Abstract:

**Purpose:** Determining the profile and the relationship between the fan's demographic profile and the way of traveling to the stadium.

**Design/Methodology/Approach:** To establish a relationship between qualitative data that allows a feature to be ordered according to its strength of that feature, the best method is to determine the Spearman's rank correlation coefficient. The use of this coefficient is the most appropriate choice, as the correlation (interdependence of qualitative characteristics) concerns the description of the transport behaviour of passengers commuting to the stadium using various means of transport. Moreover, by analyzing the relationships between these variables by the correspondence method, a similar profile of passengers using a given means of transport is obtained.

**Findings:** The analysis of the presented data in the figures and tables that the analyzed features do not show a large correlation on the way fans commute to the football stadium.

**Originality/Value:** The originality of the article is related to the innovative approach to the issues affecting the management of the offer of transport services - the use of the Spearman's rank correlation method and correspondence analysis in the research of passengers commuting to the football stadium.

**Keywords:** Polsat Plus Arena Gdańsk Stadium, public transport, football fans, football match.

**JEL codes:** O18.

**Paper type:** Research article.
1. Introduction

As a result of the COVID-19 pandemic, in the spring of 2020 the stadiums were closed to football fans. Partial abolishment of the restrictions took place in the summer of 2020. From June 19, 2020, fans could take up to 25% of seats in football stadiums, and football fans could take a seat in such a way that there was no grouping. At the same time, the organizer of the event was obliged to apply appropriate rules regarding disinfection, purchase of tickets and cooperation with transport operators in the field of transport to the stadium (Hassanzadeh-Rad and Halabchi, 2020).

However, due to the spread of the pandemic, in the autumn of 2020, the Polish government closed the stadiums to spectators. From May 15, 2021, stadium operators were able to make 25% of the stadium's capacity available, and from May 28, 2021 - 50%. From March 1, 2022, the number of spectators equal to 100% of the stadium's capacity may take part in mass events at the stadium.

Based on such a government decision, the organizer of the football event will be able to organize events without additional restrictions (McCarrick, Bilalic, Neave, and Wolfson, 2021). However, due to the small number of meetings with fans, no marketing research was conducted during the pandemic (García Bulle, Shen, Shah, and Hosoi, 2022).

Based on the above conditions, a research problem was formulated concerning the premises for the selection of public transport by fans visiting football stadiums in Poland. The stadium in Gdańsk was chosen as the case study. For the purposes of the research, a hypothesis was adopted that the fan's profile, age, sex, education and place of residence do not affect the choice of the means of transport.

For the purposes of the article, the ways in which passengers commute to the football stadium in Gdańsk and the profile of passengers using the services of the transport system were examined. The research problem presented in this way is particularly important in the situation of managing the offer of transport services performed in a decisive manner by public entities.

2. Literature Review

The need, and especially the desire to watch and support their favorite team, make some fans able to make a long journey, just to enjoy and cheer their fans (Green and Chalip, 1998; Osti, Cicero, and Moreschini, 2018; Jones, 2008; Fairley, Gibson, and Lamont, 2018). The decision-making by fans related to walking distances has become the subject of research on sports tourism by, inter alia, (Mirehie and Gibson 2020; Yang, Zhang, and Rickly, 2021; Lebrun, Su, and Bouchet, 2021; Perić, Đurkin, and Vitezić, 2018) or other authors. Of course, scientists also study scattered phenomena such as, the social impact of sports events (Chalip, 2006), aspects of
shopping among sports fans (Smith and Stewart, 2007), or maybe also the motivation to implement sports travel (McCartney, 2005; Weed, 2012) or changes in the approach to sports tourism (Higham and Hinch, 2002; Weed, 2008; Rodrigues, Valdunciel, and Miguel-Dávila, 2014).

In this context, a literature review was carried out on the preferences of residents travelling to a specific point. This assumption is important because fans who want to watch sports events have to travel home-stage in spatial terms: city, regional or even national. However, based on the research carried out, by far the largest group are fans living in the agglomeration where the stadium is located (Herold, Schulenkorf, Breitbarth, and Bongiovanni, 2021).

(Soza-Parra, Raveau, and Muñoz, 2021; 2022) analyses the transport preferences of city residents, finding that in a frequency-based system (without timetables), even with perfect regularity, users are not sure when the next vehicle will arrive. The analysis of passengers' attitudes regarding the choice of various means of transport was carried out by (Ortúzar and Willumsen, 2011), they indicated the time spent in the vehicle, waiting time, fare, availability. In turn (Carrion and Levinson, 2010; Kouwenhoven, de Jong, Koster, van den Berg, Verhoef, Bates, and Warffemius, 2014) observed that passengers choosing public means of transport also take into account the variability of time attributes. On the other hand (Allen, Browne, Woodburn, and Leonardi, 2012; Soza-Parra, Raveau, Muñoz, and Cats, 2019), in their research, assessed the time in relation to the declared satisfaction with the public transport service.

While reviewing the literature, a gap was found in transport research on the behaviour and expectations of passengers going to football events.

### 3. Research Methodology

The research focused on the behaviour of passengers using public transport on their way to the Polsat Plus Arena Gdańsk football stadium. It is the fifth stadium infrastructure facility in terms of capacity for fans - 41,620 (Polsat Plus Arena Gdańsk 2022). Lechia Gdańsk is playing its games as part of Top League (Ekstraklasa). In the 2018/19 season, the average number of spectators at the stadium was 14,264, while in the 2020/21 season - 987 people.

Based on the research of fans during one of the football matches in the 2018/19 season, a study was carried out on the way of arriving at 472 fans travelling to Lechia Gdańsk football matches at the Polsat Plus Arena Gdańsk stadium.

To establish the relationship between what factors affect and to what extent Spearman’s rank correlation and correspondence analysis were used. The use of the Spearman coefficient is most appropriate as it is used to describe the strength of several measurable features.
**Figure 1. The number of fans at Lechia Gdańsk matches in the 2018/19 season.**

*Source: Ekstraklasa 2022.*

**Table 1. The structure of passengers going to the stadium by various means of transport in %**

<table>
<thead>
<tr>
<th>Factor</th>
<th>2018/19 season</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age:</td>
<td></td>
</tr>
<tr>
<td>Up to 24</td>
<td>14.80</td>
</tr>
<tr>
<td>25-34</td>
<td>24.00</td>
</tr>
<tr>
<td>35-44</td>
<td>32.60</td>
</tr>
<tr>
<td>45-54</td>
<td>13.00</td>
</tr>
<tr>
<td>55-64</td>
<td>8.40</td>
</tr>
<tr>
<td>64 and more</td>
<td>7.20</td>
</tr>
<tr>
<td>Sex:</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>17.58</td>
</tr>
<tr>
<td>male</td>
<td>82.42</td>
</tr>
<tr>
<td>Transport systems:</td>
<td></td>
</tr>
<tr>
<td>car</td>
<td>32.60</td>
</tr>
<tr>
<td>taxi</td>
<td>6.40</td>
</tr>
<tr>
<td>suburban train</td>
<td>24.80</td>
</tr>
<tr>
<td>bus, tram</td>
<td>24.20</td>
</tr>
<tr>
<td>other</td>
<td>12.00</td>
</tr>
<tr>
<td>Zoner:</td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>31.35</td>
</tr>
<tr>
<td>yellow</td>
<td>24.15</td>
</tr>
<tr>
<td>red</td>
<td>3.82</td>
</tr>
<tr>
<td>blue</td>
<td>40.68</td>
</tr>
<tr>
<td>Accommodation:</td>
<td></td>
</tr>
<tr>
<td>Tricity</td>
<td>62.92</td>
</tr>
<tr>
<td>Pomeranian</td>
<td>13.34</td>
</tr>
<tr>
<td>Poland</td>
<td>21.62</td>
</tr>
<tr>
<td>Abroad</td>
<td>2.12</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
</tr>
<tr>
<td>primary</td>
<td>1.69</td>
</tr>
<tr>
<td>secondary</td>
<td>19.27</td>
</tr>
<tr>
<td>bachelor</td>
<td>43.64</td>
</tr>
<tr>
<td>higher</td>
<td>35.40</td>
</tr>
</tbody>
</table>

*Source: Based on data obtained from an entity researching the Tri-City tourist market established by Pomeranian Scientific Institute.*
According to the literature on the subject, the Spearman correlation coefficient is a non-parametric equivalent of the Pearson correlation coefficient, with the use of ordinal variables. It is used in the analysis of low-quality data because it is not very sensitive to outliers. It can be used in comparisons of variables with significance tests. The relationship between random variables does not necessarily mean a cause-effect relationship, it only shows that there is a relationship, which can be described as the slope of the line fit for a set of rank pairs (Discovering Statistics Using IBM SPSS Statistics Authors Andy Field).

In the case of studies with quantitative variables, they should be regrouped. An important element when ranking the variables is the fact that one convention of the scale, ranked in the right order, is maintained - the increase or decrease of a given feature as successive rank parameters. With such an assumption, the Spearman correlation enables obtaining results adequate to the examined issue. One of the added benefits of overgrowing quantitative observations is that it reduces the impact of outliers on the test score. In addition, ranking is necessary when the correlation between the quantitative and ordinal variables is checked, where the solution based on the Pearson index is not able to give us the correct answer for such a problem (Statystyka, 2022).

The Spearman rank correlation coefficient used was converted from the Pearson correlation coefficient for the case of sequences of a pair of natural numbers with n observations. In a situation where the observations on the variables \( X_i \) and \( X_j \) are natural numbers, i.e. \( x_{it} = 1, ..., n, x_{jt} = 1, ..., n \) (\( t = 1, ..., n \)), then the correlation coefficient Pearson's transformed into Spearman's coefficient:

\[
r_s = 1 - \frac{6 \sum_{t=1}^{n} d_t^2}{n(n^2 - 1)}
\]

where:
- \( dt \) denotes the difference between the observations of simultaneous values of a pair of random variables in the form of natural numbers (\( t = 1, ..., n \)).
- \( t \) - number of observations
- \( dt \) - differences between ranks (consecutive numbers of observations corresponding to the variable \( x \) in order of the decreasing value of the variable \( y \)
- \( n \) – number of community

Therefore, this coefficient can be used when the observations on each of the pair of variables are natural numbers belonging to the results of the ratio measurement (Wiśniewski, 2014).

Collected data obtained from one of the tourist organizations researching city events were ordered by variables (descending, ascending) and then natural numbers corresponding to their place in the order were assigned (ranks were given). For the
calculation, the data in the form of a spreadsheet was processed with the use of the Statistica program. The literature on statistical analysis shows that the rank correlation coefficient assumes numerical values in the range \([-1; 1]\).

The same ranks of the values of the studied variables prove the existence of a positive correlation between them \((X = Y = 1)\), i.e. \(Y\) always increases when \(X\) and vice versa. The opposite numbering suggests a negative correlation. A positive sign of the coefficient indicates the existence of a positive correlation, a negative one indicates a negative correlation. The closer the correlation coefficient is to one, the stronger the correlation is. In the case when \(rs = 0\), it means that there is no correlation relationship between the examined variables.

As a result of the calculations, the information presented in the tables was obtained.

**Table 2. Spearman's rank order correlation with significance coefficient \(p < 0.0500\)**

<table>
<thead>
<tr>
<th>A pair of variables</th>
<th>N important</th>
<th>R Spearman</th>
<th>T(N-2)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age and travel to the stadium</td>
<td>472</td>
<td>-0.127807</td>
<td>-2.79371</td>
<td>0.005424</td>
</tr>
<tr>
<td>Sex and travel to the stadium</td>
<td>472</td>
<td>-0.117644</td>
<td>-2.56829</td>
<td>0.010528</td>
</tr>
<tr>
<td>Education and travel to the stadium</td>
<td>472</td>
<td>-0.128662</td>
<td>-2.81271</td>
<td>0.005118</td>
</tr>
<tr>
<td>Accommodation and travel to the stadium</td>
<td>472</td>
<td>0.281694</td>
<td>6.364714</td>
<td>0.000000</td>
</tr>
<tr>
<td>Zone on the stadium and travel to the stadium</td>
<td>472</td>
<td>-0.067144</td>
<td>-1.45894</td>
<td>0.15249</td>
</tr>
</tbody>
</table>

*Source: Based on data obtained from an entity researching the Tri-City tourist market.*

The highest positive correlation is between the place of residence and the transport to the stadium - moderate positive correlation \((0.281694)\), with the probability coefficient close to 0. However, the sector occupied in the stadium does not affect the travel to the stadium - no correlation. Spearman's R coefficient is the closest similar to 0. Similarly, education, age and sex have very little correlation with the transport to the stadium, with the probability \(p\) for the t statistic in the range from \((0.005\) to \(0.01\)).

Statistically significant correlations are marked in red, while correlations not statistically significant- in black. Table 3 also shows the correlation between all the factors included in the study. The greatest positive correlation occurs between education and age.

Based on the Statistica program, which enables the creation of categorized scatter plots, the relationship between the two travel to the stadium and age in relation to the accommodation was presented (Figure 3).

The graph shows a two-dimensional plane for the three variables, which shows that there is a weak correlation between these elements. In the graph, individual points
correspond to the examined entities. Thus, each of the examined persons is assigned a score for the variables on the x and y axes.

By means of correspondence analysis, the profile of the person using the selected means of transport was determined when commuting to the stadium. According to (Stanimir 2002). The use of correspondence analysis in marketing research), the attractiveness of using this method results, among others, from with the possibility of graphical presentation of the results. At the same time, the analysis of correspondence makes it possible to accurately recognize the existence of interdependencies for several categories of variables, measured on a nominal scale.

**Figure 3.** A scatterplot of the travel to the stadium and age in relation to the place of residence.

![Figure 3](image)

**Source:** Based on data obtained from an entity researching the Tri-City tourist market.

**Figure 4.** Relationship between age, place of residence and the transport to the stadium.

![Figure 4](image)

**Source:** Based on data obtained from an entity researching the Tri-City tourist market.
The graphical presentation of three factors shows that the residents of the Pomeranian Voivodeship up to 24 years old most often use the tram and bus to commute to the stadium. The train to the stadium is selected by people aged 25-34 and people aged 65-74, most often living in the Tri-City. However, the vast majority of cars are used by people aged 45-64, living in the Tri-City.

Based on the data in Figure 3, it can be seen that the bus and tram are used by passengers living in the Pomeranian region and other Polish residents, with secondary education, aged 25-44, occupying the blue zone at the stadium. The most frequently chosen means of commuting to the stadium - the train - was declared by people with a bachelor degree, aged 25-34 and 65-74 years old using the green and yellow sectors. On the other hand, the car is most often used by people over 45-64 years old who attend the yellow sector at the stadium, and in this case education was not of significant importance. The fans who choose the yellow zone and the green one, who had a bachelor's education, aged 45-74, declared a different travel (by bike, on foot).

4. Discussion and Conclusions

As a result of the analyzes carried out using the Spearman's rank correlation method and the correspondence method, it should be stated that there is no strong correlation between age, sex, education and place of residence and the choice of the means of transport. The greatest relationship, understandably, occurred between the place of residence and the choice of the means of transport. However, despite the positive correlation, the dependence coefficient is weak - 0.28.

On the other hand, on the basis of the correspondence method, profiles of people who most frequently use the train, bus / tram, car or go to the stadium on foot or by bike were established. Accurate segmentation is necessary due to the process of managing the logistics process in cities.

There is no doubt that such research was carried out for the first time in Poland, which indicates the need for further verification of the attitudes and behaviour of fans commuting to the stadium. Thus, a gap in the literature on the subject was filled.

The content of the article shows a picture of the close relationship between the presented content and practice. The usefulness of the conducted research for transport operators is not subject to discussion, however, it should be clearly emphasized that there may be significant problems in the application of the proposed solutions due to the diversified ownership structure and highly diversified decision-making centers. It seems that the easiest solution to carry out may be the modernization of the stadium's infrastructure aimed at ensuring an appropriate number of parking spaces for bicycles. It becomes much more difficult to introduce optimization processes in logistics of deliveries using railways and public transport.
Railway logistics is provided by SKM Trójmiasto Ltd., in which the dominant shareholder is PKP Co. based in Warsaw, owned by the State Treasury. Urban logistics is handled by the Gdańsk commune through the entity Gdańskie Autobusy i Tramwaje Ltd. The authors’ research indicates the need to establish broad cooperation between entities dealing with delivery logistics in order to achieve synergy effects, and thus optimize logistics costs.

References:


