



**Bartosz Gimzicki**

Prognozowanie liczby rowerów miejskich dla  
stacji Brama Poznania w Poznaniu

Forecasting the number of city bikes for the  
Brama Poznania station

Praca licencjacka

Promotor: dr Maciej Beręsewicz

Pracę przyjęto dnia:

Podpis promotora

Kierunek: Informatyka i ekonometria

Specjalność: Analityka gospodarcza

Poznań 2019



# Spis treści

|   |           |
|---|-----------|
| <b>Wstęp</b>  | <b>1</b>  |
| <b>1 Poznański Rower Miejski</b>                              | <b>2</b>  |
| 1.1 Smart Cities . . . . .                                    | 2         |
| 1.1.1 Rozwój miast i Open Data . . . . .                      | 2         |
| 1.1.2 Definicje Smart Cities . . . . .                        | 3         |
| 1.1.3 Klasyfikacja Smart Cities . . . . .                     | 4         |
| 1.1.4 Smart Cities w Polsce . . . . .                         | 4         |
| 1.1.5 Ekonomia współdzielenia . . . . .                       | 5         |
| 1.2 Miejskie systemy wypożyczania rowerów . . . . .           | 6         |
| 1.2.1 Historia koncepcji rowerów miejskich . . . . .          | 6         |
| 1.2.2 Charakterystyka systemu rowerów miejskich . . . . .     | 6         |
| 1.2.3 Rowery miejskie w różnych miastach na świecie . . . . . | 7         |
| 1.2.4 Zalety rowerów miejskich . . . . .                      | 8         |
| 1.2.5 Wady rowerów miejskich . . . . .                        | 8         |
| 1.3 Rowery w Poznaniu . . . . .                               | 9         |
| 1.3.1 Infrastruktura rowerowa w Poznaniu . . . . .            | 9         |
| 1.3.2 Poznański Rower Miejski . . . . .                       | 9         |
| 1.4 Podsumowanie . . . . .                                    | 10        |
| <b>2 Prognozowanie szeregów czasowych</b>                     | <b>11</b> |
| 2.1 Pojęcie prognozy . . . . .                                | 11        |
| 2.2 Rodzaje prognoz . . . . .                                 | 12        |
| 2.3 Prognozowanie szeregów czasowych . . . . .                | 14        |
| 2.3.1 Modele szeregów czasowych . . . . .                     | 15        |

|          |   |           |
|----------|---|-----------|
| 2.3.2    | Podział danych na część uczącą i testową . . . . .                  | 16        |
| 2.4      | Analiza szeregu czasowego . . . . .                                 | 16        |
| 2.4.1    | Funkcja autokorelacji ACF i autokorelacji cząstkowej PACF . . . . . | 16        |
| 2.4.2    | Różnicowanie . . . . .  | 17        |
| 2.5      | Metody prognozowania . . . . .                                      | 18        |
| 2.5.1    | Metoda naiwna . . . . .   | 18        |
| 2.5.2    | Metoda średniej ruchomej . . . . .                                  | 18        |
| 2.5.3    | Modele ARMA i ARIMA . . . . .                                       | 19        |
| 2.5.4    | Analiza reszt . . . . .   | 22        |
| 2.5.5    | Kryteria oceny modelu . . . . .                                     | 23        |
| 2.5.6    | Kryteria oceny prognozy . . . . .                                   | 24        |
| 2.6      | Podsumowanie . . . . .  | 25        |
| <b>3</b> | <b>Konstrukcja prognozy dla stacji Brama Poznania</b>               | <b>26</b> |
| 3.1      | Pobieranie i analiza danych . . . . .                               | 26        |
| 3.1.1    | Eksploracyjna analiza danych . . . . .                              | 27        |
| 3.2      | Modele ARIMA . . . . .  | 29        |
| 3.2.1    | Model ARIMA(20,1,0) . . . . .                                       | 30        |
| 3.2.2    | Model ARIMA(2,1,2) . . . . .  | 31        |
| 3.2.3    | Diagnoza dopasowania modelu . . . . .                               | 32        |
| 3.3      | Prognoza . . . . .  | 32        |
| 3.3.1    | Wybór odpowiedniego modelu . . . . .                                | 33        |
| 3.3.2    | Prognoza dla stacji Brama Poznania . . . . .                        | 34        |
|          | <b>Zakończenie</b>  | <b>35</b> |
|          | <b>Biografia</b>  | <b>37</b> |
|          | <b>Spis Tablic</b>  | <b>38</b> |
|          | <b>Spis Rysunków</b>  | <b>39</b> |
|          | <b>Spis Programów</b>   | <b>40</b> |

# Wstęp

Technologie Informacyjno-Komunikacyjne znacząco wpłynęły na rozwój miast w XXI wieku. Dzięki postępowi technologicznemu, który ułatwił gromadzenie i przesyłanie informacji, a także komunikację pojawiło się wiele nowych dziedzin gospodarki takich jak np. ekonomia współdzielenia. Nowym paradygmatem w rozwoju miast jest dążenie do bycia miastem inteligentnym, które zwiększa interaktywność oraz wydajność swojej infrastruktury i jej składowych.

W polskich miastach coraz częściej można zauważyć elementy Smart Cities. Do jednego z podstawowych elementów można zaliczyć rozbudowany system rowerów miejskich. Ma on na celu usprawnienie poruszania się na obszarze danej miejscowości, a także promowanie zdrowego trybu życia pośród jej mieszkańców.

Celem tej pracy jest zbudowanie odpowiedniego modelu prognostycznego dla stacji Brama Poznania, który usprawni zarządzanie siecią rowerów miejskich i pozwoli przewidzieć ich zapotrzebowanie wśród mieszkańców Poznania.

Dane zebrane zostały z oficjalnego API firmy NextBike udostępnionego publicznie na jej stronie internetowej. Zbudowane zostaną odpowiednie modele prognostyczne, zbadana zostanie jakość ich dopasowania, następnie wybrany zostanie dokładniejszy z nich i na jego podstawie skonstruowana zostanie prognoza na następne 12 godzin.

Praca składa się z trzech rozdziałów. W pierwszym z nich omówione zostaną zjawiska związane z pojawieniem się sieci rowerów miejskich w wielu miastach na świecie oraz ich wpływ na mieszkańców. Przedstawiona zostanie również historia tych systemów oraz przedstawiony sposób jego działania. W drugim rozdziale poruszony zostanie temat konstrukcji prognoz na podstawie szeregów czasowych oraz ich analizy. Ostatni rozdział ma charakter aplikacyjny, w którym przedstawiono zastosowane narzędzia oraz wybrane modele ekonometryczne. Pracę kończy dyskusja uzyskanych wyników oraz podsumowanie.

# Rozdział 1

## Poznański Rower Miejski

### 1.1 Smart Cities

#### 1.1.1 Rozwój miast i Open Data

Miasta w dwudziestym wieku przeszły gwałtowny rozwój, a koncepcje urbanistyczne zmieniały się wielokrotnie. Na początku stulecia społeczeństwa za główny cel postawiły sobie poprawę warunków życiowych w tym między innymi powiększenie powierzchni mieszkaniowej na osobę czy szerszy dostęp do urządzeń sanitarnych. Kierunek ten nazwano modernizmem i funkcjonował on na świecie aż do lat siedemdziesiątych. Po 1945 roku urbanistyka europejska stanęła przed nowym wyzwaniem - odbudowaniu miast zniszczonych w wyniku drugiej wojny światowej. Rozwój motoryzacji w latach 50-tych i 60-tych wymógł przystosowanie ich do tego środka transportu - obwodnice, nowe drogi i parkingi. Rozpoczęto również budowę blokowisk, które miały poprawić warunki mieszkaniowe mieszkańców.

Dzięki gwałtownemu rozwojowi technologicznemu oraz gospodarce opartej na wiedzy powstał nowy paradygmat w planowaniu miast. Polega on na wykorzystaniu technologii informacyjnych i komunikacyjnych (z angielskiego *ICT – information and communication technologies*) do tworzenia bardziej wydajnych i innowacyjnych aglomeracji. Pojęcie to oznacza techniki służące do gromadzenia, przetwarzania i wysyłania danych w formie elektronicznej.

Obecnie wiele miast, w tym również w Polsce, prowadzi portale z otwartymi danymi (ang. *Open Data*), do swobodnego wykorzystania przez każdego, również w celach komercyjnych. Ma to na celu zwiększenie zaangażowania społeczności w życie miasta, inspirowanie przedsiębiorców i organizacje pozarządowe np. do tworzenia nowych usług czy wykazanie transpa-

rentności działań władz. Aby można było mówić o otwartych danych muszą one spełniać szereg właściwości takich jak na przykład: swobodny, stały dostęp bez żadnych barier, otwarty format, udostępniane w postaci możliwie najmniej przetworzonej, być aktualne.

### 1.1.2 Definicje Smart Cities

Śledząc dyskusje na temat współczesnego rozwoju miast często możemy natrafić na sformułowania takie jak Smart City, Inteligentne Miasto, miasta uczące się, oparte na wiedzy.

Definicja Smart City została przedstawiona między innymi w raporcie Wiedeńskiego Uniwersytetu Technologicznego(VUT) „*Smart cities - ranking of European medium-sized cities*” z 2007 roku. Autorzy przytaczają różne ujęcia tego pojęcia w literaturze sprzed roku jego publikacji. Są to np. „Inteligentny” przemysł, w szczególności w dziedzinie technik informacyjnych i komunikacyjnych lub taki wymagający ich w procesie produkcyjnym. Innym ujęciem tego zagadnienia dotyczy podnoszenia poziomu wykształcenia mieszkańców miast. W innej literaturze termin ten odnosił się do relacji między administracją i obywatelami, w szczególności do wykorzystania nowych kanałów komunikacji między nimi (elektroniczna administracja). Jeszcze inne spojrzenie dotyczy nie tyle technik ICT, co wykorzystania nowoczesnych technologii transportowych ułatwiających poruszanie się po mieście. Istnieje również wiele innych aspektów odnoszących się do Smart City, takie jak bezpieczeństwo, ekologiczność, wydajny i zrównoważony rozwój. Patrząc na zróżnicowanie odniesień do tego zagadnienia autorzy raportu wysunęli następującą definicję:

*Smart City to miasto, które uzyskuje dobre wyniki i patrzy w przyszłość odnośnie sześciu obszarów, stworzone dzięki inteligentnemu połączeniu zasobów i aktywności niezależnych i świadomych obywateli.*

Sześć kategorii wymienionych w powyższej definicji to: 1) gospodarka, 2) ludzie, 3) zarządzanie, 4) mobilność, 5) środowisko naturalne oraz 6) jakość życia.

Inną definicję przedstawiają T. Bakici, E. Almirall i J. Wareham w swojej pracy odnośnie studium przypadku Barcelony. Zdaniem autorów są to miasta, które Swoją strategię opierają na wykorzystaniu technik informacyjnych i komunikacyjnych na wielu płaszczyznach takich jak: ekonomia, środowisko, mobilność i zarządzanie w celu zmiany infrastruktury i usług”. Wskazują oni na znaczącą rolę ICT w przypadku koncepcji Miasta Inteligentnego. Zdaniem autorów Barcelona może być uważana za modelowe miasto wykorzystujące te technologie (Bakici, Almirall & Wareham, 2012).

Komisja Europejska definiuje je jako miasta, w których tradycyjne sieci i usługi, dzięki wykorzystaniu technologii cyfrowych i teleinformacyjnych są bardziej wydajne z korzyścią dla mieszkańców i biznesu.

### 1.1.3 Klasyfikacja Smart Cities

Według Boyda Cohena wyróżniamy trzy kategorie Smart Cities według ich rozwoju (Cohen, 2015):

- *Smart Cities 1.0* – charakteryzuje je to, że twórcy technologii zachęcają administrację do wprowadzenia ich rozwiązań w celu zwiększenia efektywności w zarządzaniu miastem. Problem jest jednak taki, że miasta te nie są odpowiednio przygotowane do ich odpowiedniego użycia oraz nie rozumieją, jak mogą one wpłynąć na jakość życia mieszkańców. Gubią one również znaczącą kwestię, jaką jest interakcja miasta z ich obywatelami.
- *Smart Cities 2.0* – charakteryzują się tym, że to administracja publiczna odgrywa kluczową rolę we wdrażaniu programów, które mają zwiększyć jakość życia mieszkańców. Miasta te cechują się dużą liczbą projektów, których celem jest wdrażanie nowych technologii w różnych dziedzinach życia.
- *Smart Cities 3.0* – charakteryzują się tym, że to obywatele mają ważną rolę w rozwoju miasta.

### 1.1.4 Smart Cities w Polsce

Dr Dorota Sikora-Fernandez z Katedry Zarządzania Miastem i Regionem Wydziału Zarządzania Uniwersytetu Łódzkiego mówi, że w przypadku Polski możemy raczej mówić o przejawach pewnych elementów *Smart Cities*, jako realizacji współczesnych standardów. Zwraca uwagę, że w literaturze mówi się o inteligentnych miastach w kontekście prospołecznym, a same nowe technologie to za mało (Wroński, 2018).

Przykładami praktyk Smart Cities w Polsce mogą być:

1. Systemy zarządzania transportem publicznym, ruchem ulicznym i parkingami;
2. Zarządzanie systemami energii, ciepła i oświetlenia ulicznego;
3. Dostępność usług online, informatyzacja miasta, otwarte dane;



4. Systemy pozwalające na dokonywanie obywatelskich interwencji w zakresie awarii bądź naruszenia; porządku publicznego, będące platformą komunikacji i wymiany informacji z mieszkańcami;
5. Rozwiązania typu Living Lab, które zakładają udział użytkowników rozwiązań w ich tworzeniu.

### **1.1.5 Ekonomia współdzielenia**

Ważnym pojęciem w dziedzinie rozwoju miast w epoce XXI wieku jest ekonomia współdzielenia (z ang. *sharing economy*). Jest to zjawisko ekonomiczne i społeczne, a przy jego rozwoju duży udział mają technologie informacyjne, które są wykorzystywane do kojarzenia ze sobą użytkowników chcących dzielić się dobrami i usługami. Zakłada ona wspólne korzystanie z danych dóbr bez posiadania ich na własność. Zyskuje ona na coraz bardziej na popularności dzięki globalizacji i tworzenia różnego rodzaju platform kojarzących ze sobą użytkowników, ułatwiając komunikację między nimi.

Popularnymi przykładami z dziedziny ekonomii współdzielenia mogą być:

- Internetowy portal AirBnB, który pozwala dzielić się zakwaterowaniem z osobami zainteresowanymi. Powstał on w 2008 roku. W przeciągu 10 lat skorzystało z niego ponad 300 milionów osób. Oferuje on zakwaterowania w ponad 81 tys. miast na całym świecie, a łatwość w obsłudze tego portalu czyni go niezwykle atrakcyjnym zarówno dla osób, które wynajmują i które chcą wynająć mieszkanie lub pokój na krótki czas.
- W wielu miastach w Polsce zaczęły pojawiać się tzw. auta na minuty (z ang. *car sharing*). Samochody te stoją w różnych częściach miasta, bez problemu możemy je wypożyczyć, przejechać w dowolne miejsce i zostawić je dla następnych użytkowników, płacąc za przejechany dystans. Ta forma transportu jest coraz bardziej popularna w polskich miastach. W samym tylko Poznaniu działają przynajmniej cztery takie wypożyczalnie – 4mobility, Traficar, Easyshare oraz Click2Go.
- Internetowy kantor Walutomat pozwala na wymianę walut kojarząc ze sobą użytkowników, chcących daną walutę sprzedać lub kupić. Pozwala to na przeprowadzenie transakcji po cenach innych niż z góry ustalony kurs. Osoba zainteresowana wystawia swoją ofertę, a system wynajduje użytkownika zainteresowanego daną transakcją.

- W wielu miastach w Polsce i na świecie istnieje system rowerów miejskich. Użytkownicy przez specjalnie przygotowaną aplikację mogą wypożyczyć rower z jednej z dużej liczby stacji i przemieszczać się nim swobodnie. Znajdują się one w różnych częściach miasta, co czyni je bardzo wygodnym i atrakcyjnym środkiem transportu. Chcąc zakończyć wypożyczenie wystarczy ponownie wpiąć rower do dowolnej stacji lub, w przypadku niektórych miast, zostawić go w dowolnym miejscu.

## **1.2 Miejskie systemy wypożyczania rowerów**

### **1.2.1 Historia koncepcji rowerów miejskich**

Pierwsze próby wprowadzenia systemu wypożyczania rowerów na małą skalę w Europie sięgają lat 60-tych ubiegłego wieku. W 1965 roku w Amsterdamie organizacja o nazwie *Provos* postanowiła stworzyć system o nazwie "White Bike Plan", który miał pomóc zlikwidować problemy komunikacyjne w mieście. W tym celu użyto 50 starych rowerów, które następnie pomalowano na kolor biały i pozostawiono niezabezpieczone w różnych częściach miasta, a każdy mieszkaniec mógł z nich swobodnie korzystać. Jednak pojazdy te były często uszkodzane i kradzione, dlatego szybko z niego zrezygnowano.

Następne próby wprowadzenia rowerów miejskich, tym razem z odpowiednim zabezpieczeniem sięgają lat 90-tych, gdzie w stolicy Danii, Kopenhadze, stworzono system który uwalniał rower ze stacji dopiero po wrzuceniu monety, która była zwracana wraz z oddaniem go. System ten działa skutecznie do dzisiejszego dnia, obsługuje ponad 2000 pojazdów (Shaheen, Guzman & Zhang, 2010).

Jednak prawdziwa moda na rowery miejskie rozpoczęła się dopiero w 2007 roku, kiedy to postanowiono uruchomić taki system w Paryżu. Składał się on z ponad 10 tys. rowerów ulokowanych na 750 stacjach, a za stolicą Francji postanowiły podążać również inne miasta.

### **1.2.2 Charakterystyka systemu rowerów miejskich**

Jednym z elementów Smart Cities, który występuje również w Polsce, jest program rowerów miejskich. Z roku na rok coraz więcej miast decyduje się na uruchomienie systemu ich wypożyczania, jednak czekają na nie nowe wyzwania. Istotnym zagadnieniem w tej kwestii jest odpowiednie skomunikowanie wszystkich stacji z istniejącą siecią transportu publicznego, co

znacząco ułatwia poruszanie się po mieście i uatrakcyjnia ten sposób komunikacji dla użytkowników (Łastowska & Bryniarska, 2015). Wymaga to odpowiedniego zagospodarowania przestrzeni publicznej, co powinno być uwzględnione już na etapie planowania odpowiedniej strategii transportu publicznego w mieście (Dębowska-Mróż, Lis, Szymanek & Zawisza, 2017).

Samo wypożyczenie roweru nie jest skomplikowanym procesem. Użytkownik powinien posiadać odpowiednie konto, zależne od operatora rowerów w danym mieście i z poziomu aplikacji wypożyczyć urządzenie. Ma on możliwość pozostawienia go na dowolnej stacji poprzez zadokowanie go lub przypadku niektórych miast w dowolnym miejscu. W tym momencie automatycznie naliczana jest opłata w zależności od długości trasy i cennika operatora. W celu zachęcenia do skorzystania z usługi pierwsze minuty jazdy najczęściej są bezpłatne. Aby uatrakcyjnić ofertę występuję również wiele modeli rowerów: z fotelikiem, dziecięce, elektryczne czy wieloosobowe tandemy.

### **1.2.3 Rowery miejskie w różnych miastach na świecie**

Systemy rowerów miejskich znajdują się obecnie w ponad 160 krajach całego świata. Miasta coraz chętniej wdrażają je, widząc zalety z jego posiadania, korzystając z bogatej oferty operatorów, którzy tym się zajmują. Przykładami miejsc, w których taki system istnieje mogą być:

- Paryż, który zapoczątkował w Europie modę na posiadanie rowerów miejskich. Obecnie liczy on ponad 24 tys rowerów rozlokowanych na 1400 stacjach czynnych całą dobę.
- Drugim co do ilości rowerów miejskich miastem w Europie jest Londyn. Ta forma transportu funkcjonuje w nim od 2010 roku. System liczy prawie 12 tys rowerów i 750 stacji.
- Kolejnym z Europejskich miast posiadającym rowery miejskie jest Barcelona, która jest uważana za jeden z głównych przykładów w dziedzinie Smart Cities, wyróżniona w 2014 roku przez Komisję Europejską nagrodą *iCapital* za wykorzystywanie nowych technologii w celu poprawy życia mieszkańców. Posiada ona 6300 pojazdów oraz 424 stacje na obszarze całego miasta.
- W Ameryce największy system rowerów miejskich znajduje się w Nowym Jorku i liczy on sobie 12 tys pojazdów rozlokowanych na 750 stacjach.
- Największe systemy rowerów miejskich na świecie obecnie znajdują się w Chińskich miastach Wuhan i Hangzhou, licząc odpowiednio 90 tys i 60 tys rowerów.

- W Polsce miastem, które może poszczycić się największą liczbą rowerów miejskich jest Warszawa. W stolicy łącznie na 380 stacjach znajduje się ponad 5500 pojazdów (31 rowerów na 10 tys mieszkańców). Drugim co do wielkości tego typu system w Polsce znajduje się w aglomeracji poznańskiej (30,6 rowerów na 10 tys mieszkańców).

#### 1.2.4 Zalety rowerów miejskich

Rowery miejskie są ważną częścią transportu komunikacyjnego, stanowiąc jego istotne uzupełnienie (Łastowska & Bryniarska, 2015). Do zalet dla mieszkańców i miasta, które zdecydowało się na takie rozwiązanie można zaliczyć:

- Możliwość szybkiego i wygodnego poruszania się po mieście, a w szczególności jego centrum. Rowerem szybciej pokonamy dystans, który mielibyśmy przebyć pieszo, czy nawet korzystając z autobusów lub tramwajów.
- Przyczynia się do redukcji hałasu oraz redukcji emisji dwutlenku węgla (Shaheen i in., 2010) zachęcając kierowców do pozostawienia samochodów na obrzeżach miasta i skorzystania z transportu publicznego.
- Wpływa pozytywnie na stan zdrowia mieszkańców zmniejszając ryzyko chorób serca lub depresji. Na to jak duży będzie pozytywny wpływ jazdy na rowerze wpływa jednak płeć i wiek użytkownika (Woodcock, Tainio, Cheshire, O'Brien & Goodman, 2014).
- Zachęca do skorzystania z rowerów osoby, które wcześniej z różnych przyczyn nie były skłonne korzystać z tego środka komunikacji.
- Pozytywnie wpływa na wizerunek miasta, czyniąc je bardziej nowoczesnym oraz przyjaznym dla mieszkańców (Bryce, 2018).
- System ten jest również bardzo atrakcyjny w oczach turystów, którzy chętnie korzystają z tego środka transportu w celu zwiedzania miasta (Bryce, 2018).

#### 1.2.5 Wady rowerów miejskich

Oprócz licznych zalet można również znaleźć wady funkcjonowania tego typu systemów w miastach. Do najważniejszych z nich możemy zaliczyć:

- Niewielu rowerzystów korzystających z rowerów miejskich jest zaopatrzona w kask, co może stanowić zagrożenie dla ich zdrowia.
- Problem z liczbą rowerów na stacji, które w szczytowych momentach dnia mogą być przepełnione lub puste, co utrudnia ich wykorzystywanie.
- W przypadku możliwości pozostawiania rowerów w dowolnym miejscu mogą one szpecić krajobraz miasta.
- Dla wielu osób może to być pierwszy kontakt z jazdą rowerem po mieście. Jeżeli infrastruktura miasta nie jest przygotowana pod ten rodzaj transportu osoby takie mogą stanowić zagrożenie dla innych uczestników ruchu oraz dla samych siebie.

## 1.3 Rowery w Poznaniu

### 1.3.1 Infrastruktura rowerowa w Poznaniu

W Poznaniu znajduje się około 140 kilometrów dróg rowerowych. Ogólny udział w ruchu tego środka transportu stanowi 4%. Program Rowerowy Miasta Poznania na lata 2017-2022 z perspektywą do 2025 roku na podstawowy problem z infrastrukturą rowerową wskazuje brak spójności oraz ciągłości sieci. Według autorów jest to główny czynnik powodujący, iż ruch rowerowy nie jest w Poznaniu masowym środkiem transportu. Dodają, że istotnym czynnikiem wpływającym na bezpieczeństwo jest niezgodność części rozwiązań z tzw. *dobrą praktyką*, czyli standardami obowiązującymi w Poznaniu oraz przepisami w Polsce.

Ponadto w Poznaniu istnieje wiele stojaków rowerowych umożliwiających bezpieczne pozostawienie pojazdu. Brak jednak jest działań w tworzeniu parkingów dla tego środka transportu w różnych miejscach miasta, odpowiadającym zgłaszanemu zapotrzebowaniu (Poznań, [2017](#)).

### 1.3.2 Poznański Rower Miejski

Poznański Rower Miejski Powstał w 2012 roku i od tej pory jest ciągle rozwijany. Od początku jego operatorem jest największa firma w Polsce zajmująca się obsługą miejskich systemów rowerowych – NextBike. W roku 2018 do dyspozycji mieszkańców oddano 113 stacji oraz ponad 1200 pojazdów. Program ten z roku na rok bije kolejne rekordy i jest coraz chętniej wykorzy-

stywany przez mieszkańców. Tylko w 2018 roku rowery były wypożyczane ponad 1,65 miliona razy, co daje nam wynik średnio cztery wypożyczenia na minutę.

Program ten jest dalej unowocześniany. W sezonie na 2019 rok pojawiło się 420 nowych rowerów czwartej generacji, których nie trzeba zostawiać na stacji, oraz 30 pojazdów piątej generacji, tym samym łączna ich liczba to 1650 („Rekordy wypożyczeń rowerów miejskich w Poznaniu i Szamotułach”, [2019](#)).

## **1.4 Podsumowanie**

Rowery miejskie stają się nieodłącznym elementem wielu miast na całym świecie. Jest to wynik kierunku, jakie obrały one w swoim rozwoju oraz pojawienie się nowych trendów na rynku takich jak ekonomia współdzielenia. W celu zapewnienia ich mieszkańcom jak najwyższego komfortu użytkowania systemów rowerów miejskich przydatna może być szczegółowa analiza danych dotyczących jego wykorzystywania. W dzisiejszych czasach nie jest problemem dostęp do potrzebnych danych, które następnie mogą być wykorzystane np. do konstrukcji prognoz, które pozwolą nam na lepsze zarządzanie siecią rowerową.

## Rozdział 2

# Prognozowanie szeregów czasowych

### 2.1 Pojęcie prognozy

Rozwój gospodarczy sprawił, że prognozy znajdują coraz szersze zastosowanie nie tylko w ekonomii, ale również w wielu innych dziedzinach takich jak nauki przyrodnicze, socjologiczne, czy w psychologii.

Wiele instytucji na ich podstawie podejmuje decyzje, które mają na celu zwiększenie ich efektywności, zmniejszenie przewidywanych strat czy przygotowują je na nadchodzące zmiany, które próbują przewidzieć. Rządy państw prognozują przyszłe dochody i wydatki, by na tej podstawie móc racjonalnie gospodarować budżetem. Samorządom lokalnym mogą pomóc ustalić przyszłe potrzeby mieszkańców wynikające z demografii np. budowę nowej filii szkoły. Uczelnie wyższe mogą próbować przewidzieć liczbę chętnych do podjęcia studiów, by lepiej przygotować się przed nadchodzącym naborem. Producenci na ich podstawie dokonują decyzji, czy wycofać dany produkt z rynku, czy zmodernizować przedsiębiorstwo lub nie dokonywać żadnych zmian. Widzimy więc, że prognozy są ważną częścią naszej gospodarki, gdyż wspomagają podejmowanie decyzji wpływających na nas wszystkich.

Nie każdą próbę przewidzenia przyszłości możemy nazwać prognozą. Gdy w procesie tym uczestniczą eksperci, którzy na podstawie własnego doświadczenia i intuicji chcą ustalić przyszły stan danych zjawisk, mówimy wtedy o przewidywaniach zdroworozsądkowych. Możemy do nich zaliczyć przypuszczanie wielkości sprzedaży przez handlowców z dużym doświadczeniem. Innym ich rodzajem są przewidywania nieracjonalne, w których nie mamy żadnych przesłanek, lub nie mają one żadnego związku z wnioskiem. Są to wszelkiego rodzaju wróżby czy prorocтва. Ostatnie są przewidywania naukowe, gdzie do tworzenia konkluzji dochodzimy przy pomocy

reguł nauki.

Według Marii Cieślak Prognozowanie to „racjonalne, naukowe przewidywanie przyszłych zdarzeń” (Cieslak, 2005). Definicja ta szczególną wagę zwraca na użycie metodologii naukowej w dochodzeniu do danych wniosków. Nie gwarantują one pewności co do poziomu przyszłych zjawisk, lecz pozwalają nam wykorzystać pewne mechanizmy w próbie ich przewidzenia.

Warto zaznaczyć również różnicę między pojęciem prognozy a planu. Próbuąc analizować rozwój danego zjawiska w przyszłości staramy się podejmować takie działania, które ukształtują to zdarzenie w najbardziej pożądanym przez nas kierunku. Temu służy właśnie aktywne planowanie, które zakłada kształtowanie kierunku przemian przez dokonywanie świadomego wyboru metod działania. Prognoza natomiast zawiera jedynie przesłanki, że dane zjawisko będzie kształtować się w taki sposób, nie wskazuje natomiast na środki, których należy użyć, by dany efekt osiągnąć.

## 2.2 Rodzaje prognoz

Istnieje kilka celów przeprowadzania tego typu badań. Podstawowym powodem prognozowania społecznego jest tak zwana funkcja preparacyjna, która wspomaga podejmowanie innym istotnych dla podmiotu działań. Może nim być pojedyncza osoba, grupa osób, instytucja lub przedsiębiorstwo. W oparciu o nią podejmuje on decyzję, która będzie za sobą niosła odpowiednie konsekwencje w przyszłości, przez co spoczywa na nim duża odpowiedzialność. W związku z tym, powinien od posiadać odpowiednie umiejętności co do oceny jakości danej prognozy i wyznaczeniu tej, która jego zdaniem jest najbardziej trafna, co pozwala zaplanować odpowiednie działania sprzyjające realizacji wyznaczonych celów. Inną z nich jest funkcja aktywizująca, którą według Marii Cieślak polega na: „pobudzeniu do podejmowania działań sprzyjających realizacji prognozy, gdy zapowiada ona zdarzenia korzystne, i przeciwstawiających się jej realizacji, gdy zapowiada ona zdarzenia niekorzystne” (Cieslak, 2005). Warto również zaznaczyć, że pojęcie tego, czy dane zjawisko jest korzystne, czy też nie zależy od systemu wartości, które wyznaje dane społeczeństwo, których ona dotyczy. Prognozy takie nazywamy badawczymi. Mają one za zadanie pokazanie wielu możliwych wersji tego, jak dane zjawisko może w przyszłości się ukształtować. Wśród nich wyróżniamy prognozy ostrzegawcze, które mają za zadanie wykryć niekorzystne kształtowanie się zdarzenia, co daje czas podmiotowi na podjęcie działań zapobiegających. Ostatnią funkcją, jaką mogą pełnić prognozy jest funkcja informacyjna, która



dostarcza informacji o danym zjawisku i pozwala na oswojenie się z nadchodzącymi zmianami.

Same prognozy możemy podzielić ze względu na różne kryteria, które przyjmujemy. Najbardziej podstawowym z nich jest horyzont czasowych, którego ona dotyczy. W literaturze możemy spotkać się z różnym podziałem tego kryterium na okresy. Umownie przyjmujemy następujące rozróżnienie (Zielaś, 2003):

- **Bezpośrednia** - nie dłuższy niż 1 miesiąc.
- **Krótkoterminowa** - od 1 do 3 miesięcy.
- **Średnioterminowa** - od 3 miesięcy do 2 lat.
- **Długoterminowa** - powyżej 2 lat.

Podział ten jest jednak umowny, gdyż zależy od charakteru badanego zjawiska, o którym decyduje wielkość badanego obiektu oraz istota zachodzących w nim zmian. Przykładowo dla prognoz demograficznych za krótkoterminowy przyjmuje się okres do 5 lat, średniookresowy od 5 do 10, a długoterminowy powyżej 10. Jeszcze inaczej jest w meteorologii, gdzie prognozy krótkookresowe dotyczą czasu do 24 godzin, średniookresowe do tygodnia, a długookresowe do 1 miesiąca. Jak więc możemy zobaczyć, nie ma jednej określonej ramy co do tego podziału. Warto również zaznaczyć, że im dalszy jest horyzont czasowy danej prognozy, tym mniejsze jest prawdopodobieństwo trafnego przewidzenia stanu zjawiska, co zmniejsza pewność prognozy (Cieslak, 2005).

Innym kryterium, które przyjmujemy, jest charakter zmiennej prognozowanej. Tutaj rozróżniamy dwa podziały – ilościowe i jakościowe. W pierwszym przypadku stan prognozowanego zjawiska wyrażamy liczbą. Jest to np. prognozowanie kursów walutowych czy stóp procentowych. Te z kolei możemy podzielić na punktowe, gdy podajemy określoną wartość zmiennej, lub przedziałowe, gdy wyznaczamy zakres, w którym się ona znajduje. O prognozach jakościowych mówimy w momencie, gdy dotyczą one zmiennej jakościowej np. inflacji czy jakości życia. Dotyczy to bowiem faktu, iż wielkość danego zjawiska wykaże wzrost lub spadek w stosunku do wartości zaobserwowanej w momencie jej formułowania.

Ciekawym kryterium jest podział na prognozy samosprawdzające się i destruktywne. Te pierwsze to takie, których opublikowanie sprawia, że dane zjawisko zaczyna kształtować się w taki sposób, w jaki wynika z badania pomimo, iż wcześniej szanse na zrealizowanie takiego

scenariusza nie były duże. W drugiej sytuacji mamy efekt odwrotny - ogłoszenie takiej prognozy zmniejsza to prawdopodobieństwo. Takie wzajemne oddziaływanie jest nazywane efektem Edypa (Cieslak, 2005).

Ostatnim omawianym kryterium jest zasięg terenowy. Stąd dzielimy je na światowe, międzynarodowe, krajowe czy regionalne.

## 2.3 Prognozowanie szeregów czasowych

**Szeregiem czasowym** (ang. *time series*) nazywamy obserwacje danego zjawiska w kolejnych, zazwyczaj jednakowych odstępach czasu.

Wyróżniamy w nich dwie składowe: systematyczną, która jest efektem stałego zestawu czynników, które oddziałują na zmienną prognozowaną, i przypadkową, często nazywaną składnikiem losowym. Pierwsza z nich może objawiać się w postaci (Cieslak, 2005):

- **Tendencji rozwojowej** tzw. trendu, czyli długookresowej skłonności zmiennej do jednokierunkowej zmiany.
- **Wahań sezonowych**, które wykazują zmianę poziomu zmiennej wokół tendencji rozwojowej o okresie nie przekraczającym roku (najczęściej miesiące, kwartały). Wpływ na nie mogą mieć np. pory roku.
- **Stałego poziomu** prognozowanej zmiennej, kiedy poziom badanego zjawiska oscyluje wokół jakiejś stałej wartości.
- **Wahań cyklicznych** są to długoterminowe wahania poziomu zmiennej wokół tendencji rozwojowej, której okres trwa dłużej niż rok. Ich najczęstszą przyczyną jest koniunktura gospodarcza.

W większości przypadków samo spojrzenie na stworzony wykres pozwala na identyfikację poszczególnych składowych szeregu. Proces ich wyodrębniania nazywamy dekompozycją szeregu. Innym sposobem ich identyfikacji jest przeprowadzenie analizy autokorelacji, czyli zależności między poszczególnymi obserwacjami tej samej zmiennej. Współczynnik autokorelacji informuje nas, czy są one istotnie związane z wartościami wcześniejszymi, przy czym obliczamy go do różnych rzędów  $(1, 2, \dots, k)$  wobec zmiennej prognozowanej  $Y$ . Oznacza to badanie korelacji między wartością  $Y$  oraz wartością  $Y$  opóźnionej o  $1, 2, \dots, k$  okresów. Jeżeli współczynnik

ten dla pierwszych kilku rzędów jest duży i statystycznie istotny, a dla następnych stopniowo maleje oznacza to występowanie trendu. W innym przypadku, jeżeli jego wartość jest statystycznie istotna dla rzędu równego liczbie faz cyklu sezonowego (np. 4 dla danych kwartalnych lub 12 dla miesięcznych) oznacza to występowanie w szeregu czasowym wahań sezonowych. Jeszcze innym sposobem na wykrycie poszczególnych składowych jest wykorzystanie testów statystycznych (np. test Coxa-Stuarta<sup>1</sup>) lub przeprowadzenie analizy wariancji.

### 2.3.1 Modele szeregów czasowych

W zależności od wpływu poszczególnych składowych oraz wzajemnych interakcji między nimi wyróżniamy dwa modele szeregu czasowego. Pierwszy z nich - model addytywny, zakłada, że obserwowane wartości są sumą poszczególnych elementów mających wpływ na jego wartość. Zakładamy w nim brak wzajemnego wpływu między poszczególnymi składowymi. Jeżeli zmienna czasowa jest jedyną zmienną mającą wpływ na wartość obserwacji to przyjmuje on następującą postać:

$$X_t = f(t) + g(t) + h(t) + \xi_t, \quad (2.1)$$

lub

$$X_t = a_0 + g(t) + h(t) + \xi_t, \quad (2.2)$$

gdzie:  $f(t)$  jest funkcją czasu, która charakteryzuje tendencję rozwojową (trend);  $g(t)$  jest funkcją czasu, która charakteryzuje wahania sezonowe,  $h(t)$  jest funkcją czasu, która charakteryzuje wahania cykliczne,  $\xi_t$  oznacza składnik losowy,  $a_0$  jest stałym poziomem prognozowanej zmiennej.

Drugim modelem nazywamy multiplikatywnym. Zakłada on, że wartość obserwowanej zmiennej jest iloczynem poszczególnych składowych elementów szeregu czasowego. Jeżeli zmienna czasowa jest jedyną zmienną mającą wpływ na jej wartość, model ten przyjmuje następującą postać:

$$X_t = f(t) \times g(t) \times h(t) \times \xi_t, \quad (2.3)$$

---

<sup>1</sup>Test ten jest używany do analizy występowania tendencji rozwojowej w danych, a proponowana metoda opiera się na rozkładzie dwumianowym. Został on zaprezentowany w 1955 roku (Cox & Stuart, 1955).

lub

$$X_t = a_0 \times g(t) \times h(t) \times \xi_t, \quad (2.4)$$

gdzie oznaczenia jak w poprzednio.

W modelu tym jedną ze składowych, najczęściej jest to tendencja rozwojowa lub stały poziom prognozowanej zmiennej, jest wyrażona w jednostkach zmiennej prognozowanej, zaś pozostałe z elementów są względnymi odchyleniami od nich.

### 2.3.2 Podział danych na część uczącą i testową

Kiedy wyznaczamy prognozy przy pomocy różnych metod w celu ich ocenienia i porównania skuteczności warto podzielić zbiór danych na część uczącą i testową. W takiej sytuacji konstruujemy predykcję dla pierwszego ze zbiorów, a ich dokładność oceniamy na podstawie drugiego z nich, porównując je z danymi rzeczywistymi.

## 2.4 Analiza szeregu czasowego

### 2.4.1 Funkcja autokorelacji ACF i autokorelacji cząstkowej PACF

Przed przystąpieniem do budowy modelu i prognozowania musimy dokładnie przyjrzeć się zarejestrowanemu szeregowi czasowemu w celu identyfikacji jego własności.

Przydatnym narzędziem przy wstępnej analizie szeregu czasowego jest funkcja autokorelacji ACF (z ang. *Autocorrelation Function*), którą definiujemy jako miarę zależności pomiędzy obserwacjami oddległymi o  $h$  jednostek czasowych, zwanych parametrami opóźnienia. Definiujemy ją jako:

$$ACF(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h = 0, 1, 2, \dots, n-1, \quad (2.5)$$

gdzie  $\hat{\gamma}(h)$  jest funkcją autokowariancji

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=h}^{n-1} (X_{j+h} - \bar{X})(X_j - \bar{X}), \quad h = 0, 1, 2, \dots, n-1, \quad (2.6)$$

gdzie  $\bar{X} = \sum_i^n X_i / n$ .

Wykres funkcji ACF, zwany korelogramem, pozwala nam na łatwą identyfikację własności szeregu. Jeżeli jej wartości są dodatnie i z czasem powoli zanikają świadczy to o występowaniu

trendu, jeżeli zanika ona powoli i cyklicznie sugeruje ona obecność sezonowości. Używamy jej do identyfikacji modelu ruchomej średniej  $MA$ .

Funkcja autokorelacji cząstkowej PACF (z ang. *Partial Autocorrelation Function*) jest związana z funkcją ACF. Za jej pomocą sprawdzamy bezpośrednią zależność między jednostkami oddległymi o  $h$  jednostek czasowych, eliminując zależności od obserwacji pośrednich. Jeżeli dla pierwszego opóźnienia jej wartość jest bliska 1, może to świadczyć o występowaniu trendu wzrostowego w analizowanym szeregu. Może ona służyć do identyfikacji modelu autoregresji  $AR(p)$ .

Przy badaniu szeregu czasowego z wykorzystaniem funkcji autokorelacji i autokorelacji cząstkowej posługujemy się również przedziałami ufności. Jeżeli dla danego opóźnienia próbkowego autokorelacja jest od nich większa uznajemy ją za statystycznie istotną. 95% przedział ufności ma postać:

$$(-1,96\sqrt{n}, 1,96\sqrt{n}) \quad (2.7)$$

gdzie 1,96 jest kwantylem z rozkładu  $N(0, 1)$ , a  $n$  jest liczbą obserwacji.

## 2.4.2 Różnicowanie

Różnicowanie jest podstawową transformacją szeregu czasowego do postaci stacjonarnej, czyli takiej, której właściwości nie zmieniają się w czasie, pozbawionej trendu i wahań sezonowych, jego właściwości możemy scharakteryzować następująco: (Zagdański & Suchwałko, 2016)

- $Corr(X_r, X_s) = Corr(X_{r+t}, X_{s+t}) = p(s - r) = p(h)$ , czyli korelacja pomiędzy obserwacjami zależy tylko od odstępu czasowego (z ang. *lag* - opóźnienie);
- $Var(X_t) = const$  - wariancja nie zmienia się w czasie;
- $EX_t = const$  - wartość oczekiwana jest stała, co oznacza, że szereg pozbawiony jest tendencji rozwojowej lub wahań sezonowych

Jej podstawową wersją jest różnicowanie z opóźnieniem różnym 1, dla którego wyznaczamy szereg przyrostów odejmując wartości z następujących po sobie okresów  $X_2 - X_1, X_3 - X_2, \dots, X_n - X_{n-1}$ , w ten sposób otrzymujemy szereg krótszy o długości  $n - 1$ , który wykorzystujemy przy dalszej analizie.

Jeżeli chcemy z szeregu usunąć sezonowość, możemy wykonać operację różnicowania z opóźnieniem sezonowym  $s$ . W tym przypadku otrzymujemy szereg różnic  $Y_t = X_t - X_{t-s}$ , gdzie  $t = s + 1, \dots, n$  o długości  $n - s$ .

## 2.5 Metody prognozowania

Jeżeli dokonamy już dekompozycji szeregu czasowego i poznamy, jakie jego składowe mają wpływ na poszczególne zmienne, musimy wtedy wybrać optymalną metodę, dzięki której unikniemy błędów przy konstrukcji prognozy.

### 2.5.1 Metoda naiwna

Metody naiwne należą do najprostszych modeli prognozowania. Zakłada się, że czynniki mające wpływ na wartość zmiennej prognozowanej w przyszłości będą oddziaływać na nią w identycznym stopniu, co w latach ubiegłych. Wykorzystuje się je do konstrukcji prognoz krótkookresowych na jeden okres do przodu, przy czym nie mogą występować duże wahania losowe w szeregu zmiennej prognozowanej. Najbardziej popularna metoda zakłada konstrukcję prognozy na czas  $t$  na poziomie zaobserwowanej wartości zmiennej prognozowanej w czasie  $t - 1$ . Model ten wygląda następująco:

$$X_t^* = X_{t-1}, \quad (2.8)$$

gdzie:  $X_t^*$  to wartość prognozy na czas  $t$ ,  $X_{t-1}$  to wartość zmiennej prognozowanej w czasie  $t - 1$ .

Najprostsza z nich wykorzystuje jedynie ostatnią znana wartość zmiennej prognozowanej, bardziej skomplikowane metody naiwne wymagają wykorzystania większej liczby obserwacji. Wykorzystuje się je do badania rynków papierów wartościowych czy kursów walut. Jakość takiej prognozy jest z reguły niska (Cieslak, 2005).

### 2.5.2 Metoda średniej ruchomej

Metoda Średniej ruchomej, nazywana inaczej metodą mechaniczną, polega na sekwencyjnym wyznaczeniu średnich arytmetycznych z kilku obserwacji i zastąpieniu nią najczęściej środkowej z obserwacji, na których podstawie były wyznaczane. W modelu tym zakładamy, że następna

obserwowana wartość będzie równa średniej arytmetycznej z  $k$  ostatnich wartości tej zmiennej. W wersji prostej przyjmuje on postać:

$$X_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} X_i, \quad (2.9)$$

gdzie:  $X_t^*$  oznacza prognozowaną wartość obserwacji  $X$  w czasie  $t$ ,  $X_i$  oznacza wartość zmiennej prognozowanej w czasie  $t$ ,  $k$  jest stałą wygładzania.

Stałą wygładzania wyznacza osoba przeprowadzająca badanie. Należy pamiętać, że im jest ona większa, tym silniej wygładza ona szereg, jednak słabiej reaguje na zmiany poziomu prognozowanej zmiennej i odwrotnie – im będzie ona mniejsza, tym bardziej będzie ona pokazywała zachodzące zmiany w prognozowanej zmiennej, jednak wygładzenie będzie słabsze, co oznacza większy wpływ wahań losowych.

Przyjęto się, że ponieważ nowsze obserwacje lepiej informują nas o obserwowanym zjawisku, powinniśmy nadawać im większą wagę niż starszym. W poprzednim wariacie tej metody wszystkie z nich mają jednakową wagę, powstał więc model średniej ruchomej ważonej, który zakłada tzw. postarzanie informacji. Przyjmuje on postać:

$$X_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} X_i w_{i-t+k+1}, \quad (2.10)$$

gdzie:  $X_t^*$  oznacza prognozowaną wartość obserwacji  $X$  w czasie  $t$ ,  $X_i$  oznacza wartość zmiennej prognozowanej w czasie  $t$ ,  $w_{i-t+k+1}$  jest wagą nadaną na wartość zmiennej prognozowanej w czasie  $t$ ,  $k$  jest stałą wygładzania.

$$0 < w_1 < w_2 < \dots < w_k \leq 1 \quad \text{oraz} \quad \sum_{i=1}^k w_i = 1. \quad (2.11)$$

Modele te stosuje się zazwyczaj w sytuacji, gdy poziom obserwowanej zmiennej w zanotowanym okresie utrzymuje się na prawie stałym poziomie i występują tylko niewielkie wahania losowe, brak jest zaś tendencji rozwojowej, wahań sezonowych i cyklicznych (Cieslak, 2005).

### 2.5.3 Modele ARMA i ARIMA

Modele stacjonarne służą do budowania modelu dla danych, w których nie występuje tendencja rozwojowa, ani wahania sezonowe czy okresowe, a wariancja jest jednorodna w czasie. W rzeczywistości rzadko mamy do czynienia z takimi szeregami, dlatego konieczne jest ich

przekształcenie, aby usunąć niepożądane własności. Do najbardziej znanych tego typu modeli zalicza się (Zagdański & Suchwałko, 2016):

- $WN(0, \sigma^2)$ : biały szum o średniej 0 i wariancji  $\sigma^2$ , zakłada brak korelacji czasowej i jedynkowy rozkład danych.
- $AR(p)$ : model autoregresji rzędu  $p$ ,
- $MA(q)$ : model Średniej ruchomej rzędu  $q$ ,
- $ARMA(p, q)$ : model mieszany ruchomej średniej rzędu  $q$  i autoregresji rzędu  $p$ ,
- $ARMA(P, Q)_s$ : sezonowy model autoregresji rzędu  $P$ , ruchomej średniej rzędu  $Q$ ,
- $ARMA(p, q) \times (P, Q)_s$ : połączony model  $ARMA(p, q)$  i  $ARMA(P, Q)_s$ .

**Model autoregresji  $AR(p)$**  – nazywamy stacjonarny szereg czasowy  $X_t$  rzędu  $p$ , który spełnia równanie:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t, \quad (2.12)$$

gdzie  $Z_t$  jest szeregiem typu biały szum.

Model ten jest uważany za jeden z najbardziej intuicyjnych, gdzie wartości szeregu czasowego w czasie  $t$  są przedstawione jako liniowa kombinacja  $p$  wcześniejszych wartości uzupełnionych o biały szum.

**Model ruchomej średniej  $MA(q)$**  – nazywamy stacjonarny szereg czasowy  $X_t$  rzędu  $q$ , który spełnia równanie:

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}. \quad (2.13)$$

Model ten pozwala na uwzględnienie korelacji czasowej. O tym, jak będzie ona silna decyduje rząd  $q$  oraz współczynniki  $\theta_1 \dots \theta_q$ . Jest to przykład modelu, gdzie dla rzędu większego niż  $q$  korelacja czasowa jest równa 0.

**Model autoregresji ruchomej średniej  $ARMA(p, q)$**  jest to szereg spełniający równanie:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}. \quad (2.14)$$

Dodatkowo w celu zapewnienia jednoznaczności reprezentacji zakładamy, że wielomiany:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad (2.15)$$



oraz

$$\theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q, \quad (2.16)$$

nie mogą być zredukowane do wielomianów o mniejszym stopniu.

Modele  $AR(p)$  i  $MA(q)$  możemy przedstawić jako szczególne przypadki modelu  $ARMA(p, q)$ , odpowiednio  $ARMA(p, 0)$  i  $ARMA(0, q)$ . Są one uznawane za wystarczająco bogatą rodzinę modeli dla szeregów stacjonarnych (Zagdański & Suchwałko, 2016).

Często zdarza się, że zależność od wcześniejszych obserwacji jest największa dla wielokrotności pewnego opóźnienia sezonowego  $s$ , np. 12 dla danych miesięcznych czy 4 dla danych kwartalnych. W celu analizy takich danych istnieje sezonowy wariant poprzedniego modelu – **sezonowy model autoregresji ruchomej średniej**  $ARMA(P, Q)_s$ , który ten związek uwzględnia. Równanie takiego szeregu ma postać:

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)Z_t, \quad (2.17)$$

gdzie  $\Phi$  jest sezonowym wielomianem autoregresji stopnia  $P$ , który definiujemy jako:

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}, \quad (2.18)$$

a  $\Theta_Q$  – jest sezonowym wielomianem ruchomej średniej stopnia  $Q$ , zdefiniowanym jako:

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (2.19)$$

**Model**  $ARMA(p, q) \times (P, Q)_s$  uwzględnia natomiast zarówno zależności zwykłe jak i sezonowe. Równanie stacjonarnego szeregu czasowego tego modelu ma postać:

$$\Phi_P(B^s)\phi(B)X_t = \Theta_Q(B^s)\theta(B)Z_t \quad (2.20)$$

Wszystkie omówione modele mogą być zastosowane w przypadku, gdy mamy do czynienia z szeregami czasowymi stacjonarnymi. Jednak jeżeli analizowane dane wykazują odstępstwa od tej zasady powinniśmy zastosować odpowiednie transformacje, które pozwolą nam przekształcić szereg z niestacjonarnego na stacjonarny. Modele, które uwzględniają operację różnicowania to modele  $ARIMA(p, d, q)$  i  $SARIMA(p, d, q)(P, D, Q)_s$ . Pierwszy z nich sprowadza się do modelu  $ARMA(p, d)$  po  $d$ -krotnym ( $d \geq 0$ ) z opóźnieniem jeden. Możemy go zdefiniować

jako:

$$\phi^*(B)X_t \equiv \phi(b)(1-B)^d X_t = \theta(B)Z_t, \quad Z_t \sim WN(0, \sigma^2), \quad (2.21)$$

$\phi(z)$  i  $\theta(z)$  są odpowiednio wielomianami stopnia  $p$  i  $q$  oraz  $\phi(z) \neq 0$  dla  $|z| \leq 1$ .

Model ten jest w szczególności przydatny do eliminowania trendu długoterminowego z analizowanego szeregu czasowego. Ogólniejszym modelem jest  $SARIMA(p, d, q)(P, D, Q)_s$ , który pozwala na eliminację wahań sezonowych oraz trendu, poprzez różnicowanie z opóźnieniem sezonowym  $s$  i  $d$ -krotnym różnicowaniu z opóźnieniem 1 dla  $d, D \geq 0$  i  $d, D \in N$ , który po wykonaniu wszystkich przekształceń sprowadza się do odpowiedniego modelu  $ARMA(p, q) \times (P, Q)_s Y_t = (1-B)^d (1-B^s)^D X_t$ . Ma on postać:

$$\phi_P(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad Z_t \sim WN(0, \sigma^2), \quad (2.22)$$

gdzie  $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ ,  $\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_P z^P$ ,  $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$  oraz  $\Theta(z) = 1 + \Theta_1 z + \dots + \Theta_Q z^Q$ .

## 2.5.4 Analiza reszt

Ocenę poprawności dopasowania modelu możemy dokonać np. przy pomocy analizy reszt. Wyznaczamy je jako odpowiednio przeskalowane różnice wartości szeregu czasowego oraz prognoz wyznaczonych na podstawie zbudowanego dopasowanego modelu. Możemy je zapisać jako:

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\phi}; \hat{\theta})}{r_{t-1}(\hat{\phi}; \hat{\theta})^{1/2}}, \quad t = 1, \dots, n, \quad (2.23)$$

gdzie  $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)$  oraz  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)$  są oszacowanymi współczynnikami modelu,  $\hat{X}_t(\hat{\phi}; \hat{\theta})$  jest optymalną prognozą dla obserwacji  $X_t$ , a  $r_{t-1}(\hat{\phi}; \hat{\theta})$  jest błędem średniokwadratowym predykcji ( $r_{t-1} = E(X_t - \hat{X}_t)^2 / \hat{\sigma}^2$ ).

Jeżeli model, który dopasowaliśmy, dobrze wyjaśnia dynamikę czasową analizowanych danych, wtedy w szeregu reszt nie powinno być widocznych żadnych regularnych zachowań (takich jak tendencja rozwojowa czy sezonowość) oraz istotnej korelacji czasowej, co oznacza, że powinien on zachowywać się jak biały szum (Zagdański & Suchwałko, 2016).

Do analizy tej cechy służy m.in. test losowości Ljunga-Boxa, opartym o statystykę testową

dla kilku lub kilkadziesiąt opóźnień. Ma ona wzór:

$$Q_{LB} = n(n+2) \sum_{j=1}^h acf^2(j)/(n-j) \quad (2.24)$$

gdzie  $h$  oznacza maksymalne opóźnienie, a  $acf(j)$  to próbkowa autokorelacja dla opóźnienia  $j$ .

Jeżeli wartość statystyki jest zbyt duża, uznajemy, że wartości ACF są zbyt duże, abyśmy mogli uznać szereg jako realizację ciągu nieskorelowanego. Możemy w tym przypadku posłużyć się wskaźnikiem p-value dla wybranego poziomu istotności (najczęściej  $\alpha = 0,05$ ). Jeżeli jego wartość jest niższa niż ustalony poziom istotności odrzucamy hipotezę o losowości reszt.

### 2.5.5 Kryteria oceny modelu

Różne kryteria informują nas o jakości dopasowania modelu przy pomocy danych historycznych, które jednocześnie kontrolują złożoność modelu. Zwiększając liczbę parametrów otrzymamy model lepiej dopasowany, jednak zbyt dobre dopasowanie może powodować problemy i prowadzić do złych prognoz. W celu uniknięcia takich sytuacji w kryteriach występuje składnik kary za jego wymiar. Do najbardziej popularnych kryteriów zalicza się (Zagdański & Suchwałko, 2016):

- **AIC (Akaike Information Criterion)** o wzorze:

$$AIC(p, q) = -2\ln(L) + 2(p + q + k + 1) \quad (2.25)$$

- **AICC (Corrected AIC)** o wzorze:

$$AICC(p, q) = -2\ln(L) + \frac{2(p + q + k + 1)n}{(n - p - q - k - 2)} \quad (2.26)$$

- **BIC (Bayesian Information Criterion)** o wzorze:

$$BIC(p, q) = -2\ln(L) + (p + q + k + 1) \ln n \quad (2.27)$$

We wzorach zmienna  $L$  oznacza współczynnik wiarygodności, z kolei  $k = 0$ , jeżeli w modelu nie uwzględniamy stałą  $c$  w innym przypadku  $k = 1$ . Wybierając odpowiedni model na podsta-

wie przedstawionych kryteriów porównujemy ich wartości i wybieramy ten, których wartości będą najmniejsze.

### 2.5.6 Kryteria oceny prognozy

Każdej skonstruowanej prognozie powinna towarzyszyć ocena jej dokładności. Pozwala to na oszacowanie ryzyka przy podejmowaniu decyzji podpartych informacjami, które z niej wyciągamy. Dodatkową zaletą jest możliwość porównania różnych zbudowanych wcześniej modeli i wyboru najbardziej optymalnego z nich. Jest to najczęściej najważniejszy wskaźnik, na podstawie którego wybieramy odpowiednią metodę prognozowania (Zagdański & Suchwałko, 2016).

Wśród wielu kryteriów służących do oceny prognozy jednymi z podstawowych są te oparte na błędach predykcji. Służą one do wyznaczania średnią wartość błędu prognozy z wartości bezwzględnych, co sprawia, że nie będą się one wzajemnie redukowały. Do tych kryteriów zaliczamy:

- **MAD/MAE - średni błąd absolutny** o wzorze:

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - F_i| \quad (2.28)$$

gdzie:  $X_i$  jest wartością rzeczywistą,  $F_i$  jest wartością prognozowaną;

- **MSE - (średni błąd kwadratowy** o wzorze:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - F_i)^2. \quad (2.29)$$

- **RMSE - Pierwiastek błędu średniokwadratowego** o wzorze:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - F_i)^2}. \quad (2.30)$$

Zaletą kryterium MAE jest jego łatwa interpretacja, z kolei kryterium MSE jest wygodne ze względów obliczeniowych. Atutem RMSE jest podanie wyniku w tej samej skali co dane, przez co jest on częściej wybierany niż MSE. Wadą obu ostatnich kryteriów jest natomiast ich duża podatność na odstające obserwacje (Zagdański & Suchwałko, 2016).

## 2.6 Podsumowanie

Rodzina modeli ARIMA jest wystarczająco bogata do konstrukcji prognoz. W następnym rozdziale przy pomocy pakietu statystycznego R dokonana zostanie analiza szeregu czasowego przygotowanego na podstawie zebranych wcześniej danych, po czym wybrane zostaną odpowiednie modele. Później zbadana zostanie jakość ich dopasowania do danych oryginalnych oraz porównane kryteria oceny prognozy dla danych uczących i testowych. Na tej podstawie wybrany zostanie dokładniejszy z nich, który posłuży do konstrukcji prognozy dla całości danych.

## Rozdział 3

# Konstrukcja prognozy dla stacji Brama

## Poznania

### 3.1 Pobieranie i analiza danych

Dane wykorzystane do konstrukcji prognozy pochodzą z oficjalnego API znajdującego się na stronie firmy NextBike, które zajmuje się prowadzeniem systemu rowerów miejskich w Poznaniu. Przedstawiają one średnią liczbę rowerów miejskich znajdujących się na stacji dla każdej godziny, zbieranych od 17 maja do 17 czerwca 2019 roku. Udostępniane są one w formacie XML i aktualizowane co 5 minut. W celu ich zebrania wykorzystane zostały pakiety XML (Wickham, Hester & Ooms, 2018), rvest (Wickham, 2019), RSQLite (Müller, Wickham, James & Falcon, 2018), tidyverse (Wickham, 2017), httr (Wickham, 2018) dostępne w pakiecie statystycznym R (R Core Team, 2013).

```
url <- 'https://nextbike.net/maps/nextbike-official.xml' 1
data <- read_xml(url) 2
dzien <- Sys.Date() 3
poznani_rowery <- xml_children(xml_child(xml_child(data, 11), 1)) 4
lista <- poznani_rowery %>% map(xml_attrs) %>% flatten 5
klucze <- unique(names(lista)) 6
lista2 <- sapply(klucze, function(name) {unlist(lista[grep(name, names( 7
  lista))])})
lista2[[7]] <- lista2[[7]][c(T,F)] 8
df <- data.frame(Stacja = lista2[[4]], row = as.character(lista2[[7]]), 9
  Dzień = dzien, Godzina = format(Sys.time(), "%H:%M"))
10
df_full <- bind_rows(df) 11
cat('Inserting into database...\n') 12
my_db <- src_sqlite("C:/Users/bartek/Desktop/Dane/dane.sqlite3") 13
db_insert_into(con = my_db$con, table = "rowery", values = df_full) 14
RSQLite::dbDisconnect(my_db$con) 15
```

**Program 3.1. Pobieranie danych z API NextBike oraz zapisywanie w formacie sqlite3**

### 3.1.1 Eksploracyjna analiza danych

Stacją wybraną do konstrukcji prognozy jest Brama Poznania. Znajduje się ona w najstarszej części miasta zlokalizowanej nad rzeką Cybiną, nieopodal wyspy katedralnej. W jej okolicy znajdują się liczne atrakcje turystyczne. W głównym gmachu modernistycznego budynku znajduje się centrum interpretacji dziedzictwa, w których można poznać historię wyspy katedralnej i rozwoju Poznania. Ponadto po drugiej stronie rzeki mieści się Śluza Katedralna, połączona kładką z pierwszym budynkiem kładką.

Brama Poznania jest również miejscem odpoczynku dla wielu mieszkańców. Specjalnie zaprojektowana przestrzeń służy spędzaniu czasu wolnego, spotkań z przyjaciółmi i organizacji różnych wydarzeń kulturowych. Miejsce to zostało wyróżnione licznymi nagrodami.

Znajduje się na niej 18 miejsc do umieszczenia rowerów (można je również zostawić podłączone do innego, stojącego już pojazdu, lub w jej pobliżu w przypadku rowerów 4 generacji), średnia ich dzienna liczba to ok. 10,22. Odchylenie standardowe wyniosło 4,77 rowery w przeciągu badanego okresu. Pod względem średniej liczby rowerów zajmuje ona 22 miejsce spośród wszystkich 140 stacji mieszczących się na terenie Poznania, gdzie tylko dla 24 z nich jest ona większa od 10.

Została ona wybrana ze względu na atrakcyjne położenie w okolicy wartościowych punktów turystycznych nieopodal centrum miasta oraz obszarów wspierających aktywny wypoczynek mieszkańców Poznania.

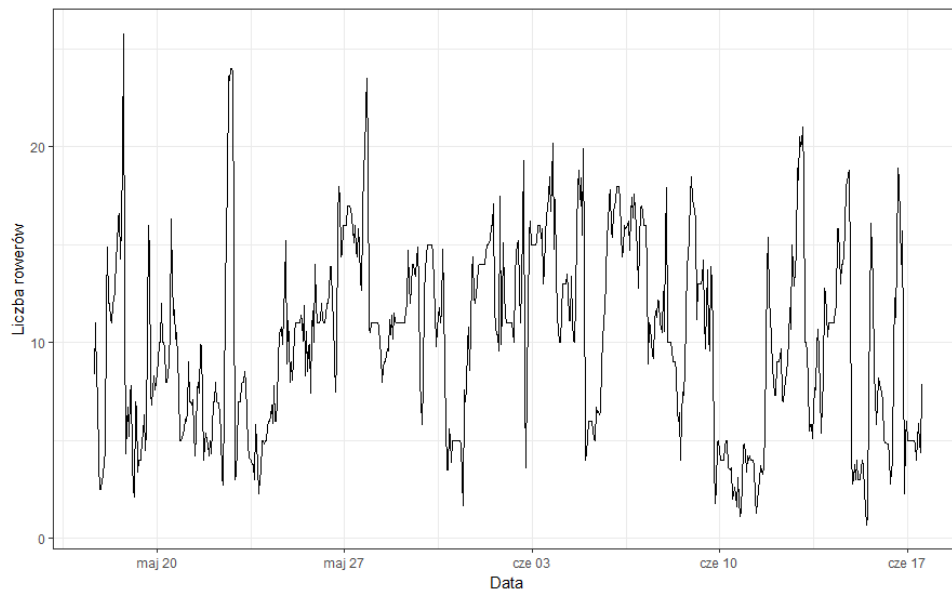
Rysunek 3.1 przedstawi całość zebranych danych.

W celu zbudowania dokładniejszej prognozy zbiór danych został podzielony na dwie części – uczącą i testową. Został on dokonany arbitralnie na podstawie horyzontu prognozy i służy zbadaniu jej dokładności dla skonstruowanego modelu.

W celu identyfikacji składowych modelu i wyboru odpowiedniego modelu, przy pomocy pakietu statystycznego R zbudowany został wykres funkcji ACF i PACF dla danych ze zbioru uczącego.

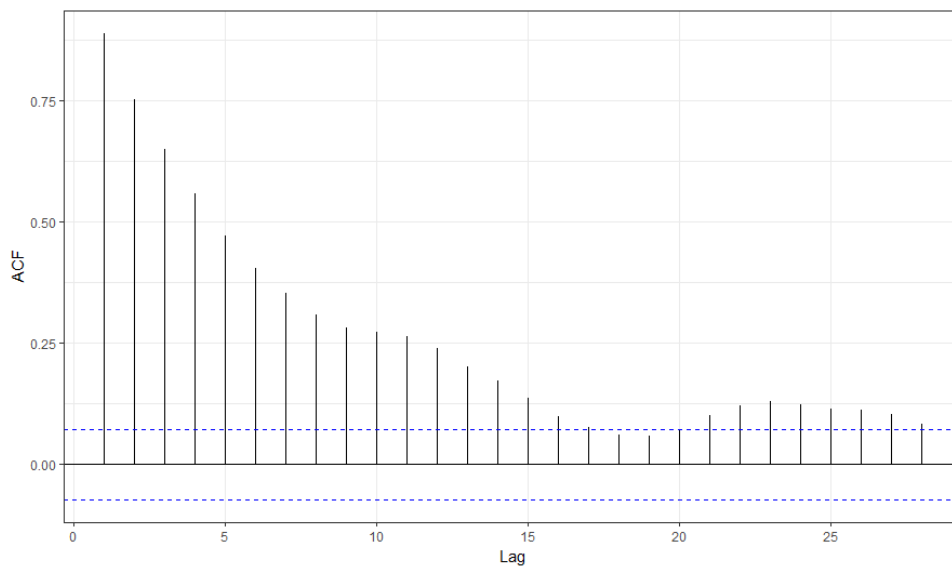
Wysoka wartość pierwszej obserwacji dla funkcji PACF wskazuje na występowanie tendencji rozwojowej, co potwierdza wykres funkcji ACF, dla którego autokorelacje zanikają powoli w czasie. W celu jej eliminacji wykonana została operacja różnicowania z opóźnieniem 1. Po niej wykresy funkcji prezentują się następująco:

Analizując wykres 3.5 po operacji różnicowania z opóźnieniem 1 dla funkcji PACF ostatnia istotna autokorelacja, która wychodzi poza przyjęty przedział ufności, występuje dla opóźnienia



**Rysunek 3.1. Liczba rowerów miejskich na stacji Brama Poznania w dniach 17 maja-17 czerwca**

Źródło: Opracowanie własne.

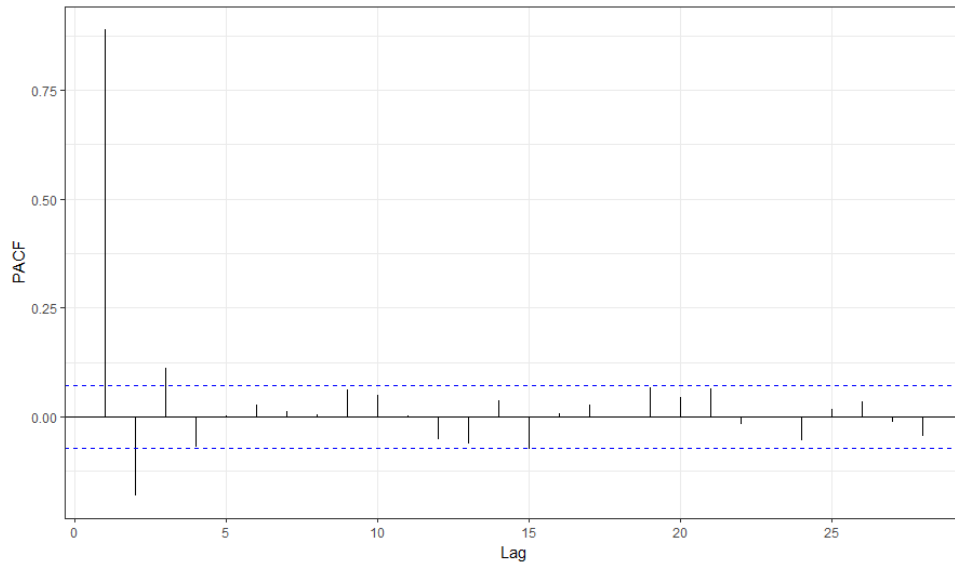


**Rysunek 3.2. Funkcja ACF zbioru uczącego dla stacji Brama Poznania**

Źródło: Opracowanie własne.

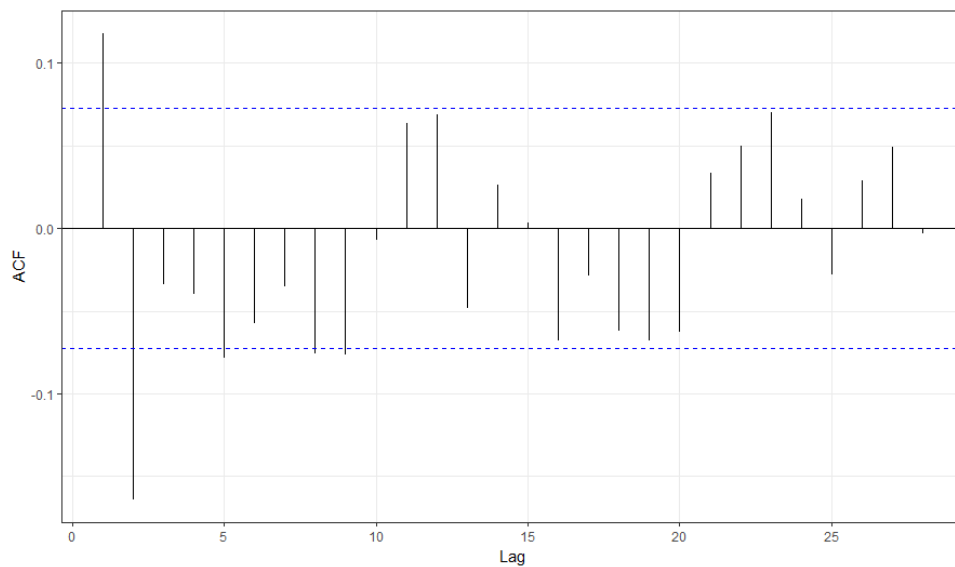
$h = 20$ . dla następnych opóźnień nie występują żadne znaczące autokorelacje, co pozwala nam założyć, że dla następnych z nich jest to realizacja szeregu typu biały szum.





**Rysunek 3.3. Funkcja PACF zbioru uczącego dla stacji Brama Poznań**

Źródło: Opracowanie własne.

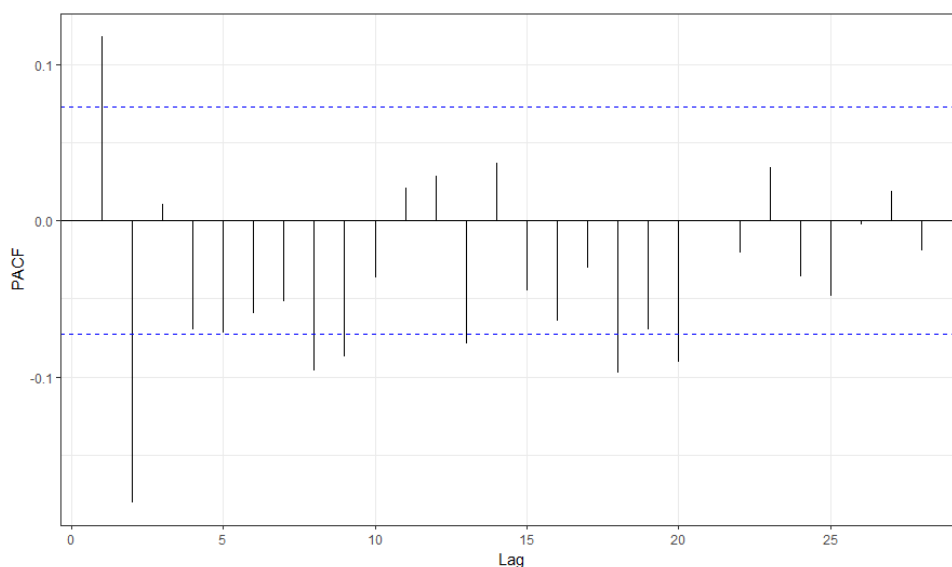


**Rysunek 3.4. Funkcja ACF zbioru uczącego po zróżnicowaniu z opóźnieniem 1 dla stacji Brama Poznań**

Źródło: Opracowanie własne.

## 3.2 Modele ARIMA

Na potrzeby prognozy zbudowane zostały dwa modele, pierwszy z nich został zbudowany na podstawie analizy skonstruowanych wcześniej wykresów autokorelacji i autokorelacji cząstko-



**Rysunek 3.5. Funkcja PACF zbioru uczącego po zróżnicowaniu z opóźnieniem 1 dla stacji Brama Poznań**

Źródło: Opracowanie własne.

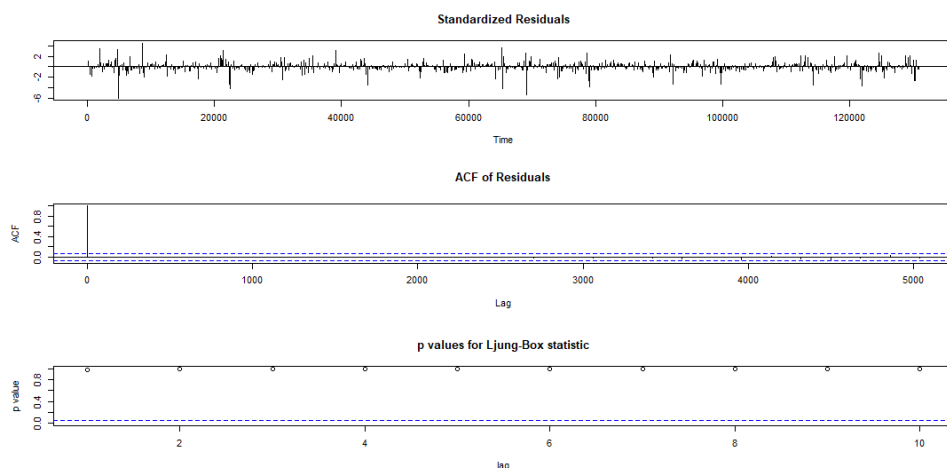
wej, natomiast drugi został zbudowany przy pomocy dostępnej w pakiecie statystycznym R funkcji `autoarima()`, która w sposób automatyczny wybiera najlepszy model bazując na minimalizacji wybranego kryterium AIC AICC (domyślne) lub BIC. Następnie przy pomocy tych kryteriów wybrany zostanie najlepiej dopasowany model.

### 3.2.1 Model ARIMA(20,1,0)

Jako, że ostatnia istotna autokorelacja w wykresie PACF po operacji różnicowania z opóźnieniem 1 została zaobserwowana dla opóźnienia  $h = 20$ , za odpowiedni model możemy uznać model autoregresji  $AR(20)$ , który odpowiada modelowi  $ARIMA(20, 1, 0)$ .

W celu dokonania diagnozy reszt wykorzystana zostanie funkcja `tsdiag()` z pakietu statystycznego R, która zwraca trzy następujące wykresy diagnostyczne: wykres standaryzowanych reszt, wykres funkcji ACF dla reszt oraz p-wartości dla testu Ljunga-Box.

Najważniejszy z nich jest jednak trzeci wykres. Na jego podstawie możemy stwierdzić, że p-wartości dla wszystkich rozważanych opóźnień są znacznie większe niż zakładany poziom istotności, co widzimy również na drugim wykresie, zwracającym funkcję ACF dla reszt. Skłania nas to do przyjęcia hipotezy o braku korelacji reszt.



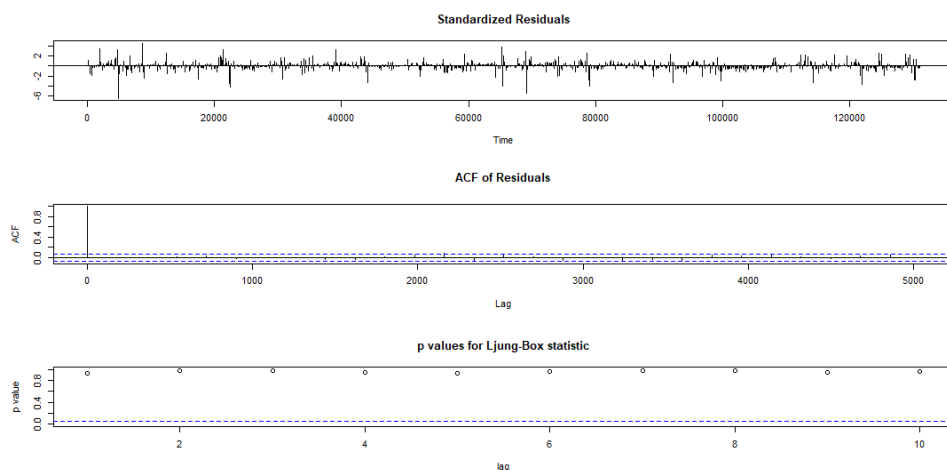
**Rysunek 3.6. Analiza losowości reszt dla modelu ARIMA(20,0,1)**

Źródło: Opracowanie własne.

### 3.2.2 Model ARIMA(2,1,2)

Funkcja `autoarima()` na podstawie minimalizacji kryterium AICC za optymalny wybrała model ARIMA(2,1,2). Wykres tego modelu prezentuje się następująco

Funkcja `tsdiag()` dla tego modelu zwraca następujące wykresy diagnostyczne, wykres standaryzowanych reszt, wykres funkcji ACF dla reszt oraz p-wartości dla testu Ljunga-Boxa:



**Rysunek 3.7. Analiza losowości reszt dla modelu ARIMA(2,1,2)**

Źródło: Opracowanie własne.

Podobnie jak w przypadku poprzedniego modelu możemy przyjąć hipotezę o braku korelacji

reszt, o czymś świadczy wykres funkcji ACF oraz wysokie p-wartości dla testu Ljunga-Boxa dla wszystkich rozważanych opóźnień.

### 3.2.3 Diagnoza dopasowania modelu

Kryteria jakości dopasowania modelu dla obu z nich wyglądają następująco:

**Tabela 3.1. Porównanie jakości dopasowania modelu ARIMA(20,1,0) oraz ARIMA(2,1,2)**

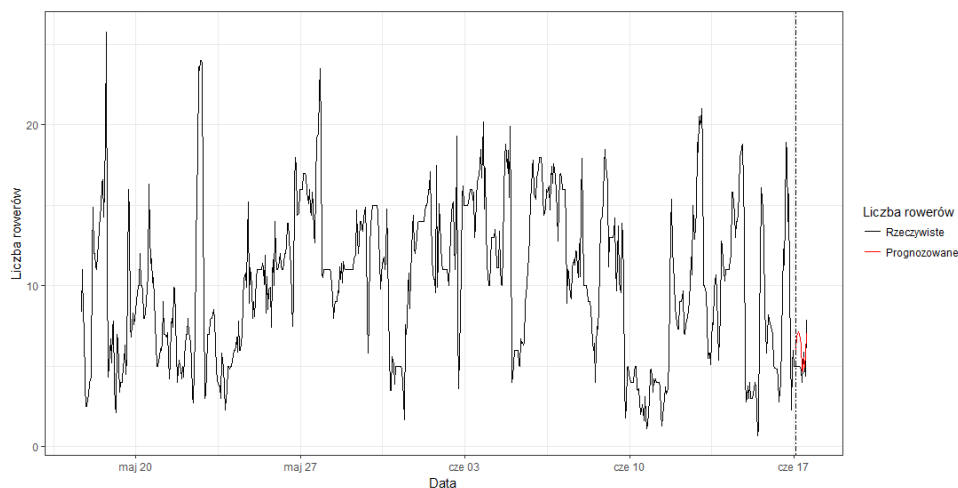
|                            | AIC     | AICc    | BIC     |
|----------------------------|---------|---------|---------|
| <b>Model ARIMA(20,1,0)</b> | 3189.01 | 3190.32 | 3285.41 |
| <b>Model ARIMA(2,1,2)</b>  | 3171.89 | 3171.97 | 3194.84 |

Źródło: Opracowanie własne.

Kierując się zasadą minimalizacji wartości wszystkich kryteriów za lepiej dopasowany model uznajemy ARIMA(2,1,2), którego wyniki w każdym przypadku były niższe.

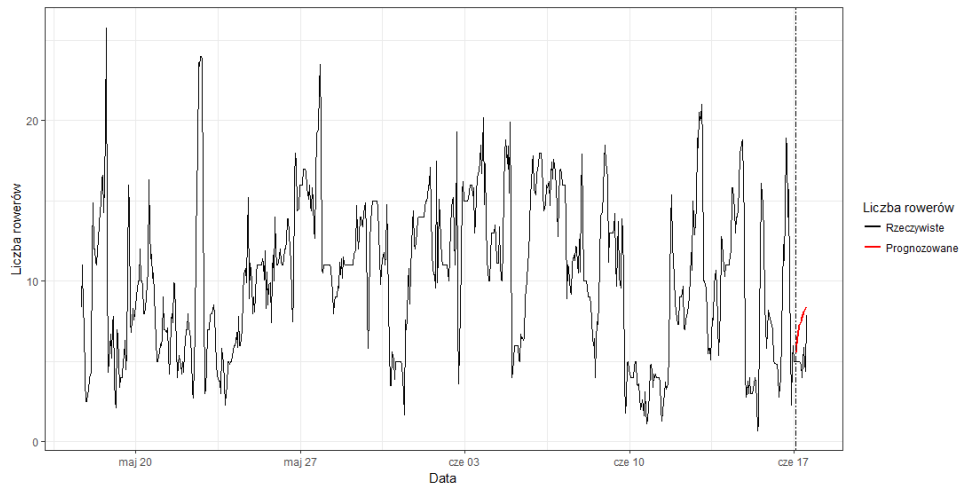
## 3.3 Prognoza

Do konstrukcji prognozy dla zbioru uczącego posłuży nam funkcja `forecast()` znajdująca się w pakiecie `forecast` (Hyndman & Khandakar, 2008) programu statystycznego R. Wykresy prognoz dla obu zbudowanych wcześniej modeli wyglądają następująco:



**Rysunek 3.8. Prognoza dla danych uczących na podstawie modelu ARIMA(20,1,0)**

Źródło: Opracowanie własne.



**Rysunek 3.9. Prognoza dla danych uczących na podstawie modelu ARIMA(2,1,2)**

Źródło: Opracowanie własne.

Prognoza dla pierwszego modelu zakłada wzrost ilości rowerów na stacji w pierwszych okresach, po czym zacznie ona spadać, zrównując się z ich rzeczywistą liczbą. Następnie zgodnie z realnymi danymi zacznie ona rosnąć. Z kolei druga prognoza zapowiada wzrost liczby rowerów we wszystkich prognozowanych okresach.

### 3.3.1 Wybór odpowiedniego modelu

W celu porównania jakości prognoz na podstawie obu tych modeli wykorzystana została funkcja `accuracy()` na zbiorze testowym dla kryteriów MAE i RMSE.

|   |   |
|---|---|
| <code>kryteria &lt;- c("RMSE", 'MAE')</code>                    | 1 |
| <code>accuracy(arima_prognoza2, stacja_test)[, kryteria]</code> | 2 |
| <code>accuracy(arima_prognoza3, stacja_test)[, kryteria]</code> | 3 |

#### Program 3.2. Wykorzystanie funkcji `accuracy()`

Ich wartości prezentują się następująco:

Lepsze wyniki uzyskał pierwszy model wybrany ręcznie – ARIMA(20,1,0). Świadczą o tym niższe wartości obu kryteriów, dlatego to jego uznajemy za dokładniejszy i wykorzystujemy przy konstrukcji prognozy. W jego przypadku na podstawie kryterium RMSE możemy stwierdzić, że dla danych uczących średnia liczba rowerów różniła się o 1,403 z kolei dla danych testowych o 1,182. W przypadku drugiego modelu te liczby to odpowiednio 1,407 oraz 2,22. Na tej podstawie możemy stwierdzić, że model wybrany ręcznie tworzy lepsze prognozy.

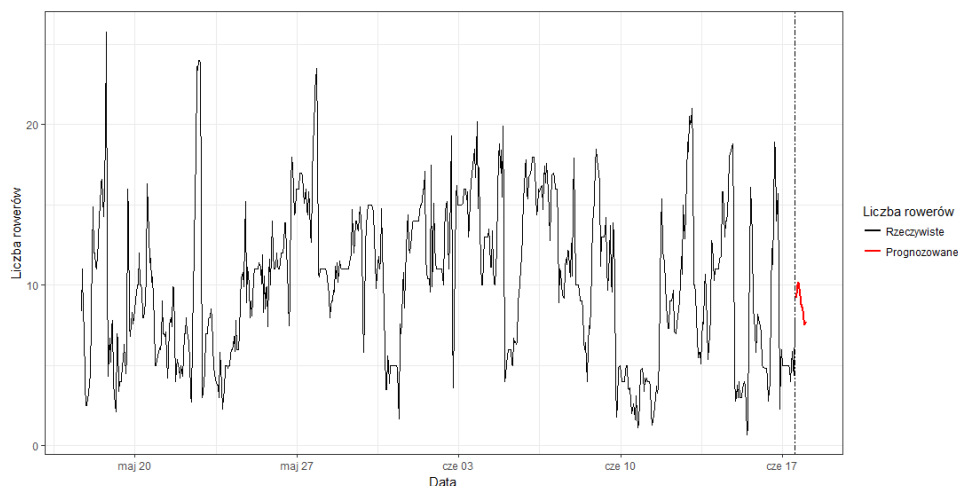
**Tabela 3.2. Kryteria oceny prognozy dla danych uczących i testowych dla modelu ARIMA(20,1,0) oraz ARIMA(2,1,2)**

| ARIMA(20,1,0) | MAE      | RMSE     |
|---------------|----------|----------|
| Dane uczące   | 2.098143 | 1.403489 |
| Dane testowe  | 1.378121 | 1.182307 |
| ARIMA(2,1,2)  | MAE      | RMSE     |
| Dane uczące   | 2.118941 | 1.407470 |
| Dane testowe  | 2.480427 | 2.222693 |

Źródło: Opracowanie własne.

### 3.3.2 Prognoza dla stacji Brama Poznania

Na podstawie wybranego modelu ARIMA(20,1,0) skonstruowana została prognoza na 12 następnych godzin dla stacji Brama Poznania.



**Rysunek 3.10. Prognoza dla całości danych na podstawie modelu ARIMA(20,1,0)**

Źródło: Opracowanie własne.

Według wyznaczonej prognozy w następnych kilku okresach liczba rowerów na stacji Brama Poznania będzie gwałtownie rosła, osiągnie ona szczyt na poziomie 10 pojazdów, po czym ponownie zacznie spadać.

## Podsumowanie

Polskie miasta coraz częściej podążają za trendem Smart Cities wprowadzając technologie związane z tym pojęciem. Podstawowym ich elementem staje się ciągle udoskonalana sieć rowerów miejskich. Celem tej pracy jest prognoza liczby tych pojazdów dla stacji Brama Poznania, znajdującej się nieopodal centrum miasta, co pozwoli na przewidzenie potrzeb użytkowników i możliwość odpowiedniej reakcji przez operatora.

Dzięki otwartym źródłom danych każdy może mieć dostęp do interesujących go informacji. Swoje API posiada również firma NextBike, który w formacie XML zawiera wszystkie podstawowe informacje odnośnie stacji rowerów miejskich w miastach, a ich liczba aktualizowana jest co 5 minut. Pobranie danych na potrzeby badań wymagało napisania odpowiedniego skryptu, który zapisywał aktualizujące się dane.

Przy pomocy narzędzi analitycznych możemy odpowiednio kontrolować wykorzystanie rowerów miejskich, aby zapewnić jak najwyższą jakość ich użytkowania dla mieszkańców. W celu prognozy tego zjawiska na następne 12 godzin zbudowane zostały różne modele prognostyczne, z których najlepsze wyniki wychodziły dla tych z rodziny ARIMA. Wybrany model dobrze dopasowuje się do rzeczywistych danych, a prognozy na nim oparte odpowiednio oddają charakter korzystania z systemu rowerów miejskich przez mieszkańców Poznania.

W dalszych badaniach warto przeprowadzić analizę oraz zbudować modele prognostyczne dla większej liczby stacji, co pozwoli na kompleksową analizę wykorzystania tego środka transportu i jego obsługę przez operatora, w celu zapewnienia jak najwyższej jakości dla jego użytkowników.

# Bibliografia

- Łastowska, A. & Bryniarska, Z. (2015). Ocena systemu bezobsługowej wypożyczalni rowerów miejskich KMK Bike w Krakowie. *Transport Miejski i Regionalny*, (8).
- Bakıcı, T., Almirall, E. & Wareham, J. (2012). A Smart City Initiative: the Case of Barcelona.
- Bryce, K. (2018). *THE MANY BENEFITS OF BIKE SHARING PROGRAMS*.
- Cieslak, M. (2005). *Prognozowanie Gospodarcze, metody i zastosowania* (4th). Warszawa, ul. Miodowa 10: Wydawnictwo naukowe PWN.
- Cohen, B. (2015). *The 3 Generations Of Smart Cities*.
- Cox, D. R. & Stuart, A. (1955). SOME QUICK SIGN TESTS FOR TREND IN LOCATION AND DISPERSION. *Biometrika*, 42(1-2), 80–95. doi:10.1093/biomet/42.1-2.80. eprint: <http://oup.prod.sis.lan/biomet/article-pdf/42/1-2/80/663980/42-1-2-80.pdf>
- Dębowska-Mróż, M., Lis, P., Szymanek, A. & Zawisza, T. (2017). The municipal bicycle as system components transport in cities. *Autobusy : technika, eksploatacja, systemy transportowe*, 18(6), 1173–1182.
- Hyndman, R. J. & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3), 1–22.
- Müller, K., Wickham, H., James, D. A. & Falcon, S. (2018). *RSQLite: 'SQLite' Interface for R*. R package version 2.1.1.
- Poznań, M. (2017). Program Rowerowy Miasta Poznania 2017-2022 z perspektywą do roku 2025.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Rekordy wypożyczeń rowerów miejskich w Poznaniu i Szamotułach*. (2019).
- Shaheen, S. A., Guzman, S. & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record*, 2143(1), 159–167.
- Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.



- Wickham, H. (2018). *httr: Tools for Working with URLs and HTTP*. R package version 1.4.0.
- Wickham, H. (2019). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.3.3.
- Wickham, H., Hester, J. & Ooms, J. (2018). *xml2: Parse XML*. R package version 1.2.0.
- Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O. & Goodman, A. (2014). Health effects of the London bicycle sharing system: health impact modelling study. *BMJ*, 348. doi:10.1136/bmj.g425. eprint: <https://www.bmj.com/content/348/bmj.g425.full.pdf>
- Wroński, M. (2018). *Smart cities w Polsce. Dobre praktyki wciąż niezbyt popularne*.
- Zagdański, A. & Suchwałko, A. (2016). *Analiza i prognozowanie szeregów czasowych. Praktyczne wprowadzenie na podstawie środowiska R*. Wydawnictwo naukowe PWN.
- Zielaś, A. (2003). *Prognozowanie Ekonomiczne*. Wydawnictwo naukowe PWN.

## Spis tabel

|     |   |    |
|-----|---|----|
| 3.1 | Porównanie jakości dopasowania modelu ARIMA(20,1,0) oraz ARIMA(2,1,2) . .                                       | 32 |
| 3.2 | Kryteria oceny prognozy dla danych uczących i testowych dla modelu<br>ARIMA(20,1,0) oraz ARIMA(2,1,2) . . . . . | 34 |

## Spis rysunków

|      |  |    |
|------|--|----|
| 3.1  | Liczba rowerów miejskich na stacji Brama Poznania w dniach 17 maja-17 czerwca                        | 28 |
| 3.2  | Funkcja ACF zbioru uczącego dla stacji Brama Poznania . . . . .                                      | 28 |
| 3.3  | Funkcja PACF zbioru uczącego dla stacji Brama Poznania . . . . .                                     | 29 |
| 3.4  | Funkcja ACF zbioru uczącego po zróżnicowaniu z opóźnieniem 1 dla stacji Brama<br>Poznania . . . . .  | 29 |
| 3.5  | Funkcja PACF zbioru uczącego po zróżnicowaniu z opóźnieniem 1 dla stacji<br>Brama Poznania . . . . . | 30 |
| 3.6  | Analiza losowości reszt dla modelu ARIMA(20,0,1) . . . . .   | 31 |
| 3.7  | Analiza losowości reszt dla modelu ARIMA(2,1,2) . . . . .  | 31 |
| 3.8  | Prognoza dla danych uczących na podstawie modelu ARIMA(20,1,0) . . . . .                             | 32 |
| 3.9  | Prognoza dla danych uczących na podstawie modelu ARIMA(2,1,2) . . . . .                              | 33 |
| 3.10 | Prognoza dla całości danych na podstawie modelu ARIMA(20,1,0) . . . . .                              | 34 |

# Spis Programów

|     |   |    |
|-----|---|----|
| 3.1 | Pobieranie danych z API NextBike oraz zapisywanie w formacie sqllite3 . . . . . | 26 |
| 3.2 | Wykorzystanie funkcji accuracy() . . . . .                                      | 33 |