

DOI: <https://doi.org/10.61016/TZAP-2956-8048-19>

Nowe źródła danych w turystyce: badania nad opiniami z serwisów społecznościowych w ujęciu administracyjnoprawnym, ekonomicznym i społecznym¹

New data sources in tourism: research on opinions from social media platforms in administrative-legal, economic, and social perspectives

Mariusz Barczak, dr²
Wyższa Szkoła Gospodarki

Dominik Borek, dr³
Szkoła Główna Handlowa

Michał Pawlak, inż. ⁴
Politechnika Bydgoska im. Jana i Jędrzeja Śniadeckich

Słowa kluczowe: big data, web scraping, turystyka
Key words: big data, web scraping, tourism

JEL K21, K23

Abstract

The article analyzes new data sources in tourism, concentrating on opinions from social networking sites, examining them from administrative-legal, economic, and social perspectives. Growing big data sets have evolved from the classic 3V model (volume, variety, velocity) to expanded concepts, emphasizing the importance of data reliability (veracity, validity) and value. Although big data analysis yields benefits, it simultaneously poses serious threats to privacy and the risk of re-identification of individuals, necessitating cohesive and ethical regulations. In the legal context, the GDPR (RODO) is the paramount EU act concerning data protection. In Poland, the basis is the Act on Public Statistics, permitting the processing of

personal data for the statistical purpose of monitoring tourism. At the EU level, Regulation (WE) No 223/2009 allows national statistical offices to obtain absolutely necessary data free of charge from private data holders if its use significantly reduces the burden on businesses. Furthermore, Regulation (UE) 2024/1028 (effective May 2026) imposes an obligation on short-term rental platform providers to transmit data concerning their activity monthly. Web scraping may prove to be the key method for acquiring big data for public statistics. This technique automates the collection of information from websites, accelerates data gathering, and allows for near real-time analysis. Scraping is a valuable resource supplementing official statistics, allowing for the assessment of the quality of the offering and the forecasting of tourist behavior. The empirical part of the article utilized web scraping to download 4,480 reviews from Google Maps for the Wawel Royal Castle. The experiment confirmed that extending the waiting time for a visit significantly lowers the rating provided. These results emphasize the crucial importance of efficient management of tourist traffic for visitor satisfaction. Future research should aim to integrate public data with scraped data, which has been termed the “Tourism Demand Horizon”.

Abstrakt

Artykuł analizuje nowe źródła danych w turystyce, koncentrując się na opiniach z serwisów społecznościowych, w ujęciu administracyjno-prawnym, ekonomicznym i społecznym. Rosnące zbiory big data ewoluowały od modelu 3V do rozbudowanych ujęć, podkreślających znaczenie wiarygodności i wartości danych. Choć analiza big data przynosi korzyści, rodzi jednocześnie poważne zagrożenia dla prywatności i ryzyko ponownej identyfikacji osób, co wymaga spójnych i etycznych regulacji. W kontekście prawnym, RODO jest nadrzędnym aktem UE w zakresie ochrony danych. W Polsce podstawę stanowi Ustawa o statystyce publicznej, zezwalająca na przetwarzanie danych osobowych do monitorowania turystyki. Na poziomie unijnym, Rozporządzenie (WE) nr 223/2009 pozwala krajowym urzędом statystycznym na bezpłatne pozyskanie absolutnie niezbędnych danych od prywatnych posiadaczy, jeśli zmniejsza to obciążenie przedsiębiorstw. Ponadto, Rozporządzenie (UE) 2024/1028 (obowiązujące od maja 2026 r.) nakłada na dostawców platform najmu krótkoterminowego obowiązek comiesięcznego przekazywania danych dotyczących działalności. Kluczową

¹ Artykuł prezentuje osobiste poglądy i nie może być interpretowany jako stanowisko jakiegokolwiek organu, podmiotu czy instytucji. Artykuł prezentuje stan prawny na dzień: 1 września 2025 r.

² mariusz.barczak@gmail.com, <https://orcid.org/0000-0001-5270-6222>

³ dpborek@wp.pl, <https://orcid.org/0000-0002-4359-9426>

⁴ michal.bartek.pawlak@gmail.com, <https://orcid.org/0009-0005-0371-5380>

metodą pozyskiwania big data dla statystyki publicznej może okazać się web scraping. Technika ta automatyzuje gromadzenie informacji z witryn, przyspiesza zbieranie danych i umożliwia analizy w czasie niemal rzeczywistym. Scraping jest cennym zasobem uzupełniającym oficjalne statystyki, pozwalającym na ocenę jakości oferty i prognozowanie zachowań turystów.

Część empiryczna artykułu wykorzystała web scraping do pobrania 4480 opinii z Google Maps dla Zamku Królewskiego na Wawelu. Eksperyment potwierdził, że wydłużanie czasu oczekiwania na wizytę znacząco obniża wystawioną ocenę. Wyniki te podkreślają kluczowe znaczenie sprawnego zarządzania ruchem turystycznym dla satysfakcji odwiedzających. Przyszłe badania powinny dążyć do integracji danych publicznych ze scrapowanymi, co nazwano „Tourism Demand Horizon”.

Wstęp

W gospodarce cyfrowej ilość i jakość tworzonej informacji stale rośnie, tworząc zbiory danych określane mianem big data, a ich analiza i przetwarzanie są cenne zarówno z perspektywy nauki jak i gospodarki, ponieważ umożliwiają pozyskanie nowej wiedzy (Kachniewska, 2014). Analiza literatury wskazuje, że pojęcie big data ewoluowało od klasycznego modelu 3V (volume, variety, velocity) do bardziej rozbudowanych ujęć, obejmujących 5V, 7V, a nawet 10V. Wspólnym elementem wszystkich definicji jest ogromna ilość danych generowanych codziennie, ich różnorodność oraz tempo przetwarzania, które umożliwiają prowadzenie analiz w czasie rzeczywistym. Rozszerzone koncepcje podkreślają dodatkowo znaczenie jakości i wiarygodności danych (veracity, validity), ich wartości dla użytkowników i organizacji (value), a także zmienności (variability) oraz kontekstu geograficznego i sektorowego (venue, vocabulary). Zwraca się również uwagę na problem niejednoznaczności danych (vagueness), który stanowi istotne wyzwanie w procesie interpretacji (Barczak 2025). Jak zauważają Cierpiat-Wolan i Stateva (2023) nowe technologie tworzenia i przetwarzania dużych zbiorów danych pozwalają na analizę zjawisk społeczno-gospodarczych w trybie rzeczywistym, co znacząco podnosi jakość prognoz. W konsekwencji big data nie można definiować wyłącznie poprzez wielkość zbiorów, lecz raczej jako złożony ekosystem cech i wyzwań, których uwzględnienie jest kluczowe dla efektywnego wykorzystania danych w analizach społeczno-gospodarczych, w tym w sektorze turystyki.

Źródła danych, dostępne w gospodarce cyfrowej, obejmują zarówno dane świadomie udostępniane przez użytkowników, jak i informacje zbierane na podstawie ich aktywności online, takie jak preferencje czy wybory zakupowe, a także dane pochodzące z danych mobilnych czy kart płatniczych. Z jednej strony ich analiza przynosi liczne korzyści społeczne i gospodarcze, z drugiej rodzi poważne zagrożenia dla prywatności, ponieważ dane te często mają charakter osobowy lub wrażliwy (Stępień, 2018). Nie ulega wątpliwości, że wraz z pojawianiem się kolejnych możliwości technologicznych należy zadbać o tworzenie spójnych regulacji zapewniających odpowiedzialne i etyczne wykorzystanie danych. Łączenie informacji z różnych baz danych umożliwia odkrywanie nowych treści niewidocznych w pierwotnych zbiorach. W przypadku danych zanonimizowanych może to prowadzić do ponownej identyfikacji osób. Wykazano bowiem, że znajomość jedynie trzech cech – kodu pocztowego, daty urodzenia i płci – pozwala na identyfikację 87% populacji USA (Tene, 2011). Tego rodzaju ryzyko ponownej identyfikacji danych osobowych powinno być zawsze impulsem do wprowadzenia regulacji prawnych. Do kluczowych regulacji prawnych zalicza się Rozporządzenie Ogólne o Ochronie Danych Osobowych (RODO), które stanowi nadrzędny akt prawny Unii Europejskiej w zakresie ochrony danych osobowych. Równolegle obowiązują również międzynarodowe kodeksy postępowania, takie jak kodeks Europejskiego Stowarzyszenia Badaczy Opinii Publicznej i Rynkowych (ESOMAR), określający standardy etyczne i metodologiczne prowadzenia badań, a także kodeks Stowarzyszenia Badań Marketingowych (Marketing Research Society, MRS), funkcjonujący w praktyce badawczej na rynku brytyjskim (Woźniczka 2018).

Celem niniejszego artykułu jest analiza potencjału wykorzystania danych pochodzących z serwisów społecznościowych i platform cyfrowych w badaniach turystycznych, ze szczególnym uwzględnieniem aspektów administracyjno-prawnych, ekonomicznych i społecznych. Badanie koncentruje się na ocenie możliwości zastosowania *big data* w statystyce publicznej oraz na identyfikacji barier etycznych i prawnych związanych z ich przetwarzaniem. Przyjęto hipotezę, że dane pozyskiwane z opinii użytkowników w serwisach społecznościowych stanowią wartościowe uzupełnienie oficjalnych statystyk turystycznych, a ich analiza pozwala na identyfikację czynników wpływających na satysfakcję odwiedzających. W części empirycznej artykułu wykorzystano technikę web scrapingu do pozyskania 4480 opinii użytkowników z serwisu Google Maps dotyczących Zamku Królewskiego na

Wawelu. Materiał źródłowy obejmuje zatem zarówno dane pierwotne pozyskane techniką web scrapingu, jak i wtórne – pochodzące z literatury naukowej, aktów prawnych i opracowań statystycznych.

Regulacje prawne w zakresie dużych zbiorów danych w turystyce

Zasadnicza podstawa prawna w regulacjach krajowych wynika z art. 35b. ust. 1 Ustawy z dnia 29 czerwca 1995 r. o statystyce publicznej, który to przepis wskazuje iż, służby statystyki publicznej przetwarzają w celu statystycznym określone dane osobowe. Wyszczególnienie tych danych zostało wskazane jako enumeratywne (numerus clausus). Dane osobowe, o których mowa powyżej są przetwarzane przez służby statystyki publicznej w ramach badań statystycznych, które dostarczają informacji o życiu i sytuacji osób fizycznych lub wybranych aspektach życia i sytuacji tych osób w obszarze m.in. uczestnictwa ludności w sporcie, rekreacji, turystyce oraz kulturze, w tym wypoczynku dzieci i młodzieży. Dodatkowo podstawą jest uchwalany rokrocznie przez Radę Ministrów program badań statystycznych statystyki publicznej, którego przyjęcie poprzedza, a najpóźniej następuje równoległe z uchwaleniem projektu ustawy budżetowej. Badania wynikające z programu badań statystycznych statystyki publicznej są niezbędne do monitorowania i prognozowania stanu sektora turystycznego w Polsce. Są one wykorzystywane przez administrację rządową i samorządową do oceny rynku usług turystycznych oraz tworzenia strategii turystycznej. Wyniki realizowanych badań są prezentowane w postaci komunikatów, informacji, raportów oraz są upowszechniane na stronach internetowych www.turystyka.gov.pl, a także www.turystyka.stat.gov.pl. Pierwszy ze wskazanych portali to System Rejestrów Publicznych w Turystyce (SRWT), gdzie publikowane są również rejestry państwowe, które umożliwiają poszerzenie zbiorów big data. Duże zbiory danych to bowiem nie tylko statystyka czy wykorzystanie nowoczesnych technologii, ale także rejestry państwowe. One umożliwiają zbieranie adekwatnych informacji i prowadzenie statystyki gospodarczej dotyczącej liczby przedsiębiorstw, ich zabezpieczeń finansowych czy adresów siedzib. W ramach portalu funkcjonuje:

- 1) Rejestr Centralny Wykaz Obiektów Hotelarskich – zamieszczany na podstawie przekazywanych w formie elektronicznej przez marszałków województw danych dotyczących skategoryzowanych w województwie obiektów hotelarskich

- 2) Rejestr Centralny Wykaz Organizatorów Szkoleń dla Kandydatów na Przewodników Górskich – zamieszczany na podstawie informacji o dokonanych wpisach do rejestru organizatorów szkoleń przekazywanych w postaci elektronicznej przez marszałków województw,
- 3) Rejestr Centralny Wykaz Przewodników Górskich – zamieszczany na podstawie danych o nadanych uprawnieniach przewodnika górskiego przekazywanych w postaci elektronicznej przez marszałków województw.
- 4) Centralna Ewidencja Organizatorów Turystyki i Przedsiębiorców Ułatwiających Nabywanie Powiązanych Usług Turystycznych – zamieszczany na podstawie informacji o dokonanych wpisach do rejestru przekazywanych w postaci elektronicznej przez marszałków województw, ewidencja prowadzona jest przez Ubezpieczeniowy Fundusz Gwarancyjny

W powyższym zakresie podstawę prawną stanowi:

- 1) Ustawa z dnia 29 sierpnia 1997 r. o usługach hotelarskich oraz usługach pilotów wycieczek i przewodników turystycznych;
- 2) Ustawa z dnia 24 listopada 2017 r. o imprezach turystycznych i powiązanych usługach turystycznych.

Drugi ze wskazanych portali to tzw. Turystyka plus, opublikowany na portalu turystyka.stat.gov.pl. Narzędzie to stanowi innowacyjny instrument analityczny opracowany przez Główny Urząd Statystyczny we współpracy z Ministerstwem Sportu i Turystyki. Portal ten pozwala na analizę danych statystycznych w różnych ujęciach, w tym prognozowanie zmian na bazie modeli predykcyjnych.

Pozostawiając na boku rozważania w zakresie krajowego porządku prawnego, warto zwrócić uwagę na regulacje występujące na poziomie unijnym. Zasadniczą regulacją jest rozporządzenie Parlamentu Europejskiego i Rady (WE) nr 223/2009 z dnia 11 marca 2009 r. w sprawie statystyki europejskiej oraz uchylające rozporządzenie Parlamentu Europejskiego i Rady (WE, Euratom) nr 1101/2008 w sprawie przekazywania do Urzędu Statystycznego Wspólnot Europejskich danych statystycznych objętych zasadą poufności, rozporządzenie Rady (WE) nr 322/97 w sprawie statystyk Wspólnoty oraz decyzję Rady 89/382/EWG, Euratom w sprawie ustanowienia Komitetu ds. Programów Statystycznych Wspólnot Europejskich (Dz.U. L 87 z 31.3.2009, s. 164). Rozporządzenie to daje fundamentalne prawo w art. 17b ust. 1 dla krajowych urzędów statystycznych, które mogą zwrócić się do prywatnego posiadacza danych (np.

operator aplikacji rezerwacyjnej) o bezpłatne udostępnienie danych i odpowiednich metadanych, jeżeli dane będące przedmiotem wniosku są absolutnie niezbędne do opracowywania, tworzenia i rozpowszechniania statystyki europejskiej i nie można ich uzyskać w inny sposób lub ich wykorzystanie spowoduje znaczne zmniejszenie obciążenia posiadaczy danych i innych przedsiębiorstw, związanego z udzielaniem odpowiedzi. Czy takie dane będą mogły być bezpłatnie uzyskiwane np. od operatorów telefonii komórkowej lub operatorów kart płatniczych, to pokaże przyszłość i interpretacja odpowiednich stanów faktycznych oraz regulacji, które zawarto w nowelizacji wskazanego rozporządzenia.

Drugim istotnym aktem prawnym na poziomie UE jest również obowiązujące wprost we wszystkich państwach członkowskich Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1028 z dnia 11 kwietnia 2024 r. w sprawie gromadzenia i udostępniania danych dotyczących usług krótkoterminowego najmu lokali mieszkalnych i zmieniające rozporządzenie (UE) 2018/1724, które wchodzi w życie 20 maja 2026 r.⁵ Kluczowym w przypadku tego aktu prawnego jest art. 9 ust. 1 wg którego co do zasady dostawcy platform internetowych pośredniczących w najmie krótkoterminowym gromadzą i co miesiąc przekazują do pojedynczego cyfrowego punktu dostępu w państwie członkowskim dane dotyczące działalności w odniesieniu do każdego podmiotu wraz z odpowiednim numerem identyfikacyjnym podanym przez wynajmującego. Ponadto dane, które będą przekazywane mają dotyczyć dokładnego adresu podmiotu oraz adresu URL oferty. Przekazywanie to odbywać się będzie za pomocą środków łączności maszyna-maszyna⁶.

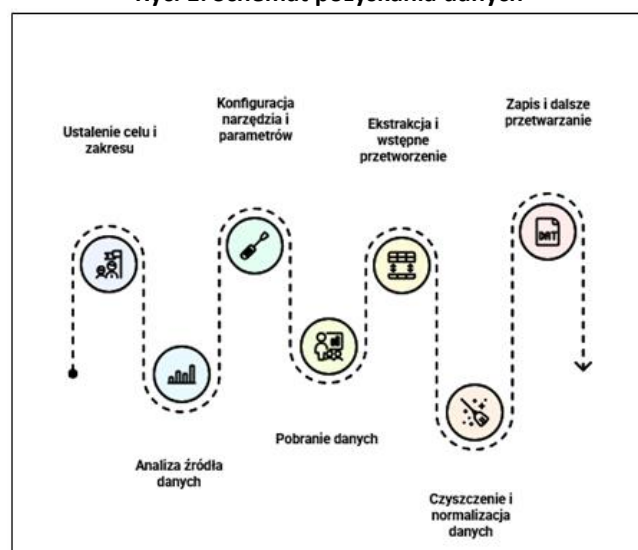
„Web scraping” jako metoda pozyskiwania danych w turystyce

„Web scraping” może być obecnie uznawany za jedną z kluczowych technik pozyskiwania danych big data na potrzeby statystyki publicznej. Polega na automatycznym gromadzeniu informacji z witryn internetowych przez specjalne programy, co umożliwia ich systematyczną analizę na przykład w czasie rzeczywistym, w tym m.in. w

sektorze turystyki. Metoda ta znalazła szerokie zastosowanie przy monitorowaniu inflacji – przykładem jest projekt *Billion Prices Project* realizowany przez MIT (Cavallo & Rigobon, 2016), a także inicjatywy Eurostatu, brytyjskiego ONS, belgijskiego Statbel (Juszczak) czy polskiego GUS, gdzie dane dotyczące opinii związanych z pobytami turystów w obiektach noclegowych pozyskano metodą web scrapingu z wybranych portali rezerwacyjnych zawierających oferty wynajmu zakwaterowania na terenie Polski. Podobnie Narodowy Bank Polski wykorzystywał metodę w prognozowaniu inflacji cen żywności (z czasem projekt rozszerzono o kolejne kategorie produktów), wskazując, że w przeciwieństwie do innych źródeł zbieranie danych online nie wiąże się z opóźnieniami, co umożliwia przygotowanie prognoz i analiz niemal w czasie rzeczywistym (Macias & Stelmasiak, 2019).

Współcześnie stwierdza się, że web scraping przyspiesza proces gromadzenia danych, ale jego ograniczenia i potencjalne wady wymagają starannego rozważenia (Thakur, Jha, Dalawat, Jha, & Khandare, 2025). Niemniej z całą pewnością jest on cennym zasobem uzupełniającym informacje dostarczane przez oficjalne statystyki, bowiem pozwala na uchwycenie jakości oferty turystycznej, a w rezultacie do przewidywania przyszłych zachowań turystów i opracowywania strategii w oparciu o przeszłe doświadczenia (Boboc, Babaligea, Ghita & Saseanu, 2025). Ponadto główną zaletą jest zbieranie danych dużych wolumenów w czasie rzeczywistym.

Ryc. 1. Schemat pozyskania danych



Źródło: opracowanie własne przy wykorzystaniu NapkinAI.

⁵ D. Borek, Management of short-term rental (str) – regulation or deregulation? – Example of Poland (w:) Sustainable Development - Circular Economy, Critical Infrastructures, Green Transition red. S.Baltowa, A. Cholakova, Sofia 2023.

⁶ D. Borek, Regulacja najmu krótkoterminowego (short-time rental STR) na poziomie krajowym i unijnym jako determinanta zarządzania obiektami zakwaterowania (w:) Europejski Przegląd Prawa i Stosunków Międzynarodowych nr 2/2024/70.

najczęściej używa się R i Pythona, które oferują liczne biblioteki, m.in. API, Selenium czy BeautifulSoup.

Przetwarzanie danych pozyskanych techniką web scrapingu realizowane jest z wykorzystaniem text miningu, czyli metody analizy i przekształcania tekstu służącej identyfikacji wzorców, trendów oraz zależności w danych tekstowych (Maślankowski, 2019).

Scraping okazał się skuteczną metodą badawczą służącą tworzeniu baz danych i weryfikacji hipotez. Literatura wskazuje na liczne projekty wykorzystujące big data pozyskiwane z portali internetowych i innych źródeł online. Przykładowo, badanie The Burning Glass Research, oparte na analizie ponad 50 mln ofert pracy zebranych z 40 tys. źródeł (2016–2017), pozwoliło zidentyfikować wymagane umiejętności i pogrupować je w 550 klastrów i 958 zawodów według klasyfikacji O*NET. Z kolei projekt Where the Works.org w Wielkiej Brytanii ujawnił niedopasowanie między kwalifikacjami kandydatów a oczekiwaniami pracodawców, analizując rynek pracy dla zawodów o średnich kwalifikacjach, jednak w ujęciu zawodów, a nie kompetencji (Vankevich & Kalinouskaya, 2021). W sektorze turystyki przykładem zastosowania web scrapingu jest analiza portalu Booking.com, z którego pozyskano dane dotyczące 8884 obiektów w Polsce i 3873 w Bułgarii, obejmujące 33 zmienne opisujące cechy obiektów i oferowane udogodnienia. Podobne badanie przeprowadzono na Hotels.com dla 4620 obiektów w Polsce i 532 w Bułgarii, gromadząc 12 zmiennych, przy czym jedynie 0,2% obiektów nie spełniało kryteriów klasyfikacyjnych. Z kolei z portalu Airbnb.com zebrano informacje o 12 556 obiektach w Polsce i 4174 w Bułgarii, obejmujące 8 podstawowych zmiennych. W przeciwieństwie do Booking.com, niemal wszystkie obiekty Airbnb (99,9%) przypisano do określonego typu, jednak klasyfikacja ta opierała się na ogólnych kategoriach, które nie odpowiadają oficjalnym statystykom (Cierpiat-Wolan & Stateva, 2023). Podobnie Barczak (2025) zaprezentował koncepcję scrapowania danych z portalu Booking w okresie jednego miesiąca w celu weryfikacji liczby dostępnych miejsc w obiektach noclegowych, z jednoczesną analizą zmian w czasie (na przykładzie województwa kujawsko-pomorskiego). W tej metodzie scrapowano dane na kolejny miesiąc, a następnie weryfikowano (raz w tygodniu) skalę zmian w dostępności miejsc noclegowych. Ostatecznie zdefiniowano listę 802 obiektów i ponad 40 tysięcy rekordów danych, wskazujących m.in. na nazwę obiektu, jego adres, standard, kordynaty GPS, opis czy rating, a przede wszystkim ilość dostępnych miejsc noclegowych.

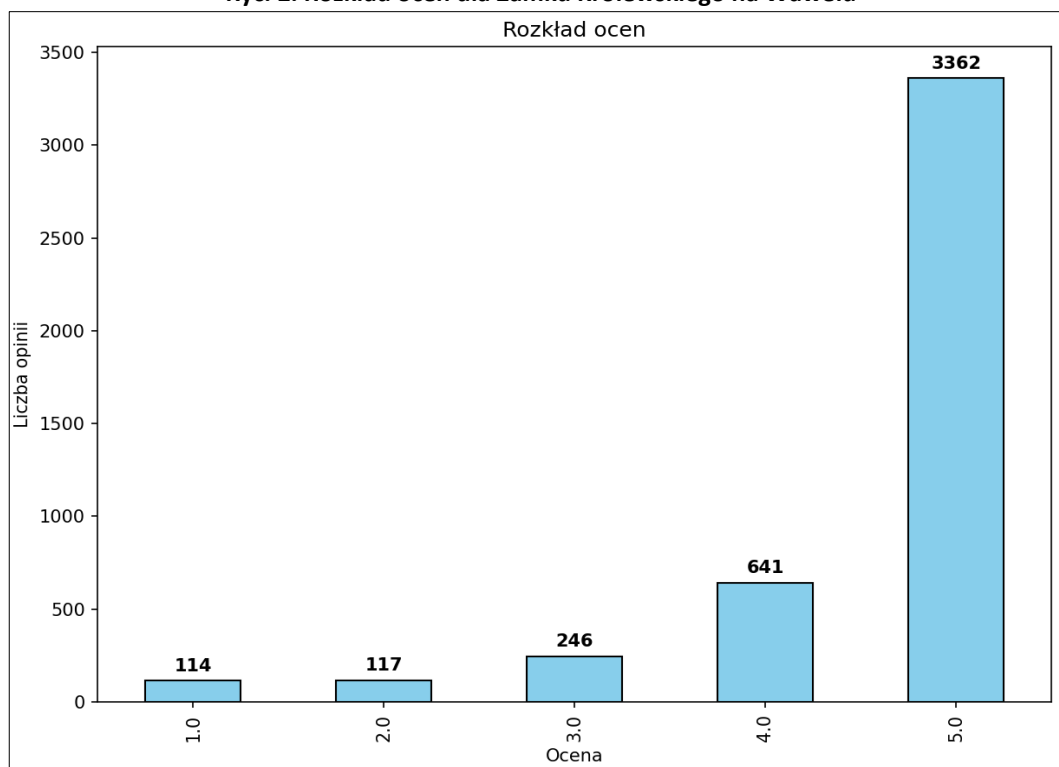
Z kolei Wąsowicz-Zaborek (2019) przeprowadziła analizę treści generowanych przez użytkowników TripAdvisora, Zoovera i Twittera (obecnie X) na temat Kołobrzegu, wykorzystując web crawlera do pozyskania danych. Narzędzie to, choć powiązane z web scrapingiem, pełni odmienną funkcję – przeszukuje wiele stron w poszukiwaniu treści, podczas gdy scraping koncentruje się na precyzyjnym pobieraniu danych z wybranych źródeł. Zebrane komentarze poddano analizie w Statistica (TextMining) i SPSS, co pozwoliło określić najczęściej występujące słowa oraz stworzyć profil semantyczny. Wyniki wskazały, że na wizerunek Kołobrzegu kluczowo wpływają atrakcje turystyczne, wydarzenia, gastronomia i baza noclegowa. Badanie, ograniczone do treści tekstowych z otwartych platform i jednej destynacji, nie daje podstaw do szerokich uogólnień, jednak może stanowić cenne źródło wiedzy dla samorządów, organizacji turystycznych i przedsiębiorców.

Wyniki eksperymentu

Celem publikacji jest ocena potencjału opinii z serwisów społecznościowych (na przykładzie Google Maps) jako alternatywnego źródła danych o atrakcyjności i jakości usług w turystyce. Metoda ta, mimo swojego potencjału analitycznego, nie gwarantuje automatycznie uzyskania wartościowych informacji – czasem dane pozyskane w ten sposób mogą być obarczone błędami, niepełne lub pochodzić z niejednoznacznych źródeł, co rodzi wyzwania związane z ich jakością i wiarygodnością.

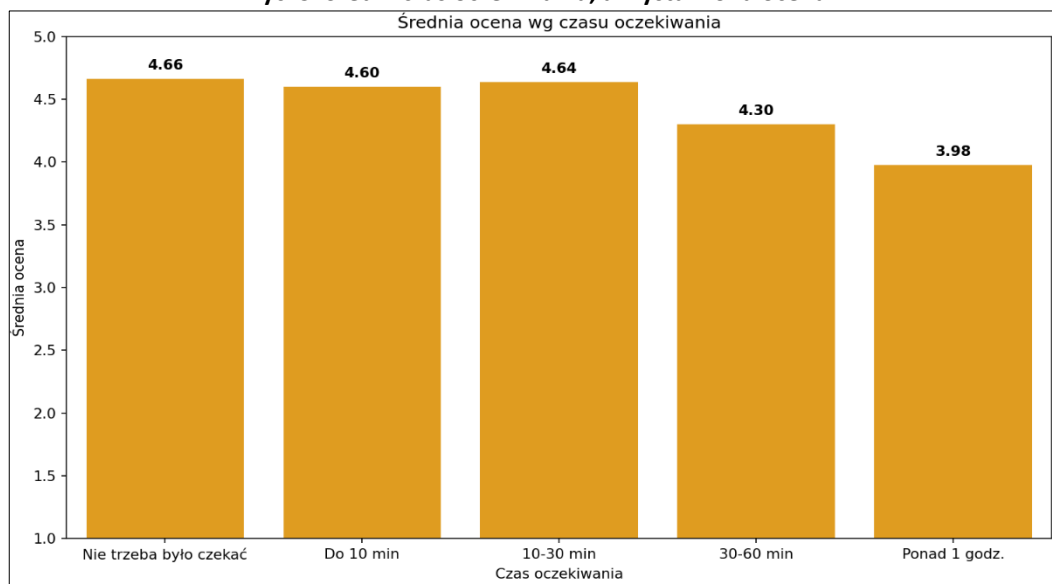
W wyniku przeprowadzonego na potrzeby niniejszej publikacji eksperymentu zdecydowano się na pobranie próbki opinii z serwisu Google Maps dla Zamku Królewskiego na Wawelu w Krakowie. Zakres danych zawierał 4480 opinii i obejmował datę wystawienia opinii, ocenę, informację o płci oceniającego, datę wizyty (weekend, dzień powszedni), czas oczekiwania, informacje o zalecanej rezerwacji, a także całą treść opinii. W pierwszej kolejności przeanalizowano rozkład ocen w prezentowanej na platformie pięciostopniowej skali (patrz ryc. nr 2), a także dokonano weryfikacji relacji pomiędzy średnią opinią a czasem oczekiwania (patrz ryc. nr 3). Z prezentowanych danych wynika, jak wydłużanie czasu oczekiwania na wizytę znacząco wpływa na pozostawioną ocenę.

Ryc. 2. Rozkład ocen dla Zamku Królewskiego na Wawelu



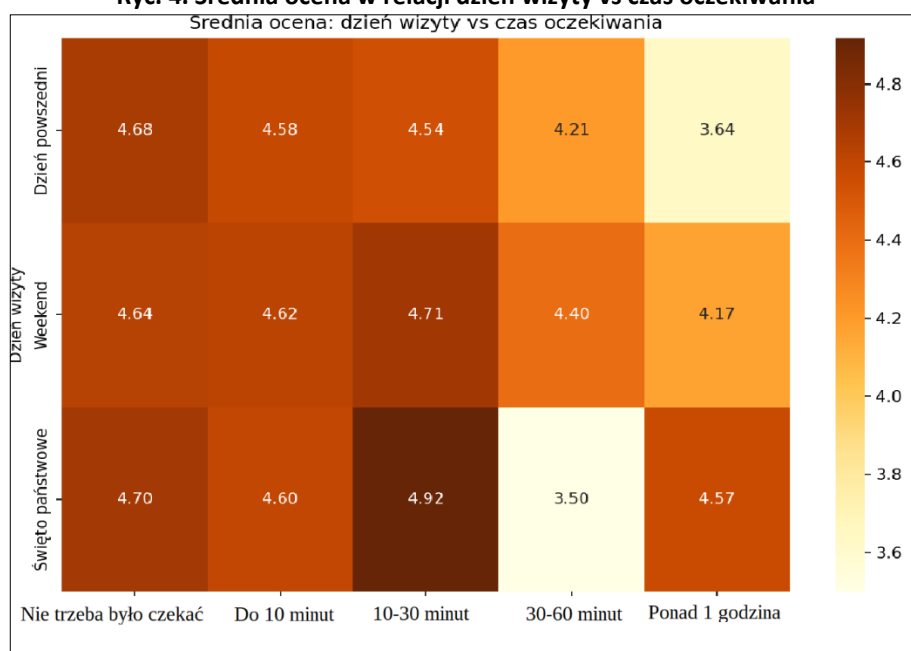
Źródło: opracowanie własne (N=4480).

Ryc. 3. Średni czas oczekiwania, a wystawiona ocena



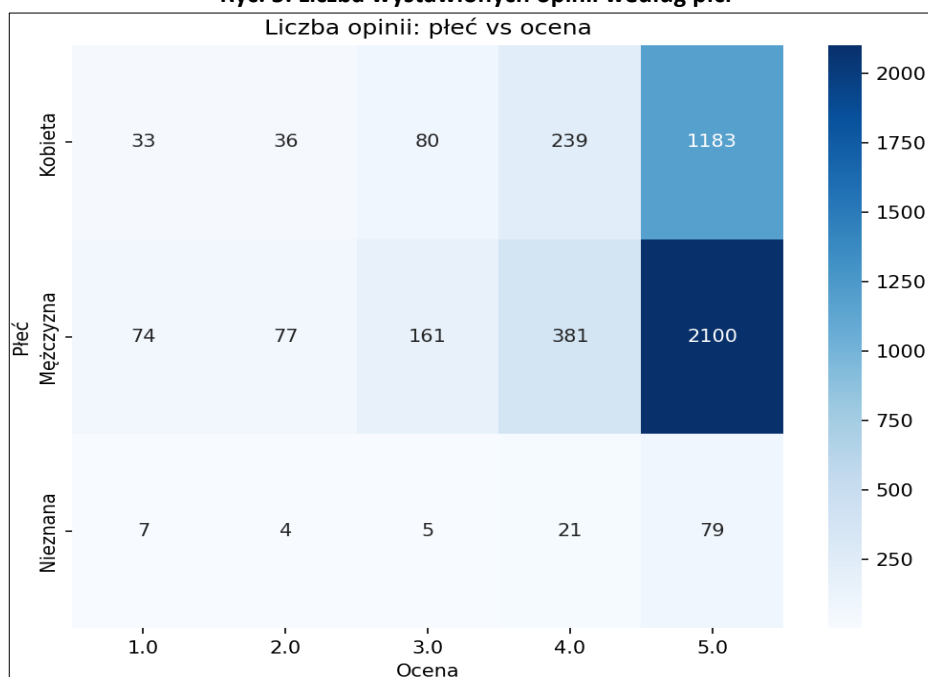
Źródło: opracowanie własne.

Ryc. 4. Średnia ocena w relacji dzień wizyty vs czas oczekiwania



Źródło: opracowanie własne.

Ryc. 5. Liczba wystawionych opinii według płci



Źródło: opracowanie własne.

Analiza przedstawiona na ryc. 4 wskazuje wyraźnie, że negatywny wpływ na ocenę obiektu pojawia się przede wszystkim w przypadku kumulacji dwóch czynników: wizyty w dniu o zwiększonym natężeniu ruchu (święto państwowe, weekend) oraz długiego czasu oczekiwania. Najniższe średnie oceny (3,5) odnotowano w sytuacji, gdy turyści odwiedzali obiekt w święto państwowe i musieli czekać 30–60 minut na wejście. Z drugiej strony, nawet w okresach dużego obłożenia, utrzymanie czasu oczekiwania na poziomie do 10–30 minut pozwalało na zachowanie wysokich ocen (4,6–4,9). Wskazuje to, że sprawne zarządzanie ruchem turystycznym i organizacja wejść mają kluczowe znaczenie dla poprawy doświadczenia odwiedzających. Wyniki te potwierdzają konieczność wdrażania rozwiązań ograniczających czas oczekiwania, takich jak systemy rezerwacji online, bilety czasowe czy dynamiczne zarządzanie dostępnością, co może w znacznym stopniu zniwelować negatywne skutki dużej frekwencji i poprawić satysfakcję turystów.

Źródło danych pozwoliło także na weryfikację ocen z perspektywy płci (patrz ryc. nr 5). W badanej próbie bowiem 1571 osób stanowiły kobiety, 2793 mężczyźni, a dla 116 osób niemożliwe było ustalenie płci, ze względu na brak podanego imienia na podstawie którego definiowano obie grupy.

Ryc. 6. Chmura słów dla oceniających kobiet



Oczywiście pobór danych z wykorzystaniem web scrapingu pozwala na szereg analiz i prognoz, niemniej w niniejszej pracy zdecydowano się także na próbę wyodrębnienia chmury słów w podziale na poszczególne płcie (patrz ryc. nr 6). Wynika z niej, że wśród trzech najczęściej pojawiających się fraz w opiniach kobiet wskazywano „robi wrażenie”, „komnaty królewskie”, czy też „obowiązkowy punkt”.

Podsumowanie i wnioski

Autorzy przewidują (Barczak 2025) znacząca ewolucję w podejściu do pozyskiwania danych na potrzeby badań sektora turystyki. W minionych latach dominowały badania oparte na tradycyjnych raportach, często o charakterze opisowym, które publikowane były ze znaczącym opóźnieniem czasowym. Współcześnie coraz częściej wykorzystywane są raporty z interaktywnym „dashboardem”, które pozwalają na bieżący

monitoring ruchu turystycznego (Barczak, Brudniki, Wierzbiańska, 2023). Tymczasem web scraping pozwala także na analizę danych w perspektywie przyszłości, bowiem umożliwia on np. pobranie danych o dostępnych pokojach hotelowych na platformach rezerwacyjnych w kolejnych miesiącach. Z kolei połączenie danych ze źródeł publicznych jak GUS czy CWOH, danych z instytucji europejskich, danych bezpośrednio pozyskanych od obiektów z danymi pochodzącymi ze scrapowania stron może zaowocować uzyskaniem pełniejszej perspektywy sytuacji na rynku, co można nazwać „Tourism Demand Horizon” (Barczak 2025). Istotne działania zostały już podjęte m.in. poprzez decyzję o uruchomieniu naboru na Zaawansowane narzędzia i algorytmy SI w oparciu o dane publiczne wspomagające rozwój gospodarczy w obszarze turystyki. Perspektywa wykorzystania sztucznej inteligencji do integracji i analiz danych otwiera drogę do ekstrapolacji trendów na podstawie kolejnych zbiorów big data, których źródłem jest także web scraping.

Część empiryczna artykułu, oparta na wykorzystaniu metody web scrapingu do pozyskiwania danych typu big data, potwierdziła przydatność nowych źródeł informacji w analizie sektora turystyki. Wyniki eksperymentów można ująć w kilku kluczowych wnioskach. W pierwszej kolejności przeprowadzone doświadczenia dowiodły, że web scraping stanowi skuteczne narzędzie tworzenia baz danych oraz ewentualnego testowania hipotez. Umożliwia zarówno ocenę jakości oferty turystycznej, jak i prognozowanie przyszłych zachowań podróżnych. Pozyskiwanie danych online przebiega bez opóźnień, co daje możliwość monitorowania zjawisk społeczno-gospodarczych na bieżąco i podnosi wartość predykcyjną analiz.

Spis literatury

- Barczak, M. (2025). *Application of Big Data in Tourism Demand Research*. Sofia: Professor Marin Drinov Publishing House of Bulgarian Academy of Sciences.
- Barczak, M., Brudnicki, R., & Wierzbiańska, S. (2023). Koncepcja metodologii badań ruchu turystycznego dla województwa kujawsko-pomorskiego uwzględniająca nowoczesne technologie pozyskiwania danych źródłowych. Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.
- Boboc, C.R., Babaligea, A.N., Ghita, S.I. and Saseanu, A. S. (2025). Leveraging Web-scraping for Tourism Data Analysis: A Case Study on Romania. In: C. Vasiliu, D.C. Dabija, A. Tziner, D. Pleșea, V. Dinu, eds. 2025. 11th BASIQ International Conference on New Trends in Sustainable Business and Consumption. Oradea, Romania, 26–28 June 2025. Bucharest: Editura ASE, 174–181.
- Borek D., Management of short-term rental (2023) – regulation or deregulation? – Example of Poland (w:) Sustainable Development - Circular Economy, Critical Infrastructures, Green Transition red. S. Baltowa, A. Cholakova.
- Borek D., (2024). *Regulacja najmu krótkoterminowego (short-time rental STR) na poziomie krajowym i unijnym jako determinanta zarządzania obiektami zakwaterowania* (w:) Europejski Przegląd Prawa i Stosunków Międzynarodowych nr 2/2024/70.
- Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2), 151–178.
- Cierpień-Wolan, M., & Stateva, G. (2023). *Ocena metod integracji danych dotyczących turystyki z uwzględnieniem big data*. Wiadomości Statystyczne. The Polish Statistician, 68(12), 25–48.
- Juszczak, A. (2021). Zastosowanie danych scrapowanych w pomiarze dynamiki cen. *Folia Oeconomica. Acta Universitatis Lodzensis*, 1(352).
- Kachniewska, M. (2014). Big data analysis jako źródło przewagi konkurencyjnej przedsiębiorstw i regionów turystycznych. *Folia Turistica*, (32), 35–54.
- Stępień, A. (2018). Bezpieczeństwo danych osobowych w cyberprzestrzeni – Big Data. *Przedsiębiorczość i Zarządzanie*, 19(2/3), 56.
- Macias, P., & Stelmasiak, D. (2019). *Food inflation nowcasting with web scraped data* (NBP Working Paper No. 302). Narodowy Bank Polski, Warsaw.
- Maślankowski, J. (2019). Pozyskiwanie i analiza danych na temat ofert pracy z wykorzystaniem big data. *Wiadomości Statystyczne. The Polish Statistician*, 64(9), 60–74.
- Tene, O. (2011). Privacy: The new generations. *International Data Privacy Law*, 1(1), 15–27.
- Thakur, S., Jha, R., Dalawat, M., Jha, P., Khandare, A. (2025). Scraping Data from Google Maps: Comparisons and Experimentations. In: Ghinea, G., Semwal, V.B., Khandare, A. (eds) *Intelligent Computing and Networking*. IC-ICN 2024. Lecture Notes in Networks and Systems, vol 1172. Springer, Singapore. https://doi.org/10.1007/978-981-97-8631-2_14.
- Vankevich, A., & Kalinouskaya, I. (2021). Better understanding of the labor market using big data. *Ekonomia i Prawo. Economics and Law*, 20(3), 677–692.
- Wąsowicz-Zaborek, E. (2019). Treści generowane przez użytkowników Internetu jako źródło wiedzy o wizerunku miejsca docelowego podróży na przykładzie miasta Kołobrzeg. *Folia Turistica*, nr 53.
- Woźniczka, J. (2018). *Big data w marketingu: Szanse i zagrożenia*. Studia Oeconomica Posnaniensia, 6(6).