Capter 149

Recife, Pernambuco, Northeast Brazil: comparison between cases of dengue, chikungunya and zikavirus in the year 2021, using the Python language

Angélica Félix de Castro
Department of ComputingFederal Rural University of the Semi-Arid Mossoró, Brazil
angelica@ufersa.edu.br

Amanda Gondim de Oliveira
Department of ComputingFederal Rural University of the Semi-Arid Mossoró, Brazil
amandagondim@ufersa.edu.br

George Felipe Fernandes Vieira
Department of Computing – UFERSA Scholarship Holder University of the Semi-Arid Mossoró, Brazil
georgeffv1@gmail.com

ABSTRACT

Arboviruses are diseases caused by so-called arboviruses, which include dengue virus, zika virus, chikungunya fever and yellow fever. They are a public health problem in the world mainly due to their potential for dispersion, their ability to adapt to new environments and hosts, the possibility of causing extensive epidemics and their universal susceptibility.

The present work consists of a study of cases of arboviruses in Recife, capital of the state of Pernambuco, Northeastern Brazil, in the year 2021. First, data were extracted from official sources of the Municipality of the city of Recife on cases of dengue, chikungunya and zika virus in the year 2021. After this extraction, Python commands were applied to understand how the outbreak of these specific diseases occurred in Recife and to generate heat maps of the neighborhoods most affected by arboviruses. It was possible to conclude that arboviruses are not worrying reasons for death and do not cause high hospitalization of patients. As for the heat maps, the neighborhoods of Varzea, Ibura, Iputinga, Imbiribeira, Vasco da Gama and Dois Unidos were the most affected by the presence of the three diseases. In comparative terms, dengue and chikungunya have similar symptoms (fever and headache).

Keywords: Arboviruses, Python, Heat Maps

1 INTRODUCTION

It is notorious that Brazil has been afflicted by diseases with serious repercussions for the health of its population according to operational and epidemiological indicators, including Arboviruses (or Arboviruses), a set of pathologies formed mainly by Dengue, Chikungunya and Zikavirus [1]. These three injuries were flagged by the Ministry of Health - the highest entity for guiding and formulating policies at the federal level that must be followed by health professionals in the State and Municipal spheres - as a neglected disease, which deserves to receive special attention from the health professionals who act in care and surveillance and managers at the federal, state and municipal levels.

The health area has been one of the areas most benefited by technology, as it is considered to have classic problems, possessing all the necessary peculiarities to be instrumentalized by such systems [2]. In this context, the evaluation of classifiers and their implementation in a system that helps in the diagnosis of

diseases such as dengue, chikungunya and zika, which today affect a large part of the world's population, is entirely possible [3].

The area of science and data analysis has also been widely used in the study and proliferation of these arboviruses. When it comes to computing applied to health, since the 2000s until today, studies have been developed using neural networks, machine learning algorithms, even statistical models, which aim to contribute to the fight against dengue and reverse this alarming situation that the disease looms over the nation [4].

Data Science can be defined as the processes, models and technologies that study data throughout its life cycle: production, storage, transformation, analysis and disposal [5]. Some operations can take place in the data analysis stage: it is possible to resort to graphs, tables and other visual aids that allow a clearer visualization of the collected data. A pivot table, for example, can help sort and filter data from different variables. During this stage of data analysis, the use of the Python language in a database can be extremely beneficial, as it has libraries with mathematical models, formulas, statistical tests and machine learning [5].

The main objective of this work is to analyze data on arboviruses in the city of Recife, capital of Pernambuco, Northeast Brazil, in the year 2021, by obtaining state government databases. The idea is to apply concepts of analysis and data science, using the Python language, to understand how these diseases behaved in Recife.

2 THEORETICAL REFERENCE

2.1 ARBOVIRUSES

In large cities, the proliferation of infectious diseases occurs with greater intensity and epidemic control is more challenging due to the increase in population and greater social interaction. Not only Brazil, but also several countries in the world, mainly with a tropical climate, have epidemics [6] of diseases transmitted by the Aedes aegypti mosquito, popularly known as the Dengue mosquito. In addition to Dengue, the mosquito also transmits diseases such as Zika and Chikungunya, all of which are part of a larger group of diseases called arboviruses.

Dengue can manifest itself in different ways. It can be asymptomatic or can progress to serious conditions, such as dengue hemorrhagic fever. Chikungunya and zika are diseases caused by viruses, called CHIKV and KIKV respectively, with a set of symptoms that is quite similar to that of dengue. Zika virus transmission can also happen through sexual contact with a person who has been infected by the disease [7] and [8].

2.2 DATA ANALYSIS

Exploratory Data Analysis (AED) had its beginnings with J.W. Tukey, in 1977, and aims to increase the researcher's knowledge about a population from a sample. In this way, it is possible to describe the AED as a set of methods suitable for the collection, exploration and description and interpretation of numerical

Development and its applications in scientific knowledge

Recife, Pernambuco, Northeast Brazil: comparison between cases of dengue, chikungunya and zikavirus in the year 2021, using the Python language

data sets. These methods allow data exploration in order to identify patterns of interest and the representation of data characterized by these patterns. Although EDA techniques are simple, they are generally robust methods (valid for a wide range of situations and models) and resistant (insensitive to gross errors or extraneous data) [9].

B.1) Use of Python in Exploratory Data Analysis

Python is a programming language widely used in data analysis activities and unlike languages like R, Python is a general purpose programming language, that is, it is not exclusive to data analysis activities [9]. The most relevant features of Python are its ability to integrate with other languages and its mature library system. Some Python libraries are strongly analytical and offer data science tools composed of highly optimized functions to work with, optimal memory configuration, ready to perform script operations with optimal performance [10].

A. *Recife, capital of Pernambuco, Northeast Brazil*

With 98,312 km², Pernambuco is one of the twenty-seven Brazilian states. Located in the east center of the Northeast Region, its coast is bathed by the Atlantic Ocean. Its capital is the city of Recife, with a total land area of 218,843 km² and an estimated population of 1,661,017 people [11]. In Fig. 1, it is possible to view: (a) the map of Brazil with emphasis on the state of Pernambuco; (b) the state of Pernambuco with emphasis on the capital Recife and (c) Metropolitan Region of Recife and Political Administrative Regions.

Fig. 1. (a) Map of Brazil highlighting the state of Pernambuco; (b) state of Pernambuco highlighting the capital Recife and (c) Metropolitan Region of Recife and political administrative regions.
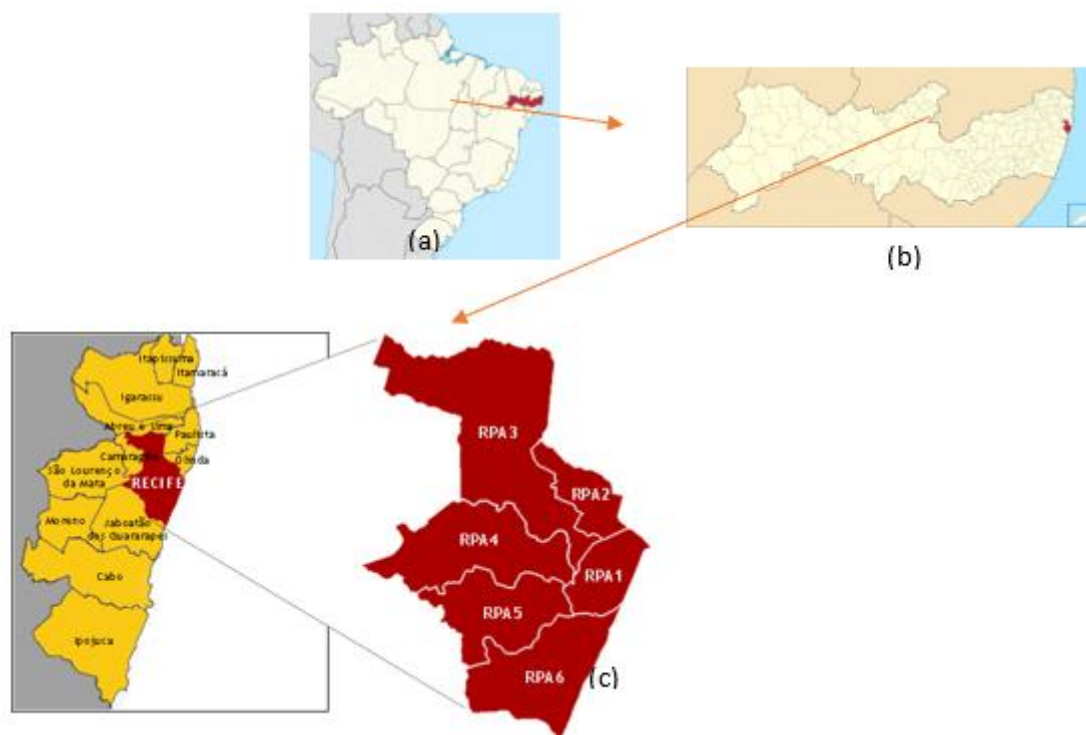
3 APPLIED METHODOLOGY

The present work is characterized as an exploratory study that evaluated the notification data of arboviruses, in the year 2021 in the metropolitan area of Recife in the state of Pernambuco. The research was carried out based on information available on the official portal of the City Hall of Recife, which registered 9217 cases of dengue, 15035 cases of chikungunya and 533 of zikavirus, in the year 2021.

For this work, the data were studied and analyzed using techniques and tools from the area of data analysis in order to provide a better visualization and interpretation of the raw data available by the City Hall.

From a practical point of view, the implementation of this work was carried out using the Python programming language due to its ability to integrate with other languages and its mature library system, which offer data science tools composed of highly optimized functions. The Python libraries used in this work were: Pandas, Matplotlib, Numpy and Pylab.

After extracting and pre-processing the data, the analysis and mining phase began. The results of these analyzes will be presented below.

Most common symptoms

According to [12], dengue in Pernambuco does not have a high fatality rate, as most people are cured and the hospitalization rate is very low (about 4%). In the entire state of Pernambuco, in the year 2021, there were 13,296 confirmed cases of dengue, 18,100 confirmed cases of chikungunya and another 11 cases of zika. Of these numbers, 4 deaths were identified [13]. It's a low mortality rate.

When analyzing dengue and chikungunya data from the official Pernambuco database, some symptoms present in both cases were identified. In Figure 2, it is possible to see these symptoms along with the presence of dengue.

Fig. 2. Sintomas presentes na dengue e número de internações
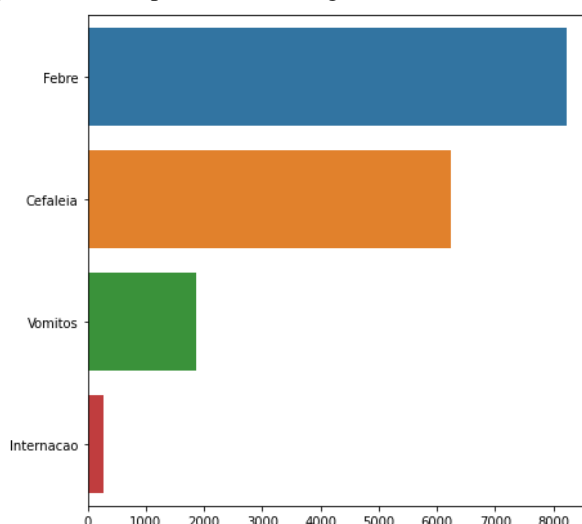
In Figure 3, the same symptoms shown in Figure 1 can be seen, but this time, in relation to chikungunya.

Fig. 3. Symptoms present in cases of chikungunya and number of hospitalizations
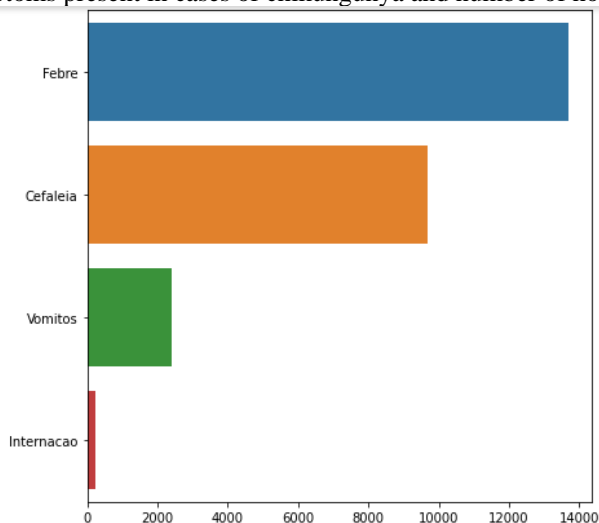
It is noticed that in both dengue and chikungunya, the symptom that most appears is fever, followed by headache. A reasonable number of people who presented vomiting throughout the disease were also identified. Even so, the numbers of hospitalizations are low and the number of deaths caused by these diseases is small. In the zika virus database, there are no symptoms recorded.

Most Affected Regions in the City

With the use of Python libraries, it was possible to identify the Recife neighborhoods that had the most detected cases of dengue, chikungunya and zikavirus in the year 2021; as well as generate heat maps of these detections.

Table 1 shows the districts, the type of arbovirus and the amount detected. In this table, only the 10 neighborhoods with the most dengue fever, the 10 neighborhoods with the most chicungunya and the 10 neighborhoods with the most zika virus were identified. Therefore, the existence of some blank cells.

Table 1. Table with the relation of arboviroses in the year 2021 x quantity of cases detected in the neighborhood

Neighborhood	Cases detected in the year 2021		
	Dengue (total of 9217 cases in 2021)	*Chikungunya (total of 15035 cases in 2021)*	*Zikavírus (total of 533 cases in 2021)*
Varzea	523	543	32
Ibura	520	855	17
Iputinga	415	456	20
Imbiribeira	402		
Vasco da Gama	304	496	22
Dois Unidos	271	707	18
Santo Amaro	271		

Neighborhood	Cases detected in the year 2021		
	Dengue *(total of 9217 cases in 2021)*	*Chikungunya* *(total of 15035 cases in 2021)*	*Zikavírus* *(total of 533 cases in 2021)*
Boa Viagem	267		
Nova Descoberta	247		21
Água Fria	245	757	
Campo Grande		584	
Cohab		493	49
Afogados		436	20
Linha do Tiro		425	
Cordeiro			28
Casa Amarela			16

In addition, it was possible to create heat maps from the data in Table 1. Figures 4, 5 and 6 show the heat maps with the neighborhoods most affected by dengue, chikungunya and zikavirus, respectively.

Fig. 4. Heat map with the neighborhoods that had the most cases of dengue in 2021, in Recife. The Varzea neighborhood was the most affected, with 523 cases.
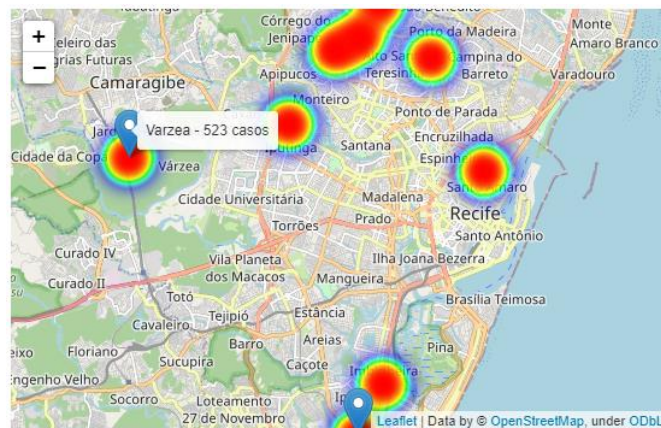

Fig. 5. Heat map with the neighborhoods that had the most cases of chikungunya in 2021, in Recife. The Ibura neighborhood was the most affected, with 855 cases.
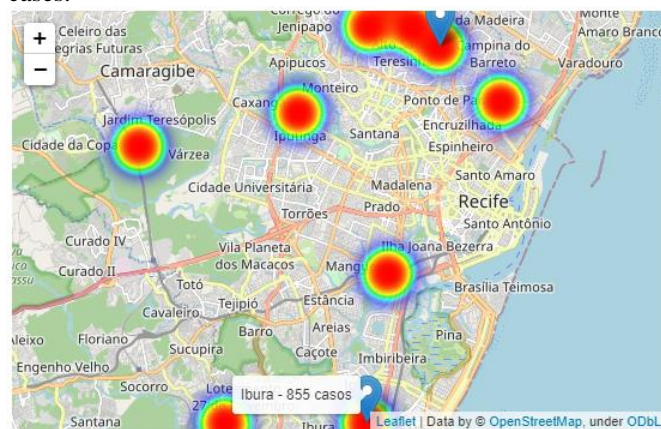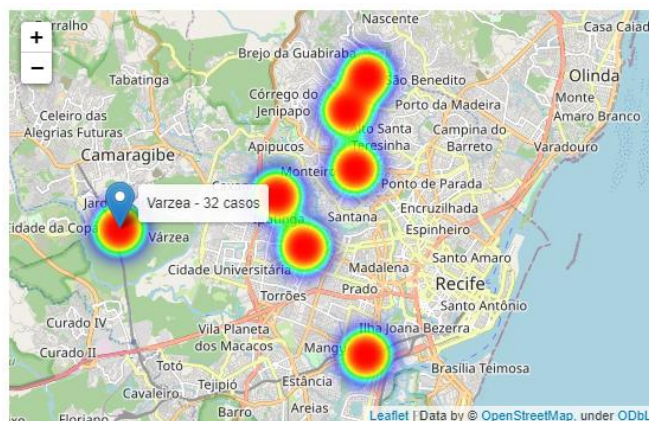
Fig. 6. Heat map with the neighborhoods that had the most zikavirus cases in 2021, in Recife. The Várzea neighborhood was the most affected, with 32 cases.

3 RESULTS AND CONCLUSION

The use of science and data analysis in healthcare has grown in recent years. Being able to extract aggregated information from data has been of great value in decision support.

Arboviruses are endemic tropical diseases that deserve special attention from health professionals who work in primary care and health surveillance and from managers at the federal, state and municipal levels. In this sense, the present work computationally evaluated data on arboviruses in Recife, PE, in the year 2021, in order to analyze this health situation.

After the application of Python commands and libraries in the official databases of the Pernambuco City Hall, some conclusions were obtained:

- Arboviruses have a low mortality rate;
- The most common symptoms in dengue and chikungunya are fever and headache;
- The hospitalization rate of people detected with arboviruses is low;
- Heat maps helped to visualize arboviruses in Recife neighborhoods;
- The Varzea, Ibura, Iputinga, Vasco da Gama and Dois Unidos neighborhoods deserve special attention as there was confirmation of the three arboviruses in 2021

As future work, it is intended to apply these techniques and methodologies in other Brazilian cities, in order to analyze the situation of the presence of arboviruses in these cities.

ACKNOWLEDGMENT

REFERENCES

[1] P.M.S. Silva e S.A.Siebr, "A construção do repositório de dados da UFPB: a experiência com o dataset de Arboviroses", Métodos e Pesquisa em Administração, Seção: Pesquisas científicas. v. 4, n. 1, p. 48-54, 2019.

[2] N.S. Nilson, Principles of Artificial Inteligence. Springer Verlag, Berlin, 1982.

[3] W.C. Lima, "Estudo de classificadores e construção de dataset para diagnóstico de dengue, chikungunya e zika", Trabalho de conclusão de curso, Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná, 2017.

[4] P. I. Andrade e L.V. T. Brilhante, "Predizendo epidemias de Dengue, no Distrito Federal, utilizando algoritmos de Regressão", Monografia de Graduação, Universidade de Brasília, 2018.

[5] F. Amaral. 2016. Introdução à Ciência de Dados: Mineração de Dados e Big Data (1st. ed.). Vol. 1. Alta Books.

[6] G. Cunha, "Surto, Epidemia, Endemia e Pandemia: Você sabe a diferença?", 2017. Disponível em: https://blog.enem.com.br/surto-epidemia-endemia-e-pandemia-voce-sabe-diferenca/. Acesso em: 07 set 2022.

[7] Rede D´or. Chikungunya. Disponível em: https://www.rededorsaoluiz.com.br/doencas/chikungunya. Acesso em: 07 set 2022.

[8] Rede D´or. Zika. Disponível em: https://www.rededorsaoluiz.com.br/doencas/zika. Acesso em: 07 set 2022.

[9] Tukey, John W. 1977. Exploratory Data Analysis. 1st Ed. Pearson. 688p.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[10] G. R. Lopes, A. W. S. Almeida, A. C. B. Delbem and, C. F. M. Toledo. 2019. "Introdução à Análise Exploratória de Dados com Python". In: Escola Regional de Computação Aplicada à Saúde (ERCAS 2019).

[11] Site oficial do IBGE. 2022. Disponível em: https://www.ibge.gov.br/cidades-e-estados/pe/recife.html. Acesso em: 16 Mar. 2023.

[12] A. F. Castro e A. G. Oliveira, "Estudo sobre os casos de dengue utilizando técnicas de Análise de Dados: Um estudo de caso no estado de Pernambuco, Brasil". In: 16th Iberian Conference on Information Systems and Technologies (CISTI), 2021, pp. 1-6, doi: 10.23919/CISTI52073.2021.9476432.

[13] Portal Eletrônico G1 - PE. Disponível em: https://g1.globo.com/pe/pernambuco/noticia/2021/11/16/recife-confirma-mais-duas-mortes-por-dengue-e-sobe-para-quatro-total-de-obitos-por-arboviroses-em-pe-em-2021.ghtml. Acesso em: 08 set. 2022

Development and its applications in scientific knowledge

Recife, Pernambuco, Northeast Brazil: comparison between cases of dengue, chikungunya and zikavirus in the year 2021, using the Python language