

FORENSIC INTELLIGENCE APPROACH ON TRAFFIC ACCIDENTS IN THE BRAZILIAN FEDERAL DISTRICT

VINÍCIUS DE OLIVEIRA LIMA & JOSÉ MARCOS V. A. GOIS

Federal District Civil Police, University of Brasilia, Brazil

ABSTRACT

Motor vehicle collisions (MVC) are a major problem in Brazil, the yearly number of deaths caused by MVC in the country is comparable to those seen in war zones around the globe. One of the first steps to deal with this problem is using appropriate tools to study it. The Civil Police of the Federal District (PCDF) is the official authority in charge of forensic analysis of MVCs that resulted in deaths or injuries. Each MVC attended by a forensic expert from PCDF generates a considerable amount of information including address, GPS coordinates, street description, type of accident, damages, vehicle identification, and others. All this data is stored in a JSON database. This work describes an interactive platform created by a forensic expert and an undergraduate student using Microsoft Power BI to extract data (through SQL queries) and generate reports from this database. These reports provide meaningful graphs such as geographical heat maps, bar plots regarding MVCs, the road speed limit and the distribution of accidents per month. When observing the first 6 months of 2020, we are able to notice three main spots of MVCs that required a reconstruction analysis. We studied each case within the three spots and a pattern for each one appeared. With this information, we provided solutions for traffic engineering in order to prevent further similar accidents. Therefore, with such insights, we proved our system can give support to government agencies to adapt their policies and introduce new ones to help curb the number of deaths and injuries caused every year due to MVCs in Brazil.

Keywords: forensic, traffic, crash collision, accident, reconstruction, data mining.

1 INTRODUCTION

In 2019, traffic accidents in Brazil occupied the 10th position in the rank of causes of deaths, in a list filled mainly with diseases, like cancer (1st) and heart pathologies (2nd) [1]. In 2018, there were 33,625 deaths in Brazilian streets and roads, according to the Health Ministry. Although this number is decreasing, the number of traffic injuries (fatalities excluded) is getting higher each year and, before the COVID-19 pandemic, in 2020, traffic accidents were responsible for 60% of public hospitals capacity in Brazil [2].

In 2018, with similar statistics, the Federal District registered 9.4 casualties per 100,000 habitants [3], [4]. While United Nations 2030 Agenda has proposed to halve traffic injuries by 2020, Brazil still have much to work on. One of the key points cited by the Agenda is data needs [5]. Data and statistics are important not only to monitor traffic accidents, but also to provide insight on road safety policies.

The Traffic Reconstruction Sector, in the Police Department of the Federal District, is responsible for the regional forensics analysis of a motor vehicle collision (MVC) that resulted in injury or death. The experts' job is to get to the scene, collect and/or describe evidences like skid marks, vehicle damage and collision fragments, in order to write a report to tell authorities what happened, in other words, a reconstruction of the accident. In 2016, all this data started to be collected digitally, exclusively for storage purposes. Data was inserted to a web server that we named *Croqui Digital* (Digital Sketch, free translation).

Nowadays, with a consisted database, the aim of this project is to gather all this data collected by field experts, and convert it into meaningful information and get a macro view of crash accidents inside the Federal District. Since one of data is the GPS coordinate of the event, a heatmap can be built, revealing where the main spots of accidents concentrate.



However, with the quantity and diversity of data acquired, not only a map is interesting, but also a correlation with any other information available. For instance, we can study the relationship between road and vehicle conditions, and the cause of accidents (according to the forensic report indexed). With the right correlation data, decision-making policy can be more accurately delivered.

Although it also carries some challenges, the most efficient and cheapest solution available was working along with Microsoft Power BI. One of the challenges was to organize data so that Power BI could use as an input. Prior to the application, there was no necessity for an organized and structured data storage. As a result, our first task was to process data in order to rearrange them into organized tables.

After accomplishing this step, the simplicity of Power BI brought a robust web-based application. Many useful forensic intelligence outputs can be extracted from it. We just needed to decide which ones we want to display. Hence, it was important to understand what parameters we had and which ones can provide us useful information. We found out that some interesting questions could be answered, for example: how mobility changed during the COVID-19 pandemic? How injuries and road speed limit are connected? What are the main causes of the accidents in a specific region? These will be discussed later on based on our system's data, evincing the applicability of it.

Finally, once the application was finalized and tested, with data displays already selected, we choose recent information, from 2020, to observe MVCs' characteristics. The heat map revealed three main concentrations of accidents. We studied them and provided solutions to each location.

2 METHODOLOGY

Although the processes themselves were simple, the right manipulations needed a close attention. First, it was necessary to understand the data available, and how to it was archived. One of the main concerns was how to deal with registered data errors, and what work should be applied around to diminish them. As mentioned, the database provided was not well organized, so it needed to be "cleaned" and rearranged. Although we could work directly with Microsoft Excel (for a more familiar application) to treat and display data, we found out the work exclusively with Power BI was easier and adequate to the purposes [6].

2.1 Data processing

Our project was divided into 6 ordered steps. Although we will explain it briefly, it is important to understand each one of them, so the output can be cohesive and the project can inspire similar studies.

- Understanding inputs.
- Data comprehension.
- Data preparing.
- Data modelling.
- Analysis.
- Implementation.

The first step was to understand how data was stored and the tools we would need to extract it. All information collected from the forensics professionals into the database consisted of a JSON format that kept everything into a single (enormous) table [7]. We realized we needed to use SQL in order to retrieve data, break the JSON and separate information.



There comes the second step. Withal, information was all scrambled in this unique table. Although JSON is a light and easy-to-read format, we were required to comprehend how the different variables were archived in the first place.

For data preparing and data modelling we went for the help of SQL. We could not just use a simple `JSON_EXTRACT` to pull information out, since it returns data jointly with quotation marks. The way to contour it was using `JSON_UNQUOTE` along with the `JSON_EXTRACT` [8].

```
"JSON_EXTRACT(<Table>, '$.DataJSON[0].SubData[0].Value')"  
"JSON_UNQUOTE(JSON_EXTRACT(croqui, 'DataJSON[0].SubData[0].Value'))"
```

The last two steps were actually creating our desire tables. We defined which parameters or which columns we want to display in our later dashboard. After selecting what was "important information", we use the `SELECT * WHERE` statements to retrieve our information [9].

Once we figure out how data is stored and how to separate them with SQL codes, it is possible to output the adequate information. When data is inside Power BI, we choose a relationship between the desired parameters, so that we get a synchronized and correct dashboard automatically [10].

The way we use to treat some of the errors was using some basic statistics concepts and identify which data was discrepant from the others. We noticed that the main cause of errors was not a system defect, but basically a misuse by the operators, who provided incorrect or unreal data. By understanding the medium and the variance of some of our parameters we could find those discrepancies and eliminate them.

Finally, when implementing our application, we opt to have our dashboard every day updated. To achieve this, we chose a personal computer that connects with the database (*Croqui Digital*), sending queries commands. When a result is received, a gateway is called, sending data to Power BI, hence, refreshing the dashboard. However, we need to have a computer turned on to keep it updated.

3 INFORMATION DISPLAY

The application's dashboard consists of six pages, being every page with a filter for the user to select the time period he/she wants. The main information of each page is as follow:

1. A heat map and a histogram linking the quantity of accidents versus the road speed limit defined (Fig. 1).
2. Histograms showing the relationship with road conditions, topography, vehicle damages and road speed limit (Fig. 2).
3. Two maps to present separately the accident spots where we have at least one car involved and at least one motorcycle involved. Also, here there is a graph of the quantity of accidents through each month of the year (Fig. 3).
4. Map revealing only the cases that already have a conclusion report. Page 4 is just like Page 3, but with a mouseover effect revealing the conclusion report of each accident spot.
5. Information from the amount of accidents from each forensic professional, as well as the time each one took to get to scene and finalize their forensic work (Fig. 4).
6. Graphs from the police stations in charge of investigating and the amount evolution through the years (Fig. 5).



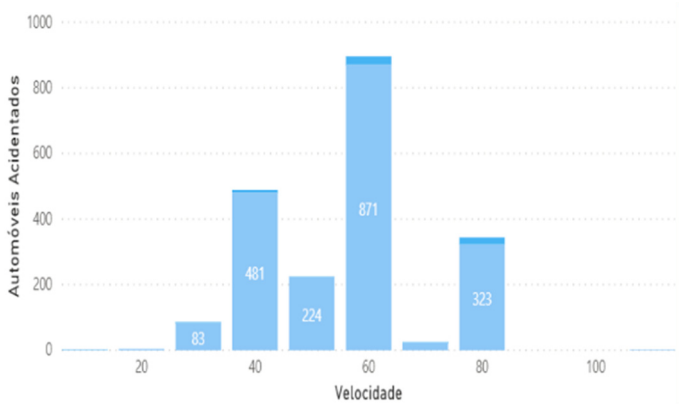
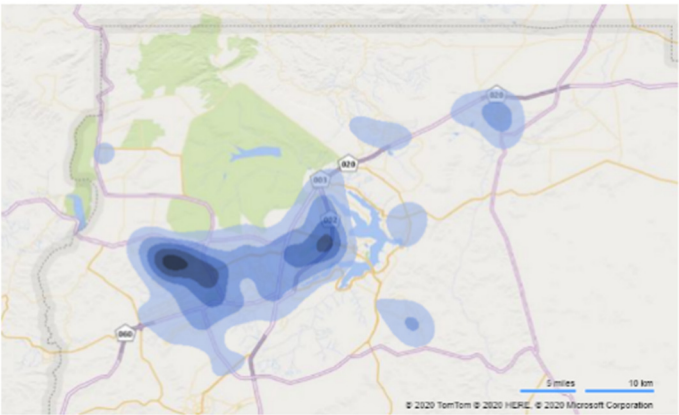


Figure 1: Main information of Page 1.

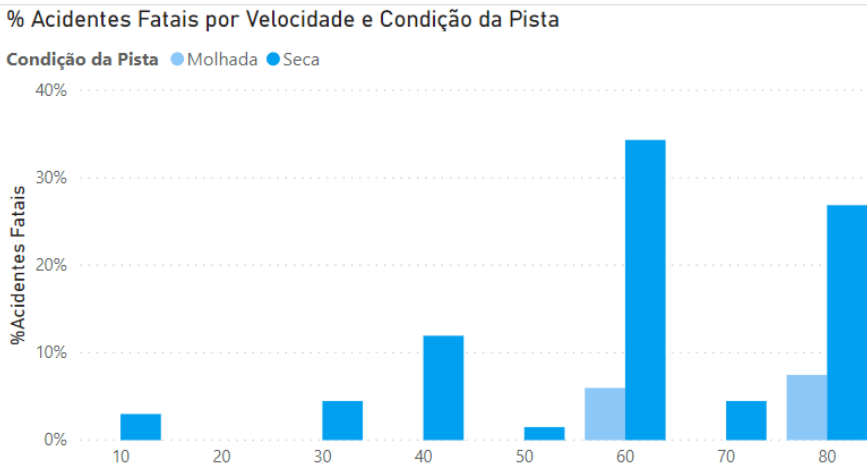


Figure 2: Example of a bar chart of Page 2, revealing information regarding road speed and their conditions (wet or dry).

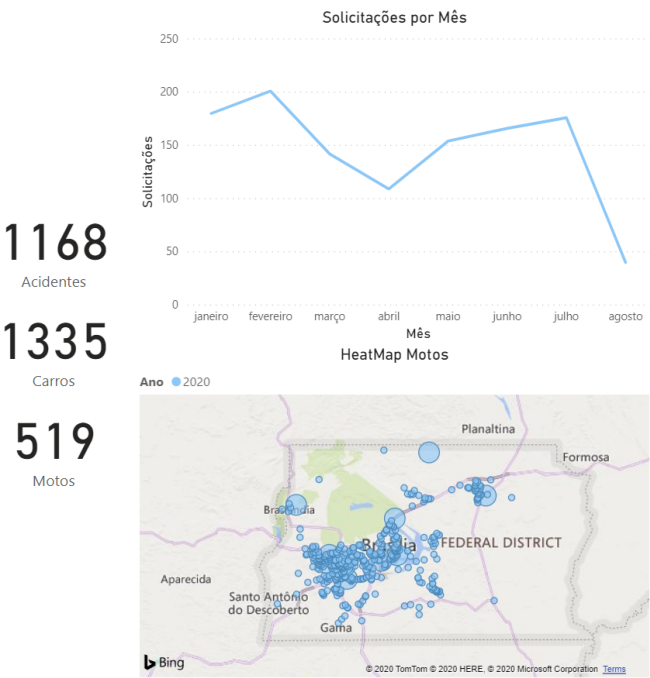


Figure 3: Part of Page 3, showing a plot of the number of accidents in each month and a map revealing accident spots where at least one motorcycle is involved.

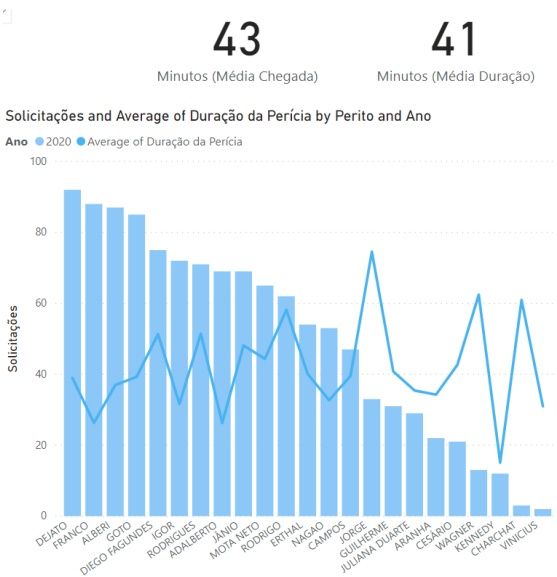


Figure 4: The bar chart shows the number of accidents each forensic professional attended and the line graph shows the average time each one took to conclude the crime scene analysis.

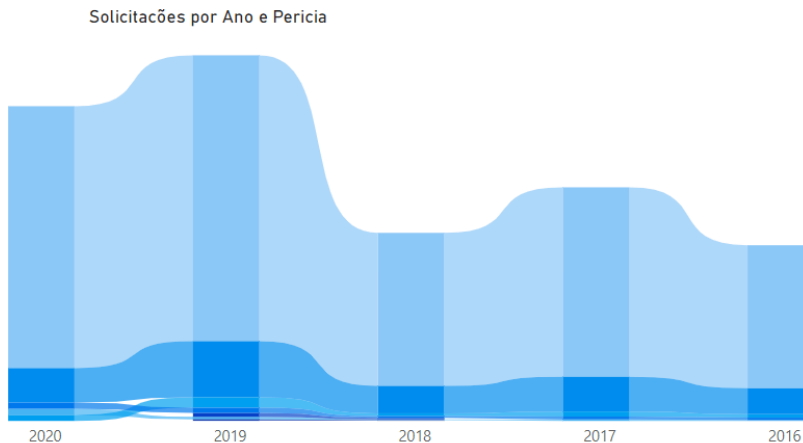


Figure 5: Example of Page 6, revealing the quantity of accidents throughout the years, separated by the type of accident (collision, runover, fatal, etc).

4 ANALYSIS AND RESULTS

The application resulted in interesting information that only georeferenced analysis can provide. The intelligence displayed is clearly useful for understanding how some parameters stimulate the occurrence of an accident. Some correlations observed were well-known and expected by experts, like speed limit, for instance. However, others were not, like the relationship between crashes and wet asphalts, as we will discourse here.

It is important to mention that database consists of data since 2016, but not all information is reliable since it was not mandatory to feed it back them. Some professionals used to collect information in the physical way (papers) and some others in the digital way (tablets). Additionally, at the begging of its creation, users tested *Croqui Digital* a lot and thus some details were not archived or actually wrongly inserted. Even though, some information was accurately used to provide trustful results. All we need to do is understand which parameters were inconsistent and which were not. Only by the end of 2019, the utilization of tablets was a must when in an accident crash scene. Therefore, we can fully trust all information after that.

Summarizing, before the end of 2019, we can say the amount of data is below the real one and thus quantity results cannot be taken for granted. Data that presented itself as inconsistent, in this period, was a few GPS locations (wrongly inserted) or empty left fields. Anyhow, in a quality perspective, the application demonstrated to be really useful as we will present a few examples in the following items.

4.1 Analysis of collisions and speed limit

One of our histograms showed that most of the accidents, in 2020, happens in a 60 km/h (37.2 mph) maximum speed road. Also, fatalities tend to occur more in that same speed limit road. But if we take all the sample we have, we will observe that fatalities are more frequent in an 80 km/h (49.7 mph) road. This make sense since we observe more severe accidents on highways. In addition, what may explain the majority of crashes in a 60 km/h road is the quantity of this sort of road in the Federal District (Table 1).



Table 1: Amount of vehicles regarding the speed limit of the road where the accident took place.

Gravity	1st semester of 2020		
	60 km/h	80 km/h	% of total
Injury	896	333	58%
Fatal	27	24	2%

It is worth highlighting that we consider a fatal event only when we have at least one corpse at the scene, since we do not have information of the victims afterwards. Hence actual fatal accidents are expected to be much higher than our data.

4.2 Analysis of cars versus motorcycles

Another interesting fact extracted from the application was the ration between cars and motorcycles involved in a collision. According to DENATRAN (National Traffic Organization of Brazil), in 2016, Federal District registered a fleet of 188,739 motorcycles, occupying the 6th place among other states, although it has the smallest area of them [11]. Therefore, we can expect a great number in motorcycles accidents.

In consonance with our expertise, we can observe in Fig. 3 that the first semester of 2020 presented a total of 1,168 injury/fatal accidents (or more accurately: accidents that called for a forensics analysis), in which involve 1,335 cars and 519 motorcycles. Because it is really rare to occur an accident with more than one motorcycle, we can estimate that collisions with this sort of vehicle represents 44% of the total. If we take only the fatal cases, the percentage is almost the same, 40%.

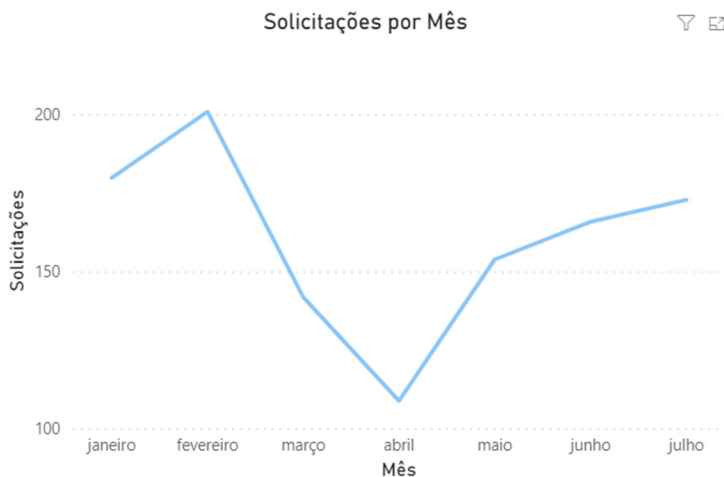


Figure 6: Crash accidents from January to July in 2020.

The numbers are disturbing, once we realize that many serious injuries and even disabilities are caused due to motorcycles crashes. The first semester of 2020 was also characterized by higher investments in the motorcycle production industry, probably due to

the pandemic situation demanding more delivery business [12]. The incentive of motorcycle production and traffic accidents are directly connected and thus need government's care and intervention.

4.3 Analysis of accidents and rain season

A common thought among population in general is that, on rain seasons, traffic accidents tend to happen more frequently. Our statistics showed differently.

Although the capital of Brazil has a typical tropical climate, with low moisture practically all year, November through March is considered the time of the year that concentrates rainfalls, with precipitations around 250 mm [13]. However, our data did not show an increase in crashes during this period. On the contrary, May was considered the month with the most registered quantity of accidents and September the least. Indeed, some other aspects that influences traffic mobility (and thus the probability of a vehicle to collide) should be taken into account, like school breaks and city events. However, if wet asphalts did contribute directly to an accident, we ought to expect higher number during the typical rain season. This study was not new, since it was already approached in another paper from other forensic researchers [14].

4.4 Analysis of mobile while in the COVID-19 isolation restrictions

From March to July in 2020, the pandemic situation forced authorities to impose partial lockdown restrictions in order to contain the spread of the coronavirus. Some of the restrictions consisted of schools and all of commercial activity closures. Also, it was highly recommended a home isolation for everyone, except healthy and security professionals.

Since less people were driving from one place to another, it was expected less traffic crashes in this period. Our application demonstrated a great decrease by the end of March and for the whole of April, revealing that people might be indeed following the isolation policy. However, for the next 3 months, numbers got higher each day, until it became similar to the numbers registered in 2019. Especially in May 2020, we had more accidents than in May 2019. Hence, even with all the strict measurements and recommendations, people only stayed home in the first days of lockdown.

Searching straight within crash accidents revealed a good way to measure city mobility. When compared to other sorts of measurements, like smart phones activity, it did demonstrate similar results. Therefore, our application could be used to secure decision-making regarding mobility issues.

4.5 Study of main spots for accidents and suggestion for an infrastructure intervention

Finally, by the end of the first semester in 2020, we used our heat map to check the main spots accidents were taking place (Fig. 7). We observed a concentration in three main sites, as the following:

- Site A: Centre of Brasília, with 18 cases;
- Site B: Centre of Taguatinga, EPTG highroad, with nine cases; and
- Site C: Hélio Prates Avenue, Ceilândia, with 17 cases.

We studied all 44 spots to understand if a pattern within which site could be observed. As mentioned before, our application permits checking the forensic report's conclusion, yet it is only available whether the report itself was properly demanded and registered in *Croqui*

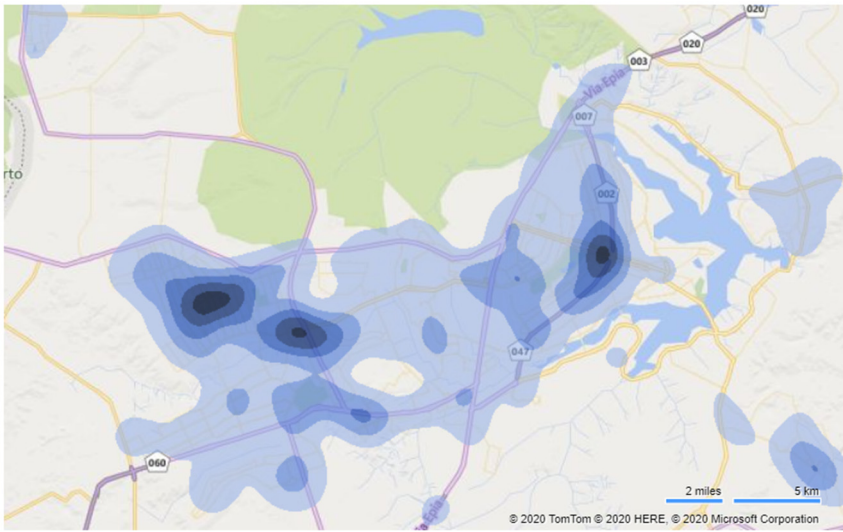


Figure 7: Three main spots revealing where most of the accidents concentrated.

Digital. Thus, many of our spots did not show a conclusion for that specific accident and we needed to look up the photos of each case and make our own interpretation. Additionally, we wanted to check how road conditions were and if traffic engineering was appropriate.

Site A was the widest area among the other two and thus it was difficult to focus on any specific road problems. Nevertheless, we observed a great quantity in accidents involving motorcycles and regarding traffic lights crossroads. Most of them were misconducts like red light traffic violations and interceptions. What we can see as an alternative solution is to place, in these specific crossroads, more red camera lights to monitor traffic violations or place a public agent to police driver's behaviour.

Site B is situated in a shopping region, close to a metro station, where it is well known for its heavy traffic and crowded sidewalks. Here we could notice two patterns: Rear-end collisions and running over accidents. The first was typically due to a U-turn that did not have a specific lane to access and thus contributed to a traffic jam just behind it. The second can be explained because pedestrians need to cross a wide road (five lanes) and sometimes are surprised by a fast vehicle unseen. Our proposed solutions are to build an underground crosswalk, a speed barrier nearby and eliminate the U-turns (access should be made a few kilometres ahead).

For Site C, almost all cases were again consequence of traffic light violations. Just like Site A, we recommend to supervise traffic with red light cameras. That Site covered a six blocks area, but only one of them already had a red-light camera. Furthermore, from midnight until 6 am, traffic lights are in an intermittent mode (yellow light blinking). Since some of the cases were at the dawn period, we rather recommend to leave the lights in the normal mode all day and night long.

All of these suggestions were putted together into a report and sent to the responsible traffic institutes of the district. Our goal is not only to provide information regarding injury accidents, but also to create support for a better prevention policy. Although 100% reliable data is only for a recent period, the application already provides meaningful information to support adequate public policy to prevent accidents.

5 CONCLUSIONS AND FUTURE PROSPECTS

The application has showed to be a great tool to understand events that resulted in a vehicle collision with a victim injury or even death. Once data is inserted into a database and adequately treated, we can output it in a graph format for an effective analysis. Our goal was to present information in a user-friendly environment in a free platform. Microsoft Power BI accomplished these requirements and therefore was our main software to work with. One difficulty was to understand how database was previous organized, since it was primarily developed only for storage purposes. With the support of SQL and some simple JSON statements we could create our structured tables and choose which parameters we want to display and how.

Since one of the principal data collected in field was a GPS coordinate, we could arrange accidents in a map view and analyse how it is distributed. With a heat map we can observe the main spots of accidents. We can check which brands and models are most involved in a collision and what is the relation between cars and motorcycles. We can also understand which elements from the road influences the most in an accident, like the topography or the speed limit.

Finally, we used the app to identify three main spots for accidents in the period from January to June 2020. These sites were studied one by one and understand what can be done to prevent similar events to occur in the future. This information was gathered and sent to traffic authorities, proving how it is a powerful tool for public policies.

Our next goal is to use some machine learning with our database and check if we can forecast when an accident is most likely to occur. The idea is not new and we already have some other articles to get support from, although we have few data, as known models recommend [15].

The web-based application proves that a forensic job is not restrained to crime scene and reports. Forensics professionals in Brazil have great amount of data that can be properly worked to provide intelligence and meaningful information and thus give support to for a better analysis of what is being investigating.

REFERENCES

- [1] Brasil reduziu, entre 2010 e 2018, em 32% a mortalidade por lesões de trânsito. www.portaldotransito.com.br/noticias/brasil-reduziu-entre-2010-e-2018-em-32-a-mortalidade-por-lesoes-de-transito/. Accessed on: 25 Jun. 2020.
- [2] Coronavírus já mata mais que acidente de trânsito por dia no Brasil. <https://veja.abril.com.br/brasil/coronavirus-ja-mata-mais-que-acidente-de-transito-por-dia-no-brasil/>. Accessed on: 25 Jun. 2020.
- [3] Estatísticas Nacionais. http://vias-seguras.com/os_acidentes/estatisticas/estatisticas_nacionais. Accessed on: 25 Jun. 2020.
- [4] Boletim anual de fatal 2018. www.detran.df.gov.br/wp-content/uploads/2018/06/df_boletim_anual_fatal_2018-1.pdf. Accessed on: 25 Jun. 2020.
- [5] United Nations, A/RES/70/1. *Transforming Our World: The 2030 Agenda for Sustainable Development*.
- [6] Ferrari, A. & Russo, M., *Introducing Microsoft Power BI*, Microsoft Press, 2016.
- [7] Functions that search JSON values. <https://dev.mysql.com/doc/refman/8.0/en/json-search-functions.html>. Accessed on: 25 Jun. 2020.
- [8] JSON_EXTRACT(), Return data from a JSON document in MySQL. https://database.guide/json_extract-return-data-from-a-json-document-in-mysql/. Accessed on: 25 Jul. 2020.



- [9] MySQL workbench documentation. <https://dev.mysql.com/doc/workbench/en/>. Accessed on: 30 Jun. 2020.
- [10] Data Analysis Expression (DAX). <https://docs.microsoft.com/pt-br/power-bi/guided-learning/introductiontodax?tutorial-step=1>. Accessed on: 30 Jun. 2020.
- [11] Wikipedia, Lista de municípios do Brasil por frota de motocicletas. https://pt.wikipedia.org/wiki/Lista_de_munic%C3%ADpios_do_Brasil_por_frota_de_motocicletas. Accessed on: 25 Jul. 2020.
- [12] Produção de Motocicletas aumenta no primeiro semestre. <https://agenciabrasil.ebc.com.br/economia/noticia/2019-07/producao-de-motocicletas-aumenta-84-no-primeiro-semester>. Accessed on: 25 Jul. 2020.
- [13] Climate-Data: Clima Brasília. <https://pt.climate-data.org/america-do-sul/brasil/distrito-federal/brasil-852/>. Accessed on: 25 Jul. 2020.
- [14] Telles, B., Kawano N., Andrade, C.A. & Vicente, A.N., Forensic intelligence as a useful tool for reducing traffic fatalities: The Brazilian Federal District case. *Revista Brasileira de Criminalística*, 5(2), p. 7, 2016.
- [15] Lee, J., Kwon, S., Yoon, T. & Lee, J., Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul City study. *Applied Sciences*, 10(1), p. 129, 2019.

