

Iwona Czerska

Uniwersytet Ekonomiczny we Wrocławiu

e-mail: iwona.czerska@ue.wroc.pl

**SEGMENTACJA FACEBOOKOWICZÓW –
UJĘCIE ILOŚCIOWE**

**SEGMENTATION OF FACEBOOK USERS –
QUANTIFICATION**

DOI: 10.15611/noz.2015.3.03

JEL Classification: C12, C38, C83, M31

Streszczenie: Przedmiotem niniejszego opracowania jest analiza statystyczna wyników ankiety internetowej, mająca na celu dokonanie segmentacji rynku użytkowników Facebooka. Zastosowano metody doboru nielosowego: dobór wygodny i dobór kuli śniegowej. Ponieważ próba nie była reprezentatywna, nie dawała podstaw do statystycznych wnioskowań o populacji użytkowników Facebooka. Wyniki badania były jedynie elementem wstępnej diagnozy, opisem zastanego stanu rzeczy. Na podstawie wybranych kryteriów segmentacyjnych oraz wykorzystując uogólnioną metodę *k*-średnich i aglomeracyjną metodę Warda, wyodrębniono trzy skupienia: „Poinformowani asekuranci” (56,31% próby), „Zaangażowani ryzykanci” (20,39% próby) oraz „Wytrwali asertywni” (23,30% próby). Profilowania segmentów dokonano na podstawie kryteriów psychograficznych i behawioralnych. Statystyczna istotność zależności między skupieniami internautów a poszczególnymi zmiennymi została potwierdzona testem chi-kwadrat.

Słowa kluczowe: segmentacja, facebookowicze, metoda Warda, metoda *k*-średnich, profilowanie segmentów.

Summary: The purpose of this study is the statistical analysis of results of the online survey in order to make market segmentation of Facebook users. Non-random sampling methods were used: convenience sampling and snowball sampling method. Because the sample was not representative, it did not give rise to statistical inferences about the population of Facebook users. The survey results were only part of the initial diagnosis, a description of the existing state of affairs. On the basis of selected market segmentation criteria and using the generalized *k*-means clustering algorithm and the Ward agglomeration method three clusters were formed: “Informed over-cautious persons” (56,31% of the sample), “Committed risk-takers” (20,39% of the sample) and “Persistent assertive people” (23,30% of the sample). Segments were profiled on the basis of psychographic and behavioral criteria. The statistical significance of the relationship between clusters of Internet users and individual variables was confirmed by the chi-square test.

Keywords: segmentation, Facebook users, Ward’s method, *k*-means method, segments profiling.

1. Wstęp

Segmentacja polega na podziale heterogenicznego (niejednorodnego) rynku na względnie homogeniczne grupy nabywców na podstawie wybranych kryteriów segmentacyjnych. Proces segmentacji pozwala wyodrębnić grupy klientów o podobnych cechach, potrzebach, motywacjach, zachowaniu na rynku. Wyodrębnione grupy odbiorców odróżniają się od siebie reakcjami na narzędzia marketingu-mix.

Najwcześniejsze podejście do analiz segmentacyjnych stanowi segmentacja normatywna, wywodząca się z badań E. Chamberlina [1933] dotyczących rynków monopolistycznych oraz badań J. Robinson [1933], związanych z konkurencją niedoskonałą. Maksymalizacja efektywności działań przedsiębiorstwa, którym odpowiadają zróżnicowane reakcje konsumentów, jest podstawą podejścia normatywnego [Dickson, Ginter 1987, s. 1-10]. Kryteria segmentacyjne, takie jak współczynniki elastyczności, współczynniki funkcji reakcji opartych na regresji wielorakiej oraz wskaźniki utargów krańcowych, okazały się trudne do praktycznego zastosowania [Sobczak 2010, s. 61-62]. Z tego względu model normatywny, stanowiący idealne podejście teoretyczne, obciążony jest wadami. Po pierwsze, ogranicza segmentację do pojedynczej zmiennej. Po drugie, w niewystarczającym stopniu uwzględnia aktualne ograniczenia kierownicze i instytucjonalne. Wreszcie, zawodzi w warunkach konkurencji [Walesiak 2000, s. 195].

Współcześnie można wyróżnić następujące modele segmentacji empirycznej: *a priori*, *post hoc*, hybrydowe (mieszane). Segmentacja *a priori*, zwana także zdroworozsądkową [Dolnicar, Leisch 2004, s. 189-207], bazuje na zmiennych obserwowalnych arbitralnie wybranych przez badacza na podstawie doświadczenia, wcześniejszych badań jakościowych i ilościowych bądź analizy zewnętrznych danych wtórnych. Zaletami modelu *a priori* są łatwo zdefiniowane segmenty, brak konieczności stosowania skomplikowanych metod statystyczno-ekonometrycznych i niska koszty- i czasochłonność. Do wad zalicza się brak charakteru eksploracyjnego oraz niski obiektywizm.

W segmentacji *post hoc*, inaczej zwanej *a posteriori* [Mazanec 2000], wykorzystuje się procedury klasyfikacyjne do podziału zbioru konsumentów na bazie danych statystycznych zgromadzonych w wyniku badań pierwotnych lub wtórnych. Najczęściej stosowane metody analizy wielowymiarowej w tym podejściu to analiza głównych składowych, analiza *conjoint*, hierarchiczna aglomeracyjna analiza skupisk, metoda *k*-średnich [Sagan 2009, s. 22-23]. Główną zaletą segmentacji *post hoc* jest charakter eksploracyjny badań. Poza tym prawidłowe zastosowanie wymienionych technik statystycznych może się przyczynić w konsekwencji do osiągnięcia przewagi konkurencyjnej. Wadą tego modelu segmentacji jest dość skomplikowane podejście metodologiczne.

Kombinacją podejścia *a priori* i *post hoc* jest model segmentacji hybrydowej. Ten rodzaj podejścia stanowi użyteczne narzędzie segmentacji wieloetapowej, jednakże wymaga dość dużej próby [Sobczak 2010, s. 63-64].

Generalnie proces segmentacji powinien być, według Marka Prymona, odkrywczy i twórczy, przeprowadzony wcześniej przez innych badaczy nie gwarantuje bowiem sukcesu w kolejnym roku oraz nie daje przewagi konkurencyjnej [Prymon 2009, s. 88]. Segmenty rynkowe rozwijają się bardzo dynamicznie ze względu na ciągły postęp techniczno-technologiczny, zmieniające się potrzeby i preferencje zakupowe nabywców oraz różne reakcje na instrumenty marketingu.

Aby właściwie dokonać podziału rynku nabywców, przeprowadza się segmentację jakościową i/lub ilościową. W ujęciu analitycznym proces ten oparty jest na wielowymiarowej analizie danych marketingowych, której złożoność, wielopłaszczyznowość i różnorodność wymaga zastosowania specjalistycznych narzędzi, systemów i programów informatycznych w celu gromadzenia, porządkowania, analizy i raportowania danych.

Do segmentacji opisowej (deskryptywnej) można zastosować metodę Warda, czyli hierarchiczną metodę aglomeracyjną analizy skupień, w której wszystkie zmienne traktuje się jako zmienne niezależne (wejściowe) [Migut 2009, s. 75-76]. Ze zbioru danych wyodrębnia się obiekty podobne do siebie i łączy w grupy. W wyniku działania tej metody z jednego niejednorodnego zbioru danych otrzymuje się grupę kilku jednorodnych podzbiorów. Obiekty znajdujące się w tym samym zbiorze uznaje się za „podobne do siebie”, natomiast obiekty z różnych zbiorów traktowane są jako „niepodobne”. Metoda Warda charakteryzuje się dużą efektywnością w wykrywaniu rzeczywistych skupień. Metoda ta minimalizuje wariancję wewnątrzgrupową w ramach każdego skupienia przy jednoczesnej maksymalizacji wariancji pomiędzy skupieniami, co zapewnia homogeniczność wewnątrz skupień i heterogeniczność między skupieniami [Sobczak 2010, s. 100].

Jedną z najważniejszych metod niehierarchicznych analizy skupień jest metoda k -średnich. Jako metoda optymalizacyjno-iteracyjna pozwala na tworzenie k skupień w możliwie największym stopniu różniących się od siebie. Liczbę skupień wybiera się na podstawie przesłanek merytorycznych bądź szacuje metodami hierarchicznymi. W wyniku grupowania bada się średnie dla każdego skupienia w każdym wymiarze celem oszacowania różnic między skupieniami. Metoda k -średnich jest wykorzystywana do analizy dużych ilości danych w celu redukcji informacji do kilku kategorii [Pietrzykowski, Kobus 2006, s. 302].

Najlepszy układ skupień otrzymuje się w wyniku zastosowania uogólnionej analizy skupień metodą k -średnich. Program statystyczny, dzięki sprawdzianowi krzyżowemu, automatycznie określa najbardziej odpowiednią liczbę skupień (segmentów) [Data Miner 2015]. Metoda ta w odróżnieniu od klasycznych metod analizy skupień pozwala obsługiwać cechy jakościowe. Proces wyodrębnienia liczby segmentów działa na podstawie algorytmu, który dzieli zbiór wejściowy kolejno na coraz większą liczbę segmentów i ocenia precyzję podziału dla każdego z nich. W tej metodzie miarą precyzji podziału jest przeciętna odległość elementów zbioru wejściowego od środka segmentu, w jakim się znajdują. Jeśli w wyniku kolejnego podziału zbudowany model poprawia się w stosunku do poprzedniego modelu

w mniejszym stopniu, niż określono to w wartości progowej (domyślnie jest to 5%), algorytm zatrzymuje swoje działanie [Migut 2009, s. 81]. Mając wyróżnioną liczbę segmentów, program podaje informacje o średnich oraz licznosci poszczególnych skupień. Owe informacje są wykorzystywane do profilowania otrzymanych segmentów.

W pracy podjęto próbę segmentacji rynku użytkowników portalu społecznościowego Facebook, wykorzystując deskryptywne podejście typu *post hoc*. Końcowym etapem procesu segmentacji w niniejszym opracowaniu było profilowanie otrzymanych segmentów.

2. Opis badania

W miesiącach luty-maj 2015 r. przeprowadzono w Internecie badanie własne, które skierowane było do polskich użytkowników Facebooka. Zastosowano metody doboru nielosowego – nieprobabilistycznego, a mianowicie dobór wygodny, polegający na wyborze osób badanych z grona znajomych autora, a następnie dobór kuli śniegowej, polegający na dotarciu za pośrednictwem znajomych do kolejnych znanych im osób użytkujących Facebooka [Metody doboru próby badawczej 2015]. W ankie-tyzacji wzięło udział 618 respondentów.

Ponieważ próba nie była reprezentatywna, nie dawała podstaw do statystycznych wnioskowań o populacji użytkowników Facebooka. Wyniki badania były jedynie elementem wstępnej diagnozy, opisem zastanego stanu rzeczy.

Jako kryteria segmentacyjne wybrano: demograficzne (płeć, wiek, wielkość gospodarstwa domowego, faza cyklu życia rodzinnego, wykształcenie), socjoekonomiczne (dochód, aktywność zawodowa), psychograficzne (cechy osobowościowe), behawioralne (częstotliwość korzystania, korzyści oczekiwane). Początkowo w celu wyodrębnienia segmentów rynku użytkowników Facebooka zastosowano dwie metody analizy skupień: hierarchiczną metodę Warda oraz niehierarchiczną metodę *k*-średnich (z zadaną liczbą skupień wynikającą z dendrogramu metody Warda). Ostatecznie jednak wykorzystano uogólnioną analizę skupień z wykorzystaniem algorytmu *k*-średnich, a następnie metodę Warda.

3. Wyniki

W badanej grupie znalazły się 472 kobiety (76,38%) i 146 mężczyzn (23,62%). Najliczniejszą grupę użytkowników portalu stanowiły osoby młode do 24. roku życia (71,52%), a najmniej liczny okazał się przedział 55-64 lata (0,65%).

Zbiór obejmował 38 zmiennych, które przekodowano i wystandaryzowano w celu możliwości dokonywania porównań między nimi. Już na wstępie wyeliminowano z analizy następujące zmienne: „płeć” – ze względu na przeważającą liczbę kobiet w badaniu, „wiek” – ze względu na dominującą liczbę młodych osób w ba-

daniu, „dochód” – ze względu na zbyt duży odsetek odmowy odpowiedzi respondentów na pytanie o dochód netto na osobę w rodzinie. Podczas analizy porównano liczne alternatywne rozwiązania, które otrzymano z wykorzystaniem różnych zestawów zmiennych, przyjmując odmienną liczbę segmentów. Ostatecznie klasyfikację zaprezentowaną w artykule wybrano z powodu istnienia wyraźnych różnic między segmentami oraz możliwości oceny i wykorzystania otrzymanych wyników w przyszłości.

W celu wyodrębnienia segmentów rynku użytkowników Facebooka początkowo zastosowano dwie metody analizy skupień: hierarchiczną metodę Warda oraz niehierarchiczną metodę k -średnich (z zadaną liczbą skupień wynikającą z dendrogramu metody Warda). Nie dały one jednak zadowalających wyników, co potwierdza fakt, iż w wypadku dużych prób, rzędu kilkuset respondentów, wymaganych w procesie segmentacji rynku, wspomniane metody nie są efektywne [Migut 2009, s. 77; Kąciak 2011, s. 191]. W takich sytuacjach zaleca się stosowanie uogólnionej analizy skupień, a następnie metody hierarchicznej.

W wyniku zastosowania (dla pięciu zmiennych wejściowych¹) uogólnionej analizy skupień z wykorzystaniem algorytmu k -średnich, przy losowym wyborze k obserwacji, odległości euklidesowej i V -krotnym sprawdzianem krzyżowym, wyodrębniono trzy skupienia. Według Adama Sagana optymalna liczba skupień wynosi od trzech do sześciu [Sagan 2010]. Statystyczna istotność zależności między skupieniami internautów a poszczególnymi zmiennymi została potwierdzona testem chi-kwadrat (tab. 1).

Tabela 1. Test chi-kwadrat

| Zmienna | <i>df</i> | Chi-kwadrat | Wartość <i>p</i> | G-kwadrat | Wartość <i>p</i> |
|------------|-----------|-------------|------------------|-----------|------------------|
| Ryzyko | 2 | 550,1611 | 0,000000 | 571,5510 | 0,000000 |
| Asertywny | 2 | 195,1273 | 0,000000 | 200,3323 | 0,000000 |
| Czas | 10 | 98,8595 | 0,000000 | 104,5617 | 0,000000 |
| Kontakty | 2 | 131,1545 | 0,000000 | 121,8472 | 0,000000 |
| Wydarzenia | 2 | 6,6019 | 0,036848 | 6,8500 | 0,032549 |

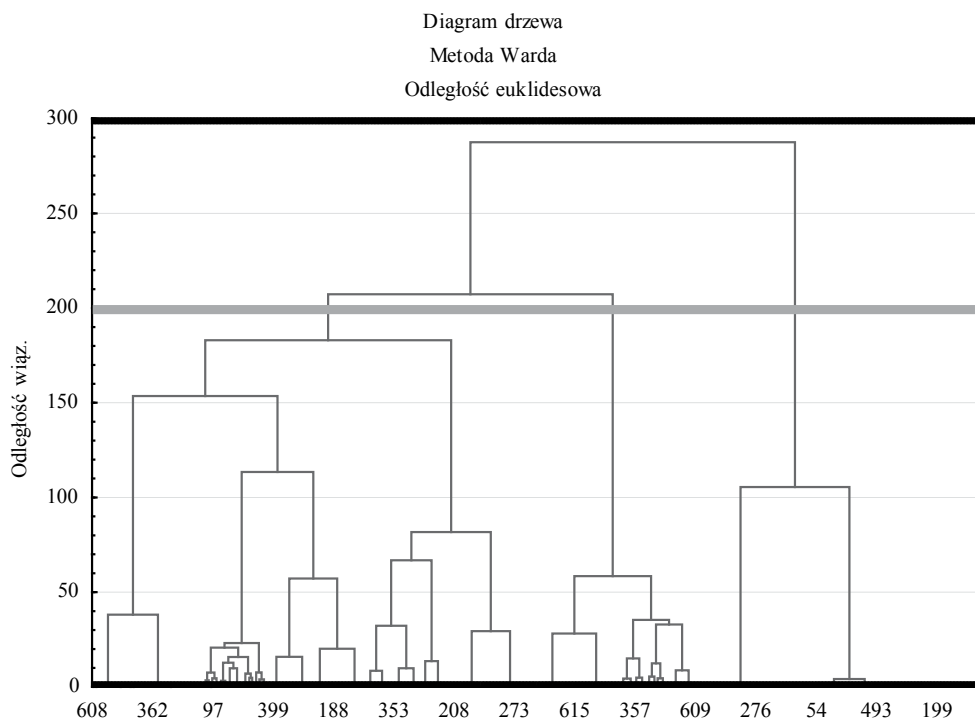
Źródło: opracowanie własne.

W kolejnym kroku zastosowano aglomeracyjną metodę Warda (dla pięciu zmiennych), aby potwierdzić słuszność wyboru trzech skupień. W jej wyniku otrzyma-

¹ Objaśnienia zmiennych wejściowych:

- „Ryzyko” – na podstawie pytania: „Proszę określić siebie przez wybranie poniższych cech osobowościowych”,
- „Asertywny” – na podstawie pytania: „Proszę określić siebie przez wybranie poniższych cech osobowościowych”,
- „Czas” – na podstawie pytania: „Od kiedy jest Pan(i) użytkownikiem Facebooka?”,
- „Kontakty” – na podstawie pytania: „Czego oczekuje Pan(i) od Facebooka?”,
- „Wydarzenia” – na podstawie pytania: „Czego oczekuje Pan(i) od Facebooka?”.

no dendrogram (pionowy wykres sopelkowy) (rys. 1). Za miarę odległości przyjęto odległość euklidesową. Do określenia liczby skupień w analizowanej zbiorowości konieczne było ustalenie punktu odcięcia dendrogramu, czyli odległości, powyżej której poszczególne elementy traktowane są jako różne. W niniejszej analizie punkt odcięcia ustalono na poziomie 200 (gruba szara linia). Dla takiego poziomu odległości wiązania można wyróżnić trzy różne grupy facebookowiczów. Gdyby jednak przyjąć granicę na poziomie 250, wówczas otrzymano by tylko dwa skupienia. Z kolei dla granicy 100 wyodrębniono by aż 7 skupień, a dla 150 – 5. Wynik aglomeracji nie jest jednoznaczny, gdyż punkt odcięcia można określić dowolnie, np. w sposób arbitralny, kierując się względami praktycznymi. W niniejszej analizie kierowano się wynikiem uogólnionej analizy skupień i ostatecznie przyjęto liczbę trzech skupień.



Rys. 1. Diagram drzewa (dendrogram) – metoda Warda

Źródło: opracowanie własne.

4. Profilowanie segmentów

Ostatnim etapem w niniejszej analizie było profilowanie otrzymanych segmentów. Na podstawie analizy średnich poszczególnych zmiennych dla każdego skupienia oraz analizy tabel liczości dla tych zmiennych opracowano poniższe profile facebookowiczów.

Profil I – segment „Poinformowani asekuranci” (56,31% respondentów)

Do tego segmentu należą osoby, które chcą być na bieżąco z nowymi wydarzeniami na Facebooku (78,16% skupienia). Nie są one natomiast zainteresowane nawiązywaniem kontaktów. Żaden respondent tego segmentu nie jest skłonny do podejmowania ryzyka, a tylko 12,64% osób uważa siebie za osobę asertywną. Najwięcej internautów spośród wszystkich segmentów (70,3%) jest użytkownikami portalu od 2 do 4 lat.

Profil II – segment „Zaangażowani ryzykanci” (20,39% respondentów)

100% przedstawicieli tej grupy internautów to osoby skłonne do podejmowania ryzyka. Na Facebooku interesują się przeważnie nowymi wydarzeniami (84,13% skupienia), a 22,22% użytkowników szuka nowych kontaktów. 43,65% skupienia określa siebie jako osoby asertywne. Przeważająca liczba respondentów użytkuje portal od 2 do 10 lat.

Profil III – segment „Wytrwali asertywni” (23,30% respondentów)

77,08% skupienia to osoby asertywne, które stanowią 52,86% osób określających siebie tą cechą we wszystkich segmentach. Zaledwie 9,03% skupienia to osoby skłonne do działań ryzykownych. Najwięcej osób w tym segmencie jest użytkownikami portalu od 5 do 10 lat. 87,5% internautów koncentruje się na nowych wydarzeniach, a 58,33% osób chce być w kontakcie ze znajomymi.

5. Zakończenie

W wyniku segmentacji facebookowiczów, przy użyciu metod wielowymiarowej analizy statystycznej, otrzymano trzy zróżnicowane skupienia: „Poinformowani asekuranci” (56,31% próby), „Zaangażowani ryzykanci” (20,39% próby) oraz „Wytrwali asertywni” (23,30% próby). Profilowania segmentów dokonano na podstawie kryterium psychograficznego i behawioralnego.

Jak wspomniano w opracowaniu, segmentację powinno się przeprowadzać regularnie, uwzględniając najnowsze dane, ze względu na dynamicznie zmieniający się rynek, zmieniające się potrzeby, preferencje i oczekiwania klientów. Bez segmentacji bowiem nie jest możliwe wyciągnięcie konstruktywnych wniosków dla biznesu.

Coraz szersze wykorzystywanie programów statystycznych opartych na skomplikowanych procedurach i algorytmach również znacznie podnosi wartość każdej próby segmentacji rynku nabywców. Zastosowane przez autora narzędzia statystyczno-ekonometryczne były niezbędne do ilościowego ujęcia procesu segmentacji rynku.

Literatura

- Chamberlin E.H., *Theory of Monopolistic Competition*, Harvard University Press, Cambridge 1933.
- Data Miner, StatSoft Polska, <http://www.statsoft.pl/Programy/Ogolna-charakterystyka/Moduly-analityczne/Data-Miner> (2.02.2015).
- Dickson P.R., Ginter J.L., *Market segmentation, product differentiation and marketing strategy*, „Journal of Marketing” 1987, no. 51.
- Dolnicar S., Leisch F., *Delivering the wright tourist service to the wright people – a comparison of segmentation approaches*, „The Journal of Quality Assurance in Hospitality and Tourism” 2004, no. 5.
- Mazanec J., *Market Segmentation*, [w:] *Encyclopedia of Tourism*, ed. J. Jafari, Routledge, London 2000.
- Metody doboru próby badawczej, Biuro Badań Marketingowych, Poland The Marketing Research Agency, <http://www.bbm.com.pl/metodologia-badawcza/metody-doboru-proby-badawczej/> (2.02.2015).
- Migut G., *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku*, StatSoft Polska, 2009, http://www.statsoft.pl/Portals/0/Downloads/Zastosowanie_techinik.pdf.
- Pietrzykowski R., Kobus P., *Zastosowanie modyfikacji metody k-średnich w analizie portfelowej*, Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego „Ekonomika i Organizacja Gospodarki Żywnościowej”, nr 60, Wydawnictwo SGGW, Warszawa 2006.
- Prymon M., *Badania marketingowe w aspektach menedżerskich*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2009.
- Robinson J., *The Economics of Imperfect Competition*, McMillan, Cambridge 1933.
- Sagan A., *Metody statystyczne w marketingu i badaniach rynku*, materiały szkoleniowe StatSoft w ramach kursu obsługi programu Statistica, Kraków 2010.
- Sagan A., *Podejście modelowe w segmentacji rynku*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie nr 800, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków 2009.
- Sobczak E., *Segmentacja rynków zagranicznych*, Prace Naukowe nr 120 Uniwersytetu Ekonomicznego we Wrocławiu, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2010.
- Walesiak M., *Segmentacja rynku. Kryteria i metody*, praca wykonana w ramach projektu badawczego nr 1 H02B 011 16 finansowanego przez Komitet Badań Naukowych w latach 1999 i 2000, 2000.