

Renata Jaroszevska-Brudnicka
Wyższa Szkoła Bankowa w Toruniu

Robert Brudnicki, Rafał Śpiewak
Wyższa Szkoła Gospodarki w Bydgoszczy

Big Data Analysis – nowa, lepsza metoda badań konsumentów usług turystycznych w regionach? Analiza porównawcza na przykładzie badań ruchu turystycznego prowadzonych w regionie kujawsko-pomorskim

Zarys treści: Zbieranie statystyk dotyczących turystyki jest „niezbędne do pomiaru wielkości, skali, wpływu i wartości turystyki na różnych szlakach geograficznych – od poziomu globalnego, poprzez krajowy aż do lokalnego miejsca docelowego” (Hall, Page 2014). Większość statystyk (wielkość turystyki, wydatki oraz cechy związane z turystami i podróżami) jest zbierana tradycyjnie za pomocą wywiadów lub ankiet (np. w gospodarstwach domowych, w regionach recepcyjnych, podczas tranzytu w podróży turystycznej). Tego typu metody pozyskiwania danych były dotychczas postrzegane jako podstawowe źródła danych dla statystyk turystyki krajowej i wyjazdowej. Celem artykułu jest identyfikacja i ocena potencjału analizy zbiorów danych typu Big Data, które w dobie rosnącej informatyzacji, zastosowania nowoczesnych technologii ICT (Information and Communications Technologies – ICT), w obszarze badań empirycznych nad turystyką wydają się być jedną z najbardziej aktualnych kwestii. Z uwagi na rosnące znaczenie, a także coraz łatwiejszy dostęp do tego typu danych, celem jest analiza porównawcza procedur badawczych oraz ocena przydatności danych zaliczanych do Big Data, jako potencjalnie wartościowych w perspektywie kolejnych planowanych badań, monitorujących ruch turystyczny w regionie (studium przypadku regionu kujawsko-pomorskiego).

Słowa kluczowe: statystyka turystyki, badania ruchu turystycznego, Big Data, marketing turystyczny, badania popytu turystycznego, zachowania turystyczne, monitoring ruchu turystycznego.

Wprowadzenie

Dopływ wartościowych informacji i obiektywnych danych wymaga zastosowania odpowiednich procedur i metod ich zbierania, a następnie przetworzenia ich w użyteczne informacje. Obiektywną charakterystykę i rozpoznanie popytu

i podaży turystycznej w naszym kraju utrudniają luki informacyjne w statystyce publicznej, która nadal w przypadku wielu diagnoz wykonywanych dla potrzeb kierunkowania rozwoju turystyki w regionach stanowi podstawowy punkt odniesienia i zarazem źródło informacji na temat regionalnego czy krajowego rynku turystycznego. Istotnym utrudnieniem w interpretacji uzyskiwanych wyników jest fakt, że ruch turystyczny oraz związana z nim konsumpcja dóbr i usług determinowane są przez wyjątkowo dużą liczbę różnych czynników, a przy tym wszystkim turystyka jest zjawiskiem wyjątkowo dynamicznym i czułym na wszelkiego rodzaju zmiany (ekonomiczne, społeczne, polityczne, pogodowe itp.) (por. Alejziak 2010). Pozyskiwanie poprawnych (obiektywnych), użytecznych danych każdorazowo wymaga również rozwiązywania trudnych problemów organizacyjnych badań (problemy metodologiczne, racjonalność ekonomiczna). Zbieranie, porządkowanie, interpretowanie informacji jest bowiem związane z pewnym wysiłkiem, przez co jest procesem zasobochłonnym (koszt, czas, odpowiednie zasoby ludzkie itp.). Z uwagi na to, że cel badawczy w tego typu badaniach często można zrealizować posługując się różnymi metodami, procesom poznawczym towarzyszy problem wyboru odpowiedniej metodyki, czyli takiej, która zapewni maksymalizację użyteczności wyników przy minimalnym wykorzystaniu dostępnych zasobów (finansowych, ludzkich, czasowych i innych).

Problemy związane z określeniem wielkości i struktury popytu turystycznego (zwłaszcza w wymiarze statystycznym) determinują stan ciągłego poszukiwania nowych źródeł i sposobów zdobywania informacji na ten temat (por. Saluveer, Raun i in. 2020). W takiej sytuacji konieczne jest sięganie do informacji pierwotnych. Poza badaniami empirycznymi (badania niewyczerpujące, oparte na technikach ankietowych i pozyskiwaniu danych pierwotnych), które w ostatnich latach na polskim rynku pełniły szczególnie ważną rolę w niwelowaniu luki informacyjnej w zakresie rozpoznawania mechanizmów popytowych turystyki, obecnie zarówno w środowisku naukowym, jak i w biznesie zauważa się również podejmowanie licznych prób wykorzystania informacji z „chmury danych”, generowanej przez użytkowników internetu i urządzeń mobilnych. Potencjał i możliwości takie dają gwałtowny rozwój technologii komunikacyjnych (Information and Communications Technologies – ICT) oraz obliczeniowych. Dzięki szybkiemu rozwojowi technologii ICT, cyfryzacji, powszechnej dostępności urządzeń generujących i rejestrujących dane oraz ich połączeniu z globalną siecią internetową powstają coraz większe zasoby informacji stwarzające nieznane dotąd możliwości analityczne (Woźniczka 2018, Śledziwska, Włoch 2021). Potencjał tkwiący w analizie zbiorów Big Data (ang. BIG-Data Analysis – BDA), w tym również do analizy popytu turystycznego, został dostrzeżony stosunkowo

niedawno, ale tempo jego wykorzystania wyraźnie rośnie (Saluveer, Raun i in. 2020). Jednym z pierwszych i zasadniczych pól wykorzystania Big Data stał się marketing (także ten „turystyczny”), a czynnikiem determinującym taki stan rzeczy były szybkie, mierzalne finansowo efekty analityki danych.

W kontekście biznesowym i marketingowym wykorzystywana analiza Big Data zmierza do jej zasadniczego zastosowania, jakim jest dostarczanie rozległej wiedzy służącej m.in. dostarczaniu wartości nabywcom usług i produktów (rosnąca personalizacja, świadczenie usług i produkcja towarów coraz lepiej dostosowanych do potrzeb i oczekiwań indywidualnych odbiorców), wytwarzaniu treści (również marketingowych), uzyskiwaniu przewagi konkurencyjnej oraz pomiarowi efektów prowadzonych działań (por. Zhan i in. 2017).

Wydaje się jednak, że na równi z rosnącym przekonaniem o wysokiej wartości informacji tkwiących w tego typu zasobach w dyskusjach nad możliwościami często również wskazuje się na wady i problemy, które w szczególności wynikać mają ze stosowanych algorytmów obliczeniowych (za którymi stoją konkretni ludzie i firmy), a które z reguły muszą opierać się na uproszczonych założeniach, pasują do określonej tezy, a tym samym schematyzują świat (rzeczywistość), nie zwracając tym samym uwagi na nietypowe zachowania człowieka (O’Neil 2017). W dyskusji naukowej podkreśla się, że cechą Big Data, utrudniającą wykorzystanie, jest stale i w bardzo szybkim tempie powiększający się wolumen oraz złożoność, co sprawia, że trudne staje się ich przetworzenie za pomocą tradycyjnych narzędzi do zarządzania bazami danych lub aplikacji do przetwarzania.

Badania ruchu turystycznego w regionach – syntetyczna ocena tradycyjnych i najnowszych rozwiązań w świetle aktualnego stanu wiedzy naukowej

Ograniczone możliwości poznawcze w oparciu o istniejące dane (braki w statystyce publicznej), złożoność badanego zjawiska oraz jego szerokie uwarunkowania w zasadniczy sposób wpłynęły na kształt dotychczasowych metodologii regionalnych i lokalnych badań ruchu turystycznego, realizowanych w naszym kraju oraz w jego regionach. Ich krytyczna analiza wykazała bardzo dużą rozbieżność metodologiczną (co dotyczy zarówno ich skali, charakteru, zakresu, częstotliwości podejmowanych badań, różnic w zakresie stosowanych metod i technik badawczych, sposobów gromadzenia materiału, jak i jego analizowania) (por. Matczak 1992, Alejski 2010, Włodarczyk 2011, Dziedzic 2013).

W świetle dotychczasowych rozwiązań oraz wytworzonego poziomu wiedzy naukowej standardem jest, że badanie ruchu turystycznego na określonym obszarze (miejscowości, rejonie, regionie) wymaga zastosowania kompleksowej

metodologii badawczej, która musi uwzględniać wiele metod oraz wykorzystanie danych pochodzących z licznych i uzupełniających się źródeł. Samo badanie wymaga doświadczenia eksperckiego i wiedzy o specyfice ruchu turystycznego na danym obszarze (Alejziak 2010). Warto podkreślić, że zdecydowaną większość badań dotyczących turystyki stanowią badania częściowe, co wynika z przekonania o braku możliwości całościowego ujęcia statystycznego zjawiska. Największe znaczenie w ramach badań częściowych należy przypisać badaniom reprezentacyjnym, w których główną rolę odgrywają badania z wykorzystaniem wywiadów i ankiet. Takie podejście (ocena całości przez poznanie części zjawiska) określa się potocznie jako zastosowanie metody ankietowo-sondażowej (por. Brudnicki 2005).

Pomijając stwierdzone w różnych ocenach naukowych sytuacje braku poprawności metodologicznej, większość z prowadzonych w ostatnich latach w Polsce regionalnych badań ruchu turystycznego i konsumpcji turystycznej każdorazowo wносиła całkiem szerokie spectrum użytecznych informacji i danych (struktura ruchu turystycznego, cechy przestrzenne, sposób organizacji, cele przyjazdu, motywy wyboru regionu/miejscowości recepcyjnej jako miejsca destynacji turystycznej, sposoby spędzania czasu podczas pobytu, wydatki w czasie pobytu i podróży, sezonowość i długość pobytu, częstotliwość przyjazdów, opinie na temat regionu/miejscowości), dawała także podstawę do szacowania wielkości całkowitego strumienia ruchu turystycznego docierającego do regionu. Warto jednak podkreślić, że pomimo wielu wartościowych analiz (np. „fotografie” przestrzennej koncentracji ruchu turystycznego, charakterystyki preferencji turystów, segmentacja, zachowania), do tej pory w sposób pogłębiony nie udało się zbadać czasoprzestrzennych wzorców zachowań turystów (nie było możliwości wykonania analizy przemieszczeń turystów przez cały czas trwania ich wizyt w danym regionie).

Niezależnie od powyższych procedur pozyskania danych, w ostatnim okresie pojawiły się nowe „niekonwencjonalne analizy” i „źródła danych”, których tempo wykorzystania w badaniach nad turystyką zaczyna wyraźnie przyspieszać (wiele z tych nowych, dużych źródeł danych mierzy ludzką aktywność lub ruch). Według coraz liczniej pojawiających się głosów w środowisku badaczy turystyki w bliskiej przyszłości mogą one w niespotykanym dotąd wymiarze poprawić naszą wiedzę na jej temat.

W ostatnim dziesięcioleciu istota turystyki i zachowań człowieka związanych z podróżowaniem bez wątpienia zmieniła się wskutek szybkiego rozwoju technologii informacyjno-komunikacyjnych (ICT). Jak powszechnie wiadomo, każde działanie w „wirtualnym świecie” pozostawia w nim cyfrowy ślad. Aktywność użytkowników internetu i urządzeń mobilnych generuje coraz bogatszy strumień danych opisujących ich zachowania, zainteresowania, grupy społeczne,

w których się obracają oraz miejsca, które odwiedzają. Ślady te mogą mieć charakter aktywny (świadomie umieszczane w sieci przez użytkownika pliki multimedialne, recenzje, posty w mediach społecznościowych) lub pasywny (automatycznie generowane logi w punktach bazowych sieci komórkowej, informacje o dokonanych płatnościach, informacje zbierane automatycznie przez system operacyjny i zainstalowane aplikacje na smartfonach) (por. Majewska, Napierała i in. 2016). Dane te są zasileniem zbiorów danych stanowiących wejście do rozwijanych wielkoskalowych systemów analityki, w tym analityki internetowej. Z racji skali wolumenu gromadzonych i przetwarzanych danych mówi się o „przetwarzaniu” tzw. dużych zbiorów danych, czyli wskazanych powyżej (Big Data) (Śledziwska, Włoch 2021).

Analiza wpisów w internecie i w mediach społecznościowych, w tym m.in. analiza tagów, słów kluczowych i „sentymentu”, są wykorzystywane do oceny wartości marki dla nabywców i jej pozycji konkurencyjnej oraz do przewidywania zmian wartości marki dla akcjonariuszy (np. Nam, Kannan 2014). Analiza wejść na strony internetowe (*clickstream data*) pozwala prześledzić zakupowe i pozazakupowe zachowania nabywcy w internecie, określić wzorce ścieżek online, co ułatwia tzw. segmentację behawioralną i pozycjonowanie reklam (por. Kachniewska 2014). Szybki postęp w technologii informacyjno-komunikacyjnej otworzył przed badaczami i statystykami kilka nowych możliwości zrozumienia zachowań czasoprzestrzennych turystów. W szczególności różne zbiory danych śledzenia są szczególnie ważne dla badania mobilności turystycznej. Szczególnie ważne znaczenie w analizowaniu wzorców mobilności ludzi, w tym w kontekście prowadzenia coraz bardziej precyzyjnych i kompleksowych badań popytu turystycznego z użyciem analityki internetowej, mają dane geolokalizacyjne pozyskiwane z sieci telefonii komórkowych oraz mediów społecznościowych. Rozwój technologii mobilnej (smartfon + aplikacje), dostęp do geograficznych systemów informacyjnych (ang. Geographic Information System – GIS) oraz upowszechnienie się usług geolokalizacji (Majewska i in. 2016, s. 309-340) powodują tworzenie licznych „śladów użytkowników” w przestrzeni geograficznej (dane z urządzeń mobilnych i lokalizatorów GPS), które są analizowane m.in. w celu tzw. geotargetowania i docierania do odbiorców z komunikacją marketingową uwzględniającą kontekst miejsca i sytuacji, w których się znajdują, a także w celu oceny skuteczności i efektywności tego typu działań.

Analiza porównawcza metodologii badań monitorujących ruch turystyczny w regionie kujawsko-pomorskim w latach 2009–2015 oraz w badaniu przeprowadzonym z wykorzystaniem analizy danych typu BIG-DATA za rok 2019

Celem podstawowym badań monitorujących przyjazdy turystyczne do regionu kujawsko-pomorskiego w latach 2009, 2010, 2015 było przeprowadzenie badania ruchu turystycznego przyjazdowego (krajowego i zagranicznego) wraz z analizą wyników, wnioskami i określeniem tendencji w tym zakresie, metodą wywiadu bezpośredniego z użyciem kwestionariusza oraz uzupełniających badań typu desk research (w fazie przygotowawczej). Badaniami byli objęci odwiedzający w wieku od 16 lat. Definicja „odwiedzający” dotyczyła turystów i odwiedzających jednodniowych (beznoclegowych), zatem zgodnie z nomenklaturą UNWTO. Za turystów uważano odwiedzających (zarówno krajowych, jak i zagranicznych, a także mieszkańców regionu), którzy spędzili przynajmniej jedną noc w publicznych lub prywatnych obiektach noclegowych w odwiedzanym regionie. W grupie tej zakładano również identyfikację osób nocujących u krewnych lub znajomych oraz spędzających czas (z noclegiem) w „drugich domach” i „innych” obiektach. Istotne znaczenie miała również grupa odwiedzających jednodniowych, a więc osób, które nie nocowały w odwiedzanym regionie, ale przebywały w nim w jasno określonych „celach turystycznych”. W świetle tak sformułowanych definicji w obu przypadkach mogły to być osoby spoza Polski, mieszkańcy innych regionów kraju oraz mieszkańcy macierzystego województwa.

Z racji prowadzenia badań częściowych w celu podniesienia poziomu reprezentatywności wyników końcowych wprowadzano procedury strukturyzacji próby. Opierając się na analizie danych zastanych (wtórnych) i metodzie *desk research* z uwagi na obserwowaną w ostatnich latach, utrzymującą się relację proporcji pomiędzy grupą turystów krajowych i zagranicznych, wprowadzano proporcje w liczbie zakładanych do pobrania ankiet na poziomie 90/10. Jednocześnie ustalano proporcje w ogólnej liczbie pozyskiwanych ankiet pomiędzy kobietami a mężczyznami na poziomie: 55% kobiet i 45% mężczyzn. Badania przeprowadzono na pokaźnej próbie co najmniej 5000 osób w zakładanych okresach sezonu turystycznego. Narzędziem badania był kwestionariusz wywiadu, który w swej konstrukcji umożliwiał porównywanie wyników uzyskiwanych w poszczególnych edycjach badań (2009, 2010, 2015) (porównywalność wyników zapewniało również stosowane podejście – koncepcję metodologiczną badań starano się zachowywać od pierwszej edycji badań, jednocześnie w kolejnych wprowadzano rozwiązania podnoszące jakość uzyskiwanego materiału ankietowego – podsta-

wowego źródła danych). Ankieta zawierała każdorazowo około 20 pytań oraz metrykę respondenta. Dominującą formą diagnozowania były pytania zamknięte, część pytań miała formę otwartą, umożliwiającą respondentowi swobodną wypowiedź w zakresie danego zagadnienia badawczego (pytania diagnozowały region oraz poszczególne jego rejony, w których przeprowadzano pobór materiału ankietowego w badaniu bezpośrednim – *face to face*).

Miejsca poboru wywiadów kwestionariuszowych w każdej z dotychczasowych edycji badań identyfikowano wykorzystując wiedzę ekspercką (opinia ekspertów wskazywała istnienie w regionie ukształtowanej i względnie stałej struktury użytkowania przestrzeni turystycznej) oraz analizę *desk research*. Podstawową jednostką odniesienia, na poziomie której dokonywano identyfikacji miejsc poboru materiału ankietowego, był powiat. Typowanie konkretnych rejonów badań, wraz z określeniem liczby ankiet, uzyskiwano poprzez przeprowadzenie procedury oceny taksonomicznej wagi obszarów na mapie turystycznej województwa. Wykorzystywano w tym celu informacje pochodzące ze statystyki publicznej, m.in. o liczbie obiektów noclegowych, miejsc noclegowych, udzielonych noclegów, ponadto atrakcyjności turystyczno-kulturowej regionu (w tym brano pod uwagę: obiekty historyczne, budynki, budowle itd.), atrakcyjności kulturalnej (w tym brano pod uwagę: współczesne wydarzenia kulturalne, festiwale, wystawy, wydarzenia historyczne, imprezy sportowe etc.), atrakcyjności turystycznej bazującej na walorach naturalnych oraz faktycznie udokumentowanej atrakcyjności, mierzonej tzw. odwiedzalnością muzeów i innych atrakcji turystycznych.

Reasumując: analiza części metodologicznej i analitycznej raportów z badań prowadzonych w latach 2009, 2010, 2015 pozwala na wyciągnięcie wniosku, że badania te były prowadzone na podstawie naukowej, obiektywnej oceny rozwiązań stosowanych w praktyce monitorowania ruchu turystycznego (zarówno zagranicznych, jak i krajowych). Koncepcja metodologiczna badań była oparta na wybranych, przyjętych w środowiskach naukowych i praktyków procedurach i rozwiązaniach metodycznych, prowadzących do pozyskania odpowiedniego do realizacji celów badania materiału źródłowego z ankiet, realizowanych metodą wywiadu bezpośredniego (dobór próby, strukturyzacja próby, dobór miejsc poboru próby, stosowanie rozwiązań mających wpływ na podnoszenie jakości materiału źródłowego pozyskiwanego w badaniach częściowych, a także procedur podnoszących reprezentatywność wyników).

Analizując i oceniając wyniki dla regionu kujawsko-pomorskiego uzyskane w procedurze badawczej opartej na analizie danych typu Big Data (badania wykonano za rok 2019 – przedpandemiczny), w pierwszej kolejności konieczne jest odwołanie do metodologii, jaką posługuje się firma – wykonawca badania.

Podstawę badań i wykonanych analiz stanowiła baza danych (o zupełnie innej genezie i strukturze), pochodząca z tzw. hurtowni danych firmy. Baza ta jest własnością wykonawcy badania, a została utworzona w wyniku zapytań oraz wyświetleń reklam w kanale mobilnym, jak również od wydawców konkretnych aplikacji. Dane tego typu zbierane są w sposób pasywny i aktualizowane w czasie rzeczywistym. Z punktu widzenia dotychczas prowadzonych badań ruchu turystycznego w regionie istotne jest to, że wśród wielu danych zgromadzonych i przetwarzanych przez firmę znajduje się również informacja o lokalizacji smartfona w danym czasie, która daje możliwość selekcji osób do analizy z punktu widzenia ich aktywności realizowanej w czasie i przestrzeni (w tym na terenie regionu). Jest to możliwe w przypadku tych osób, których smartfony udostępniały takie dane oraz tych użytkowników, których dane ostatecznie trafiły do bazy danych zarządzanej przez podmiot. Odpowiednio przygotowane zapytanie do bazy danych (każda z charakterystyk, każdy z profili użytkowników smartfona są szacowane na podstawie modelu uczenia maszynowego) jest w stanie wyselekcjonować osoby spełniające warunek bycia w określonym miejscu i czasie. Wykonawca oferuje usługę opartą na tym mechanizmie, stosując zastrzeżoną nazwę własną „geotrapping” (znak towarowy słowny, zastrzeżony przez Urząd Unii Europejskiej ds. Własności Intelektualnej). Obecnie według informacji firmy wykonawcy badań, w bazie firmy znajduje się 24 mln unikalnych użytkowników smartfonów w Polsce. Do jednego unikalnego użytkownika – średnio – w bazie przypisane są 362 informacje (np. lokalizacja – koordynaty GPS, język systemu operacyjnego telefonu, aplikacje, jakie stosuje użytkownik, strony internetowe, które odwiedza, kod pocztowy). Baza danych wykonawcy jest zintegrowana z 21 globalnymi sieciami reklamowymi. Zbiera ona i analizuje dane pochodzące z ponad 200 000 aplikacji z reklamami oraz około 14 000 000 mobilnych stron www. Dzięki mnogości „tytułów” możliwe jest dotarcie do maksymalnej liczby użytkowników w jak najkrótszym czasie. Ekosystem wykonawcy odczytuje do 30 000 zapytań reklamowych na sekundę na rynku polskim. Ilość zapytań i wyświetleń reklam jest zależna od pory dnia oraz aktywności użytkowników smartfonów oraz tabletów.

Firma gromadzi dane dotyczące wyświetlania reklam dla unikalnych użytkowników, dzięki czemu – jak podaje – można m.in. określić miejsce zamieszkania, miejsce pracy, czy kierunki podróży danego użytkownika. Dane mobilne pozwalają na oszacowanie grupy wiekowej, płci, miejsca zamieszkania, czy liczby dni pobytu użytkownika odwiedzającego określone miejsce. W tym celu firma posługuje się informacjami o języku telefonu, koordynatach GPS, w jakich rejestrowane jest urządzenie oraz aplikacjach, z jakich korzysta użytkownik telefonu i stronach mobilnych, jakie odwiedza.

W opisie metodologii „Badania ruchu turystycznego w województwie kujawsko-pomorskim za rok 2019 r.” wykonawca wskazuje jednoznacznie, że źródłem danych wykorzystanych do analizy są informacje pozyskane w wyniku zapytań oraz wyświetleń reklam w kanale mobilnym, jak również od wydawców konkretnych aplikacji, co wskutek przyjętych algorytmów obliczeniowych pozwoliło na: dokonanie oszacowania m.in. liczebności turystów i odwiedzających jednodniowych (na podstawie liczby użytkowników telefonów komórkowych zarejestrowanych w koordynatach GPS województwa kujawsko-pomorskiego), wyodrębnienie grupy turystów krajowych i zagranicznych (na podstawie kraju, w którym najczęściej rejestrowane jest urządzenie mobilne odwiedzających województwo kujawsko-pomorskie), wskazanie regionów pochodzenia turystów krajowych (na podstawie koordynatów GPS miejsca, w którym przez 5 z 7 dni użytkownik przebywał między godziną 23:00 a 7:00), określenie płci osób (odwiedzających) (na podstawie klasyfikatora, który estymuje płeć użytkownika telefonu komórkowego na podstawie informacji o używanych przez niego aplikacjach), identyfikację długości pobytu (1 dzień bez noclegu, 1 nocleg, 2-3 noclegi, 4-7 noclegów, 8-14 noclegów, 15-21 noclegów oraz 21-90 noclegów – poprzez zastosowanie klasyfikatora estymującego rzeczywisty czas pobytu).

Jednocześnie firma udostępniła stosowane definicje („słownik pojęć”), których zakres należy uznać za istotny również z punktu widzenia oceny wyników badania. Pojęcia zawarte w „słowniku pojęć” określały m.in. sposób wyodrębniania odwiedzających (w tym turystów zagranicznych, krajowych, krajowych wewnątrzregionalnych), odwiedzających jednodniowych (bez noclegu), w tym: zagranicznych, krajowych, wewnątrzregionalnych. Znamiennym jest, że identyfikatorem „turysty” była osoba posiadająca smartfon („użytkownik smartfona”). W charakterystyce uzyskanych wyników posługiwano się również określeniem „unikalny turysta i odwiedzający jednodniowy”.

W celu porównania i oceny zastosowanych procedur badawczych (identyfikacji różnic i ich możliwego wpływu na uzyskane wyniki) posłużono się kryteriami przyjętymi w badaniach przez J. Ryśnika, D. Żylak, P. Gibasa (2019). Wyniki oceny przedstawiono w poniższym zestawieniu tabelarycznym (tab. 1).

Tab. 1. Porównanie procedur

Postępowanie badawcze stosowane w badaniach ruchu turystycznego w regionie kujawsko-pomorskim w latach 2009, 2010, 2015	Postępowanie badawcze – firma wykonująca badania (wyniki uzyskane za rok 2019)
Postępowanie badawcze – etap projektowania badania	
Na etapie projektowania badań priorytetowym założeniem (co ważne – każdorazowo wypełnianym) było dążenie do porównywalności wyników pomiędzy poszczególnymi edycjami badań (2009, 2010, 2015). Podejście wymagało jasno sprecyzowanych celów badań i ich utrzymania w kolejnych edycjach, ale przede wszystkim stosowania bardzo zbliżonej metodyki pozyskiwania materiału źródłowego. Podmiot badań był jasno określony – do identyfikacji ruchu turystycznego (odwiedzających, turystów i odwiedzających jednodniowych stosowano terminologię UNWTO). Badanie pozwalało na uzyskanie danych o charakterze ilościowym i jakościowym. Zespół badawczy każdorazowo posiadał zdecydowanie większą swobodę w formułowaniu celów badawczych (np. w ich uszczegółowianiu), przebiegu badań, doborze metod (w szczególności dbając o podwyższenie jakości wyników) oraz w analizie wyników. Kwestionariusz ankiety (wywiadu) stanowił potężne narzędzie badawcze (np. można było każdorazowo rozszerzyć zakres o nowe dane empiryczne). Ograniczeniem zespołu od strony organizacyjnej były niewątpliwie środki finansowe (były to badanie niewyczerpujące, ale jednak prowadzono je na stosunkowo dużych próbach oraz na dużym obszarze, co generowało znaczne koszty). Uwagi szczególne: • Terminologia – podstawy teoretyczne: definicje zjawisk przyjęte w badaniach były ogólnie uznane, a zatem nie stanowiły podstaw do krytyki. Przyjęto zapisy „Metodologii Unii Europejskiej w dziedzinie statystyki turystyki” (1998). • <u>Cele badań:</u> celem głównym badań nie było szacowanie wielkości strumienia ruchu turystycznego, lecz przede wszystkim poznanie struktury ruchu, a także szeregu parametrów opisowych, istotnych z punktu widzenia rozpoznania jego właściwości. Badano również zachowania turystyczne, każdorazowo ważnym celem była ocena satysfakcji i atrakcyjności turystycznej. • <u>Zakres poznawczy badań:</u> w kontakcie bezpośrednim z turystą pozyskiwano dane pierwotne o zdecydowanie szerszym zakresie poznawczym w relacji do pozyskanych w 2019 r., tj.: profil społeczno-demograficzny i ekonomiczny	Badanie posiada zupełnie odmienne założenia i cele. Wykorzystano inny, nowoczesny, ale także dotychczas na szerszą skalę niedostępny dla badaczy sposób pozyskiwania informacji o popycie turystycznym. Celem badań w głównej mierze było oszacowanie wielkości strumienia ruchu turystycznego, który dotarł do regionu kujawsko-pomorskiego w 2019 r. (na podstawie danych historycznych zapisanych w bazie typu Big Data). Cel ten został osiągnięty (jego poprawność trudno jednak obiektywnie ocenić). Wykonawca pozyskał i wstępnie opracował dane pochodzące z tzw. „śladów” używania urządzeń mobilnych, tj. informacje charakteryzujące użytkowników smartfonów oraz tabletów w określonych lokalizacjach, a na ich podstawie przedstawił oszacowania wielkości ruchu turystycznego, informacje o koncentracji przyjazdów w wybranych lokalizacjach oraz zbudował profile turystów. W raporcie scharakteryzowano wybrane parametry ruchu turystycznego (długość pobytu, częstotliwość wizyt, pochodzenie terytorialne, wyniki dla wybranych wydarzeń itp.). W przypadku zastosowania identycznego zestawu zapytań do bazy danych – podczas ewentualnych przyszłych badań – badania niewątpliwie będą posiadać wysoki poziom porównywalności. Warto jednak doprecyzować, że będzie to możliwe tylko wtedy, kiedy badania prowadzić będzie ta sama firma (istotne jest wykorzystanie tej samej bazy danych). Przy założeniu, że badanie popytu turystycznego powinno odbywać się systematycznie, opierać się na powtarzalności stosowanych metod i procedur oraz co wydaje się najważniejsze – powinno pokrywać jak największą populację turystów – taka sytuacja dla regionu oznacza konieczność związania się z danym usługodawcą (firmą, właścicielem bazy danych typu Big Data). Uwagi szczególne: • Terminologia – podstawy teoretyczne: zastosowana terminologia wśród badaczy turystyki może budzić wątpliwości i pewien niepokój (np. wg przyjętych założeń można wnioskować, że „turystami” są tylko użytkownicy smartfonów). • Wyniki nie obejmują analizy celów przyjazdów turystów (!?). • Należy podkreślić, że pozyskane dane w zasadzie nie mogą podlegać innym analizom (chodzi

turystów i odwiedzających jednodniowych (w tym: płeć, wiek, wykształcenie, sytuacja materialna, status zawodowy, kraj stałego zamieszkania), powtarzalność i częstotliwość wizyt, cele przyjazdów, długość pobytów, wydatkowane kwoty, sposoby organizacji przyjazdów, preferencje w wyborze bazy noclegowej, osoby towarzyszące w podróży, wykorzystywane środki transportu, wykorzystywane źródła informacji nt. regionu itp.), pochodzenie terytorialne, opinie nt. elementów wpływających na postrzeganie atrakcyjności turystycznej regionu i poszczególnych regionów (analizowano bardzo szczegółowo wybrane cechy niematerialne, np.: atmosferę miejscowości, gościnność, bezpieczeństwo, jak i materialne: np. oznakowanie turystyczne, transport lokalny itp.), jak również relacje jakości świadczonych usług do obowiązujących cen, pozyskiwano oceny atrakcyjności turystycznej regionu i badanych rejonów z uwzględnieniem słabych i mocnych stron rozwoju turystyki oraz deklaracji chęci powrotu (ponownego odwiedzenia regionu i badanego rejonu). Należy podkreślić, że pozyskane dane mogły podlegać różnorodnym analizom (możliwość prezentacji wyników w różnych konfiguracjach, z użyciem wielu zadanych kryteriów, z odniesieniem do różnych jednostek przestrzennych (np. w analizie pochodzenia terytorialnego odwiedzających).	o możliwość prezentacji wyników w różnych konfiguracjach, z użyciem wielu zadanych kryteriów, z odniesieniem do różnych jednostek przestrzennych, np. w analizie pochodzenia terytorialnego odwiedzających) Zakres danych – zapewne rozszerzenie zakresu danych byłoby możliwe, jednakże w sposób znaczący wpłynęłoby na ostateczne koszty badania.
Powtarzalność badań	
W przypadku ponowienia procedury badawczej w nowej edycji badań niestety nie ma możliwości ponownego dotarcia do tych samych respondentów, a zatem ponownego wykonania takich samych (o takim samym zakresie) lub innych badań. Osoba badana jest anonimowa, a możliwość badania jednorazowa. Generalnie jasno trzeba stwierdzić, że w tym przypadku nie ma żadnej możliwości dokładnego powtórzenia sytuacji badawczej. Istnieje natomiast możliwość porównywania wyników, jeśli są one uzyskane tą samą metodą w innej grupie badanych w innej/podobnej sytuacji badawczej (np. w trakcie innego sezonu turystycznego).	Badanie z wykorzystaniem analizy Big Data ma zasadniczą przewagę nad sposobem dotychczas wykorzystywanym w badaniach ruchu turystycznego regionu kujawsko-pomorskiego. W tym przypadku przewaga polega na tym, że praktycznie zawsze jest możliwe ponowne dotarcie do tych samych danych (a dokładnie mówiąc – do tych osób), np. w celu ponownego wykonania takiej samej lub innej analizy badawczej. Przypomina to metodę badań panelowych (jest to typ badania dynamicznego, polegającego na zbieraniu danych od tej samej grupy ludzi – próby, panelu – w kilku punktach czasowych). Dane wykorzystane przez wykonawcę są oznaczone unikalnym identyfikatorem treści reklamowych, stanowiącym jednocześnie unikalny klucz identyfikujący daną osobę lub urządzenie. Baza wykonawcy w sposób ciągły podlega rozszerzaniu (jest uzupełniana o nowe informacje, tj. „ślady” o jej użytkownikach). Daje to możliwość wykonania tej samej analizy na tych samych osobach, ale na uzupełnionym w czasie zbiorze (dzięki temu można przeprowadzić analizę zmian ich profilów w czasie).

	Sytuacja ta wpływa na wysoką ocenę potencjału naukowego i zarazem praktycznego tego typu analiz. Potencjał ten jest już dostrzegany, także w badaniach popytu turystycznego (ocena na podstawie różnych case studies z rynków zagranicznych).
Kwalifikacje zespołu badawczego	
Każdorazowo zespoły badawcze musiały wykazać się posiadanym doświadczeniem eksperckim (kierownik badań, analitycy, zespół ankierski).	Brak danych o predyspozycjach naukowych zespołu wykonującego badania (co dotyczy wiedzy naukowej nt. podstaw teoretycznych i praktyki badań ruchu turystycznego, wiedzy nt. uwarunkowań rozwoju turystyki w regionie, znajomości tendencji i czynników determinujących popyt turystyczny itp.). Opracowanie „Raportu...” (w zasadzie zestawienia danych bez ich interpretacji) wymagało wyłącznie zastosowania odpowiedniego algorytmu obliczeniowego (odpowiedniego zadania pytań do bazy) oraz interpretacji wyników.
Dostęp do danych źródłowych	
Każdy członek zespołu badawczego dysponował dostępem do bazy pierwotnych danych, co wpływało na możliwości ich eksplorowania. Zamawiający otrzymywał pełne prawo do wglądu do bazy danych oraz stawał się właścicielem bazy danych. Objętość i lokalizacja bazy danych nie powodowała znaczących problemów w przypadku wykonywania analiz (dane nie były rozproszone, a do ich obróbki wystarczało ogólnodostępne oprogramowanie, powszechnie stosowane w badaniach naukowych).	Region nie posiada w tym przypadku dostępu do danych źródłowych. Właścicielem bazy danych jest podmiot wykonujący badania (zmuszony do stosowania zasad RODO).
Organizacja badań – koszty badań	
200-300 tys. zł za badanie – dane szacunkowe	Dane poufne

Źródło: opracowanie własne.

Podsumowanie

Dostępność zbiorów Big Data, również na rynku turystycznym, ze względu na stosunkowo łatwe możliwości pozyskania odpowiednich danych, będzie się zapewne szybko zwiększać. Rodzi się zatem pytanie: czy pozyskiwanie danych naukowych tą drogą może skutecznie konkurować z klasycznymi, realizowanymi głównie z użyciem metod ankietowo-sondażowych? Pomimo pewnych wątpliwości metodologicznych (aby je rozwiązać, niezbędny jest dialog z zespołami, które w podmiotach komercyjnych oferujących usługi analiz danych Big Data odpowiadają za generowanie wyników) oraz zidentyfikowaniu luk w wynikach (co w analizowanym konkretnym przypadku może wynikać z przyjętej specyfikacji zamówienia oraz listy oczekiwanych wyników) można przyjąć, że analizy danych z tzw. „chmur danych” w zakresie popytu turystycznego powinny stać się stałym

elementem w badaniach ruchu turystycznego w regionach (poszerzenie katalogu metod stosowanych w badaniach konsumentów usług turystycznych w regionie powinno się jednak wiązać z przyjęciem pewnych założeń). W obecnych realiach rynkowych analiza danych Big Data powinna mieć przede wszystkim charakter uzupełniający, a zatem nie powinno się odchodzić od badań klasycznych (ponieważ nadal przynoszą one wiele wartości poznawczych). Stosowanie rozwiązania polegającego na wykorzystaniu obu procedur (klasycznej – z wykorzystaniem wywiadów kwestionariuszowych oraz opartej na analizie danych z „chmur danych”) wydaje się być na tę chwilę najkorzystniejsze. W efekcie wyniki uzyskiwane w taki sposób mogłyby pokrywać znacznie większy niż dotychczas fragment pola badawczego i jednocześnie poszerzać efekty poznawcze – co może przynieść zdecydowanie większe korzyści w przypadku zastosowania praktycznego tej wiedzy). Ponieważ w badaniach opartych na Big Data istnieje możliwość docierania z badaniem do tych samych konsumentów, w szczególności korzystnym rozwiązaniem byłoby wykorzystanie Big Data do badania turystów *post factum*, z wykorzystaniem techniki ankietowej (dodatkowe badanie z wykorzystaniem techniki ankietowej, z możliwością pozyskania danych jakościowych).

Bibliografia

- Alejziak W., 2010, *Aktualny stan badań konsumentów usług turystycznych w regionach*, [w:] E. Dziedzic (red.), *Regionalne badania ruchu turystycznego i konsumpcji usług turystycznych w Polsce*, POT, Warszawa [online] <https://www.pot.gov.pl> [data dostępu: 15.05.2022].
- Brudnicki R., 2005, *Kierunki i metody badań ruchu turystycznego*, Seria „Turystyka i Rekreacja”, nr 5, WSG, Bydgoszcz.
- Dziedzic T., 2013, *Funkcje i użyteczność pomiarów ruchu turystycznego*, „Problemy Turystyki i Rekreacji”, 2013 (4), SGTIR, Warszawa.
- Hall C.M., Page S., 2014, *The geography of tourism and recreation: Environment, place and space* (4th ed.), Routledge, New York [online] <https://www.taylorfrancis.com> [data dostępu: 15.05.2022].
- Kachniewska M., 2014, *Big Data Analysis jako źródło przewagi konkurencyjnej przedsiębiorstw i regionów turystycznych* [online] <https://cor.sgh.waw.pl> [data dostępu: 15.05.2022].

Majewska J., Napierała T., Adamiak M., 2016, *Wykorzystanie nowych technologii i informacji do opisu przestrzeni turystycznej*, „Folia Turistica”, nr 41, s. 309-340, Kraków [online] <http://yadda.icm.edu.pl> [data dostępu: 15.05.2022].

Matczak A., 1992, *Model badania ruchu turystycznego. Studium metodologiczne*, Acta Universitatis Lodziensis, Wyd. UŁ, Łódź.

Nam H., Kannan P.K., 2014, *Informational Value of Social Tagging Networks*, „Journal of Marketing”, 78 (4) [online] <https://www.researchgate.net> [data dostępu: 15.05.2022].

O’Neil C., 2017, *Broń matematycznej zagłady – jak algorytmy zwiększają nierówności i zagrażają demokracji*, Wydawnictwo Naukowe PWN, Warszawa.

Ryśnik J., Żylak D., Gibas P., 2019, *Tradycyjne ankietowanie vs. wykorzystanie Big Data danych geolokalizacyjnych – analiza porównawcza na przykładzie badań turystów-kibiców sportowych uczestniczących w ME U21 2017 w Tychach*, „Folia Turistica”, nr 53, Kraków [online] <http://www.folia-turistica.pl> [data dostępu: 15.05.2022].

Saluveer E., Raun J. i in., 2020, *Methodological framework for producing national tourism statistics from mobile positioning data*, „Annals of Tourism Research”, Vol. 81, March 2020, 102895, [online] <https://www.sciencedirect.com> [data dostępu: 15.05.2022].

Śledziwska K., Włoch R., 2021, *Gospodarka cyfrowa – jak technologie zmieniają świat*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa [online] <https://www.delab.uw.edu.pl> [data dostępu: 15.05.2022].

Włodarczyk B. (red.), 2011, *Ruch turystyczny w Łodzi i województwie łódzkim w roku 2011*, Wyd. ROTWŁ i IGMiTUŁ, Łódź.

Woźniczka J., 2018, *Big Data i ich wykorzystanie w analityce marketingowej. Wybrane problemy badawcze*, „Marketing i Rynek”, 3/2018 [online] <https://www.researchgate.net> [data dostępu: 15.05.2022].

Zhan Y., Hua Tan K., Ji G., Chung L., Tseng M., 2021, *A Big Data Framework for Facilitating Product Innovation Processes*, „Business Process Management Journal”, 23 (3) [online] <https://sm.xmu.edu.cn> [data dostępu: 15.05.2021].

BIG-Data Analysis – is a new, better method of researching tourism service consumers in regions? Comparative analysis on the example of tourism research conducted in the Kujawsko-Pomorskie region?

Abstract: Collecting tourism statistics is “necessary to measure the size, scale, impact and value of tourism at various geographic scales – from global to national to local destination” (Hall and Page 2014). BIG-Data sets in research on tourism demand, which, along with the growing computerization and the use of modern ICT (Information and Communications Technologies – ICT) technologies, in the field of empirical research on tourism seem to be one of the most current issues. In the article, using the example of tourism research conducted in the Kujawsko-Pomorskie region, the authors compared standard and BIG-Data-based research procedures and assessed the usefulness of this type of data. An attempt was made to answer the following question – can the scientific data obtained from the BIG-Data collections be able to effectively compete with the data on tourism demand obtained in classic research (conducted mainly with the use of survey and survey methods)? Despite some methodological doubts and the identification of gaps in the compared results, it can be assumed that the analysis of data on tourism demand from the so-called “data clouds” should become a permanent feature of regional tourism research. However, the extension of the catalog of methods and data used in research on tourist services consumers to include BIG-Data data analysis should be associated with the adoption of certain assumptions. In the current market realities, the analysis of BIG-Data should primarily be complementary, and therefore one should not depart from classical research (because they still bring a lot of cognitive value). As a result, the results obtained in this way could cover a much larger fragment of the research field than before and at the same time broaden the cognitive effects – which may bring much greater benefits when this knowledge is applied in practice.

Key words: tourism statistics, tourist traffic research, BIG-Data, tourist marketing, tourist demand research, tourist behavior, tourist traffic monitoring.