

## **kim są moi najlepsi klienci?**

**użyj SPSS, by pozyskać istotne informacje  
z twoich danych marketingowych**



## KIM SĄ MOI NAJLEPSI KLIENTY?

Jest to jedno z kluczowych pytań, z odpowiedzi na które zadowolony byłby każdy szef działu – marketingu, sprzedaży, rozwoju produktu czy serwisu.

Dziś, przy wciąż zmieniającym się rynku, coraz trudniejsze i droższe staje się zainteresowanie, pozyskanie i utrzymanie klientów. W wyniku presji rynku coraz więcej przedsiębiorstw korzysta z analiz marketingowych na własnych bazach danych, by maksymalizować efektywność promocji wśród własnych klientów. Przekształcenie danych w pogłębioną wiedzę o klientach staje się przysłowiowym „być albo nie być” firmy na rynku.

Poznanie charakterystyki twoich klientów daje możliwość wejścia „w głębię” tzn. zrozumienia klienta. Wiedza o najbardziej i najmniej wartościowych klientach, ich wzorcach zakupów, zachowaniach konsumpcyjnych i profilu demograficznym jest kluczem do rozwoju strategii marketingowych. Lepsze zrozumienie profilu typowego klienta pomoże zwiększyć ich lojalność, utrzymać zainteresowanie, zaplanować strategię łączenia produktów w wiązki, czy lepiej określić docelowe rynki. Decyzje poparte rzetelną wiedzą będą podstawą efektywnych kampanii reklamowych i promocyjnych, marketingu bezpośredniego oraz innych sposobów komunikacji marketingowej.

Istnieje wiele możliwości definiowania profili najlepszych klientów i równie dużo sposobów na pomiar tych profili. Poniższe opracowanie to tylko skrótowe objaśnienie jednej z wielu metod analizy klientów przy użyciu pakietu SPSS.

Z bazy danych przedsiębiorstwa wylosowano 2070 klientów i w oparciu o badanie ankietowe uzupełniono informacje społeczno-demograficzne o tych klientach. W tym przykładzie wykorzystano:

- datę dokonania pierwszego zakupu przez klienta,
- historię zakupów według wartości kolejnych zamówień,
- odpowiedzi na poszczególne oferty,
- poziom dochodu gospodarstwa domowego,
- region zamieszkania klienta,
- płeć i inne dane demograficzne.

Celem analizy jest przyjrzenie się ilości pieniędzy wydanych przez klientów przedsiębiorstwa w czasie, w celu zidentyfikowania różnych segmentów klientów, określonych cechami demograficznymi. W poniższym opracowaniu skorzystamy z różnych technik analizy danych – od podstawowych do wyrafinowanych – w celu wydobycia użytecznych informacji.

Informacje jakie można uzyskać, dzięki nawet najbardziej elementarnym procedurom analitycznym, mogą mieć głęboki wpływ na zrozumienie klientów, a tym samym na trafne ocenienie ich wartości dla przedsiębiorstwa. Połączenie wiedzy o własnej branży z elastycznymi i potężnymi narzędziami analitycznymi jest najlepszą drogą do wyciągnięcia maksimum informacji z dostępnych danych.

## WSTĘPNA EKSPLORACJA: KIM JEST MÓJ TYPOWY KLIENT?

Zacniemy od wstępnej analizy różnych zmiennych z naszej bazy danych, w celu uzyskania odpowiedzi na pytania, takie jak:

- Z których regionów pochodzą nasi klienci?
- Jaki jest przeciętny dochód gospodarstwa domowego naszych klientów?
- Jak długo poszczególni klienci kupują nasze produkty?
- Jaki jest przeciętny poziom odpowiedzi na poszczególne oferty? Ile osób odpowiedziało na ofertę nr 1?
- Ile pieniędzy wydają kupujący na nasze produkty?

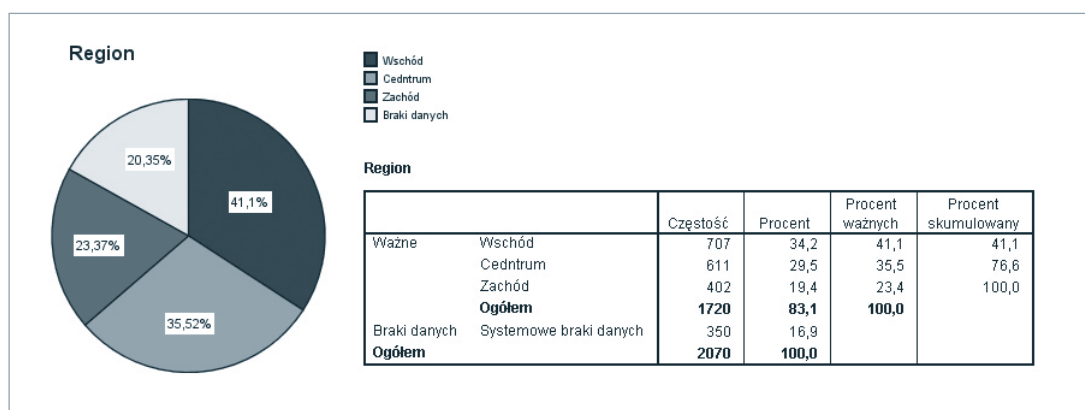
*by tworzyć  
efektywną  
politykę  
marketingową  
musisz zrozumieć  
swoich klientów*

SPSS oferuje wiele metod pozwalających na szybkie uzyskanie odpowiedzi na tak postawione pytania. *Tabele częstości* i *Statystyki opisowe* dają pierwszy wgląd w nasze dane i jednocześnie sugerują jakiego rodzaju analiz trzeba użyć w dalszych krokach.

Spojrzenie na zmienną określającą miejsce zamieszkania pomoże określić liczbę klientów w każdym z regionów. Tabele częstości SPSS podają liczebności i procenty w poszczególnych kategoriach analizowanych zmiennych, jednocześnie prezentując wyniki na wykresach słupkowych, kołowych czy histogramach. SPSS automatycznie prezentuje wyniki w tabelach i na wykresach od razu w formie gotowej do umieszczenia w raporcie.

Z tej analizy otrzymujemy informacje, które mogą okazać się bardzo ważne. Z wykresu kołowego i zawartości tabeli [Rysunek 1.] widzimy, że najwięcej naszych klientów (ok. 34%) mieszka na obszarze wschodnim, a najmniej (ok. 19%) w regionie zachodnim. Równocześnie nie znamy miejsca zamieszkania ok. 17% naszych klientów.

**Rysunek 1.**  
Tabela i wykres utworzone w SPSS, wskazują, że najwięcej osób kupujących nasze produkty mieszka w regionie wschodnim (ok. 34%).



SPSS potrafi rozróżniać różne rodzaje braków danych traktując je w sposób szczególny. Jest to ważne, gdyż wiedza – kiedy i dlaczego pewnej informacji brakuje – może okazać się bardzo użyteczna. Możemy np. chcieć rozróżnić braki informacji w sytuacjach, w których pytanie respondent nie dotyczy, od sytuacji kiedy pewne informacje są dla nas niedostępne. W tabeli [Rysunek 1.] kolumna *Procent* jako bazę do procentowania przyjmuje wszystkich klientów z analizowanego zbioru danych, w tym także tych, którzy nie udzieli odpowiedzi na to pytanie, natomiast kolumna *Procent ważnych* za bazę do procentowania przyjmuje tylko tych klientów, którzy odpowiedzieli na to pytanie. Umożliwia to zaobserwowanie wpływu braków danych na wyniki analizy.

Aby uzyskać informacje o dochodach gospodarstwa domowego wystarczy skorzystać z podstawowych statystyk takich jak: średnia, wartość minimalna i maksymalna. Zmienne ilorazowe lub interwałowe, takie jak dochód mierzony w złotych lub wiek mierzony w latach, najlepiej wstępnie analizować właśnie z użyciem tego typu statystyk opisowych.

Procedura *Statystyki opisowe* oferuje cały zestaw statystyk takich jak: średnia, mediana, odchylenie standardowe itp. Z tabeli [Rysunek 2.] możemy odczytać, że średni dochód gospodarstwa domowego naszych klientów wynosi ok. 30700,- PLN, a większość wartości tej zmiennej waha się od 25200,- do 36200,- PLN (plus/minus jedno odchylenie standardowe).

Aby dowiedzieć się, jak długo kupujący są naszymi klientami, musimy przeprowadzić pewne transformacje danych i następnie zliczyć ilość klientów przybywających w każdym okresie. Na bazie zmiennej zawierającej datę wprowadzenia klienta do naszej bazy (czyli datę dokonania przez niego pierwszego zakupu), stworzymy nową zmienną: czas bycia klientem. Używając jednej z wielu dostępnych w SPSS funkcji operujących na datach, możemy łatwo przetransformować datę pierwszego zakupu na, podawany w latach, czas jaki upłynął od momentu pierwszego zakupu. Po utworzeniu takiej zmiennej możemy stworzyć tabelę częstości czasu bycia klientem.

**Statystyki opisowe**

	Minimum	Maksimum	Średnia	Odchylenie standardowe
Dochód gospodarstwa domowego	19276.0 PLN	42785.5 PLN	30,693.19 PLN	5503.4 PLN

**Czas pozostawania klientem (w latach)**

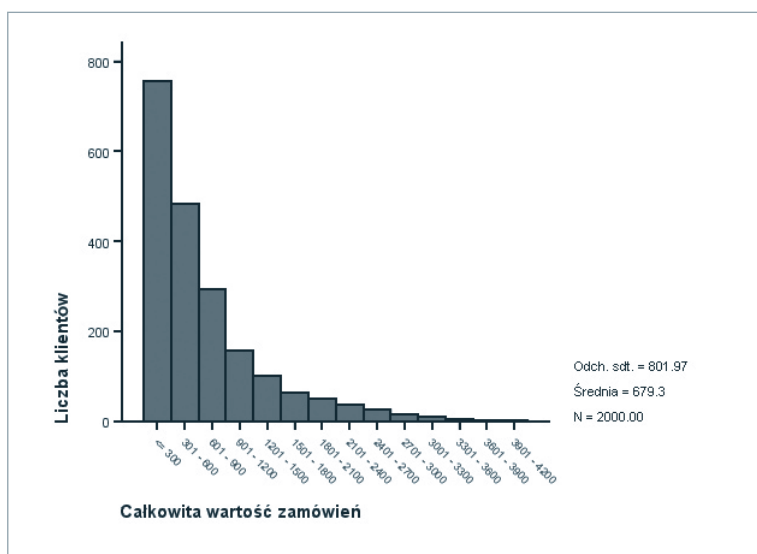
		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważne	1	169	8,2	8,5	8,5
	2	172	8,3	8,7	17,2
	3	147	7,1	7,4	24,6
	4	144	7,0	7,3	31,9
	5	132	6,4	6,7	38,5
	6	121	5,8	6,1	44,6
	7	145	7,0	7,3	51,9
	8	135	6,5	6,8	58,7
	9	112	5,4	5,6	64,4
	10	136	6,6	6,9	71,2
	11	117	5,7	5,9	77,1
	12	128	6,2	6,5	83,6
	13	105	5,1	5,3	88,9
	14	110	5,3	5,5	94,4
	15	111	5,4	5,6	100,0
	<b>Ogółem</b>	<b>1984</b>	<b>95,8</b>	<b>100,0</b>	
Braki danych	Systemowe braki danych	86	4,2		
	<b>Ogółem</b>	<b>2070</b>	<b>100,0</b>		

**Rysunek 2.**  
Procedura Statystyki opisowe dostarcza szybkiego opisu danych, wskazującego, iż średni dochód gospodarstwa domowego wynosi ok. 30 700,-PLN.

**Rysunek 3.**  
Tabela częstości wskazuje, że ok. 52% naszych klientów kupuje u nas nie dłużej niż 7 lat.

Z tabeli [Rysunek 3.] możemy odczytać, że 29% naszych klientów znajduje się w bazie danych od ponad 10 lat, a ponad połowa klientów jest z nami nie dłużej niż 7 lat.

Następnym pytaniem, na które będziemy chcieli odpowiedzieć jest: Kto wydaje u nas najwięcej pieniędzy? Za najlepszych klientów uważani są zazwyczaj ci, którzy tworzą największy dochód lub ci, którzy wydają najwięcej pieniędzy w naszej firmie.



**Rysunek 4.**  
Z histogramu widać, że duża grupa naszych klientów wydała u nas nie więcej niż 300,- PLN, a tylko nieliczni wydali znacznie więcej pieniędzy.

Po pierwsze stworzymy nową zmienną określającą, dla każdego klienta, łączną kwotę wydanych w naszej firmie pieniędzy, poprzez podsumowanie poszczególnych zakupów (zmiennie „zakup1”, „zakup2” itd.). Skoro łączna kwota wydanych pieniędzy jest zmienną ciągłą to najlepszym sposobem graficznej prezentacji tych wyników będzie histogram. W histogramie każdy słupek odpowiada pewnemu zakresowi danych. Z histogramu [Rysunek 4.] można łatwo odczytać, iż duża grupa naszych klientów wydała na nasze produkty nie więcej niż 300,- PLN i równocześnie bardzo niewielu kupuje u nas za znacznie wyższe kwoty. Średnia kwota zakupów wynosi ok. 680,- PLN, a tylko nieliczni klienci wydali więcej niż 3300,- PLN.

Podsumowując, na razie wiemy, że przeciętny klient:

- mieszka na wschodzie,
- dochód jego gospodarstwa domowego wynosi ok. 30700,- PLN,
- kupuje nasze produkty od 7 lat,
- wydaje ok. 680,- PLN na nasze produkty i usługi.
- Jak klienci odpowiadają na różne oferty?

*analiza promocji  
jest następnym  
ważnym krokiem  
w kierunku  
zrozumienia  
naszych klientów*

Ocena skutków programów marketingowych pomoże nam określić czy dana oferta „działa”. Pozwoli to w przyszłości powtarzać te oferty, które zaowocowały sukcesem komercyjnym i uczyć się na błędach, poprawiać te, które nie przyniosły oczekiwanych rezultatów. Aby uzyskać odpowiedzi na pytania „Ile osób odpowiedziało na poszczególne oferty?” i „Jaki jest przeciętny poziom odpowiedzi na nasze promocje?”, użyjemy procedury *Tabela częstości* dostępnej w programie SPSS, by sprawdzić częstość odpowiedzi na poszczególne oferty oraz procedury *Statystyki opisowe* by oszacować wartość zamówień wygenerowanych przez każdą z czterech ofert.

Z tabeli [Rysunek 5.] wynika, że 890 osób, czyli prawie 45% naszych klientów, odpowiedziało na ofertę nr 1. Analogicznie, wyniki dla następnych ofert wskazują, że 39% klientów odpowiedziało na ofertę nr 2, ok. 37% odpowiedziało na ofertę nr 3 i ok. 17% osób na ofertę nr 4.

Rezultaty nasuwają nowe pytania: Co było szczególnego w ofercie nr 1, że wywołała ona tak silny odzew wśród naszych klientów? Innymi słowy, czym różniła się ona od pozostałych ofert i czy otrzymane różnice są istotne?

**Rysunek 5.**  
Prawie 45%, czyli 890 klientów z naszej bazy odpowiedziało na ofertę nr 1.

Oferta 1		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważnych	Nie odpowiedział	1110	53.6	55.5	
	Odpowiedział	890	43.0	44.5	100,0
	Ogółem	<b>2000</b>	<b>96.6</b>	<b>100,0</b>	
Braki danych	Systemowe braki danych	70	3.4		
	<b>Ogółem</b>	<b>2070</b>	<b>100,0</b>		

**Rysunek 6.**  
Analiza sprzedaży wykazała, że średnia wartość sprzedaży w Ofercie nr 3 wynosiła ok. 147,-PLN i była niższa niż we wszystkich innych ofertach.

Statystyki opisowe				
	Minimum	Maksimum	Średnia	Odchylenie standardowe
Wartość zamówienia (oferta 1)	.00 PLN	1,640.12 PLN	188.32 PLN	301.16 PLN
Wartość zamówienia (oferta 2)	.00 PLN	1,738.66 PLN	156.18 PLN	284.64 PLN
Wartość zamówienia (oferta 3)	.00 PLN	1,688.59 PLN	146.99 PLN	268.24 PLN
Wartość zamówienia (oferta 4)	.00 PLN	1,669.79 PLN	187.81 PLN	297.86 PLN

Spojrzenie na informacje dotyczące historii zakupów [Rysunek 6.] wskazuje, że średnia wartość sprzedaży w ofercie nr 3 wynosiła ok. 147,- PLN i była niższa niż we wszystkich innych ofertach. W dalszych analizach sprawdzimy czy różnice w stosunku do pozostałych ofert są znaczące.

**POGŁĘBIONE ANALIZY:****CZYM RÓŻNIĄ SIĘ MOI KLIENTI MIĘDZY SOBĄ? W CZYM SĄ DO SIEBIE PODOBNI?**

Teraz kiedy dysponujemy podstawową wiedzą o zachowaniach naszych klientów i sukcesach poszczególnych ofert, możemy osiągnąć jeszcze więcej dzięki analizie współzależności dwóch lub więcej zmiennych z naszej bazy.

SPSS pomaga odnaleźć ukryte relacje między zjawiskami, trudno dostrzegalne w inny sposób, np. na tym etapie analizy wiemy już, w których regionach mamy najwięcej klientów i ile osób odpowiedziało na ofertę nr 1.

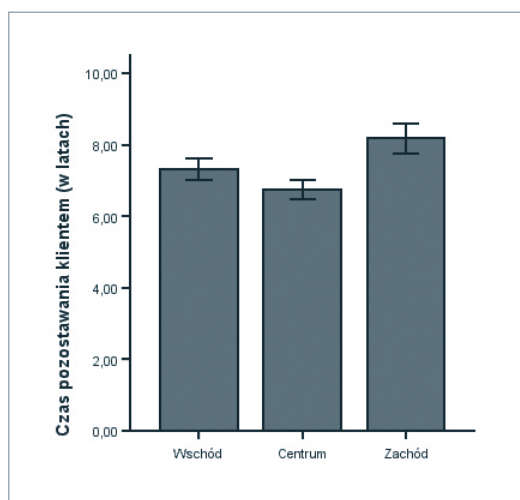
Następnym krokiem będzie zbadanie jak reagowali na ofertę nr 1 klienci w poszczególnych regionach. W tym celu prześledzimy odpowiedzi na pytania:

- Jaki jest przeciętny czas bycia klientem w poszczególnych regionach?
- Jak na ofertę nr 1 reagowali klienci w poszczególnych regionach?

SPSS ułatwia porównywanie różnych grup danych. Procedury: *Tabele krzyżowe* i *Porównywanie średnich*, oraz *Zgrupowany wykres słupkowy* czy *Wykres słupków błęd* jasno prezentują wyniki; statystyka Chi-kwadrat, *Analiza wariancji* i moduł SPSS Classification Trees wskazują kiedy wyniki analiz i nasze spostrzeżenia są istotne statystycznie. Jest to ważne, gdyż pozwala skoncentrować się na rzeczach istotnych.

W następnej kolejności warto znaleźć odpowiedź na pytanie: „Jaki jest przeciętny czas bycia klientem w poszczególnych regionach?”. Dobrym narzędziem, z którego można skorzystać, dostarczającym jednocześnie wielu informacji, tzn. pokazującym równocześnie średnią i 95% przedział ufności dla danych, jest wykres słupkowy ze słupkami błęd [Rysunek 7.]. Łatwo można spostrzec, że wysokość słupka (czyli przeciętny czas bycia klientem) dla regionu zachodniego jest większa niż w pozostałych regionach.

Porównywanie średnich daje możliwość liczenia statystyk opisowych dla połączonych rozkładów. Raport w tabeli [Rysunek 8.] – zawierający te same informacje co wykres słupkowy, ale w formie tabeli – wskazuje, iż ogółem przeciętny czas bycia klientem wynosi 7,49 lat, ale w regionie zachodnim jest dłuższy niż w regionach wschodnim i środkowym. Czy jednak to spostrzeżenie jest znaczące?



*SPSS ułatwia  
porównywanie  
różnych grup  
danych*

**Rysunek 7.**  
*Wykres słupkowy ze słupkami błędów pokazuje równocześnie średnią i 95% przedział ufności dla danych. Łatwo można spostrzec, że wysokość słupka dla regionu zachodniego jest większa niż dla pozostałych regionów, oraz że jest to różnica istotna statystycznie: pozostałe średnie znajdują się poza granicami wąsów dla tego regionu.*

**Rysunek 8.**  
Porównywanie średnich wskazuje, iż ogólnie przeciętny czas bycia klientem wynosi 7,49 lat, ale w regionie zachodnim jest dłuższy niż średnia ogółem i dłuższy niż w regionach wschodnim i środkowym.

**Raport**  
Czas pozostawania klientem (w latach)

Wschód	N	7,38
	Średnia	703
	Odchylenie standardowe	4,40
Centrum	N	6,96
	Średnia	607
	Odchylenie standardowe	4,27
Zachód	N	8,48
	Średnia	398
	Odchylenie standardowe	4,31
Ogółem	N	7,49
	Średnia	1708
	Odchylenie standardowe	4,37

Istotność statystyczna mówi czy dostrzeżone różnice mogą mieć charakter losowy, czy też są wystarczająco duże, aby warto było poszukiwać merytorycznego ich wytłumaczenia. Jeżeli jednak różnice mają charakter losowy znaczy to, że nie warto szukać ich teoretycznego wytłumaczenia. Oznacza to, że żadna z wybranych zmiennych nie ma istotnego wpływu na wyniki. Jeżeli różnice są istotne statystycznie znaczy to, że są one większe niż można by oczekiwać w sytuacji braku związku i wskazują one najprawdopodobniej na nieprzypadkowy wpływ danego czynnika. Kiedy między zmiennymi występuje istotna statystycznie zależność jest to silna zachęta do dalszych analiz.

*kiedy występują zależności istotne statystycznie są one silną zachętą do dalszych badań*

Wyniki procedury *Analiza wariancji* [Rysunek 9.] pokazują, że zróżnicowanie pomiędzy regionami, jeśli chodzi o czas bycia klientem, jest istotne statystycznie. Kiedy istotność jest mniejsza od 0,05 możemy przyjmować, że różnice między średnimi są istotne statystycznie, czyli czas pozostawania klientem jest przynajmniej częściowo zależny od regionu zamieszkania. Wyjaśnienia takiej sytuacji mogą być różne: być może w regionie zachodnim najwcześniej rozpoczęliśmy działalność, albo po prostu zapotrzebowanie na nasze produkty jest różne w różnych regionach, bądź cechy naszego produktu szczególnie dobrze „trafiały” w gust klientów w tym jednym regionie. Jednak by ostatecznie to wyjaśnić, ważna jest także znajomość własnych produktów, otoczenia rynkowego itp., by przejść od danych do weryfikacji własnych przeczuć.

**Rysunek 9.**  
Wyniki procedury ANOVA wskazują, że różnice są istotne statystycznie, co jest zachętą do prowadzenia dalszych badań.

**Analiza wariancji**

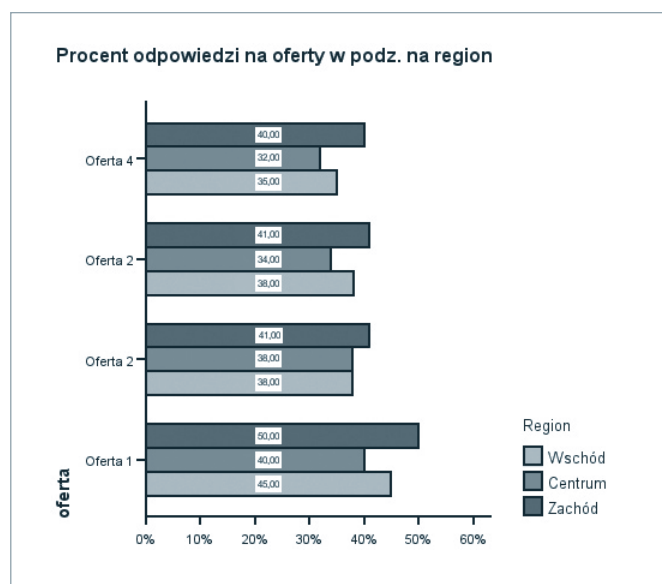
		Suma kwadratów	df	Średni kwadrat	F	Istotność
Czas pozostawania klientem (w latach * region)	Międzygrupowa	569.499	2	284.749	15.188	.000
	Wewnątrzgrupowa	31965.164	1705	18.748		
	Ogółem	32534.663	1707			

Warto przyrzeć się także reakcji klientów na poszczególne oferty promocyjne. W SPSS, za pomocą *Zgrupowanego wykresu słupkowego*, możemy szybko zaprezentować informacje dla wszystkich czterech ofert równocześnie. Wykres [Rysunek 10.] pokazuje wzorce odpowiedzi na promocję w poszczególnych regionach. Widzimy, że w regionie centralnym odsetek zamówień jest niższy niż w dwóch pozostałych regionach, a zwłaszcza w regionie zachodnim. Takiego spostrzeżenia nie moglibyśmy dokonać wyłącznie w oparciu o liczbę odpowiedzi na tę promocję w poszczególnych regionach, która pokazuje jedynie, iż w obszarze zachodnim odpowiedziało na nią najmniej osób.

Aby zbadać czy powyższa zależność jest istotna, warto spojrzeć głębiej na skuteczność poszczególnych ofert w podziale na regiony. Aby odpowiedzieć na pytanie „Jak klienci reagowali na ofertę nr 1 w różnych regionach?” można posłużyć się tabelą krzyżową dla zmiennych „Oferta nr 1” i „Region”. Tabela [Rysunek 11.] wskazuje, iż spośród respondentów którzy odpowiedzieli na tę ofertę ok. 41% osób mieszka na wschodzie. Co prawda zaledwie ok. 27% osób, które odpowiedziały na ofertę nr 1 pochodzi z zachodu, ale jest to ponad

polowa (50,5%) adresatów tej oferty z regionu zachodniego. Aby zrozumieć czy miejsce zamieszkania determinuje prawdopodobieństwo zareagowania na ofertę nr 1, porównujemy procentowanie w wierszach (procent w regionie) i zauważamy, że na ofertę nr 1 odpowiedziało 45% klientów ze wschodu, oraz 40% klientów z regionu centralnego. W oparciu o te informacje możemy stwierdzić, że region zachodni jest obszarem dogodnym dla ofert podobnych do oferty nr 1.

*statystyka  
Chi-kwadrat  
wyjaśni czy  
różnice, które  
dostrzegasz, nie  
są przypadkowe*



**Rysunek 10.**  
Wykres pokazuje wzorce odpowiedzi na promocje w poszczególnych regionach. Jak widać najwyższy odsetek odpowiedzi jest w regionie zachodnim.

		Oferta 1		Ogółem
		Nie odpowiedział	Odpowiedział	
Region	Wschód	Liczebność 391	316	707
		% Oferta 1 41,0%	41,3%	41,1%
		% Region 55,3%	44,4%	100,0%
Centrum	Liczebność	364	247	611
		% Oferta 1 38,2%	32,2%	35,5%
		% Region 59,6%	40,4%	100,0%
Zachód	Liczebność	199	203	402
		% Oferta 1 20,9%	26,5%	23,4%
		% Region 49,5%	50,5%	100,0%
Ogółem	Liczebność	954	766	1720
		% Oferta 1 100,0%	100,0%	100,0%
		% Region 55,5%	44,5%	100,0%

**Rysunek 11.**  
Tabela krzyżowa wskazuje, iż spośród respondentów którzy odpowiedzieli na ofertę nr 1 ok. 41% mieszka w regionie wschodnim.

Jednakże pomimo, iż odsetek reakcji na ofertę różni się w poszczególnych regionach, nie możemy jeszcze z całą pewnością stwierdzić, że w regionie zachodnim należy powtarzać podobne oferty. Musimy mieć świadomość, że ze względu na stosunkowo niski udział klientów z tego regionu wśród wszystkich osób, które odpowiedziały na tę promocję, proste porównywanie procentów może prowadzić do błędnych wniosków. Dlatego przy pomocy statystyki Chi-kwadrat warto sprawdzić, czy różnice w procentowaniu są istotne statystycznie. Warto więc przyjrzeć się zaobserwowanym różnicom korzystając ze statystyki Chi-kwadrat.

Po pierwsze, na bazie testu Chi-kwadrat, sprawdzimy czy zaobserwowane różnice są istotne statystycznie. Tabela [Rysunek 12.] zawiera wartości testu Chi-kwadrat dla zmiennych „Region” i „Oferta nr 1”. Zwykle przyjmuje się, że poziom istotności statystyki Chi-kwadrat Pearson’a powinien być niższy od 0,05 (czyli przy założeniu 95-procentowego przedziału ufności), aby uznać różnice za istotne. W naszym przypadku poziom istotności Chi-kwadrat wynosi 0,007, a więc zależność jest istotna statystycznie.

**Rysunek 12.**  
Poziom istotności  
Chi-kwadrat 0,007 dla  
zmiennych: „Region”  
i „Oferta nr 1” wskazuje,  
że różnice pomiędzy  
regionami są istotne  
statystycznie.

**Testy Chi-kwadrat**

	Wartość	df	Istotność asymptotyczn a (dwustronna)
Chi-kwadrat Pearsona	9.970 <sup>a</sup>	2	,007
Iloraz wiarygodności	9.964	2	,007
Test związku liniowego	1.999	1	,157
N Ważnych obserwacji	1720		

a. 0 komórek (0%) ma liczebność oczekiwaną mniejszą niż 5.  
Minimalna liczebność oczekiwana wynosi 179.03.

Co powoduje, że poziom odpowiedzi na ofertę nr 1 jest różny w poszczególnych regionach? Być może istnieje pewna wyjątkowa, możliwa do zidentyfikowania przyczyna, czyniąca ofertę nr 1 szczególnie skuteczną na zachodzie, na przykład oferta ta lepiej zaspokaja potrzeby klientów z tego regionu, lub rodzaj kampanii medialnej został lepiej dobrany do pozyskania i utrzymania zainteresowania klientów.

Poprzez identyfikację przyczyn skuteczności na zachodzie możemy przenieść naszą wiedzę na sukces przy konstruowaniu następnych ofert w tym regionie. Możemy także badać inne zależności występujące w danym regionie.

**KTÓRZY KLIENTY WYDAJĄ NAJWIĘCEJ PIENIĘDZY?**

Innym sposobem spojrzenia na historię sprzedaży jest przyjrzenie się sprzedaży ogółem, a nie badanie wielkości sprzedaży dla poszczególnych zamówień. Być może znajomość zależności między sprzedażą ogółem i regionem wniesie coś nowego.

Jednoczynnikowa analiza wariancji dostarcza informacji na temat istotności obserwowanych różnic pomiędzy porównywanymi wartościami średnich.

Pierwszym wynikiem uzyskanym z procedury Jednoczynnikowa analiza wariancji jest tabela *Statystyki opisowe* [Rysunek 13.] pokazująca, iż średnia ilość pieniędzy wydawanych ogółem dla czterech promocji w poszczególnych regionach znacznie się różni. W regionie centralnym wydawano średnio 603,- PLN, na wschodzie 696,- PLN, zaś na zachodzie średnia przekraczała 800,- PLN.

**Rysunek 13.**  
Statystyki pokazują  
różnicę w ilości  
w wydawanych pieniędzy  
przez klientów z różnych  
regionów.

**Statystyki opisowe**  
Zmienna zależna: Sprzedaż

Region	Średnia	Odchylenie standardowe	N
Wschód	695.84 PLN	221.60	673
Centrum	602.75 PLN	198.65	411
Zachód	800.18 PLN	234.61	420
Ogółem	699.54 PLN	218.62	1504

**Rysunek 14.**  
Jednoczynnikowa analiza  
wariancji pokazuje,  
że różnice w ilości  
wydawanych pieniędzy  
pomiędzy regionami  
wschodnim i zachodnim  
nie są istotne  
statystycznie, podczas gdy  
różnice między regionami  
zachodnim i centralnym  
są istotne.

**Porównania wielokrotne**  
Zmienna zależna: Sprzedaż  
Test Scheffe

(I) Region	(J) Region	Różnica średnich (I-J)	Błąd standardowy	Istotność	95% przedział ufności	
					Dolna granica	Górna granica
Wschód	Centrum	92.8 PLN	44.354	.112	-15.8 PLN	202 PLN
	Zachód	-113 PLN	50.159	.078	-236 PLN	9.60 PLN
Centrum	Wschód	-92.8 PLN	44.354	.112	-202 PLN	15.8 PLN
	Zachód	-206 PLN*	51.568	.000	-332 PLN	-79.8 PLN
Zachód	Wschód	113 PLN	50.159	.078	-9.60 PLN	236 PLN
	Centrum	206 PLN*	51.568	.000	79.8 PLN	332 PLN

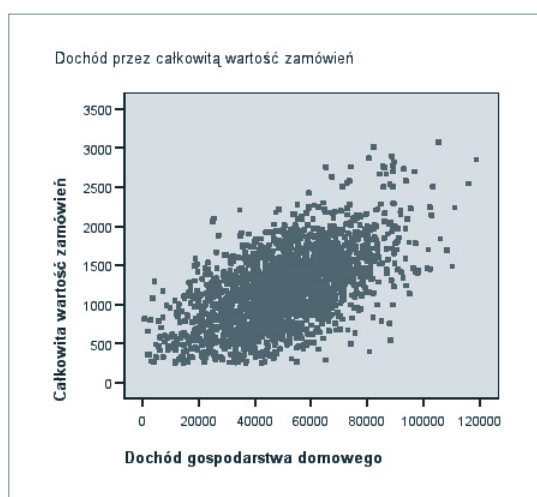
\* Różnica średnich jest istotna na poziomie < 0.05

Natomiast końcowy fragment raportu z *Jednoczynnikowej analizy wariancji* [Rysunek 14.] uwiadamia, że różnice pomiędzy średnią wydawanych pieniędzy w regionie wschodnim, a średnią wydawanych pieniędzy w regionie zachodnim nie są statystycznie istotne. Z drugiej strony różnice między regionami zachodnim i centralnym są istotne statystycznie.

Te istotne informacje można wykorzystać w celu dalszego rozpoznania czym i dlaczego różnią się poszczególne regiony, oraz do tworzenia docelowych planów marketingowych uwzględniających odkryte różnice. Na przykład, na obszarze centralnym lepsze rezultaty powinny dać: inny marketing, inne oferty lub sprzedaże wiązane produktów i usług. Zaś w regionie zachodnim dotychczasowa polityka marketingowa może być powtórzona z podobnie dobrym skutkiem.

## PROGNOZOWANIE CAŁKOWITEJ SPRZEDAŻY

Modele predykcyjne są potężnym narzędziem służącym lepszemu ukierunkowaniu planów marketingowych i optymalizacji nakładów. Pomagają one w uzyskaniu odpowiedzi na pytania typu: „Jaką część swoich dochodów wydaje przeciętne gospodarstwo domowe na nasze produkty i usługi?”



*modele  
predykcyjne  
są potężnym  
narzędziem  
służącym  
lepszemu  
ukierunkowaniu  
planów  
marketingowych  
i optymalizacji  
nakładów*

**Rysunek 15.**  
*Wykres ten pokazuje kształt zależności pomiędzy dwiema zmiennymi. Wykres rozrzutu jest najlepszym sposobem przedstawienia łącznego rozkładu zmiennych ilorazowych lub interwałowych.*

Celem wielu badań statystycznych jest ustalenie wyrażonej równaniem relacji pomiędzy zmiennymi, by móc przewidywać typowe wartości jednej zmiennej dla zadanej wartości innej zmiennej. Pakiet SPSS oferuje wiele procedur pozwalających na ustalenie zależności między zmiennymi czy zdefiniowanie modeli predykcyjnych: od wykresów rozrzutu [Rysunek 15.] i korelacji, poprzez regresję liniową i logistyczną analizę regresji, do analizy CHAID. SPSS wspomaga użytkownika oferując pomoc podręczną „Co to jest?”, dającą wyjaśnienie każdego terminu, każdej statystyki oraz Samouczek, który prowadzi użytkownika krok po kroku, umożliwiając wykonywanie procedur także osobom nie będącym statystykami.

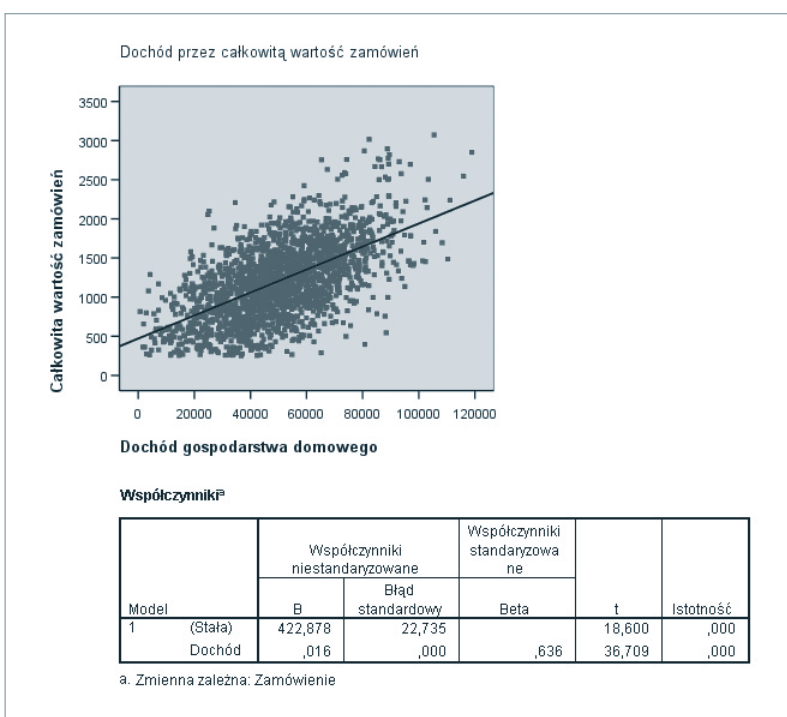
Wartość 0,636 współczynnika korelacji Pearsona [Rysunek 16.], wskazuje na silną zależność pomiędzy dochodem gospodarstwa domowego, a całkowitą ilością wydawanych pieniędzy. Analiza regresji określa zależności za pomocą modelu, jak pokazują to tabela i wykres na Rysunku 17. Ta zależność wskazuje, iż w miarę wzrostu dochodu gospodarstwa domowego rosną jego wydatki na nasze produkty. To spostrzeżenie możemy wykorzystać w celu lepszego prognozowania sprzedaży i udoskonalania działań marketingowych. Przykładowy program może zawierać: dotarcie z większą ilością produktów i usług do gospodarstw domowych o wyższym dochodzie, rozwijanie polityki utrzymania klienta, która pomoże zatrzymać klientów o wysokim dochodzie gospodarstwa domowego, pozyskiwanie stałych klientów. Jak dotąd zauważyliśmy wyraźną zależność pomiędzy miejscem zamieszkania klienta, a prawdopodobieństwem dokonania przez niego zakupu naszych produktów. Dodatkowo, widzimy, że wysokość dochodu jest dodatnio skorelowana z całkowitą ilością wydawanych przez klientów pieniędzy.

**Rysunek 16.**  
Wartość 0,636 współczynnika korelacji zmiennych wskazuje na silną zależność: w miarę wzrostu przychodu gospodarstwa domowego wzrasta ilość pieniędzy wydawanych na nasze produkty.

**Korelacje (CORRELATIONS)**

		Całkowita zamiennosc (w PLN)	Dochód gospodarstwa domowego
Korelacja Pearsona	Całkowita zamiennosc (w PLN)	1,000	,636**
	Dochód gospodarstwa domowego	,636**	1,000
Istotnosc (dwustronna)	Całkowita zamiennosc (w PLN)		,000
	Dochód gospodarstwa domowego	,000	
N	Całkowita zamiennosc (w PLN)	2000	1984
	Dochód gospodarstwa domowego	1984	1984

**Rysunek 17.**  
Regresja liniowa określa związek między dochodem gospodarstwa domowego, a całkowitą ilością wydawanych pieniędzy. Im więcej pieniędzy zarabiamy, tym więcej ich wydajemy.



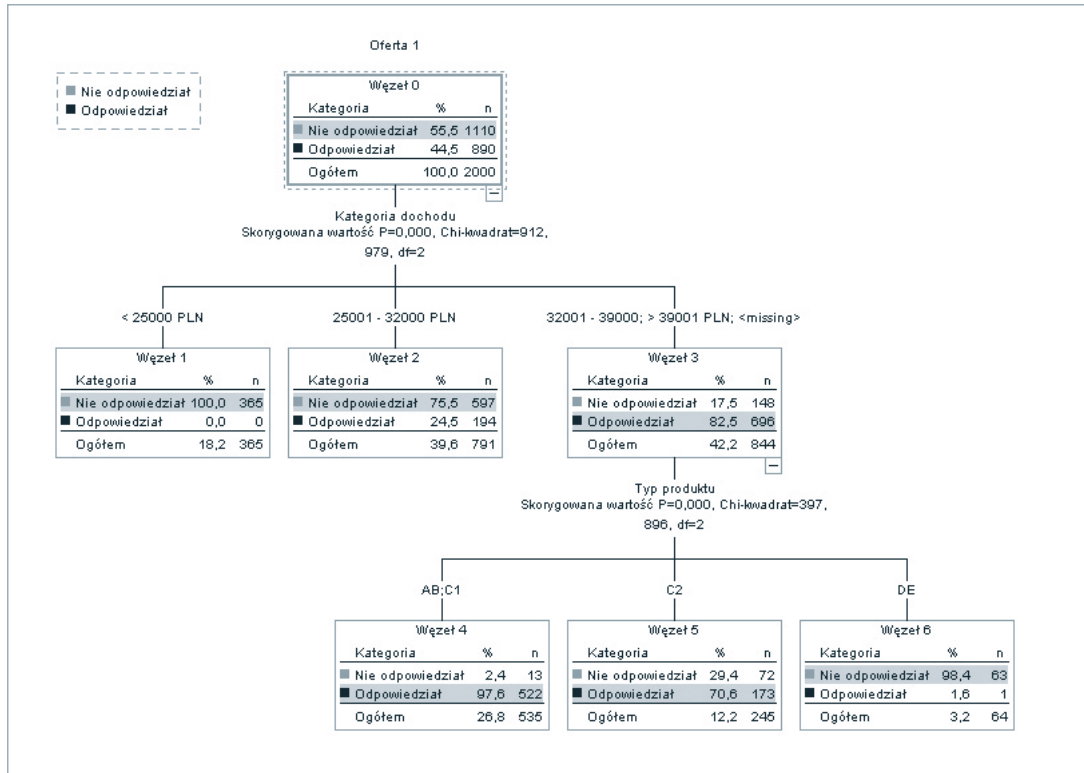
**SPSS Classification**

**SEGMENTACJA KLIENTÓW JAKO SPOSÓB NA BARDZIEJ DOCHODOWY I EFEKTYWNY MARKETING**

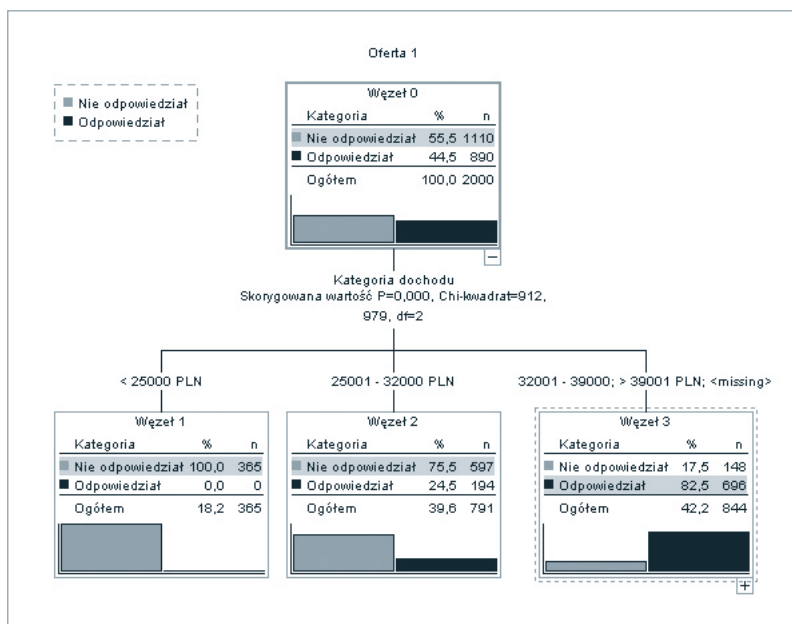
Trees umożliwia identyfikację unikalnych segmentów w danych, dzięki czemu możesz osiągnąć lepsze wyniki swoich działań marketingowych

Analitycy często korzystają z automatycznej detekcji interakcji na bazie testu Chi-kwadrat, w skrócie CHAID. Jest to jeden z czterech algorytmów drzew klasyfikacyjnych zawartych w module SPSS Classification Trees. Metoda ta zwykle nie jest stosowana do sprawdzania czy związek między dwiema zmiennymi jest istotny statystycznie, ale raczej pomaga określić jaka kombinacja charakterystyk kilku zmiennych ma największy wpływ na zmienną wynikową (np. poziom odpowiedzi na nasze oferty). Do modelu budowanego w SPSS Classification Trees włączamy zmienne „region”, „typ produktu” i „dochód gospodarstwa domowego” (w podziale na kategorie), aby dowiedzieć się, którzy klienci najchętniej odpowiadali na ofertę nr 1. SPSS Classification Trees automatycznie buduje drzewo wynikowe, takie jak przedstawione na Rysunku 18.

Rysunek 19. pokazuje szczegóły górnej gałęzi drzewa klasyfikacyjnego, zawierającej zmienne mające najmniejszy wpływ na odpowiedzi na naszą ofertę. Jak widać dochód gospodarstwa domowego jest najlepszym predyktorem (co potwierdza wcześniejsze spostrzeżenia z analizy regresji o silnym wpływie dochodu). W tym przypadku SPSS Classification Trees sięga dalej niż analiza regresji badając zależności głębiej.

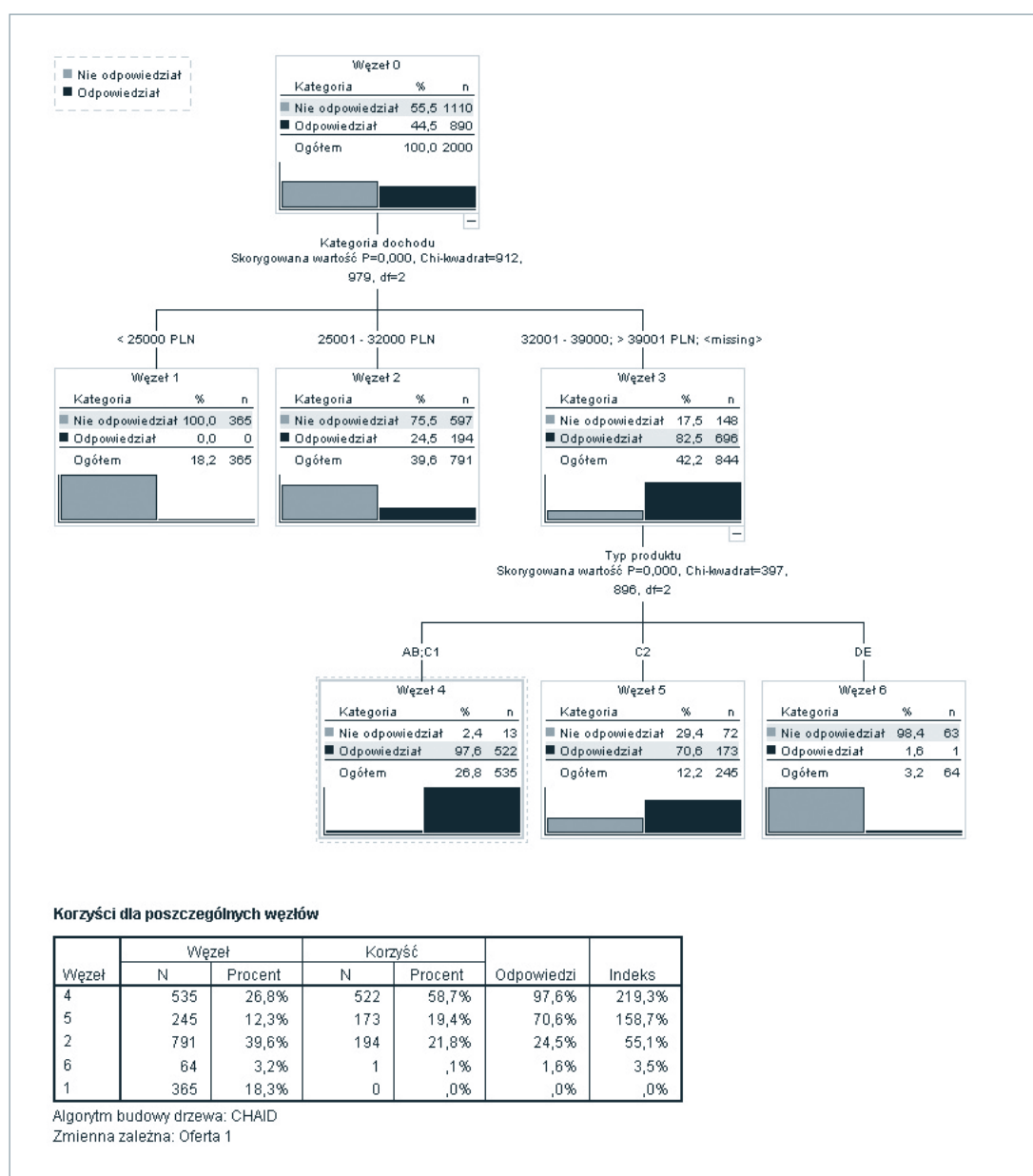


**Rysunek 18.**  
SPSS Classification Trees tworzy model opisujący jakie kombinacje cech mają najchętniej odpowiadający na ofertę nr 1.



**Rysunek 19.**  
Górna gałąź drzewa wynikowego w SPSS Classification Trees wskazuje na dochód jako najmocniejszy predyktor odpowiedzi na naszą ofertę.

Kolejna gałąź drzewa klasyfikacyjnego [Rysunek 20.] można zauważyć, że w przypadku gdy klient pochodzi z gospodarstwa domowego o dochodzie równym lub większym 32 000,- PLN, a nasza oferta dotyczy produktu klasy AB lub C1, mamy 98% poziom odpowiedzi na ofertę. SPSS Classification Trees identyfikuje unikalne segmenty w danych, dzięki czemu, poprzez możliwość ustalenia kombinacji cech klientów, możemy optymalizować politykę marketingową. Dzięki drzewu wykonanemu w SPSS Classification Trees odkryliśmy, iż osoby z gospodarstw domowych o dochodzie 32 000,- PLN lub wyższym, które kupują produkty typu AB lub C1, częściej niż jakakolwiek inna grupa naszych klientów kupowały produkty w odpowiedzi na ofertę nr 1. Analiza z dodatkowymi zmiennymi może prowadzić do kolejnych wniosków np.: może uświadomić nam, że wprawdzie region centralny nie odpowiedział na ofertę tak jak się tego spodziewaliśmy, jednakże w tym regionie kobiety w pewnych kategoriach dochodu gospodarstwa domowego pozytywnie reagują na nasze oferty i być może warto zaadresować do nich następną kampanię promocyjną.



**Rysunek 20.**  
SPSS Classification Trees identyfikuje szczególnie atrakcyjny segment klientów z dochodem ponad 32 000,- PLN i kupujących produkty grupy AB lub C1.

## PODEJMOWANIE DZIAŁAŃ

SPSS pozwala nam szybko określać średnie i rozkłady danych. Dzięki temu zdobywamy wiele ważnych informacji o naszych typowych klientach jak np. to, że pochodzą oni z obszaru wschodniego, zwykle są klientami długoterminowymi, niechętnie reagują na ofertę typu 3, oraz legitymują się wysokim dochodem. Poznanie profilu typowego klienta pozwoli nam w przyszłości lepiej ukierunkowywać promocje i oferty specjalne. Przez równoczesne porównanie wielu charakterystyk i grup SPSS pomaga badać ukryte informacje: np. nie tylko, że oferta nr 3 przyniosła nam najmniejsze korzyści, ale także że była ona szczególnie nieskuteczna w regionie centralnym, regionie który zwykle słabiej reaguje na nasze promocje od dwóch pozostałych. Ponieważ klienci w regionie centralnym mają najniższy średni dochód, łatwiej nam zrozumieć dlaczego relatywnie słabiej reagują na nasze oferty. Dzięki zidentyfikowaniu tych grup będziemy mogli precyzyjniej adresować nasze działania marketingowe i skuteczniej walczyć o utrzymanie klientów.

Podsumowując, można stwierdzić, że modelowanie predykcyjne i techniki segmentacyjne dostępne w SPSS dają możliwość budowania modeli opisujących zależności pomiędzy dochodami, a całkowitą ilością pieniędzy wydawanych na nasze produkty, a tym samym pozwalają lepiej prognozować przyszłą sprzedaż. Możemy także identyfikować specyficzne segmenty klientów w oparciu o odsetek odpowiedzi na ofertę nr 1. Korzystanie z segmentacji opartej na przewidywanym poziomie odpowiedzi na promocję jest kluczem do tworzenia efektywnych programów marketingowych. Kiedy charakterystyki segmentów przypiszemy do poszczególnych klientów, możemy powtórzyć efektywne programy, a wyeliminować lub poprawić te, które nie dały oczekiwanych wyników.

W wyniku powyższej analizy możemy zaplanować:

- stworzenie nowych programów utrzymania naszych najlepszych klientów w segmentach definiowanych przez wysoki dochód, zamieszkiwanie w regionie zachodnim, bycie stałym klientem, oraz kupowanie produktów z oferty nr 1
- rozwijanie i testowanie nowych rodzajów sprzedaży i usług wiązanych w celu lepszego zaspokojenia oczekiwań klientów w obszarze centralnym, stosownie do niskiego poziomu dochodu klientów z tego regionu
- powtórzenie strategii rozwoju sprzedaży z regionu zachodniego w obszarach centralnym i wschodnim, aby zbudować grupę stałych klientów
- powtórzenie oferty nr 1 w regionie zachodnim
- skoncentrowanie wydatków na przyszłe kampanie reklamowe wokół prognozowanych najatrakcyjniejszych segmentów (zbudowanych głównie w oparciu o dochód gospodarstwa domowego).

Przez rozszerzenie analiz można by prawdopodobnie jeszcze pogłębić naszą wiedzę. Jednak w tym opracowaniu chcieliśmy jedynie pokazać, iż nie trzeba być statystykiem, ani znać wyrafinowanych metod statystycznych, aby dzięki SPSS zdobyć wiele cennych informacji, mających rzeczywiste zastosowanie biznesowe.

*SPSS dostarcza  
informacji  
o ukrytych  
współzależnościach*

## white PAPER

SPSS dostarcza wiedzę i narzędzia, które pozwalają na efektywną realizację projektów badawczych. Dostarczamy rozwiązań z zakresu zarządzania relacjami z klientem (CRM) i business intelligence, które umożliwiają użytkownikom systemów SPSS bardziej dochodową współpracę z ich klientami. Narzędzia SPSS pozwalają scalać i analizować dane marketingowe, dane o klientach i dane operacyjne w obrębie najważniejszych branż na całym świecie – między innymi w telekomunikacji, ochronie zdrowia, bankowości, finansach, ubezpieczeniach, produkcji, handlu, badaniach rynku, administracji, edukacji i sektorze publicznym. Poza centralą w Chicago (USA) SPSS posiada blisko 170 biur na całym świecie.

SPSS Polska zapewnia pełną informację o produktach SPSS, prowadzi kursy i szkolenia z zakresu analizy danych oraz obsługi i zastosowań programów SPSS. Użytkownikom zapewnia serwis i pomoc techniczną. Więcej informacji znajdą Państwo na stronach SPSS Polska, dostępnych pod adresem [www.spss.pl](http://www.spss.pl).

---

### **SPSS Polska**

ul. Raclawicka 58  
30-017 Kraków  
tel./faks 012.636.96.80  
tel./faks 012.636.07.91  
tel./faks 012.636.45.35  
e-mail: [info@spss.pl](mailto:info@spss.pl)  
[www.spss.pl](http://www.spss.pl)  
[www.analizadanych.pl](http://www.analizadanych.pl)  
[www.webmining.pl](http://www.webmining.pl)