



## **STATISTICA W MARKETINGU I BADANIACH RYNKU**

*Piotr Wójtowicz, StatSoft Polska Sp. z o.o.*

Większość osób zgodzi się ze stwierdzeniem, że dzisiejsza gospodarka w sposób szczególny wymaga profesjonalnej sprzedaży, bo zazwyczaj łatwiej jest wyprodukować produkt niż go sprzedać. Dlatego tak ważna jest rola działów marketingu wspierających procesy sprzedażowe oraz dogłębna znajomość rynku i zachowań klientów. Właściwe zaplanowanie działań wymaga przeprowadzenia odpowiednich badań i interpretacji ich wyników. Powoduje to, że szeroko rozumiany marketing to obszar, gdzie odpowiednio zastosowane metody analizy danych przynoszą wymierne korzyści. Dlatego też narzędzia *STATISTICA* są często wykorzystywane w tej dziedzinie z bardzo dobrym skutkiem. W niniejszym artykule przedstawię najważniejsze obszary, w jakich analiza danych i *STATISTICA* przynoszą największe korzyści.

Na kolejnych stronach znajdziecie Państwo bliższe informacje na następujące tematy:

Badania ankietowe .....	5
Marketing relacyjny .....	7
<i>Business Intelligence</i> .....	11
Usługi StatSoft Polska .....	11
Przykłady i opinie klientów .....	12
Rozwiązania StatSoft .....	18
StatSoft Polska .....	36

### **Badania ankietowe**

Badania ankietowe wymagają odpowiednich narzędzi analitycznych wspierających wszystkie etapy opracowywania wyników. W naszej ofercie znajdują się zaawansowane narzędzia wspierające proces analizy danych ankietowych, system umożliwiający wprowadzanie danych ankietowych, a także usługi szkoleniowe i doradcze z zakresu opracowania wyników badań.

Narzędzia *STATISTICA* wspierają cały proces opracowywania wyników badań ankietowych, począwszy od wprowadzania danych do systemu informatycznego, poprzez wstępne



opracowanie wyników, aż po końcowy raport z badań. Umożliwiają także automatyzację procesu analizy i raportowania.

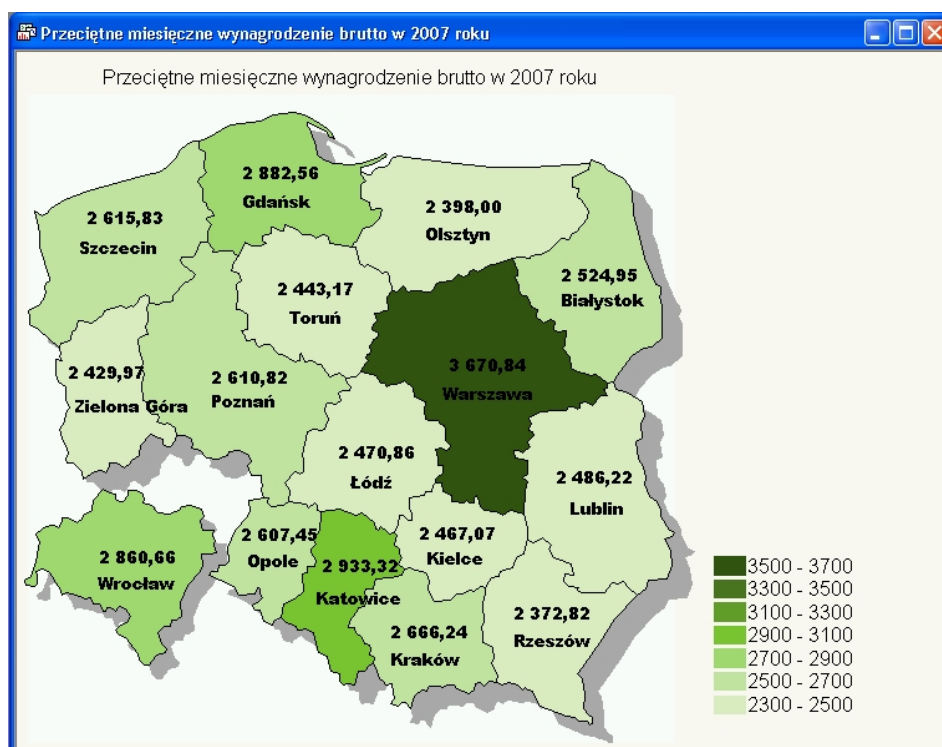
*STATISTICA* może współpracować z systemami wspomaganych komputerowo wywiadów telefonicznych (CATI), ale oferuje również dedykowane narzędzie zarówno do wypełniania ankiet przez respondentów on-line, jak i do wprowadzania danych przez koderów z ankiet już wypełnionych.

## Czyszczenie danych

Czyszczenie danych to najbardziej czasochłonny i żmudny etap każdej analizy danych. Dla badacza oznacza on często wiele godzin lub dni spędzonych nad sprawdzaniem poprawności zgromadzonych danych, usuwaniem błędów i pomyłek występujących w danych. *STATISTICA* wspomaga proces czyszczenia danych za pomocą szeregu modułów służących do przygotowania danych, które umożliwiają między innymi sprawdzanie ich poprawności, zmianę sposobu kodowania, zarządzanie brakami danych oraz radzenie sobie z przypadkami odstającymi. Co ważne, wszystkie czynności związane z przygotowaniem danych można zautomatyzować za pomocą skryptów nagranych podczas analizy, które mogą być dalej modyfikowane przez użytkownika.

## Wizualizacja danych

Jeden wykres potrafi przekazać więcej niż dziesiątki tabel. *STATISTICA* od lat słynie z wysokiej jakości wykresów, które pozwalają na przygotowanie nie tylko merytorycznie zaawansowanych, ale także efektownych raportów. Dodatkowo, możliwość prezentacji zależności znalezionych w danych za pomocą wykresów na mapach, są szczególnie przydatne przy prezentacji danych „geograficznych”.





## *Analizy przekrojowe*

Dzięki wykorzystaniu narzędzi analitycznych *STATISTICA* można przeprowadzać analizy przekrojowe, analizy w grupach, porównywać różne grupy. Dzięki specjalnie zaprojektowanemu interfejsowi wykonywanie tego typu analiz nie wymaga rozległej wiedzy na temat statystyki.

## **Marketing relacyjny**

Marketing relacyjny umieszcza potrzeby klienta w centrum zainteresowania organizacji. Zakłada konieczność budowania indywidualnych, trwałych kontaktów pomiędzy dostawcą a klientem. Budowanie trwałych relacji z klientem przynosi wiele wymiernych korzyści:

- ◆ stali klienci przynoszą firmie zdecydowanie większe korzyści niż klienci incydentalni,
- ◆ pozyskanie nowych klientów jest bardziej kosztowne niż zatrzymanie obecnych,
- ◆ odzyskanie klienta, który odszedł do konkurencji, kosztuje znacznie więcej niż jego zatrzymanie,
- ◆ łatwiej jest sprzedać nowy produkt osobie, która jest już naszym klientem niż osobie zupełnie nieznaną.

Budowanie takich relacji w przypadku produktów masowych jest trudne ze względu na konieczność „indywidualnego” traktowania tysięcy klientów. W takiej sytuacji niezbędne staje się wykorzystanie metod statystycznych i data mining, które znakomicie sprawdzają się w przypadku dużej liczby klientów, gdzie nawiązywanie indywidualnych relacji z pojedynczym klientem jest praktycznie niemożliwe. Dzięki odpowiednio zastosowanym metodom analitycznym *STATISTICA* możliwa jest analiza i interpretacja danych gromadzonych o klientach i podejmowanie na tej podstawie takich działań, które nieróżnią się od działań podejmowanych w stosunku do klientów, z którymi utrzymujemy rzeczywiste relacje.

## *CRM analityczny*

Z filozofią marketingu relacyjnego nieodłącznie związana jest metodyka CRM wymuszająca budowę indywidualnych, trwałych relacji z aktualnymi klientami i maksymalizację korzyści płynących z tej relacji.

CRM analityczny to podejście, które nie ogranicza się jedynie do gromadzenia danych o klientach, ale dzięki technikom statystycznym i data mining pozwala na przewidywanie zachowań oraz traktowanie w sposób „indywidualny” tysięcy klientów.

Z reguły relacja z klientem ewoluuje i zmienia się w czasie. Potencjalni klienci, dokonując zakupu, zmieniają się w nowych klientów, ci zaś, jeśli przez pewien czas korzystają z usług, mogą być postrzegani, jako stali klienci. Następnie pewna ich część może zrezygnować, stając się byłymi klientami. Budowa w *STATISTICA* modeli predykcyjnych związanych z cyklem życia klienta to typowy obszar wykorzystywania metod data mining



w zarządzaniu relacjami z klientem. Oczywiście w zależności od etapu cyklu zmieniają się typy analiz, jakie można przeprowadzić.

Do najpowszechniej stosowanych można zaliczyć:

- ◆ segmentację,
- ◆ budowę modeli predykcyjnych odnoszących się do etapów cyklu życia klienta,
- ◆ analizy dotyczące sekwencji zakupów oraz podobieństwa i powiązań między produktami.

W poniższych punktach znajdują się krótkie opisy poszczególnych obszarów wykorzystania narzędzi analitycznych *STATISTICA*.

### ***Segmentacja***

Analiza data mining pozwala zidentyfikować grupy podobnych do siebie klientów, innymi słowy dokonać ich segmentacji. Tego typu analiza jest przydatna do optymalizacji naszych działań i „indywidualnego” traktowania klientów. Wszystkie osoby znajdujące się w tej samej grupie uważane są za podobne do siebie, osoby znajdujące się w różnych grupach są różne. Dzięki tego typu podziałowi nie trzeba określać odrębnych działań dla każdego z klientów. Wystarczy, jeśli dany sposób postępowania zostanie określony w odniesieniu do całej grupy (segmentu) podobnych osób.

Innymi zaletami segmentacji jest możliwość patrzenia na klientów z nieco innej perspektywy, umożliwiającej lepsze zrozumienie ich cech i zachowań. Wyniki segmentacji mogą również wskazać działania, jakie należy podjąć w stosunku do określonych grup. Na przykład: jeśli osoby znajdujące się w dwóch różnych segmentach mają bardzo podobne cechy demograficzne, a tylko jeden reprezentuje wartościowych klientów, istnieje szansa, że klienci z drugiego segmentu mogą stać się równie wartościowi, jeśli podejmiemy w stosunku do nich odpowiednie działania. *STATISTICA* umożliwia szybką i efektywną realizację segmentacji z wykorzystaniem najnowszych algorytmów analitycznych.

### ***Skoring marketingowy***

Modele skoringowe są jednym z najpopularniejszych rodzajów modeli analitycznych wykorzystywanych w biznesie. Są znakomitym wsparciem menedżerów, którzy chcą zarządzać relacjami ze swoimi klientami i optymalizować kampanie skierowane w ich kierunku. Najczęstszym wykorzystaniem modeli skoringowych w zarządzaniu relacjami z klientem są modele oceniające skłonność klientów do rezygnacji z usług (churn) oraz modele pozwalające optymalizować kampanie sprzedażowe - oceniające skłonność klientów do zakupu określonych produktów.

Dzięki wykorzystaniu modeli skoringowych uzyskuje się wymierne korzyści:

- ◆ zmniejszenie kosztów kontaktów z klientami dzięki ograniczeniu kontaktów do grupy klientów potencjalnie zainteresowanych ofertą;
- ◆ wspomaganie i zwiększanie efektywności procesu sprzedażowego, dzięki wiedzy o tym, jakie produkty i w jakiej kolejności należy oferować danemu klientowi;



- ◆ zwiększenie zadowolenia klientów, ponieważ klienci doceniają traktowanie w sposób indywidualny i kierowanie do nich ofert, którymi naprawdę są zainteresowani.

*STATISTICA* oferuje dedykowane rozwiązanie do budowy modeli i kart skoringowych – *Zestaw Skoringowy STATISTICA*, dzięki któremu modele skoringowe można budować przy użyciu szeregu mniej lub bardziej zaawansowanych metod statystycznych i data mining. Do najpopularniejszych należy zaliczyć regresję logistyczną, drzewa klasyfikacyjne oraz różne „odmiany” drzew – drzewa wzmacniane (*boosting trees*) czy losowy las (*random forest*). Wykorzystywać można również sieci neuronowe, metodę wektorów nośnych (*support vector machines*) czy k-najbliższych sąsiadów i szereg innych.

*STATISTICA* zawiera odpowiednie narzędzia, które pozwolą nawet mniej doświadczonym analitykom szybko i efektywnie budować, oceniać i wdrażać modele i karty skoringowe z wykorzystaniem zaawansowanych metod analitycznych.

### ***Optymalizacja kampanii marketingowych***

Zdobywanie nowych klientów jest ważnym źródłem wzrostu sprzedaży. Skuteczność kampanii mającej na celu zdobycie nowych klientów jest więc bardzo istotna. Bardzo ważnym zadaniem stojącym przed pracownikami działu marketingu jest wybranie odpowiednich osób, do których zostanie skierowana przygotowana oferta. Bardzo pomocne w wyborze segmentu potencjalnych klientów są narzędzia data mining zawarte w *STATISTICA Data Minerze*. Sprawdzają się one zwłaszcza w sytuacji, gdy dysponujemy bazą potencjalnych klientów, z których musimy wskazać jedynie tych najbardziej odpowiednich, co może w dużym stopniu ograniczyć koszty kampanii, nie zmniejszając jej skuteczności. Zob. także *Skoring marketingowy*.

### ***Cross- i up-selling***

Jedną z możliwości maksymalizacji korzyści z nawiązania relacji z klientami jest stosowanie tak zwanej sprzedaży krzyżowej (*cross-selling*). Polega ona na oferowaniu klientom pewnych dodatkowych usług i produktów związanych z wcześniejszymi zakupami (np. posiadacz rachunku w banku decyduje się na korzystanie z karty kredytowej). Można również starać się, by klienci zaczęli korzystać z poszerzonych wersji zakupionych wcześniej produktów i usług (*up-selling*). *STATISTICA Data Miner* umożliwia budowę modeli predykcyjnych przewidujących skłonności zakupowe. Możliwe jest także budowanie z wykorzystaniem dedykowanych narzędzi *STATISTICA* specjalnych modeli i kart skoringowych (zob. *Skoring marketingowy*). Zbudowane modele służą do prognozowania, którzy klienci są skłonni zakupić dany produkt z największym prawdopodobieństwem.

### ***Analiza lojalności i migracji klientów (churn analysis)***

Na wartość klienta wpływa nie tylko wielkość dokonywanych przez niego zakupów. Innym istotnym czynnikiem jest lojalność klienta. Klient regularnie korzystający z usług jest bardzo cenny, dlatego zapewnienie jego lojalności jest jednym z kluczowych zadań, jakie stoją przed osobami zajmującymi się zarządzaniem relacjami z klientami. W sytuacji, gdy rynek



nasyca się produktem lub usługą i jest coraz mniej potencjalnych klientów, znaczenie utrzymania klienta staje się coraz większe. Waga zagadnienia zapobiegania odejściom klientów jest coraz większa również ze względu na konieczność sprostania z jednej strony coraz silniejszej konkurencji, a z drugiej coraz większym wymaganiom klientów domagających się wyższej jakości za niższą cenę.

Metody analizy danych, a w szczególności data mining, stosowane są w zapobieganiu migracji klientów w rozmaity sposób. Techniki analityczne *STATISTICA* wykorzystuje się np. do badania i monitorowania satysfakcji klienta (np. z wykorzystaniem kart kontrolnych), wykrywania przyczyn decydujących o zadowoleniu klientów i ich lojalności, monitorowania zmian stopnia satysfakcji klientów. Do tego celu są wykorzystywane m.in. badania potrzeb i preferencji klientów za pomocą badań ankietowych oraz badania reklamacji.

Inne podejście to znalezienie tych klientów, dla których prawdopodobieństwo odejścia jest największe. *STATISTICA Data Miner* pozwala na zbudowanie modelu oceniającego skłonność do rezygnacji z usług każdego z klientów. Dzięki technikom data mining znajdowane są czynniki zwiększające ryzyko odejścia, a także sposób, w jaki najlepiej zapobiec odejściu. Wykorzystując dedykowane narzędzia *STATISTICA*, można budować modele i karty skoringowe (zob. *Skoring marketingowy*). Ponadto przewidywanie odejść klientów pozwala wyznaczać spodziewany całkowity czas pozostawania klientem i łączną wartość klienta.

Kolejnym zadaniem jest identyfikacja przyczyn potencjalnego odejścia. Wiedza ta jest konieczna, by podjąć w stosunku do zagrożonych klientów optymalne kroki zapobiegawcze (odejścia dobrowolne), bądź optymalizujące działania dotyczące potencjalnie niewyplacalnych klientów.

### ***Analizy sekwencji zakupów i analizy koszykowe***

Data mining pomaga również w określeniu powiązań między poszczególnymi produktami, a dokładniej daje odpowiedź na pytanie: „Jakie produkty są zwykle kupowane razem lub w pewnych odstępach czasu oraz w jakiej kolejności?”. Odpowiedzią na to pytanie jest zwykle zestaw pewnych reguł zbudowany na podstawie historycznych danych (transakcyjnych), zawierających zapis dokonywanych zakupów. Znajomość takich reguł może być bardzo przydatna w różnego rodzaju działaniach skierowanych na klienta. Wiedza na temat kupowanych produktów umożliwia także uzyskanie informacji odnośnie naszych klientów oraz motywów, jakimi się kierują, dokonując określonych zakupów. Tego typu analizy poprzedzają często opisywane powyżej kampanie sprzedaży krzyżowej. Narzędzia *STATISTICA* zawierają zarówno implementację tradycyjnej analizy koszykowej, jak i umożliwiają analizę sekwencji zakupowych.

### ***Systemy lojalnościowe***

Popularnymi narzędziami wykorzystywanymi do wzmocnienia relacji z klientami są systemy lojalnościowe. Narzędzia te w swojej istocie mają za zadanie skłaniać aktualnych klientów do częstszych, większych zakupów czy rozszerzenia koszyka zakupowego. Jednak samo wprowadzenie systemu lojalnościowego jest środkiem niewystarczającym, by



zapewnić lojalność klientów. Nawet atrakcyjny prezent zwykle nie jest wystarczającą motywacją do zakupu, jeśli oferta nie zaspokaja w pełni potrzeb i oczekiwań klienta.

Stąd też poza samym stosowaniem systemów lojalnościowych konieczne jest podjęcie działań mających na celu lepsze rozumienie potrzeb i oczekiwań klientów. Znakomitym źródłem informacji o klientach są dane gromadzone w związku z transakcjami klientów korzystających z systemu lojalnościowego. Wartość tych danych jest wprost nieoceniona. Dzięki wykorzystaniu narzędzi analitycznych *STATISTICA* możemy określać preferencje i zwyczaje klientów, które mogą z znacznym stopniem wpłynąć na efektywność działań marketingowych, zmniejszyć ich koszty, a także pozwalają lepiej dopasować naszą ofertę do potrzeb naszych klientów.

## ***Business Intelligence***

Systemy *Business Intelligence* umożliwiają przekształcenie danych w informację, a informacji w wiedzę, która może być wykorzystana do zwiększenia konkurencyjności przedsiębiorstwa. Składają się z reguły z systemu bazodanowego gromadzącego dane przedsiębiorstwa (najczęściej jest to hurtownia danych) oraz z warstwy analityczno-raportującej, analizującej dane zawarte w hurtowni danych, pozwalającej na wizualizację danych, automatyzację procesu raportowania oraz dającej możliwość przewidywania przyszłości na podstawie wzorców zawartych w danych historycznych.

Ze względu na otwartą architekturę systemy *STATISTICA* w naturalny sposób integrują się z bazami i hurtowniami danych przedsiębiorstw, tworząc warstwę systemu *Business Intelligence* odpowiedzialną za analizę i raportowanie. Wykorzystanie w warstwie analityczno-raportującej systemu *STATISTICA Enterprise* umożliwia ponadto automatyczne generowanie raportów podsumowujących stan obecny, również dla tzw. kokpitów menedżerskich oraz budowę korporacyjnego portalu wiedzy o przedsiębiorstwie i jego procesach.

## **Usługi StatSoft Polska**

StatSoft Polska nie poprzestaje na oferowaniu swoim klientom narzędzi analitycznych z rodziny *STATISTICA*. Bardzo ważnym elementem naszej działalności jest prowadzenie szkoleń, konsulting w zakresie analizy danych oraz budowa kompletnych systemów analitycznych.

### ***Szkolenia***

Ważnym elementem naszej oferty są szkolenia, podczas których można zapoznać się z metodyką prowadzenia badań marketingowych oraz sposobami opracowywania wyników. Są to kursy autorskie, a nie kalki szkoleń przygotowanych za granicą i przetłumaczone na język polski. Dzięki temu, że programy naszych kursów są przygotowywane od początku do końca przez prowadzących, specjalistów w zakresie analizy danych w różnych zastosowaniach, są one doskonale dopasowane do polskich realiów i spełniają oczekiwania



uczestników. Oferujemy cały cykl szkoleń poświęconych badaniom rynkowym i marketingowym. Uczestnicy mają możliwość zapoznania się z zagadnieniami analizy danych ankietowych, metodami statystycznymi wykorzystywanymi w badaniach marketingowych, podstawami modelowania, technikami segmentacji rynku oraz badaniem satysfakcji i wartości klienta. Cykl kursów przeznaczony jest dla przedstawicieli agencji badań rynku i działów marketingowych przedsiębiorstw, firm badawczych, uczelni, instytutów naukowych oraz instytucji samorządowych. Oferujemy także kursy na zamówienie, które są dostosowywane do indywidualnych potrzeb zamawiającego. Potwierdzeniem najwyższej jakości szkoleń StatSoft Polska są opinie uczestników, które zamieszczone są na stronie internetowej [www.StatSoft.pl/kursy.html](http://www.StatSoft.pl/kursy.html). Są tam też szczegółowe informacje o kursach.

Ponadto współpracujemy z Krakowską Szkołą Biznesu Uniwersytetu Ekonomicznego przy realizacji studiów podyplomowych: „Analizy Marketingowe z wykorzystaniem data mining”. Zajęcia z praktycznej analizy danych z wykorzystaniem *STATISTICA* obejmują zarówno klasyczne metody analizy danych (m.in. modele regresyjne, modele strukturalne, analizę skupień), jak i narzędzia eksploracji danych (data mining) - drzewa klasyfikacyjne i regresyjne, sieci neuronowe, reguły asocjacyjne (analiza koszykowa).

## ***Konsulting***

StatSoft świadczy również usługi konsultingowe, wspierając badaczy w realizacji całego procesu analizy danych ankietowych, począwszy od opracowania metodologii badań, aż po analizę wyników. Współpraca z nami udowadnia, że korzystanie z zaawansowanego systemu analitycznego nie musi być trudne - pozostaje tylko odpowiedzieć na pytanie, gdzie analiza danych może przynieść wymierne korzyści.

StatSoft Polska oferuje usługi konsultingowe w zakresie analizy danych i budowy systemów analitycznych. Mogą one obejmować bardzo szeroki zakres zagadnień, np. analizę zgromadzonych danych, budowę modeli analitycznych, opracowanie i implementację kompletnych systemów analitycznych (od zbierania i gromadzenia danych, przez zaprojektowanie szablonów analiz, po automatyczne generowanie raportów), doradztwo w zakresie gromadzenia i analizy danych, tworzenie dedykowanych aplikacji działających w środowisku *STATISTICA* i wykorzystujących jej unikalne właściwości analityczne. Do StatSoft Polska może zwrócić się każdy, kto chce zasięgnąć porady w zakresie analizy danych lub chce skorzystać z pomocy naszych doświadczonych specjalistów. Więcej informacji o usługach konsultingowych StatSoft Polska, np. opinie o wybranych wdrożeniach, można znaleźć na stronie [www.StatSoft.pl/konsulting.html](http://www.StatSoft.pl/konsulting.html).

## **Przykłady i opinie klientów**

Od wielu lat *STATISTICA* i StatSoft cieszą się dużym uznaniem u klientów, obszar badań marketingowych i rynkowych nie jest tu wyjątkiem. Między innymi dlatego *STATISTICA* tak często jest wykorzystywana do analizy danych w tym obszarze, a pracownicy działów marketingowych tak chętnie korzystają z usług StatSoft Polska.



## **Grupa Wydawnicza INFOR S.A.**

*W ramach szeroko zakrojonych działań mających na celu poprawę jakości jednego z produktów Grupy Wydawniczej INFOR S.A. wspólnie z firmą StatSoft Polska został opracowany i wdrożony system badań monitoringowych. Ponieważ z naszej perspektywy był to jeden z kluczowych czynników sukcesu całego przedsięwzięcia, potencjalnym kooperantom zostały postawione bardzo wysokie wymagania.*

*Złożoność oferty produktowej tytułu (czasopismo plus multimedia plus usługi okołoprojektowe) oraz potrzeba pomiaru dynamicznie zmieniających się parametrów związanych z poszczególnymi jej elementami z jednej strony, z drugiej zaś oczekiwania związane z możliwością wykorzystywania wyników przez różne struktury firmy powodowały, że poszukiwanie złotego środka było drogą trudną i najeżoną pułapkami.*

*W efekcie współpracy udało się stworzyć efektywne narzędzie do monitoringu jakościowego naszego dwutygodnika, które ze względu na swoje zalety znajduje zastosowanie w badaniach jakościowych innych naszych produktów. Otrzymane wyniki są wykorzystywane przez redakcje na poziomie operacyjnym, a dzięki możliwości tworzenia wskaźników syntetycznych także jako dane wejściowe do podejmowania decyzji o charakterze strategicznym.*

*Należy podkreślić, że przez cały okres współpracy konsultanci ze StatSoft Polska wykazywali nie tylko zaangażowanie, ale przede wszystkim wiedzę i doświadczenie, dzięki którym projekt zakończył się sukcesem.*

*Marek Korzyński  
Menedżer Rozwoju Rynku  
Grupa Wydawnicza INFOR S.A.*

## **Wydawnictwo Szkolne PWN**

*Serdecznie dziękuję za przesłanie nowej wersji programu STATISTICA. Gratuluję, rewelacyjne zmiany, trudno mi się teraz oderwać od pracy ze STATISTICA. Cieszę się, że jestem użytkownikiem tego programu.*

*Piotr Dobrodziej  
Konsultant ds. Badań Rynku  
Wydawnictwo Szkolne PWN*

## **PKN ORLEN**

PKN ORLEN, dysponując największą w Polsce siecią stacji benzynowych (ok. 2 tys. stacji), posiada także największy w Polsce program lojalnościowy – VITAY. W programie tym bierze już udział ponad 5,5 miliona polskich kierowców. Aby zmaksymalizować korzyści płynące z programu lojalnościowego, PKN ORLEN wdrożył zaawansowany system analityczny, pozwalający na gromadzenie i analizę tak dużej ilości informacji i uzyskanie wiedzy na temat zachowań swoich klientów.



Do budowy zaawansowanych modeli analitycznych data mining wykorzystany został *STATISTICA Data Miner*. Zadania wdrożonego systemu to: segmentacja klientów, analiza efektywności programu lojalnościowego, badanie efektywności promocji i ich wpływu na zachowania klientów, planowanie kampanii marketingowych itp.

System pozwolił na wyodrębnienie homogenicznych grup klientów (segmentów), różniących się mobilnością, zachowaniami zakupowymi, lojalnością itp., dzięki czemu można prowadzić kampanie marketingowe skierowane do ściśle określonej grupy docelowej, ograniczyć koszty związane z prowadzonymi działaniami marketingowymi i jeszcze lepiej spełniać oczekiwania poszczególnych segmentów klientów. Zaawansowane modele analityczne zbudowane przy wykorzystaniu *STATISTICA Data Minera* pozwalają także przewidzieć efekty planowanych kampanii marketingowych przed ich uruchomieniem.

Wdrożenie systemu zostało uzupełnione przeprowadzeniem przez StatSoft Polska cyklu szkoleniowego, pozwalającego pracownikom PKN ORLEN na wykonywanie w systemie *STATISTICA* analiz ad-hoc, niezaplanowanych w systemie w postaci gotowych, automatycznie aktualizowanych raportów.

## **Netia**

W celu zwiększenia retencji klientów. Netia wybrała rozwiązanie firmy StatSoft Polska, które za pomocą najnowocześniejszych technik analizy danych zawartych w *STATISTICA Data Miner* umożliwiło budowę modeli prognostycznych wspierających działania związane z utrzymaniem klientów.

Dzięki opracowanym przez StatSoft modelom utrzymania zostały wygenerowane listy osób o najniższej skłonności do kontynuowania korzystania z usług, a także za pomocą metod analizy danych zidentyfikowane zostały czynniki istotnie wpływające na podjęcie przez klienta decyzji o zmianie operatora.

*StatSoft zaoferował nam rozwiązanie gwarantujące wysoką skuteczność. Wraz z zaostrzeniem konkurencji w branży telekomunikacyjnej coraz bardziej istotna staje się troska o klientów już pozyskanych. Staramy się dostosowywać ofertę do bieżących potrzeb klienta i dbać o to, by był zadowolony z naszych usług.*

*Remigiusz Siudziński  
Starszy Kierownik ds. Utrzymania Klientów  
Netia S.A.*

## **Fundacja Inicjatyw Menedżerskich**

*We wspólnej opinii ekspertów i konsultantów Fundacji STATISTICA to program oferujący procedury obliczeniowe o olbrzymiej precyzji i efektywności. Oprogramowanie STATISTICA jest wykorzystywane do planowania, zbierania, przetwarzania i obliczania danych statystycznych w zakresie zadań realizowanych w ramach projektu finansowanego ze środków Unii Europejskiej tj. analizy popytu na pracę (metoda heurystyczna), analizy szkolnictwa (modele statystyczno-ekonometryczne), badań prognoz gospodarczych (metody*



taksonometryczne) oraz opracowania analizy strategicznej w obszarze przekształceń ekonomiczno-społecznych.

Zakup oprogramowania STATISTICA możemy z czystym sumieniem polecić każdej instytucji zajmującej się analizą danych.

Anna Bielak  
Dyrektor Biura  
Fundacja Inicjatyw Menedżerskich

### **BP Polska Sp. z o.o.**

Firma BP Polska Sp. z o.o. współpracowała z firmą StatSoft Polska w ramach projektu związanego z analizą wybranego fragmentu transakcyjnej bazy danych BP partnerclub. Preferencje zakupowe Klientów są niezmiernie istotne w procesie przygotowania promocji i akcji marketingowych BTC zarówno w zakresie samego mechanizmu jak również dopasowania ATL/BTL i dlatego też BP postanowiło wykorzystać doświadczenie i wiedzę ekspertów ze StatSoft w zakresie analizy takiej bazy.

W ramach wspólnie przeprowadzonego projektu badawczego pracownicy firmy StatSoft wykonali statystyczne opracowanie 2% próby z bazy danych programu BP partnerclub przekazanej przez BP. Raport końcowy z badań przedstawiony został gronu osób zaangażowanych w marketing i sprzedaż w BP.

Profesjonalne podejście oraz propozycje i wsparcie udzielane w trakcie trwania projektu dowiodły, że Konsultanci StatSoft to wysokiej klasy fachowcy. Na podstawie naszych doświadczeń przy tym projekcie polecamy wszystkim zainteresowanym współpracę z Konsultantami firmy StatSoft.

Paweł Dziadkowiec  
Loyalty Manager  
BP Polska Sp. z o.o.

### **GlaxoSmithKline Pharmaceuticals**

Serdecznie dziękuję za niezwykle profesjonalne i interesujące przeprowadzenie szkolenia. To przyjemność współpracować z Wami - na pewno w najbliższej przyszłości w znacznie szerszym zakresie.

Inga Kleniewska-Wieczór  
Starszy Specjalista CRM  
GlaxoSmithKline Pharmaceuticals

### **Wydawnictwa Szkolne i Pedagogiczne**

Jestem bardzo zadowolony ze szkolenia, w którym brałem udział. Zarówno poziom merytoryczny, sposób prezentacji wiedzy, jak i organizacja kursu stały na wysokim poziomie.



*Wykładowcy wykazali wysoką umiejętność mówienia w sposób prosty i zrozumiały o rzeczach skomplikowanych i trudnych. Polecam kursy StatSoft Polska.*

*Jarosław Stawarz  
Specjalista d/s Badań i Analiz Rynkowych  
Wydawnictwa Szkolne i Pedagogiczne*

## **DEVK**

Firma Ubezpieczeniowa DEVK optymalizuje marketing bezpośredni, wykorzystując *STATISTICA Data Miner*. Dzięki temu został poprawiony wskaźnik liczby odpowiedzi na kampanie i uzyskano 6-cyfrowe oszczędności w budżecie.

Firma ubezpieczeniowa DEVK, jeden z liderów niemieckiego rynku ubezpieczeń, rozsyła rocznie kilka milionów listów do klientów z informacją o nowych produktach oraz z ofertami uzupełnienia istniejącej ochrony ubezpieczeniowej. Dla wzrostu efektywności takich akcji marketingowych bardzo duże znaczenie przy określaniu odbiorców kampanii ma (obok optymalizacji treści) identyfikacja odpowiednich grup klientów. Właśnie w tym celu zdecydowano się na zakup nowoczesnego narzędzia data mining. Głównymi kryteriami wyboru narzędzia były nie tylko stosunek cena - funkcjonalność i łatwa obsługa, lecz także prosta instalacja i wysoka jakość wsparcia technicznego do programu. System *STATISTICA Data Miner* spełnił najlepiej wszystkie wymagania.

*Natychmiastowa gotowość do pracy i łatwa eksploatacja STATISTICA Data Miner umożliwiły nam szybki postęp projektu.*

*Bruno Kuepper  
Specjalista ds. technik komunikacji  
DEVK*

Już na samym początku stosowania *STATISTICA Data Miner* uzyskano interesującą informację. Analizy przeprowadzone za pomocą drzew C&RT ukazały, że cecha, która do tej pory określała podobieństwo klientów na potrzeby marketingu bezpośredniego, w rzeczywistości nie umożliwiała ich różnicowania. Podczas dalszej analizy uzyskano inne ciekawe informacje na temat zachowania klientów.

Wyniki uzyskanie dzięki zastosowaniu nowych metod częściowo potwierdziły prawidłowości, których istnienia domyślano się, jednak przede wszystkim wskazały nowe, interesujące zależności.

Zróżnicowanie odpowiednich cech klientów zostało przełożone na wartość wskaźnika określającego prawdopodobieństwo odpowiedzi na ofertę. Z około 330.000 klientów wybrano 95 000 z najwyższym prawdopodobieństwem odpowiedzi. Po zrealizowaniu zaplanowanej wysyłki i ocenie jej efektywności stwierdzono, że w porównaniu z przeprowadzonym w poprzednim roku mailingiem uzyskano o 41% większą liczbę odpowiedzi.



Oszczędności DEVK wynoszą już przy jednej wysyłce więcej niż koszty licencji dla *STATISTICA Data Miner* na kilka lat.

### ***argonauten360°***

Niemiecka firma konsultingowa argonauten360° używa systemu *STATISTICA Data Miner* do tworzenia właściwych ofert produktów dostosowanych do potrzeb klientów.

argonauten360° jest wiodącą firmą konsultingową działającą w sektorze telekomunikacyjnym (oraz innych). Zastosowanie efektywnych i zaawansowanych technologii do skoringu klientów, analizy skupień oraz obliczenia wartości życiowej klienta należą do rutynowych, codziennych działań tej firmy. Wymagane są do tego elastyczne i potężne narzędzia analityczne. Istniejące zestawy narzędzi należało więc poszerzyć o najnowocześniejsze zaawansowane, a jednocześnie elastyczne, możliwości data mining. Innym krytycznym warunkiem narzuconym na rozwiązanie było uzyskanie szybkiego ROI - rozwiązanie musiało być łatwe w zastosowaniu i pozwalające szybko nauczyć się jego obsługi, tak by nawet najbardziej zaawansowane procedury analityczne mogły być szybko wykorzystywane przez analityków.

*STATISTICA Data Miner* to idealne połączenie cech, których oczekiwał klient. Decyzja zastosowania *STATISTICA* była więc jednogłówna, mimo że rozważane były też produkty innych producentów data mining - proste narzędzia analityczne, których już używano.

W Europie, wśród użytkowników telefonów komórkowych oraz linii tradycyjnych, popularne są usługi typu „Call-by-Call”. W tych planach opłaty abonamentowe za podstawową usługę są bardzo niskie lub zerowe, zaś rozliczenie tworzone jest na podstawie rzeczywistego czasu rozmów. Na tym polu istnieje bardzo silna konkurencja, a sukces dostawcy zależy w dużej mierze od atrakcyjności stawek za minutę rozmowy.

Z powodu silnej konkurencji panuje przekonanie, że na tym rynku nie ma praktycznie możliwości ruchów cenowych (dostawcy nie mogą zwiększyć marginesu zysku bez utraty klientów), a nawet jeśli taka możliwość istnieje to na pewno nie podlega predykcji. Konsultanci argonauten360° przeanalizowali dostępne dane w *STATISTICA Data Miner* i udowodnili, że powszechne przekonanie jest błędne! W efekcie argonauten360°, dzięki analizom uwieńczonym sukcesem, pozyskał wiodącego dostawcę usług typu Call-by-Call.

Przed stosowaniem modeli uzyskanych metodami data mining, wykonywano heurezy w postaci “opinii ekspertów” w celu oszacowania przewidywanego czasu połączeń (w minutach) na najbliższe dwa miesiące. Po zastosowaniu *STATISTICA Data Miner*, trafność prognoz znacznie się poprawiła, natomiast stopa błędów zmniejszyła się o połowę!

Za pomocą tego rozwiązania dostawca może obecnie przewidywać z większą trafnością zapotrzebowanie na czas rozmów na rynku bardzo czułym na ceny i oferować „właściwe” plany taryfowe. Jest to doskonały przykład udanego wdrożenia technologii data mining w celu uzyskania przewagi nad konkurencją w wysoce konkurencyjnym środowisku.



## **GfK Market Measures**

Agencja badawcza GfK wykorzystuje *STATISTICA* do opracowywania wyników badań rynku.

GfK Market Measures (GfK) oferuje usługi w zakresie analizy grup leków oraz narzędzia pozwalające mierzyć efektywność kampanii promocyjnych. GfK dostarcza również klientom raporty z okresowych badań rynku. Na podstawie informacji przechowywanych w repozytoriach danych GfK, które dotyczą przepisywanych leków, wykonywane są podsumowania i porównania potrzebne do przygotowania raportów. W związku z coraz większą liczbą wymaganych analiz oraz ich pracochłonnością pojawiła się potrzeba automatyzacji wykonywania analiz i samego procesu generowania raportów z badań rynku.

W tym celu w GfK Market Measures powstała specjalna aplikacja, umożliwiająca automatyzację przygotowywania raportów. Tworzenie raportu zawierającego wyniki analiz statystycznych stało się możliwe dzięki wdrożeniu oprogramowania *STATISTICA*. Aplikacja wykorzystuje model obiektowy *STATISTICA*, aby zautomatyzować wykonywanie testów statystycznych, których wyniki umieszczane są w raportach.

## **Rozwiązania StatSoft**

*STATISTICA* zapewnia kompletne rozwiązania w zakresie analizy danych – najlepsze narzędzia analityczne, systemy korporacyjne z elementami pracy grupowej i kontroli dostępu oraz możliwością zdalnej pracy przez Internet, mechanizmy dostępu do baz i hurtowni danych, systemy zbierania informacji z badań ankietowych oraz systemy rozpowszechniania informacji. Wybierając poszczególne produkty z rodziny *STATISTICA*, można zbudować rozwiązanie dostosowane do aktualnych potrzeb użytkowników.

*STATISTICA* spełnia wszystkie wymagania, jakie specjaliści zajmujący się marketingiem i badaniami rynku stawiają systemom analitycznym:

- ◆ dostarcza proste w obsłudze, intuicyjne środowisko pracy;
- ◆ zapewnia otwartość i elastyczność umożliwiającą wymianę danych pomiędzy różnymi aplikacjami, w tym MS Excel;
- ◆ zapewnia narzędzia do zbierania danych ankietowych przez Internet;
- ◆ udostępnia zarówno proste, jak i złożone narzędzia analizy danych w tym szeroki zakres analiz wielowymiarowych;
- ◆ ma ogromne możliwości wizualizacji danych (w tym z wykorzystaniem map), umożliwiające tworzenie atrakcyjnych graficznie raportów i zestawień;
- ◆ zawiera narzędzia pozwalające łatwo zautomatyzować rutynowo wykonywane prace z wykorzystaniem standardowego języka MS Visual Basic.



## ***Ogólne cechy programów z rodziny STATISTICA***

*STATISTICA* to rodzina programów wspomagających analizy danych wszelkiego rodzaju, od obliczania podstawowych charakterystyk i wizualizacji danych, poprzez modelowanie klasycznymi metodami statystycznymi, po wykorzystanie nowoczesnych technik optymalizujących pracę z bardzo dużymi zbiorami danych.

System analizy danych *STATISTICA* można zbudować, zestawiając takie narzędzie analityczne, jakie jest potrzebne do konkretnego przypadku: od zestawu do prostych zastosowań *STATISTICA Pakiet Podstawowy* po obszerne i zaawansowane narzędzie data mining *STATISTICA Data Miner*. *STATISTICA* to także możliwość zdalnej pracy przez internet, współpracy z intranetem oraz elektronicznym obiegiem dokumentów.

### **Środowisko pracy**

*STATISTICA* znana jest z prostego w użyciu, intuicyjnego, graficznego środowiska pracy. Można je z łatwością dostosować do własnych potrzeb i preferencji. Możliwe jest również definiowanie całkowicie nowego środowiska, w tym także interfejsów bazujących na przeglądarce internetowej. Wysokiej jakości narzędzia wizualizacji wyników analiz umożliwiają czytelną prezentację wyników i danych. Na każdym etapie analizy dostępny jest „pod ręką” zestaw charakterystycznych dla danej metody wykresów analitycznych, które umożliwiają przedstawienie w przejrzysty sposób wyników analizy. Oprócz wykresów wyników dostępny jest również bogaty zestaw wykresów prezentujących zgromadzone dane.

### **Architektura**

Podstawowa architektura rozwiązania, niezależnie od wybranego zestawu technik analitycznych (zob. *Narzędzia analityczne*, poniżej), to wersja sieciowa concurrent. Zapewnia ona efektywne wykorzystanie licencji – limitowana jest liczba użytkowników w danym momencie, a nie całkowita liczba użytkowników (oprogramowanie można zainstalować na dowolnej liczbie stacji roboczych połączonych z serwerem). *STATISTICA* w wersji concurrent jest programem pozwalającym korzystać jednocześnie z systemu wielu użytkownikom, pracującym w lokalnej sieci Windows.

Dla instytucji posiadających mniejsze wymagania w zakresie korzystania z rozwiązania alternatywnie proponujemy wersję jednostanowiskową. Można ją zainstalować i wykorzystywać na jednym komputerze.

Firmom o większych wymaganiach proponujemy wersje korporacyjne: *STATISTICA Enterprise* oraz *WebSTATISTICA Server* (zob. poniżej). Łącząc wszystkie lub wybrane narzędzia analityczne z jednym z systemów korporacyjnych, otrzymujemy rozwiązanie, które nie tylko pozwala na korzystanie z odpowiednich narzędzi analitycznych przez wielu użytkowników, ale także organizuje wspólną pracę różnych osób. W szczególności systemy korporacyjne *STATISTICA* pozwalają wielu użytkownikom korzystać ze wspólnych szablonów zapytań do baz danych, analiz, raportów przechowywanych w jednym miejscu z zapewnieniem kontroli dostępu itp.



## STATISTICA Enterprise

*STATISTICA Enterprise* to system analizy danych ogólnego przeznaczenia, m.in. dla zastosowań typu *business intelligence* w marketingu, finansach oraz w innych dziedzinach. Jest to kompleksowy system analityczno-raportujący. Użytkownicy *STATISTICA Enterprise* mogą korzystać ze wspólnych, nawet bardzo złożonych zapytań, pozwalających na przeglądanie wybranych danych z repozytoriów centralnych, oraz współużytkowanie centralnie uaktualnianych skryptów analiz (np. predefiniowane raporty dostosowywane przez administratorów lub analityków). *STATISTICA Enterprise* ma własny system użytkowników i grup z możliwością zarządzania uprawnieniami do zasobów systemu, narzędzie do definiowania szablonów raportów oraz generator raportów (raporty generowane automatycznie co pewien czas lub o określonym czasie).

*STATISTICA Enterprise* zorganizowany jest wokół centralnej hurtowni danych, którą można zainstalować na wszystkich popularnych systemach zarządzania bazami danych, w tym na popularnych skalowalnych systemach takich jak Sybase, Oracle, MS SQL Server czy IBM DB2.

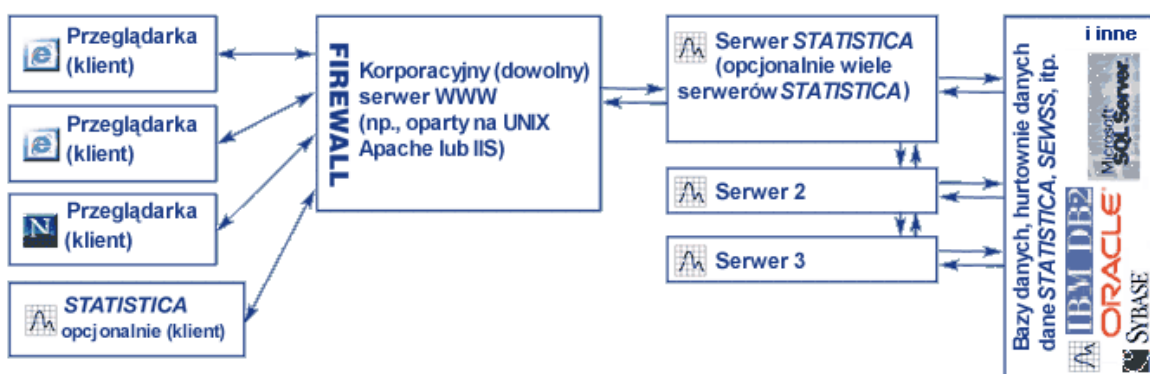
Wyniki pracy mogą być udostępniane w środowisku lokalnym (innym użytkownikom *STATISTICA Enterprise* posiadającym odpowiednie uprawnienia) oraz rozpowszechniane w szerszym zakresie (np. jako raporty publikowane w Internecie lub Intranecie).



## WebSTATISTICA Server

jest w pełni skalowalnym internetowym systemem udostępniania i analizy danych w skali całej organizacji. Umożliwia on zdalne uruchamianie *STATISTICA* w ramach przeglądarki internetowej (wszystkie obliczenia są wykonywane na serwerze), a także szybki i łatwy dostęp do danych, raportów i narzędzi analitycznych z praktycznie każdego komputera połączony z Internetem. Pozwala to na współpracę osób z tego samego biura, jak i z różnych kontyentów.

System ten w pełni wykorzystuje wielowarstwową architekturę klient-serwer. *WebSTATISTICA* wykorzystuje przetwarzanie rozproszone i wielowątkowość, co pozwala optymalnie wykonywać bardzo duże zadania obliczeniowe. Technologia ta umożliwia szybkie wykonanie nawet bardzo dużych projektów, z pełnym wykorzystaniem wielu procesorów serwera lub wielu serwerów pracujących równolegle.





Z systemu można korzystać z każdego komputera na świecie połączonego z Internetem, wyposażonego w przeglądarkę internetową. Nie ma także ograniczeń związanych z systemem bazy danych i platformą, na której on pracuje. *WebSTATISTICA Server* może pracować na danych pochodzących ze wszystkich popularnych systemów baz danych.

## Dane

*STATISTICA* w każdej wersji zawiera *STATISTICA Query*, czyli narzędzie, które umożliwia łatwy dostęp do danych gromadzonych w różnego typu bazach danych (np. Sybase, Oracle, MS SQL Server itp.) za pomocą standardu OLE DB lub ODBC. *STATISTICA* pozwala szybko i w prosty sposób tworzyć rozbudowane zapytania bez znajomości języka SQL.

*STATISTICA* pozwala także na pracę z plikami danych z rozmaitych aplikacji, np. MS Excel, pliki tekstowe, .xml i wiele innych.

*STATISTICA* jest zoptymalizowana pod kątem pracy z dużymi plikami danych. Dane przechowywane są w arkuszach o praktycznie nieograniczonej pojemności. Nie ma również ograniczeń co do zawartości poszczególnych komórek arkusza, można do nich dołączać pliki dźwiękowe, wideo, wykresy, animacje oraz inne obiekty zgodne z ActiveX. Dostępny opcjonalnie mechanizm IDP pozwala także na pracę z danymi (zgromadzonymi np. w bazie czy hurtowni danych) bez konieczności tworzenia ich lokalnej kopii.

## Gromadzenie i analiza danych ankietowych

W ofercie StatSoft dostępne są także dedykowane narzędzia, które wspierają prowadzenie badań ankietowych.

System wprowadzania i gromadzenia danych ankietowych umożliwia wprowadzanie danych z uzupełnionych ankiet lub ich wypełnianie on-line za pośrednictwem komputera wyposażonego w łącze internetowe i przeglądarkę. Dane te są przechowywane w bazie danych będącej częścią systemu, specjalnie zaprojektowanej, zoptymalizowanej pod kątem wykonywania analiz i zestawień. Przed każdym nowym badaniem, oprócz wprowadzenia do systemu schematu odpowiedniej ankiety, określamy listę uprawnionych użytkowników, dzięki czemu kontrolujemy dostęp odpowiednich osób do ankiet, a także uzyskujemy informację o jakości danych wprowadzonych przez poszczególnych ankietowanych. Dzięki wdrożeniu systemu w jednym miejscu tworzona jest składnica informacji, gromadząca wyniki wszystkich przeprowadzanych badań ankietowych, do których będzie można w każdym momencie sięgnąć, a centralne składowanie danych zapewnia zgodność uzyskanych wyników.

Zaprojektowany przez StatSoft system wprowadzania i gromadzenia danych ankietowych jest z powodzeniem stosowany m.in. przez administrację publiczną do realizacji dużych, ogólnopolskich projektów badawczych, w których gromadzone są informacje z kilkudziesięciu tysięcy ankiet.



## Automatyzacja

Ponadto aplikacja ma wbudowane środowisko programistyczne *STATISTICA Visual Basic* (zgodne ze standardowym MS Visual Basic), które w pełni wykorzystuje zalety modelu obiektowego *STATISTICA* i umożliwia programowy dostęp do praktycznie każdego elementu i każdej funkcji programu.

## Otwartość

Otwarta architektura pozwala na integrację *STATISTICA* z praktycznie dowolnym systemem gromadzenia danych, StatSoft oferuje budowę dedykowanych systemów analitycznych i ich integrację z istniejącymi systemami.

## Prezentacja wyników

W trakcie przeprowadzania analiz w *STATISTICA* na wyjściu otrzymujemy wyniki w postaci multimedialnych tabel (arkuszy) i wykresów. Tabele wynikowe mogą być bezpośrednio wykorzystane jako źródło danych dla kolejnych analiz.

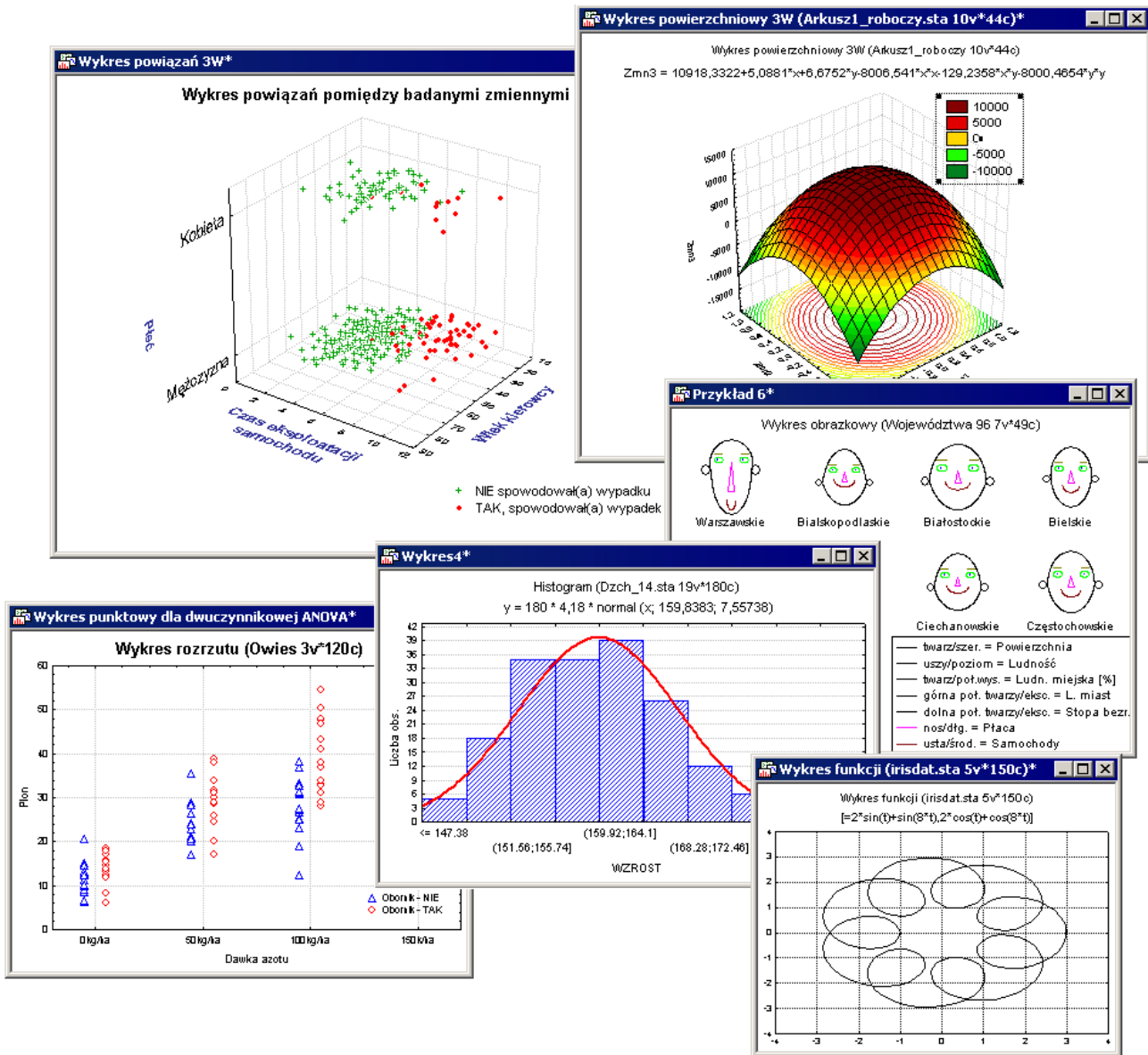
*STATISTICA* od lat jest bezkonkurencyjna wśród narzędzi do statystycznej analizy danych i data mining pod względem narzędzi wizualizacji danych i tworzenia wykresów. *STATISTICA* oferuje setki typów dwuwymiarowych i trójwymiarowych wykresów oraz wygodne narzędzia do projektowania nowych typów własnych wykresów i dodawania ich do menu lub pasków narzędzi. Oprócz ogólnych typów wykresów istnieje wiele specjalistycznych wykresów, zintegrowanych ze specjalistycznymi procedurami analitycznymi i dostępnych w odpowiednich oknach wyboru wyników lub w podręcznych menu. Praktycznie każdy szczegół wyglądu wykresu (a są to setki cech) może być ustawiany przez użytkownika. Wszystkie wykresy utworzone na podstawie arkusza danych mogą być automatycznie odświeżane przy każdej modyfikacji danych lub na życzenie użytkownika.

Raport *STATISTICA* przypomina dokument edytora tekstu, a poszczególne obiekty (np. wykresy, arkusze, arkusz czy wykres MS Excel) umieszczane są w nim kolejno, jeden za drugim. Wszystkie raporty mogą być zapisywane nie tylko we własnym formacie *STATISTICA*, ale także w postaci plików DOC, RTF, HTML i PDF.

## Najwyższej jakości grafika

Dostępne w każdym produkcie z rodziny *STATISTICA* wysokiej jakości narzędzia wizualizacji wyników analiz umożliwiają czytelną prezentację wyników i danych.

Na każdym etapie analizy mamy „pod ręką” zestaw charakterystycznych dla danej metody wykresów analitycznych, które umożliwiają przedstawienie w przejrzysty sposób wyników danej analizy. Oprócz wykresów wynikowych mamy również do wyboru bogaty zestaw wykresów prezentujących zgromadzone dane. Dodatkowo dostępne są (w postaci dodatku, który można pobrać ze stron internetowych StatSoft) narzędzia do wizualizacji danych z wykorzystaniem map.



## Skalowalność

Rozwiązania z rodziny *STATISTICA* są skalowalne pod każdym względem:

- ◆ liczby użytkowników – od wersji jednostanowiskowych po korporacyjne systemy wieloużytkownikowe z kontrolą dostępu i pracą przez Internet,
- ◆ wielkości przetwarzanych danych – *STATISTICA* jest zoptymalizowana pod kątem szybkości wykonywanych analiz, specjalna technologia IDP pozwala na analizę ekstremalnie dużych zbiorów danych bez konieczności tworzenia lokalnej kopii całego analizowanego zbioru,
- ◆ zakresu procedur analitycznych – w każdym momencie można rozszerzyć licencję o dodatkowe metody analityczne, cały czas pracując w tym samym środowisku, bez konieczności zmiany interfejsu użytkownika i związanych z nim przyzwyczajzeń.



## Przetwarzanie równoległe i rozproszone

*STATISTICA* w wersji klient-serwer obsługuje przetwarzanie równoległe i rozproszone po stronie serwera. Technologia w pełni równoległego przetwarzania rozdziela zadania nie tylko pomiędzy wiele procesorów serwera, ale również na wiele komputerów, co radykalnie zwiększa wydajność w projektach o wysokich wymaganiach obliczeniowych, czy takich, w których wykorzystywane są wielkie zbiory danych.

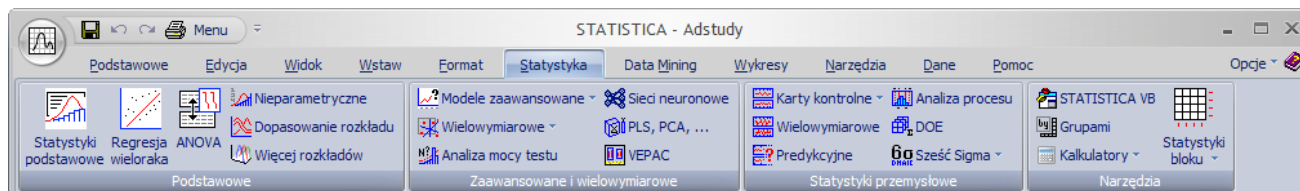
## Programowalność i otwarta architektura

Każdy z programów z rodziny *STATISTICA* ma wbudowane środowisko programistyczne *STATISTICA Visual Basic* (zgodne ze standardowym MS Visual Basic), które w pełni wykorzystuje zalety modelu obiektowego *STATISTICA* i umożliwia programowy dostęp do praktycznie każdego elementu i każdej funkcji programu. *STATISTICA Visual Basic* rozszerza standardową składnię języka Visual Basic o ponad 11 tysięcy nowych funkcji (w wersji 8 *STATISTICA* dodane zostały kolejne procedury), co daje bardzo rozbudowane i bogate środowisko programistyczne. Umożliwia ono tworzenie specjalizowanych systemów analitycznych, które łączą potężne możliwości zoptymalizowanych procedur *STATISTICA* z różnymi rozszerzeniami i własnym interfejsem. Środowisko *STATISTICA Visual Basic* oferuje programiście elastyczny edytor i debugger, edytor okien dialogowych, a także przeglądarki dostępnych funkcji i obiektów.

Programy *STATISTICA Visual Basic* można uruchamiać nie tylko w *STATISTICA*, lecz (dzięki zgodności ze standardami) także w innych środowiskach bazujących na konwencjach Microsoft Visual Basic (np. MS Excel, MS Word lub w samodzielnych instalacjach Visual Basic). Należy zauważyć, że ta bogata biblioteka funkcji *STATISTICA* jest dostępna nie tylko dla aplikacji w języku Visual Basic (wbudowanego w program lub też innego), lecz również z dowolnego innego zgodnego środowiska programistycznego (np. C/C++, Java, Delphi).

## Narzędzia analityczne

Każda z aplikacji analitycznych może pracować zarówno jako wersja jednostanowiskowa (na pojedynczym komputerze), w sieci (na wielu komputerach użytkowników), jak i może być elementem korporacyjnego systemu analizy danych.



Zakres metod analitycznych oferowanych przez *STATISTICA* jest niezwykle bogaty. Metody analityczne zostały podzielone na różne programy, z których można zbudować odpowiedni zestaw analityczny. Poszczególne programy z rodziny *STATISTICA* można łączyć w jednym pakiecie, który może z czasem być rozszerzany o dodatkowe moduły. Dzięki

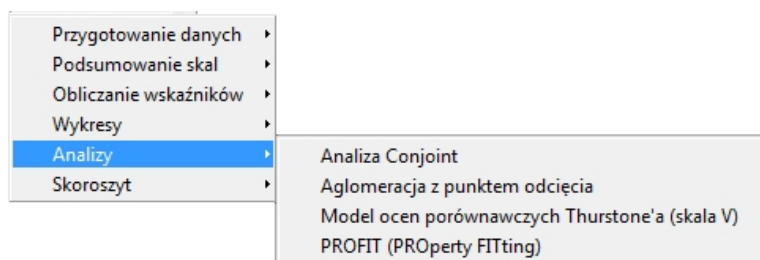


temu możemy mieć dostęp do bardzo wielu różnych metod analitycznych zawartych w jednej aplikacji *STATISTICA*.

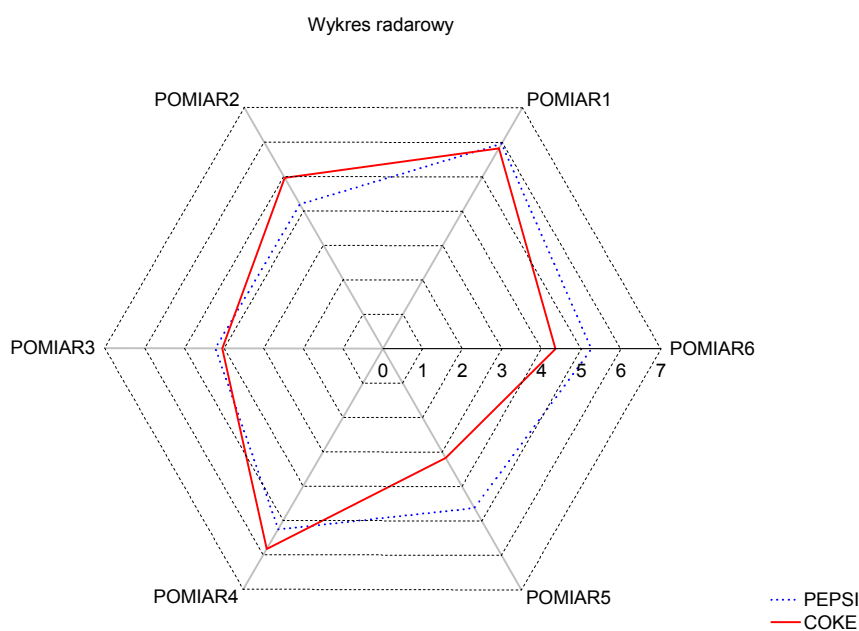
Poniżej znajdują się opisy wybranych aplikacji *STATISTICA*, najczęściej stosowanych w marketingu i badaniach rynku. Opisy wszystkich programów z rodziny *STATISTICA* dostępne są na stronie internetowej [www.StatSoft.pl](http://www.StatSoft.pl).

### ***STATISTICA dla analiz marketingowych i rynkowych***

*STATISTICA dla analiz marketingowych i rynkowych* jest programem ułatwiającym i przyspieszającym proces opracowania wyników badań, ze szczególnym uwzględnieniem badań ankietowych. Zawiera wszystkie moduły analityczne oraz narzędzia obróbki danych zawarte w *STATISTICA Pakiet Zaawansowany* (zob. poniżej), a także dodatkowe funkcje przydatne podczas przygotowania danych do specyfiki analiz pochodzących z badań marketingowych. Za jego pomocą możliwe jest szybkie rekodowanie danych, przekodowanie na zmienne sztuczne, zliczanie wystąpień określonego zakresu wartości w wybranych zmiennych. Dostępne są także zaawansowane narzędzia do sprawdzania poprawności danych.

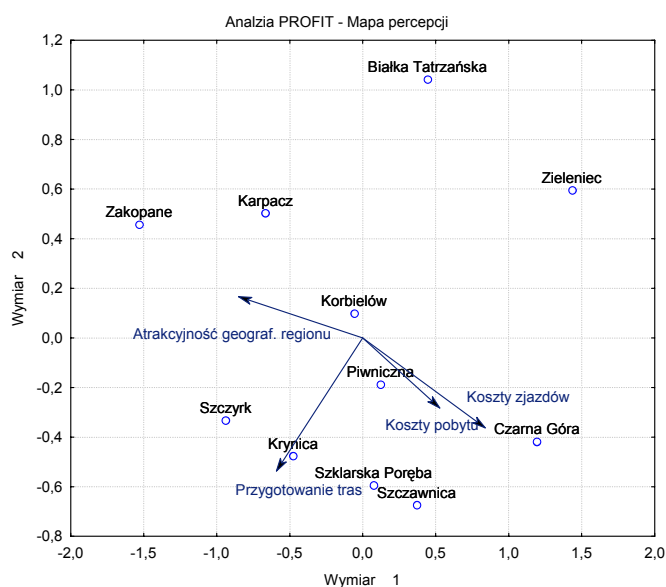
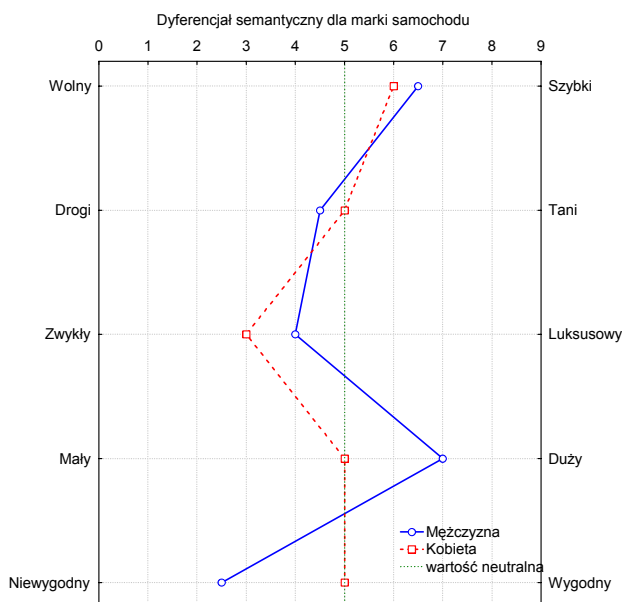


Badacze za pomocą jednej opcji mogą uzyskać raporty podsumowujące skale: pozycyjne, rangowe, semantyczne (generowany jest wykres dyferencjału semantycznego) oraz Stappela. Dodatkowo duży wybór wykresów zawarty w *STATISTICA Pakiet Zaawansowany* uzupełniony został o wykres radarowy oraz możliwość generowania krzywej Lorenza.





*STATISTICA dla analiz marketingowych i rynkowych* zawiera także zbiór zaawansowanych narzędzi umożliwiających wykonanie analizy conjoint, zbudowanie modelu ocen porównawczych Thurstone'a (skala V), wykonanie mapy percepcji dla różnych reprezentacji danych wraz z uwzględnieniem dodatkowych wymiarów (analiza PROFIT), a także przeprowadzenie grupowania metodą aglomeracyjną z możliwością wyboru optymalnego punktu odcięcia.

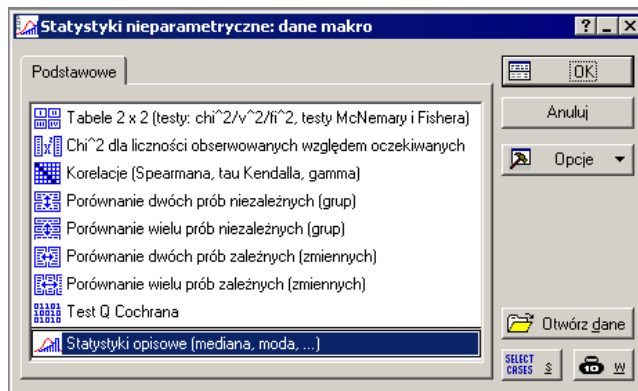
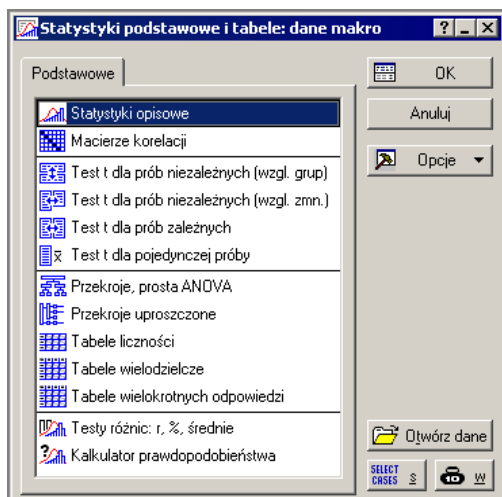


Zestaw uzupełniają narzędzia ułatwiające publikację uzyskanych wyników w raporcie (publikacja w MS Word czy też zapis całego skoroszytu do Excela).



### **STATISTICA Pakiet Podstawowy PL**

*STATISTICA Pakiet Podstawowy* to obszerny zestaw najpopularniejszych metod statystycznych, obejmujący: *Statystyki opisowe, Tabele liczebności, Przekroje, Tabele wielodzienne, Korelacje, Testy parametryczne, ANOVA, Analiza regresji, Dopasowanie rozkładów, Testy nieparametryczne i Kalkulator prawdopodobieństwa.*





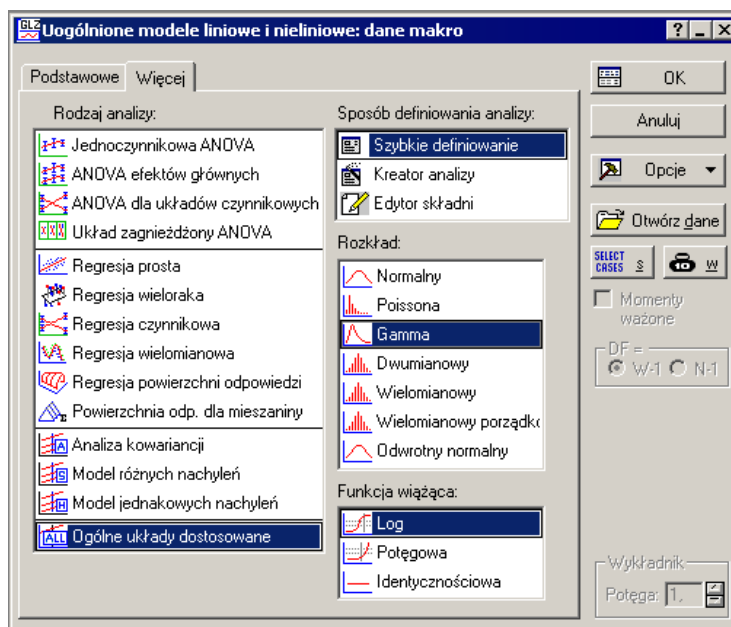
Zawiera pełne środowisko *STATISTICA*: modyfikowalny interfejs, elastyczne narzędzia zarządzania wynikami i raportowania, interakcyjne narzędzia do tworzenia zapytań do baz danych (*STATISTICA Query*), wszechstronne narzędzia importu i eksportu danych, komplet narzędzi wizualizacji danych i wyników analiz (wysokiej jakości wykresów) oraz w pełni zintegrowany język programowania *STATISTICA Visual Basic*.

### **STATISTICA Pakiet Zaawansowany PL**

*STATISTICA Pakiet Zaawansowany* zawiera wszystkie funkcjonalności programu *STATISTICA Pakiet Podstawowy*, a także moduły *Modele Zaawansowane*, *Analizy Wielowymiarowe* oraz *Analiza Mocy Testów*.



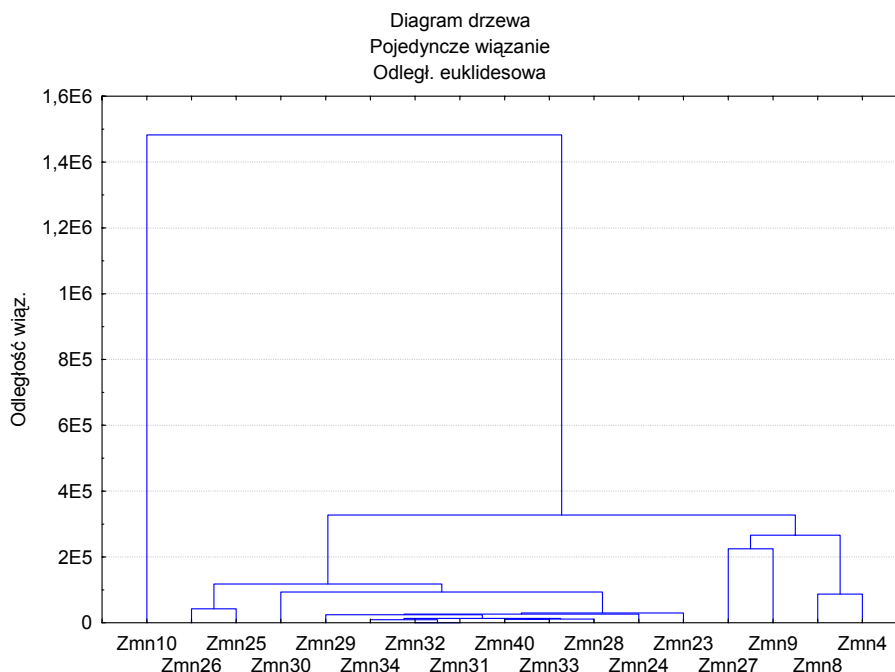
*Modele Zaawansowane* to szeroki wybór zaawansowanych narzędzi do budowy modeli liniowych i nieliniowych. Umożliwiają one wykorzystanie predyktorów ciągłych i jakościowych, uwzględnianie interakcji, tworzenie modeli hierarchicznych oraz automatyczny wybór modeli. Ponadto pozwalają obliczać komponenty wariancyjne, analizować dane o postaci szeregów czasowych oraz stosować wiele innych metod. Moduł obejmuje następujące grupy procedur: *Ogólne modele liniowe (GLM)*, *Uogólnione modele liniowe i nieliniowe*, *Ogólne modele regresji*, *Szeregi czasowe i prognozowanie*, *Modele cząstkowych najmniejszych kwadratów*, *Algorytm NIPALS (PCA/PLS)*, *Komponenty wariancyjne i model mieszany ANOVA/ANCOVA*, *Analiza czasu przeżycia i awaryjności*, *Estymacja nieliniowa*, *Linearyzowana regresja liniowa*, *Analiza log-liniowa tabel liczebności*, *Modelowanie równań strukturalnych i analiza ścieżkowa (SEPATH)*.



*Analizy Wielowymiarowe* to bogaty zestaw technik eksploracji danych różnego typu, począwszy od analizy skupień, a skończywszy na wybranych metodach drzew klasyfikacyjnych, wraz z obszernym zestawem interakcyjnych narzędzi wizualizacji, służących do badania współzależności i występowania regularnych układów danych. Moduł obejmuje



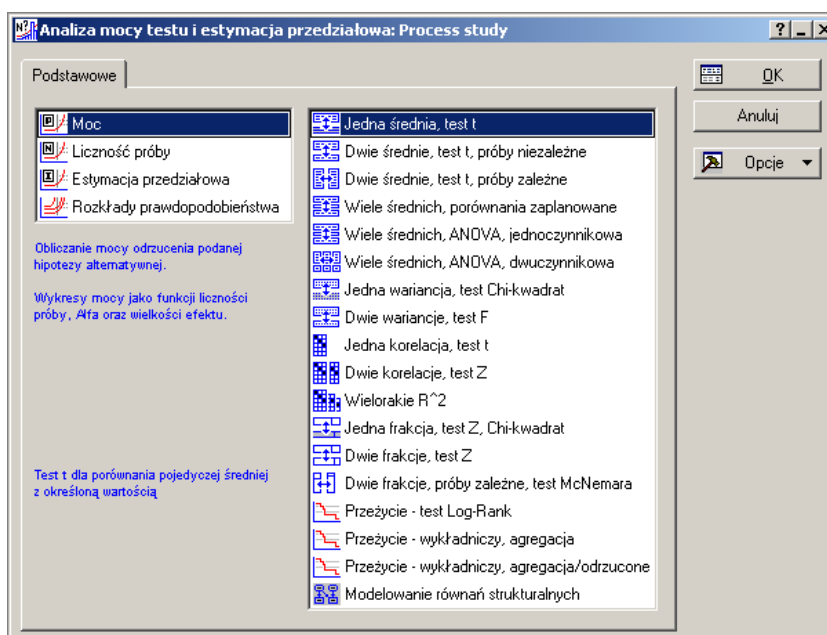
następujące grupy procedur: *Analiza skupień*, *Analiza czynnikowa*, *Analiza składowych głównych i klasyfikacja*, *Analiza kanoniczna*, *Analiza rzetelności i pozycji*, *Drzewa klasyfikacyjne*, *Analiza korespondencji*, *Skalowanie wielowymiarowe*, *Analiza dyskryminacyjna*, *Ogólne modele analizy dyskryminacyjnej (GDA)*.



Przedstawiony powyżej wykres przedstawia przykładowy diagram drzewa będący wynikiem *Analizy skupień*.



*Analiza Mocy Testów* to specjalistyczne, a jednocześnie przyjazne dla użytkownika narzędzie do analizowania wszelkich aspektów mocy testów statystycznych i wyznaczania liczności próby.





Moduł ten obejmuje obszerny zestaw narzędzi ułatwiających takie zaplanowanie badań, aby wielkość próby odpowiadała ich celowi, a także analizy związane z obliczaniem mocy statystycznej i wielkości próby. *Analiza Mocy Testów* wychodzi poza standardowe testy „braku różnic” i wykorzystuje także techniki bazujące na estymacji przedziałowej.

### ***Zestaw Skoringowy STATISTICA PL***

*Zestaw Skoringowy STATISTICA* jest rozwiązaniem wspierającym proces budowy, oceny i utrzymania modeli i kart skoringowych stosowanym do różnego rodzaju skoringów. Modele skoringowe są jednym z najpopularniejszych rodzajów modeli statystycznych wykorzystywanych w biznesie. Są wykorzystywane wszędzie tam, gdzie z jednej strony dysponujemy odpowiednią ilością danych na temat klientów, a z drugiej strony istnieje potrzeba optymalizacji podejmowanych w ich kierunku działań. Modele skoringowe wykorzystywane są w instytucjach finansowych i ubezpieczeniowych, firmach telekomunikacyjnych, sprzedaży detalicznej, a nawet w medycynie.

Modele skoringowe zbudowane w *Zestawie Skoringowym STATISTICA* wykorzystuje się np. w procesie utrzymania klienta, gdzie wskazują osoby najbardziej zagrożone odejściem. Kolejny obszar zastosowań *Zestawu Skoringowego STATISTICA* to wsparcie procesu sprzedaży - za ich pomocą można optymalizować ofertę sprzedaży dla konkretnej grupy klientów. Model skoringowy potrafi wskazać osoby, które z największym prawdopodobieństwem odpowiedzą na ofertę poszczególnych produktów. Inne rodzaje zadań, w których stosowany jest *Zestaw Skoringowy STATISTICA*, to skoring kredytowy, skoring wyłudzeń (antyfraud), skoring windykacyjny oraz skoring medyczny. Krótko mówiąc, skoring sprawdza się, kiedy chcemy podzielić klientów na dwie kategorie: spłaci kredyt/nie spłaci, odpowie na ofertę/nie odpowie, przyniesie zysk/nie będzie dochodowy, zagrożony odejściem/pozostanie klientem.

Analizy danych, które wykonywać można w *Zestawie Skoringowym STATISTICA*, można podzielić na następujące etapy:

1. *Wybór zmiennych do analizy* umożliwia wykonanie rankingu istotności cech i selekcję jedynie tych, które są istotnie powiązane z badanym zjawiskiem. Dodatkowo istnieje możliwość identyfikacji grup skorelowanych ze sobą zmiennych i wybór ich reprezentantów.
2. *Dyskretyzacja zmiennych* umożliwia automatyczne bądź interakcyjne przygotowanie przedziałów analizowanych zmiennych jednorodnych ze względu na badane zjawisko. Moduł pozwala uwzględnić w analizie braki danych oraz obserwacje nietypowe.
3. *Budowa modeli i tworzenie tablicy skoringowej*. Modele skoringowe można budować przy użyciu szeregu mniej lub bardziej zaawansowanych metod statystycznych i data mining. Do najpopularniejszych należą: regresja logistyczna, drzewa klasyfikacyjne oraz różne „odmiany” drzew – drzewa wzmacniane (*boosting trees*) czy losowy las (*random forest*). Wykorzystywać można również sieci neuronowe, metodę wektorów nośnych (*support vector machines*) czy k-najbliższych sąsiadów i szereg innych.

Na podstawie przygotowanych danych model skoringowy może być zbudowany automatycznie za pomocą regresji logistycznej, z uwzględnieniem zaawansowanych



strategii doboru najlepszego podzbioru predyktorów. Można także budować modele skoringowe typu SURVIVAL (z wykorzystaniem analizy przeżycia za pomocą proporcjonalnego hazardu Coxa), które pozwalają nie tylko określić prawdopodobieństwo zajścia danego zdarzenia, ale również czas, w jakim osiągnie ono dany (krytyczny) poziom – np. kiedy dana osoba odejdzie do konkurencji lub przestanie spłacać kredyt. W przypadku budowy modeli „małej próby” możliwe jest wykorzystanie techniki bootstrap.

Model zbudowany z wykorzystaniem regresji logistycznej możemy również zamienić jednym kliknięciem na tablicę skoringową (wykorzystywaną powszechnie w skoringu kredytowym), którą następnie można zapisać w postaci pliku Excela, XML i innych formatów.

4. *Ocena modeli*: w tym module możemy ocenić zarówno tablice skoringowe, jak również modele zbudowane za pomocą innych technik (np. drzewa wzmacniane, sieci neuronowe). Dostępnych jest szereg miar jakości modeli, takich jak: IV (*information value*), statystyka KS (Kolmogorow-Smirnow), statystyka Hosmera-Lemeshowa, dywergencja, indeks Giniego, oraz wykresów: krzywa ROC, wykres przyrostu (*Lift*), wykres zysku (*Gain*).
5. *Obliczanie skoringu* daje możliwość obliczania punktacji skoringowej oraz prawdopodobieństwa dla nowych danych na podstawie zapisanego modelu.
6. *Zarządzane punktem odcięcia* umożliwia wybór optymalnego punktu odcięcia odpowiedzi modelu (możliwość wyboru do 3 różnych punktów odcięcia). Podjęcie optymalnej decyzji jest wspomagane przez zestaw wykresów i raportów, a także możliwość uwzględnienia czynnika finansowego (jaki punkt odcięcia jest najbardziej opłacalny).

Poza powyższymi modułami ogólnego przeznaczenia użytkownicy zajmujący się skoringiem kredytowym mogą skorzystać z dodatkowych modułów pozwalających na dodatkowe analizy związane z monitorowaniem portfela kredytowego (analiza *Vintage*) badaniem stabilności populacji i cech oraz analizą wniosków odrzuconych (*reject inference*).

Powyższy zakres *Zestawu Skoringowego STATISTICA* można w razie indywidualnych potrzeb i wymagań rozszerzyć i dostosować.



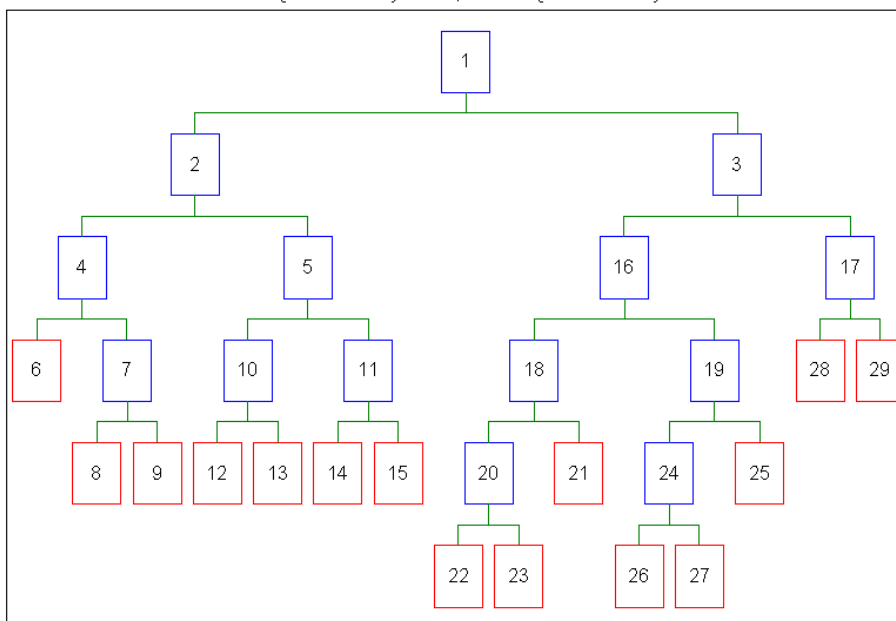
### **STATISTICA Trees PL**

Program *STATISTICA Trees* jest doskonały dla analityków chcących przeprowadzać analizy data mining przy wykorzystaniu modeli drzew decyzyjnych. W programie tym znajduje się część metod dostępnych także w *STATISTICA Data Miner*. Są to: *Ogólne modele drzew klasyfikacyjnych i regresyjnych (G-Trees)*, *Ogólne modele CHAID*, *drzewa interakcyjne i wzmacniane*. *STATISTICA Trees* umożliwia wykonywanie analiz z zastosowaniem algorytmu budowy drzew, które prowadzą do znalezienia zbioru logicznych warunków podziału typu „jeżeli, to”, pomagają jednoznacznie zaklasyfikować badane obiekty. Mechanizm eksploracji drzewa jest bardzo przydatny przy szczegółowej analizie wyników.



Układ drzewa 1 dla Zmn2

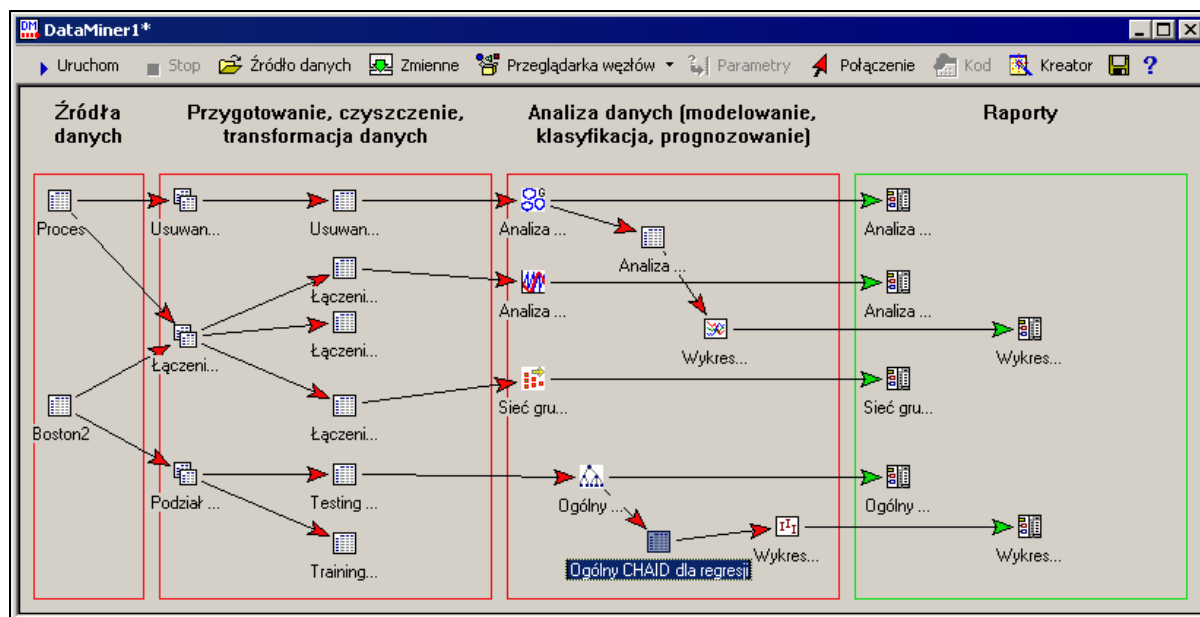
Liczba węzłów dzielonych: 14, liczba węzłów końcowych: 15

**STATISTICA Data Miner PL**

*STATISTICA Data Miner* to kompletny zestaw narzędzi data mining, zaprojektowany tak, aby umożliwić łatwe i szybkie wykonywanie analiz danych i stosowanie wyników we wspomaganie podejmowania decyzji. *STATISTICA Data Miner* jest łatwym w użyciu, a jednocześnie bardzo wydajnym narzędziem do realizacji wszystkich etapów wydobywania z danych użytecznej wiedzy: począwszy od pobierania danych z baz i hurtowni danych, poprzez etap czyszczenia i transformacji danych, wykonywanie analiz danych, a na tworzeniu raportów skończywszy.

System *STATISTICA Data Miner* jest wyposażony w bardzo bogaty wybór metod analitycznych, zarówno wbudowanych w gotowe projekty data mining (szeroki wybór predefiniowanych, gotowych do użycia projektów dla różnego typu zagadnień), jak i dostępnych do budowy własnych projektów. *STATISTICA Data Miner* zawiera pełną funkcjonalność programów: *STATISTICA Pakiet Podstawowy*, *Modele Zaawansowane*, *Analizy Wielowymiarowe*, *Automatyczne Sieci Neuronowe* (z generatorem kodu C), *Trees* i *Analiza Mocy Testów*, składnik IDP, a przede wszystkim kilkanaście specjalnie zaprojektowanych modułów data mining.

Specjalne środowisko graficzne tworzenia projektów analitycznych (patrz poniżej) bazuje na konwencji *przeciągnij i upuść*, dzięki czemu tworzenie projektów data mining jest intuicyjne, a praca z systemem nie wymaga długich przygotowań. Sprawia to, że z systemu *STATISTICA Data Miner* korzystać mogą nawet nowicjusze. Jest on jednocześnie wysoce elastyczny i dostosowywalny. Jak widać poniżej, obszar roboczy w *STATISTICA Data Miner* podzielony jest na cztery sekcje. Podział ten odpowiada czterem etapom projektów data mining.



W jednym projekcie data mining mamy możliwość stosowania różnych technik i modeli. Po opracowaniu gotowych modeli *STATISTICA Data Miner* oferuje elastyczne narzędzia do wdrażania i stosowania modeli, zintegrowane ze środowiskiem tworzenia projektów, które umożliwia sterowanie przetwarzaniem danych w obiektach analizy (węzłach) za pomocą poręcznych, zgodnych ze standardami przemysłu informatycznego skryptów *STATISTICA Visual Basic*.

Dane przetwarzane w *STATISTICA* mogą pochodzić z wielu różnych źródeł. System *STATISTICA Data Miner* cechuje łatwość w korzystaniu z danych zewnętrznych. Mamy możliwość podłączenia różnych źródeł danych, tak że jakakolwiek zmiana danych na wejściu powoduje przeliczenie całego projektu, tak aby wyniki odpowiadały aktualnym danym.

Wielkość danych przetwarzanych przez systemy *STATISTICA* jest praktycznie nieograniczona. Specjalne techniki data mining w *STATISTICA Data Miner* są zoptymalizowane pod kątem przetwarzania olbrzymich zbiorów danych.

W jednym projekcie data mining możemy używać różnych technik i modeli. Stosowanie modeli dla nowych danych jest w *STATISTICA Data Miner* bardzo wydajne, dzięki zastosowaniu przenaszalnego, bazującego na XML, języka PMML (*Predictive Models Markup Language*), który można stosować zarówno dla modeli klasyfikacyjnych, jak i regresyjnych. Raz oszacowane modele mogą być później wielokrotnie stosowane dla nowych danych. Możemy także na bieżąco modyfikować ich parametry w miarę napływu danych. Oprócz języka PMML projekty i szacowane modele możemy zapisywać w postaci kodu w językach programowania SVB (zgodnym ze standardowym Visual Basic) oraz C/C++. Dzięki temu analizy przeprowadzane w *STATISTICA Data Miner* mogą być wykorzystywane także w innych aplikacjach. Najczęstszym jednak zastosowaniem dla modeli zapisywanych w tych różnych językach jest ich ponowne wykorzystanie w projektach data mining w *STATISTICA Data Miner*. Zapisane wcześniej analizy mogą być wstawiane do przestrzeni projektowej jako nowe węzły i łączone z pozostałymi elementami projektu.



*STATISTICA Data Miner* zawiera również *Przepisy Data Miner* – kompletne rozwiązanie, dzięki któremu tworzenie modeli predykcyjnych staje się systematycznym procesem wykonywanym krok po kroku według ustalonej procedury. Celem *Przepisów Data Miner* jest umożliwienie łatwego i szybkiego wydobywania z danych użytecznych informacji. W *Przepisach* użytkownik tworzy model, wykonując krok po kroku pewien ogólny przepis obejmujący przygotowanie i oczyszczenie danych, a następnie użycie ich do stworzenia modeli predykcyjnych i zastosowanie modeli dla nowych obiektów. Wiele domyślnych ustawień zależy od danych, np. jeśli w pliku mamy wiele zmiennych, to musimy zastosować dobór zmiennych do modelowania, a jeśli zmiennych jest niewiele, to dobór zmiennych jest opcjonalny. Domyślne ustawienia systemu zazwyczaj dają zadowalające wyniki, ale zaawansowani użytkownicy mają do dyspozycji dodatkowe ustawienia i mogą zmieniać domyślne ustawienia.

*STATISTICA Data Miner* zawiera kilkanaście specjalnie zaprojektowanych modułów data mining:

- ◆ **Interakcyjne drażenie danych:** Zadaniem tego modułu jest umożliwienie jednoczesnego tworzenia tabel i wykonywania graficznej, eksploracyjnej analizy danych, tak aby można było szybko przyrzeć się rozkładowi badanych zmiennych, ich związkom z innymi zmiennymi i wykrycia obserwacji należących do konkretnych grup.
- ◆ **Automatyczne sieci neuronowe** to kompletne, korzystające z technik sztucznej inteligencji narzędzia modelowania i prognozowania. Metody dostępne w *STATISTICA Automatyczne Sieci Neuronowe* znajdują tu zastosowanie w projektach data mining.
- ◆ **Analiza składowych niezależnych** służy do filtrowania sygnałów za pomocą dobrze sprawdzonej i niezawodnej techniki statystycznej znanej jako *Independent Component Analysis (ICA)*. Zastosowana implementacja tej metody pozwala rozwiązywać



problemy separacji sygnałów metodami jednoczesnego wyodrębniania i deflacji, wstępnie przetwarzać dane, a także stosować rozwiązanie dla nowych danych.

- ◆ **Analiza skupień uogólnioną metodą EM i k-średnich** to metody uczenia bez nauczyciela (nieukierunkowane), stosowane do rozpoznawania wzorców i segmentacji, dla których można jednak zastosować metody wdrażania zazwyczaj wykorzystywane w modelach predykcyjnych. Skupienia mogą być definiowane w oparciu o wartości zmiennych ciągłych (liczbowych, ilościowych) i skategoryzowanych (jakościowych).
- ◆ **Analiza koszykowa** jest implementacją tzw. „algorytmu wykrywania a priori” (poszukiwania) reguł powiązania typu: „klienci, którzy zamówili produkt A, często zamawiają produkt B lub C”, „pracownicy, którzy dobrze oceniają rozwiązanie X, również często narzekają na rozwiązanie Y, ale zadowala ich rozwiązanie Z”. Moduł ten umożliwia przetwarzanie olbrzymich zbiorów danych w poszukiwaniu powiązań (współzależności), z wykorzystaniem zdefiniowanych wstępnie wartości progowych do wykrywania.
- ◆ **Ogólne modele drzew klasyfikacyjnych i regresyjnych (G-Trees)**: Moduł ten zawiera kompletną implementację metod CART®. Program ma bardzo wiele opcji sterujących budową drzew, przycinania, jak i wyboru najlepiej dopasowanego rozwiązania.
- ◆ **Ogólne modele CHAID** (*Chi-squared Automatic Interaction Detector*) to jedna z najstarszych metod drzew klasyfikacyjnych. CHAID buduje drzewa, w których z węzłów mogą wychodzić więcej niż dwie gałęzie i wykonuje to za pomocą algorytmu, który nadaje się zwłaszcza do analizy dużych zbiorów danych.
- ◆ **Drzewa interakcyjne (C&RT, CHAID)** W tym module budowane są (rosną) automatycznie lub za pomocą algorytmu sterowanego regułami i kryteriami określonymi przez użytkownika, poprzez interaktywny, graficzny interfejs, lub obydwoma metodami.
- ◆ **Drzewa wzmacniane**: Moduł ten pozwala zastosować wzmacnianie (*boosting*) do zadań predykcyjnego data mining. Użytkownik może sterować procedurą budowy modelu, na każdym jej etapie dostępne są szczegółowe podsumowania, pozwalające obserwować i oceniać postęp uzyskiwany na każdym kroku estymacji.
- ◆ **Losowy las** (regresja i klasyfikacja) służy do budowy złożonych modeli klasyfikacji i regresji. Jest szczególnie odpowiedni dla skrajnie dużych liczb predyktorów. *STATISTICA Losowy las* jest realizacją algorytmu Random Forest opracowanego przez Breimana. *STATISTICA Losowy las* daje użytkownikowi pełną kontrolę nad kluczowymi ustawieniami procedury dopasowania modelu.
- ◆ **Uogólnione modele addytywne** to implementacja metod rozwiniętych i spopularyzowanych przez Hastiego i Tibshiraniego. W programie można stosować wiele rozkładów dla zmiennej zależnej. Podobnie wiele funkcji wiążących (ang. *link function*) może być użytych do opisu zależności zmiennej zależnej od predyktorów.
- ◆ **MARSplines** (*Multivariate Adaptive Regression Splines* – Wielozmienna regresja adaptacyjna z użyciem funkcji sklepanych) jest implementacją uogólnienia techniki służącej do rozwiązywania zarówno problemów regresyjnych, jak i klasyfikacyjnych,



w których celem jest znalezienie wartości zmiennych wyjściowych (zależnych) na podstawie zmiennych wejściowych (predykcyjnych).

- ◆ **Inne metody uczenia maszyn** (*Machine Learning*) to zestaw zaawansowanych metod modelowania dla zagadnień regresyjnych i klasyfikacyjnych przy wielu zmiennych niezależnych (predyktorach) i zależnych. Te metody to:
  - wektory nośne (*Support Vector Machines*),
  - naiwny klasyfikator Bayesa,
  - K-najbliższych sąsiadów (*K-Nearest Neighbors*).
- ◆ **Szybkie wdrażanie modeli predykcyjnych**: W tym module można szybko wygenerować predykcje jednego lub wielu uprzednio zbudowanych modeli zapisanych w standardowym formacie PMML (*Predictive Model Markup Language*).
- ◆ **Dobroć dopasowania, klasyfikacji, predykcji** (*Goodness of fit*). Przeznaczeniem tego modułu jest ocena modeli predykcyjnych, zarówno zmiennych zależnych ciągłych, jak i skategoryzowanych. Obliczane są różne statystyki dobroci dopasowania na podstawie przewidywanych i obserwowanych wartości lub klasyfikacji. Powstaje też wiele podsumowujących wykresów ilustrujących wyniki badania.
- ◆ **Dobór i eliminacja zmiennych** umożliwia przeprowadzanie automatycznego doboru podzbiorów zmiennych ze skrajnie dużych zbiorów danych lub baz danych, połączonych w celu zdalnego przetwarzania (IDP). Może on operować na praktycznie nieograniczonej liczbie zmiennych; można sprawdzić dosłownie miliony zmiennych wejściowych w celu wyszukania predyktorów dla zagadnień regresyjnych lub klasyfikacyjnych.
- ◆ **Łączenie klas (grup) do predykcyjnego data mining** jest metodą przetwarzania predyktorów jakościowych o dużej liczbie kategorii.

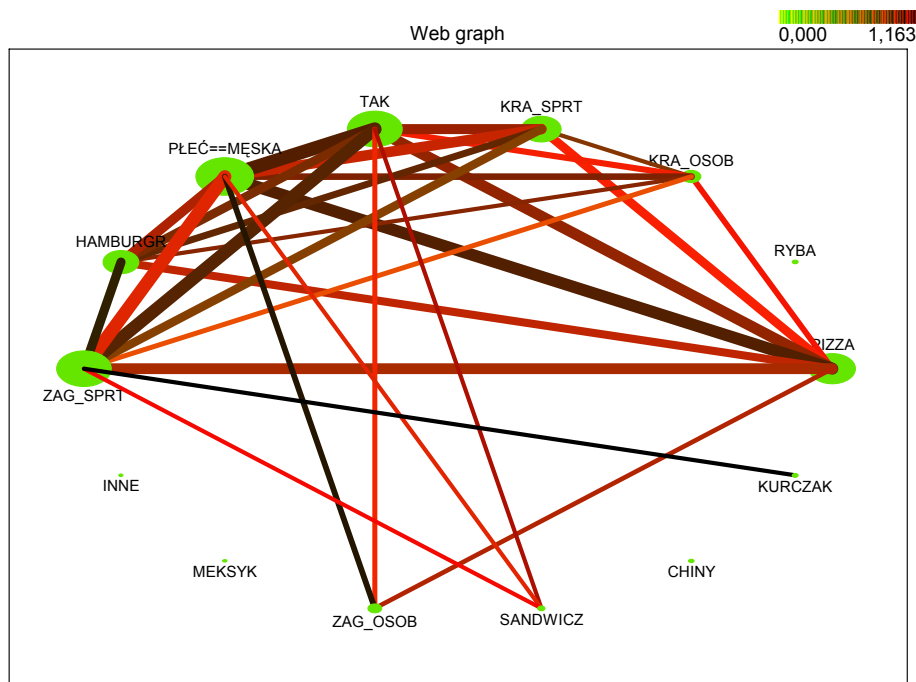


### **STATISTICA Sequence, Association and Link Analysis (SAL) PL**

*STATISTICA SAL* wykorzystuje szereg najnowszych technik analitycznych przeznaczonych do znajdowania w zbiorach danych reguł, jakim podlegają „koszyki zakupów”. Program ten można traktować jako rozszerzenie technik dostępnych w *Analizie koszykowej* zawartej w *STATISTICA Data Miner*. *STATISTICA SAL* pomaga w następujących analizach:

- ◆ wykrywanie reguł asocjacji (ustalenie reguł, które produkty kupowane są razem),
- ◆ analiza sekwencji (badanie, w jakiej kolejności zjawiska zachodzą po sobie),
- ◆ analiza skojarzeń (dzięki niej można, np. analizując dane dotyczące klientów, wnioskować o tym, co będą oni chcieli kupić, bazując na informacjach o wcześniejszych ich zakupach).

*STATISTICA SAL* stanowi doskonale rozszerzenie pakietu *STATISTICA Data Miner*, może też działać samodzielnie.



## StatSoft Polska

StatSoft Polska jest działającym od 1995 roku biurem amerykańskiej firmy StatSoft Inc., która ma już ponad dwudziestopięcioletnie doświadczenie w zakresie dostarczania systemów analitycznych i związanych z nimi usług. Główną domeną działalności StatSoft Polska jest szeroko rozumiana analiza danych. Profesjonalne oprogramowanie analityczne *STATISTICA* oraz kompleksowe usługi doradcze i szkoleniowe realizowane przez doświadczonych specjalistów stanowią gwarancję zadowolenia klientów.

StatSoft Polska zajmuje się dystrybucją i sprzedażą produktów firmy StatSoft w Polsce, wdrażaniem i implementacją rozwiązań przygotowanych w oparciu o narzędzia z rodziny *STATISTICA* oraz konsultingiem w zakresie analizy danych i budowy systemów analitycznych. Oprócz tego organizuje kursy i szkolenia związane z analizą danych oraz uczestniczy w przygotowaniu polskich wersji programów z rodziny *STATISTICA*.

StatSoft Polska jest jednym z największych w Polsce dostawców programów do statystycznej analizy danych i data mining, a także największym w Polsce organizatorem specjalistycznych kursów i szkoleń dotyczących zastosowań analizy danych (statystycznej, data mining) w różnych dziedzinach. Członkowie zespołu StatSoft Polska brali udział w wielu projektach analitycznych i wdrożeniowych.

Rozwiązania StatSoft są powszechnie wykorzystywane ze względu na uniwersalność i nowoczesność zaimplementowanych metod, łatwość ich stosowania w specjalnie zaprojektowanym na potrzeby analizy danych interfejsie, bogatym możliwościami wizualizacji danych. Szacuje się, że liczba użytkowników *STATISTICA* na całym świecie przekracza



600 000. W Polsce są to tysiące użytkowników w setkach firm, instytucji rządowych i samorządowych, uczelni, instytutów badawczych.

Jednym z elementów dopasowania oferty StatSoft do potrzeb klientów są usługi konsultingowe w zakresie gromadzenia i analizy danych, projektowania systemów analizy danych i współdzielenia wiedzy, budowy modeli analitycznych (modeli prognostycznych, klasyfikacyjnych, segmentacyjnych itp.) oraz ich wdrażania w środowisku *STATISTICA*. StatSoft posiada zaplecze doświadczonych pracowników, specjalistów w dziedzinie opracowywania i implementacji systemów analitycznych na bazie programów *STATISTICA*. Korzysta też z doświadczenia konsultantów biur StatSoft na świecie. Członkowie zespołu konsultantów i pracowników StatSoft Polska brali udział m.in. w projektach realizowanych dla: Netia (budowa modeli *anty-churn*, wspomagających utrzymanie klienta), PKN ORLEN (analiza danych z programu lojalnościowego VITAY), BP Polska (segmentacja klientów), Goodyear Dunlop Tires Polska (prognozowanie potencjału polskiego rynku opon), Grupy Wydawniczej INFOR (badanie satysfakcji klienta) i innych.