

Ćwiczenie: Prosty projekt tworzący standardowe raporty i podsumowania

Plik danych *Adstudy.sta* zawiera:

- nazwiska (w kolumnie z nazwami przypadków) osób oceniających skuteczność reklamy napoju orzeźwiających
- w kolumnie *płeć* odnotowana jest płeć osoby
- kolumna *Reklama* zawiera nazwę ocenianego napoju *Coke* lub *Pepsi*
- w zmiennych od *Pomiar1* do *Pomiar23* zawarta jest ocena (w skali od 1 do 9; 0 oznacza brak zdania) skuteczności reklamy jako np. ilość zakupów (z ostatnich 23 miesięcy), przeprowadzonej przez pewną firmę.

Cel analiz: utworzenie projektu data mining, który umożliwi szybkie wskazanie:

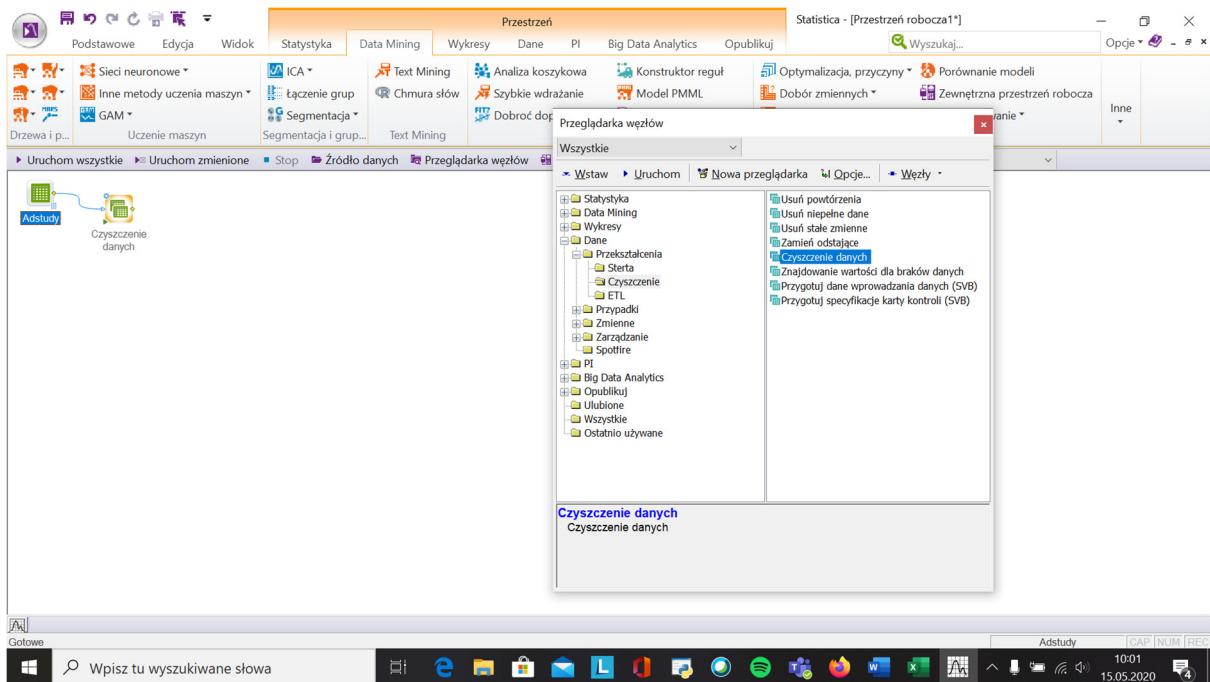
- osób najwyżej oceniających reklamę oraz
- skuteczności reklamy z okresu 3 miesiące i 6 miesięcy (i porównanie czy zmieniała się w czasie), z uwzględnieniem podziału na dwa badane produkty *Coke* i *Pepsi*

1. Otwórz Plik danych *Adstudy.sta*

The screenshot shows the Statistica software interface with the title bar "Statistica - [Dane: Adstudy (25 zmn. * 50 prz.)]". The menu bar includes "Podstawowe", "Edycja", "Widok", "Format", "Statystyka", "Data Mining", "Wykresy", "Narzędzia", and "Dane". The "Dane" tab is selected. The toolbar below the menu includes icons for "Nowy", "Otwórz", "Zapisz", "Projekt", "Dodaj do skoroszytu", "Dodaj do raportu", "Dodaj do MS Word", "Wstaw do przestrzeni", "Analizy", "Makro", "Integracja", "Opcje", "Wyewidencjonuj", "Sąsiadujące w poziomie", "Zamknij wszystkie", "Zaewidencjonuj", "Sąsiadujące w pionie", "Uporządkuj ikony", "Odrzuć", "SharePoint", "Kaskada", "Okna", and "Wyszukaj...". The main window displays a table titled "Badanie efektywności reklamy." with 18 columns labeled 1 through 18. The first column is "Odpowiedzi" and the second is "Płeć". The table contains data for 25 individuals, each with a name, gender, product preference, and 23 measurement values. The names listed include J. Baker, A. Smith, M. Brown, C. Mayer, M. West, D. Young, S. Bird, D. Flynd, J. Owen, H. Morrow, F. East, C. Clint, I. Neil, G. Boss, K. Record, T. Bush, P. Squire, L. Mynard, E. Bynum, M. Quick, J. Liu, M. Good, B. Quale, U. Andy, T. Hill, and T. Patere.

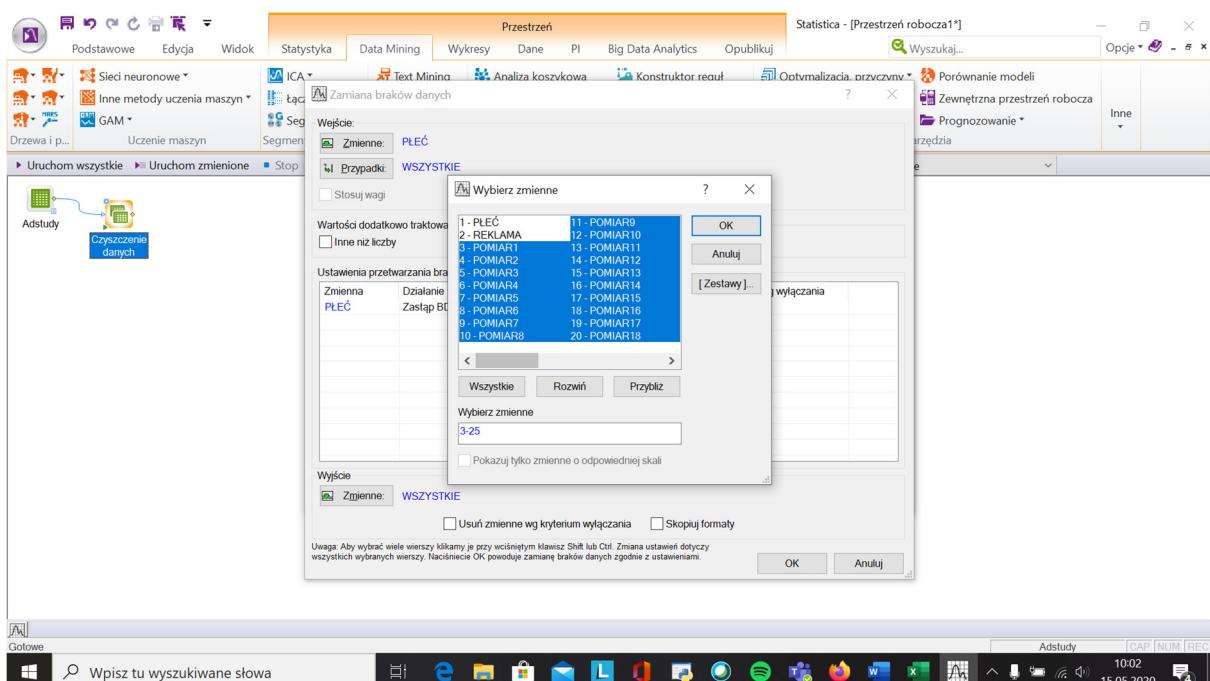
Odpowiedzi	Płeć	Badanie efektywności reklamy.															
		1 REKLAMA	2 POMIAR	3 POMIAR	4 POMIAR	5 POMIAR	6 POMIAR	7 POMIAR	8 POMIAR	9 POMIAR	10 POMIAR	11 POMIAR	12 POMIAR	13 POMIAR	14 POMIAR	15 POMIAR	16 POMIAR
J. Baker	Mężczyzna	PEPSI	9	1	6	8	1	2	1	3	8	2	6	7	0	3	0
A. Smith	Mężczyzna	COKE	6	7	1	8	0	0	6	8	9	9	4	9	5	0	0
M. Brown	Kobieta	COKE	9	8	2	9	8	8	0	3	2	2	1	5	7	4	
C. Mayer	Mężczyzna	PEPSI	7	9	0	5	9	9	6	6	0	7	0	3	7	3	2
M. West	Mężczyzna	PEPSI	7	1	6	2	8	9	6	4	0	7	9	6	3	3	7
D. Young	Kobieta	COKE	6	0	0	8	3	1	0	7	8	7	8	2	9	0	2
S. Bird	Kobieta	COKE	7	4	3	2	5	7	1	6	3	9	4	3	4	0	2
D. Flynd	Mężczyzna	PEPSI	9	9	2	6	6	8	7	5	2	4	7	7	7	6	
J. Owen	Kobieta	PEPSI	7	8	2	3	6	9	1	3	2	9	8	9	0	3	4
H. Morrow	Mężczyzna	PEPSI	6	6	2	8	3	6	4	2	3	0	9	7	4	7	5
F. East	Kobieta	PEPSI	4	6	6	5	6	8	7	2	5	5	8	4	5	4	6
C. Clint	Mężczyzna	COKE	7	3	3	7	0	6	5	0	4	1	3	3	9	1	0
I. Neil	Mężczyzna	PEPSI	6	2	3	1	8	1	4	2	6	2	9	9	2	5	4
G. Boss	Mężczyzna	COKE	7	2	4	8	1	2	6	8	9	4	8	9	5	5	3
K. Record	Kobieta	PEPSI	6	2	7	5	7	2	5	0	6	5	5	6	4	2	8
T. Bush	Kobieta	PEPSI	3	2	5	4	4	4	3	7	6	5	0	7	3	9	0
P. Squire	Mężczyzna	COKE	2	9	9	3	1	4	5	5	4	3	6	2	6	1	8
L. Mynard	Kobieta	PEPSI	1	0	7	5	2	4	2	6	9	2	2	4	0	8	5
E. Bynum	Kobieta	COKE	0	6	2	3	2	4	9	2	6	7	0	3	8	0	3
M. Quick	Mężczyzna	COKE	6	8	1	9	5	7	8	9	5	8	3	6	4	1	2
J. Liu	Mężczyzna	PEPSI	9	2	7	0	2	4	9	0	9	9	2	0	5	0	
M. Good	Kobieta	PEPSI	7	0	1	8	5	2	6	7	7	1	4	1	0	9	3
B. Quale	Mężczyzna	PEPSI	5	7	8	8	6	8	6	0	2	0	6	0	3	0	5
U. Andy	Kobieta	COKE	4	4	7	4	7	3	1	1	3	7	9	0	5	7	8
T. Hill	Mężczyzna	COKE	7	0	6	8	5	8	7	1	5	4	5	1	4	7	3
T. Patere	Mężczyzna	COKE	6	8	9	3	0	3	4	6	9	8	8	7	0	n	

- Otwórz przestrzeń roboczą i wybierz to źródło danych
- Otwórz przeglądarkę węzłów
- Wstaw węzeł Dane->Przekształcenia->Czyszczenie -> *Czyszczenie danych*

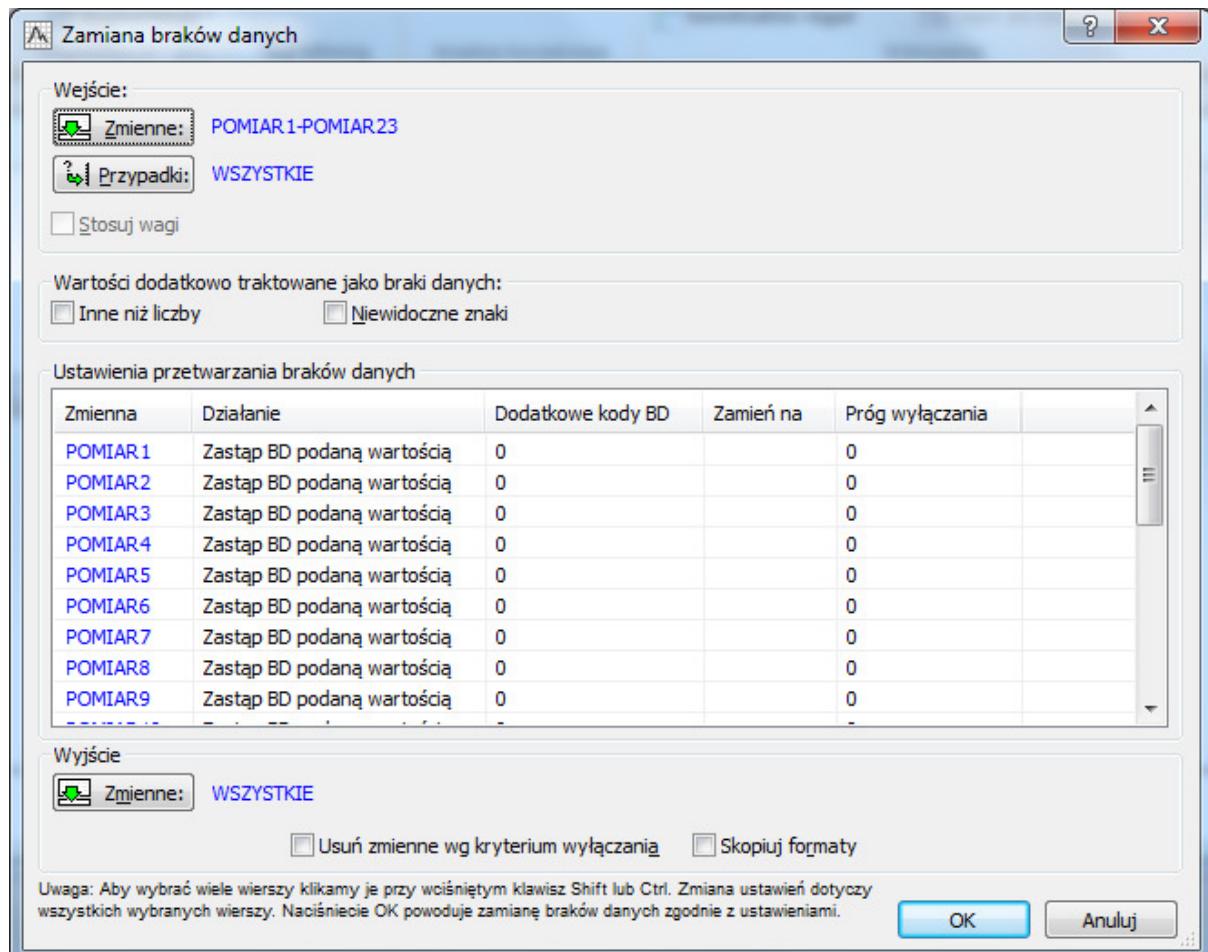


5. Ustaw parametry węzła:

- wybierz zmienne na wejściu (Edytuj parametry): POMIAR1-POMIAR23



- na wyjściu: WSZYSTKIE
- Wszędzie tam, gdzie w pomiarze jest w arkusz 0, należy je zastąpić brakiem danych, żeby nie liczył się do średniej:
- „Ustawienia przetwarzania braków danych” wybierz wszystkie zmienne, w „Działanie” wybierz „Zastąp BD podaną wartością” i w „Dodatkowe kody BD” wpisz wszędzie „0”



6. Uruchom węzeł i otwórz dokument:

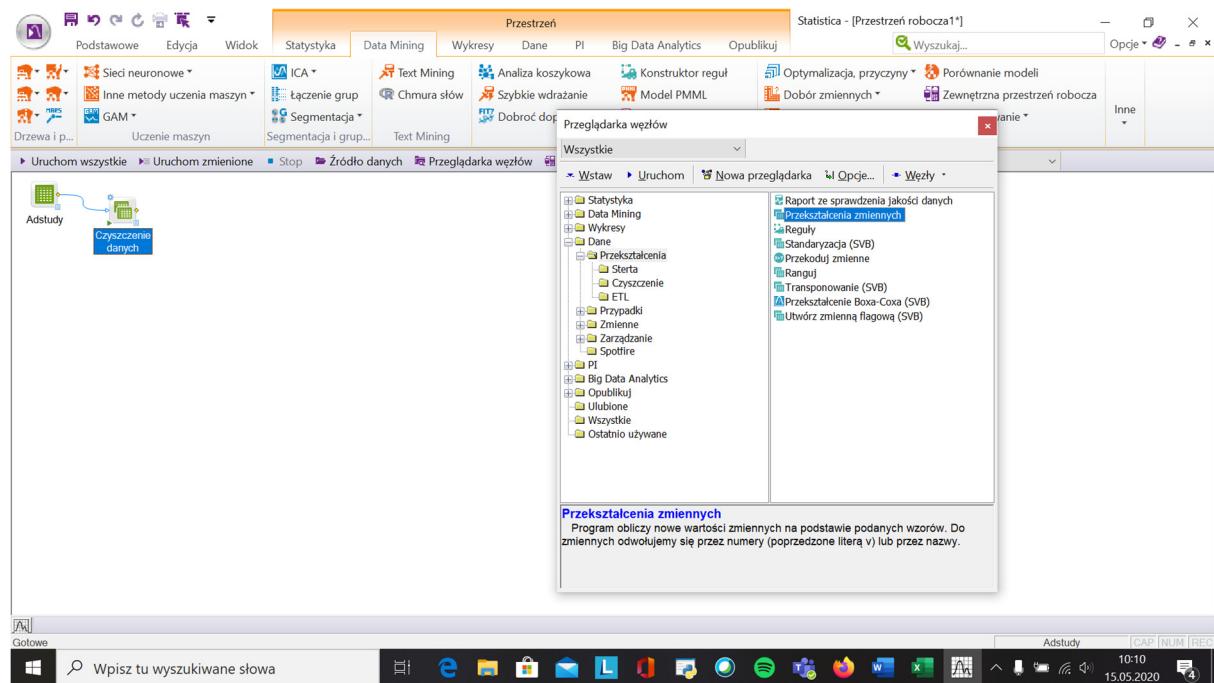
7. Po sprawdzeniu powstałego dokumentu widzimy, że wszystkie zera z pliku z danymi zostały usunięte i zastąpione brakami danych.

8. Teraz możemy przystąpić do wyliczenia średnich.

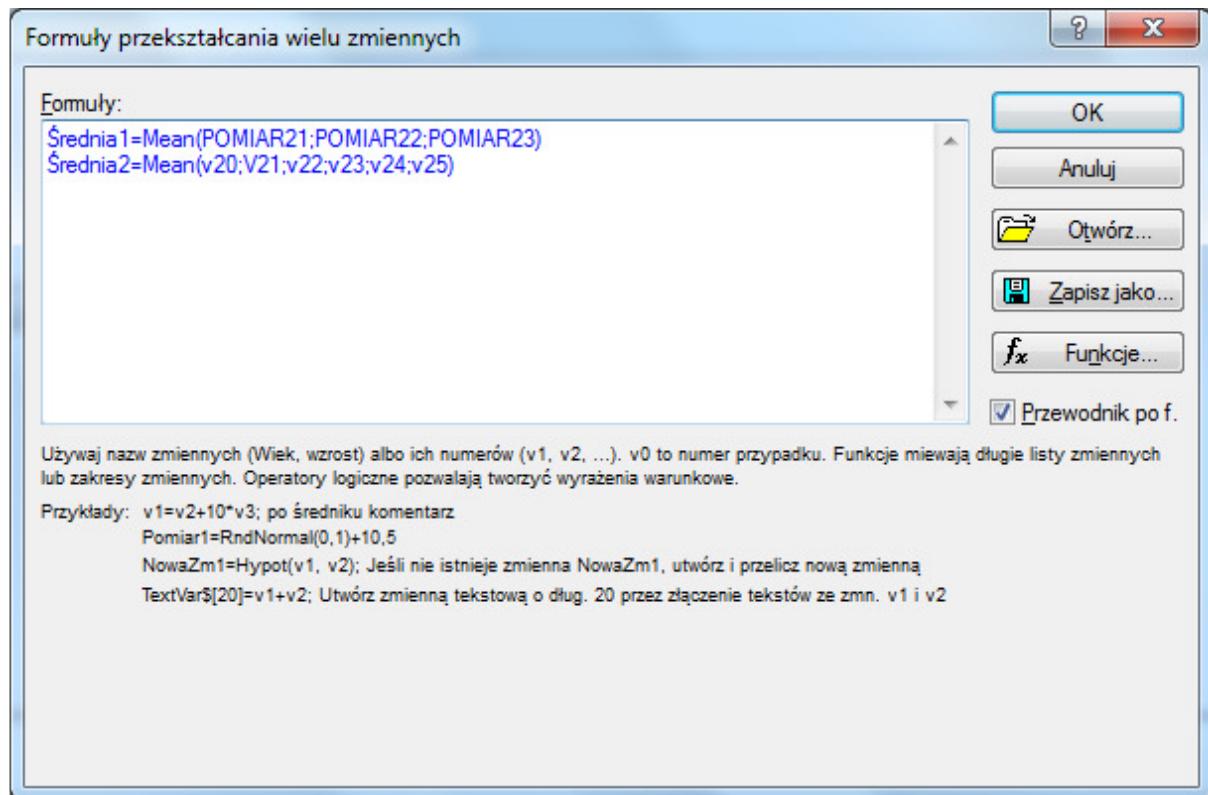
Transformacja danych.

Obliczmy średnią ocenę w ostatnich trzech oraz sześciu miesiącach, tzn. średnią z wartości zmiennych od *POMIAR21* (zmienna 23) do *POMIAR23* (zmienna 25) oraz od *POMIAR18* (zmienna 20) do *POMIAR23* (zmienna 25).

1. Wstaw węzeł Dane -> Przekształcenia -> *Przekształcenia zmiennych*



2. Wpisz przekształcenia (wybierz Edytuj parametry w tym węźle). Można wpisywać nazwy zmiennych, czyli POMIAR21, albo numery zmiennych, czyli v20:



3. Potwierdź dodanie nowych zmiennych do arkusza:

4. Uruchom węzeł i sprawdź, że dodały się nowe zmienne:

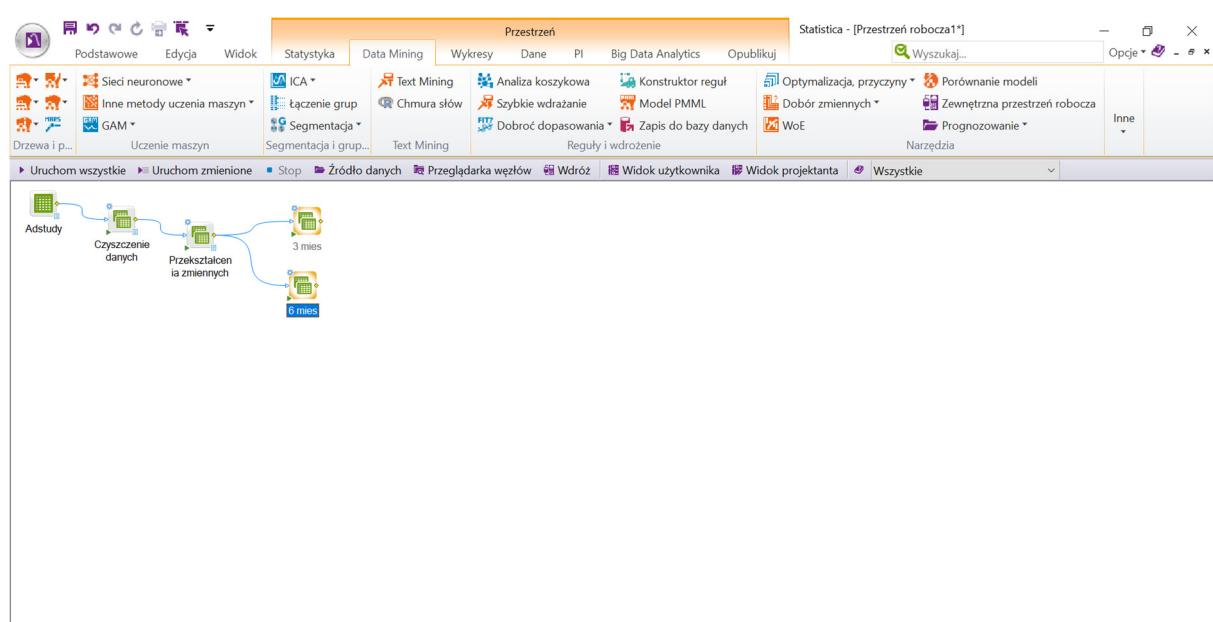
Statistica - [Dane: Przekształcenia zmiennych* (27 zmn. * 50 prz.)]

Odpowiedzi	Badanie efektywności reklamy.																													
	15 POMIAR13	16 POMIAR14	17 POMIAR15	18 POMIAR16	19 POMIAR17	20 POMIAR18	21 POMIAR19	22 POMIAR20	23 POMIAR21	24 POMIAR22	25 POMIAR23	26 Średnia1	27 Średnia2																	
J. Baker	3	5	1	4	1	8	2	3	3	3	7	7,5	6,75																	
A. Smith	9	5	1	3	6	4	6	2	1	8	4	8	7	6	4	3,4	3,4	3,4	3,4	3,4	3,4	3,4	3,4	3,4	3,4	3,4				
M. Brown	5	7	4	6	4	7	4	2	1	8	1	7	7	6	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333			
C. Mayer	7	3	2	8	1	8	1	2	1	7	4	4	7	9	9	9,333333333	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667		
M. West	3	3	7	6	6	7	4	2	1	7	4	4	7	9	9	9,333333333	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667	5,166666667		
D. Young	9		2	7	8	3	7	2	1	7	4	4	7	9	9	9,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333		
S. Bird	4		2	9	2	7	2	1	2	7	4	4	7	9	9	9	9,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	
D. Flynn	7	7	6	9	9	7	2	8	8	7	2	8	7	8	8	8,5	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	
J. Owen	3	4	8	7	3	3	3	3	3	9	3	3	9	3	8	5,5	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	5,333333333	
H. Morrow	4	7	5	1	4	7	5	5	9	3	5	5	9	4	4	4,5	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	
F. East	4	6		5	9	7	2	8	3	2	8	3	2	8	3	5	5	5	5	5	5	5	5	5	5	5	5			
C. Clint	9	1		1	7	7	2	8	3	4	7	6	7	6	8	6	6	6	6	6	6	6	6	6	6	6	6			
I. Neil	2	5	4	4	4	3	6	7	6	7	6	7	6	7	6	6	6	6	6	6	6	6	6	6	6	6	6			
G. Boss	5	5	3	1	4	8	1	7	6	6	7	1	7	6	6	6	5	5	5	5	5	5	5	5	5	5	5			
K. Record	4	2	8	5	6	4	1	7	6	7	3	4	6	6	1	3,3	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	
T. Bush	3	9		2	6	6	3	4	6	7	3	5	5	5	5	3	4	4	4	4	4	4	4	4	4	4	4			
P. Squire	6	1	8	4	9	6	7	2	2	2	2	2	1	1	4	5	5,333333333	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	
L. Mynard	8	3	9	7	5	7	2	2	2	2	2	2	1	1	4	5	5,333333333	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	5,666666667	
E. Bynum	8		3	9	9	9	8	9	9	9	9	9	9	9	9	9	5	5	5	5	5	5	5	5	5	5	5			
M. Quick	4	1	2	9	7	6	3	2	9	7	6	3	2	9	6	6	6	6	6	6	6	6	6	6	6	6	6			
J. Liu	5		2	9	2	9	3	2	9	3	2	9	5	9	4	7,6	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	6,666666667	
M. Good	9		3	9	2	9	2	8	8	2	8	8	4	4	5	5	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	1,333333333	
B. Quale	3		5	6	3	3	3	3	6	8	8	7	7	7	8	8,7	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	3,333333333	
U. Andy	5	7	8	3	7	7	7	7	7	4	4	4	6	6	5	5	7	7	6	6	5	5	5	5	5	5	5			
T. Hill	4	7	3	9	9	9	7	7	7	5	5	5	1	1	8	8	4,5	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25	5,25
T. Peters	7		8	5	3	2	9	2	9	9	9	9	9	9	9	9	1	1	5	5	5	5	5	5	5	5	5	5		
G. Honan	7	1	9	9	1	1	8	3	8	5	1	1	6	6	6	6	4,5	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667	1,666666667

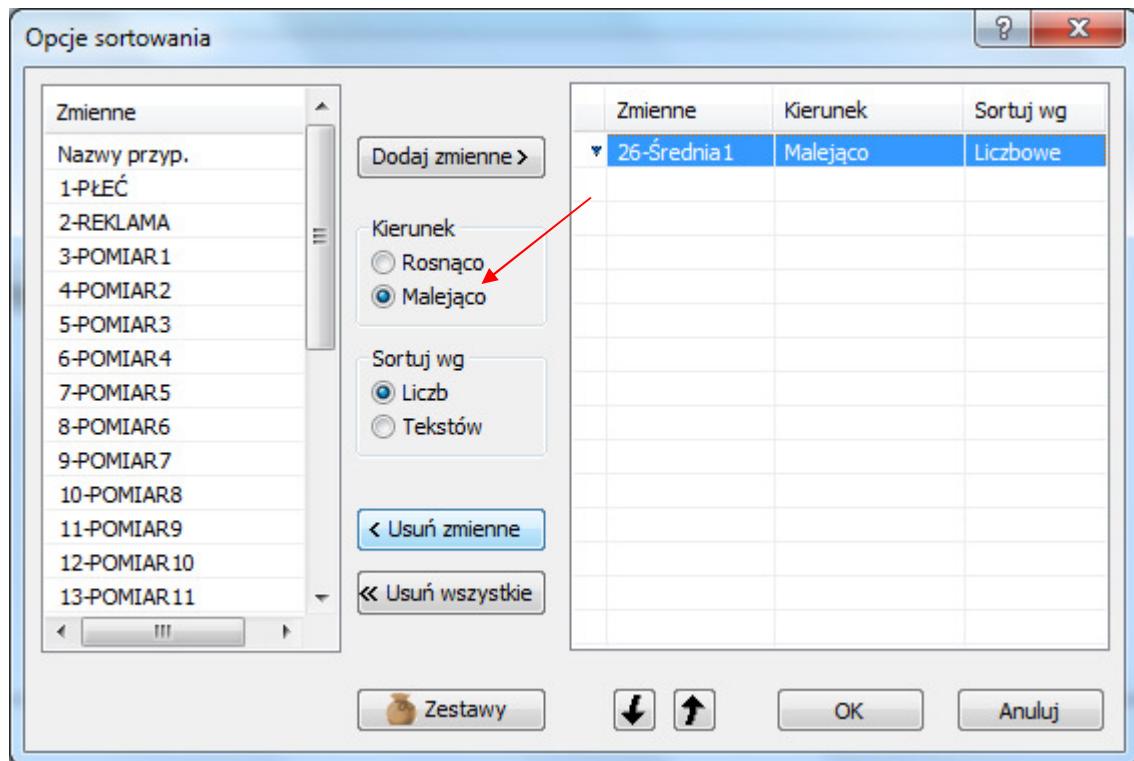
Sortowanie danych.

Teraz posortujemy dane względem dwóch utworzonych zmiennych, czyli średnich ocen z ostatnich 3 oraz 6 miesięcy.

1. Wstaw 2 węzły sortowania: Dane -> Przypadki -> Sortuj
2. Pierwszemu węźłowi zmień nazwę na „3 mies” a drugiemu na „6 mies” (na węźle prawy przycisk):



3. W pierwszym węźle wybierz parametry (Edytuj parametry) (Średnia1 to średnia z 3 mies):



4. A w drugim sortuj według Średnia2
 5. Uruchom i sprawdź wyniki:

Odpowiedzi	Badanie efektywności reklamy.																									
	15 POMIAR13	16 POMIAR14	17 POMIAR15	18 POMIAR16	19 POMIAR17	20 POMIAR18	21 POMIAR19	22 POMIAR20	23 POMIAR21	24 POMIAR22	25 POMIAR23	26 Średnia1	27 Średnia2													
N. Crust	6	4	9	3		2	2	5	9			9	6,66666667													
M. West	3	3	7	6		7	4	4	7			9	6,83333333													
M. Quick	4	1	2	9	7	6	3	2	9	6	9	8	6,16666667													
F. Joyce	7	5		8	6	1	9	3	6	9	9	8	6,16666667													
A. Smith	9	5		1	3		8	4	8			7	7,5													
D. Flynd	7	7	6	9	9	7	2	8				8	7,5													
N. Segal		7	7	4	2	5	6					7	7,5													
B. Quale	3		5	6	3	3	6	8				8	7,33333333													
I. Neil	2	5	4	4	4	3	6	7	6	8		6	6,66666667													
J. Liu		5		2	9		3	5	9	4		7	6,66666667													
R. Jerin	5	5	6	5	5	8	6	1	6	8		5	6,33333333 5,66666667													
R. Irving	9	9	3		2	5	7	6	7	6		6	6,33333333 6,16666667													
U. Andy	5	7	8	3	7	7	4	4	6	5		7	6													
J. Harper		4	8		2	5		6	8	5		5	6													
H. Morrow	4	7	5	1	4	7	5		9	4		4	5,66666667													
N. Luce	7	1	3		7		3	8	2	9		6	5,66666667													
M. Crow	7	2	4	3		7	1	2	2	6		9	5,66666667													
J. Hardy	2	1	5	4	3	8	3	5	1	8		8	5,66666667													
F. Wind	2	9	3	2	8	4		4	9	6		2	5,66666667													
C. Mayer	7	3	2	8	1	8	1		3	7		6	5,33333333													
D. Young	9		2	7	8	3	7	7	4	4		8	5,33333333													
S. Bird	4		2	9	2		1	2	9	4		3	5,33333333													
J. Owen		3	4	8	7	3	3	9	3	8		5	5,33333333 5,16666667													
D. Hartel	1		2	6	3	4	9	1	2	5		9	5,33333333													
L. Swinger		3	8	1	9		8	7	6	3		5	5,33333333													
E. Bynum	8		3	9	9	8	9	2	3	7		5	5,66666667													
T. Patore		7		8	5	2	2	0		0		1	5,66666667													

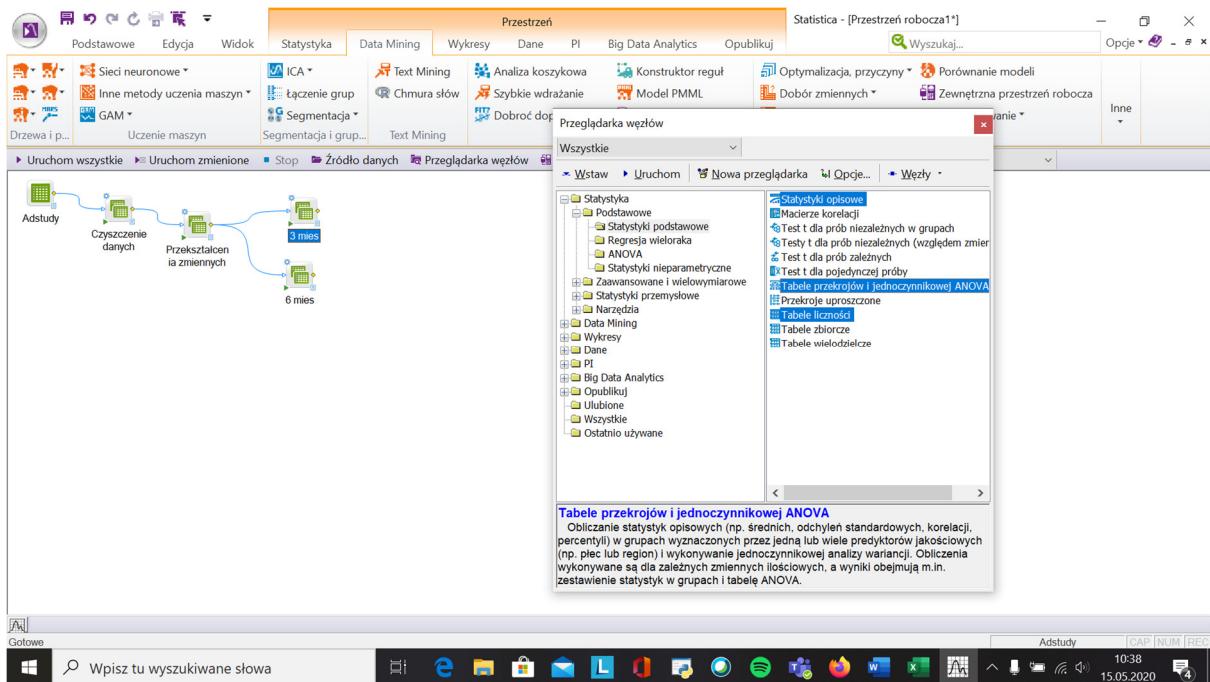
Odpowiedzi	Badanie efektywnosci reklamy.												
	15 POMIAR13	16 POMIAR14	17 POMIAR15	18 POMIAR16	19 POMIAR17	20 POMIAR18	21 POMIAR19	22 POMIAR20	23 POMIAR21	24 POMIAR22	25 POMIAR23	26 Średnia1	27 Średnia2
A. Smith	9	5		1	3	8	4	8	7	7,5	6,75		
M. West	3	3	7	6		7	4	4	7	9	9,833333333	6,666666667	
B. Quale	3		5	6	3	3	6	8	7	8,733333333	6,5		
F. Pono	7	1	4	7	8	9	9	6	9	2	4	5	6,5
N. Segal		7	7	4	2	5	6		8	7	7,5	6,5	
D. Flynn	7	7	6	9	9	7	2	8		7	8,7,5	6,4	
R. Irving	9	9	3		2	5	7	6	7	6	6,333333333	6,166666667	
F. Joyce	7	5		8	6	1	9	3	6	9	8,6,166666667		
I. Neil	2	5	4	4	4	3	6	7	6	8	6,666666667	6	
L. Swinger	3	8	1	9				8	7	6	3,5,333333333	6	
M. Quick	4	1	2	9	7	6	3	2	9	6	9	8,5,833333333	
H. Morrow	4	7	5	1	4	7	5		9	4	4,5,666666667	5,8	
J. Harper		4	8		2	5	6	8	5	5	6	5,8	
E. Bynum	8		3	9	9	8	9	2	3	7	5	5,5,666666667	
R. Jerin	5	5	6	5	5	8	6	1	6	8	5,6,333333333	5,666666667	
J. Liu		5		2	9		3	5	9	4	7,6,666666667	5,6	
N. Luce	7	1	3		7		3	8	2	9	6,5,666666667	5,6	
D. Young	9		2	7	8	3	7	7	4	4	8,5,333333333	5,5	
U. Andy	5	7	8	3	7	7	4	4	6	5	7,6	5,5	
S. Reber	1	9	1	0	1	8	6	5	3	7	4,4,666666667	5,5	
J. Hardy	2	1	5	4	3	8	3	5	1	8	8,5,666666667	5,5	
T. Hill	4	7	3	9	9		7	5	1	8	4,5,166666667	5,25	
J. Owen	3	4	8	7	3	3	9	3	3	8	5,5,333333333	5,166666667	
G. Hogan	7	1	9	9	1	8	3	8	5	1	6	4,5,166666667	
C. Mayer	7	3	2	8	1	8	1		3	7	6,5,333333333	5	
D. Hartel	1		2	6	3	4	9	1	2	5	9,5,333333333	5	
I. Preston		4	6	1	2	7	6	6	3	7	5	5,5	

Teraz możemy już wykorzystać posortowane arkusze danych do kontaktów z odpowiednimi osobami, "najbardziej zadowolone" (np. robiące najwięcej zakupów) osoby są na górze.

Podłączanie węzłów analiz statystycznych

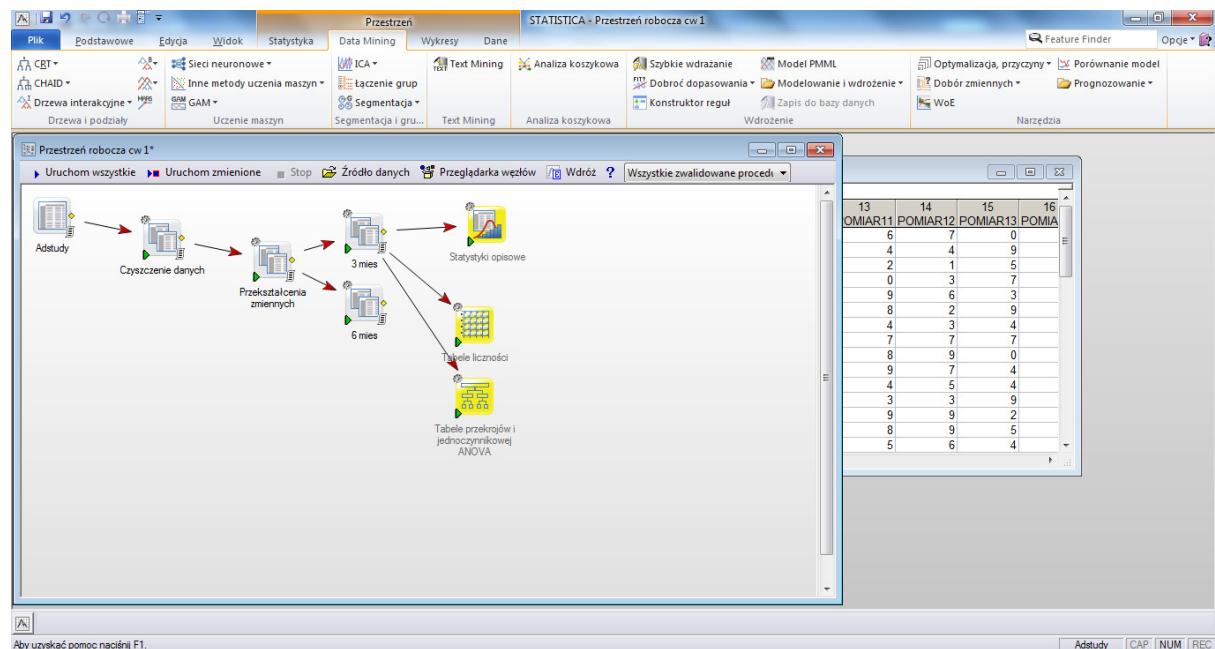
Dołączymy teraz do tych plików danych pewne statystyki opisowe i wykresy. Przypuśćmy, że interesuje nas ilość zakupów w ostatnich 3 oraz 6 miesiącach dla wszystkich obserwacji z pliku danych (czyli wszystkich kierowników zaopatrzenia), oraz średnie, w podziale względem zmiennej *REKLAMA* (podział na *Coke* oraz *Pepsi*).

1. Wstaw węzły: Statystyka -> Podstawowe -> Statystyki podstawowe -> *Statystyki opisowe, Tabele liczności, Tabele przekrojów i jednoczynnikowej ANOVA*



2. Skonfiguruj węzły (Edytuj parametry w każdym węźle):

- Statystyki opisowe* i *Tabele liczności* wybierz Średnia1, Średnia2, REKLAMA
- Tabele przekrojów i jednoczynnikowej ANOVA* wybierz jako zmienne zależne: Średnia 1, Średnia2, a zmienne grupujące: REKLAMA



Przykładowo, reklama *Pepsi* na przestrzeni ostatnich 3 miesięcy jest bardziej skuteczna niż na przestrzeni 6 miesięcy. Oprócz tego podane są tabele liczności itp.

Statistica - [Dokumenty raportu* - Tabela przekrojów statystyk opisowych (3 mies)]

Wyszukaj... Opcje

Podstawowe Edycja Widok Format Statystyka Data Mining Wykresy Narzędzia Dane Skoroszyt

Nowy Otwórz Zapisz Projekt Plik Dodaj do skoroszytu * Dodaj do raportu * Dodaj do MS Word * Wstaw do przestrzeni * Analizy Makro Integracja Opcje Wyewidencjonuj Sąsiadujące w poziomie Zamknij wszystkie Zawiadomienia Sąsiadujące w pionie Uporządkuj ikony Kaskada Okna

Dokumenty raportu*
Statystyki opisowe
Tabele przekrojów i jednociągowe
Tabele licznosci
Tabela licznosci: REKLAMA
Tabela licznosci: Średnia1
Tabela licznosci: Średnia2

Tabela przekrojów statystyk opisowych (3 mies)
N=50 (Zmienne zależne nie zawierają BD)

REKLAMA	Srednia1 Średnie	Srednia1 Ważnych	Srednia1 Odch.std	Srednia2 Średnie	Srednia2 Ważnych	Srednia2 Odch.std
PEPSI	5.327160	27	1.628723	5.248765	27	1.082208
COKE	5.072464	23	1.564139	4.699275	23	0.973376
Ogół	5.210000	50	1.588226	4.996000	50	1.059888

Adstudy P1.Z1 PEPSI 10:48 15.05.2020 CAP NUM REC

Gotowe Wpisz tu wyszukiwane słowa