

# Usage of BI Tools in Analysis of the Tomas Croatia 2019 Survey

I. Beroš<sup>\*</sup>, N. Hlupić<sup>\*\*</sup> and F. Džapo<sup>\*\*</sup>

<sup>\*</sup> Institute for Tourism, Zagreb, Croatia

<sup>\*\*</sup> Faculty of Electrical Engineering and Computing, Zagreb, Croatia

ivo.beros@iztg.hr, nikica.hlupic@fer.hr, frane.dzapo@fer.hr

**Abstract** - The managers in the tourism industry, as in any other industry, rely on the data and information to support the decision-making process. The data can be generated from within the business or it can be obtained from publicly accessible databases, various research, questionnaires, and surveys. Among the most important surveys for Croatian tourism are TOMAS surveys on attitudes and expenditures of tourists. The results of these surveys provide a relevant basis for tourism management at the national and regional level. In order to provide better insight into the survey findings, the Institute for Tourism created an interactive website using Microsoft Power BI platform. This paper describes the main steps required to transform data prepared for classical statistical presentation to data suitable for the use in the data visualization software like Power BI. The specified steps can be used as the basis for a systematic approach to the preparation of the survey results for the use in the Microsoft Power BI platform.

**Keywords** – Power BI; Data Visualisation, TOMAS

## I. INTRODUCTION

The business environment constantly changes. Private or public business organizations, government and public administration are under pressure to respond quickly to different conditions and to adjust the way they operate. Such activities require organizations to make frequent strategic and operational decisions. Making such decisions is impossible without acquiring relevant data, information, and knowledge.

For data and information to be useful, they should be presented in appropriate manner and access to them should be quick and simple. There is a general agreement nowadays that visual presentations of data serve as fundamental components of cognitive functioning and behavior [1]. According to Gelman and Unwin [2], there are at least six reasons to apply data visualization: giving an overview, to show scale and complexity of the data, to allow data exploration, to communicate findings, to tell a story, and to attract attention and stimulate interest. The first three reasons are about discovery – showing data without interpretation and allowing the reader to conceive own interpretation, while the last three reasons are about communicating a message.

The recent advances in data visualization have been in large part driven by the development of computer applications which allow interactive content. Data visualization can be shown in a web browser and users are enabled to interact with it in many different forms [3]. For

example, hovering the mouse over an object to bring up a tooltip with additional information. A click on the object can serve as a trigger for showing additional information, or to replace content providing more details on the clicked object (“drilling down”). There are many commercial tools for creating interactive data visualization such as Tableau, Power BI, SAS Analytics Cloud, or FineReports and most of them have free versions with some limitations.

Tourism is one of Croatia’s most important economic sectors. In 2019, the value of foreign exchange income from tourism reached 10.5 billion Euro and accounted for 19.4% of the GDP [4]. Thus, it is vital for all decision-makers linked with tourism sector to have access to all data and information required to make decisions. The sources like eVisitor [5] give precise information about several aspects of tourism supply (number of beds) and demand (number of arrivals, number of overnights) which can be grouped by time period, region or the country of origin. However, it is widely recognized today that, in order to succeed, business must be focused on satisfying the customer needs [6]. In order to find why tourists come in Croatia, what their motives are, which activities they prefer or what the factors that disturb them are, the Institute for Tourism has conducted the TOMAS surveys [7]. This paper deals with the problem of visualizing the data of the TOMAS surveys results.

The paper is organized in the following manner. The introductory section is followed by the section which describes the methodology, significance and structure of TOMAS survey as well as the Power BI platform used for data visualization. The details about data preparation and report preparation are provided in the third section whereas the final section provides the conclusion and recommendations for further research.

## II. METHODOLOGY

### A. The TOMAS Survey

TOMAS is a common name for the repeated cross-sectional surveys of tourist profile, behavior, satisfaction, and expenditures in Croatia that have been conducted since 1987 [7]. All TOMAS surveys share the same methodological framework established by the Institute for Tourism. The most prominent TOMAS survey is TOMAS Summer which was conducted ten times in the period from 1987 to 2017. These surveys gathered data about the profile, satisfaction, and consumption of tourists in Adriatic counties of Croatia during the main seasonal months.

The results obtained from TOMAS surveys help tourist destinations recognize the profile and expectations of their visitors and adjust their strategies accordingly.

In Croatia, many important tourism-related data, e.g., the number and types of accommodation facilities, the number of beds, the number of arrivals and overnights, can be found in eVisitor or the Croatian Bureau of Statistics. eVisitor (<https://www.eVisitor.hr>) is tourist check-in and check-out information system. It collects the data from the accommodation providers who can check their guests in and out at any time [5]. From the collected data it is possible to extract information about length of the visit, location, age, country of origin and other. The data from eVisitor allow tracing tourists' movements almost in real time and enable the national or local tourist boards to implement active marketing policy.

However, the data from eVisitor cannot provide insight into the data collected and analyzed by TOMAS: opinions of guests about elements of tourism supply, the reasons of their arrival, their consumption, activities, and others. The obtained research results, together with other tourism-related data, are used in design and implementation of key strategic activities related to national tourism brand of the country.

In 2019, the Institute for Tourism conducted TOMAS Croatia 2019 survey [8]. The full name of the survey was *Survey on attitudes and expenditures of tourists in Croatia - TOMAS 2019*. That survey, for the first time, introduced the profile, consumption, and satisfaction of tourists in the whole of Croatia during the whole year, including the period out of the main season. The survey was conducted from May 2019 to March 2020, when it was stopped due to the COVID-19 pandemic. The main characteristics of the survey are [8]:

- 13,582 respondents in hotels, hostels, camp sites and family accommodation in 143 destinations along the country,
- domestic tourists and foreign tourists form 24 main generating markets,
- the results are relevant for demands of Croatian tourism according to region, type of accommodation facility, season, and country of origin,
- the results are weighted by data on tourists' overnights provided by eVisitor information system.

The survey results are provided in forms of the book and e-book (pdf document). The first part of the book presents the research findings and comparison with findings of the previous TOMAS surveys. The second part of the book consists of the tables in which the results of each question are grouped by the country of origin, region (whole year, summer season and out of summer season), type of accommodation facility, season, and the main motive of visit.

To facilitate the presentation of research results, the Institute for tourism decided to develop an interactive website. Considering budget limitations as well as the

format and size of the data, Microsoft Power BI has been chosen as the data visualization platform. The website enables an overview of the results grouped by region for three periods: summer (main) season, out of summer season, and the whole year. The website navigation is easy as well as intuitive and it allows users to quickly find the desired information.

To avoid any confusion, in the rest of the paper the term "collected data" refers to the data collected by the questionnaire, whilst the term "survey results" refers to the processed data, i.e., the numbers calculated from the collected data. Also, the term "report" will be used instead of "website" to emphasize the usage of the Power BI platform.

It is important to notice that the website serves only as a presentation platform of survey results and it does not store the collected data. The main reasons for that decision are speed and simplicity. Namely, all survey results are calculated from the collected data weighted depending on actual number of tourist overnights in survey period [7]. Therefore, to avoid unnecessary calculation and querying, the survey results prepared for the book are used.

### B. Power BI platform

Power BI is a cloud-based service that lets users store and share essential business data in the form of dashboards and reports [9], [10]. Power BI consists of several tools [10]:

- Query Editor is Power BI tool for data gathering and transformation. Using Query Editor, users can import data from multiple sources like text files, Excel workbooks, important databases (SQL server, Oracle, IBM Db2 and so on) and various on-line services (e.g., Microsoft Dynamics 365 or Salesforce Reports). The data can also be transformed and all steps in data transformations are saved and can be used over again.
- The xVelocity In-memory Analytical Engine is the technology that manages the storage and retrieval of the data within Power BI. This technology, also known as the VertiPaq engine, implements powerful in-memory database, which stores and organizes the data tables in an optimized and
- The Tabular Model is applied to define tables of data, the relationships between the tables and hierarchies among the tables. Business calculations in the Tabular model are created using measures. A measure is an expression written in DAX (Data Analysis Expressions) language that returns a single value [12] and can be used in reports or as a part of calculations in other measures. Measures react and respond to user interaction (filter context) and recalculate appropriately.
- Power View is responsible for data visualization and allows users to create interactive reports and dashboards in an easy way.
- Power BI reports and dashboards can be published to the Power BI service or can be saved to the Power BI Report Server. For organization-level

content sharing, paid license is required. The Power BI Report Server is used when organizations do not want to share content using cloud-service. The Power BI Service can be used free of charge, but users can only publish report on Internet where everyone can access it.

Application Power BI Desktop provides the best environment for the reports and dashboard creating. Typically, the author prepares report in Power BI Desktop and, upon completion, publishes finished report on Power BI service (powerbi.com) where certain additional modifications can be made.

Power BI Desktop can work with data stored inside or outside Power BI file. In situations where analysis must be performed in real time, Power BI Desktop can directly execute a query against the source data to render visualization. Otherwise, the data can be imported and stored within Power BI file.

Note that the Query Editor, the VertiPaq Engine and Power View are also parts of Microsoft Power Pivot. Power Pivot is an Excel extension for doing self-service business intelligence. The main Power Pivot feature is ability to store and work with complex data models and the large amount of data [13].

### III. DATA VISUALIZATION PROCEDURE

In this section, the necessary steps in the data visualization process using Power BI platform are described. Roughly, the process can be divided into two phases: data preparation and report preparation. Data preparation includes all actions required to obtain data model, while report preparation is related to the report design, i.e., the presentation of the data.

#### A. Data Preparation

To efficiently use Power BI platform, it is essential to build a suitable data model. Usually, data are modeled as a star scheme [11] with one fact table and several dimension tables. Before explanation of the data model used, brief description of data source is given.

For each question in the questionnaire there is a table, formatted as it is shown on Fig. 1. There are 84 tables in total and all tables are in the Excel file. For nearly all questions, value in the intersection of row A and column B represent percentage of tourists in region B who gave answer *Answer XA*. The exception is the question about expenditures, where value in the intersection of row A and column A represents how much tourist in region B spent on the particular service (*Answer XA*).

Question X for period Y

Description	Region 1	...	Region B	...	Region M
Answer X1					
⋮					
Answer XA					
⋮					
Answer XN					

Figure 1. Typical report table for the survey question.

The questionnaire results formatted as on the Fig.1 are not suitable for Power BI and must be transformed in a more appropriate data model, shown on Fig. 2.

The data model presented on Fig. 2 is a variation of the star scheme model [11]. It consists of five tables. The dimensions in the model are *tRegion* (description of region), *tPeriod* (description of period), *tQuestion* (description of questions) and *tAnswer* (description of question answer). The fact table in the model is *DataAll* and it is directly related with tables *tRegion*, *tPeriod*, and *tAnswer*. It is noteworthy that *DataAll* and *tQuestion* are related via *tAnswer*, not directly. The attributes of table *tQuestion* are: *idQuestion* – primary key, *questionText* – question text, *answerHeader* – description of answer like years or motives (cell Description in the upper-left corner on Fig. 1), *measuringUnit* – measuring unit such as % or €, and *questionType* – type of question (single response, multiple response etc.). Other tables have expected attributes.

In the presented case, we deal with same set of questions observed in three periods and the same column descriptors. In general case, similar model can be used. Tables *tQuestion* and *tAnswer* will remain unchanged, table *tRegion* have to be adapted to describe column variable, while *DataAll* is connected only with *tRegion* and *tAnswer*. Depending on concrete case, more attributes can be added to the tables.

Although transformation of the survey results formatted as on the Fig. 1 can be accomplished by Query Editor, in this work all preparation was made in Excel. The preparation took three steps:

- The first step was manual preparation of the tables with few records, *tRegion* and *tPeriod*.
- The second step was preparation of the tables *tQuestion* and *tAnswer*. For this step VBA macro is used. Question information (text, measurement unit, description of answers) and answers are extracted from the table containing question results and stored in proper tables. Information about question type is manually set.
- In the final step table *DataAll* is assembled. VBA macro is used again. The values in the table containing question results are processed and for each entry the identifiers of region, period and answer are obtained.

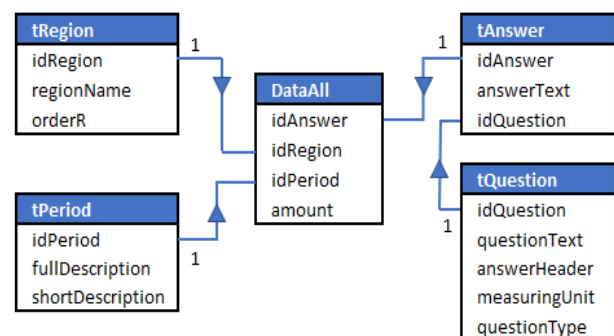


Figure 2. Data model for the presentation of the survey results

After preparation, the tables are loaded into Power BI Desktop by Query Editor and connection among tables are made. The approach that uses data stored in Excel file and is afterwards loaded into Power BI desktop is suitable when the data are static. Also, possible manual modification of dataset is easier in Excel than Power BI Desktop.

### B. Report preparation

In this section the report making procedure is outlined.

Data visualization in Power BI is achieved by visuals. Visuals are individual presentation elements created on the basis of underlying data [9]. Currently, there are more than 30 standard visuals in Power BI and many more custom visuals can be imported from the MS Office Store. Visuals are arranged on a single page (dashboard), while Power BI report is a collection of pages in a single file, all using the same dataset [9]. The data visualization of the survey results' goal is to produce the Power BI report and publish it on the web. In the rest of the section the term "report" indicates Power BI report made for the visualization of the TOMAS Survey results.

Basic visuals used in the visualization of the survey results are table, column and a bar chart, slicer, card, and shape map.

Table visual provides simple row groups of dimension values and measures [14]. It enables the users to look up specific data points. Also, the users can easily sort and filter the table to get a better view of the data. Column chart visual is used to show structure of the answers. Although pie chart or doughnut chart visuals can be used for the same purpose, column chart visual gives a more comprehensible display. Bar chart visual is used to show distribution of data. Card visual enables the user to show measure value connected with the data.

Shape map visual is used to compare regions on a map using colors. Its main purpose is to color the map regions according to the chosen values. As opposed to the map visual, shape map visual cannot show precise geographical locations. Also, shape map visual is not limited to geographical maps, it can be used on any custom map users can create such as a floor plan or seating arrangements.

While aforementioned visuals are used to display data, a slicer visual is used to control which data will be displayed or used for evaluation. The slicer visual is used to enable report users to apply their own filter selections [14].

Page navigation in the report is provided by bookmarks. In Power BI, a bookmark is the saved representation of the entire state of the page, including all filter and slicers selections, and properties of the visual, like hidden or not [14]. The bookmarks are a great way to make a presentation with the Power BI and can also be used for the report navigation together with images, shapes, and buttons. Namely, action property of these objects can be set to bookmark, and after the click on the object the specified bookmark will be retrieved, i.e., the saved page representation is recreated. Working with bookmarks can be tedious because in the settings of the action property bookmark name should be set explicitly and not as a



Figure 3. Report home page

measure result. On the other hand, created reports are easy to navigate and use.

The design of report is straightforward – the report home page (Fig 3) is used as the main menu. The report users choose a question group and then select a question. For the majority of questions there are three views to present results. In the first view (Fig 4), the results for the specified answer and regions for the chosen period are shown on the map. In the second view (Fig 5), the results for the specified region, all answers, and for the chosen period are presented in the table and in the chart. In the case of the three questions, table and chart are displayed separately since there are too many answers (categories). In the last view (Fig 6), the results for the specified answer and for the chosen period are shown for all the regions in the bar chart visual.

The first view is shown on the figure 4. The mayor parts of the view are card visuals where question text is shown (1), slicer for the period (2), slicer for the answer together with card visual for header (3), and 10 card visuals for displaying the results for the regions (4). But there is another, invisible slicer for the questions which ensures that

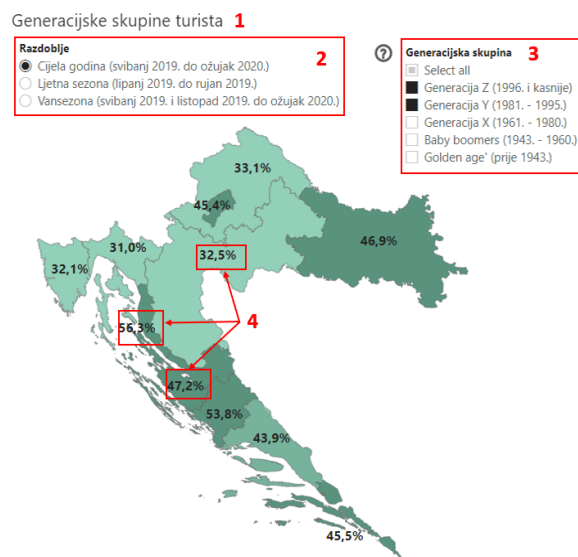


Figure 4. The map view on the results of the specified question.

information shown on page, like answers, question text or answer header, are related only to specified question.

The working principle is simple: invisible slicer is set on specified question. Since the tables `tQuestion` and `tAnswer` are related, invisible slicer affects slicer for the answer and only answers for the chosen question are displayed. After the initial values of slicers for period and answers have been set, the current state of visuals on the page is saved as a bookmark. When the report user clicks the right button, the status of all visuals is set on previously saved state. The report author only needs to make a bookmark for each question.

Measure `cQuestText`, defined by

$$cQuestText = \text{Max}(tQuestion[questionText])$$

is used as an input parameter for a card visual (1). Note that the invisible slicer allows only one selected question. The input parameters for the cards in part (4) are measures `amountR1`, ..., `amountR10` defined by

$$amountRx = \text{Calculate}(\text{Sum}(\text{DataAll}[amount]), \text{DataAll}[idRegion]=x)$$

where `x` is a number between 1 and 10. The `amountRx` measures are easy to explain: sum all amounts from table `DataAll` where `idRegion` is equal to `x`. But, the state of the period slicer and the answer slicer, also affects measures. The card visuals have to be used because shape map visual cannot display numbers.

Slicer for the answers is set to allow multiple selection if the question type is a single response. For the multiple response questions, slicer is set to allow single selection only. Those settings cannot be done dynamically, therefore two similar pages, differing only in settings of answer slicer, are created. In that manner, all similar situations are resolved.

The second view on the survey results is shown on the figure 5. The page consists of five main parts: question text (1), slicer for the period (2), slicer for the region (3), table (4) and chart (5) visual for the question results. As before, there is invisible slicer for the question on the page. Technically, there is not any significant difference between

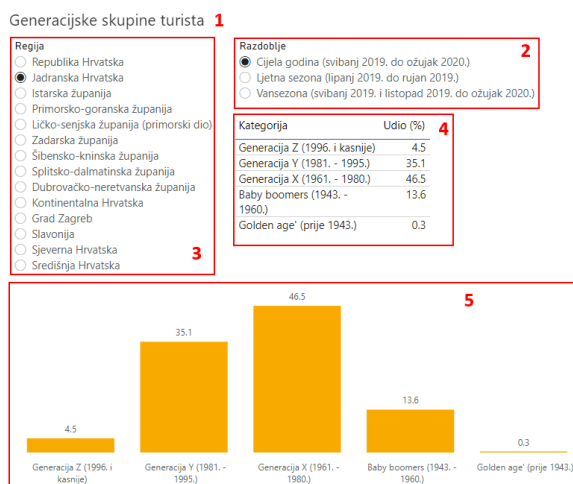


Figure 5. The second view on the survey results – the question results for the chosen region and period.

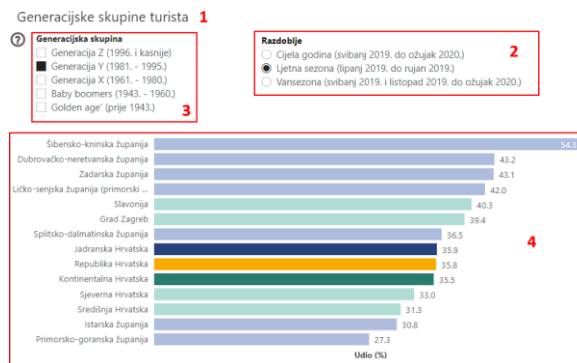


Figure 6. The third view on the survey results – answer results for all regions are displayed.

the first and the second view. The table uses `amount` measure defined by

$$amount = \text{Sum}(\text{DataAll}[amount])$$

The `amount` measure is affected implicitly by the state of the region slicer and the period slicer and by the value of the answer in the first column of the table. The `amount` measure is also used as a value in the chart visual. To distinguish single from multiple response questions, two pages, with different chart visual, are created. The page with the column chart visual is used for single response questions, while the page with the bar chart visual is used for multiple response questions. For questions with many answers, the pages with a table and a bar chart visual are created separately.

The last view is shown in figure 6. The only notable difference regarding previous views is the use of colors in the bar chart visuals (part 4). The bar chart displays results of the selected answers for the chosen period for all regions. Additionally, the results for Adriatic Croatia, Continental Croatia and Croatia are displayed, all sorted descending. The results for regions of the Adriatic Croatia are colored light blue, while the results for Adriatic Croatia as a whole are colored blue. The similar coloring is used for the regions belonging to Continental Croatia (light green – green). The report user now can easily see how the results for region relate to the results for the corresponding part of Croatia, and Croatia at large.

The page navigation in each view is achieved using buttons, but they are omitted from figures 4 – 6 for the reason of simplicity. The report user can navigate to other views for the same question or go back to the report home page.

#### IV. CONCLUSION

Any quality decision-making relies on the relevant data and its appropriate presentation, so does the decision-making processes in tourism industry as well. The main sources of data on tourism industry in Croatia are eVisitor and Croatian Bureau of Statistics. They provide the information on the tourism demand and supply such as the number of beds, overnights, and arrivals. The TOMAS survey, conducted by the Institute for Tourism, Zagreb provides additional data about the tourist profile,

motivation, behavior, and satisfaction. Traditionally, the results of the TOMAS surveys were printed as a book. In order to make the results more accessible the Institute for Tourism has visualized the latest results and made them available in an interactive online report on <http://www.iztztg.hr/hr/tomas-interaktivna-stranica>.

The report has been produced using the Power BI platform. The process of preparation was conducted in two main phases: data preparation and report preparation. In the data preparation phase, the data was loaded into the Power BI and organized as a star scheme. In the second phase, the data were visualized by tables, graphs, and slicers. The navigation through the report has been enabled by bookmarks, i.e. the saved representation of the entire state of the page, including all filter and slicer selections.

Main steps necessary for the use of Power BI in the data visualization of the survey results are outlined in this paper. Our study shows that it is possible to prepare survey results for data visualization with little effort. General suggestions are:

- In the data model, information about questions (text, measure unit etc.) and question answers should be stored in the separate tables. Every dimension should have its own table.
- The control of the visuals on a page is better when invisible slicer is used.
- It is advisable to use bookmarks for page navigation.
- Different views for the presentation of question results should be prepared. The different views can contain various visuals.

In the future, we intend to visualize other tables from the survey TOMAS CROATIA 2019 as well as the data of previous surveys. Another project is the data visualization of previous surveys. The data from several surveys, joined

together, will enable better insight into trends inherent to motives, attitudes and behavior of the tourists visiting Croatia.

## REFERENCES

- [1] K. Sedig and P. Parsons, *Design of Visualizations for Human-Information Interaction: A Pattern-Based Framework*. Morgan & Claypool, 2016.
- [2] A. Gelman and A. Unwin, "Infovis and statistical graphics: different goals, different looks," *J. Comput. Graph. Stat.*, vol. 22, no. 1, pp. 2–28, Jan. 2013.
- [3] R. Grant, *Data Visualization: Charts, Maps, and Interactive Graphics*. Boca Raton: CRC Press, Taylor & Francis Group, 2019.
- [4] I. Rašić, "Sektorske analize: Turizam," *Sekt. Anal.*, vol. 9, no. 81, pp. 1–49.
- [5] Croatian National Tourist Board, "eVisitor.," Accessed: Jan. 21, 2021. [Online]. Available: <https://www.htz.hr/hr-HR/projekti-i-potpore/evisitor>.
- [6] P. Atrill and E. McLaney, *Management Accounting for Decision Makers*, 6th ed. New York: Financial Times/Prentice Hall, 2009.
- [7] Z. Marušić, I. Sever, and S. Čorak, "Identifying Trends in Tourism Demand Using Longitudinal Survey," in *Evolution of destination planning and strategy*, Springer, 2017, pp. 221–243.
- [8] Z. Marušić, S. Čorak, N. Ivandić, I. Beroš, and M. Ambrušec, *TOMAS Hrvatska 2019.: stavovi i potrošnja turista u Hrvatskoj*. Zagreb: Institut za turizam, 2020.
- [9] A. Aspin, *Pro Power BI Desktop*. Staffordshire [England]: Apress, 2016.
- [10] B. Larson, *Data Analysis with Microsoft Power BI*. Mc Graw Hill, 2020.
- [11] M. Russo and A. Ferrari, *Microsoft PowerPivot for Excel 2010: Give Your Data Meaning*. Redmond, Wash.: Microsoft, 2011.
- [12] P. Seamark, *Beginning Dax with Power BI*. New York, NY: Springer Science+Business Media, 2018.
- [13] I. Beroš, N. Hlupić, and D. Butković, "Power Pivot for Excel as a tool for descriptive analytics," in *Proceedings of the ISCCRO – International Statistical Conference in Croatia - ISCCRO'20: The 3rd International Statistical Conference in Croatia / Žrnjani, Berislav; Čeh Časni, Anita, Zagreb, Croatia, 2020*, vol. 3, pp. 29–34.
- [14] B. Powell, *Mastering Microsoft Power BI*. Packt Publishing, 2018.