

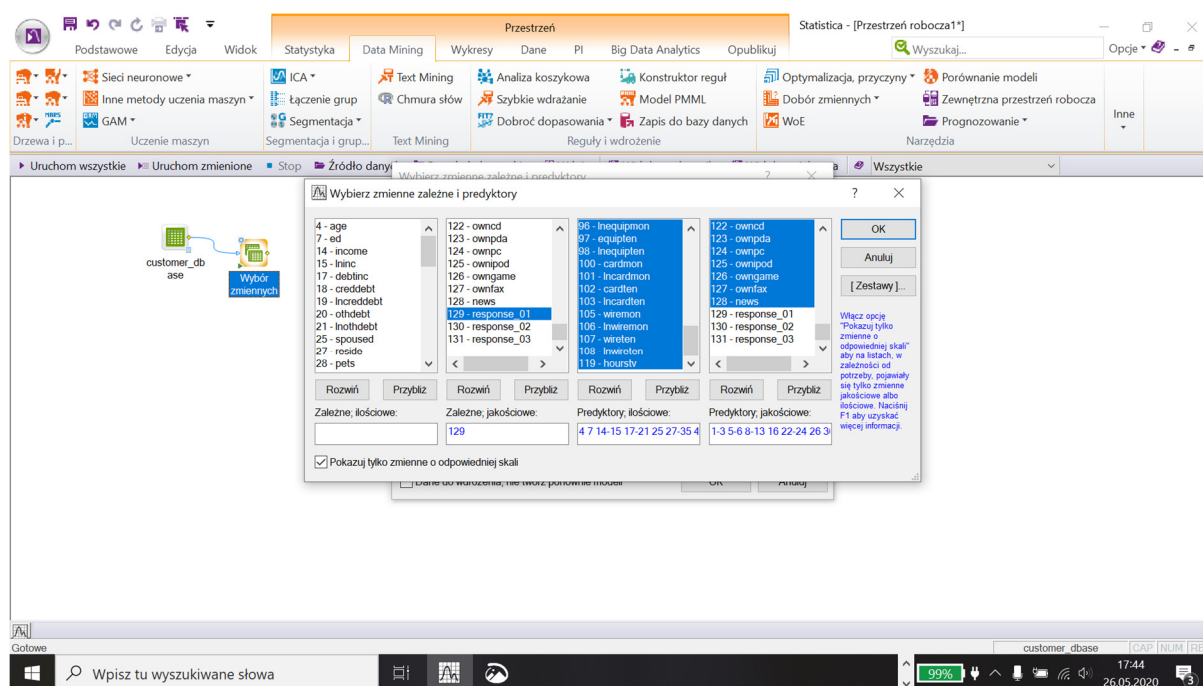
Ćwiczenie: Badanie predyktorów

Wybór cech (*feature selection*) pozwala zidentyfikować atrybuty, które mają największy wpływ na predykcję atrybutów wyjściowych.

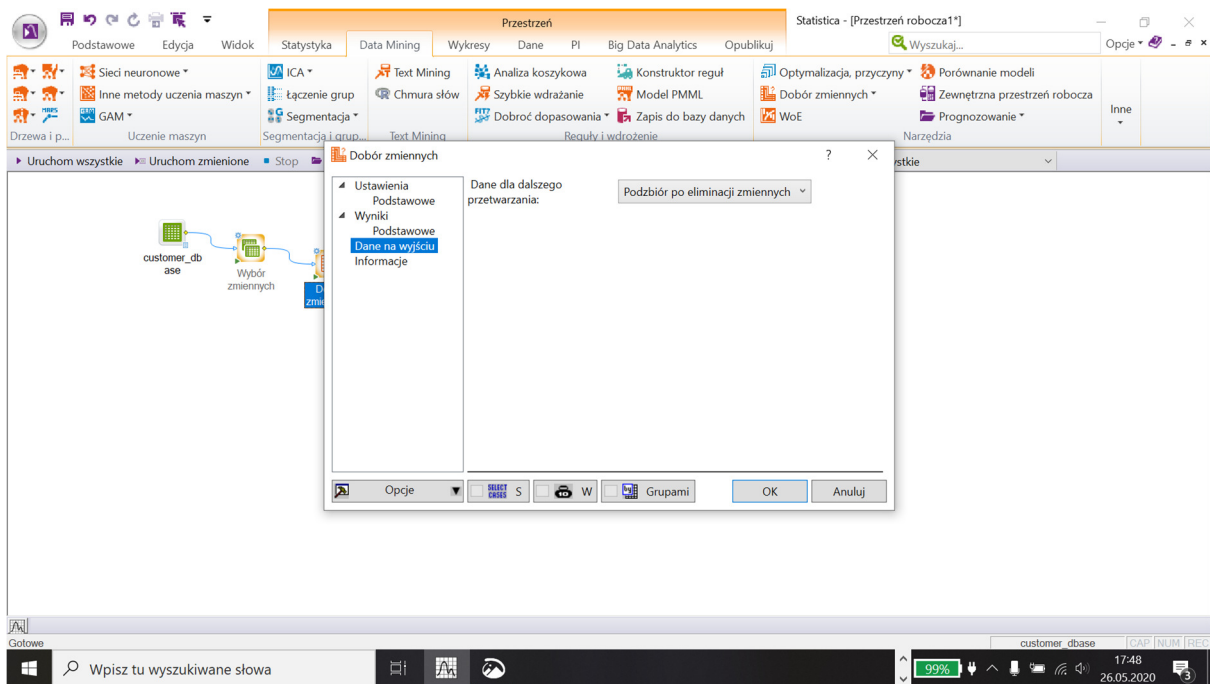
Dane *customer_dbsave.sav* zawierają informacje o odpowiedziach na pytania ankietowe udzielonych przez 5000 klientów firmy telefonicznej. Dane zawierają informacje o wieku klienta, zatrudnieniu, dochodach i statystyki wykorzystania telefonu. Trzy „docelowe” pola wskazują, czy klient odpowiedział na zapytania ofertowe. Firma chce wykorzystać te dane do wspomagania predykcji którzy klienci odpowiedzą w przyszłości na podobne oferty.

Przykład wykorzystuje tylko jedną ofertę jako cel. Wykorzystany zostanie model drzewa decyzyjnego CHAID: bez wyboru cech i z wyborem cech (10 najlepszych predyktorów). W drugim przypadku otrzymane wyniki okażą się efektywniejsze.

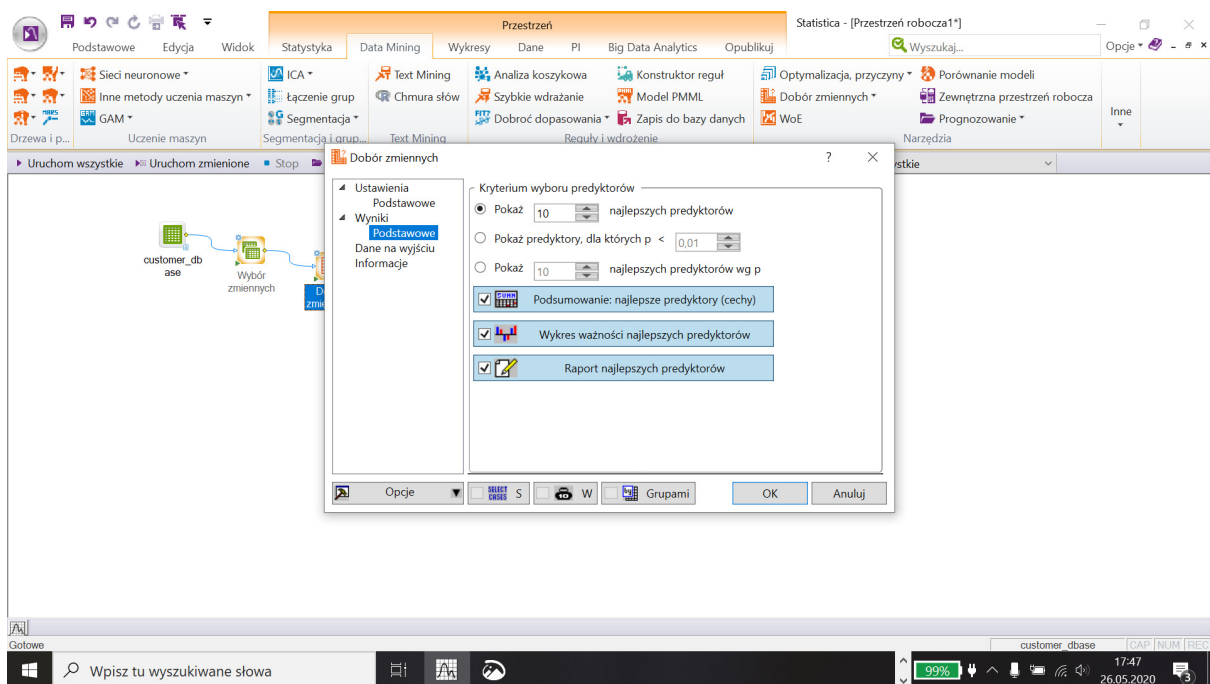
1. Otwórz nową przestrzeń roboczą
2. Wstaw źródło *customer_db.sav*
 - a. zaznacz *weź nazwy przypadków z 1 kolumny*
3. Otwórz przeglądarkę węzłów
4. Wstaw węzeł *Wybór zmiennych* i wybierz zmienną *response_01* jako zależną jakościową, pozostałe jako predyktory



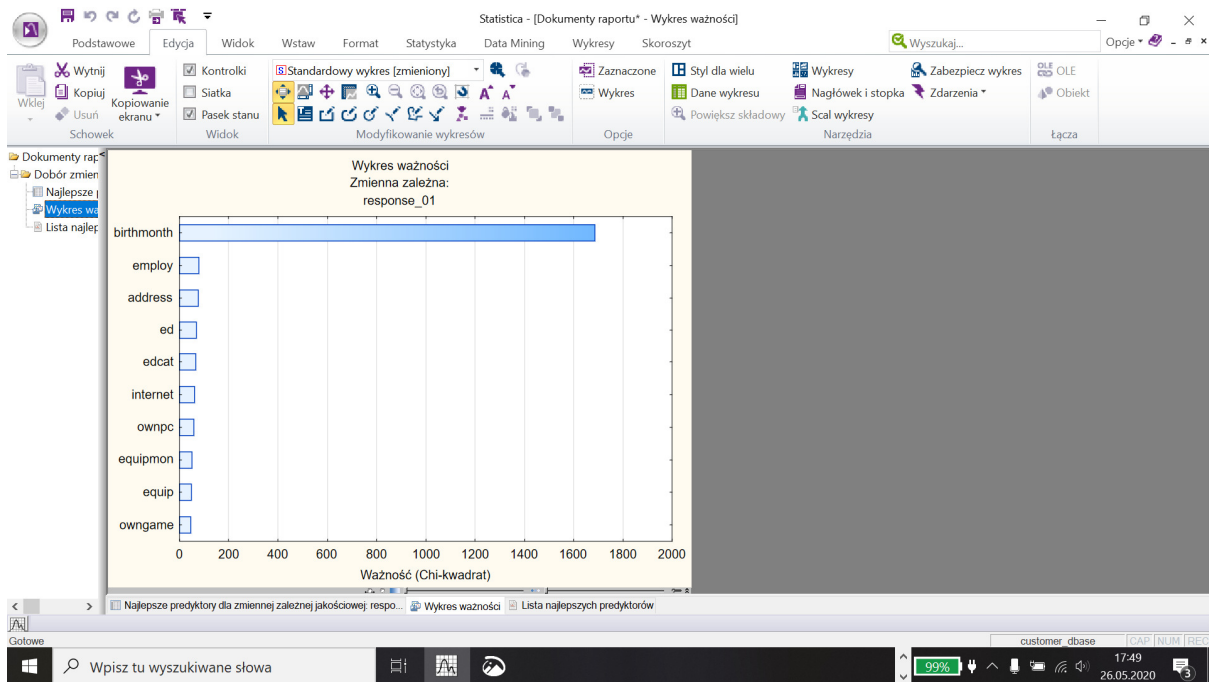
5. Wstaw węzeł *Dobór zmiennych* i ustaw
- a. Wyniki -> Dane na wyjściu na *Podzbiór po eliminacji zmiennych*



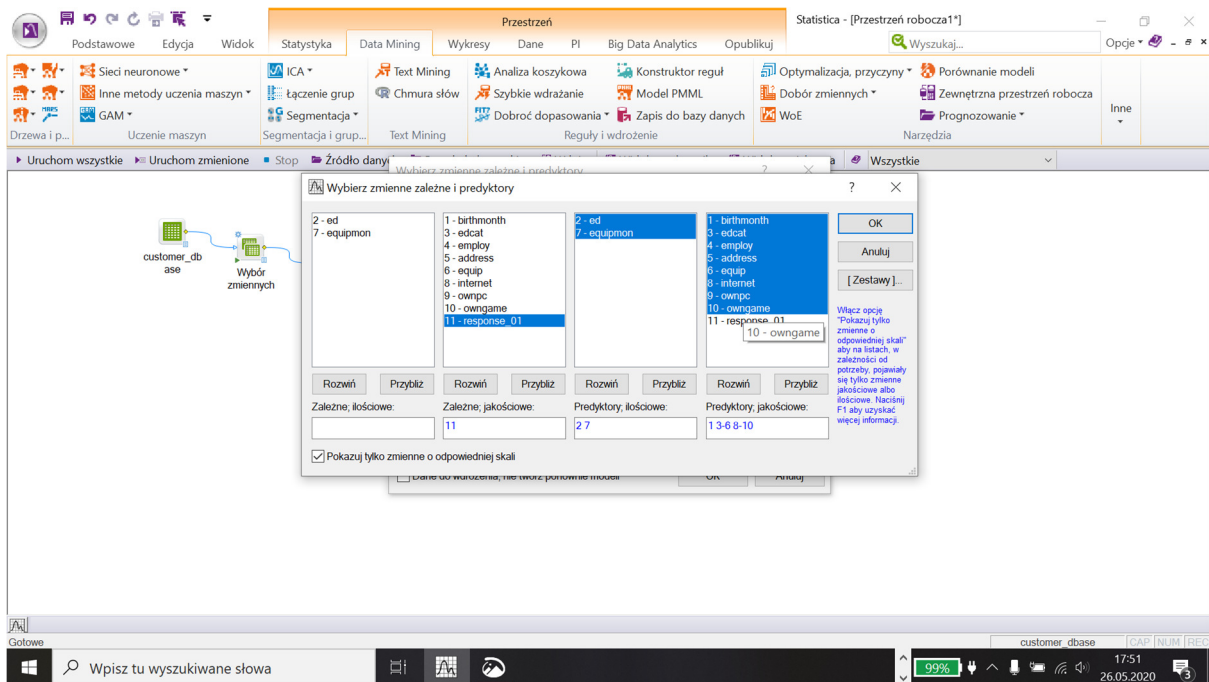
b. Wyniki -> Podstawowe: Pokaż 10 najlepszych predyktorów



6. Uruchom węzeł (pojawi się arkusz wyjściowy tylko z 10 zmiennymi):

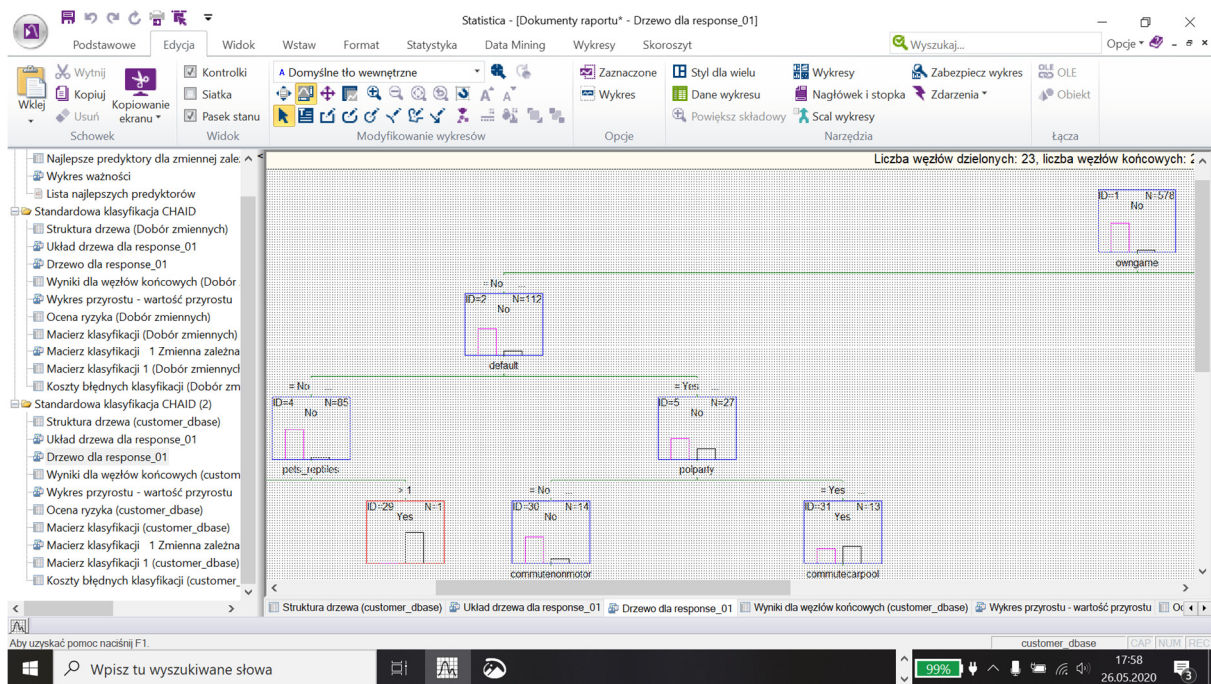


7. Podłącz do niego na wyjściu węzeł **Wybór zmiennych** i wybierz: **zależne jakościowe** = **response01**, a pozostałe 10 jako predyktory



8. Podłącz 2 węzły **Standardowa Klasyfikacja CHAID**:
- jeden podłącz do węzła **Wybór zmiennych**(na wejściu)
 - drugi do arkusza **Wybór zmiennych(2)**
 - w obydwu ustaw **Zakres wyników** na **rozszerzony**

a...bez wyboru cech:



Przeanalizuj obydwa drzewa. Które jest łatwiej analizować.

Zadanie:

Powtórz analizy dla *response_02* i *response_03*