

# Ćwiczenie: Analiza ryzyka kredytowego

## Dane do analizy

Informacje o klientach banku: *tree\_credit*:

1. Uruchom program Statistica i otwórz plik:

The screenshot shows the Statistica software interface with the following details:

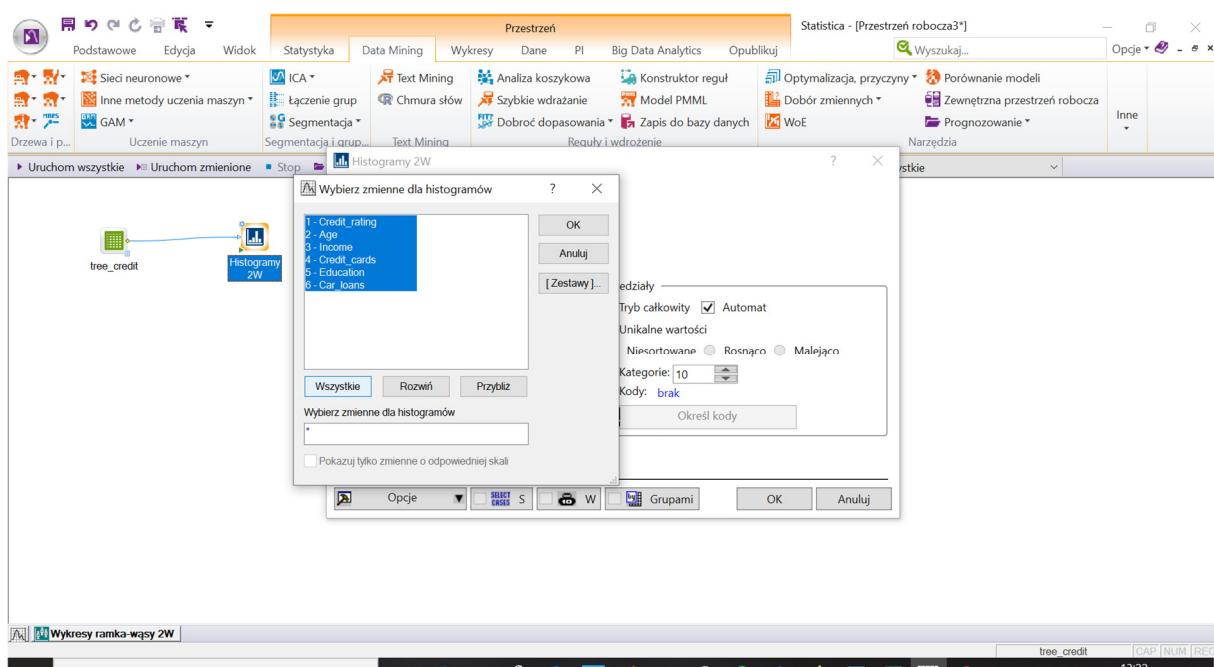
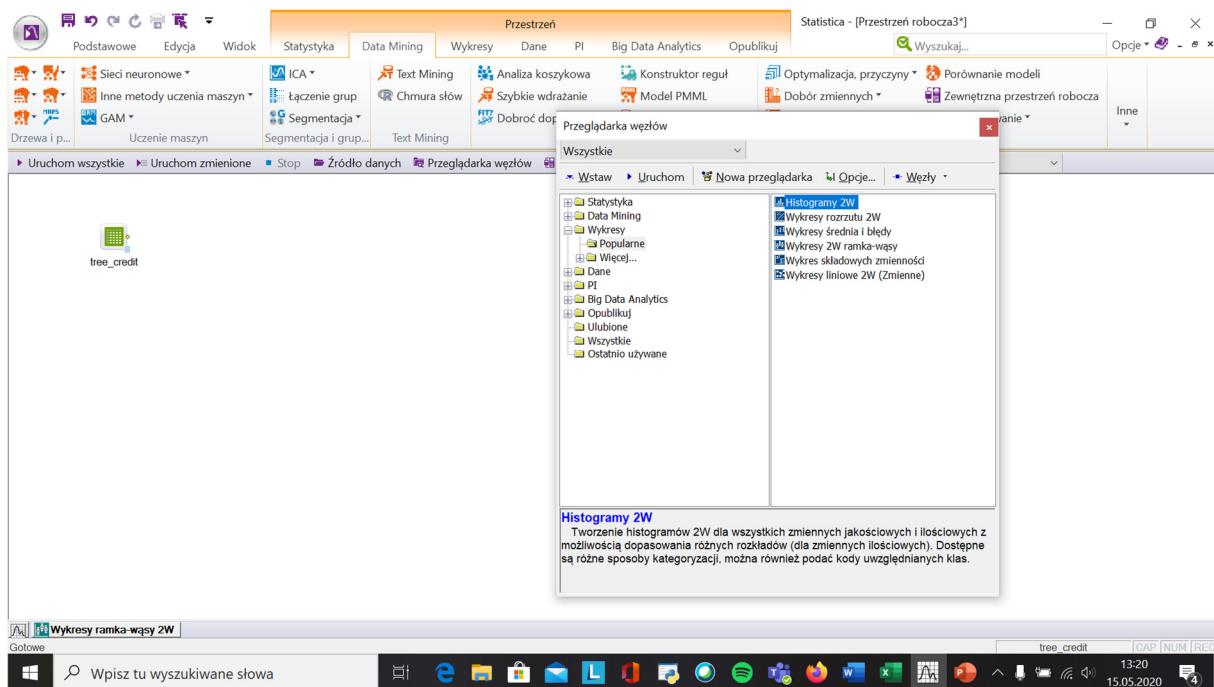
- Menu Bar:** Podstawowe, Edycja, Widok, Format, Statystyka, Data Mining, Wykresy, Narzędzia, Dane.
- Toolbars:** Przepisy, Data Miner, Drzewa, Uczenie maszyn, Segmentacja i grupowanie, Text Mining, Analiza skupień, Wykresy ramka-wąsy 2W.
- Data Grid:** Shows 2464 rows of data with columns: Credit\_rating, Age, Income, Credit\_cards, Education, and Car\_loans.
- Status Bar:** Wykresy ramka-wąsy 2W, Gotowe, Wpisz tu wyszukiwane słowa, tree credit, P1 Z1, Bad, Selekcja Nie, Waaga Nie, 13:06, 15.05.2020.

Nazwa pola	Opis
Credit_rating	Zdolność kredytowa: Bad, Good
Age	Wiek w latach
Income	Poziom dochodu Low, Medium, High
Credit_cards	Liczba posiadanych kart kredytowych: Less than five, Five or more
Education	Poziom wykształcenia: High school, College
Car_loans	Liczba zaciągniętych kredytów samochodowych: None or one, More than two

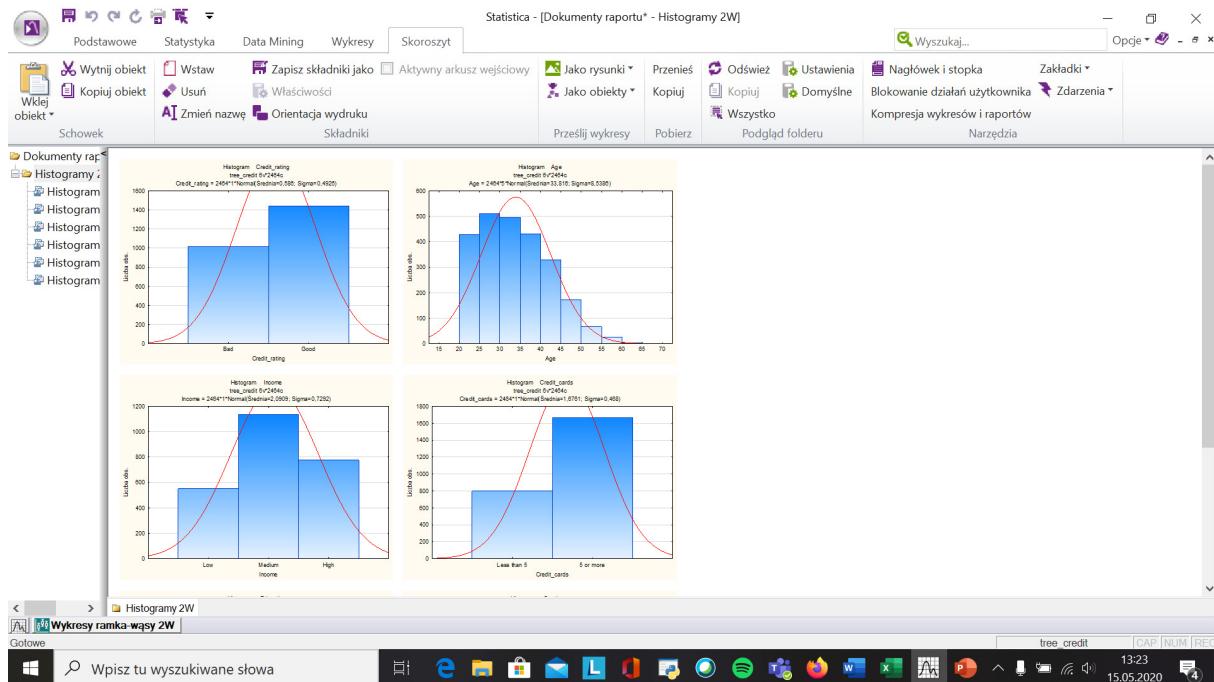
2. Otwórz nową przestrzeń i wybierz ten plik jako źródło

3. Analizujemy dane wejściowe:

- histogramy dla zmiennych kategorycznych: Wykresy->Popularne->*Histogramy 2W*



Uruchamiamy strumień i otrzymujemy:

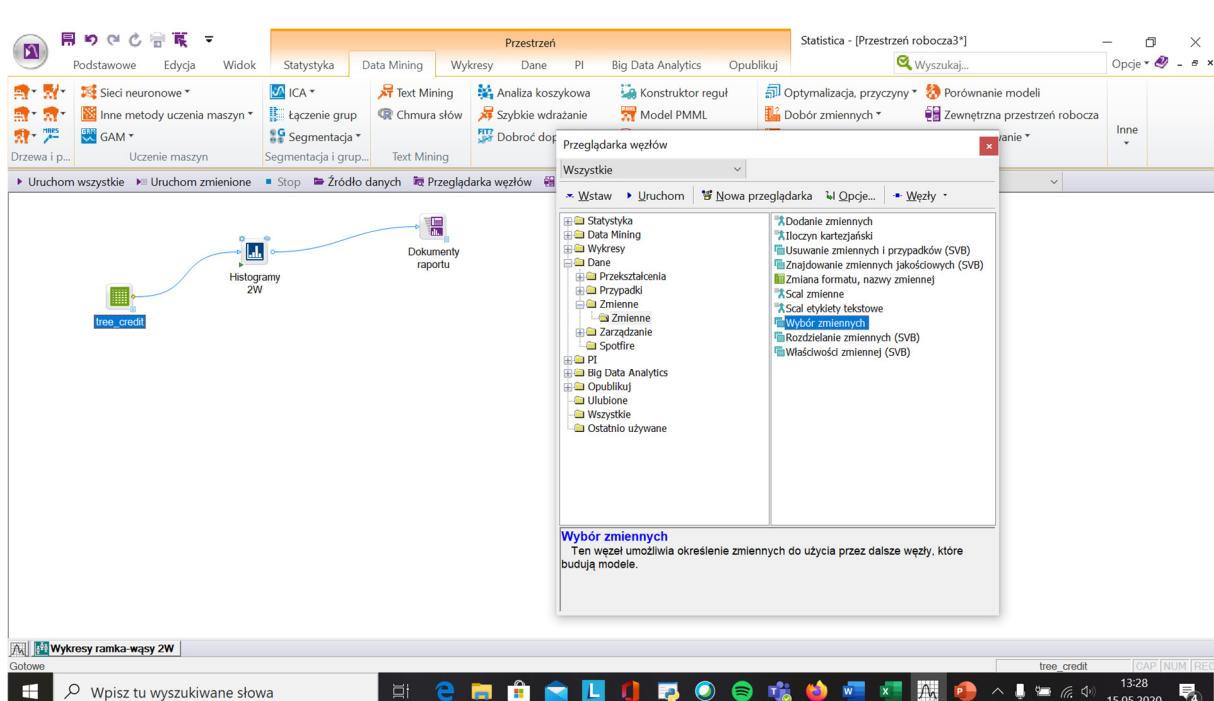


## Wnioski z analiz?

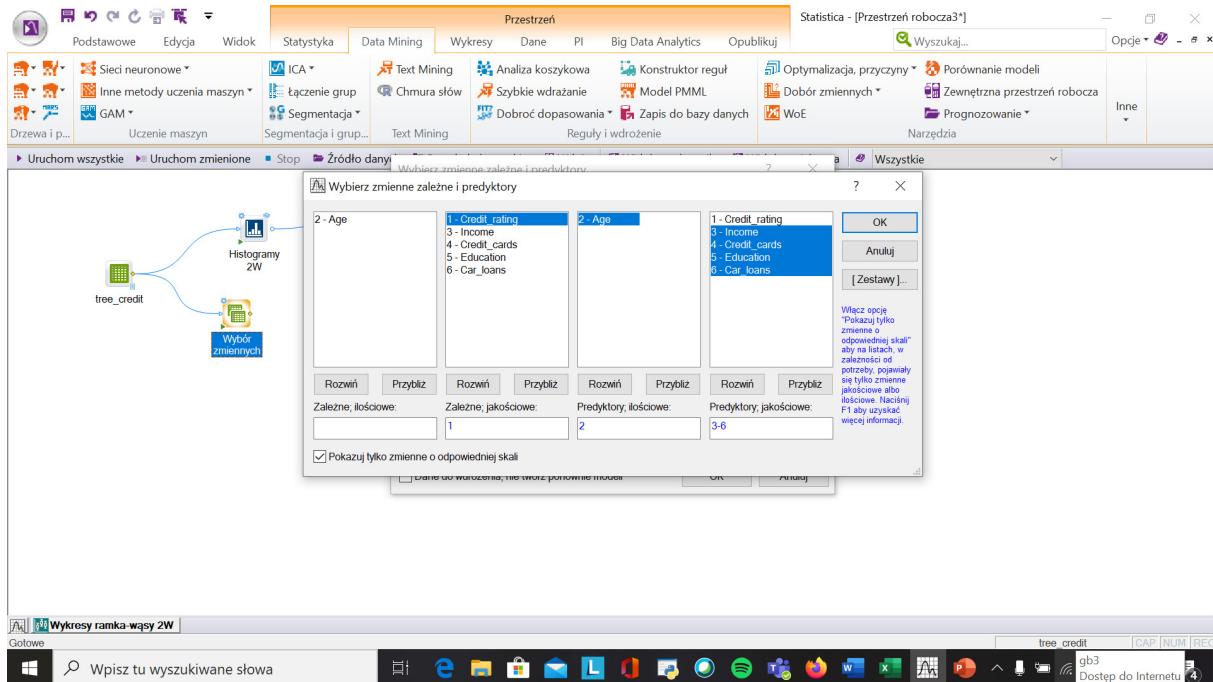
4. Przystępujemy do określenia którą zmienną będą przewidywali (zmienna predykcyjna), a które zmienne będą zmiennymi zależnymi

5. Zaznacz węzeł ze źródłem danych

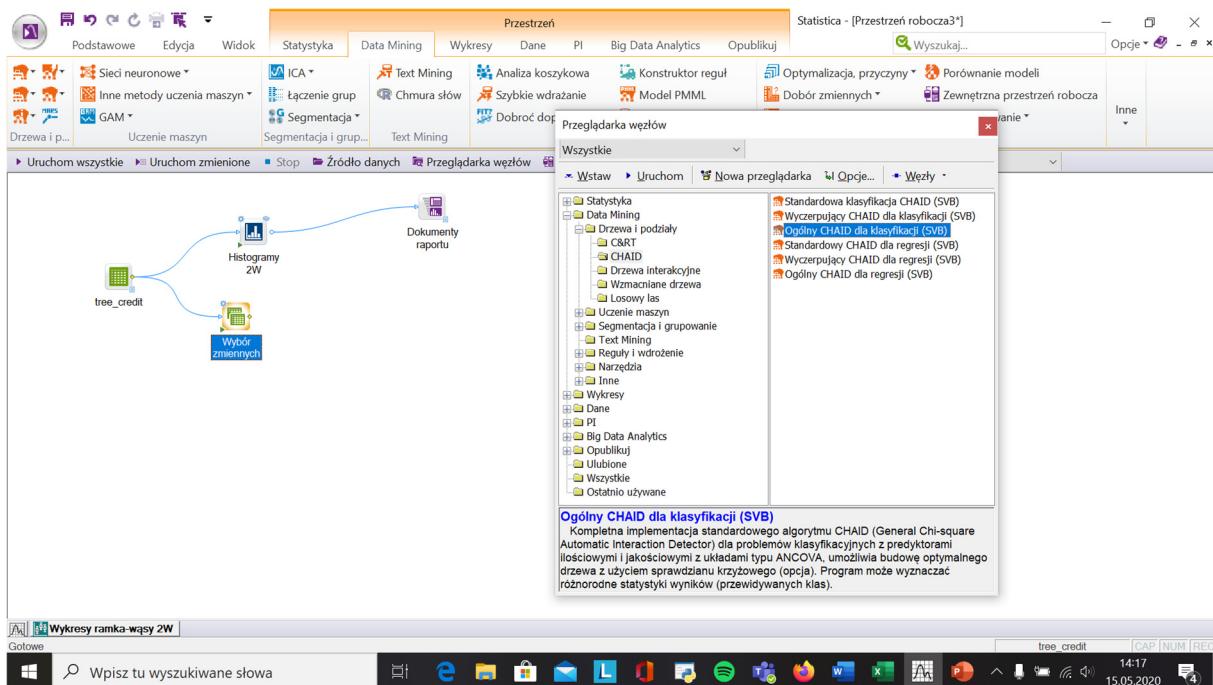
6. Z Przeglądarki węzłów wybierz: Dane->Zmienne->*Wybór zmiennych*:



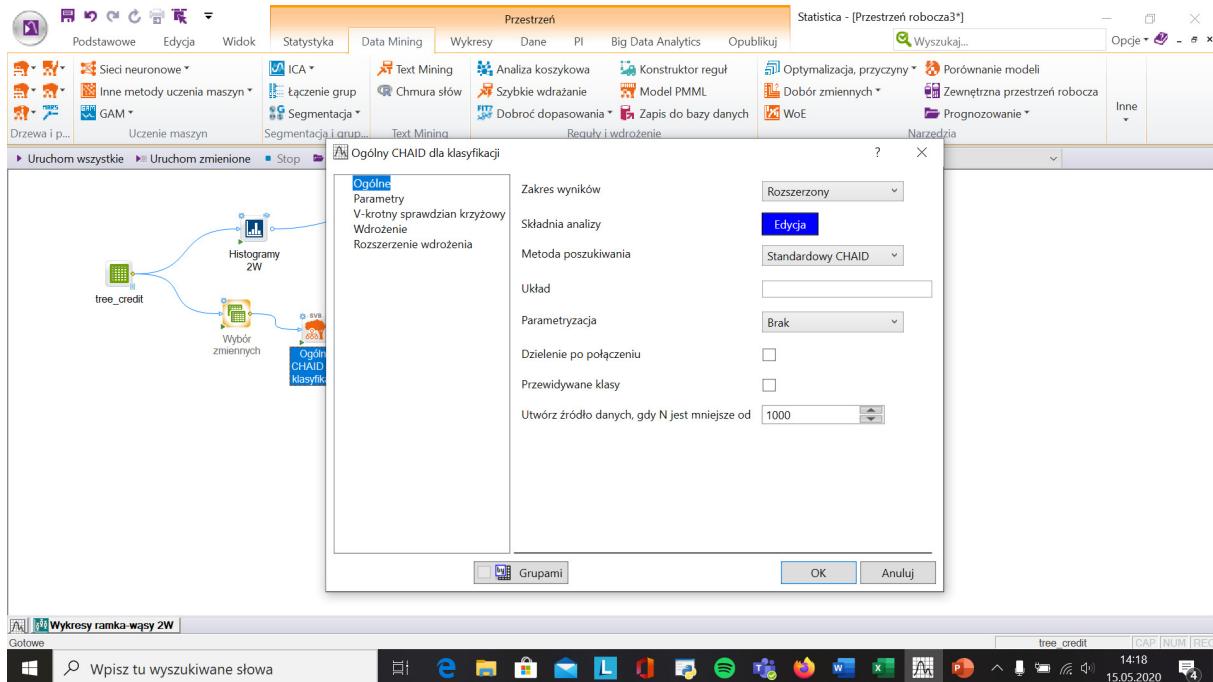
7. Wybierz następujące zmienne:



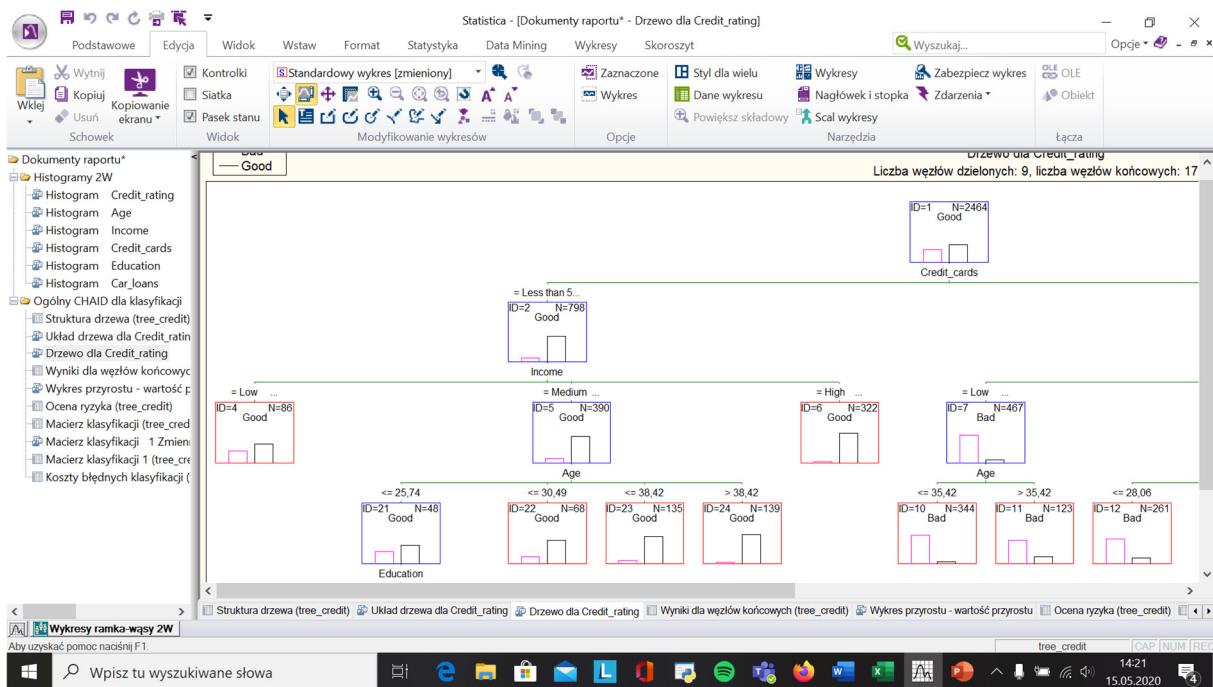
## 8. Wybierz węzeł modelowania: Data Mining -> CHAID->*Ogólny CHAID dla klasyfikacji*:



## 9. Kliknij 2x w węzeł i skonfiguruj go:



## 10. Uruchom węzeł i oglądni skoroszyt – przenalizuj drzewo:



## 11. Możemy sprawdzić, które zmienne są najważniejsze dla zmiennej predykcyjnej

12. Zaznacz w strumieniu węzeł Wybór zmiennych i wstaw Wszystkie->Dobór zmiennych:

The screenshot shows the Statistica software interface. At the top, there's a ribbon menu with tabs like 'Podstawowe', 'Edycja', 'Widok', 'Statystyka', 'Data Mining', 'Wykresy', 'Dane', 'PI', 'Big Data Analytics', and 'Opublikuj'. Below the ribbon, there's a search bar and a 'Wyszukaj...' button. The main area features a flowchart diagram on the left and a detailed search results window on the right.

**Flowchart Diagram:**

- A node labeled 'tree\_credit' has arrows pointing to 'Histogramy 2W' and 'Wybór zmiennych'.
- 'Histogramy 2W' points to 'Dokumenty raportu'.
- 'Wybór zmiennych' points to 'Ogólny CHAID dla klasyfikacji'.
- 'Ogólny CHAID dla klasyfikacji' points to 'Dokumenty raportu'.

**Search Results Window:**

**Dobór zmiennych**

Program wykonyuje następujące działania:

1. Jedna lista zmiennych zależnych:
  1. Jeśli dla źródła danych określono jedną listę zmiennych zależnych, to utworzone

### 13. Uruchom węzeł:

Statistica - [Dokumenty raportu\* - Wykres ważności]

Wyszukaj... Opcje

Wytnij Kopiuj Kopiowanie Usuń ekranu Schowek

Kontrolki Standardowy wykres [zmieniony] Siatka Wykres Zaznaczone Wykresy Wykresy Nagłówki i stopka Zabezpiecz wykres Zabezpiecz wykres Nagłówki i stopka Zdarzenia Obiekt Obiekt

Widok Pasek stanu Modyfikowanie wykresów Opcje Powiększ składowy Scal wykres Narzędzia Łącza

Dokumenty raportu\*

- Histogramy 2W
  - Histogram Credit\_rating
  - Histogram Age
  - Histogram Income
  - Histogram Credit\_cards
  - Histogram Education
  - Histogram Car\_loans
- Ogólny CHAID dla klasyfikacji
  - Struktura drzewa (tree\_credit)
  - Układ drzewa dla Credit\_ratin
  - Drzewo dla Credit\_rating
  - Wykres przyrostu - wartość p
  - Ocena ryzyka (tree\_credit)
  - Macierz klasyfikacji (tree\_cred)
  - Macierz klasyfikacji 1 Zmien
  - Macierz klasyfikacji 1 (tree\_cra
  - Koszty błędnych klasyfikacji (
- Dobór zmiennych
  - Najlepsze predyktory dla zmi
  - Wykres ważności
  - Lista najlepszych predyktorów

Wykres ważności  
Zmienna zależna:  
Credit\_rating

Zmienna	Ważność (Chi-kwadrat)
Income	~650
Credit_cards	~420
Age	~350
Car_loans	~280
Education	0

Najlepsze predyktory dla zmiennej zależnej jakościowej: Credit\_rating Wykres ważności Lista najlepszych predyktorów

## Zadanie:

Wyeliminuj zmienna Education i sprawdź jak teraz wygląda wynik.