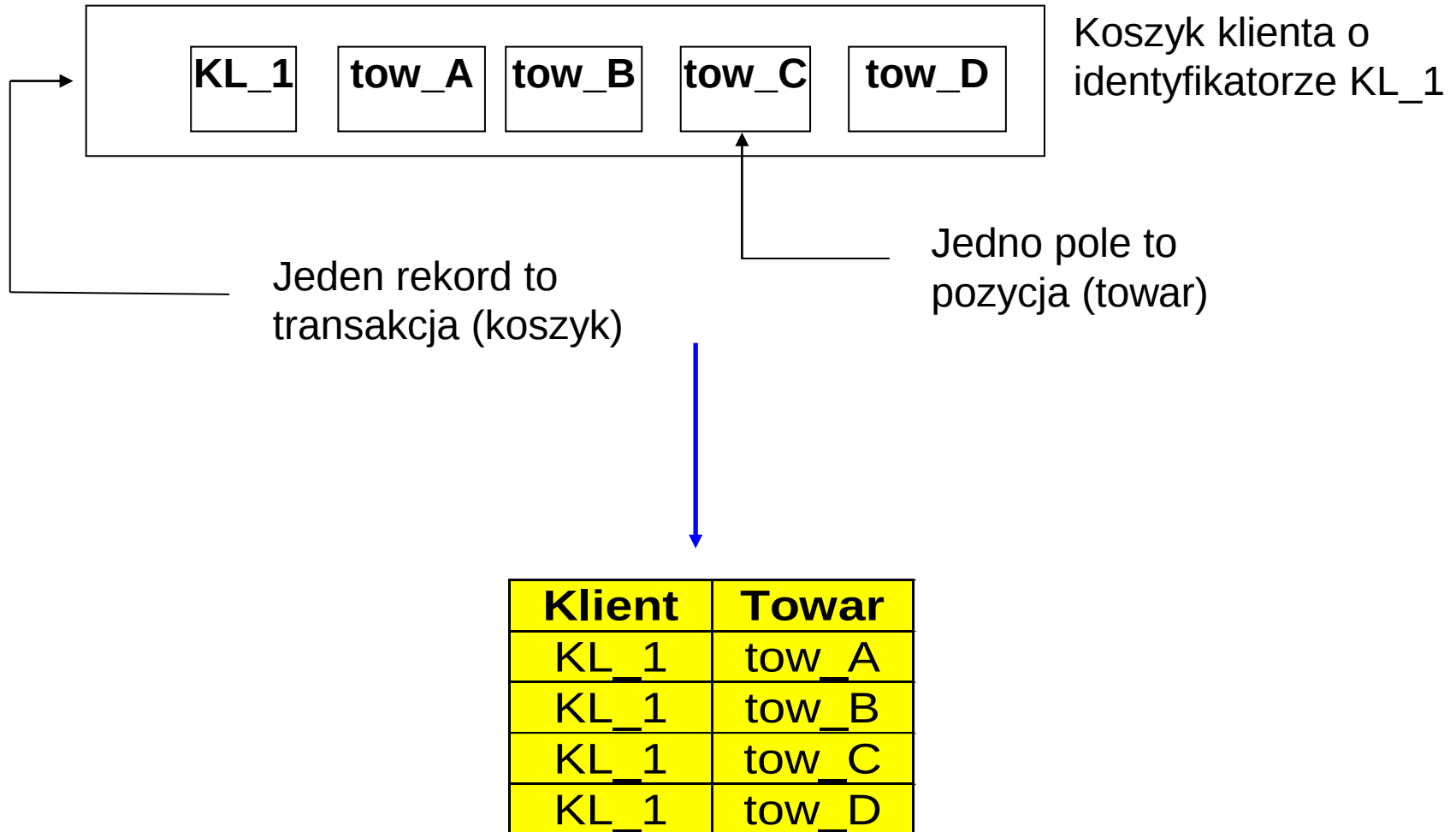
Wydział Zarządzania i Modelowania Komputerowego  
Politechnika Świętokrzyska

# Zadania analizy asocjacji

Celem analizy asocjacji (związków ) jest **wykrywanie współwystępowania pewnych wartości atrybutów w zbiorze danych**

**Reguły asocjacyjne** reprezentują wiedzę, że pewne wartości atrybutów zazwyczaj występują łącznie z pewnymi wartościami innych atrybutów

# Struktura zbioru dla analizy asocjacji



# Reguła asocjacyjna

*Lista\_wartości\_warunkujących => Lista\_wartości\_warunkowanych*

**Jeśli A jest częścią jakiegoś zdarzenia,  
to B jest również jego częścią:**

$$A \Rightarrow B$$

Przykładowa reguła analizy koszykowej:

*Klient kupuje towar A => Klient kupuje towar B*

# Jakość reguł asocjacyjnych

## Wsparcie

Niech będzie dana reguła  $A \Rightarrow B$

**Wsparcie reguły** (*support*) - prawdopodobieństwo wspólnego wystąpienia zdarzeń A i B - prawdopodobieństwo jednoczesnego zakupu przez klienta towarów A i B:

$$\text{Support}(A \Rightarrow B) = P(A \cap B)$$

Wskaźnik symetryczny:  $\text{Support}(A \Rightarrow B) = \text{Support}(B \Rightarrow A)$

*Częstość „koszyków”, w których występuje jednocześnie A i B.*

**Interpretacja:** im większe wsparcie tym silniejsza reguła; reguły o bardzo niskim wsparciu są zwykle odrzucane, bo są zbyt rzadkie.

# Jakość reguł asocjacyjnych

## Wiarygodność

Niech będzie dana reguła  $A \Rightarrow B$

**Ufność reguły** (*confidence*, pewność, wiarygodność) - prawdopodobieństwo warunkowe zajścia zdarzenia B pod warunkiem, że zaszło zdarzenia A - prawdopodobieństwo zakupu przez klienta towaru B, jeśli zakupił towar A: **Confidence** ( $A \Rightarrow B$ ) =  **$P(B/A)$**

Wskaźnik niesymetryczny:

$$\text{Confidence}(A \Rightarrow B) \neq \text{Confidence}(B \Rightarrow A)$$

*Częstość „koszyków” z B pod warunkiem, że jest tam też A.*

**Interpretacja:** im wyższa wiarygodność, tym bardziej silniejsza reguła.

# Jakość reguł asocjacyjnych

## Siła asocjacji

Niech będzie dana reguła  $A \Rightarrow B$

**Siła asocjacji reguły** (*lift*) - iloraz wiarygodności (*confidence*) oraz oczekiwanego wsparcia (*expected support*) B, czyli prawdopodobieństwa zajścia zdarzenia B. Wskaźnik siły asocjacji informuje ile razy częściej kupują towar B klienci, którzy kupili towar A w porównaniu z klientami wybranymi losowo:

$$\text{Lift}(A \Rightarrow B) = P(B/A) / P(B)$$

Wskaźnik symetryczny:  $\text{Lift}(A \Rightarrow B) = \text{Lift}(B \Rightarrow A)$ .

### Interpretacja

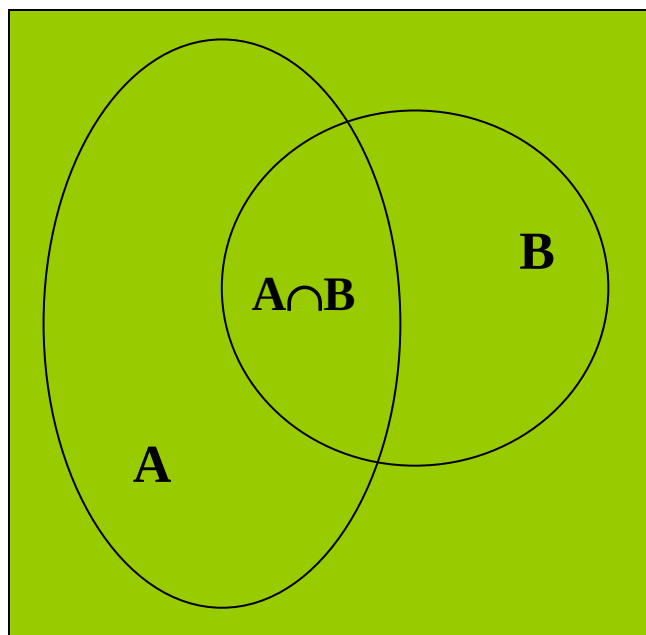
Lift = 1 – brak związku między A i B (niezależność)

Lift > 1 – pozytywna asocjacja (A sprzyja B)

Lift < 1 – negatywna asocjacja (A zniechęca B)

# Ilustracja miar siły związków w analizie asocjacji

Niech będzie dana reguła  $A \Rightarrow B$



$n$  – całkowita liczba przypadków (transakcji)

$n_A$  – liczba przypadków zawierających wartość A

$n_B$  – liczba przypadków zawierających wartość B

$n_{A \cap B}$  – liczba przypadków zawierających jednocześnie wartości A i B

$$\text{Support} = P(A \cap B) \approx n_{A \cap B} / n$$

$$\text{Confidence} = P(B/A) \approx n_{A \cap B} / n_A$$

$$\text{Expected confidence} = P(B) \approx n_B / n$$

$$\begin{aligned} \text{Lift} &= P(B/A) / P(B) = P(A \cap B) / (P(A) \cdot P(B)) \\ &\approx n_{A \cap B} \cdot n / (n_A \cdot n_B) \end{aligned}$$

# Obliczanie reguł asocjacyjnych

|         | Towar B | Towar D | Razem |
|---------|---------|---------|-------|
| Towar A | 5000    | 1000    | 6000  |
| Towar C | 3500    | 500     | 4000  |
| Razem   | 8500    | 1500    | 10000 |

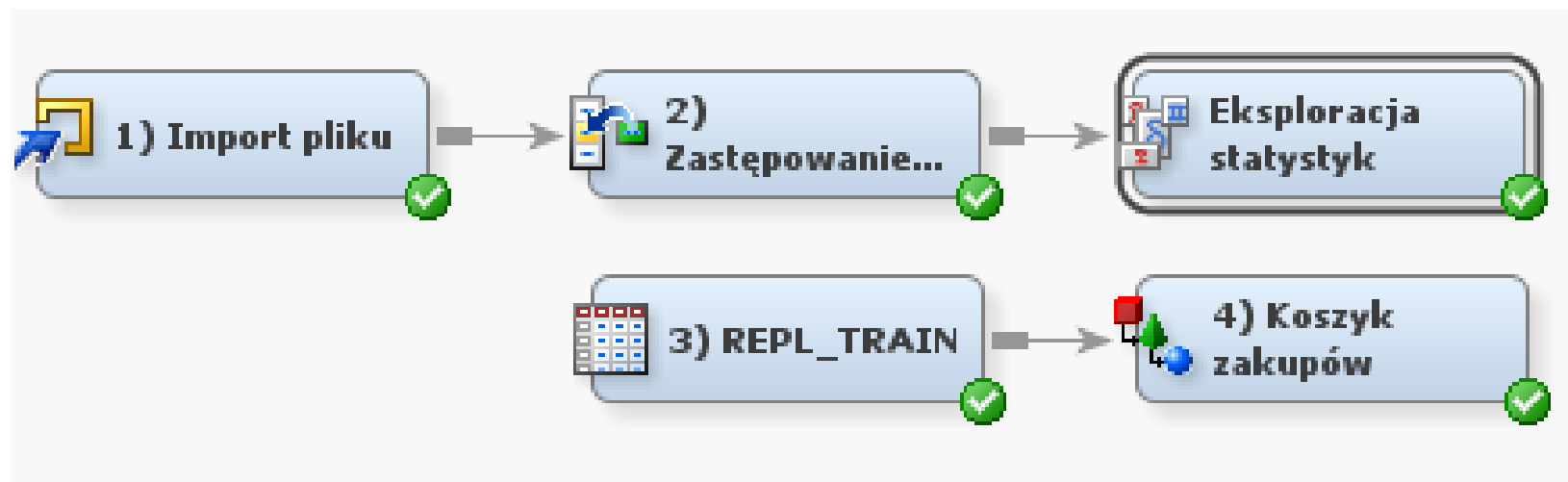
- **Wsparcie** =  $P(A \cap B) = N_{A \cap B} / N = 5000/10000 = 50\%$ .
  - **Wiarygodność** (uwaga:  $\#A < \#B$ ):
    - dla reguły  $A \Rightarrow B$  wiarygodność jest równa  $P(B/A) = N_{A \cap B} / N_A = 5000/6000 = 83\%$
    - dla reguły  $B \Rightarrow A$  wiarygodność jest równa  $P(A/B) = N_{A \cap B} / N_B = 5000/8500 = 58,8\%$
- $P(B/A) = 83\%$                       ale                       $P(B/\sim A) = 3500/4000 = 87,5\%$
- **Siła asocjacji**: dla reguły  $A \Rightarrow B$  jest równa  $P(B/A) / P(B) = 83\%/85\% = 0,98$ . Taka ułamkowa wartość (mniejsza od 1) wskazuje na korelację negatywną.

# Analiza koszykowa danych o zakupach klientów sklepu spożywczego

Duży sklep spożywczy zamawia analizę danych dotyczących zarejestrowanych zakupów swoich klientów w celu wykrycia zależności jakie grupy produktów łącznie klient zazwyczaj kupuje. W ten sposób sklep chce określić preferencje zakupowe swoich klientów. Jednym z celów badań jest uzyskanie informacji jak zaplanować ekspozycję towarów, aby sprowokować klienta do tzw. transakcji wiązanej. Dzięki temu istnieje szansa na zwiększenie obrotów sklepu.

W celu określenia preferencji zakupowych klienta należy przeprowadzić analizę koszykową danych o zakupach klientów sklepu na podstawie informacji zawartej w pliku formatu SAS o nazwie *assocs.sas7bdat*.

# Diagram procesu dla analizy asocjacji



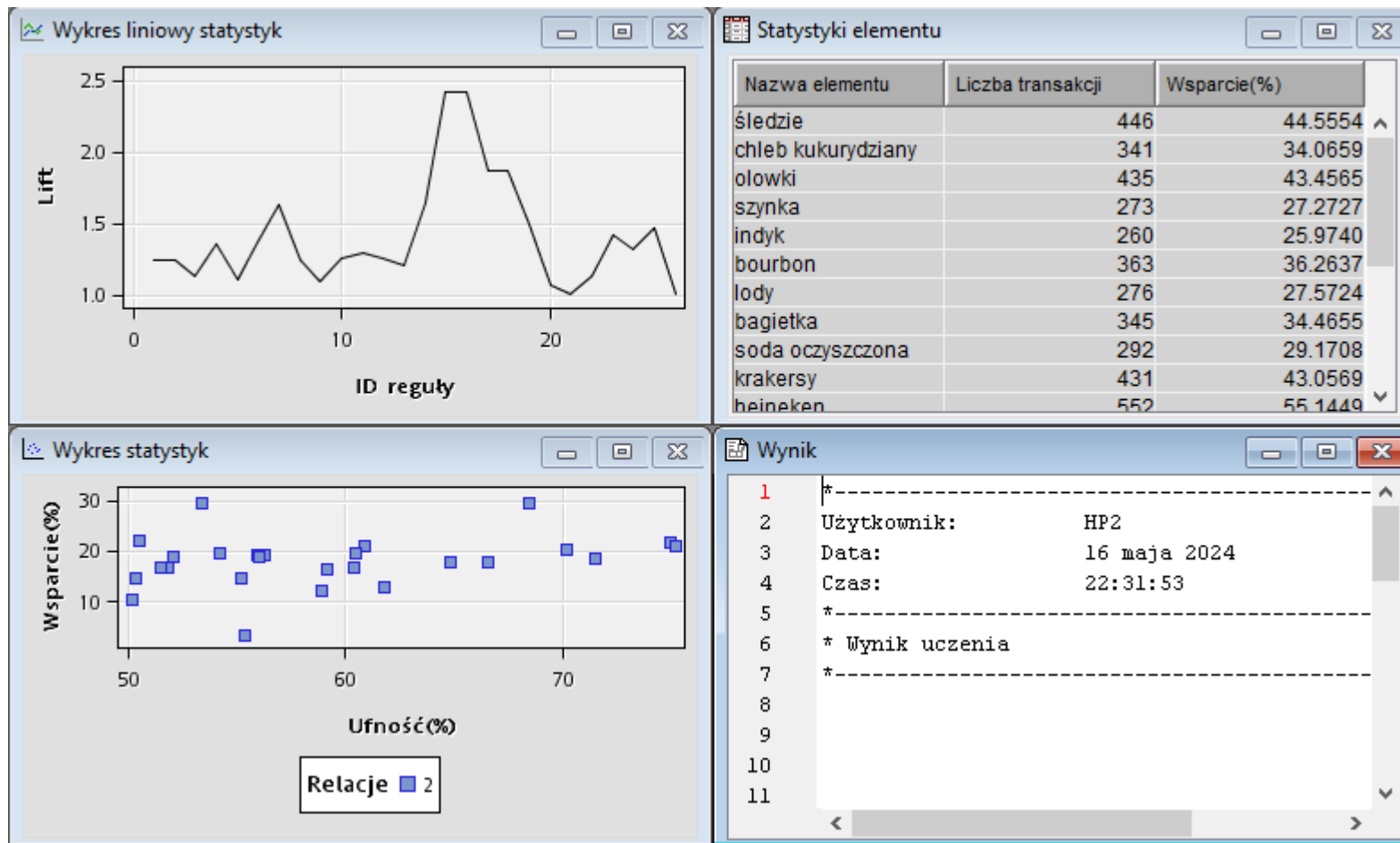
# Parametry węzła *Koszyk zakupów*

| Importowane dane - 4) Koszyk zakupów |        |                 |              |               |
|--------------------------------------|--------|-----------------|--------------|---------------|
| Port                                 | Źródło | Tabela          | Rola         | Dane istnieją |
| DATA                                 |        |                 | Uczące       | Nie           |
| VALIDATE                             |        |                 |              | Nie           |
| TEST                                 |        |                 |              | Nie           |
| SCORE                                |        |                 |              | Nie           |
| TRANSACTION                          |        | EMWS1.Ids4_DATA | Transakcyjne | Tak           |
| RULES                                |        |                 |              | Nie           |

| Więzy                    |         |
|--------------------------|---------|
| Maksymalnie elementów    | 2       |
| Minimalny poziom ufności | 50      |
| Minimalny lift           | .       |
| Minimalny lift wsparcia  | .       |
| Typ wsparcia             | Procent |
| Liczba wspierających     | 5       |
| Procent wspierających    | 2.0     |

1. Maksymalna liczba elementów w regule
2. Minimalny poziom wiarygodności dla wygenerowania reguły
3. Minimalna wartość Lift
4. Minimalny poziom wsparcia dla asocjacji

# Wyniki analizy asocjacji



# Statystyki reguł asocjacyjnych

Tabela: Wykres liniowy statystyk

| ID reguły | Wielkość<br>LHS reguły | Wielkość<br>RHS reguły | Liczba<br>transakcji | Wsparcie(%)<br>) | Ufność(%) | Lift   | Element 1     | Element 2 | Reguła          | Relacje |
|-----------|------------------------|------------------------|----------------------|------------------|-----------|--------|---------------|-----------|-----------------|---------|
| 1         | 1                      | 1                      | 295                  | 29.4705          | 53.4420   | 1.2412 | heineken      | krakersy  | heineken =...   | 29      |
| 2         | 1                      | 1                      | 295                  | 29.4705          | 68.4455   | 1.2412 | krakersy      | heineken  | krakersy ==...  | 29      |
| 3         | 1                      | 1                      | 220                  | 21.9780          | 50.5747   | 1.1351 | olowki        | śledzie   | olowki ==> ...  | 29      |
| 4         | 1                      | 1                      | 219                  | 21.8781          | 75.0000   | 1.3601 | soda oczys... | heineken  | soda oczys...   | 29      |
| 5         | 1                      | 1                      | 210                  | 20.9790          | 60.8696   | 1.1038 | bagietka      | heineken  | bagietka ==...  | 29      |
| 6         | 1                      | 1                      | 209                  | 20.8791          | 75.1799   | 1.3633 | karczochoy    | heineken  | karczochoy =... | 29      |
| 7         | 1                      | 1                      | 205                  | 20.4795          | 70.2055   | 1.6305 | soda oczys... | krakersy  | soda oczys...   | 29      |
| 8         | 1                      | 1                      | 197                  | 19.6803          | 54.2700   | 1.2488 | bourbon       | olowki    | bourbon ==...   | 29      |

# Interpretacja miar dla reguły o największym wsparciu

Wsparcie reguły ( $\text{Support} (Krakersy \rightarrow Piwo) = 37\%$ ) informuje, że blisko 37% klientów zakupiło jednocześnie krakersy i piwo heineken (reguła 1); liczba transakcji jest równa 366 na 1001.

Wśród klientów, którzy kupili krakersy 75% kupiło również piwo heineken ( $\text{Confidence} (Krakersy \rightarrow Piwo) = 75\%$  - reguła 1), a wśród klientów, którzy kupili piwo tylko 61% kupiło krakersy ( $\text{Confidence} (Piwo \rightarrow Krakersy) = 61\%$  reguła 2). Może to sugerować, że zakup krakersów w większym stopniu prowokuje również zakup piwa niż odwrotnie.

Wartość siły asocjacji ( $Lift$ ) dla reguł 1 i 2 równa 1,25 informuje, że klienci którzy kupili krakersy (piwo) 1.25 razy częściej kupują również piwo (krakersy) niż klienci wybrania losowo (klienci przypadkowi, o których nie wiadomo co innego jeszcze kupują).