DE GRUYTER OPEN

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 37 (50) 2014

DOI: 10.2478/slgr-2014-0026

Market Basket Analysis as a Support Tool for The Management of Public Transport

Dorota Sokołowska[1]

[1] Higher Vocational School in Suwalki, Poland, dorotasokolowska@gazeta.pl

Abstract. The aim of this paper is to characterize a non-standard use of the method of market basket analysis in one of the areas of economy, i.e. public transport. Generally, one of the aims of the market basket analysis method is associating the consumer's market basket – in the case of public transport this being the choice of bus stops in the city area made by passengers. Owing to a new, practical use of this method, it was possible to build an efficient model characterizing the movement of flows of public transport passengers, and assess the degree of transferring (changing lines), thus making it possible to adapt the routes of buses to the needs of people using this particular means of transport, as well as to plot new communication lines.

The data analysis was performed using the Statistica statistical package and its SAL application, i.e. the algorithms used in Data Mining.

Keywords: market basket analysis, data mining, SAL, associations and sequences, flows of passengers, decision

1. Introduction

Public transport – its organization and effective management – is a challenge for both transporters and municipal authorities, the trust and satisfaction of the client – a potential citizen – being the primary concern. In times of rapid economic changes, when shifts in population density due to, for example, the construction of new housing estates or "disappearance" of large enterprises replaced by smaller ones, there is a tendency for population centers to concentrate or deconcentrate. A public transport system in use must, on the one hand, ensure citizen safety and be stable, and respond to changes and be adapted to emerging needs, on the other. From the economic point of view, the difficulties that need to be considered in the area of transport have multiple facets, such as the financial aspects (including the distribution of costs, donations, used and refinanced allowances), the aspects of organization arising from demand, or service quality control.

ISBN 978–83–7431–413–8 ISSN 0860-150X

A basis for solving problems in this field is the broadly understood knowledge in the area of employing means essential for transporting passengers in the area of the city and the adjacent communities. The matter of projecting the flow of passengers has already been studied with the use of various methods: the nested logit model (Gangrade, Pendyala, McCullough, 2002, p. 17–26), Markov's chains (Lerman, 1979, p. 273–291), through estimating the value of the expected number of passengers transported yearly by a single communication line (Dytkowski, Gamrot, Tomanek, 2009, p. 75–98), or by studying patterns of passengers' behavior and, perhaps the most importantly, the distribution of travel routes with the use of data collected by electronic ticket vending devices (Navick, Furth, 2002, p. 107–113).

The subject of the discussion presented in this article is the issue of the estimation of flows of passengers using the public transport buses in Białystok. The population of the urban agglomeration is approx. 370,000 persons, of which 295,000 live in the city proper. The city of Białystok ranks fourth in Poland (among cities with district rights) in terms of population density, amounting to 2888 persons/km^2 (Area and Population in the Territorial Profile in 2013, Central Statistical Office). The Białystok Public Transport Company (Białostocka Komunikacja Miejska – BKM) has a fleet of 268 buses, transporting 100,000,000 passengers a year. (Information obtained from the Białystok Public Transport Company). The modern passenger counting systems mounted in the buses, make it possible to generate large databases and allow to perform multi-criteria analyses, which are an essential support tool for decisions connected with the organization of transport in general.

2. Market basket analysis

In order to perform an effective data analysis, one of the most popular association rules, i.e. the market basket analysis, was used. Its underlying idea is to determine the market basket that the customer collects in a single place (e.g. a shop, a company, the exchange) or within a specified time period (e.g. a day, a month, a year). Among the aims of the market basket analysis, the most important ones are (Łapczyński, 2009):

- to offer optimization in order to increase the sale of a product (determination of the market basket purchased by customers);
- to find all the rules containing characteristics of association between products (detecting characteristic traits of behavior of consumers purchasing goods);
- to determine the sequence of purchase of goods.

The market basket analysis belongs to the group of models based on association. Association models take the form of the "if A then B" conditional sentences, which means that there is a co-existence of two sentence elements in a specific order. This sequence is a simple implication, which when used for the market basket analysis, describes the purchase of a specific Commodity A, which implies the purchase of Commodity B with a certain probability.

The basic qualitative measures describing the association rules are the support ratio and the confidence ratio. These measures can be expressed as:

$$\text{support} = P(A \cap A) = \frac{\text{number of transactions containing } A \text{ and } B}{\text{the total number of transactions}}$$

which is the probability of a customer purchasing Products A and B at the same time, and

$$\text{confidence} = P(A/B) = \frac{P(A \cap B)}{P(A)}$$

where $P(A)$ is the number of transactions including A
which is the probability of purchase of Product B among those customers that have already purchased Product A.

An innovative use of the market basket analysis is the analysis of passenger traffic in public transport, where passengers are customers by analogy, while the commodities they choose are the bus stops they use when they start and finish their journeys. Positive results of the use of the market basket analysis are only possible for large data sets. (Harańczyk, 2010). Creating such sets in the era of digitization is not a difficult undertaking. Examples of such sets are product barcodes, addresses, bank card numbers, or bus ticket numbers collected in databases or data banks. The enormity of data may give the impression of chaos at first glance, but for a researcher, this constitutes an endless possibility for gaining knowledge. Data browsing and exploration, called Data Mining, is a relatively new branch of science, rapidly developing together with the development of digitization. (Migut, 2009). Data exploration techniques make it possible to find, or discover, previously unknown correlations and patterns, which may be used to facilitate the decision-making process or to describe a database. By means of the right algorithm, rules, decision trees, or neural nets may be generated.

In the undertaken research, due to the volume of the database and the specific character of the data, association rules available in the SAL module (Sequence, Association and Link Analysis) included in the Statistica v.10

package were used. This application provides the user with a set of computational and graphic tools for mining association rules and visualizing them in the form of graphs.

3. Preliminary data analysis

In order to effectively manage the public transport system and for its effective optimization, readers, whose tasks include counting passengers in an automated manner, were mounted in public transport buses in Białystok. Data is collected continually in the form of user-friendly tables and graphs, and then sent to the person managing the vehicle fleet. Owing to its automated character, the system makes it possible to obtain exact data, characterized by repeatability. The data pertains to, for instance, the structure and number of passengers, duration of a journey, or its starting place, and thus to the flow on individual routes, the bus number, and the driver number.

When applying the market data analysis to investigating flows of passenger movement, two crucial problems appeared. The first one was of technical nature, the other was content-related. Firstly, using the market basket analysis, one has to create a matrix, whose rows correspond to successive passengers, while the columns describe bus stops. Due to the multimillion-record database, such analysis would be extremely difficult. The other problem was the database itself, which was created as the result of recording impulses generated from electronic cards used on entering the bus at the moment the journey started. There is, however, no data that would unambiguously indicate where the passenger finished their journey.

Due to the lack of information concerning the place the journey finishes and considering the size of the database, advanced statistical methods with the use of the latest version of the Statistica (v10) package were employed. Methods of association and sequence analysis (SAL) were especially important in the calculations. This is a type of statistical analysis whose aim is to identify the combination of values of variables occurring with a certain frequency and regularity in a data set. In other words, the SAL application "counts" passengers with the same characteristic features, e.g. those using the same bus stops. In the case of an analysis of urban connections, this means projecting with a certain probability the places where a passenger will appear, knowing where they entered the bus.

Sequence analysis, used to study transferring, also played an important role in the description of flows of public transport passengers. It was also

used to describe directions of movement of passengers riding on public transport buses starting their journey in a specific place by taking into account the chronological order of e-card log-ins in successive locations. Additionally, data at the level of individual bus stops was analyzed, indicating those that were the most popular, and pairs of bus stops at which the same cards were logged-in were pointed out.

For the purpose of the research, information from the counting system of the passengers using the Białystok Public Transport and registering their journeys by means of the e-card was selected. The range of data covered two one-month periods: October 2012 and March 2013. The choice of months for the analysis was not accidental. Its aim was to single out the most representative months that would reflect typical behavior of passengers using the public transport in Białystok – this is why holiday and school break months were excluded. The database contained 3,943,664 records, describing the journeys of 68,097 passengers in October 2012 and 3,418,235 records, describing the journeys of 69,060 passengers in March 2013. For the purpose of the calculations, information such as e-card ID, date and time, and the place (line umber and bus stop name) of registration of each of the journeys was used. An assumption was made that each e-card corresponds to a single passenger using the public transport. The analysis does not contain information concerning the passengers riding on the public transport in Białystok who use allowances or privileges (e.g. pensioners or annuitants), nor persons using traditional (paper) tickets. It is assumed that the proportion of persons not registered by means of the card readers is approx. 50% of all the passengers. (Information obtained from BKM Management). In addition, the city was divided into 36 parts, consistent geographically, administratively, communication-wise. Together with the 6 communes serviced by BKM, the total number of areas, called Regions, was 41. Then, the bus stops located in individual regions were grouped.

Preliminary data investigation made it possible to:

- analyze the number of electronic cards (equivalent to passengers) logged-in in the area of the individual regions serviced by BKM;
- determine the number of passengers moving between pairs of regions;
- identify the most popular connections;
- perform an analysis of the movement of passengers from individual residential areas.

The total numbers of recorded journeys, as well as all e-cards used, in the analyzed months on all lines and all bus stops serviced by BKM are presented in the table below and on the graphs.

Table 1
Number of recorded e-cards and journeys in the analyzed months

Information	Oct 2012	Mar 2013
number of journeys (log-ins)	3,943,664	3,418,235
number of passengers (cards used)	68,097	69,060
number of bus stops (according to recorded log-ins)	836	817
Number of lines (according to recorded log-ins)	39	43

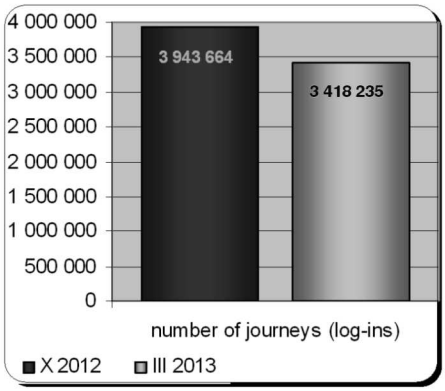


Figure 1. Recorded journeys

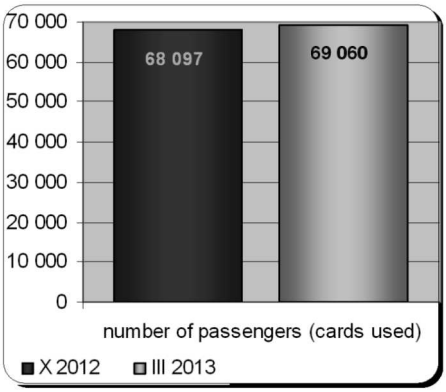


Figure 2. Recorded e-cards

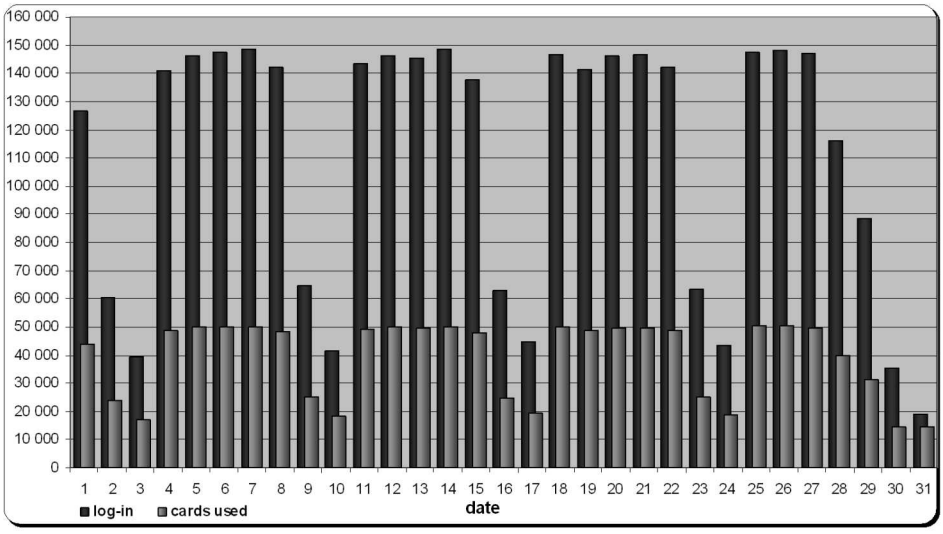


Figure 3. Distribution of recorded e-cards and journeys on individual days of March 2013

When analyzing the data from October 2012 and March 2013, a greater number of e-cards in March 2013, by almost 1000 units (+1.4%), can be observed. Despite the greater number of passengers using the e-card, a smaller number of journeys of these passengers, by over 500,000 (–13.3%), was recorded. However, on the basis of the data from other months, it can be stated that the numbers of journeys of passengers differ only slightly in fact, therefore, all further analysis was based on the data from March 2013, as it is the month with the highest number of persons traveling, but also because it was more recent in comparison with the October of the previous year.

On the basis of the analysis of the public transport bus journeys recorded by means of the e-cards on individual days of March 2013, it can be stated that the numbers of journeys on individual workdays (from Monday to Friday) are similar; as are the numbers of journeys on individual Saturdays and Sundays. A significant difference in the number of recorded e-cards and journeys (log-ins) can be noticed, on the other hand, between workdays and Saturdays and Sundays.

4. Analysis of passenger traffic between Regions

Due to the small difference between the number of journeys in successive workdays in March, shown above, a detailed analysis of passenger movement on a single workday allows to estimate the whole structure of traffic from Monday do Friday. An assumption was made that the analyzed day will be March 7, 2013, owing to the highest number of journeys out of all the days of this month. All the data below pertains to this chosen day.

The system of data concerning BKM passengers using e-cards allows to accurately ascribe each log-in (on entering the bus) in the system to individual bus stops, whereas ascribing bus stops to the appointed Regions (parts of the city and the surrounding communes) makes it possible, in turn, to ascribe all e-card log-ins to them.

5. Daily data from individual Regions

Using the Statistica package and by ascribing bus stops to individual Regions, daily numbers of passengers using e-cards recording their journeys in the area of each of the Regions were determined.

a

Table 2
Daily number of log-ins (passengers) in the area of individual Regions. (The names of Regions are left in their original form)

Region	Number of passengers
Centrum Wschód	17 500
Wiejska_Kawaleryjska	8 029
Centrum Zachód	7 430
Piaski	6 430
Wysoki Stoczek	6 293
Osiedle Młodych	6 014
Sienkiewicza_Wasilkowska	5 990
Sady Antoniukowskie	5 571
PKP_PKS	4 944
Branickiego_Warszawska	4 496
Dziesięciny	4 367
Piasta	4 056
Zielone Wzgórza	3 895
Mickiewicza_urzędy	3 787
Nowe Miasto	3 513
Słoneczny Stok	3 281
Bażantarnia	3 164
TBS	2 933
Kawaleryjska_Kleosin	2 690
Białostoczek	2 489
Skorupy_Przemysłowe	2 420
Leśna Dolina	2 073
Dojlidy	2 027
Starosielce	1 802
Wygoda_Chełmońskiego	1 684
Pietrasze	1 407
Jaroszówka_Wyżyny	1 198
Białostoczek przemysłowy	1 113
DOBRZYNIEWO	821
Bacieczki	788
Dojlidy Górne	612
WASILKÓW	587
SUPRAŚL	585
Starosielce przemysłowe	579
Jurowiecka_Poleska_BohGetta	493
CHOROSZCZ	410
Pieczurki	404
JUCHNOWIEC	391
Mickiewicza_osiedle	369
Zawady	335
ZABŁUDÓW	29

6. Flows of passengers and determining the most popular pairs of Regions

When entering the bus, each of the passengers must register their ride; when they leave the bus, they are not obliged to do the same. In this study, an assumption was formulated whereby the finishing point of a journey was the same place where the return journey of a specific passenger started (recorded by means of the same e-card). Based on the association of a specific e-card number, the numbers of passengers traveling between two points can be counted. It must be noticed, however, that the analysis pertains to a single day only, thus the passengers who started their journey on the 7th of March and returned on the following day (or the other way around – finished their journey in the analyzed day) are excluded from the study for technical reasons, i.e. the lack of a "pair" for the e-card number that could be associated with it. The number of such passengers is about 7,000, which amounts to approximately 10% of those traveling on the studied day.

Hence, with the use of the association analysis (in the investigated set of passenger log-ins in individual Regions of the city), the probability of where a passenger will appear, knowing where they entered the bus could be determined. Converting the probability of a passenger traveling in a certain direction into the number of passengers, an estimated number of these passengers traveling between two specific locations in the course of a day can be determined.

Using the aforementioned method, the numbers of passengers traveling between the assigned regions were determined, without taking into account their order.

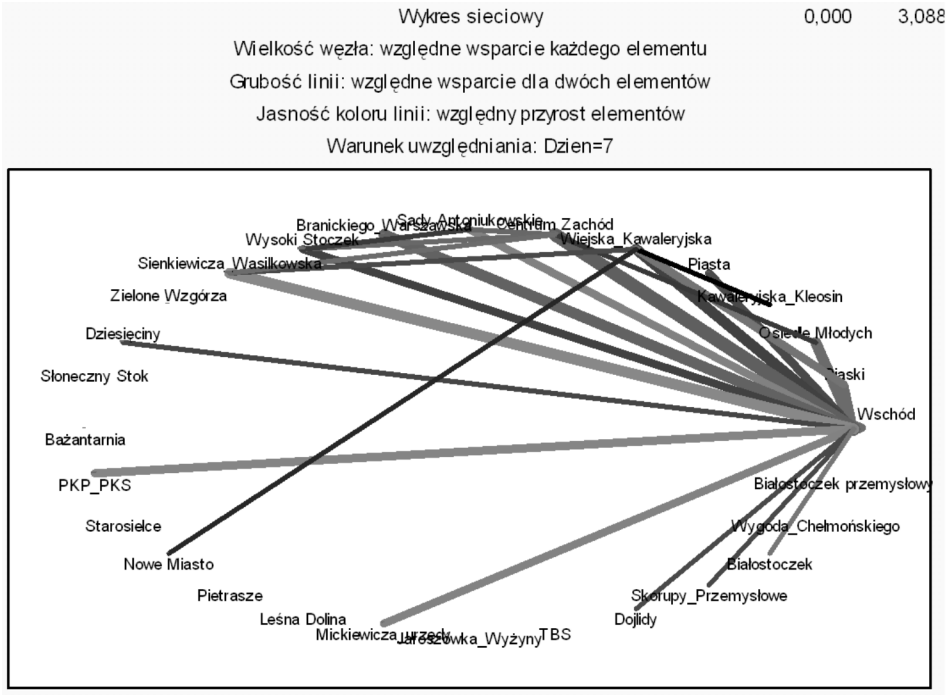
Table 3

The most popular pairs of Regions

Region	Region	Number of passengers
Centrum Wschód	Centrum Zachód	2 831
Centrum Wschód	Sienkiewicza_Wasilkowska	2 245
Centrum Wschód	Wiejska_Kawaleryjska	2 162
Branickiego_Warszawska	Centrum Wschód	2 033
Centrum Wschód	Osiedle Młodych	1 935
Centrum Wschód	Piaski	1 907
Centrum Wschód	Piasta	1 839
Centrum Wschód	Mickiewicza_urzędy	1 754
Centrum Wschód	Wysoki Stoczek	1 721

Region	Region	Number of passengers
Centrum Wschód	Sady Antoniukowskie	1 693
Wiejska_Kawaleryjska	Piaski	1 616
Centrum Wschód	PKP_PKS	1 460
Białostoczek	Centrum Wschód	1 363
Centrum Wschód	Skorupy_Przemysłowe	1 351
Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	1 330
Centrum Wschód	Dziesięciny	1 230
Centrum Zachód	Sienkiewicza_Wasilkowska	1 202
Centrum Wschód	Dojlidy	1 157
Sady Antoniukowskie	Wysoki Stoczek	1 145
Centrum Zachód	Osiedle Młodych	1 117
Nowe Miasto	Wiejska_Kawaleryjska	1 085
Centrum Zachód	Sady Antoniukowskie	1 059
Centrum Zachód	Wysoki Stoczek	1 028
Sienkiewicza_Wasilkowska	Wiejska_Kawaleryjska	1 004
Centrum Wschód	Słoneczny Stok	977
Centrum Wschód	Zielone Wzgórza	928
Dziesięciny	Sady Antoniukowskie	912
Wysoki Stoczek	TBS	911
Centrum Wschód	TBS	906
Centrum Wschód	Nowe Miasto	868
Bażantarnia	Wiejska_Kawaleryjska	849
Osiedle Młodych	Wysoki Stoczek	830
Dziesięciny	Wysoki Stoczek	822
Branickiego_Warszawska	Centrum Zachód	792
Osiedle Młodych	Słoneczny Stok	781
Osiedle Młodych	PKP_PKS	751
Centrum Zachód	Piaski	743
Sienkiewicza_Wasilkowska	Piaski	732
Centrum Wschód	Wygoda_Chełmońskiego	729
Centrum Wschód	Leśna Dolina	719
Centrum Zachód	PKP_PKS	688
Osiedle Młodych	Sady Antoniukowskie	687
Branickiego_Warszawska	Skorupy_Przemysłowe	685
Centrum Wschód	Jaroszówka_Wyżyny	668
Bażantarnia	Centrum Wschód	666
Osiedle Młodych	Zielone Wzgórza	663
PKP_PKS	Wiejska_Kawaleryjska	645
Dziesięciny	Sienkiewicza_Wasilkowska	630
Centrum Zachód	Mickiewicza_urzędy	628

Using the package, the data from Table 3 was visualized on a net graph.



[Net graph]
[Node size: relative support for each element]
[Line thickness: relative support for two elements]
[Brightness of line color: relative increment of elements]
[Condition for inclusion: Day=7th]

Figure 4. The largest flows of passengers between Regions

Lack of connection between Regions indicates the number of passengers below the assumed support level.

As shown by the tables and the graph above, among the most popular directions of movement of passengers, the Region specified as Centrum Wschód is the dominant one. It is safe to assume that this place plays the role of the city's "transfer centre". Paths of most of the passengers using BKM services cross at this location. This is even more visible when the main flows of passengers are plotted on the map of Białystok. (The map was used with permission of the Białystok Municipal Office).

In order to better assess traffic distribution outside the dominant place, an analysis identical to the earlier one was performed, this time with the dominant point excluded. Using the Statistica package, the numbers of passengers traveling between the appointed regions (with the exclusion of the Centrum Wschód Region) were determined.

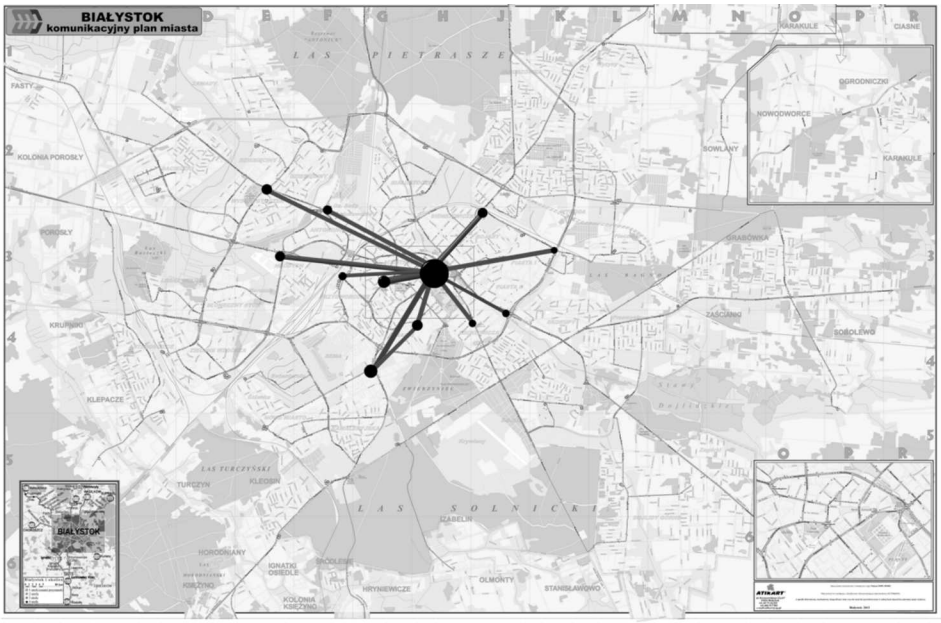


Figure 05. Main flows of passengers in Białystok

Table 4
The most popular pairs of Regions (without the Centrum Wschód Region)

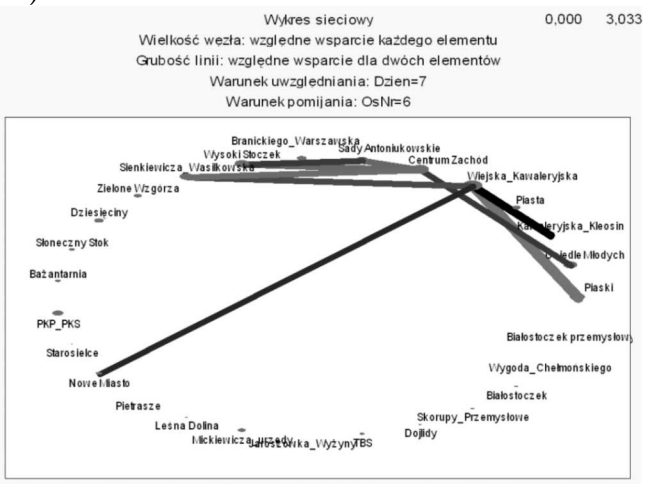
Region	Region	Number of passengers
Wiejska_Kawaleryjska	Piaski	1 616
Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	1 330
Centrum Zachód	Sienkiewicza_Wasilkowska	1 202
Sady Antoniukowskie	Wysoki Stoczek	1 145
Centrum Zachód	Osiedle Młodych	1 117
Nowe Miasto	Wiejska_Kawaleryjska	1 085
Centrum Zachód	Sady Antoniukowskie	1 059
Centrum Zachód	Wysoki Stoczek	1 028
Sienkiewicza_Wasilkowska	Wiejska_Kawaleryjska	1 004
Dziesięciny	Sady Antoniukowskie	912
Wysoki Stoczek	TBS	911
Bażantarnia	Wiejska_Kawaleryjska	849
Osiedle Młodych	Wysoki Stoczek	830
Dziesięciny	Wysoki Stoczek	822
Branickiego_Warszawska	Centrum Zachód	792
Osiedle Młodych	Słoneczny Stok	781

Region	Region	Number of passengers
Osiedle Młodych	PKP_PKS	751
Centrum Zachód	Piaski	743
Sienkiewicza_Wasilkowska	Piaski	732
Centrum Zachód	PKP_PKS	688
Osiedle Młodych	Sady Antoniukowskie	687
Branickiego_Warszawska	Skorupy_Przemysłowe	685
Osiedle Młodych	Zielone Wzgórza	663
PKP_PKS	Wiejska_Kawaleryjska	645
Dziesięciny	Sienkiewicza_Wasilkowska	630
Centrum Zachód	Mickiewicza_urzędy	628
Kawaleryjska_Kleosin	Piaski	616
Sady Antoniukowskie	Wiejska_Kawaleryjska	600
Centrum Zachód	Wiejska_Kawaleryjska	594
Nowe Miasto	Piaski	593
Branickiego_Warszawska	Osiedle Młodych	592
Sienkiewicza_Wasilkowska	Wygoda_Chełmońskiego	582
Osiedle Młodych	Piaski	580
Bażantarnia	Osiedle Młodych	559
Osiedle Młodych	Bażantarnia	559
PKP_PKS	Wysoki Stoczek	548
Centrum Zachód	Piasta	547
Pietrasze	Sienkiewicza_Wasilkowska	542
PKP_PKS	Sady Antoniukowskie	541
Centrum Zachód	Pietrasze	537
Sienkiewicza_Wasilkowska	Wysoki Stoczek	535
Sady Antoniukowskie	Piaski	527
PKP_PKS	Sienkiewicza_Wasilkowska	509
Centrum Zachód	Zielone Wzgórza	496
Mickiewicza_urzędy	Piaski	488
PKP_PKS	Piaski	485
Centrum Zachód	Dziesięciny	482
Osiedle Młodych	TBS	482
Wiejska_Kawaleryjska	Wysoki Stoczek	480
Sady Antoniukowskie	Sienkiewicza_Wasilkowska	474

In order to visualize the main flows of passengers on net graphs in a legible manner, support estimate values of 2% and 1.5% were used.

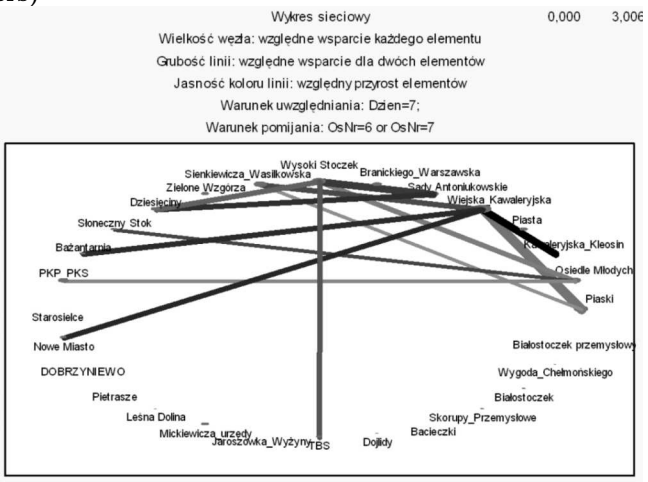
An analysis of data with the exclusion of the "transfer" Region of Centrum Wschód showed that the greatest number of passenger journeys to other Regions are started from: Wiejska_Kawaleryjska, Centrum_Zachód, Sienkiewicza_Wasilkowska, and Sady Antoniukowskie.

a) a journey of at least 2% of the total number of passengers (approx. 1000 passengers)



[Net graph]
[Node size: relative support for each element]
[Line thickness: relative support for two elements]
[Condition for inclusion: Day=7th]

b) a journey of at least 1,5% of the total number of passengers (approx. 780 passengers)



[Net graph]
[Node size: relative support for each element]
[Line thickness: relative support for two elements]
[Brightness of line color: relative increment of elements]
[Condition for inclusion: Day=7th]

Figure 6. The largest flows of passengers between Regions (without Centrum Wschód)

7. Passenger traffic sequence analysis – a study of transferring

Sequence analysis, unlike association analysis, allows to analyze, in addition to their number, the directions of movement of passengers, by taking into account the chronological order of e-card log-ins in successive locations on the investigated day.

An example visualization of the direction of the flow of passengers from the Centrum Wschód Region (the middle part of the graph) is presented in the radar graph below.

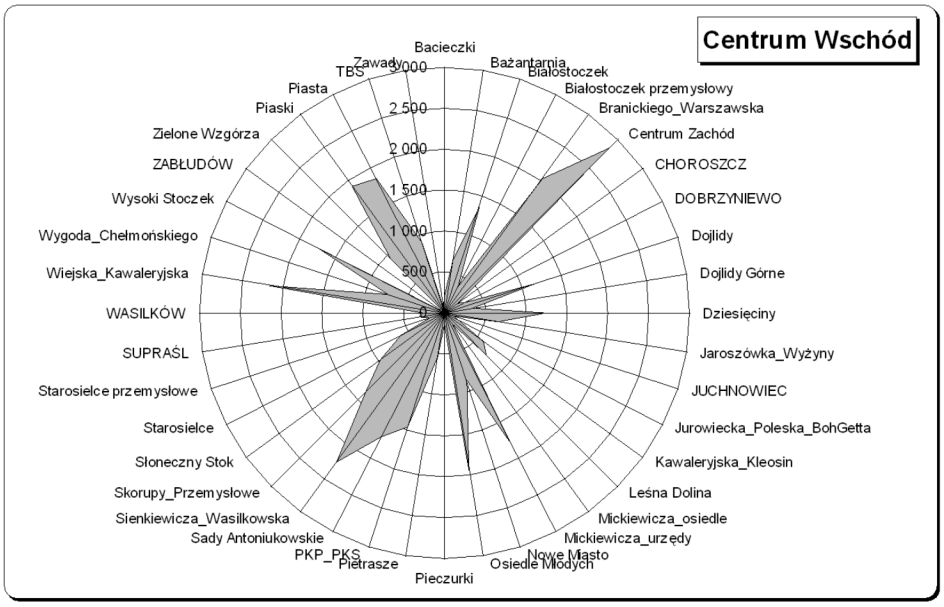


Figure 7. Numbers of passengers traveling from the Centrum Wschód Region

As can be noticed, Centrum Wschód is an important point on the public transport map. This is the result of the radial arrangement of the city on the one hand, and of thus organized network of connections of bus communication on the other.

Using sequence analysis in the Statistica package, an investigation of the movement of passengers between two, three and four successive locations respectively was performed, compiled on the example spreadsheet.

Table 5

E-card log-ins in two successive Regions
(example numbers from an Excel spreadsheet, sorted in descending order)

Region 1	Region 2	Number of passengers
Sienkiewicza_Wasilkowska	Centrum Wschód	1 671
Centrum Wschód	Centrum Zachód	1 652
Wiejska_Kawaleryjska	Centrum Wschód	1 609
Centrum Zachód	Centrum Wschód	1 556
Piasta	Centrum Wschód	1 481
Branickiego_Warszawska	Centrum Wschód	1 415
Wysoki Stoczek	Centrum Wschód	1 318
Piaski	Centrum Wschód	1 291
Mickiewicza_urzędy	Centrum Wschód	1 172
Osiedle Młodych	Centrum Wschód	1 144
Białostoczek	Centrum Wschód	1 123
Centrum Wschód	Osiedle Młodych	1 094
Sady Antoniukowskie	Centrum Wschód	1 090
Skorupy_Przemysłowe	Centrum Wschód	1 084
Dziesięciny	Centrum Wschód	1 050
Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	1 029
Centrum Wschód	Branickiego_Warszawska	1 025
Wiejska_Kawaleryjska	Piaski	1 021
Centrum Wschód	Sienkiewicza_Wasilkowska	979
Centrum Wschód	Wiejska_Kawaleryjska	974

Table 6

E-card log-ins in three successive Regions
(example numbers from an Excel spreadsheet, sorted in descending order)

Region 1	Region 2	Region 3	Number of passengers
Centrum Wschód	Centrum Wschód	Centrum Zachód	269
Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	268
Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	265
Wysoki Stoczek	Wysoki Stoczek	Wysoki Stoczek	262
Wiejska_Kawaleryjska	Centrum Wschód	Wiejska_Kawaleryjska	259
Wiejska_Kawaleryjska	Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	253
Wiejska_Kawaleryjska	Piaski	Wiejska_Kawaleryjska	247
Sienkiewicza_Wasilkowska	Centrum Wschód	Centrum Zachód	235
Piasta	Centrum Wschód	Piasta	226
Sienkiewicza_Wasilkowska	Centrum Wschód	Sienkiewicza_Wasilkowska	219
Centrum Wschód	Centrum Wschód	Sienkiewicza_Wasilkowska	204
Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	Piaski	189
Centrum Wschód	Centrum Wschód	Branickiego_Warszawska	179
Nowe Miasto	Wiejska_Kawaleryjska	Nowe Miasto	172
Centrum Wschód	Centrum Wschód	Mickiewicza_urzędy	171
Centrum Wschód	Centrum Wschód	Osiedle Młodych	169
Wysoki Stoczek	Centrum Wschód	Wysoki Stoczek	168
Centrum Wschód	Centrum Wschód	Wiejska_Kawaleryjska	162
Centrum Wschód	Centrum Wschód	Sady Antoniukowskie	157

Table 7

E-card log-ins in four successive Regions (example numbers from an Excel spreadsheet, sorted in descending order)

Region 1	Region 2	Region 3	Region 4	Number of passengers
Wysoki Stoczek	Wysoki Stoczek	Wysoki Stoczek	Wysoki Stoczek	53
Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	Kawaleryjska_Kleosin	49
Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	Wiejska_Kawaleryjska	Piaski	40
Sienkiewicza_Wasilkowska	Centrum Wschód	Centrum Zachód	Sienkiewicza_Wasilkowska	37
Sienkiewicza_Wasilkowska	Centrum Wschód	Sienkiewicza_Wasilkowska	Centrum Zachód	36
Piasta	Centrum Wschód	Piasta	Piasta	35
Słoneczny Stok	Osiedle Młodych	Słoneczny Stok	Osiedle Młodych	31
Dziesięciny	Centrum Wschód	Centrum Zachód	Sady Antoniukowskie	30
Leśna Dolina	Bażantarnia	Wiejska_Kawaleryjska	Bażantarnia	30
TBS	Wysoki Stoczek	Wysoki Stoczek	Wysoki Stoczek	30
Wiejska_Kawaleryjska	Piaski	Wiejska_Kawaleryjska	Piaski	30
Nowe Miasto	Piaski	Wiejska_Kawaleryjska	Nowe Miasto	29
Wiejska_Kawaleryjska	Kawaleryjska_Kleosin	Wiejska_Kawaleryjska	Piaski	29
Sienkiewicza_Wasilkowska	Centrum Wschód	Sienkiewicza_Wasilkowska	Sienkiewicza_Wasilkowska	28
Wysoki Stoczek	Centrum Wschód	Wysoki Stoczek	Wysoki Stoczek	28
Centrum Wschód	Centrum Wschód	Mickiewicza_urzędy	Centrum Zachód	27

An example of the use of sequence analysis for determining flows of passengers traveling from a specific location, in this case this is the Bacieczki Region.

Table 8

Example distribution of directions of movement of passengers

Region 1	Region 2	Region 3	Region 4	Number of passengers
Bacieczki				788
Bacieczki	Pietrasze			12
Bacieczki	Pietrasze	Centrum Wschód		5
Bacieczki	Pietrasze	Centrum Wschód	Skorupy_Przemysłowe	2
Bacieczki	Pietrasze	Centrum Zachód		5
Bacieczki	Pietrasze	Centrum Zachód	Centrum Wschód	2
Bacieczki	Pietrasze	Centrum Zachód	Skorupy_Przemysłowe	2
Bacieczki	PKP_PKS			34
Bacieczki	PKP_PKS	Branickiego_Warszawska		2
Bacieczki	PKP_PKS	Centrum Wschód		8
Bacieczki	PKP_PKS	Centrum Wschód	Centrum Zachód	2
Bacieczki	PKP_PKS	Centrum Zachód		2
Bacieczki	PKP_PKS	Osiedle Młodych		3
Bacieczki	PKP_PKS	Sady Antoniukowskie		7
Bacieczki	PKP_PKS	Wysoki Stoczek		4

An analysis of the data from the table above (which is a fragment of a complete compilation consisting of over 46,000 records) allows to make an observation concerning passenger traffic. The above example describes a fragment of the flow of passengers traveling from the Bacieczki Region in the direction of Pietrasze, the train station and the bus station. Out of the 788 passengers logging-in in Bacieczki, 12 went in the direction of Pietrasze, out of whom 5 went to Centrum Wschód (out of whom 2 persons went to Skorupy Przemysłowe), another 5 to Centrum Zachód (out of whom 2 persons to Centrum Wschód and 2 to Skorupy Przemysłowe); 34 passengers went from Bacieczki in the direction of the stations, out of whom 2 persons went further in the direction of Branickiego_Warszawska, 8 persons went in the direction of Centrum Wschód (out of whom 2 then went to Centrum Zachód), 2 persons went directly to Centrum Zachód, 3 persons went to Osiedle Młodych, etc.).

8. Conclusions

The organization and management of transport in municipal agglomerations is a huge challenge for the city officials and nowadays it requires knowledge gained by means of modern technologies and innovative uses of methods from various branches of science – statistical knowledge being the key area. Research in the field of public transport management is conducted all over the world and published in thousands of publications.

Despite the fact that the analysis was limited to passengers with electronic tickets (they could be called BKM's regular clients), the size of the database allows to determine the directions of movement of Białystok citizens with a high probability.

The city's public transport is arranged radially with a single transfer center, so it is possible to trace new bus routes in such a way that they would bypass the city center, which would allow to avoid traffic concentration in the center of the city and would certainly reduce the duration of a direct journey between selected Regions.

Knowing also the numbers of passengers and the main directions of flows of citizen movement, a possible construction of new roads, or a renovation of old ones, can be planned in order to improve the movement of BKM's fleet.

By selecting the most popular pairs of bus stops and thus the highest number of persons traveling between specific bus stops, new lines could be opened, or express buses could be arranged on the existing lines.

The conducted analysis may serve as a support tool for the management of a public transport network, which, in addition to the aforementioned conclusions, would yield benefits in the form of an improvement of the comfort of traveling, or a reduction of the costs of management and energy consumption. This would also have a positive effect on environment protection and increase the competitiveness of bus transport over private car transport.

REFERENCES

Area and Population in the Territorial Profile in 2013. *Statistical Information and Elaborations* (2013), Warszawa: Central Statistical Office. Publication available on CD-ROM and at website – www.stat.gov.pl

Dytkowski, G., Gamrot, W., Tomanek, R. (2009). Wykorzystanie metod statystycznych w badaniu popytu na usługi transportu miejskiego. *Prace Naukowe Akademii Ekonomicznej im. Karola Adamieckiego.* Katowice: Wydawnictwo Akademii Ekonomicznej w Katowicach.

Gangrade, S., Pendyala, R. M., McCullough, R. G. (2002). A Nested Logit Model of Commuters' Activity Schedules, *Journal of Transportation and Statistics*, No 2/3. Bureau of Transportation Statistics Research and Innovative Technology Administration.

Harańczyk, G. (2010). Zastosowanie technik Data Mining w badaniach naukowych. Kraków: StatSoft Polska, www.statsoft.pl/czytelnia.html

Lerman, S. R. (1979). The use of Disaggregate Choice Models in Semi-Markov Process Models of Trip Chaining Behavior. *Transportation Science*, No 13. Massachusetts: Massachusetts Institute of Technology.

Linoff, G., Berry, M. (2011), Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Indiana: Wiley Publishing, Inc., Indianapolis.

Łapczyński, M. (2009). Analiza koszykowa i analiza sekwencji – wielki brat czuwa Kraków: StatSoft Polska, www.statsoft.pl/czytelnia.html.

Migut, G. (2009). Czy stosowanie metod Data Mining może przynieść korzyści w badaniach naukowych? Kraków: StatSoft Polska, www.statsoft.pl/czytelnia.html.

Navick, D. S., Furth, P. G (2002). Estimating passengers miles, origin – destination patterns, and loads with location – stamped farebox data. *Transportation Research Record.* No 1799.

Zagóżdżon, B. (2003). Komunikacja miejska jako element sektora publicznego, w: *Liberalizacja transportu w warunkach transformacji gospodarczej*, praca zbiorowa pod red. G. Dydkowskiego i R. Tomanka, *Prace Naukowe Akademii Ekonomicznej im. Karola Adamieckiego*, Katowice: Wydawn. Akad. Ekonomicznej im. Karola Adamieckiego w Katowicach.

http://www.stat.gov.pl/cps/rde/xbcr/gus/l_powierzchnia_i_ludnosc_przekroj_terytorialny_2013.pdf.