

Bartosz Kozicki

E-mail: bartosz.kozicki@wat.edu.pl; nr ORCID: 0000-0001-6089-952x
Wojskowa Akademia Techniczna, Wydział Logistyki

Jarosław Tomaszewski

E-mail: jarekt7@wp.pl; nr ORCID: 0000-0003-2365-0797
Ministerstwo Obrony Narodowej

Metodyka wyboru metody prognozowania wizyt na stronie internetowej w przedsiębiorstwie turystycznym

*Methodology for the selection of the forecasting method
of visits on the website in the tourist enterprise*

W artykule poruszono problem badawczy związany z analizą danych pierwotnych, jej oceną, wyborem najlepszej metody do prognozy na przyszłość. Badania rozpoczęto od wykorzystania narzędzi badawczych do poszukiwania istnienia w szeregu czasowym pierwotnym zależności. Następnym krokiem było pogrupowanie danych, ich analiza i ocena. Uzyskane oceny stały się przesłanką zbudowania modelu zerjedynkowego regresji wielorakiej w celu potwierdzenia wykrytych zależności. Stwierdzone zależności pozwoliły na sprowadzenie szeregu pierwotnego do stacjonarności. Szereg czasowy został podzielony na dwie części: uczącą i testową. Wskutek krytycznej analizy literatury i uzyskanych zależności wybrano trzy najlepsze metody do prognozy szeregu uczącego na okres równy szeregowi testowemu. Uzyskane prognozy zostały poddane ocenie przy zastosowaniu obserwacji wzrokowej i MAPE. Wybrano najlepszą metodę, którą wykonano prognozowanie szeregu czasowego pierwotnego na 2019 rok (202 przyszłe okresy).

Słowa kluczowe:

analiza wielowymiarowa, szereg czasowy, seria, prognozowanie, przedsiębiorstwo turystyczne

In this article the author raises the research problem regarding the analysis of original data, its evaluation and the selection of the best forecasting method for the future. The research was initiated with the application of research tools in order to search for the relationships within the original time series. The following step was to group data, analyze and evaluate them. The results obtained were the premise for the construction of a zero-one model of multiple regression in order to confirm the relationships found. The detected relationships enabled to bring down the original series to stationarity. Time series was divided into two parts: teaching and testing ones. Due to the critical analysis of literature and the relationships obtained, three best forecasting methods of teaching series were selected for the same period of testing series. The forecasting obtained were evaluated by means of visual observation and MAPE. The best method was selected for the forecasting of the original time series for 2019 (202 future periods).

Key words:

multidimensional analysis, time series, series, forecasting, tourist enterprise

JEL: M15

Wstęp

W ostatnich latach obserwuje się trend rosnący liczby turystów wypoczywających w Polsce. Jest to związane z wzrostem gospodarczym, jak również rozwojem infrastruktury, a także świadomości właścicieli ośrodków, którzy inwestują w poprawę warunków socjalno-bytowych, reklamę i własną markę.

Celem głównym opracowania jest przedstawienie metodyki wyboru najlepszej metody do prognozy na przyszłość rozpatrywanych danych retrospektywnych dotyczących liczby wejść na stronę internetową przedsiębiorstwa turystycznego.

W artykule zastosowano metody badawcze w postaci analizy literatury, która dotyczy zagadnień związanych z analizą szeregów czasowych, wielowymiaro-

wą analizą danych, prognozowaniem, analizą dokumentów źródłowych, metodą symulacji komputerowej oraz porównaniem. Dodatkowo użyto techniki badawczej w postaci programu komputerowego Statistica. Co więcej, zastosowano następujące narzędzia badawcze: wykres kwartyłowy, autokorelację, autokorelację cząstkową, regresję wieloraką, histogram, test Shapiro-Wilka.

Uzyskanie poprawnych wyników prognozy pozwala podjąć wiele decyzji związanych z planowaniem dotyczącym wydatkowania środków niezbędnych do efektywnego pozyskiwania oraz obsługi turystów.

Przedsiębiorstwo turystyczne i prognozowanie

Z przeprowadzonej analizy literatury i badań wstępnych wynika, że w Polsce obserwuje się przyrost liczby turystów korzystających z przedsiębiorstw turystycznych. Przedsiębiorstwo definiowane jest jako jednostka organizacyjna gospodarki narodowej służąca zaspokajaniu potrzeb społecznych, której celem jest osiąganie efektów ekonomicznych poprzez produkcję dóbr, świadczenie usług lub inną działalność (Gołębowski, 1997, s. 9). Natomiast przedsiębiorstwo turystyczne uznawane jest za wyspecjalizowaną jednostkę działalności gospodarczej zajmującą się obsługą ruchu turystycznego i zaspokajaniem potrzeb turystów.

Ważnym aspektem związanym z funkcjonowaniem przedsiębiorstw turystycznych staje się poszukiwanie nowych klientów. Są oni źródłem przyszłego przychodu. Jedną z najskuteczniejszych form pozyskiwania nowych klientów są dobrze przygotowane strony internetowe. W dzisiejszych czasach obserwuje się dużą konkurencję w tej kwestii i tylko najlepsi, ci, którzy potrafią działać dynamicznie, obserwując zmieniające się potrzeby, wygrywają i odnoszą sukcesy. W artykule poruszono ważny problem związany z metodyką wyboru metody prognozowania, opartej na dokładnej analizie i ocenie danych retrospektywnych wizyt na stronie internetowej przedsiębiorstwa turystycznego.

Szczególnym rodzajem analiz, dzięki którym można zaobserwować zależności w danych retrospektywnych, są analizy wielowymiarowe. Pozwalają one na zwiększenie efektywności działań poprzez zaobserwowanie zależności niewidocznych z perspektywy danych surowych, wykorzystywanych do budowy modeli prognozowania. Uznaje się je za grupę metod statystycznych, za pomocą których jednocześnie analizuje się co najmniej dwie zmienne opisujące zmienną zależną (Łuniewska, 2006, s. 9). W artykule zastosowano metody wielowymiarowe grupowania statystycznego. Wyodrębniono trzy zbiory grupujące w postaci lat, miesięcy, dni tygodnia i na ich podstawie dobrano metody prognozowania, które poddano

analizie i ocenie. Najlepszą metodą wykonano prognozowanie na 2019 r. szeregu czasowego pierwotnego. Uzyskane prognozy są niezwykle ważne, gdyż wskażą, ile w przyszłości klientów będzie korzystało z usług analizowanego przedsiębiorstwa.

Zdaniem P. Dittmanna prognozowanie jest racjonalnym, naukowym przewidywaniem przyszłych zdarzeń (Dittmann, 2016, s. 20). Prognozowanie powinno być oparte na poprawnej analizie szeregowych czasowych, która jest jedną z części statystyki stosowanej w praktyce gospodarczej. Pozwala ona na poznanie mechanizmów, które powodowały, że zjawisko przebiegało tak, a nie inaczej, oraz na przeprowadzenie symulacji związanej z przewidywaniem zaobserwowanych tendencji w przyszłości. Szereg czasowy w literaturze jest definiowany jako nielosowa funkcja $x(t)$, będąca realizacją procesu stochastycznego $X(w, t)$ dla ustalonego zdarzenia elementarnego w_0 (Kot, Jakubowski, Sokołowski, 2011, s. 366).

Analiza i ocena danych pierwotnych

Na rysunku 1 zestawiono dane pozyskane z przedsiębiorstwa turystycznego dotyczące liczby wejść na stronę internetową rozpatrywanego ośrodka w latach 2017–2019 w ujęciu dziennym. Średnia arytmetyczna analizowanego szeregu pierwotnego wynosi 48,89, a mediana 29. Szereg składa się z 893 elementów. Najniższy element szeregu to 2, a najwyższy 247. Kwartył dolny wyniósł 15, a trzeci 58. Rozstęp międzykwartyłowy wynosi 43. Oceną wstępnie przeprowadzonej statystyki opisowej szeregu przedstawionego na rysunku 1 jest przesłanka występowania wartości odstających i ekstremalnych, co sugeruje zastosowanie narzędzia badawczego w postaci wykresu ramka-wąsy.

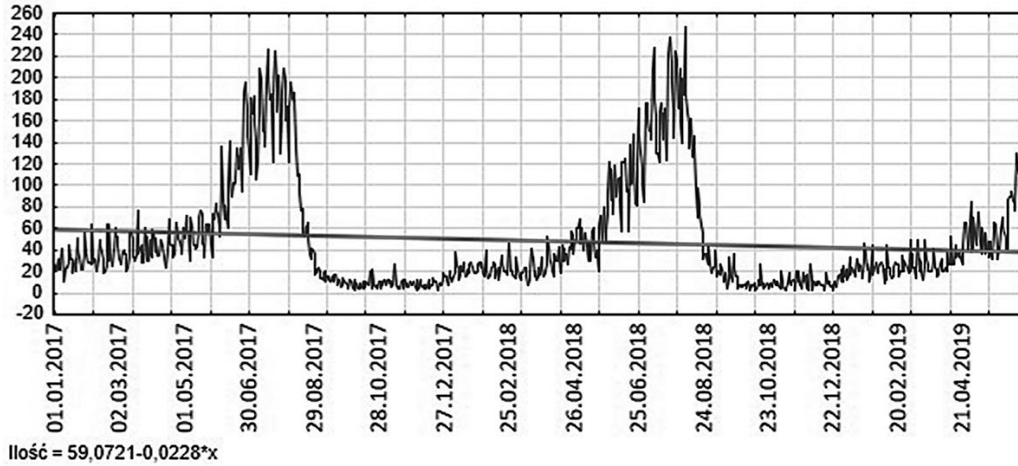
Z obserwacji wzrokowej rysunku 1 wynika, że widoczna jest sezonowość. Co więcej, zaobserwowano trend, czego przesłanką było jego nakreślenie i opisanie wzorem. Pierwszym etapem badań było poszukiwanie istnienia wartości odstających i ekstremalnych w analizowanym szeregu czasowym pierwotnym. Do tego celu użyto narzędzia badawczego w postaci wykresu ramka-wąsy wraz z nakreślonymi danymi surowymi, wartościami odstającymi, ekstremalnymi oraz kwartyłami i medianą.

Ocena rysunku 2 prowadzi do zaobserwowania wartości odstających i ekstremalnych, które widoczne są w miesiącach od maja do sierpnia w ujęciu dynamicznym. Mediana jest na poziomie 29.

Następnie wykonano badanie rozkładu analizowanego szeregu czasowego pierwotnego. Do tego celu użyto następujących narzędzi badawczych: histogramu z testem Shapiro-Wilka (rysunek 3) i wykresu normalności (rysunek 4).

Rysunek 1

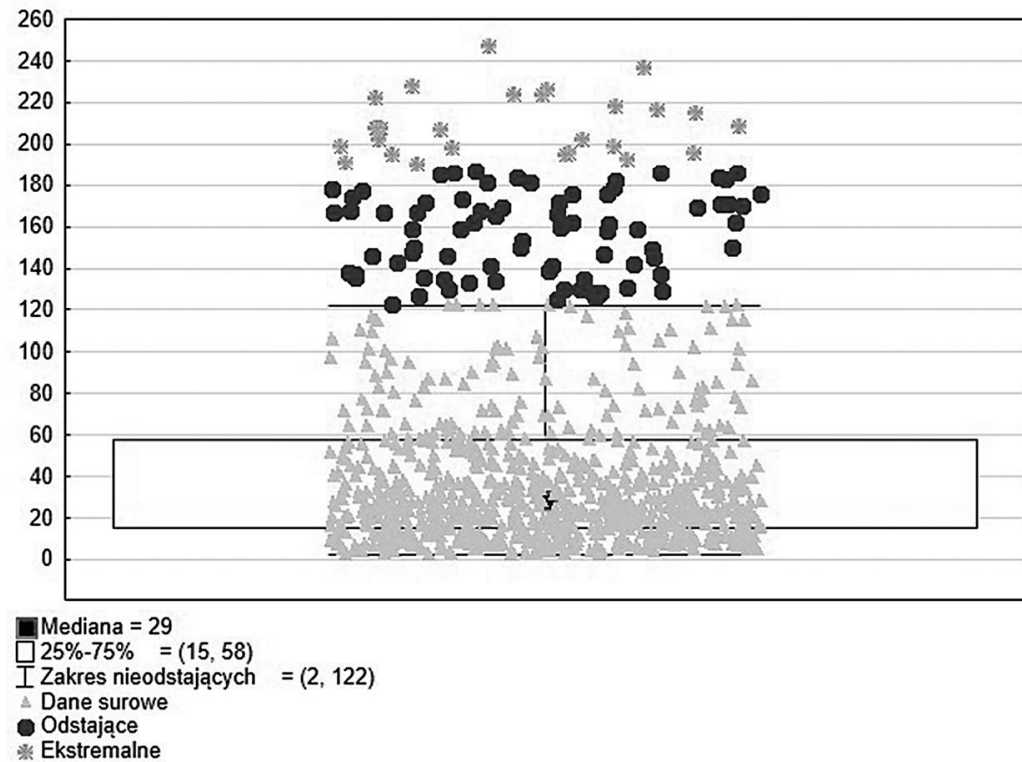
Zestawienie danych pierwotnych dotyczących wejść na stronę internetową przedsiębiorstwa turystycznego w latach 2017–2019



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 2

Wykres ramka-wąsy wraz z nakreślonymi danymi surowymi i medianą



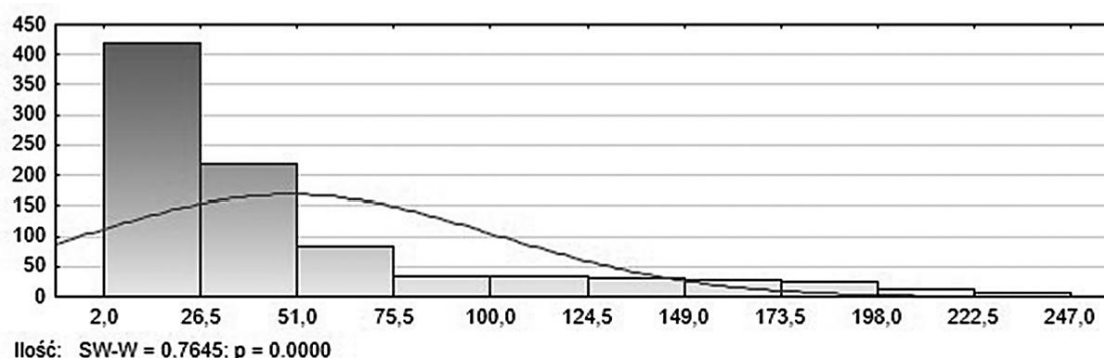
Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Ocena rozkładu szeregu czasowego pierwotnego każe stwierdzić, że rozkład nie jest normalny. Wynika to z dużej ilości danych zawartych w przedziale od

2 do 26,5 i wielu pojedynczych będących wartościami odstającymi i ekstremalnymi. Kwantyle odbiegają od nakreślonej linii prostej, a test Shapiro-Wilka jedno-

Rysunek 3

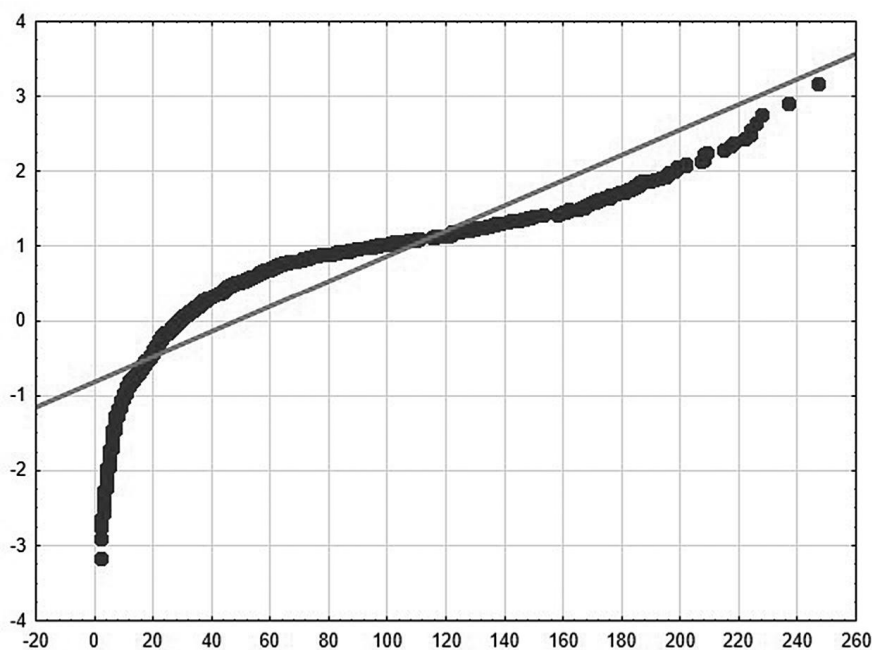
Histogram i test Shapiro-Wilka danych pierwotnych



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 4

Wykres normalności danych pierwotnych



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

znacznie potwierdza przyjęte stwierdzenie. Skośność analizowanego szeregu to 1,72, a kurtoza 2,2. Świadczy to o tym, że jest bardziej wysmukły od normalnego, prawostronnie asymetryczny.

Następnie przeprowadzono badanie istnienia zależności w analizowanym szeregu czasowym pierwotnym. Do tego celu użyto: autokorelacji (rysunek 5) i autokorelacji cząstkowej (rysunek 6).

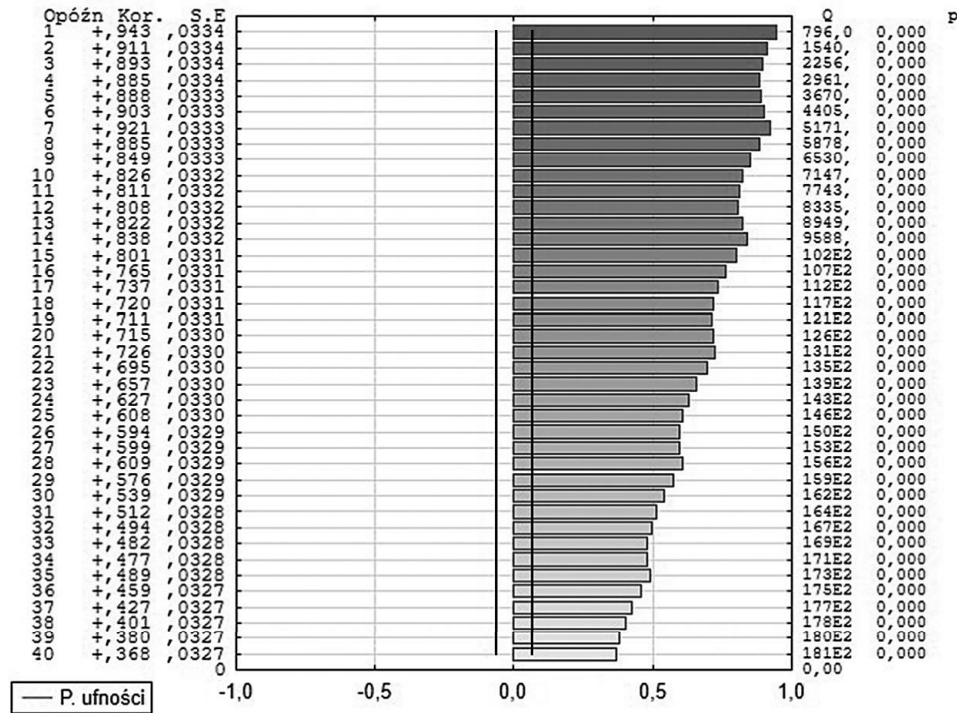
W wyniku oceny użytych narzędzi badawczych w postaci autokorelacji i autokorelacji cząstkowej

stwierdzono istnienie zależności w opóźnieniach szeregu czasowego pierwotnego. Co więcej, istnieje prawdopodobieństwo istnienia trendu oraz sezonowości w ujęciu miesięcznym i rocznym. Stało się to przesłanką pogrupowania danych pierwotnych w celu przeprowadzenia dalszych ich analiz i ocen.

Pierwszym przeprowadzonym grupowaniem było grupowanie ze względu na miesiące. Następnie zastosowano narzędzie badawcze w postaci skategoryzowanego wykresu ramka-wąsy wraz z nakre-

Rysunek 5

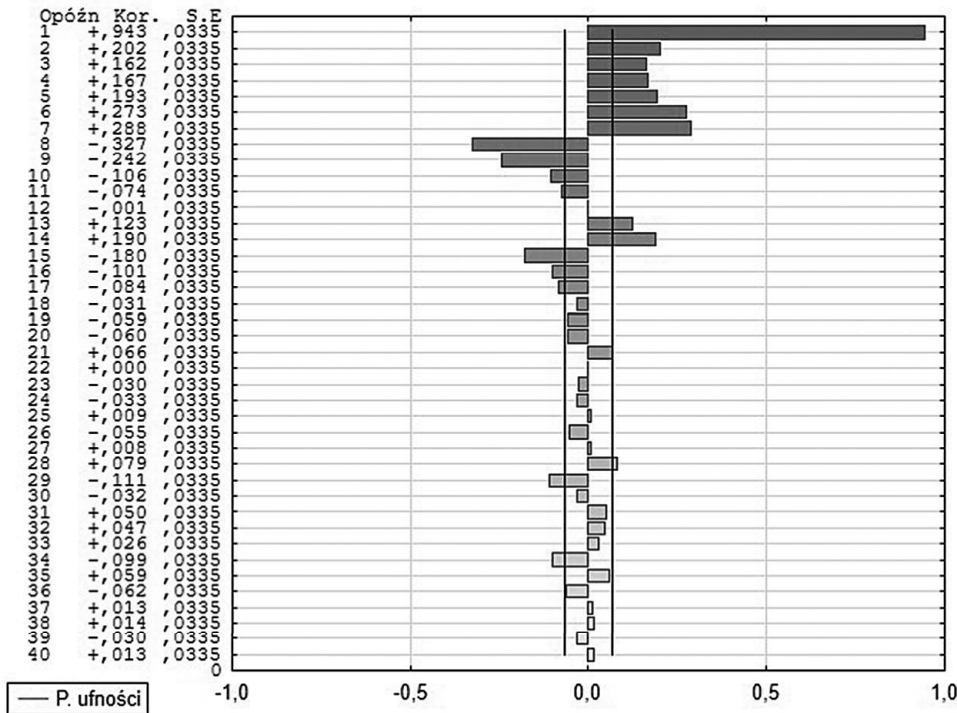
Autokorelacja danych pierwotnych



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 6

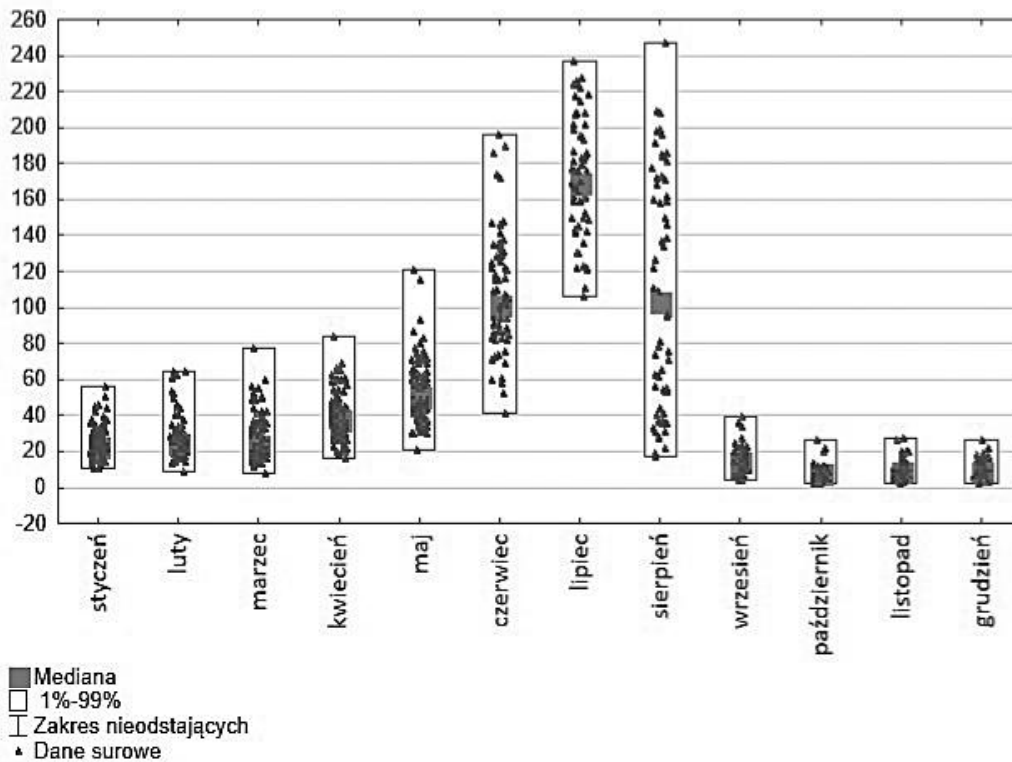
Autokorelacja cząstkowa danych pierwotnych



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 7

Skategoryzowany wykres ramka-wąsy wraz z nakreśloną medianą i percentylami danych pierwotnych z podziałem na miesiące w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

ślonymi medianami i percentylami w ujęciu dynamicznym.

Ocena rysunku 7 pozwala zaobserwować zależności w postaci sezonowości w ujęciu miesięcznym. Najdłuższy rozstęp międzypercentylowy widoczny jest w sierpniu, czerwcu i lipcu. Najwyższa mediana jest w lipcu, następnie w sierpniu i czerwcu.

W kolejnym kroku użyto narzędzia badawczego w postaci średnich arytmetycznych wraz z nakreślonymi przedziałami ufności $\pm 95\%$.

Użycie narzędzia badawczego w postaci skategoryzowanych średnich arytmetycznych wraz z przedziałami ufności danych pierwotnych potwierdziło zaobserwowanie sezonowości w ujęciu miesięcznym. Najszerzy przedziały ufności widoczny jest w sierpniu. Średnie arytmetyczne najwyższe widoczne są w lipcu, sierpniu i czerwcu.

Dalszym etapem badań było przeprowadzenie statystyki opisowej analizowanych danych w grupie miesiące (tabela 1).

Największa ilość danych w grupie miesiące poddana badaniu była w następujących przypadkach:

- 93 — styczeń, marzec i maj,
- 90 — kwiecień,
- 84 — luty,

- 72 — czerwiec,
- 62 — lipiec, sierpień, październik, grudzień,
- 60 — wrzesień i listopad.

Ocena przeprowadzonej analizy w grupie miesięcy prowadzi do zaobserwowania, że we wszystkich przypadkach (dwunastu miesiącach) średnia arytmetyczna jest większa od mediany. Ogółem na stronę w ujęciu dynamicznym weszło około 24 650 osób. Najwięcej osób oglądało stronę w lipcu — 10 693, a najmniej w październiku — 518. Największe odchylenie standardowe od średniej arytmetycznej widoczne jest w sierpniu (63,5), a najmniejsze w październiku (4,2).

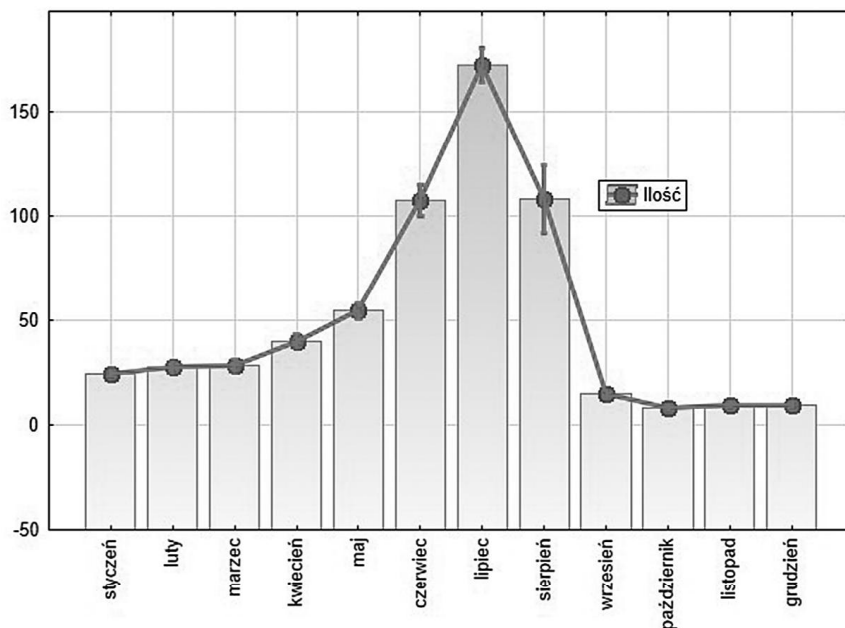
Dalszym etapem badań była analiza i ocena rozkładu analizowanych danych w ujęciu dynamicznym. Do tego celu użyto następujących narzędzi badawczych: wykresu normalności, histogramu i testów Shapiro-Wilka.

Z użytych narzędzi badawczych przedstawionych na rysunkach 9–10 i w tabeli 2 wynika, że rozkład normalny widoczny był tylko w dwóch miesiącach: czerwcu i lipcu.

Dalszym etapem badań była analiza grupy lat. Jako pierwszego użyto narzędzia w postaci skategoryzowanego wykresu ramka-wąsy wraz z nakre-

Rysunek 8

Skategoryzowany wykres średnich arytmetycznych wraz z przedziałami ufności
+–95% danych pierwotnych z podziałem na miesiące w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 1

Analiza statystyki opisowej — grupa: miesiące

	Średnia arytmetyczna	Ilość	Suma	Odchylenie standardowe	Wariancja	Mediana	średnia > mediana
styczeń	24,7	93	2297,0	8,8	77,5	23,0	1
luty	27,5	84	2314,0	11,9	142,1	25,0	1
marzec	28,4	93	2639,0	12,6	158,9	24,0	1
kwiecień	40,3	90	3624,0	13,4	178,6	38,0	1
maj	54,5	93	5071,0	17,4	301,6	51,0	1
czerwiec	107,4	72	7731,0	32,1	1031,7	102,0	1
lipiec	172,5	62	10693,0	32,3	1046,1	169,5	1
sierpień	108,3	62	6716,0	63,5	4033,0	103,0	1
wrzesień	14,7	60	880,0	7,5	56,4	14,0	1
październik	8,4	62	518,0	4,2	17,2	8,0	1
listopad	9,8	60	586,0	5,4	29,7	9,0	1
grudzień	9,5	62	587,0	4,8	22,9	8,5	1
Ogółem	48,9	893	43656,0	51,7	2670,1		

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

ślonymi percentylami, danymi surowymi i medianami.

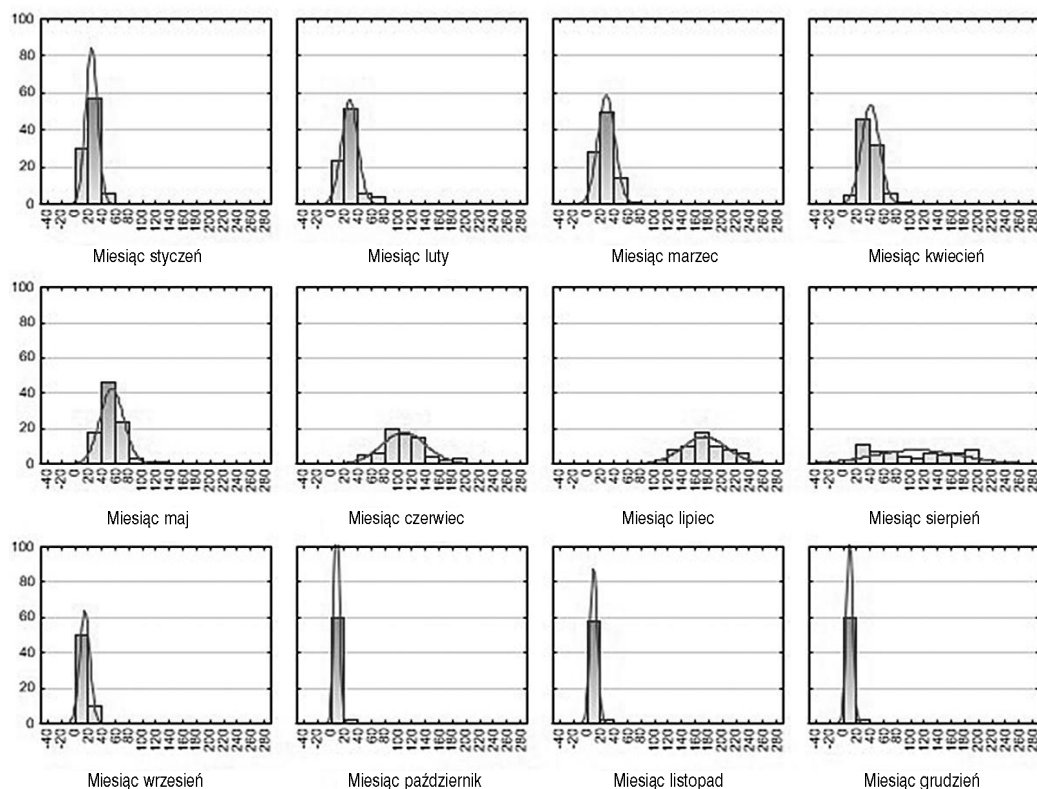
Ocena rysunku 11 pozwala na zaobserwowanie największego rozstępu międzypercentylowego dla

2018 r. Największa mediana była w 2017, następnie w 2019 i 2018 r. Najgęstsza ilość danych surowych widoczna jest w 2018 r.

Dalszym etapem badań była analiza średnich ary-

Rysunek 9

Kategoryzowane histogramy danych pierwotnych z podziałem na miesiące w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 2

Testy Shapiro-Wilka — grupa: miesiące

Miesiące	Pvalue	Test SW-W
styczeń	0,00004	0,9238
luty	0,00000	0,8580
marzec	0,00002	0,9175
kwiecień	0,02820	0,9686
maj	0,00020	0,9346
czerwiec	0,05360	0,9667
lipiec	0,42340	0,9804
sierpień	0,00140	0,9283
wrzesień	0,00060	0,9166
październik	0,00000	0,8219
listopad	0,00003	0,8813
grudzień	0,00140	0,9288

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

metrycznych wraz z przedziałami ufności $\pm 95\%$ (rysunek 12).

Najwyższa średnia arytmetyczna widoczna była w 2017 r., następnie w 2018 i 2019. Przedziały ufności najwyższe były w latach 2018 i 2017.

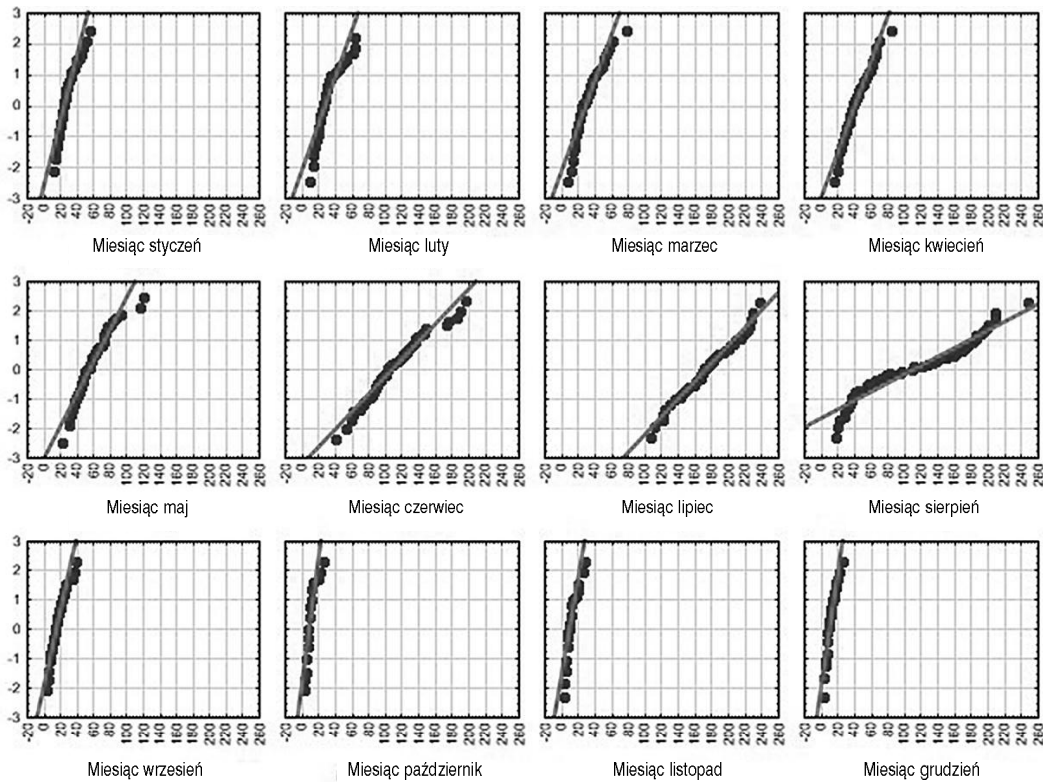
Następnie w tabeli 3 przeprowadzono analizę statystyki opisowej analizowanej grupy danych pierwotnych (lat).

W statystyce opisowej rozpatrywano 893 przypadki (tabela 3). Należy podkreślić, że w latach 2017 i 2018 było ich 365, a w roku 2019 — 163. Najwyższa średnia arytmetyczna była w 2017 r. i wyniosła 53, natomiast mediana w 2017 była na niższym poziomie — 34. W 2018 średnia arytmetyczna zmalała do 50,2, a mediana do 23. Najwyższe odchylenie standardowe od średniej arytmetycznej analizowanych danych widoczne było w 2018 roku i wyniosło 56,9. Najwyższa zmienność widoczna była w 2018 r. i wyniosła 3237.

Dalszym etapem badań była analiza oraz ocena rozkładu danych pierwotnych grupy lat. Do tego celu użyto następujących narzędzi badaw-

Rysunek 10

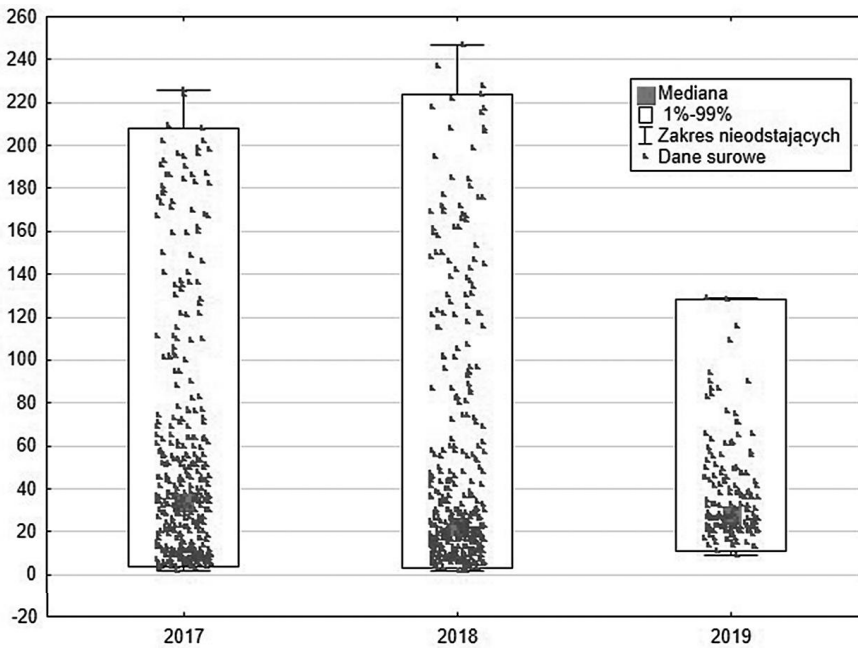
Skategoryzowany wykres normalności danych pierwotnych z podziałem na miesiące w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 11

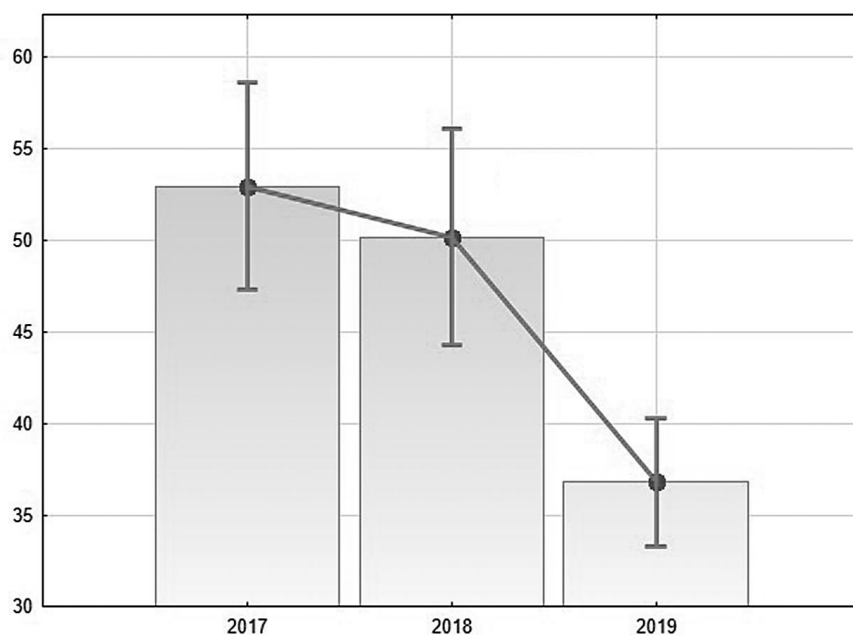
Skategoryzowany wykres ramka-wąsy wraz z nakreśloną medianą i percentylami danych pierwotnych z podziałem na miesiące w ujęciu rocznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 12

Skategoryzowany wykres średnich arytmetycznych wraz z przedziałami ufności $\pm 95\%$ danych pierwotnych z podziałem na lata w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 3

Analiza statystyki opisowej — grupa: lata

	Średnia arytmetyczna	Ilość	Suma	Odchylenie standardowe	Wariancja	Mediana	średnia > mediana
2017	53,0	365	19331,0	54,7	2993,2	34,0	1
2018	50,2	365	18322,0	56,9	3237,0	23,0	1
2019	36,8	163	6003,0	22,7	515,5	28,0	1
Ogółem	48,9	893	43656,0	51,7	2670,1		

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

czych: histogramów (rysunek 13), wykresu normalności (rysunek 14) oraz testów Shapiro-Wilka (tabela 4).

Tabela 4

Analiza testów Shapiro-Wilka — grupa: lata

Lata	Pvalue	Test SW-W
2017	0,79	0,00
2018	0,75	0,00
2019	200,82	0,00

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Ocena przeprowadzonej analizy w zakresie rozkładu kaže jednoznacznie stwierdzić brak normalności analizowanych danych pierwotnych w grupie lat.

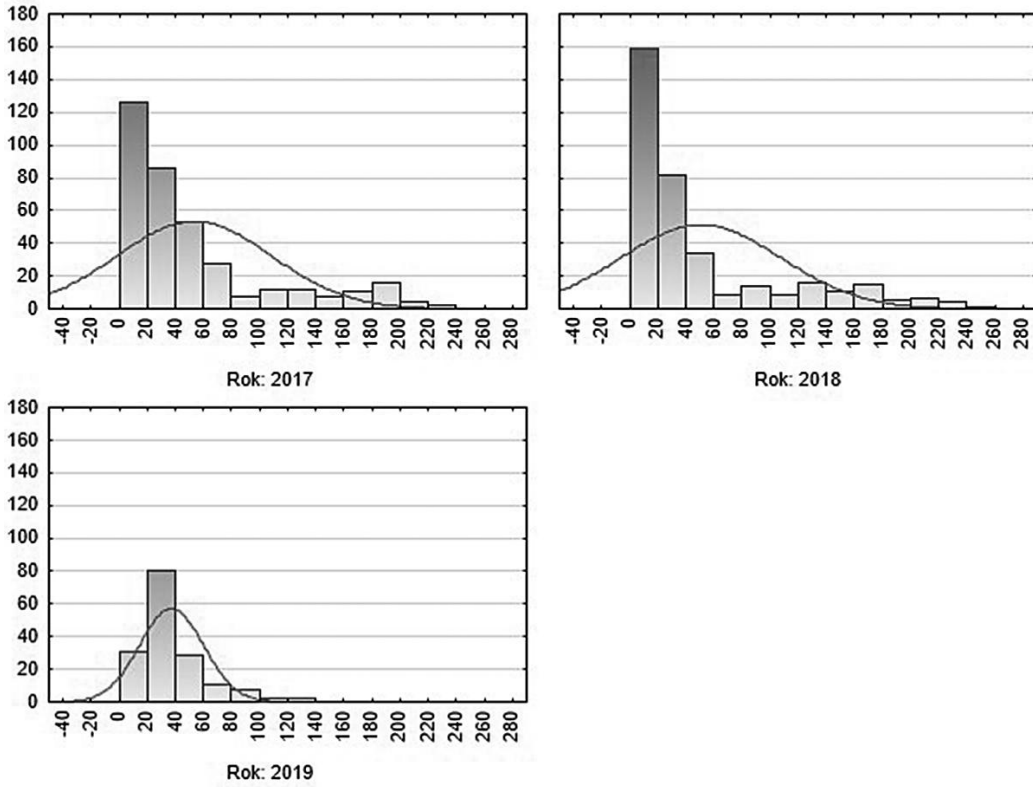
Ocena przeprowadzonej całej analizy grupy danych pierwotnych dotyczących lat prowadzi do stwierdzenia istnienia trendu malejącego.

Dalszym etapem badań jest analiza i ocena grupy danych pierwotnych dotyczących dni tygodnia. Pierwszym użytym narzędziem był skategoryzowany wykres ramka-wąsy wraz z nakreśloną medianą i percentylami danych pierwotnych.

Dokonując oceny danych z rysunku 15, zaobserwowano najwyższy poziom mediany dla niedzieli. Najwyższy rozstęp międzykwartyłowy widoczny jest we wtorek, a zdecydowanie najmniejszy w sobotę.

Rysunek 13

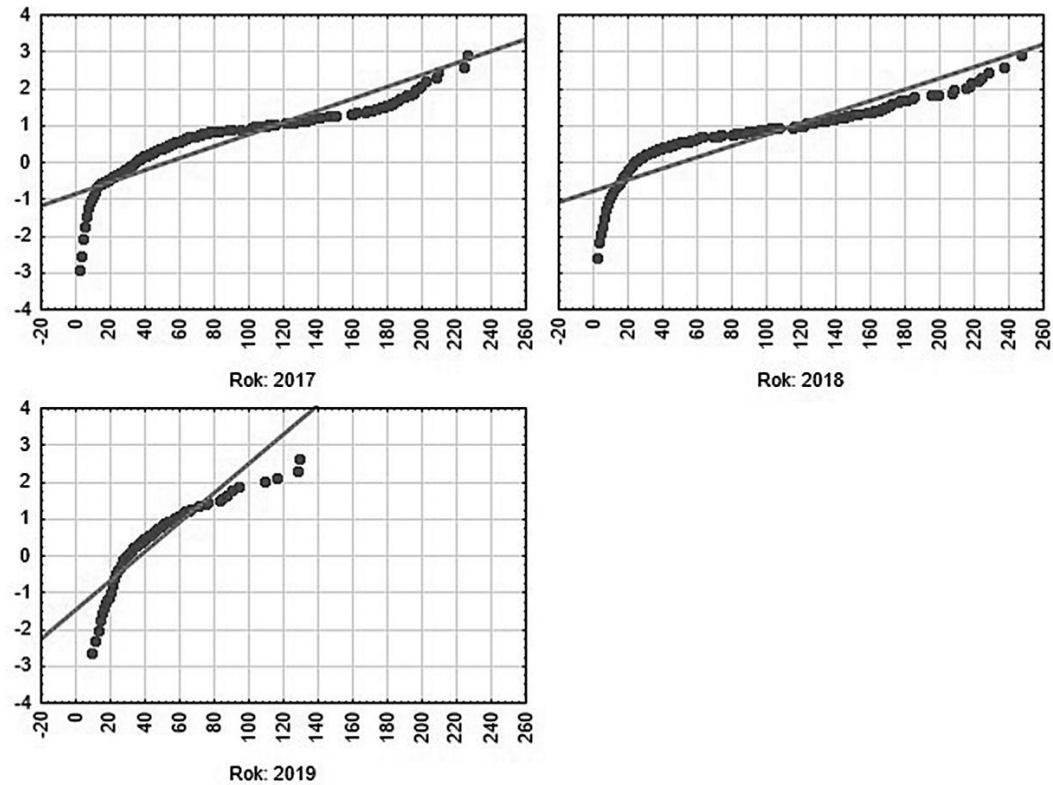
Skategoryzowane histogramy danych pierwotnych z podziałem na lata w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 14

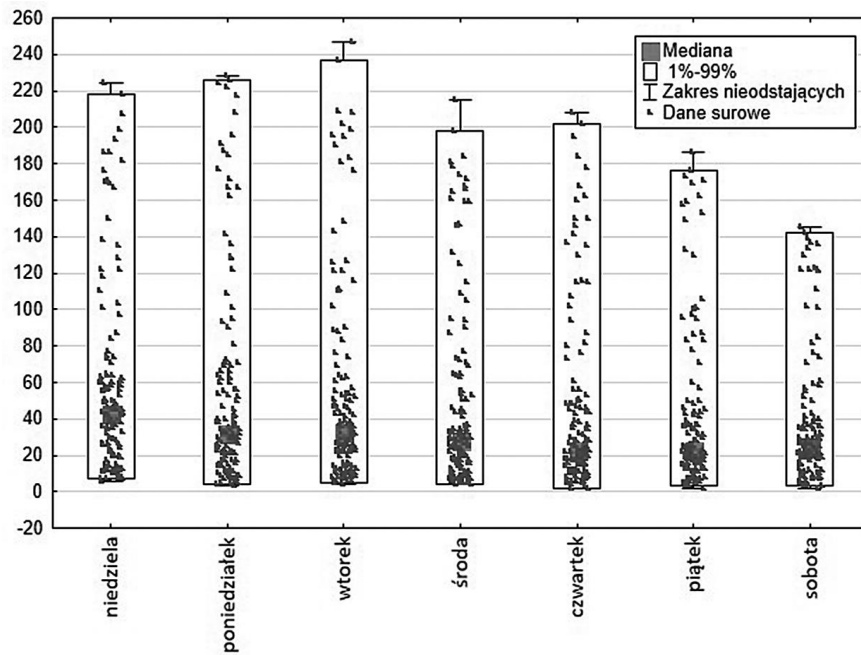
Skategoryzowany wykres normalności danych pierwotnych z podziałem na lata w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 15

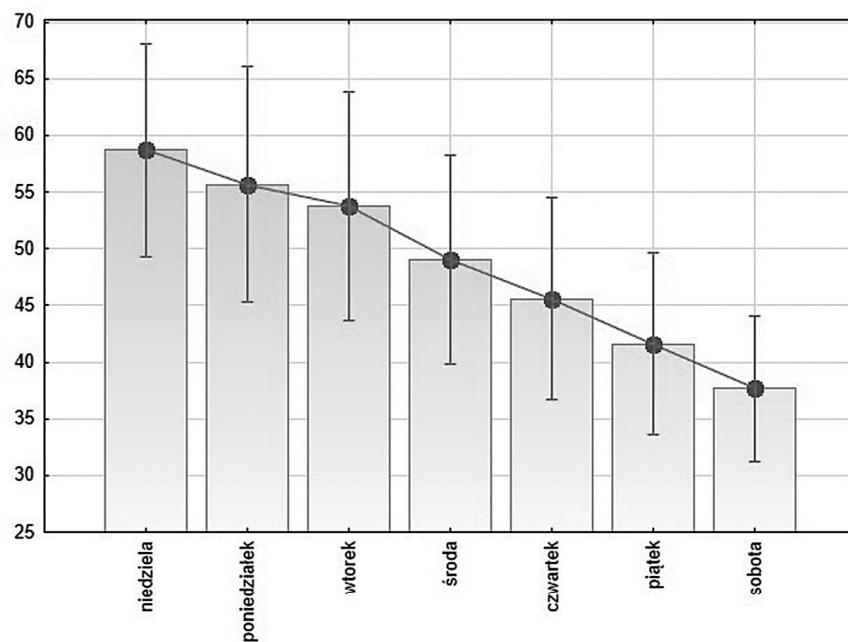
Skategoryzowany wykres ramka-wąsy wraz z nakreśloną medianą i percentylami danych pierwotnych z podziałem na dni tygodnia w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 16

Skategoryzowany wykres średnich arytmetycznych wraz z przedziałami ufności $\pm 95\%$ danych pierwotnych z podziałem na dni tygodnia w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 5

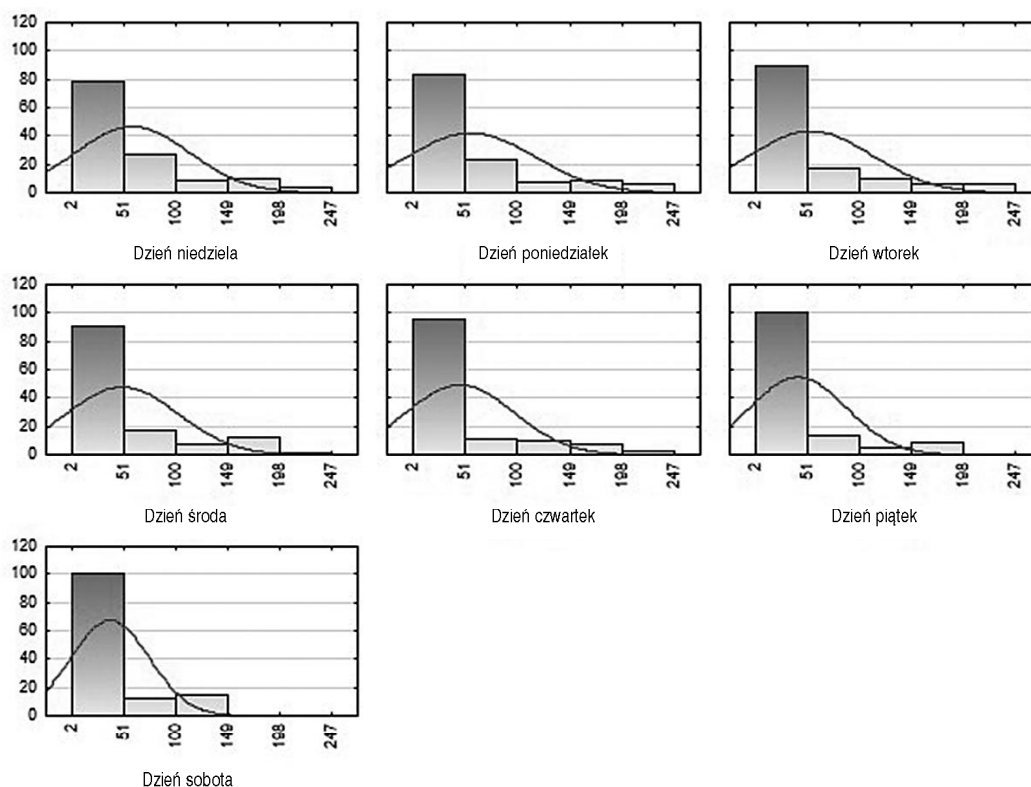
Analiza statystyki opisowej — grupa: dni tygodnia

	Średnia arytmetyczna	Ilość	Odchylenie standardowe	Wariancja	Mediana	średnia>mediany
niedziela	58,7	128	53,6	2877,5	42,5	1
poniedziałek	55,7	128	59,4	3531,4	32,0	1
wtorek	53,8	128	57,6	3316,5	32,5	1
środa	49,0	128	52,5	2752,4	28,5	1
czwartek	45,6	127	50,5	2548,2	23,0	1
piątek	41,6	127	45,4	2061,0	23,0	1
sobota	37,6	127	36,8	1355,7	25,0	1
Ogółem	48,9	893	51,7	2670,1		

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 17

Skategoryzowane histogramy danych pierwotnych z podziałem na dni tygodnia w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Dalszym etapem badań była analiza średnich arytmetycznych i przedziałów ufności (rysunek 16).

Oceną rysunku 16 jest zaobserwowanie, że najwyższa średnia arytmetyczna widoczna była w kolejności malejącej w następujących dniach tygodnia: niedziela, poniedziałek, wtorek, środa, czwartek, piątek i w sobotę.

Dalszym etapem badań było przeprowadzenie analizy statystyki opisowej grupy danych pierwotnych dotyczących dni tygodnia.

W analizie statystyki opisowej (tabela 5) użyto 893 przypadków, z czego po 128 przypadko na: niedzielę, poniedziałek, wtorek i środę, a w pozostałe rozpatry-

wane dni było ich 127. We wszystkich analizowanych przypadkach średnia arytmetyczna była wyższa od mediany. Największe odchylenie standardowe od średniej arytmetycznej widoczne było w poniedziałek i wyniosło 59,4. W poniedziałek widoczna była również najwyższa zmienność, która kształtowała się na poziomie 3531,4. Najwyższy poziom mediany był w niedzielę i wyniósł 42,5. Co więcej, średnia arytmetyczna w niedzielę była najwyższa — 58,7. Najniższy poziom mediany był odnotowany w dwóch przypadkach: w czwartek i piątek. Natomiast najniższa średnia arytmetyczna w sobotę — 37,6.

Dalszym etapem badań była analiza i ocena rozkładu analizowanych danych pierwotnych dotyczących grupy dni tygodnia (rysunki 17–18 i tabela 6).

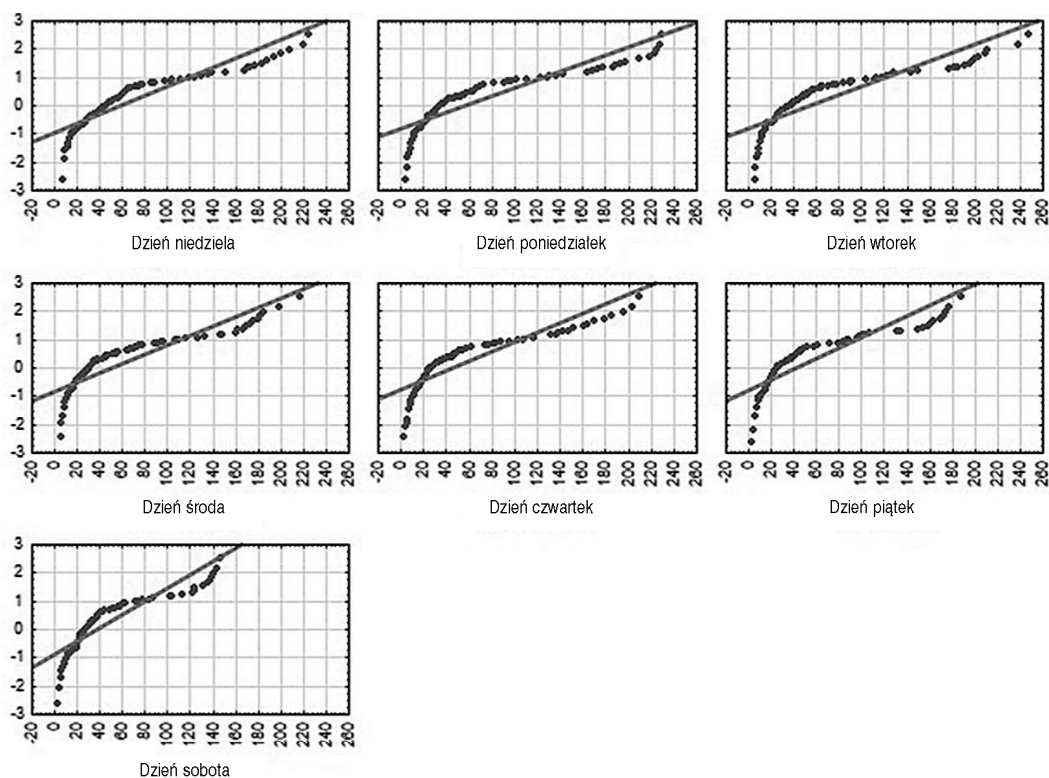
Ocena wykonanej analizy rozkładu przedstawionej na rysunkach od 17 do 18 i w tabeli 6 prowadzi do stwierdzenia braku rozkładu o charakterze normalnym.

Ocną wykonanej analizy jest przesłanka istnienia sezonowości w grupie dni tygodnia.

Oceniając wykonane analizy w trzech grupach danych, zaobserwowano trend i sezonowość w ujęciu miesięcznym i dziennym w szeregu czasowym pierwotnym. Stało się to bezpośrednią przesłanką zbudowania modelu zerojedynkowego regresji wielorakiej w celu potwierdzenia postawionej hipotezy.

Rysunek 18

Kategoryzowany wykres normalności danych pierwotnych z podziałem na dni tygodnia w ujęciu dynamicznym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 6

Analiza testów Shapiro-Wilka — grupa: dni tygodnia

Dzień	Pvalue	Test SW-W
niedziela	0,80	0,00
poniedziałek	0,76	0,00
wtorek	0,75	0,00
środa	0,76	0,00
czwartek	0,75	0,00
piątek	0,74	0,00
sobota	0,77	0,00

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Zbudowano model złożony z dwudziestu jeden predyktorów, w tym zmiennych w postaci: t , t^2 , t^3 , $\ln t$, zmiennych zerojedynkowych dla miesięcy styczeń — listopad oraz dni tygodnia poniedziałek–piątek i niedziela. Model był macierzą złożoną z 19 643 elementów. Istotne predyktory zestawiono w tabeli 7 (tylko te poniżej przyjętego poziomu istotności).

Ocena tabeli 7 pozwala stwierdzić, że postawiona hipotez jest poprawna. W analizowanym szeregu cza-

sowym pierwotnym widoczny jest słaby trend malejący oraz sezonowość w ujęciu miesięcznym i dziennym.

Dalszym etapem badań była analiza i ocena reszt zbudowanego modelu zerojedynkowego regresji wielorakiej (rysunki 19–24).

Ocena rysunku 19 kaže stwierdzić dobre dopasowanie wartości przewidywanych i obserwowanych.

Reszty zbudowanego modelu regresji wielorakiej są zarówno dodatnie, jak i ujemne o długich wąsach, co może być przesłanką istnienia w nich zależności.

Tablica 7

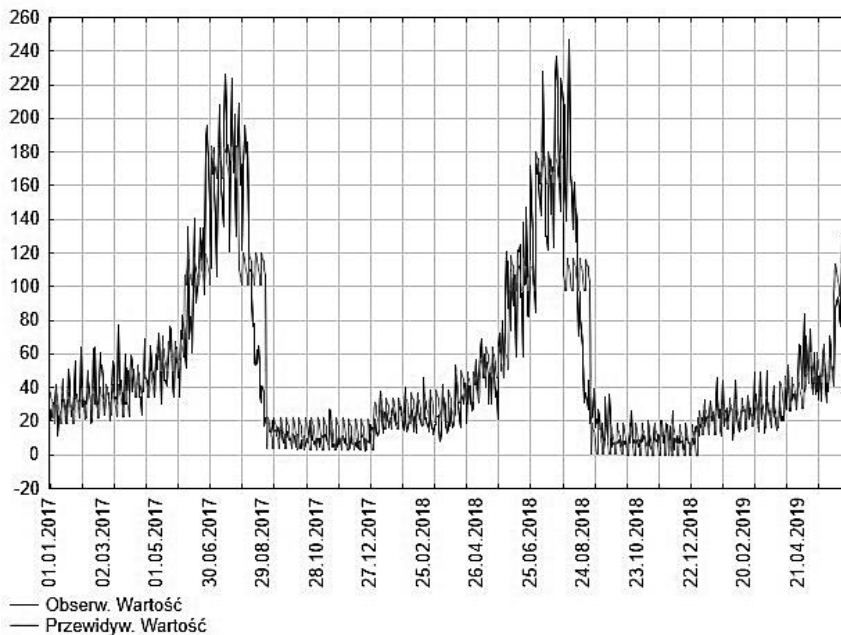
Model regresji wielorakiej

N = 893	R = ,90682877 R ² = ,82233842 Popraw. R2 = ,81950555 Błąd std. estymacji: 21,953					
	b*	Bł. std.	b	Bł. std.	t (878)	p
W. wolny			6,1345	2,297075	2,67058	0,007712
t	-0,045468	0,014421	-0,0091	0,002889	-3,15290	0,001671
styczeń	0,075532	0,015933	12,7707	2,693886	4,74062	0,000002
luty	0,091871	0,015748	16,2530	2,786004	5,83381	0,000000
marzec	0,105223	0,015846	17,7907	2,679141	6,64046	0,000000
kwiecień	0,170276	0,015786	29,2108	2,708163	10,78622	0,000000
maj	0,260480	0,015831	44,0410	2,676686	16,45354	0,000000
czerwiec	0,509188	0,015540	96,5849	2,947658	32,76667	0,000000
lipiec	0,789680	0,015446	160,4440	3,138236	51,12555	0,000000
sierpień	0,478342	0,015411	97,1876	3,131135	31,03911	0,000000
poniedziałek	0,103093	0,016144	15,1936	2,379247	6,38590	0,000000
wtorek	0,092522	0,016136	13,6803	2,385815	5,73403	0,000000
środa	0,062296	0,016144	9,1811	2,379319	3,85870	0,000122
czwartek	0,038734	0,016130	5,7272	2,384951	2,40140	0,016540
niedziela	0,125614	0,016141	18,5128	2,378759	7,78255	0,000000

Źródło: Opracowanie własne na podstawie

Rysunek 19

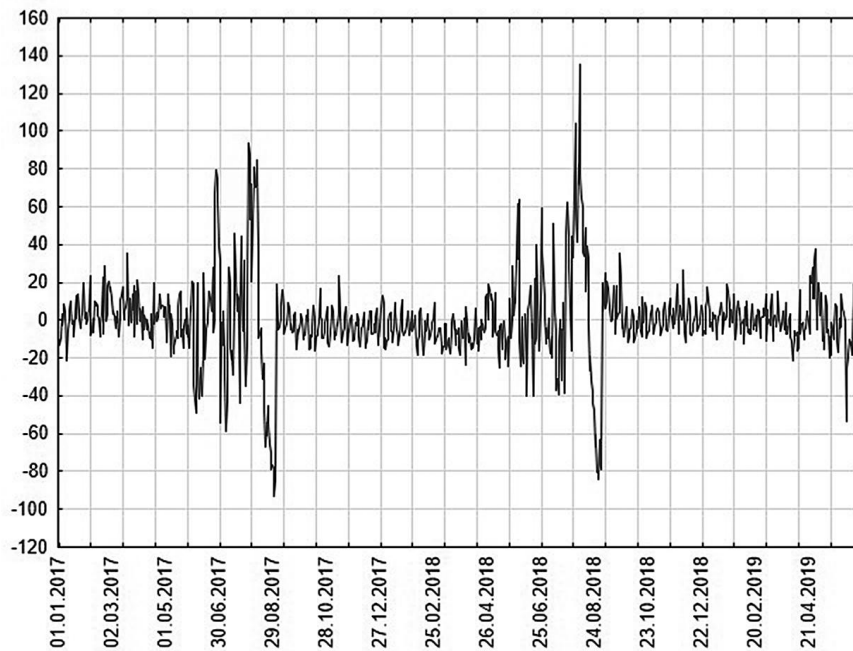
Wykres wartości przewidywanych i obserwowanych



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 20

Wykres liniowy reszt modelu regresji wielorakiej



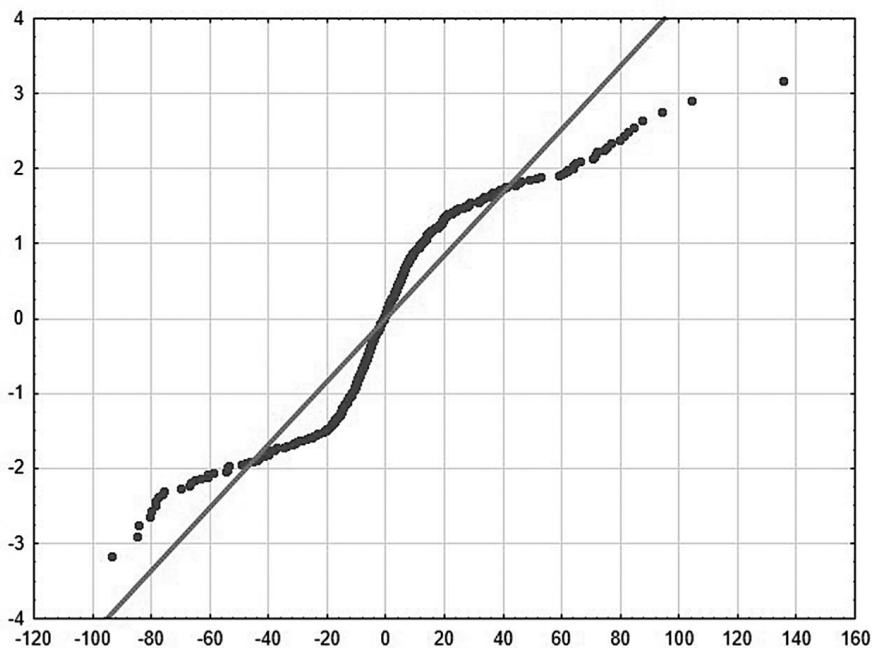
Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Kwantyle odstają od nakreślonej linii prostej. Histogram i test Shapiro-Wilka wskazują na brak rozkładu o charakterze normalnym.

Autokorelacja (rysunek 23) potwierdza istnienie zależności w resztach zbudowanego modelu regresji wielorakiej.

Rysunek 21

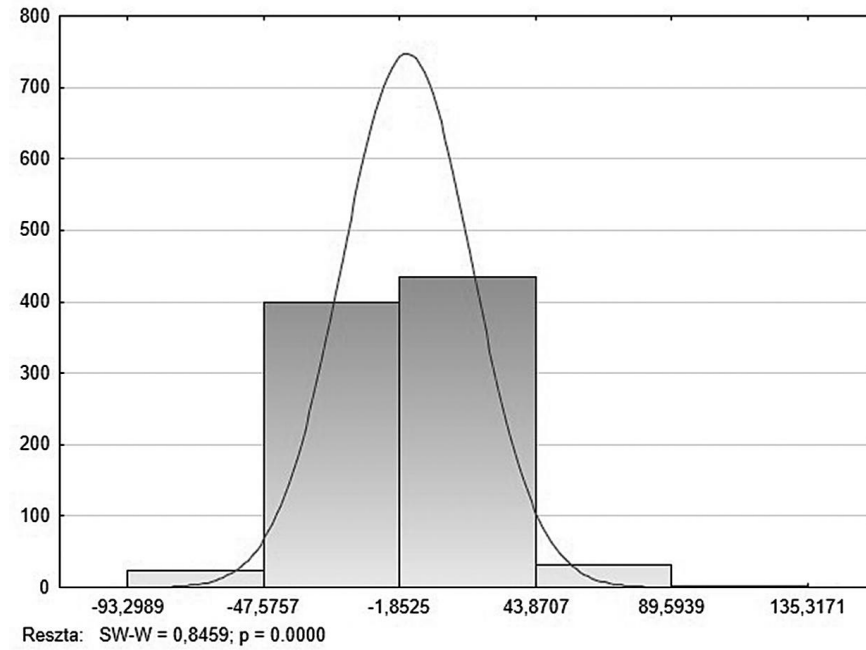
Wykres normalności reszt modelu regresji wielorakiej



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 22

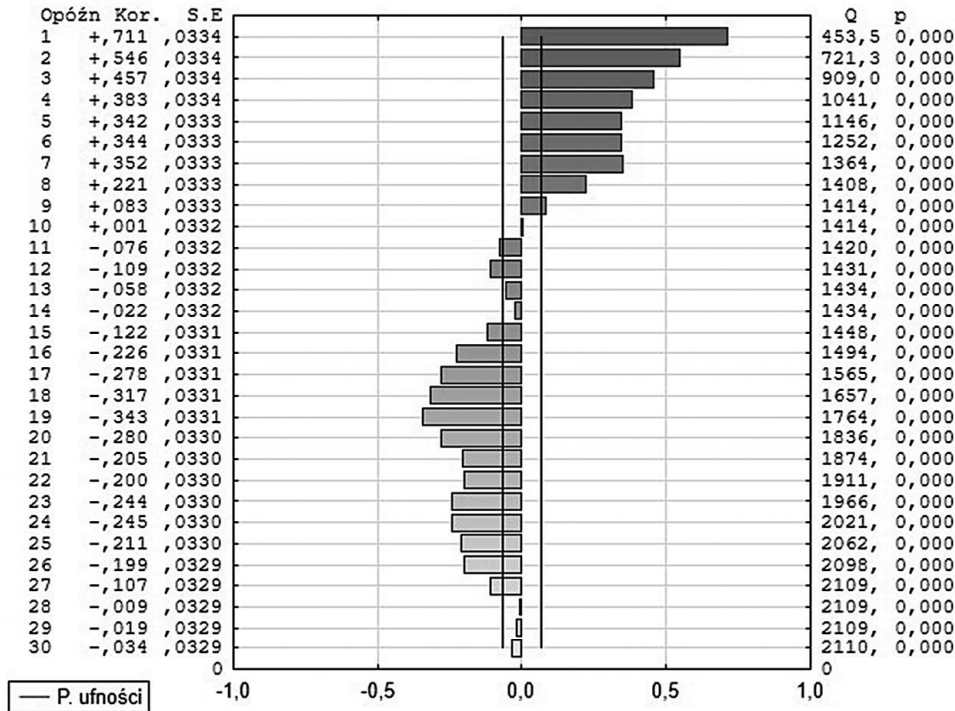
Histogram reszt modelu regresji wielorakiej



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 23

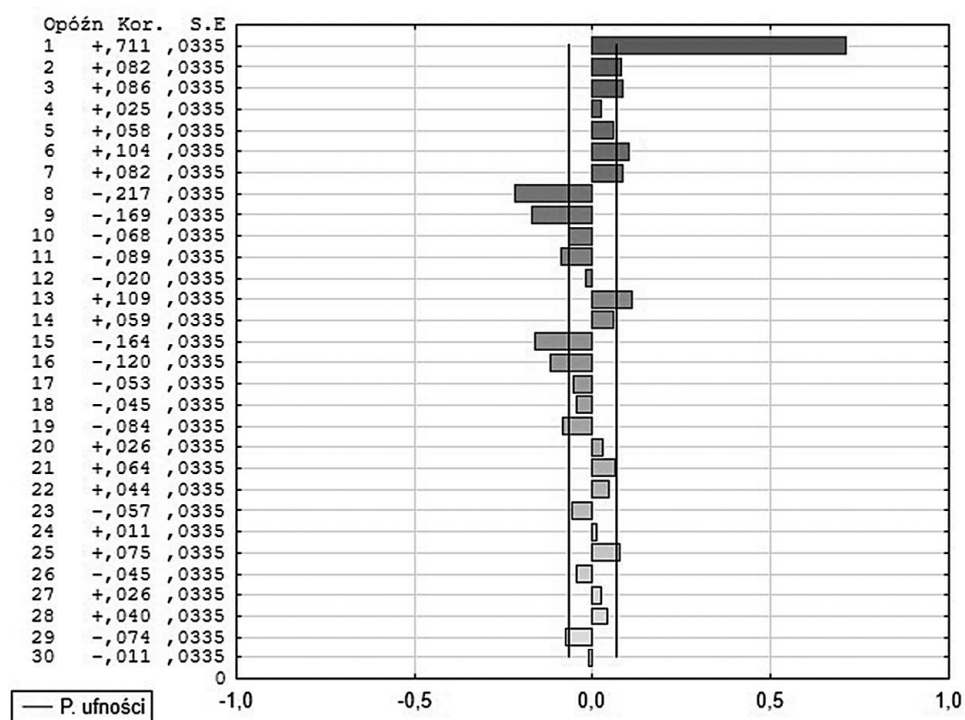
Autokorelacja reszt modelu regresji wielorakiej



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 24

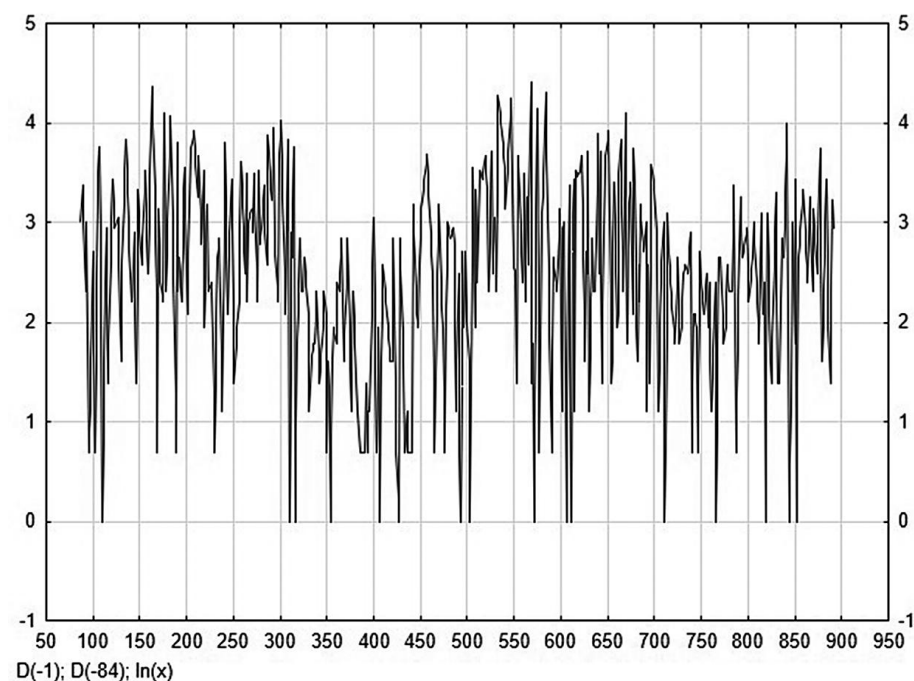
Autokorelacja cząstkowa reszt modelu regresji wielorakiej



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 25

Różnicowanie i logarytmowanie szeregu czasowego pierwotnego



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Autokorelacja cząstkowa również wskazuje na istnienie zależności w postaci widocznej sezonowości.

Wykonane oceny przeprowadzonych analiz stały się przesłanką sprowadzenia analizowanego szeregu czasowego pierwotnego do postaci stacjonarnej. W tym celu analizowany szereg pierwotny zróżnicowano na poziomie 1, następnie usuwając sezonowość na poziomie 84 i logarytmując (rysunek 25).

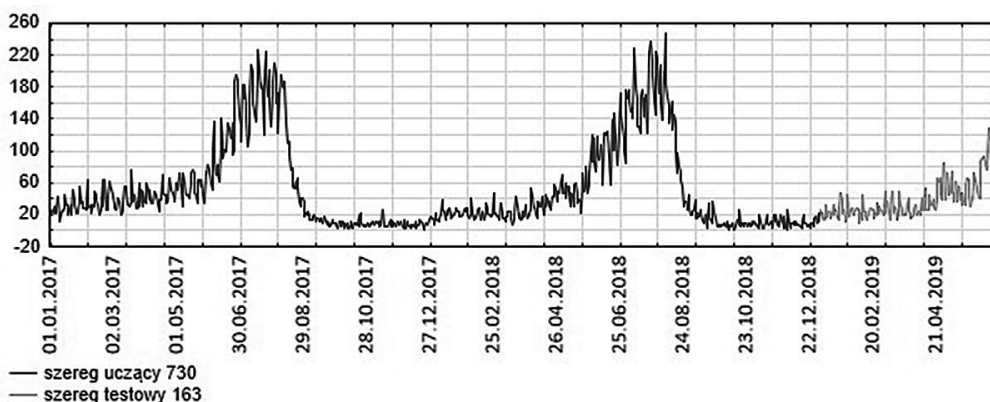
Obserwacja rysunku 25 prowadzi do stwierdzenia, że, że analizowany szereg czasowy pierwotny został sprowadzony do postaci stacjonarnej. Zabieg ten będzie miał duże znaczenie podczas budowy modelu typu SARIMA.

Analiza metod prognozowania i prognozowanie

Dalszym etapem badań w celu znalezienia najlepszej metody prognozy szeregu czasowego pierwotnego na przyszłość było jego podzielenie na dwie części: uczącą — 730 i testową 163 (rysunek 26).

Rysunek 26

Podział danych pierwotnych na część uczącą (730 elementów) i testową (163 elementy)



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Szereg czasowy uczący posłuży do budowy modeli prognozowania, a testowy do ocen wykonanych prognoz.

Wskutek krytycznej analizy literatury do prognozy analizowanych danych retrospektywnych na przyszłość (163 okresy) użyto trzech metod badawczych: metody wykładniczego Holta-Wintersa (rysunek 27), modelu SARIMA (rysunek 28) i metody naiwnej (rysunek 29).

Prognoza wykonana metodą Holta-Wintersa

(rysunek 27) nie jest zadowalająca. Nie wykazuje kontynuacji tendencji zaobserwowanych w przeszłości.

Podobnie zbudowany model SARIMA (rysunek 28) wykazuje słabą prognozę, która nie zachowuje zaobserwowanych serii z przeszłości.

Prognoza naiwna wydaje się najlepsza. Dla celów badawczych wykonane prognozy zestawiono z szeregiem czasowym testowym na rysunku 30.

Obserwacja wzrokowa rysunku 30 jednoznacznie wskazuje, że najlepszą metodą do prognozy szeregu czasowego pierwotnego jest metoda naiwna. Dla celów badawczych przeprowadzono analizę i ocenę MAPE (tabela 8).

Analiza przeprowadzona w tabeli 8 prowadzi do obserwacji, że, że najniższy średni absolutny błąd prognozy widoczny jest w metodzie naiwnej i wynosi 0,33%.

Stało się to przesłanką użycia metody naiwnej do prognozy szeregu czasowego pierwotnego na 202 przyszłe okresy. Wyniki przedstawiono na rysunku 31 i w tabeli 9.

Ocena prognozy przedstawionej na rysunku 31 także stwierdzić zachowanie zachowania tendencji z przeszłości. Szczegółowe wyniki prognozy przedstawia tabela 9.

Tabela 8

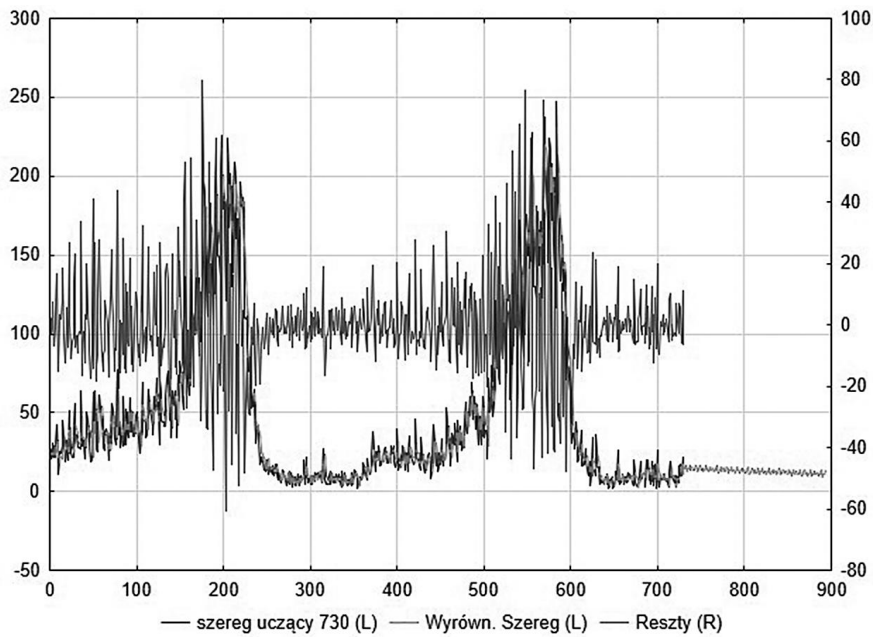
Analiza MAPE

	MAPE HW	MAPE SARIMA	MAPE NAIWNA
ŚREDNIA	0,54	0,67	0,33

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 27

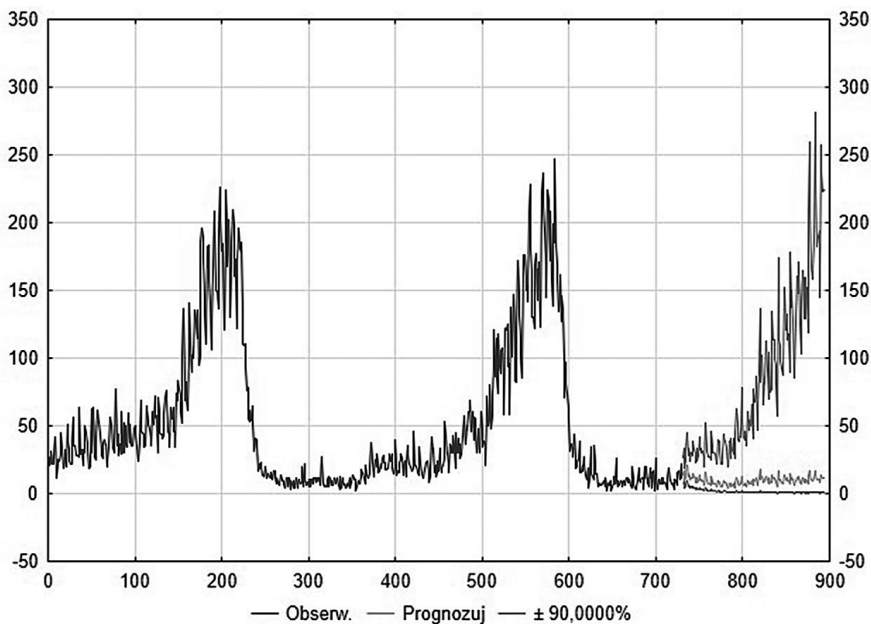
Prognozowanie szeregu uczącego metodą wygładzania wykładniczego
 Holta-Wintersa na 163 przyszłe okresy



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 28

Prognozowanie szeregu uczącego modelem SARIMA (6,1,0) (6,1,0)
 z opóźnieniem sezonowym 84 na 163 przyszłe okresy

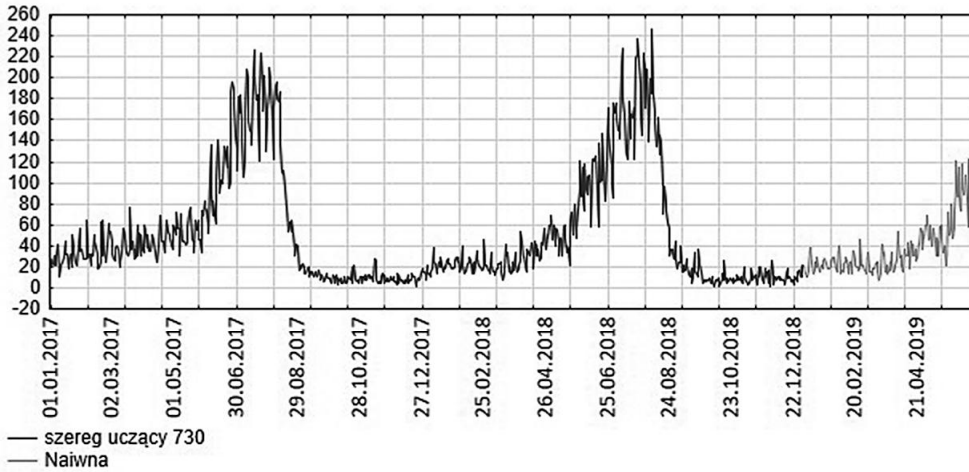


Prognoza; Model: (6,1,0)(6,1,0) Opóź. sezon.: 84

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 29

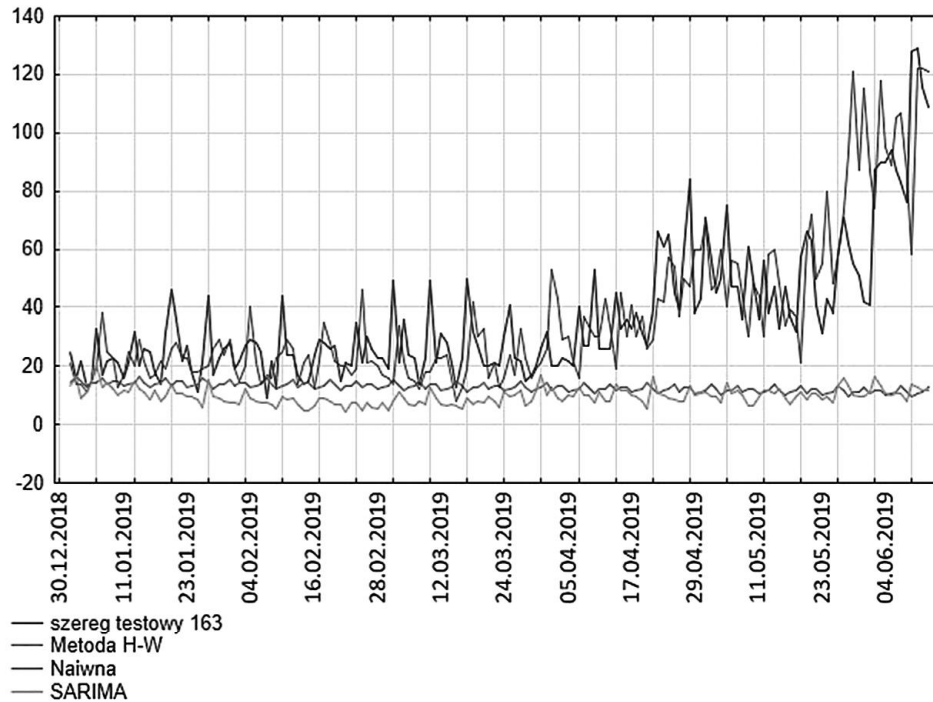
Prognozowanie szeregu uczącego metodą naiwną na 163 przyszłe okresy



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 30

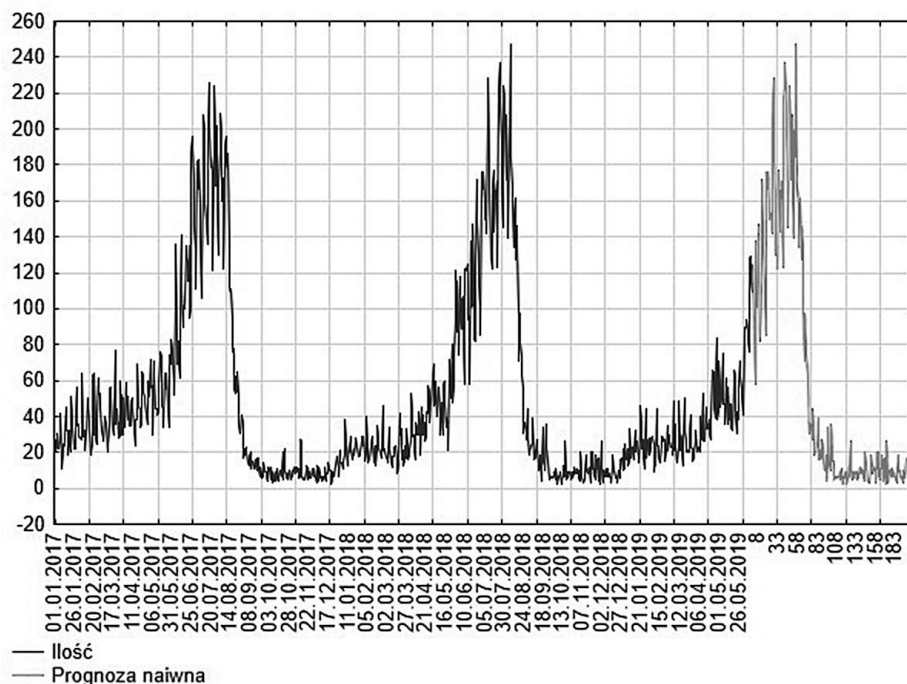
Zestawienie prognoz wykonanych metodami: SARIMA, Holta-Wintersa, Naiwną z szeregiem uczącym



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Rysunek 31

Prognozowanie szeregu pierwotnego modelem naiwnym na 202 przyszłe okresy (rok 2019)



Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Tabela 9

Prognoza z wykorzystaniem modelu SARIMA, Holta-Wintersa i oceny ekspertów danych pierwotnych na 202 przyszłe okresy (2019 rok)

czas	prognoza	czas	prognoza	czas	prognoza	czas	prognoza	czas	prognoza
894	125	936	215	978	16	1020	9	1062	17
895	94	937	195	979	16	1021	5	1063	6
896	96	938	169	980	21	1022	9	1064	2
897	58	939	145	981	17	1023	6	1065	26
898	138	940	224	982	27	1024	9	1066	12
899	101	941	217	983	22	1025	9	1067	3
900	102	942	208	984	14	1026	8	1068	10
901	147	943	172	985	10	1027	5	1069	9
902	115	944	208	986	12	1028	9	1070	7
903	83	945	159	987	19	1029	6	1071	11
904	82	946	139	988	4	1030	8	1072	7
905	122	947	199	989	12	1031	12	1073	11
906	172	948	185	990	34	1032	8	1074	7
907	148	949	247	991	15	1033	11	1075	8
908	131	950	184	992	13	1034	8	1076	7

Cd. tabeli 9

czas	prognoza	czas	prognoza	czas	prognoza	czas	prognoza	czas	prognoza
909	116	951	168	993	10	1035	10	1077	6
910	101	952	158	994	36	1036	6	1078	3
911	85	953	134	995	23	1037	4	1079	12
912	176	954	150	996	15	1038	20	1080	19
913	167	955	162	997	15	1039	11	1081	9
914	176	956	127	998	5	1040	6	1082	10
915	161	957	146	999	7	1041	4	1083	7
916	150	958	137	1000	6	1042	3	1084	6
917	153	959	97	1001	7	1043	5	1085	7
918	142	960	71	1002	6	1044	9	1086	3
919	207	961	97	1003	7	1045	19	1087	10
920	228	962	81	1004	7	1046	10	1088	9
921	181	963	74	1005	9	1047	8	1089	6
922	165	964	63	1006	5	1048	10	1090	16
923	130	965	56	1007	11	1049	12	1091	16
924	130	966	31	1008	6	1050	11	1092	17
925	122	967	37	1009	2	1051	3	1093	10
926	169	968	36	1010	7	1052	20	1094	22
927	177	969	33	1011	6	1053	11	1095	11
928	143	970	27	1012	8	1054	20		
929	166	971	44	1013	9	1055	7		
930	162	972	28	1014	2	1056	11		
931	171	973	19	1015	5	1057	5		
932	123	974	20	1016	7	1058	19		
933	218	975	24	1017	13	1059	11		
934	222	976	25	1018	7	1060	16		
935	237	977	39	1019	26	1061	4		

Źródło: opracowanie własne na podstawie danych pierwotnych uzyskanych z przedsiębiorstwa turystycznego.

Ocena prognoz przedstawionych w tabeli 9 prowadzi do stwierdzenia, że zostały one wykonane na 202 przyszłe okresy, w których średnia arytmetyczna wyniosła 60,89, a odchylenie standardowe od średniej arytmetycznej 70,13. Największa wartość w 202 elementach prognozy to 247, a najmniejsza 2. Mediana była na poziomie 17. Rozkład był prawostronnie asymetryczny, gdyż skośność wyniosła 1. Co więcej, rozkład był bardziej spłaszczony od normalnego. Dolny kwartył prognoz naiwnych był na poziomie 8, a górny wyniósł 125. W uzyskanych prognozach brak jest wartości odstających i ekstremalnych.

Podsumowanie i wnioski

Cel artykułu został osiągnięty. Przeprowadzono analizę i ocenę szeregu czasowego pierwotnego. Na podstawie uzyskanych ocen i krytycznej literatury wybrano metody prognozowania, które poddano analizie i ocenie. Najlepszą metodą wykonano prognozowanie szeregu czasowego pierwotnego na 2019 r.

Ważnym aspektem poruszonym w artykule jest przedstawiona metodyka wyboru metod prognozowania poprzez poprawnie przeprowadzoną analizę szeregu czasowego pierwotnego. Ważną kwestią sta-

je się zastosowanie właściwego narzędzia badawczego poprzez ciągle przeprowadzaną obserwację wzrokową i ocenę. Oceny są przesłankami stosowania kolejnych analiz i ocen z wykorzystaniem różnych narzędzi badawczych. Ostatnim etapem jest dobór metod prognozowania na podstawie zaobserwowanych zależności i krytycznej analizy literatury.

Uzyskane informacje z przedstawionej metodyki wyboru metody prognozowania wizyt na stro-

nach internetowych są niezwykle istotne z punktu widzenia szeroko rozumianej turystyki i jej rozwoju. Uzyskane poprawne wyniki prognozy pozwalają podjąć wiele decyzji związanych z planowaniem dotyczącym wydatkowania środków niezbędnych do przeprowadzenia realizowanych w przyszłości przedsięwzięć w całym procesie związanym z pozyskiwaniem oraz obsługą turystów w ujęciu dynamicznym.

Bibliografia

- Dittmann, P. (2016). *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*. Warszawa: Wolters Kluwer.
- Dittmann, P., Szabelska-Pasierbińska, E., Dittmann, I., Szpulak, A. (2016). *Prognozowanie w zarządzaniu przedsiębiorstwem*. Warszawa: Wolters Kluwer.
- Dittmann, P., Szabelska-Pasierbińska, E., Dittmann, I., Szpulak, A. (2016). *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*. Warszawa: Wolters Kluwer.
- Gołembski, G. (1997). *Przedsiębiorstwo turystyczne w gospodarce wolnorynkowej*. Poznań: Wydawnictwo Akademii Ekonomicznej.
- Kot, S. M., Jakubowski, J., Sokołowski, A. (2011). *Statystyka*. Warszawa: Difin.
- Kozicki, B. (2018). Metodyka prognozowania zysku. *Systemy Logistyczne Wojsk*, (49), 138–157.
- Kozicki, B., Waściński, T., Brzeziński, M., Lisowska, A. (2018). Cost forecast in a shipping company. *Transport Means*, (1), 1235–1241.
- Kozicki, B., Waściński, T., Brzeziński, M., Tomaszewski, J. (2018). Zastosowanie prognozy do planowania przychodów przedsiębiorstwa. *Gospodarka Materialowa i Logistyka*, (5), 332–343.
- Kozicki, B., Waściński, T., Lisowska, A. (2018). Selection of optimal forecasting method for a CPI inflation measure in Poland. W: E. Skrzypek (red.), *Zarządzanie organizacją w warunkach różnorodności* (565–574). Lublin: Katedra Zarządzania Jakością i Wiedzą. Wydział Ekonomiczny. Uniwersytet Marii Curie-Skłodowskiej.
- Łuniewska, M., Tarczyński, W. (2006). *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*. Warszawa: PWN.
- Tomaszewski, J. (2018). Prognozowanie przyjęć do pracy jako element budowy strategii personalnej. *Systemy Logistyczne Wojsk*, (49), 206–218.



Społeczność wokół marki wzbudza zainteresowanie podmiotów rynkowych. Geneza powstania tej formy społeczności sięga czasów, kiedy to lokalnie funkcjonujące społeczności inspirowały się wzajemnie do podejmowania decyzji zakupowych. Dziś członkowie grupy gromadzącej się wokół marki dzielą się między sobą informacjami o marce, inspirowały się jej symboliką, przesłaniem, misją. Kreuje to w efekcie lojalność klientów do marki.

W prezentowanej książce została opisana społeczność wokół marki, mechanizm jej powstawania i korzyści, jakie z niej czerpią poszczególne podmioty rynkowe. Na podstawie badań przeprowadzonych wśród fanów Legii Warszawa pokazane zostały niuanse kreowania lojalności wobec marki w odniesieniu do rynku sportowego.

Adresatami publikacji są wszyscy czytelnicy zainteresowani tematyką marki, zwłaszcza menedżerowie oraz pracownicy uczestniczący w budowaniu strategii marki, osoby z agencji reklamowych, badawczych oraz studenci marketingu uczelni ekonomicznych.

www.pwe.com.pl