



OD KWESTIONARIUSZA DO RAPORTU – SYSTEM ZBIERANIA DANYCH I NARZĘDZIA ANALITYCZNE W BADANIACH ANKIETOWYCH

Michał Kusy, StatSoft Polska Sp. z o.o.

Wprowadzenie

Badania ankietowe stanowią obecnie jeden z najpopularniejszych sposobów gromadzenia informacji. Jest to złożony proces wymagający zarówno znajomości metodyki prowadzenia badań, jak i dysponowania odpowiednimi narzędziami analitycznymi wspierającymi badacza na każdym etapie opracowywania wyników. Ostatnio w badaniach socjologicznych i marketingowych coraz większe znaczenie zyskują sondaże elektroniczne, które są najtańszym sposobem zbierania danych. W porównaniu z kwestionariuszami w wersji papierowej pozwalają na uzyskanie rzetelniejszych i obarczonych mniejszą liczbą błędów informacji, głównie dzięki ustaleniu formatu odpowiedzi oraz ominięciu etapu ich przepisywania do wersji elektronicznej.

Oprócz podstawowych i bardziej zaawansowanych narzędzi wspierających proces analizy danych ankietowych, w ofercie StatSoft znajduje się również system umożliwiający wypełnianie kwestionariuszy przez Internet. W artykule przedstawimy pokrótce kolejne etapy badań ankietowych - od przygotowania kwestionariusza i gromadzenia informacji on-line w zintegrowanym systemie wprowadzania danych, po czyszczenie danych i wykonanie podstawowych analiz. Przykład bazuje na fikcyjnych danych, przygotowanych na podstawie opisu badania satysfakcji i preferencji klienta (Hayes [1]) oraz doświadczenia konsultantów StatSoft w przeprowadzaniu tego typu analiz.

Kwestionariusze

Sieć kawiarni *KAWAnaŁAWĘ* zdecydowała się przeprowadzić analizę preferencji swoich klientów oraz ich zadowolenia z dotychczasowych usług kawiarni. Badanie miało uwzględnić płeć badanych oraz porę dnia, w której respondenci odwiedzili lokal. Przy okazji postanowiono zbadać skuteczność przeprowadzonej kampanii reklamowej.

Przygotowane zostały pytania dotyczące czynników decydujących o wyborze kawiarni. Listę odpowiedzi stworzono, wykorzystując wiedzę i doświadczenie pracowników firmy. Rozbudowano ją o pytanie otwarte, co miało umożliwić analizę czynników nie ujętych w zaproponowanych odpowiedziach. Zadaniem ankiety była również ocena wybranych



cech kawiarni, takich jak: czas oczekiwania, życzliwość obsługi, smak kawy, jakość jedzenia, cena oraz menu. Klienci mogli oceniać te cechy na skali od *bardzo niezadowolony* (1) do *bardzo zadowolony* (5).

Proszę wskazać cztery najważniejsze czynniki decydujące o wyborze kawiarni:

Życzliwość obsługi	<input checked="" type="checkbox"/>
Cena	<input checked="" type="checkbox"/>
Wiedza obsługi	<input checked="" type="checkbox"/>
Długość oczekiwania	<input type="checkbox"/>
Smak kawy	<input checked="" type="checkbox"/>
Miejsce	<input checked="" type="checkbox"/>
Jakość jedzenia	<input checked="" type="checkbox"/>
Specjalne produkty	<input type="checkbox"/>
Występy muzyczne	<input checked="" type="checkbox"/>
Publikacje o kawie	<input checked="" type="checkbox"/>
Reputacja	<input checked="" type="checkbox"/>
Inne (jakie):

W celu analizy najczęściej kupowanych produktów i ich zestawów poproszono respondentów o zaznaczenie zamówionych przez nich pozycji spośród pięciu napojów (kawa biała, kawa czarna, cappuccino, espresso i herbata) oraz pięciu deserów (croissant, sernik, kremówka, szarlotka i lody). Ponadto, w związku z planami utworzenia w lokalach osobnych pomieszczeń dla palaczy, postanowiono zbadać opinię badanych również w tej kwestii.

Przed wprowadzeniem ankiet w całej sieci kawiarni, w kilku wybranych lokalach przeprowadzono badanie pilotażowe. Następnie dane z ankiet wypełnianych przez respondentów wprowadzane były ręcznie do komputera. Pilotaż pokazał kilka potencjalnych problemów, które mogły pojawić się podczas badania całej sieci *KAWAnaŁAWĘ*. Po pierwsze ankietowani niezbyt chętnie udzielali odpowiedzi na pytania i oddawali tylko częściowo wypełnione ankiety. Niektóre ankiety wypełnione zostały nierzetelnie - wystąpiły serie regularnych wzorów odpowiedzi i odpowiedzi identycznych lub nielogicznych. Ankietowani udzielali również sprzecznych deklaracji - np. że są „raczej niezadowoleni” z ceny i jednocześnie, że „nie mają zdania” w tej kwestii. Pojawił się również problem na etapie samego wprowadzania danych do komputera. Okazało się, że osoby przepisujące ankiety do wersji elektronicznej stosowały różne sposoby kodowania - np. *TAK* (1), *NIE* (2), *nie mam zdania* (3) i *NIE* (1), *nie mam zdania* (2), *TAK* (3), co prowadziło do nieprawidłowego scalenia danych pochodzących z różnych lokali i, w efekcie, wyciągnięcia błędnych wniosków. Późniejsze porównanie ankiet papierowych i danych elektronicznych ujawniło również



pewne niezgodności, które były wynikiem nieuwagi lub zmęczenia przy przepisywaniu ankiet. Pojawiły się kody odpowiedzi nie istniejących w ankiecie - np. 33 lub 222, które były efektem dłuższego przytrzymania odpowiedniego klawisza oraz odpowiedzi przesunięte względem pytań wskutek pominięcia pytania lub dwukrotnego wpisania tej samej odpowiedzi do różnych pytań.

Proszę ocenić wymienione poniżej cechy poprzez zaznaczenie jednej z ocen:

	Bardzo niezadowolony	Raczej niezadowolony	Nie mam zdania	Raczej zadowolony	Bardzo zadowolony
Czas oczekiwania	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Życzliwość obsługi	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Smak kawy	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jakość jedzenia	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cena	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Menu	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Aby zaradzić tym problemom, w badaniach na szerszą skalę postanowiono wprowadzić pewne zmiany. Po pierwsze zdecydowano, że zamiast ankietowania losowych klientów kawiarni o wypełnienie ankiety proszone będą osoby starające się o kartę rabatową sieci *KAWAnaŁAWE*, co ma zapewnić lepszą rzetelność odpowiedzi i pominięcie opinii bardziej przypadkowych klientów. Po drugie, aby ograniczyć liczbę błędów i czas potrzebny na czyszczenie danych, postanowiono wykorzystać system internetowy służący do wprowadzania danych ankietowych. Za wdrożeniem takiego systemu przemawiał również fakt, że sieć *KAWAnaŁAWE* składa się z wielu kawiarni ulokowanych w różnych miejscowościach. Wprowadzanie danych on-line umożliwiło bezpośrednie ich gromadzenie w centralnej bazie danych.

System wprowadzania danych

W trakcie przygotowywania karty rabatowej dla klienta był on proszony o wypełnienie ankiety on-line na stronie internetowej kawiarni. Każdy lokal miał w tym celu przygotowane stanowisko komputerowe, przy którym klienci mogli dodatkowo skorzystać z Internetu. Ankieta była obsługiwana przez zintegrowany system wprowadzania danych, który zapewnił odpowiedni format odpowiedzi, ograniczył liczbę błędów i umożliwił błyskawiczny dostęp do wyników.



Strona startowa - Mozilla Firefox

Plik Edycja Widok Historia Zakładki Narzędzia Pomoc

http://localhost:1676/FormStart.aspx

Google

KAWAnaŁAWĘ

Strona startowa

Użytkownik: user

Wypełnianie ankiet Zarządzanie systemem Informacje Logowanie

Utwórz nową

Edytuj ankietę

Badanie preferencji i satysfakcji klientów sieci kawiarni KAWAnaŁAWĘ

Szanowni Państwo,

Serdecznie zapraszamy do wzięcia udziału w naszym sondażu. Ankieta, którą Państwo wypełnicie składa się z kilku pytań, dzięki którym będziemy mogli poznać oczekiwania naszych klientów oraz poprawić jakość świadczonych przez nas usług. Każda Państwa opinia jest dla nas bardzo cenna. Z góry dziękujemy za rzetelne wypełnienie ankiety.

Aby wypełnić nową ankietę lub poprawić wprowadzone wcześniej dane prosimy wybrać odpowiednią opcję menu *Wypełnianie ankiet*

Copyright © StatSoft Polska, 2009

StatSoft Polska StatSoft Polska Sp. z o.o. ul. Kraszewskiego 36, 30-110 Kraków
tel. (12) 428 43 00, (601) 41 41 51, faks (12) 428 43 01, e-mail: info@statsoft.pl, www.StatSoft.pl

http://localhost:1676/FormLabels.aspx

System jest dostępny z dowolnego miejsca przez przeglądarkę internetową, a mechanizm zarządzania użytkownikami umożliwia między innymi nadawanie innych praw dostępu dla administratorów systemu oraz osób przeprowadzających ankietę. Po zalogowaniu się użytkownik (w sieci *KAWAnaŁAWĘ* był to pracownik przygotowujący kartę rabatową) ma możliwość utworzenia nowej ankiety lub aktualizacji już istniejącego kwestionariusza (na przykład jeśli z jakiejś przyczyny ankieta nie została wypełniona do końca). W bazie danych zapisywany jest automatycznie identyfikator użytkownika oraz czas wprowadzenia ankiety, co w przypadku sieci kawiarni umożliwiło późniejsze porównanie opinii klientów w poszczególnych lokalach oraz o różnych porach dnia.



Wprowadzanie odpowiedzi - Mozilla Firefox

Plik Edycja Widok Historia Zakładki Narzędzia Pomoc

http://localhost:1676/FormQuest.aspx

Google

KAWA na TAWĘ

Wprowadzanie odpowiedzi

Ankieta nr: 01_1094 Użytkownik: user

Wypełnianie ankiet Zarządzanie systemem Informacje Logowanie

Przejdź do:

<< Wstecz Dalej >> Zakończ

Pytanie 3: Proszę wskazać cztery najważniejsze czynniki decydujące o wyborze kawiarni:


- życzliwość obsługi
- cena
- wiedza obsługi
- długość oczekiwania
- smak kawy
- miejsce
- jakość jedzenia
- specjalne produkty
- występy muzyczne
- publikacje o kawie
- reputacja
- inne (jakie):

Zakończono

System uwzględnia pełny zakres rodzajów pytań – od pytań jednokrotnego wyboru, wielokrotnego wyboru, po pytania otwarte, skale i pytania z kilkoma odpowiedziami różnego typu. Zastosowanie kwestionariusza elektronicznego umożliwia określenie formatu odpowiedzi oraz wprowadzenie pytań filtrujących. Dzięki temu osoba wypełniająca ankietę nie ma możliwości zaznaczenia wielu odpowiedzi w przypadku pytań jednokrotnego wyboru lub udzielenia odpowiedzi na pytanie, które jej nie dotyczy. Wszystkie dane są zapisywane w centralnej bazie danych, co w znacznym stopniu usprawnia przeprowadzanie późniejszych analiz. W bazie danych przechowywany jest również szablon kwestionariusza, co ułatwia modyfikację ankiety i dostosowanie jej schematu na potrzeby kolejnych sondaży.



Wprowadzanie odpowiedzi - Mozilla Firefox
Plik Edycja Widok Historia Zakładki Narzędzia Pomoc
http://localhost:1676/FormQuest.aspx

 Wprowadzanie odpowiedzi
Ankieta nr: 01_1094 Użytkownik: user

Wypełnianie ankiet Zarządzanie systemem Informacje Logowanie

Przejdź do: << Wstecz [] Dalej >> Zakończ

Pytanie 4: Proszę ocenić wymienione poniżej cechy kawiarni poprzez zaznaczenie jednej z ocen:

	Bardzo niezadowolony(a)	Raczej niezadowolony(a)	Nie mam zdania	Raczej zadowolony(a)	Bardzo zadowolony(a)
czas oczekiwania	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
zyczliwość obsługi	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
smak kawy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
jakość jedzenia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
cena	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
menu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Zakończono

Sprawdzanie poprawności danych

W wyniku badania przeprowadzonego w sieci kawiarni *KAWAnaŁAWĘ* zebrano 993 ankiety. Dane ankietowe zostały pobrane z bazy danych do arkusza programu *STATISTICA*, w którym wykonamy odpowiednie analizy. Przed przystąpieniem do analizy zebranych odpowiedzi należy jednak sprawdzić ich poprawność, przeanalizować pojawiające się braki danych i wykonać czyszczenie danych, aby uzyskane wyniki były obarczone jak najmniejszym błędem. Poniżej zamieszczono tabelę zawierającą odsetki ważnych odpowiedzi udzielonych na poszczególne pytania. Jak widać, w większości ankiet uzyskano odpowiedzi na wszystkie pytania. Największy odsetek braków danych (10%) występuje w przypadku pytania o zamawiane produkty. Wynika to najprawdopodobniej z kolejności pytań – pytanie o zamawiane produkty zostało zadane jako ostatnie, a część osób nie miała ochoty wypełniać ankiety do końca.



Zmienna	N ważnych	% Ważnych
ID	993	100,0
Godzina	993	100,0
Menu	993	100,0
Płeć	976	98,3
Osobno palacze	981	98,8
Jak trafił do kawiarni	964	97,1
Czynniki wyboru kawiarni	984	99,1
Ocena cech kawiarni	945	95,2
Zamówione produkty	897	90,3

Ponieważ dane zbierane były przez system, liczbę błędów związanych z przepisywaniem odpowiedzi lub ich formatem ograniczyliśmy do minimum. Wolną rękę zostawiliśmy jednak ankietowanym w przypadku pytań otwartych i właśnie tam rozpoczniemy sprawdzanie poprawności danych. Za pomocą opcji *Autofiltr*, dostępnej na karcie *Przekształcenia*, uzyskujemy szybki podgląd przypadków o określonych wartościach zmiennej *Ważne inne (jakie)*. Już pobieżna obserwacja wartości przyjmowanych przez tę zmienną pozwala zauważyć kilka nieprawidłowości. Niektórzy ankietowani wpisywali na przykład „cenę” jako inny czynnik decydujący o wyborze kawiarni, podczas gdy została już ona uwzględniona na liście odpowiedzi (zmienna *Ważna cena*). W takich przypadkach musimy ręcznie zmienić wartość tej zmiennej i usunąć niepotrzebną kategorię *cena* spośród czynników określonych jako „inne”.

Kilku ankietowanych wskazało z kolei ogródek przy kawiarni jako czynnik wpływający na wybór danego lokalu. Kategorii tej odpowiadają jednak różne, nadane przez klientów określenia: *czy jest ogródek, ogródek, ogródki*, które należało by ujednoclić. Pojawił się również podejrzany wpis składający się z ciągu takich samych liter. Okazuje się, że ten sam ankietowany ocenił jednocześnie wszystkie cechy kawiarni na 1 (*bardzo niezadowolony*) i pominął pytanie o zamówione produkty. W celu identyfikacji podobnych przypadków rozpatrzymy badanych, dla których średnia ocena kawiarni wynosi 1 lub 5 (co będzie oznaczało, że ze wszystkich cech kawiarni klient jest bardzo niezadowolony lub odpowiednio bardzo zadowolony). Okazuje się, że w tej grupie są jeszcze dwie osoby, które zgodnie z wynikami ankiety, zamówiły wszystkie produkty z listy, uważają, że każdy czynnik ma wpływ na wybór kawiarni i jednoznacznie źle oceniają wszystkie cechy kawiarni. Zakładając, że nie mieliśmy do czynienia z wyjątkowo niezadowolonymi krytykami kulinarnymi, możemy przyjąć, że był to żart i wyłączyć te przypadki z dalszych analiz.



	10 Ważne oczekiwa nie	11 Ważny smak kawy	12 Ważne miejsce	13 Ważna jakość jedzenia	14 Ważne specjalne produkty	15 Ważne występy	16 Ważne publikacj e	17 Ważna reputacja	18 Ważne inne	19 Ważne inne (jakie)	20 Ocena oczekiwa nie	21 Ocena życziwo ść	22 Ocena smak kawy
1	TAK	TAK	NIE	TAK	NIE	NIE	NIE	NIE	NIE	Sortuj rosnąco			
2	NIE	TAK	NIE	TAK	NIE	NIE	NIE	NIE	NIE	Sortuj malejąco			
3	TAK	NIE	TAK	NIE	TAK	NIE	NIE	NIE	NIE	(Wszystkie)			
4	NIE	TAK	NIE	TAK	NIE	NIE	NIE	TAK	NIE	(Inne...)			
5	NIE	NIE	NIE	TAK	NIE	NIE	NIE	TAK	NIE	aaaaaaaaaaaaa			
6	TAK	NIE	NIE	NIE	NIE	NIE	NIE	TAK	NIE	barista			
7	TAK	TAK	NIE	NIE	NIE	NIE	NIE	NIE	TAK	brak klimatyzacji			
8	TAK	NIE	TAK	TAK	NIE	NIE	NIE	NIE	NIE	cena			
9	TAK	NIE	NIE	NIE	NIE	NIE	NIE	NIE	TAK	cena deserów			
10	TAK	TAK	NIE	TAK	NIE	NIE	NIE	NIE	NIE	ciekawe smaki kawy			
11	NIE	TAK	TAK	TAK	NIE	NIE	NIE	NIE	NIE	czas oczekiwania na kawę			
12	TAK	NIE	TAK	TAK	NIE	NIE	NIE	NIE	NIE	czas po jakim dostaje zamówienie			
13	TAK	NIE	TAK	TAK	NIE	NIE	NIE	NIE	NIE	czekanie na zamówienie			
14	TAK	TAK	NIE	TAK	NIE	NIE	NIE	NIE	NIE	czy jest ogródek			
15	TAK	NIE	NIE	TAK	NIE	NIE	NIE	NIE	NIE	czystość			
16	NIE	TAK	TAK	TAK	NIE	NIE	NIE	NIE	NIE	dobre menu			
17	NIE	TAK	NIE	TAK	NIE	NIE	NIE	NIE	NIE	duzo miejsca			
18	TAK	TAK	NIE	NIE	NIE	NIE	NIE	NIE	NIE	kawa latte			
19	TAK	TAK	TAK	NIE	NIE	NIE	NIE	NIE	NIE	kawa po zbojnicku			
20	TAK	TAK	NIE	NIE	NIE	NIE	NIE	NIE	NIE	kelnerka			

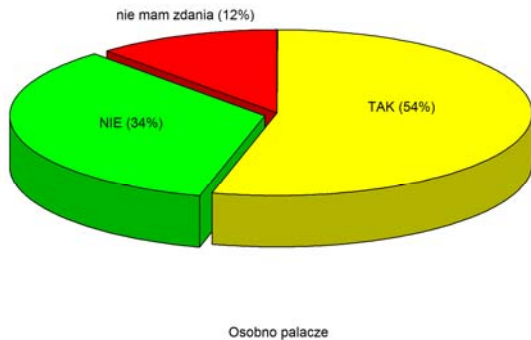
Analiza danych

Program *STATISTICA* oferuje pełen zakres narzędzi statystycznych i graficznych, dzięki którym można wydobyć z danych ankietowych potrzebną wiedzę. *STATISTICA* dostarcza zarówno podstawowych narzędzi analitycznych, takich jak: szereg metod do analizy graficznej, statystyki opisowe, testy statystyczne, jak i bardziej zaawansowanych technik analizy wielowymiarowej - między innymi: analizy korespondencji, analizy skupień, analizy sekwencji, asocjacji i połączeń oraz drzew klasyfikacyjnych.

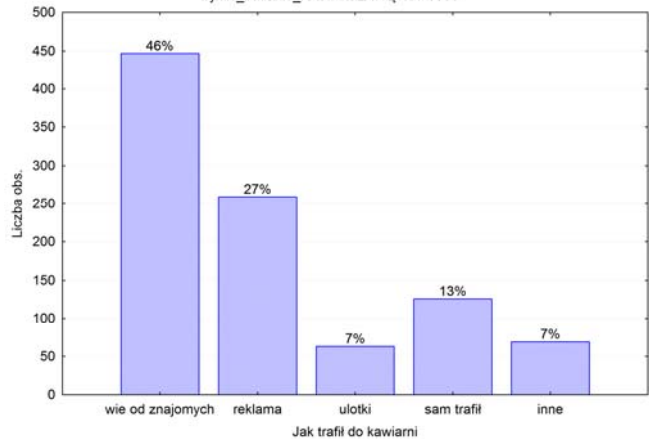
Wstępny etap analizy danych ankietowych obejmuje zazwyczaj proste zestawienia oraz graficzny i liczbowy opis danych zebranych w kwestionariuszach. Do badania empirycznych rozkładów zmiennych można wykorzystać tabele licznosci lub ich graficzną prezentację w postaci wykresów kołowych, histogramów, czy wykresów ramka-wąsy.



Wykres kołowy Osobno palacze
wyniki_badanie_KAWAnaŁAWE_40v*993c



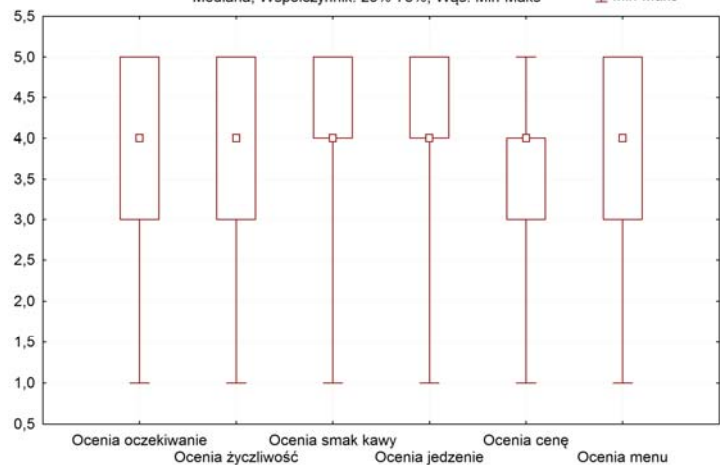
Histogram Jak trafił do kawiarni
wyniki_badanie_KAWAnaŁAWE_40v*993c



Zestawienie odpowiedzi z kwestionariusza sieci *KAWAnaŁAWE* rozpoczniemy od pytania o wydzielenie osobnego miejsca dla palaczy. Jak widać, ponad połowa ankietowanych jest zdania, że należy takie pomieszczenie utworzyć. Jeśli chodzi o skuteczność kampanii reklamowej, to wydaje się, że są powody do optymizmu - prawie co trzecia osoba twierdzi, że dowiedziała się o kawiarni właśnie dzięki niej. Z kolei co druga osoba trafiła do *KAWAnaŁAWE* dzięki znajomym, zgodnie z wynikami przedstawionymi na histogramie.

Dane: Liczności (Identyczne odp. były ignor. ...)			
Klasa	Liczności (Identyczne odp. I Zmienna: Ważna życzliwość (Wiel. dychotomia; zliczana)		
	Liczba	Procent odpow.	Procent przypad.
N=981			
Ważna życzliwość	506	13,18	51,58
Ważna cena	621	16,18	63,30
Ważna wiedza	279	7,27	28,44
Ważne oczekiwanie	729	18,99	74,31
Ważny smak kawy	550	14,33	56,07
Ważne miejsce	401	10,45	40,88
Ważna jakość jedzenia	577	15,03	58,82
Ważne specjalne produkty	47	1,22	4,79
Ważne występy	12	0,31	1,22
Ważne publikacje	3	0,08	0,31
Ważna reputacja	49	1,28	4,99
Ważne inne	64	1,67	6,52
Ogół	3838	100,00	391,23

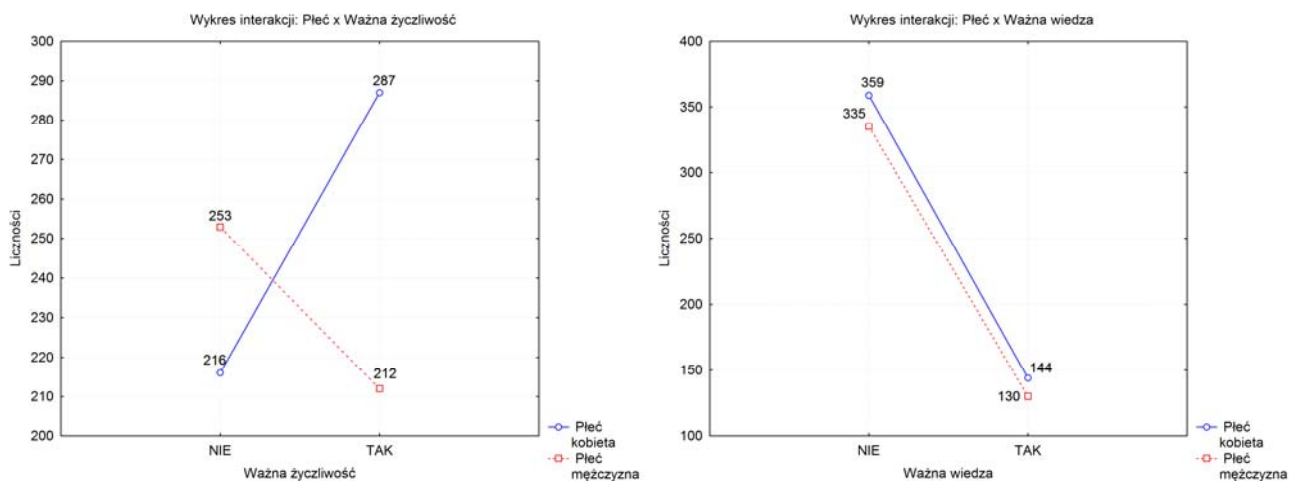
Ramkowy wiele zmiennych
wyniki_badanie_KAWAnaŁAWE_40v*993c
Mediana; Współczynnik: 25%-75%; Wąs: Min-Maks



Zobaczmy teraz, jak kształtują się odpowiedzi na pytanie o najważniejsze czynniki decydujące o wyborze kawiarni. Ponieważ w tym przypadku ankietowani udzielali więcej niż jednej odpowiedzi, wykorzystamy opcję programu *STATISTICA* umożliwiającą tworzenie tabel wielokrotnych odpowiedzi. W związku z tym, że odpowiedzi na pytanie zakodowane zostały w odpowiednich zmiennych binarnie (*TAK* - 1, *NIE* - 0), utworzymy tabelę wielokrotnych dychotomii. Tabela ta składa się z trzech kolumn zawierających odpowiednio: liczbę osób, które zaznaczyły daną odpowiedź, procenty odnoszone do liczby odpowiedzi (sumujące się do 100%) oraz do liczby respondentów (które zwykle nie sumują się do 100%). Okazuje się, że ankietowani najczęściej wybierali czas oczekiwania na zamówienie oraz cenę produktów. Z kolei konfrontacja tych wyników z oceną wybranych cech kawiarni pokazuje, że gorsze noty uzyskały właśnie oczekiwanie na zamówienie, życzliwość

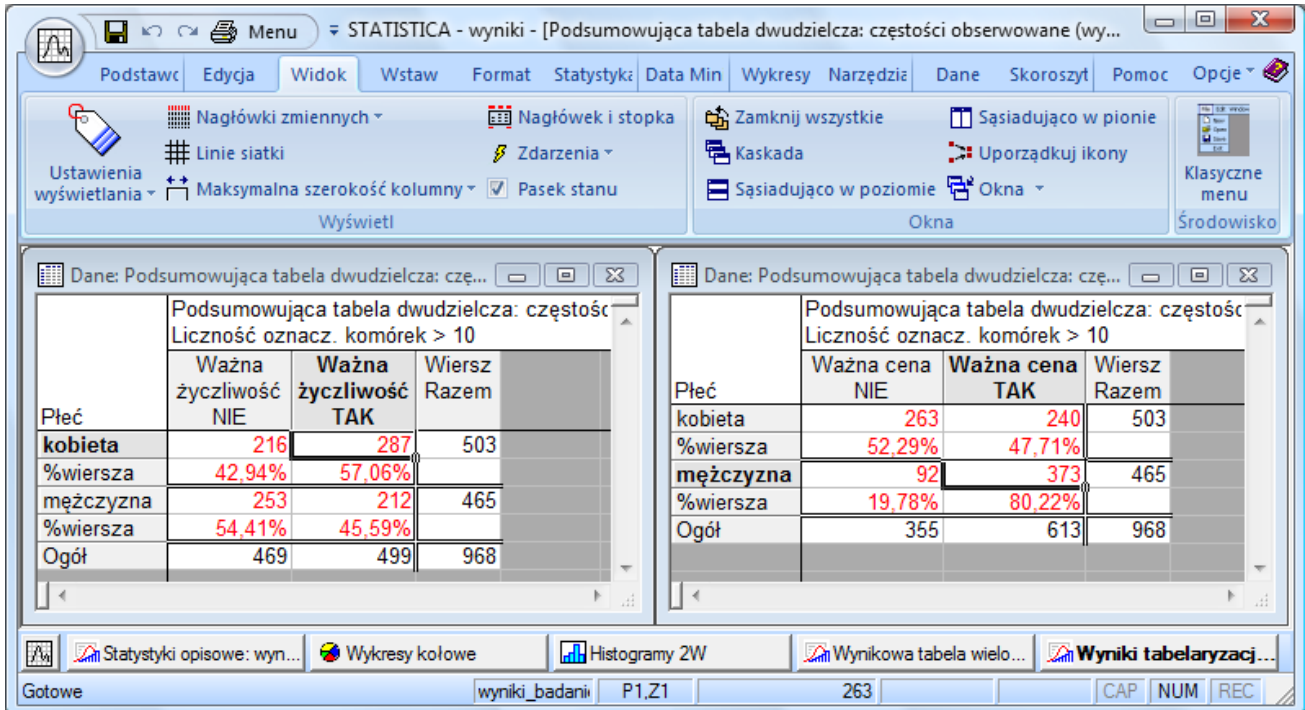


obsługi i cena. Porównanie takie zilustrowane zostało na wielokrotnym wykresie ramka wąsy, zestawiającym medianę oraz pierwszy i trzeci kwartył dla wszystkich ocen. Widać dodatkowo, że klienci byli najbardziej zgodni w ocenie jakości produktów (smak kawy i jedzenie) oraz ceny - dla tych zmiennych rozstęp międzykwartyłowy opisujący zróżnicowanie odpowiedzi jest najmniejszy.

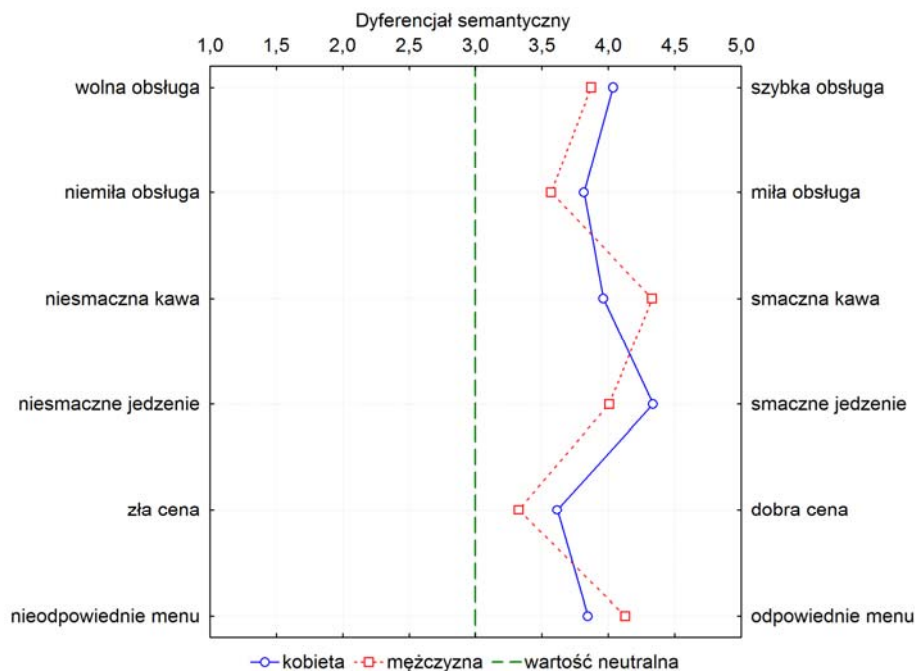


W badaniach ankietowych mamy na ogół do czynienia ze zbiorem danych, który nie jest całkiem jednorodny. W związku z tym zwykle zachodzi potrzeba przeprowadzenia analiz w określonych grupach przypadków. W naszym badaniu chcielibyśmy sprawdzić, czy występują jakieś różnice w ocenie cech kawiarni i czynnikach decydujących o wyborze danego lokalu z punktu widzenia kobiet oraz mężczyzn i w zależności od pory dnia. W tym celu możemy wykorzystać wykresy skategoryzowane umożliwiające jednoczesne porównanie rozkładów zmiennych w poszczególnych grupach oraz wykresy interakcji licznosci, które są graficznym odpowiednikiem tabel dwudzielczych. Dla przykładu porównajmy wykresy interakcji zmiennej określającej płeć i czynników: zyczliwość i wiedza. Pierwszy z nich sugeruje zróżnicowanie oceny zyczliwości między kobietami a mężczyznami. Z kolei drugi wykres pokazuje, że kobiety, na równi z mężczyznami, rzadko kierują się wiedzą obsługi podczas wyboru kawiarni.

Tego typu zestawienia są najczęściej punktem wyjścia do testowania istotności różnic między grupami i uogólniania wniosków z przeprowadzonych badań na całą badaną populację. Zidentyfikowanie prostych relacji między zmiennymi umożliwiają tabele dwudzielcze (lub wielodzielcze), często stosowane w analizie danych ankietowych. Tabele dwudzielcze powstają w oparciu o podział badanej grupy na kategorie wyznaczone przez dwie zmienne i zawierają licznosci odpowiadające poszczególnym kategoriom. Do oceny statystycznej istotności związku między zmiennymi najczęściej stosuje się test niezależności χ^2 Pearsona lub dokładny test niezależności Fishera (w przypadku tabel 2x2). Z kolei siłę powiązań między zmiennymi pozwalają określić współczynniki korelacji, na przykład współczynnik kontyngencji C Pearsona, współczynnik V Cramera i inne.



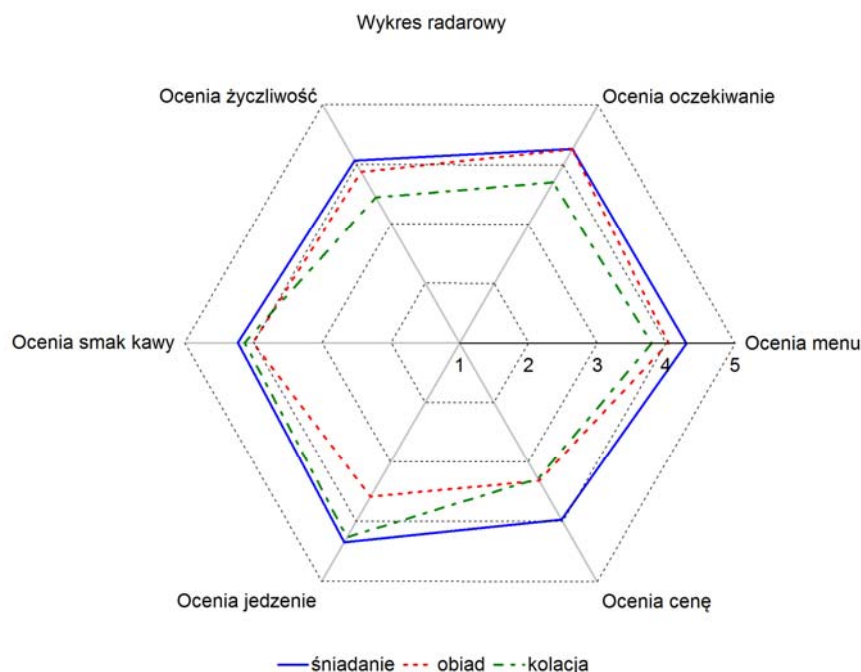
Zgodnie z naszymi przewidywaniami zależność między płcią ankietowanego a kierowaniem się życzliwością obsługi przy wyborze kawiarni okazała się istotna statystycznie. Prawdopodobieństwo testowe dla testu χ^2 jest niższe od zakładanego poziomu istotności 0,05 i wynosi 0,0004. Okazuje się, że płeć jeszcze silniej różnicuje fakt kierowania się ceną podczas wyboru lokalu. Współczynnik kontyngencji, przyjmujący wartości od 0 (brak zależności) do 1 (bardzo silna zależność), jest równy 0,32 dla czynnika: cena i 0,11 dla czynnika: życzliwość.





Na podstawie wyników zawartych w tabelach dwudzielnych możemy stwierdzić, że ceną produktów kieruje się 80% mężczyzn i 48% kobiet. Z kolei życzliwość częściej pod uwagę biorą kobiety - 57% w stosunku do 46% mężczyzn.

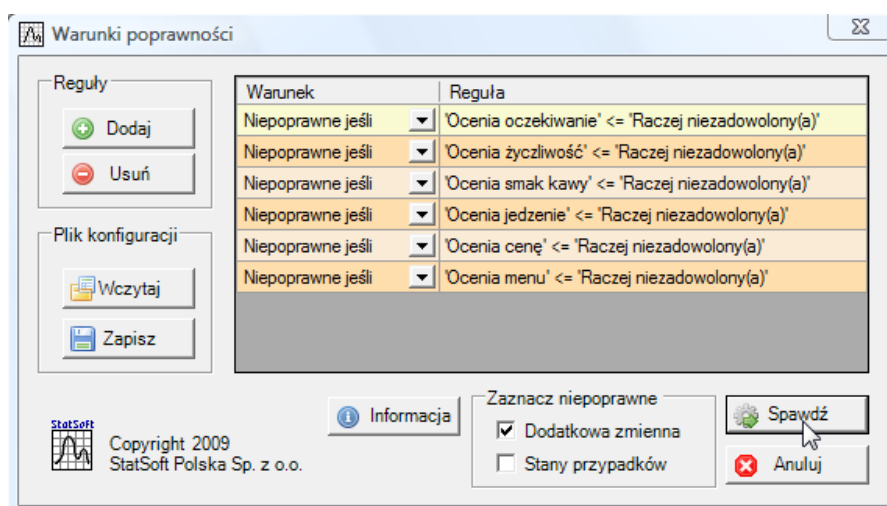
Skoro kobiety kierują się innymi czynnikami przy wyborze kawiarni, być może inaczej od mężczyzn oceniają również wybrane cechy lokalu. Cechy te oceniane były w pięciostopniowej skali: od *bardzo niezadowolony* do *bardzo zadowolony*. Do zestawienia tych zależności wykorzystamy pomocniczo wykres przedstawiający inną skalę porządkową - tzw. dyferencjał semantyczny, dostępny w *STATISTICA dla badań marketingowych i rynkowych*, który jest dodatkiem do programu *STATISTICA*. Wykres ten pokazuje poziom natężenia danej cechy rozpiętej między dwiema przeciwstawnymi kategoriami (np. ocenami: *niesmaczna kawa/smaczna kawa*). Za wartość neutralną przyjmiemy w tej sytuacji kategorię 3 - *nie mam zdania*. Na podstawie wykresu można wyciągnąć wniosek, że badani mężczyźni ogólnie gorzej od kobiet ocenili obsługę lokalu (czas oczekiwania na zamówienie i życzliwość), jakość jedzenia oraz ceny produktów. Kobiety przeważnie gorzej od mężczyzn oceniały z kolei smak kawy i menu.



System wprowadzania danych ankietowych umożliwił automatyczne zapisywanie czasu. Dzięki temu można było utworzyć skategoryzowaną zmienną informującą o porze dnia, w jakiej klient wypełnił ankietę. Wyznaczono następujące kategorie: *śniadanie* (7:00-11:00), *obiad* (11:00-15:00), *kolacja* (15:00-21:00), które uwzględniały zarówno specyficzne dla danej pory pozycje w menu kawiarni, jak i zmiany obsługi. Tym razem przeciętny poziom ocen poszczególnych cech w podziale na porę dnia przedstawimy za pomocą jeszcze innej opcji wizualizacji danych, którą również można znaleźć w *STATISTICA dla badań marketingowych i rynkowych*, a mianowicie wykresu radarowego. Na wykresie tym każda cecha jest przedstawiona na oddzielnej osi, a linie określające porę dnia łączą punkty odpowiadające ocenom poszczególnych cech. Można powiedzieć, że ogólna ocena obsługi



sieci *KAWAnaŁAWĘ* jest zaniżana przez stosunkowo niską ocenę czasu oczekiwania oraz życzliwości obsługi przez klientów zarejestrowanych po godzinie 15:00. Taki wniosek nie dziwi - o tej porze kawiarnie przeżywają natłok klientów, a obsługa jest bardziej zmęczona. Ankietowani najlepiej oceniali menu w porze śniadania, gorzej w porze obiadowej i kolacji, co może mieć związek częściowo ze zniecierpliwieniem związanym z dłuższym oczekiwaniem na zamówienie, a częściowo z niższą oceną jakości jedzenia w porze obiadowej. Widać, że z porą dnia jest również powiązana opinia o cenach produktów.

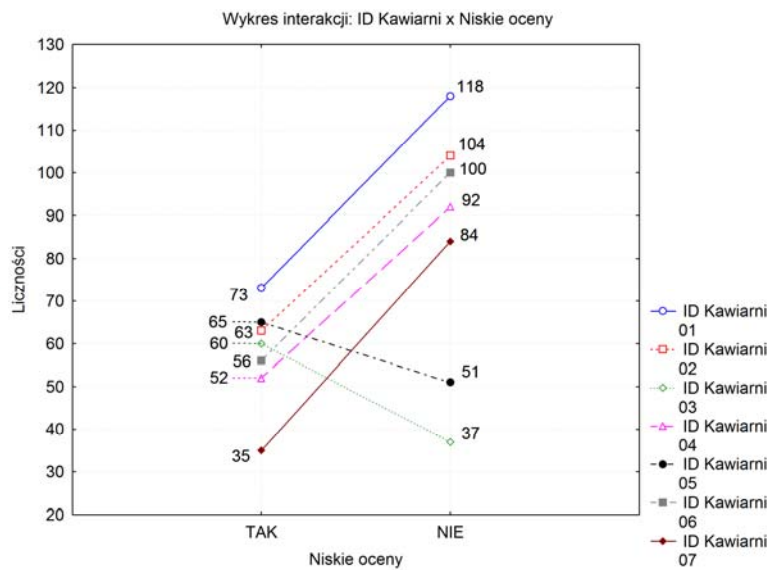


Analiza satysfakcji klientów pokazała, że część z nich znacznie niżej oceniła takie cechy jak: życzliwość obsługi, czas oczekiwania na zamówienie czy jakość jedzenia. Naturalnie nasuwa się zatem pytanie, czy taka ocena dotyczy wszystkich kawiarni, czy jest może związana z jakimś konkretnym lokalem. Aby to sprawdzić wykorzystamy ponownie program *STATISTICA dla badań marketingowych i rynkowych*, a konkretnie opcję umożliwiającą sprawdzenie reguł poprawności danych. Po utworzeniu zmiennej, która wskazuje przypadki „bardzo niezadowolone” lub „raczej niezadowolone” ze wszystkich cech kawiarni zbadamy jej związek ze zmienną *ID kawiarni*. Okazuje się, że związek jest statystycznie istotny, choć na pierwszy rzut oka liczba takich osób w każdej z badanych kawiarni jest zbliżona. Liczności te musimy jednak odnieść do łącznej liczby klientów ankietowanych w danej kawiarni. Takie zestawienie nie pozostawia złudzeń - najgorzej wypadły trzecia i piąta kawiarnia (odpowiednio 62% i 56% „niezadowolonych” klientów). Jeszcze lepiej pokazuje to wykres interakcji. Lokale te charakteryzują się dodatkowo najmniejszą liczbą respondentów, co jest w pewnym stopniu związane z niezadowolaniem klientów - ankietowane były przecież osoby starające się o kartę rabatową lokalu.

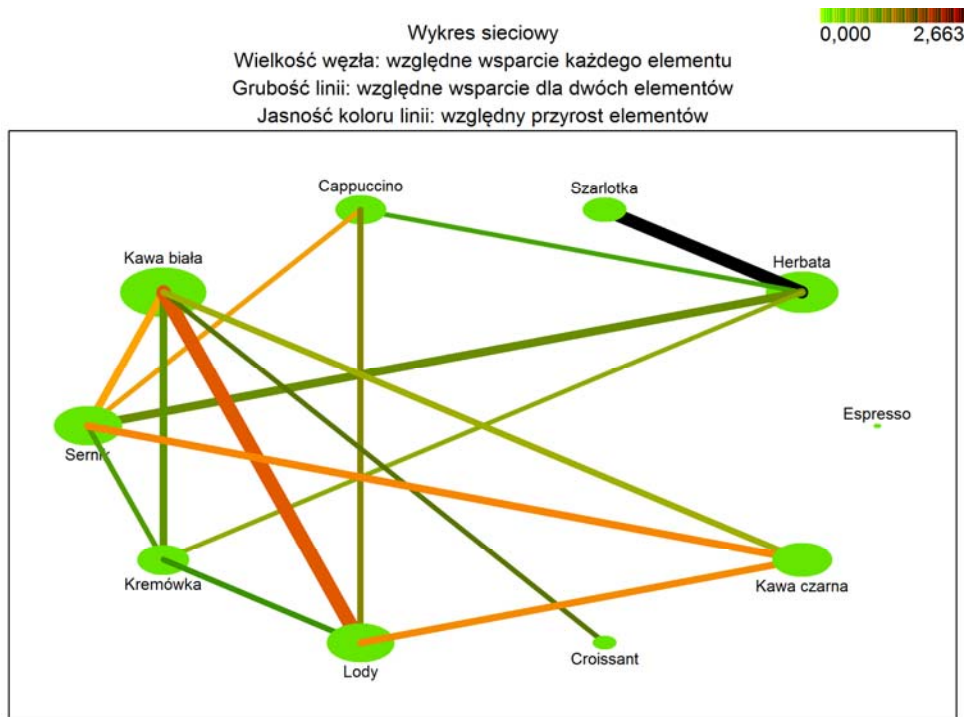
W ostatnim etapie badania przeprowadzimy analizę najczęściej wybieranych produktów i ich zestawów. Wykorzystamy do tego celu moduł *Analiza sekwencji, asocjacji i połączeń STATISTICA (SAL)* zawierający implementację różnych nowoczesnych technik analitycznych, zaprojektowanych do wyszukiwania w dużych zbiorach danych wzorców, zwanych ogólnie „koszykami”. Dzięki niemu możemy wyszukać reguły asocjacji mówiące, które produkty zamawiane są często równocześnie.



Dane: Podsumowująca tabela dwudzi...			
Podsumowująca tabela dwudzielcza: c Liczność oznacz. komórek > 10			
ID Kawiarni	Niskie oceny TAK	Niskie oceny NIE	Wiersz Razem
01	73	118	191
%wiersza	38,22%	61,78%	
02	63	104	167
%wiersza	37,72%	62,28%	
03	60	37	97
%wiersza	61,86%	38,14%	
04	52	92	144
%wiersza	36,11%	63,89%	
05	65	51	116
%wiersza	56,03%	43,97%	
06	56	100	156
%wiersza	35,90%	64,10%	
07	35	84	119
%wiersza	29,41%	70,59%	
Ogół	404	586	990



Graficzne podsumowanie reguł asocjacji stanowi tzw. wykres sieciowy. Wielkość węzłów (elips) ilustruje popularność odpowiedniego produktu, a grubość każdej linii określa udział dwóch połączonych tą linią produktów (zamawianych razem). Z kolei kolor linii świadczy o mierze skorelowania - im ciemniejsza barwa, tym silniejsza korelacja między danymi produktami. Najczęściej wskazywanymi przez ankietowanych napojami są: kawa biała (wskazało tak 32% badanych), herbata (27%) i kawa czarna (23%), natomiast najpopularniejsze desery to sernik (26%) i lody (26%). Jeśli chodzi o zestawy, to najczęściej wybierane były: kawa biała z lodami oraz herbata z szarlotką (zmiennie te były dodatkowo najsilniej skorelowane, czyli jeśli klient zamawiał szarlotkę, to najczęściej z herbatą).





Wyniki wszystkich przeprowadzonych wcześniej analiz zostały automatycznie umieszczone w skoroszybie, który stanowi domyślny sposób zarządzania wynikami w programie *STATISTICA*. Skoroszyt może następnie zostać zapisany w formacie HTML, co umożliwi przekazanie wyników osobom nieposiadającym programu *STATISTICA*, a nawet publikowanie ich w portalach internetowych (na przykład na stronie internetowej kawiarni). Program daje również możliwość automatycznego umieszczania wyników analizy bezpośrednio w dokumencie programu Word, dzięki czemu od razu uzyskaliśmy raport z prowadzonych badań.

Podsumowanie

Badanie satysfakcji klientów sieci *KAWAnaŁAWĘ* rozpoczęto od przygotowania kwestionariusza i przeprowadzenia badania pilotażowego. Pilotaż pokazał szereg problemów, których udało się później uniknąć dzięki zmianie doboru próby oraz wykorzystaniu internetowego systemu do wprowadzania danych ankietowych. System umożliwił bezpośrednio gromadzenie informacji z wszystkich ankiet w centralnej bazie danych i ograniczył do minimum czyszczenie danych, dzięki czemu etap ten został sprowadzony do analizy braków danych i niepoprawnych odpowiedzi na pytania otwarte.

Na początku analizy, w oparciu o szereg podstawowych narzędzi analitycznych i graficznych z rodziny *STATISTICA*, utworzone zostały proste zestawienia. Pokazały one między innymi, że ponad połowa klientów chciałaby utworzenia w kawiarni osobnych pomieszczeń dla palaczy. Skuteczność kampanii reklamowej potwierdził fakt, że co trzecia osoba dowiedziała się o kawiarni właśnie dzięki niej. Najważniejszymi czynnikami decydującymi o wyborze kawiarni okazały się czas oczekiwania na zamówienie oraz cena.

W dalszej kolejności przeprowadzono analizę zależności między płcią klienta oraz porą dnia a zadowoleniem z wybranych cech kawiarni. Zwykle kobiety niżej od mężczyzn oceniały smak kawy oraz menu. Z kolei analiza ocen w zależności od pory dnia pokazała, że najgorzej kawiarnie są oceniane w porze kolacji. Do przedstawienia tych zależności wykorzystano opcje graficzne dostępne w programie *STATISTICA dla badań marketingowych i rynkowych*. Zawarty tam moduł do sprawdzania poprawności danych pomógł również zidentyfikować lokale, w których oceny były dużo niższe niż w pozostałych kawiarniach. W ostatnim etapie analizy wykorzystano bardziej zaawansowany moduł - *Analiza sekwencji, asocjacji i połączeń STATISTICA (SAL)*. Umożliwił on znalezienie reguł mówiących, jakie zestawy produktów są zamawiane najczęściej.

Zobaczyliśmy zatem, jak można zapewnić wsparcie każdego etapu badania ankietowego – od gromadzenia i przygotowania danych, przez cały proces analityczny, obejmujący zarówno proste zestawienia, jak i zaawansowane techniki analizy wielowymiarowej, aż po automatyczne generowanie raportów.



Literatura

1. G.A. Churchill, *Badania marketingowe. Podstawy metodologiczne*, Wydawnictwo Naukowe PWN, Warszawa 2002.
2. B.E. Hayes, *Measuring customer satisfaction*, ASQ Quality Press, 1997.
3. N. Hill, J. Alexander, *Pomiar satysfakcji i lojalności klientów*, Oficyna Ekonomiczna, Kraków 2003.
4. M. Łapczyński, A. Sagan, *Analiza danych ankietowych*, Materiały kursowe StatSoft Polska, 2009.