Police Force - Traffic Accidents

Michal Kvet, Marek Kvet
University of Žilina
Univerzitná 8215/1, 01026 Žilina, Slovakia
{michal.kvet, marek.kvet}@fri.uniza.sk

Lucia Konštiaková, Branislav Dupák, Ján Dunaj
University of Žilina
Univerzitná 8215/1, 01026 Žilina, Slovakia
{konstiakova, dupak, dunaj5}@stud.uniza.sk

***Abstract*—This article focuses on the topic of the Police Corps project. In the following sections of the article, we will deal with the field of traffic accident data processing. This article describes the goal of the project, the analysis, and processing of data provided for this project, the creation of an information system for processing this data, and the subsequent basic analysis of the data we process. The data on which the analysis is performed is realistic, provided by the police force of the Czech Republic.**

I. Introduction

The main goal of the project is to create a sustainable information system for selected departments of the police force of the Slovak Republic. The project will be solved in cooperation with PZ (Policajný zbor = Police force) SR. It follows on from several of the bachelor's theses, which at least partially covered those areas that required the most support for information technology. This project is divided into two parts, traffic accidents and crimes in the Slovak Republic. In this article, we will focus on the topic of traffic accidents in the Slovak Republic.

The primary goal of this project is to analyze the existing data on traffic accidents recorded in the Czech Republic, where we have enough freely accessible data. Another task is to think about how to reasonably download and subsequently process the provided data. After this step, it is necessary to load this data in a suitable environment selected by us, display it, and perform its subsequent analysis.

Chapter II. Analysis of the provided data is focused on the analysis of real data from the police force of the Czech Republic. This chapter describes the individual items of the forms that were used to record traffic accidents. In Chapter III. Data processing, we focused on the analysis of how to process the data and design properly the database for the data. Chapter IV. The display of data is focused on the topic of how we already displayed the processed data. This is the data we can perform analysis on. Chapter V. Data analysis is a chapter in which we describe the resulting processed data and individual analyses of them. In Chapter VI. The conclusion is a general summary of the project and a description of what we expect from the project in the future.

II. Analysis of the proivided data

All collected data on accidents in the Czech Republic is available on particular Czech police website. These data are recorded from 2016 to the present in a monthly interval. All data is written to .csv files marked 00.csv it 19.csv and one PEDESTRIANS.csv file. The regions of Czech republic are depicted in Fig. 1.

Fig. 1. Map of the regions of the Czech Republic

TABLE I. Description of .csv files for regions

Filename	File description
00.csv	Accidents on the territory of the capital city Prague
01.csv	Accidents in the Central Bohemian region
02.csv	Accidents in the South Bohemian region
03.csv	Accidents in the Pilsen region
04.csv	Accidents in the Ústí nad Labem region
05.csv	Accidents in the Hradec Králové region
06.csv	Accidents in the South Moravian region
07.csv	Accidents in the Moravian-Silesian region
14.csv	accidents in the Olomouc region
15.csv	accidents in the Zlín region
16.csv	accidents in the Vysočina region
17.csv	accidents in the Pardubice region
18.csv	accidents in the Liberec region
19.csv	accidents in the Karlovy Vary region
PEDESTRIANS.CSV	pedestrian accidents

The files 00.csv 19.csv contain data on traffic accidents in individual parts of the regions of the Czech Republic and the file PEDESTRIANS.csv contains data on pedestrian accidents.

TABLE II. DESCRIPTION OF .CSV FILE FORM

Item	Description
p1	Identification number
p2	Accident time data
p5a	Accident location
p6	Type of accident
p7	Type of collision of oncoming vehicles
p8	Type of fixed obstacle
p9	The nature of the accident
p10	Accident fault
p11	Alcohol present in the accident culprit
p12	The main causes of the accident
p13	Consequences of the accident p13a - people killed, p13b - severely injured people, p13c - slightly injured people
p14	Total material damage
p15	Type of road surface
p16	The condition of the road surface at the time of the accident
p17	Road status
p18	Weather conditions at the time of the accident
p19	Visibility
p20	Prospective conditions
p21	Road division
p22	The situation of an accident on the road
p23	Traffic management at the time of the accident
p24	Local adjustment of driving priorities
p27	Specific places and objects at the accident site
p28	Directional conditions
p33	Consequences on the lives and health of pedestrians p33c - sex of the person, p33d year of birth of the pedestrian, p33e - nationality, p33f - first aid, p33g -
p34	Number of participating vehicles
p35	Accident site
p36	Type of road
p37	Road number
p39	Kind of crossing the road
p44	Vehicle type
p45a	Vehicle make
p47	Year of vehicle production
p48a	Vehicle characteristics
p49	Skid
p50a	Vehicle after accident
p50b	A leak of mobile, transported materials
p51	Method of releasing persons from the vehicle
p52	The direction of travel or position of the vehicle
p53	Damage to the vehicle
p55a	Driver category
p57	Driver condition
p58	External influence of the driver
d	X coordinate
e	Y coordinate

All these files contain several columns that describe the accident. Each column item contains a number to which a description is assigned. Then, when all these numbers match the description of the form, we get the whole accident described.

This form is used to record every accident. Based on all these items, it is possible to perform a different amount of analysis regarding the common causes of accidents in different combinations.

TABLE III. DESCRIPTION OF THE PEDESTRIAN .CSV FILE FORM

Item	Description
p1	Identification number
p29	Pedestrian category
p30	Pedestrian condition
p31	Pedestrian behavior
p32	The situation at the scene of the accident

The items on the form describe the record of pedestrian traffic incidents. Every item contains an integer. To describe properly the entire pedestrian traffic incident, we must match the items of the form with the individual descriptions.

III. DATA PROCESSING

When we performed the overall analysis of all the provided .csv files, we had to decide how to process the individual data. First of all, it was necessary to figure out how to store the individual data recorded in files. Since the data are recorded as code lists, it was appropriate to create a separate table for each item in the form, which contains an identification number and a description that is defined by that number. This created the tables as listed in Table II. Another problem was how we would store the individual data. Since the data are divided by regions, it was appropriate to create several tables, which were named among the individual regions of the Czech Republic. This created tables where we could divide the individual data horizontally based on the areas. Furthermore, each drought is divided over the year in which the accident occurred, using range partitioning. This is what the SQL statement looks like when creating a table, which divides the table into regions and individual years for each region [11]- partition by range (id_kraj) – see Fig. 2.

```
subpartition by range (datum) (
partition praha values less than(1)
(
 subpartition rok_2017 values less than (to_date('2018-01-01','YYYY-MM-DD')),
 subpartition rok_2018 values less than (to_date('2019-01-01','YYYY-MM-DD')),
 subpartition rok_2019 values less than (to_date('2020-01-01','YYYY-MM-DD')),
 subpartition rok_2020 values less than (to_date('2021-01-01','YYYY-MM-DD')),
 subpartition rok_2021 values less than (to_date('2022-01-01','YYYY-MM-DD'))
```

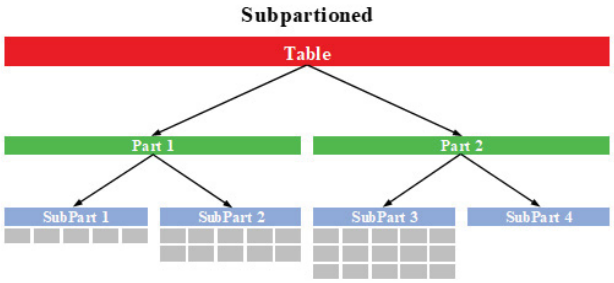

Fig. 2. Example of subpartition in Oracle

Thanks to dry partitions, we have achieved a more efficient search of data on a per-year basis. An example of multiplicity can be seen in Table IV.

TABLE IV. NUMBER OF PARTITIONS FOR THE PRAGUE REGION

Region	Year	Multiplicity
Prague	2021	17 503
	2020	16 919
	2019	21 453
	2018	22 758
	2017	23 027
	2016	22 870

Each such table contains descriptive numbers that refer to individual item descriptions from the form. When we had created a dry database, we started processing all the data into the database.

The first problem was how to store the collected data in a database efficiently. To solve this problem, we have created an information system that is able to download all records and then process and store all of them in a database. The initial problem occurred when reading the data from files. Not all the items were filled out. The errors began to occur while we were loading and writing data. We solved the problem by adding an undefined value for the item from the form where the data was missing.

Another problem that occurred when entering data was that some of the data did not match the values defined in the table that was describing the form of the data to fill in. Errors occurred where the number one was entered in the format 01 and not 1, as specified by the rules of the form. We solved this problem by adding additional values with the same description to the individual tables of the database that describes the form items. [2]. Over time a new problem arose when updating data. This problem was due to the fact that every month the newly added data was cumulatively written to the original files. As a result, we were unable to read and write the entire file in rows, as this result would be in duplicate data. It was necessary to solve this problem by filtering out the data that has been already read from the database and writing only those which were newly added. [3]

After solving all the problems, we successfully filled the database with all the data with the possibility of updating them every month in the data provided on traffic accidents by the police force of the Czech Republic.

IV. DATA DISPLAY

The issue we addressed in this topic was how to retrieve and display the data we process in a database. First, we must decide what kind of information system we are going to define, whether it will be a desktop application, mobile application, or web application. Finally, we decided to create a desktop application. To create this application, we chose to use the C# programming language and .net technology for graphical data display. We used the Oracle database system to store the data. [13]

To be able to perform various analyses on the data, it was necessary to create an information system that will be able to retrieve this data and then display it for individual regions of the Czech Republic. It was also necessary for this information system to display the data in graphical form, in form of suitable graphs. Thanks to them we could analyze and compare the individual data. [12]

The problem with retrieving data was that in the case of a slower Internet connection on the part of the user, it took a long time to retrieve all the data with all the retrieved items from the form. To make data retrieval more efficient we had to solve this problem by reading only the code lists of individual traffic accidents at the beginning of the program. This reduced number of data to read from different tables. Now it is possible to read data only from the tables where a traffic accident is recorded in the form of a code list. If the user decides to display the details of the accident, then all the items of the form will be loaded into the accident and displayed. In this case, it is fast as only one accident has to be paired at a time, as it is matched by a unique identification number.

Another feature needed for this application was the graphical representation of the place where the accident occurred. This is due to a deeper analysis, in terms of solving problematic transport hubs, intersections, parking lots, etc., where there were frequent accidents. Thanks to analysis and changes in the traffic rules, there could be greater road safety in the future. In this case, there was a problem with the processing of coordinates and their displaying. Coordinates that were entered have the form of the coordinate system S-JTSK, which is a system of a unified cadastral trigonometric network. This network is used in the geodesy of the Czech Republic and Slovakia. Since most of the freely available map views don't know read this type of coordinates, we used the website https://en.mapy.cz/, which is able to convert, and display entered coordinates on the map.

Another element of the information system is basic data filtering. The user can choose to view accident data by region or display them all. It is possible also to combine additional filters above this data. [14]

V. ANALYSIS OF PROCESSED DATA

In the data analysis, we had to decide what part of the processed data we would analyze. First, we chose to analyze the total number of traffic accidents in individual regions of

the Czech Republic. In the case of data containing traffic accident records from the capital Prague, we decided to take the data as a separate entity since the city of Prague is treated as a separate administrative unit.

TABLE V. ROAD ACCIDENTS IN THE CZECH REPUBLIC FOR 2021

Region	Statistics for regions		
	Number of traffic accidents	The population	% per capita
The territory of the capital city of Prague	16 122	1 251 072	1,29
Central Bohemian region	13 671	1 256 850	1,09
South Bohemian region	3 950	637 723	0,62
Pilsen region	3570	571 831	0,62
Ústí nad Labem region	10 104	835 814	1,21
Hradec Králové region	4 472	547 903	0,82
South Moravian region	5 833	1 152 819	0,51
Moravian-Silesian region	9 099	1 244 837	0,73
Olomouc region	4 851	640 410	0,76
Zlin region	4 123	590 527	0,70
Vysočona region	4 404	513 677	0,86
Pardubice region	4 009	520 584	0,77
Liberec region	4 200	433 948	0,97
Karlovy Vary region	2 245	304 602	0,74

Based on the recorded data for 2021, we found that the highest accident rates in the Czech Republic are in the regions of Prague, Central Bohemia, Ústí nad Labem region, and the Moravian-Silesian region. On the other hand, the Karlovy Vary region, the Pilsen region, and the South Bohemian region have the lowest accident rates. The following Fig. 3 describes the number of traffic accidents in selected regions.

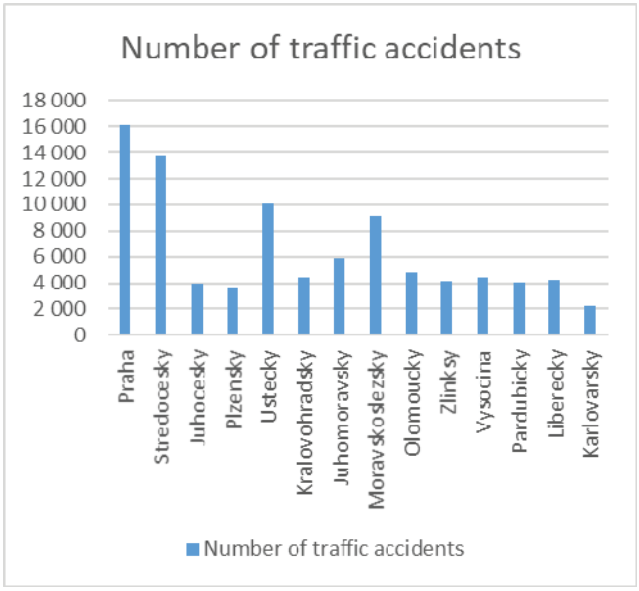

Fig. 3. Number of traffic accidents - graph

Another analysis that we dealt with was the analysis of the number of health consequences for people in individual regions caused by traffic accidents. We have focused on the numbers of lightly and severely injured people and people killed.

TABLE VI. CONSEQUENCES OF TRAFFIC ACCIDENTS IN THE CZECH REPUBLIC IN 2021

Region	Statistics for regions			
	Number of traffic accidents	Slightly injured	Severely injured	Deaths
The territory of the capital city of Prague	16 122	1 286	110	18
Central Bohemian region	13 673	1 895	226	61
South Bohemian region	3 950	1 270	154	40
Pilsen region	3570	1 088	57	30
Ústí nad Labem region	10 104	1 139	124	31
Hradec Králové region	4 472	817	73	34
South Moravian Region	5 833	1 799	161	40
Moravian-Silesian region	9 099	1 501	119	27
Olomouc region	4 851	864	66	33
Zlin region	4 123	901	60	23
Vysočina region	4 404	831	58	20
Pardubice region	4 009	828	51	17
Liberec region	4 200	775	37	15
Karlovy Vary region	2 245	438	37	12

Based on these data, we concluded that in the Central Bohemian, South Bohemian, and South Moravian regions there is a high number of traffic accidents with consequences for lives, but also a high number of easily and severely injured people. Based on the result we recommend that vigilance in these areas of the region should be increased. The graph of consequences of traffic accidents follows in Fig. 4.

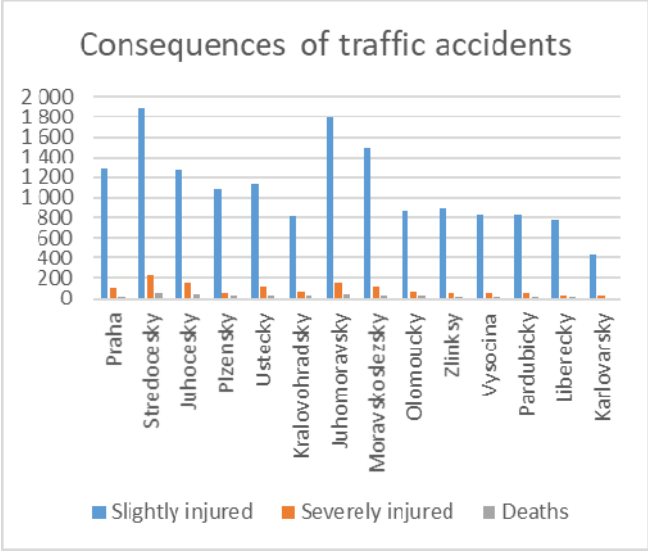

Fig. 4. Consequences of traffic accidents - graph

We also focused on the effects of weather on traffic accidents. We compared the number of accidents under different conditions such as fog, light rain, rain, snow, icing, gusts of wind, or otherwise.

TABLE VII. IMPACT OF WEATHER ON TRAFFIC ACCIDENTS IN THE CZECH REPUBLIC FOR 2021

Weather	Number of traffic accidents
Fog	709
Light rain	2 952
Rain	3 145
Snowfall	2 559
Icing, ice	1 076
Gusting wind	135
Other problems	181

According to this data, we concluded that the biggest traffic accident was affected by adverse conditions during the rain when the road is wet. Other unfavorable conditions are snowing and icing. In these circumstances, drivers should be extra careful in avoiding such traffic incidents that might have unpleasant consequences. We also had these data divided for individual regions of the Czech Republic. In this case, the Prague and Central Bohemian regions had the highest accident rate in rainy conditions. All regions had the largest accident rate in rainy conditions, except the Moravian-Silesian region, which had the largest accident rate during snowfall. Interestingly, drivers had more accidents during the rain than during the snowing or icing. This is probably because the drivers are more careful in the worse conditions than during the rain. A graphical representation is depicted in Fig. 5.

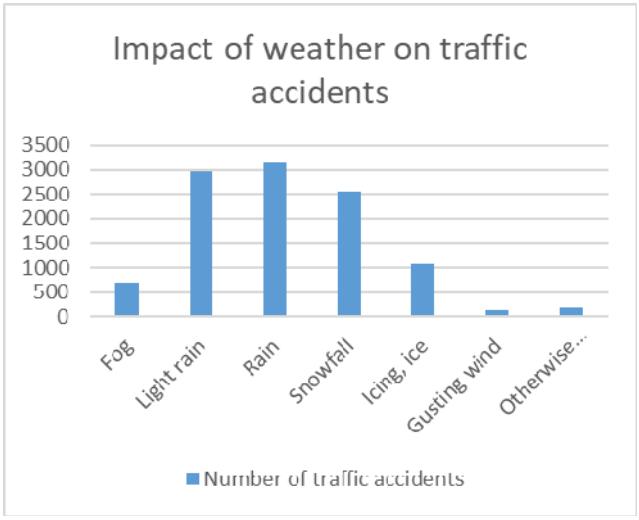

Fig. 5. Impact of weather on traffic accidents - graph

As a further analysis, we chose an analysis of the number of accidents in individual months for each region of the Czech Republic and the total number of traffic accidents in the entire territory of the Czech Republic. In the case of December for the year 2021, there are only a few data. Since our analysis took place this month, not all the records of this month were added.

TABLE VIII. OCCURRENCE OF TRAFFIC ACCIDENTS ON THE TERRITORY OF THE CZECH REPUBLIC IN INDIVIDUAL MONTHS IN 2021

Month	Number of traffic accidents
January	7 325
February	6 747
March	6 096
April	7 043
May	8 792
June	9 580
July	8 737
August	8 933
September	9 091
October	9 666
November	8 560
December	75

Based on these data, we can find that number of traffic accidents that occurred on the roads averaged 7,387 each month. Since the average deviation of individual months is 1,230 (if we do not consider the month of December), we can say that individual month did not affect the occurrence of traffic accidents. Since the data were also divided into individual regions, we found in their analysis that the largest accident rate occurred mainly in the autumn and winter for the regions of Prague, Central Bohemia, Ustecky region, South Moravian region. In other regions, the incidence of traffic accidents was almost the same for each month – see Fig. 6.

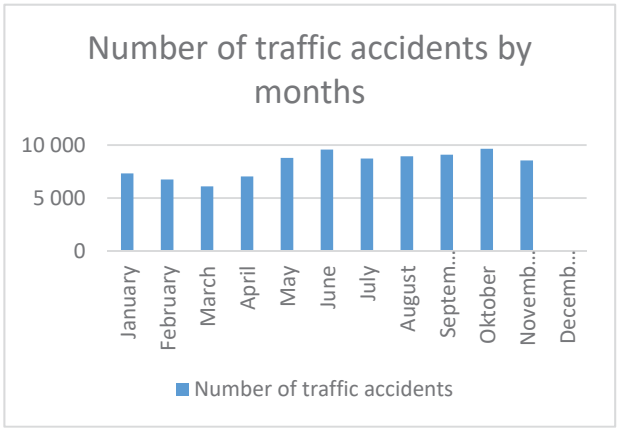

Fig. 6. Number of traffic accidents by months - graph

Further analysis of the issue of road accidents caused by alcohol and drugs. In this case, we compared data for individual regions of the Czech Republic but also in the whole territory for the year 2021.

TABLE IX. OCCURRENCE OF TRAFFIC ACCIDENTS IN THE CZECH REPUBLIC CAUSED BY THE INFLUENCE OF ALCOHOL AND DRUGS ON DRIVERS IN 2021

Alcohol and drug use	Number of traffic accidents
Not	46 910
Yes, blood alcohol content up to 0.24 ‰	166
Yes, blood alcohol content from 0.24 ‰ to 0.5 ‰	253
Yes, blood alcohol content from 0.5 ‰ to 0.8 ‰	217
Yes, blood alcohol content from 0.8 ‰ to 1.0 ‰	185
Yes, blood alcohol content from 1.0 ‰ to 1.5 ‰	604
Yes, blood alcohol content 1.5 ‰ and more	2 476
Under the influence of drugs	293
Under the influence of alcohol and drugs	74
Not detected	39 465

According to the recorded data, most traffic accidents were not caused by any drugs. If the traffic accident was caused by alcohol, it occurred in the case of alcohol with a blood content of more than 1.0 ‰. It was also noted that in many cases it was not established whether the accident was caused by drugs, which could affect the recorded data on the use of these substances. For a graph representation see Fig. 7.

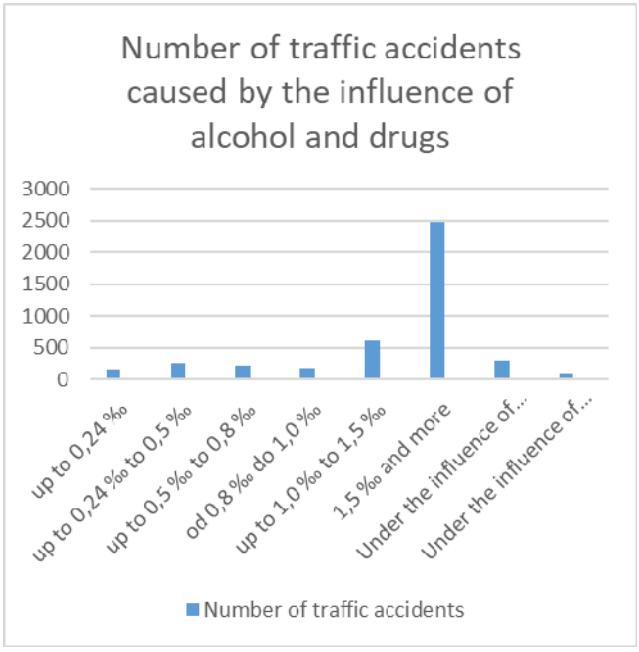

Fig. 7. Number of traffic accidents caused by the influence of alcohol and drugs - graph

Another analysis we were dealing with was the total material damage on the roads caused by traffic accidents. We analyzed them from the point of view of individual regions. We found out which were the most expensive and on the other hand the cheapest ones in individual months. We also compared material damage in the whole territory of the Czech Republic for individual months.

TABLE X. MATERIAL DAMAGES FOR EACH MONTH FOR 2021

Moon	Total material damages for each month	
	Total material damages for 2021 CZK	Total material damage for 2021 €
January	4 895 232	196 927,83
February	4 397 919	176 921,68
March	4 024 483	161 898,91
April	4 590 849	184 682,96
May	5 946 966	239 237,51
June	6 259 749	251 820,30
July	5 601 202	225 327,94
August	6 183 445	248 750,70
September	5 952 897	239 476,10
October	6 740 047	271 141,97
November	6 228 825	250 576,27
December	56 157	2 259,11

Based on these data, we found out that the most expensive month for 2021 was month of October with total material damage of CZK 6,740,047, which is calculated as CZK 271,141.97. However, high total material damage for 2021 also included the months of June, August, and November. The months with the lowest total material damage are February, March, and April if we do not take into account the month of December due to small number of recorded data – see Fig. 8.

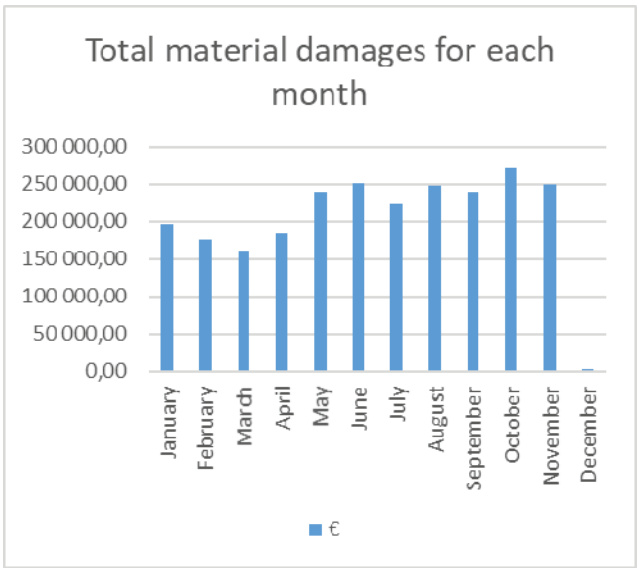

Fig. 8. Total material damages for each month - graph

As regards the comparison of total material damage for each region of the Czech Republic, the biggest material damage was caused in the regions of Prague, the Central

Bohemian region, the Ustecky region, and the Moravian-Silesian region.

TABLE XI. MATERIAL DAMAGES FOR EACH REGION FOR 2021

Country	Total material damage for individual regions		
	Total material damage CZK	Population	Conversion per inhabitant of CZK
territory of the capital city of Prague	12 218 740	1 251 072	9,76
Central Bohemian region	10 387 835	1 256 850	8,26
South Bohemian region	3 265 231	637 723	5,12
Pilsen region	2 932 782	571 831	5,13
Ústí nad Labem region	5 472 310	835 814	6,55
Hradec Králové region	3 199 191	547 903	5,84
South Moravian region	3 884 805	1 152 819	3,37
Moravian-Silesian region	5 005 186	1 244 837	4,02
Olomouc region	3 107 294	640 410	4,85
Zlin region	2 346 069	590 527	3,97
Vysočina region	3 163 677	513 677	6,16
Pardubice region	2 557 616	520 584	4,91
Liberec region	1 994 906	433 948	4,60
Karlovy Vary region	1 342 129	304 602	4,41

Looking at the total material damage in terms of the total population of the region, we found out that even the region of Vysočina had high total material damage despite lower total material damage calculated per citizen.

VII. Conclusion

In the first semester, we have done a lot of work on the project. We successfully analyzed all the form items, decided which data is needed, created a database in which all the data was entered, created a program that displayed individual data in the form of graphs, and perform basic analysis of individual data.

This project is not short-term work. It is possible to continue to work on this project for several semesters. The number of people working on this project should increase over time. It would be appropriate to consider further analyses, but not only the basic ones above the separate items of the form. It would be interesting to perform a deeper analysis of different combinations of individual items. These options could identify how individual factors affect each other. Analyses of this project can be helpful to prevent any road accidents from fatal consequences. Furthermore, we are considering various analyses of individual roads (motorway, road 1. 2nd, and 3rd class). We can consider the comparisons of the traffic situation in individual years that means how the situation on the roads has changed over the years.

In the future, we would like to replace the recording of data into .csv files with this information system because of the errors we found in .csv files (missing data or wrong form of data). We would like to have the possibility to enter the data using this information system, which would automatically check the entered data and thus avoid unnecessary inconsistencies.

As another improvement of this information system, we propose to focus on the improvement of the filtration of the data displayed. There are some possibilities to make it more detailed. With the data filtered by better filtration of data, the possible analyses could be performed more effectively.

Acknowledgment

This publication was realized with the support of Operational Program Integrated Infrastructure 2014 - 2020 of the project: Intelligent operating and processing systems for UAVs, code ITMS 313011V422, co-financed by the European Regional Development Fund.

References

[1] Behounek, L., Novák, V.: Towards Fuzzy Patrial Logic. In 2015 IEEE Internal Symposium on Multiple-Valued Logic, 2015.

[2] Bryla, B.: Oracle Database 12c The Complete Reference, Oracle Press, 2013, ISBN – 978-0071801751

[3] Burleson, D. K.: Oracle High-Performance SQL Tuning, Oracle Press, 2001, ISBN - 9780072190588

[4] Delplanque, J., Etien, A., Anquetil, N., Auverlot, O.: Relational database schema evolution: An industrial case study, IEEE International Conference on Software Maintenance and Evolution, ICSME 2018, Spain, 2018, pp. 635-644

[5] Eisa, I., Salem, R., Abdelkader, H.: A fragmentation algorithm for storage management in cloud database environment, Proceedings of ICCES 2017 12th International Conference on Computer Engineering and Systems, Egypt, 2018

[6] Kriegel, H., Kunath, P., Pfeifle, M., Renz, M.: Acceleration of relational index structures based on statistics, 15th International Conference on Scientific and Statistical Database Management, 2003

[7] Lien, Y.: Multivalued Dependencies With Null Values In Relational Data Bases. In Fifth International Conference on Very Large Data Base, 1979.

[8] Mirza, G.: Null Value Conflict: Formal Definition and Resolution, 13th International Conference on Frontiers of Information Technology (FIT), 2015.

[9] Moreira, J., Duarte, J., Dias, P.: Modeling and representing real-world spatio-temporal data in databases, Leibniz International Proceedings in Informatics, LIPIcs, Volume 142, 2019

[10] Schreiner, W., Steingartner, W. and Novitzká, V.: A Novel Categorical Approach to Semantics of Relational First-Order Logic, Symmetry-Basel, Vol. 12, No. 10, MDPI, OCT 2020, doi: 10.3390/sym12101584

[11] Sawant, M., Kinage, K., Pilankar, P., Chaudhari, N.: Database Partitioning: A Review Paper, International Journal of Innovative Technology and Exploring Engineering, 2013

[12] Kuhn, D., Kyte, T., Expert Oracle Database Architecture: Techniques and Solutions for High Performance and Producitivity. 2021, ISBN – 9781484274989

[13] MacDonald M., Pro WPF 4.5 in C#: Windows Presentantion Foundation in .NET 4.5. 2013, ISBN – 9781430243656

[14] Troelsen, A., Japikse P., Pro C# 9 with .NET 5: Foundational Principles and Practices in Programming. 2021, ISBN - 9781484269381.