eMOTIONAL Cities

Mapping the cities through the senses of those who make them

METHODOLOGICAL REPORT FOR MAPPING HOTSPOTS IN LISBON

Authors

Paulo Morgado (supervisor)

João Reis (Spatial analysis)

Rita Morais (report drafting)

19/09/2023

ULISBOA UNIVERSIDADE DE LISBOA

IGOT Instituto de Geografia e Ordenamento do Território UNIVERSIDADE DE LISBOA

CEG Centro de Estudos Geográficos

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 945307.

Index

Index of Figures

Index of Tables

1. Introduction

This report aims to provide an overview of the variables, metrics, indexes, and analytical methods used by the Institute of Geography and Spatial Planning (IGOT), regarding the methodological procedures for mapping the hotspots following the guidelines detailed in the **deliverable 4.3** – *'Mapping of cities based on cognitive aspects and emotional responses triggered by the built environment'*, led by the University of Cambridge.

Considering the project's goals, we are examining four distinct types of urban health variables: urban physical environment, health-related variables, socioeconomic-related variables, and perception-related variables. While some of them were obtained straightforwardly, being readily available for use, others required pre-processing to provide the information needed.

In this report, we will first describe the procedures taken for data acquisition and selection of the variables, presenting all the 37 variables considered in this study. Then, we will describe the pre-processing steps undertaken to make the variables spatially coherent and statistically accurate and lastly, we will explore the statistical analysis already performed to our data and present some results and conclusions.

2. Data collection

The identification and selection of the relevant variables for this study considered the literature review (LR) conducted in the WP2, and our own LR regarding the subject of urban health and wellbeing, focus only on evidence-based articles. The primary aim of this task was to identify both individual's and urban environment's aspects that have the greatest impact on health outcomes, in accordance with our theoretical framework (**deliverable 2.2**. *'Conceptual framework'*). Following the identification of the major subjects of urban health-related data, we selected the outlined variables and proceeded with their collection.

Given our interest in encouraging Open Geospatial Consortium (OGC) guidelines regarding data (FAIR; Findable, Accessible, Interoperable, and Reusable), we prioritized the use of open data, namely government organizations and other entities with a public service, data providers which are subject to quality control and standardization of methods for data acquisition and processing, making their data reliable, consistent, and representative. While relying primarily on public data, we also made use of free geodata created and made available by private projects / companies, and crowdsourcing data to ensure we included all the relevant information in our analysis. In **Table 2.1**, we present the list of the public and private sources, which provided information for this study.

Table 2.1. Type of sources and respective level of application of the provided data.

	Level of application
Public Organisms	
Área Metropolitana de Lisboa	Intermunicipal
Câmara Municipal de Lisboa	Municipal
European Space Agency/Copernicus	European
European Environmental Agency / Copernicus	European
European Union / European Space Agency / Copernicus	European
Instituto Geográfico do Exército	National
Serviço Nacional de Saúde	National
Statistics Portugal	National
Turismo de Portugal	National
Private Organisms	
Centro de Estudos e Avaliação em Saúde / Associação Nacional de Farmácias	National
ESRI, Michael Bauer Research GmbH	Global
OpenstreetMap and GeoFabrik	Global
Twitter	Global

Apart from the upper list, the only data produced directly by the authors was the location of fast-food outlets, due to the lack of credible sourced. This involved georeferencing fast-food establishments — '100 Montaditos', 'Burger King', 'Burger Ranch', 'Domino's Pizza', 'McDonald's', 'Pizza Hut', 'Telepizza', 'KFC', 'Taco Bell', 'Subway', 'Pans & Company', and 'Papa John's' brands — via chain's own websites and Google Earth, and then performing field work to validate the correct coordinates and state of activity.

A total of 37 variables were collected, which were considered to better describe dimensions and aspects of urban health data. Eight of these variables — '*patients with obesity'*, *'patients with hypertension'*, *'patients diagnosed with dementia'*, *'patients diagnosed with anxiety disorder'*, *'patients diagnosed with depressive disorder'*, *'drug prescription of anxiolytics'*, *'drug prescription of antidepressants'* and *'drug prescription of antidementia'* — are the dependent variables chosen to reflect the health outcomes subject to investigation. **Figure 2.1** illustrates all the variables, organised by 4 dimensions — **Urban Health Data**, **Physical Environmental Data**, **Socioeconomic Environment Data**, and **Perception Data** — and corresponding aspects, with the respective processing methodology that will be explained in the following points.

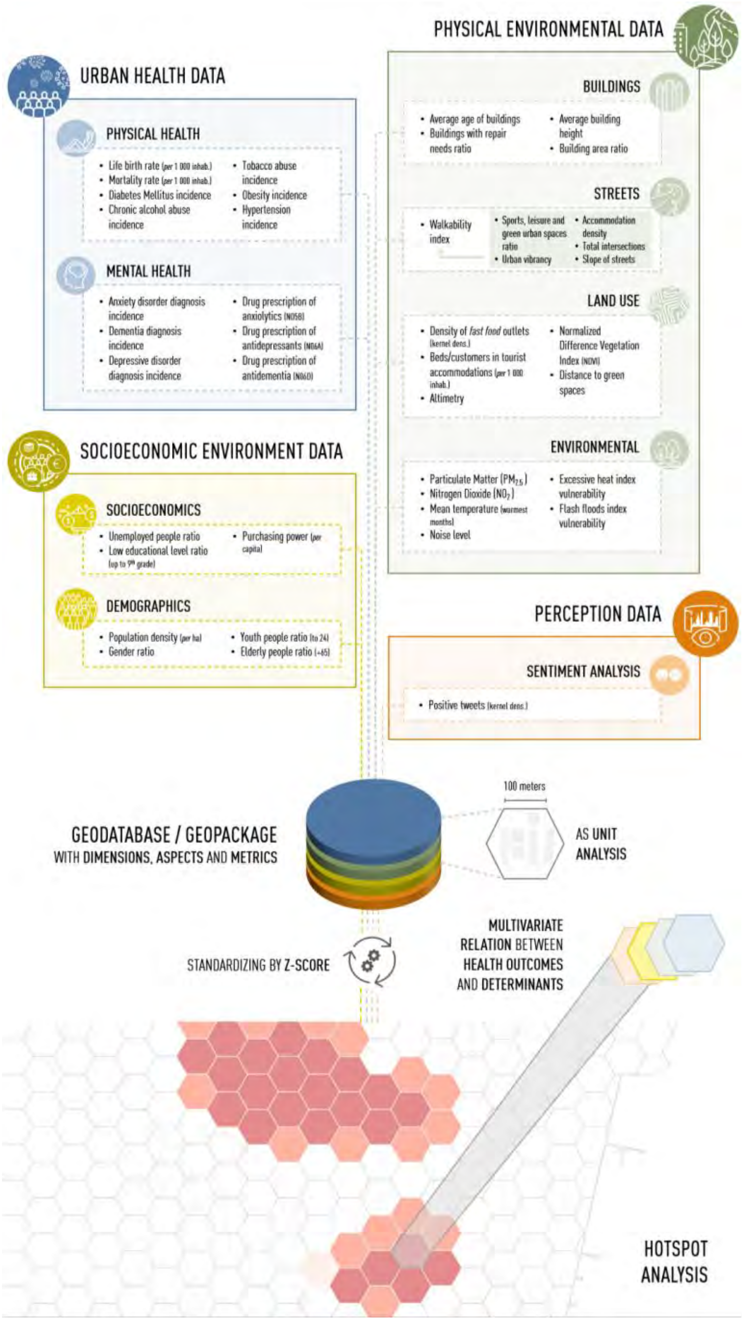


Figure 2.1. List of selected variables by each dimension and adopted general methodology.

For example, *'life birth rate'*, '*mortality rate'* and '*Diabetes Mellitus incidence'* are used to assess physical health aspects, while variables such as '*anxiety disorder diagnosis incidence'* and '*depressive disorder diagnosis incidence'* are used for assessing mental health aspects in the urban context**.** The majority of the variables gathered were related to both **Physical Environmental Data** (16) and **Urban Health Data** (13). The complete list of variables considered for our analysis are available in **Appendix 1**, along with additional information on source, date time and spatial resolution.

3. Data pre-processing

To harmonize the results, and to make all variables spatially comparable, a hexagonal grid was adopted as the unit of analysis, as in the work developed by Cambridge for London (in **deliverable 4.3.** *'Mapping of cities based on cognitive aspects and emotional responses triggered by the built environment'*). Unlike traditional square grids, hexagonal grids maintain the distance between the centroid of each cell and its limits, reducing sampling bias (Birch *et al.*, 2007).

In the case of Lisbon, a hexagonal grid with an edge of 100 meters was adopted (**Figure 3.1**). In London, it was adopted a hexagonal grid where the distance between each centroid in two nearby hexagons was 350 meters; since the study area in London is about 3.5x larger compared to Lisbon, we considered a 3.5x smaller value of edges in Lisbon. Moreover, we had spatial detailed data (*subseção estatística*, which represents a block) that justifies the 100m hexagon.

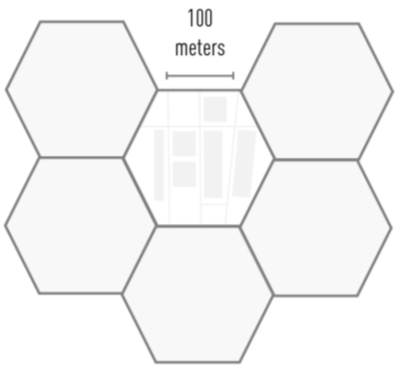


Figure 3.1. Analysis unit for Lisbon, in hexagons.

Additionally, at the boundaries of the study area, and unlike in London, the hexagons were clipped by the inland boundary of Lisbon. According to the *'Official Administrative Map of Portugal'* (*'Carta Administrativa Oficial de Portugal'* in Portuguese), in its 2022 version, the municipality of Lisbon has 100.1 km^2 of area, of which 13.2 km^2 correspond to the Tagus River wetlands. The statistical data of the Census 2021, from Statistics Portugal, considered these areas, which biases the results when it is necessary to divide a value by the respective area of analysis. As such, these areas were excluded from the analysis, and only the inland area of Lisbon was considered.

When applicable, the variables were clipped, by the study area boundaries:

> *Geoprocessing* > *Analysis Tools* > *Clip* (for vectorial features) or *Geoprocessing* > *Spatial Analyst Tools* > *Extract by Mask* (for raster features)

and were converted to 'ETRS 1989 Portugal TM06', a local coordinate system, which is the default coordinate system in Portugal:

> *Geoprocessing* > *Data Management Tools* > *Project* (for vectorial features) or *Geoprocessing* > *Data Management Tools* > *Project Raster* (for raster features)

In this study, none of the variables were originally provided at hexagon level (in the case of *'walkability index'*, although calculated directly by hexagon, the variables integrated into it were also not provided at hexagon level), therefore all variables required spatial processes to obtain the respective value at hexagon level. In the following points, these procedures will be developed by type of unit of analysis and / or by type of information.

3.1 Data in parish / block / class-break level

In this section, we will explore the pre-processing of data collected at parish, block, or class-break level, corresponding about half of the total number of variables (**Table 3.1**). In the case of the variables from **Socioeconomic Environment Data** dimension, since they mostly correspond to variables from the 2021 Census, except for *'tweets'*, all other variables were provided at parish or subsection scale.

Table 3.1. Variables at parish, subsection or class-break level, by dimension.

Dimensions	Metrics	Data Source	Datetime	Scale
Urban Health Data	Life births rate	Statistics Portugal (2022)	2021	Parish
	Mortality rate	Statistics Portugal (2022)	2021	Parish
Physical Environment Data	Average age of buildings	2021 Census, Statistics Portugal (2022)	2021	Block
	Buildings with repair needs ratio	2021 Census, Statistics Portugal (2022)	2021	Block
	Beds / customers in tourist accommodations	*Turismo de Portugal* (2021)	2021	Parish
	Noise level	*Câmara Municipal de Lisboa* (2021)	2020	Class-break
	Vulnerability to excessive heat index	PMAAC-AML (2018)	Actual vulnerability	Parish
	Vulnerability to flash floods index	PMAAC-AML (2018)	Actual vulnerability	Parish
Socioeconomic Environment Data	Purchasing power	Esri, Michael Bauer Research GmbH (2022)	2021	Parish
	Unemployed people ratio	2021 Census, Statistics Portugal (2022)	2021	Block
	People with low literacy level ratio	2021 Census, Statistics Portugal (2022)	2021	Block
	Population density	2021 Census, Statistics Portugal (2022)	2021	Block
	Gender ratio	2021 Census, Statistics Portugal (2022)	2021	Block
	Youth people ratio	2021 Census, Statistics Portugal (2022)	2021	Block
	Elderly people ratio	2021 Census, Statistics Portugal (2022)	2021	Block

In the case of *'noise level'* variable, provided at class-break level, the maximum value of each class was considered as being the corresponding value of that class, in order to be able to make calculations (e.g., for the class between 80 and 85 dB[A] (A-weighted decibel), the value '85' was considered). Additionally, in the hexagons with information gaps, the average of the values of the adjacent hexagons was considered.

The methodological procedure for the conversion of the variable at class-break level, and most of the variables at block level, to hexagon level, is represented in **Figure 3.2**: in ArcGIS Pro, with both features at block or class-break level, and at hexagon level, we initially apply a geoprocessing intersect analysis to both of them:

Geoprocessing > *Analysis Tools*

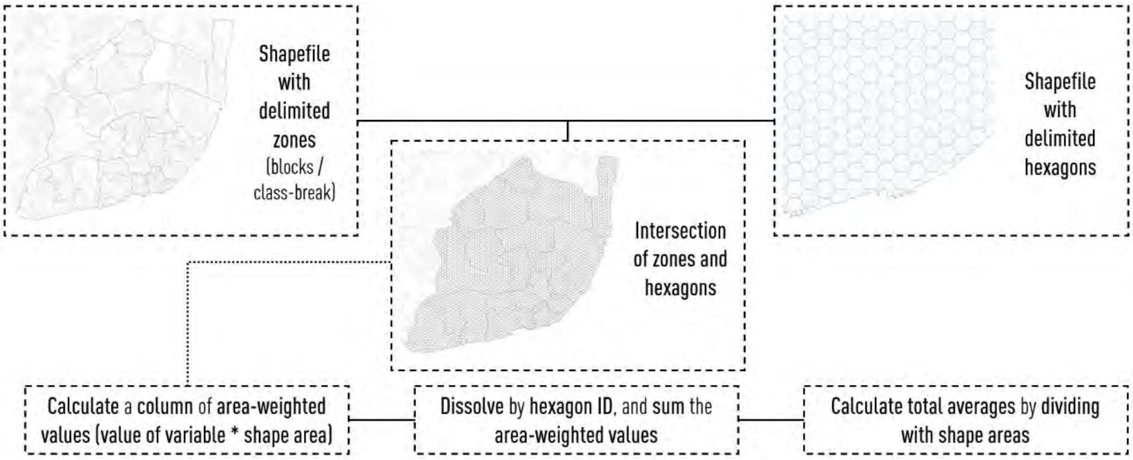


Figure 3.2. Methodological process to obtain subsection / class-break level values in hexagon grid.

Subsequently, it was created a column where the values of the respective variable were weighted:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Field*, where the field with the value of respective variable is multiplied by the shape area.

This feature is then dissolved by the identifier of each hexagon, summing the field with weighted respective variable values by area:

Geoprocessing > *Data Management Tools* > *Dissolve*

Finally, a column is created where the total average *per* hexagon is calculated:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Field*, where the field with the area-weighted values is divided by the shape area.

Regarding the parish-level variables, the method used to convert data at block or class-break level into hexagons results in halos in the hexagons between parish boundaries. Moreover, for the *'population density'* and *'gender ratio'* variables, the adoption of this method, due to the nature of the variable itself and the existence of hexagons with very small dimensions within the boundaries of the study area, results in extremely high values in some of the hexagons. In such cases, to convert the data into hexagons, it was used the *Spatial Join* tool, in:

Geoprocessing > Analysis Tools, where the value in each hexagon will correspond to the value of each parish / subsection that has the highest percentage of area in each hexagon.

The results of *'purchasing power'*, *'people with low literacy level ratio'* and *'noise level'* variables, originally at parish, block, or class-break level, respectively, are represented in **Figure 3.3** to **Figure 3.5**; the other variables are represented in **Appendix 2**.

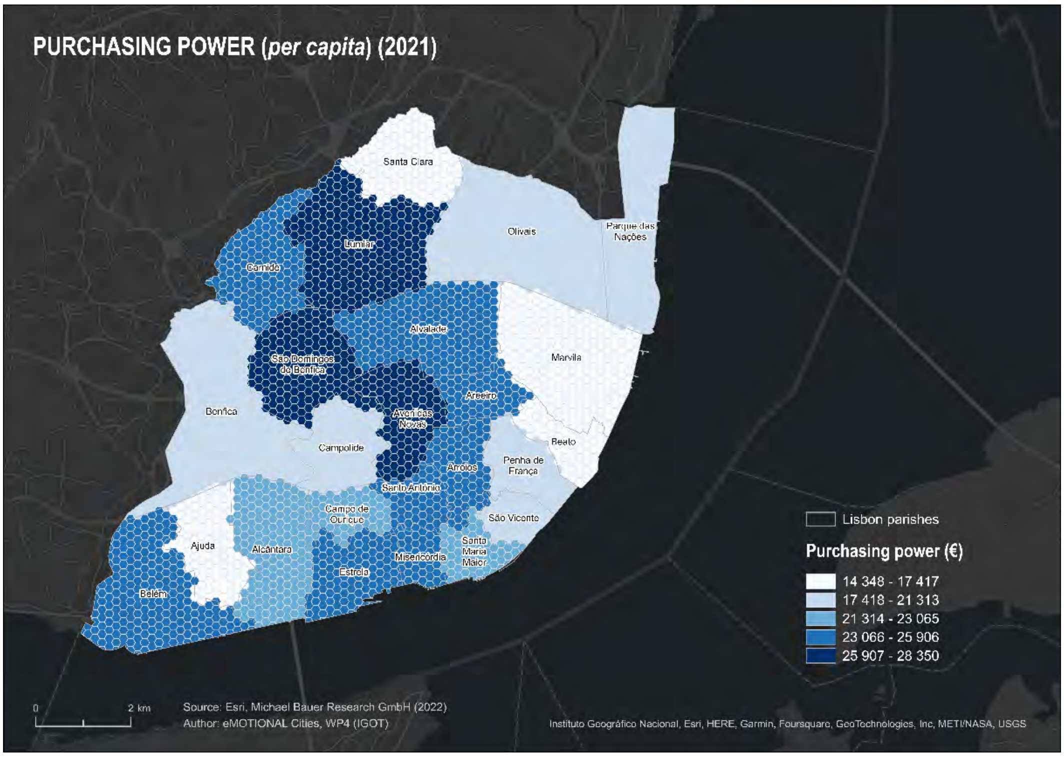


Figure 3.3. Purchasing power *per* capita in Lisbon.

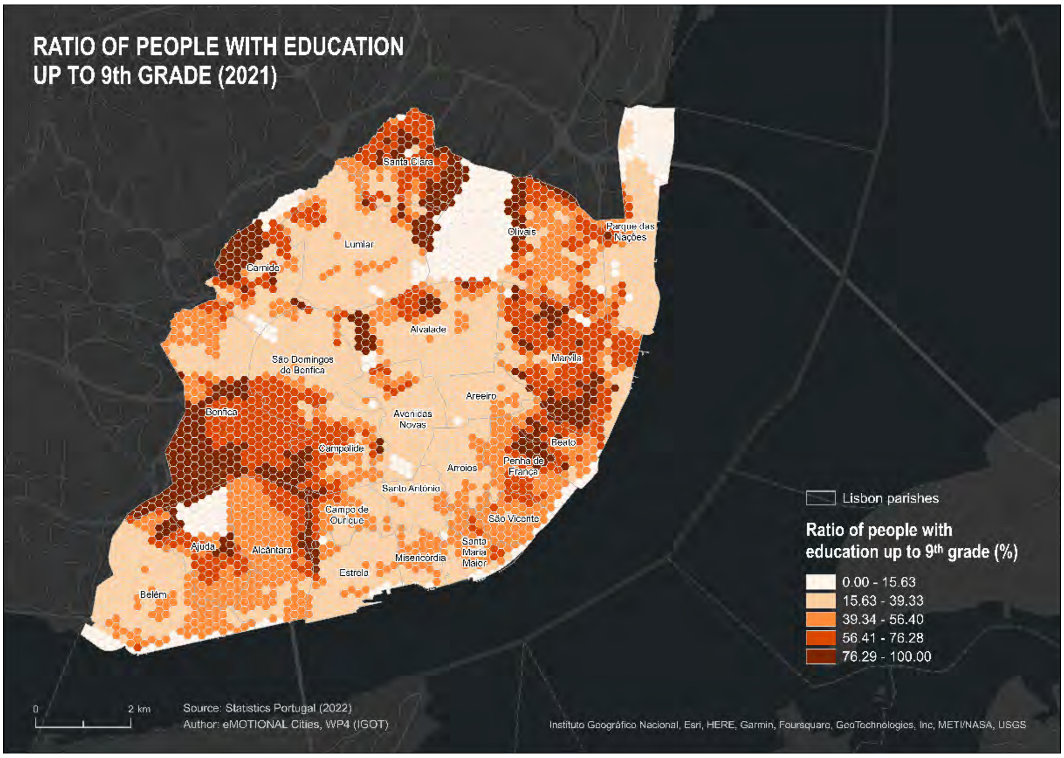


Figure 3.4. Ratio of people with low literacy level in Lisbon.

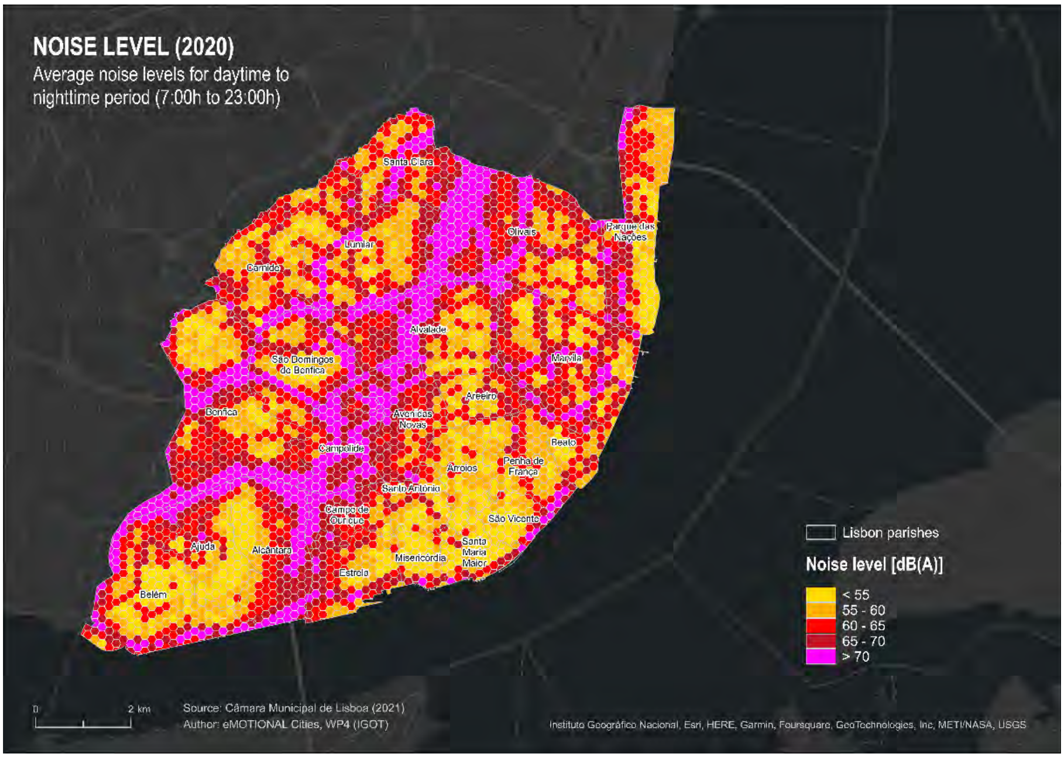


Figure 3.5. Average noise levels for 7h until 23h in Lisbon.

3.2 Data in raster format

Raster data is used to represent geographic information as a grid of cells (pixels), with specific dimension (spatial resolution), each containing information about a location. In this section, we will go over the pre-processing of the data we collected as raster data. Our emphasis will be on the methodologies used to derive certain variables that involve more intricate procedures. From the seven variables considered (six from **Physical Environment Data** dimension and one from **Perception Data** dimension), and with the exception of the *'density of fast-food outlets'* and *'density of positive tweets'* variables (explained in **3.2.1 and 3.2.2**), they were originally provided in raster format (**Table 3.2**).

Table 3.2. Variables in raster format, by dimension.

Dimensions	Metrics	Data Source	Datetime	Resolution
Physical Environment Data	Particulate Matter ($PM_{2.5}$)	Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., ... & Martin, R. V. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, 55(22)	2021	0.01° × 0.01°
	Nitrogen Dioxide (NO_2)	Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., ... & Lamsal, L. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *The Lancet Planetary Health*, 6(1), 49-58	2020	0.0083 ° x 0.0083 °
	Altimetry	*Instituto Geográfico do Exército* (n.d.)	-	25x25 m
	Mean temperature	Copernicus Climate Change Service (2019)	2017	100x100 m
	Normalized Difference Vegetation Index (NDVI)	European Union / ESA / Copernicus (2022)	2021	10x10 m
	Density of fast-food outlets	Elaborated by the authors (2022)	17 to 21 March 2022	10x10 m
Perception data	Density of positive tweets	Twitter (2022)	2018 to 2021	10x10 m

To obtain the same cell size for all variables, a 1x1 meter resample was executed for each raster, using ArcGIS Pro in:

> *Geoprocessing > Data Management Tools > Resample*

The methodological process of converting each final raster to hexagon level was identical for all variables. However, and except for the *'altimetry'* variable, it was necessary to do some procedures to obtain the raster with the final values; these processes will be detailed in the following points. From *'altimetry'* raster file, and the feature containing hexagon delimitation, by using the *Zonal Statistics as Table* tool of ArcGIS Pro, in:

> *Geoprocessing > Image Analyst Tools*

it was obtained the average altitude in each hexagon. Subsequently, the table obtained in the previous step was joined through the *Join Field* tool in ArcGIS Pro:

> *Geoprocessing > Data Management Tools*, using as a common field the identifier of each hexagon

The results of *'altimetry'* are represented in **Figure 3.6**.

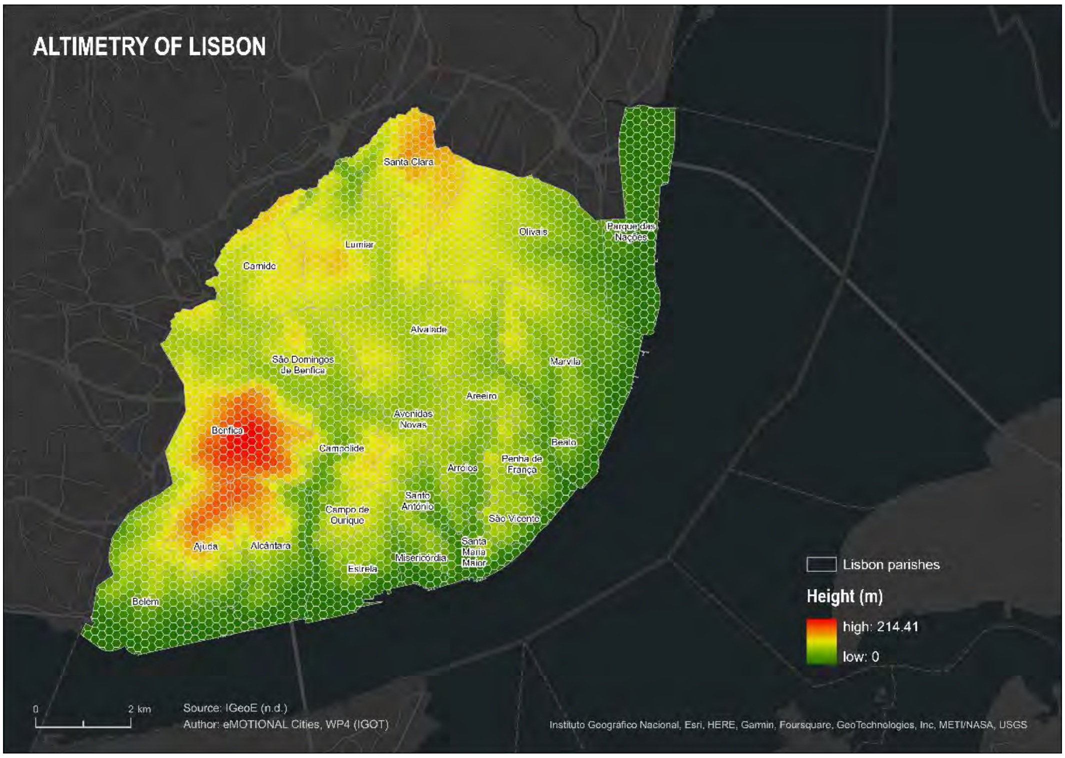


Figure 3.6. Altimetry in Lisbon.

3.2.1 Density of fast-food outlets

Due to the lack of credible sources, this indicator was produced by the authors. The problem identified with the available data was related to the data being obsolete, not considering the relocation or emergence of new ones — as well as the opposite. This involved creating a full and newer inventory of all fast-food establishments in Lisbon, relying on remote sensing (e.g., Google Earth's satellite and street view images) and validating locations and state of activity with fieldwork. The outlets gathered include '100 Montaditos', 'Burger King', 'Burger Ranch', 'Domino's Pizza', 'McDonald's', 'Pizza Hut', 'Telepizza', 'KFC', 'Taco Bell', 'Subway', 'Pans & Company', and 'Papa John's'. The inventory resulted in a layer of points corresponding to each fast-food establishment's locations on the city.

In order to generate a continuous surface representing the density of fast-food outlets (points) — and to, subsequently, convert to hexagon level —, we applied a kernel function estimator to our point layer, in ArcGIS Pro, in:

Geoprocessing > Spatial Analyst Tools > Kernel Density

using the default option for searching radius, which computes a spatial variant of 'Silverman's Rule of Thumb' (Silverman, 1986). This procedure is considered to be robust enough for outliers in the sample (ESRI, https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-kernel-density-works.htm). Subsequently, it was adopted the methodology developed in **3.2**, to obtain the *'density of fast-food* *outlets'* variable in hexagon grid. The results of this variable are represented in **Figure 3.7**, by high-low density.

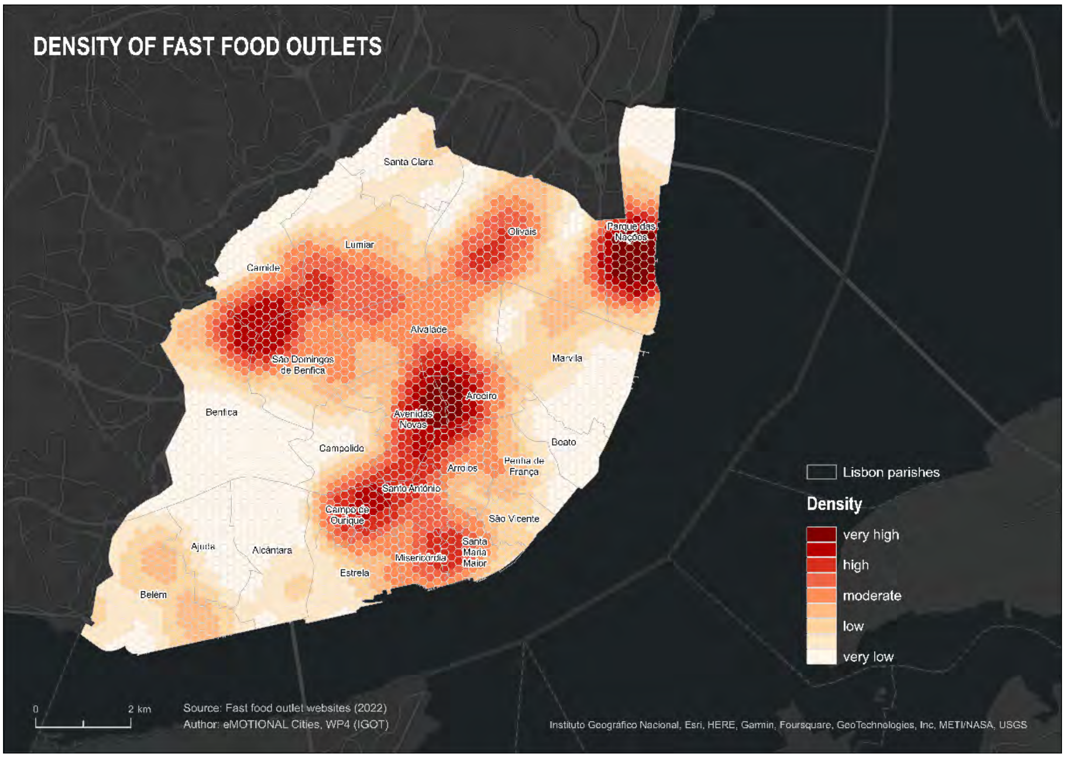


Figure 3.7. Density of fast-food outlets in Lisbon.

3.2.2 Sentiment analysis — Density of positive tweets

The entire process for collecting and processing data to measure and map emotions with social media data — specifically through tweet analysis — has already been described in previously published reports (in **deliverable 4.3.** *'Mapping of cities based on cognitive aspects and emotional responses triggered by the built environment'*) for London; in Lisbon, the adopted methodology is the same. More details about the methodology can be found in that report.

To obtain the density of positive tweets *per* hexagon (**Figure 3.8**) in ArcGIS Pro, it was applied the same procedures described in the previous point (**3.2.1**).

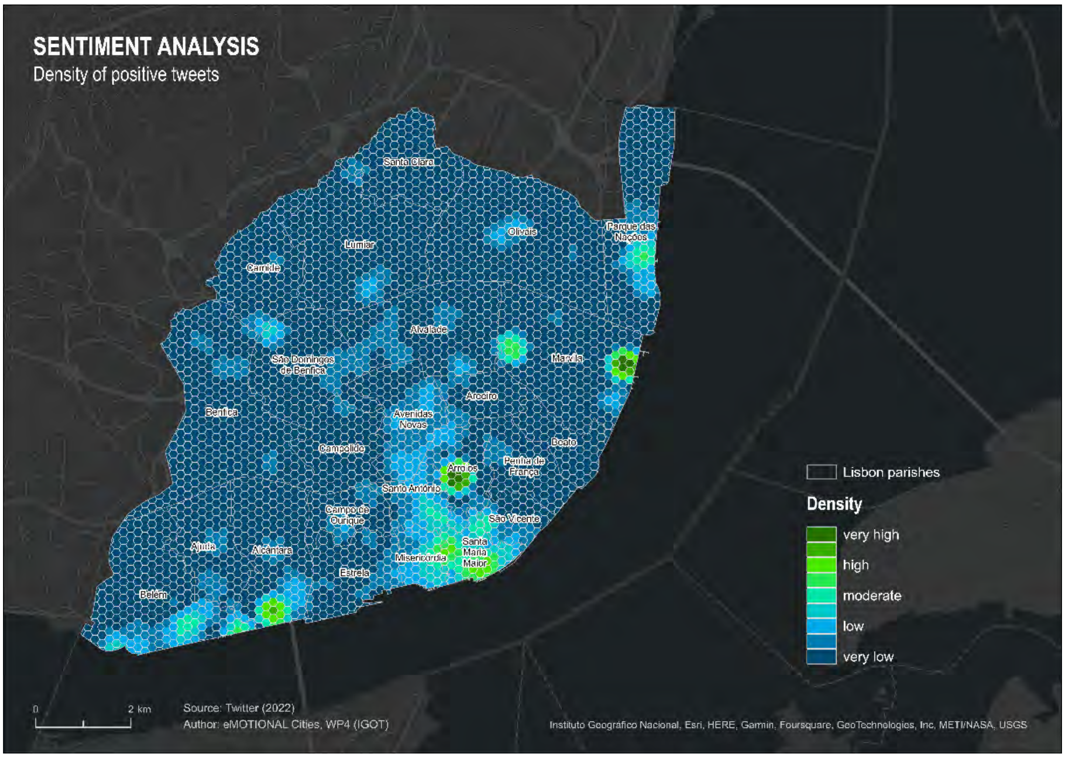


Figure 3.8. Density of positive tweets in Lisbon.

3.2.3 Normalized Difference Vegetation Index (NDVI)

The NDVI (or Normalized Difference Vegetation Index) is an indicator that is widely used in agriculture, forestry, and land management, as it provides information on vegetation health and productivity. It is obtained using the formula:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - RED}{\mathrm{NIR} + \mathrm{RED}}$$

where 'NIR' corresponds to the spectral reflectance measured in Near-Infrared waveband (reflected by plant leaves) and 'RED' corresponds to the spectral reflectance measured in Red waveband (absorbed by plant leaves) (Pettorelli *et al.*, 2005).

The output NDVI values range from -1 to 1, with higher positive values indicating a higher density of green vegetation (green vegetation is represented when NDVI is greater or equal to 0.1), values close to zero indicating low vegetation cover, and negative values indicating water or snow cover (USGS, https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology).

There are several ways to calculate this indicator, using different Geographic Information System (GIS) tools. In our case, it was calculated via Google Earth Engine, which provides access to a wide range of satellite imagery, such as Landsat, Sentinel, and MODIS. The calculation of NDVI was obtained directly using code in JavaScript, as showed in **Figure 3.9**.

```
function maskS2clouds(image) {
  var qa = image.select('QA60');

  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
      .and(qa.bitwiseAnd(cirrusBitMask).eq(0));

  return image.updateMask(mask).divide(10000);
}

var ae = ee.Geometry.Rectangle({
coords: [[-9.573819,39.085103], [-8.431422,38.380030]],
geodesic: false
});

var dataset = ee.ImageCollection('COPERNICUS/S2_SR')
                  .filterDate('2021-01-01', '2021-12-31')
                  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE',1))
                  .map(maskS2clouds);

var addNDVI = function(image) {
return image.addBands(image.normalizedDifference(['B8', 'B4']));
};

var dataset = dataset.map(addNDVI);

var NDVI = dataset.select(['nd']);
var NDVImean = NDVI.mean();

var ndvi_pal = ['FFFFFF', 'CE7E45', 'DF923D', 'F1B555', 'FCD163', '99B718', '74A901', '66A000', '529400',
'3E8601', '207401', '056201', '004C00', '023B01', '012E01', '011D01', '011301'];

Map.addLayer(NDVImean.clip(ae), {palette: ndvi_pal}, 'NDVI');

Export.image.toDrive({
 image: NDVImean.clip(ae),
 description: 'NDVI_2021',
 folder: 'GEE',
 fileNamePrefix: 'NDVI',
 region: ae,
 fileFormat: 'GEOTIFF',
 scale: 10
});
```

Figure 3.9. Code introduced in Google Earth Engine to obtain NDVI raster.

In the variable *'ae'*, the bounding box in which the NDVI will be obtained is defined (the coordinates are in the 'WGS 1984' system). In case of the variable *'dataset'*, the satellite used to obtain the images is specified (it was considered the Sentinel-2 satellite due to the higher spatial resolution — 10x10 meters), as well as the temporal space for analysis). Lastly, in the variable *'NDVImean'*, it is requested to calculate the average of the values of all images obtained over the defined time period, obtaining — in this case — an annual average.

After running the code above — which applied directly a 'cloud mask' in all obtained images —, the average NDVI in 2021 was exported in GeoTIFF format, then imported into ArcGIS Pro, where it was adopted the methodology developed in **3.2**, in order to obtain the *'NDVI'* variable in hexagon grid (**Figure 3.10**).

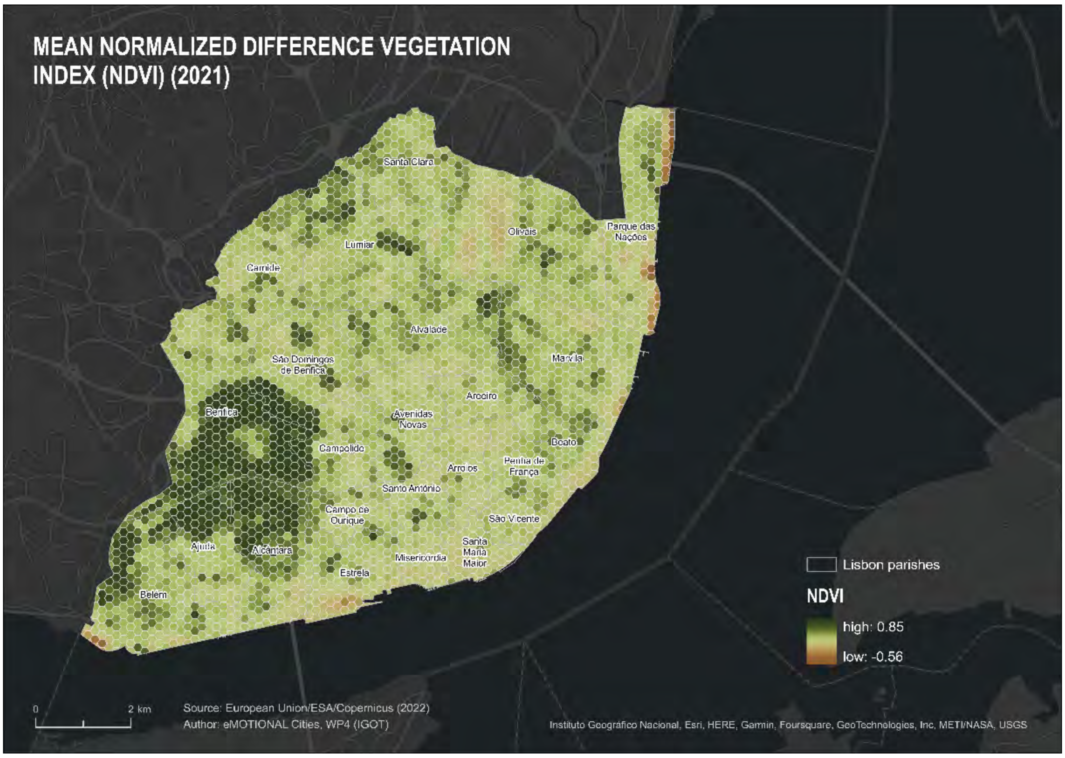


Figure 3.10. Mean NDVI in Lisbon.

3.2.4 Particulate Matter ($PM_{2.5}$) and Nitrogen Dioxide (NO_2)

Particulate Matter (PM) is a complex collection of constituents with varying chemical and physical properties. Aerodynamic diameter is used as an indicator of particle size to classify particles and determine their transport and removal processes in the air, deposition sites, and clearance pathways within the respiratory tract [World Health Organization (WHO), 2021]. While new research findings highlight the dynamic and complex nature of PM, understanding its concentration remains critical for assessing individuals' exposure to air pollutants, particularly in urban environments, where concentrations are frequently higher than in rural areas, due to human activities. The focus in recent decades has been on particles having aerodynamic dimensions of less than or equal to 2.5 micrometre (µm) ($PM_{2.5}$) or 10 µm (PM_{10}).

As $PM_{2.5}$ and PM_{10}, nitrogen dioxide (NO_2) is also one of the main air pollutants with harmful effects on human health. Its chemical properties mean that this pollutant plays a harmful role in climate change and when exposed to solar radiation, it triggers photochemical reactions that generate organic particles, nitrate, and sulphate, which are measured as $PM_{2.5}$ or PM_{10} (WHO, 2021).

Respecting the theoretical framework, we incorporated two variables measuring $PM_{2.5}$ and NO_2. These variables were obtained from tw0.o different sources: the $PM_{2.5}$ was obtained from a predictive model conducted by van Donkelaar et al. (2021) for 2021, with an R^2 of 0.68 for Europe; the NO_2 was obtained from a predictive model conducted by Anenberg et al. (2022) for 2020, with an R^2 of 0.52 for Europe.

Both variables are in NetCDF format, which ArcGIS Pro can read and convert to raster; for that, it is used the *Make NetCDF Raster Layer* tool, in:

> *Geoprocessing* > *Multidimension Tools*

Despite the high geographical granularity of the original information, given that it is a global database, it is of must importance to have geodata with even higher granularity for both variables. To do so, we applied geostatistical interpolation methods, which uses point files, i.e., the raster files are converted to point files by using the *Raster to Point* tool, in:

> *Geoprocessing* > *Conversion Tools*

For both variables, and to generate a continuous surface representing pollutant concentration in Lisbon with higher granularity, it was adopted the Empirical Bayesian Kriging (EBK) geostatistical interpolation method, using the *Geostatistical Wizard* tool. According to Mejía et al. (2023) and Morillo et al. (2022) studies, which compared different spatial interpolation methods to obtain a continuous surface representing NO_2 and PM_{10} concentrations in Guayaquil (Ecuador) and Madrid (Spain), respectively, the method with better results was the EBK; Banerjee et al. (2018) also concluded that EBK is a good method for interpolating PM.

After obtaining the continuous surface with $PM_{2.5}$ and NO_2 concentrations in Lisbon, with a R^2 of 0.853 and 0.995 respectively, the results, in vector format, were converted to raster using the *GA Layer to Rasters* tool, in:

> *Geoprocessing* > *Geostatistical Analyst Tools*, where the "surface output" with prediction values is selected.

Afterwards, it was adopted the methodology developed in **3.2**, in order to obtain the *'$PM_{2.5}$'* and *'NO_2'* variables (**Figure 3.11** and **Figure 3.12** respectively) in hexagon grid.

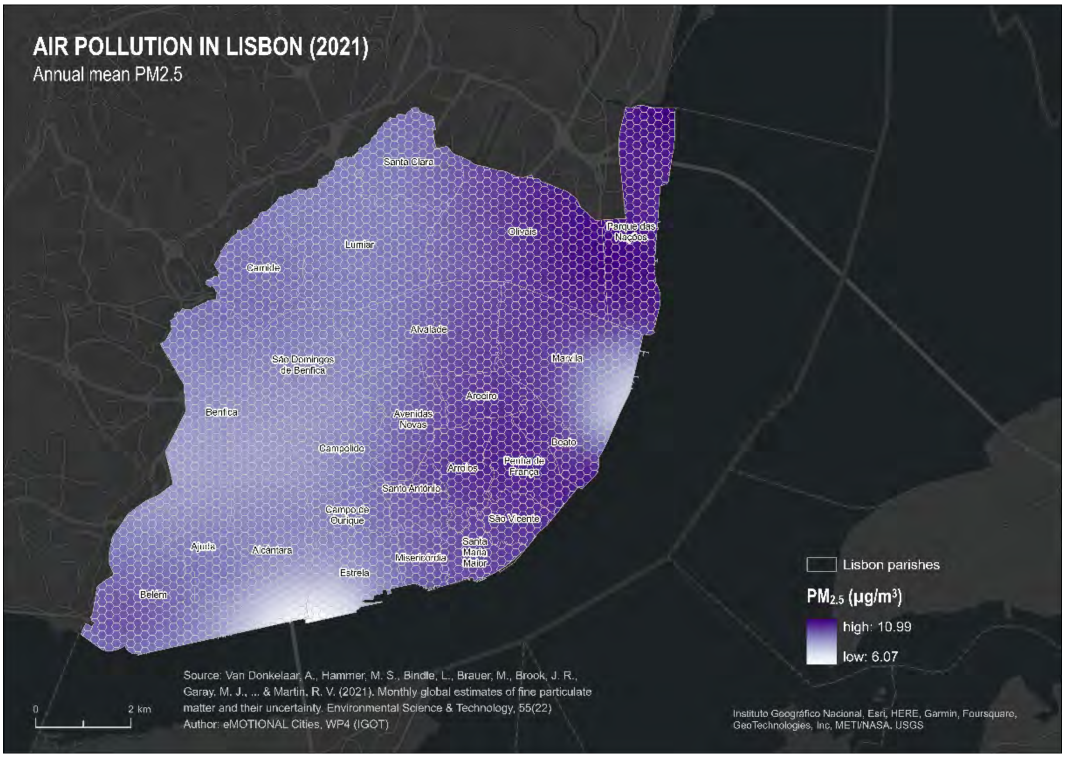


Figure 3.11. Mean $PM_{2.5}$ in Lisbon.

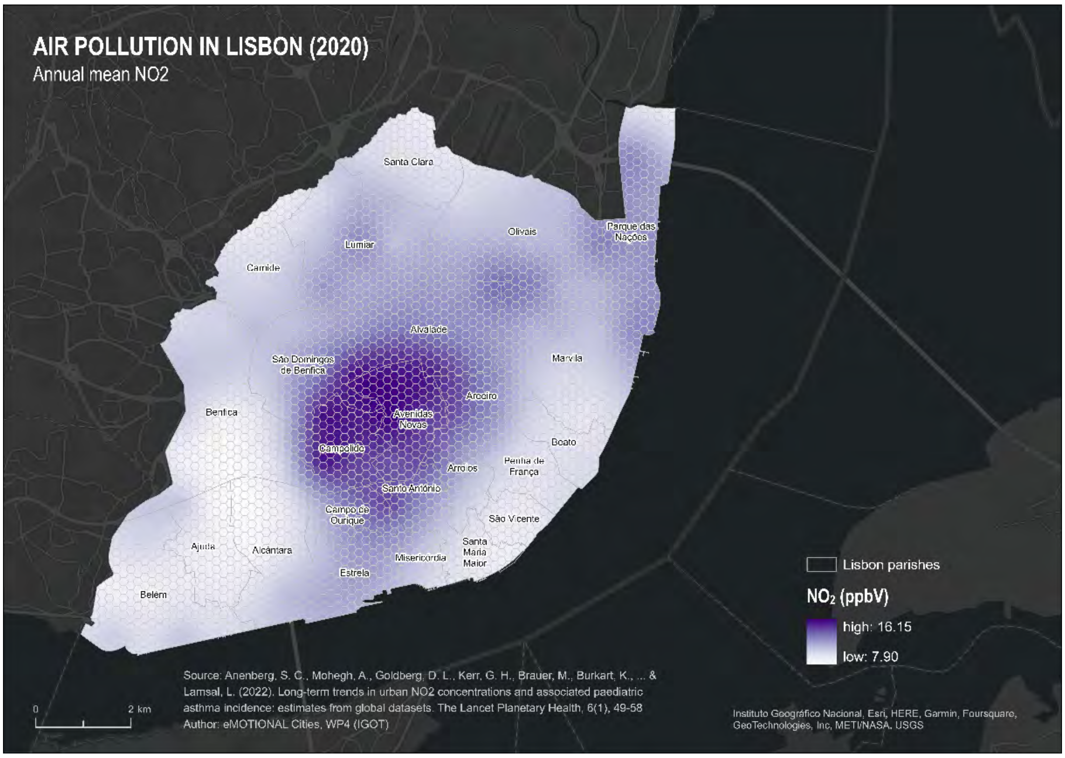


Figure 3.12. Mean NO_2 in Lisbon.

3.2.5 Mean temperature

To obtain the *'mean temperature'* indicator, we used the *'UrbClim'* climate model data for the year 2017. Developed by VITO and Copernicus Climate Change Service, this model generates hourly data for climate parameters — including atmospheric temperature — and releases them in NetCDF format monthly.

To convert the files to raster format, it is used the *Make NetCDF Raster Layer* tool, in:

> *Geoprocessing* > *Multidimension Tools*

where a raster layer is obtained, by month, and each layer contains hourly bands of the downloaded climate parameter; for a month with 31 days, the layer contains in total 744 bands. Then, to get the average monthly temperature, the *Cell Statistics* tool is used, in:

> *Geoprocessing* > *Image Analyst Tools*, where it is selected each raster layer, averaging all bands

According to the obtained results, June and August were the two warmest months of that year. We omitted the rest of the months, and determined the mean temperature based on those two, using the same tool of previous step. The result was, then, converted from degree Kelvin to degree Celsius, using the *Raster Calculator* tool, in:

> *Geoprocessing* > *Image Analyst Tools*, where it was applied the formula:

$$\text{feature in degree Celsius} = '\text{feature in degree Kelvin}' - 273.15$$

To conclude, it was adopted the methodology developed in **3.2**, in order to obtain the *'mean temperature'* variable in hexagon grid (**Figure 3.13**).

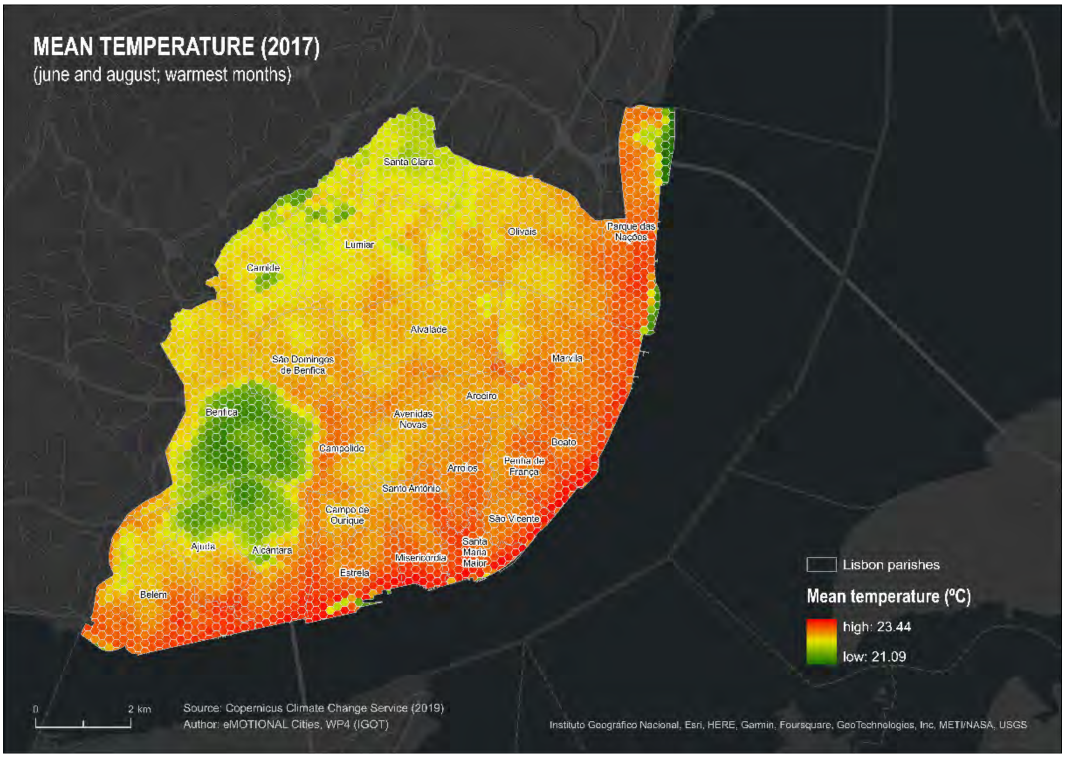


Figure 3.13. Mean air temperature in Lisbon.

3.3 Other data

3.3.1 Physical and mental health

The variables related to physical and mental health come mostly from *Sistema de Informação e Monitorização do Serviço Nacional de Saúde* (Portuguese National Health Service's Monitoring and Information System). Unlike the variables discussed above, these are provided by percentage of incidence at health unit level, which, by turn, is at the level of one or more parishes. To overcome this limitation and obtain the absolute values *per* parish, we performed a spatial transformation process to aggregate the indicator's values by the desirable spatial units. Using the total number of people registered in each health centre and the health centre's catchment area, we recalculated the indicator at parish level and, subsequently, at block level, as demonstrated in

Figure *3.15*. This harmonization process ensured that all the variables accurately reflected the health outcomes of each parish and each block.

The methodological process in ArcGIS is very similar to the one explained and performed in **3.1.**, except for the last step. At the end, the different physical and mental health variables, such as *'patients with hypertension'* (**Figure 3.14**) or *'patients diagnosed with depressive disorder'* (**Figure 3.16**) were obtained at hexagon grid level; the other variables are represented in **Appendix 2**.

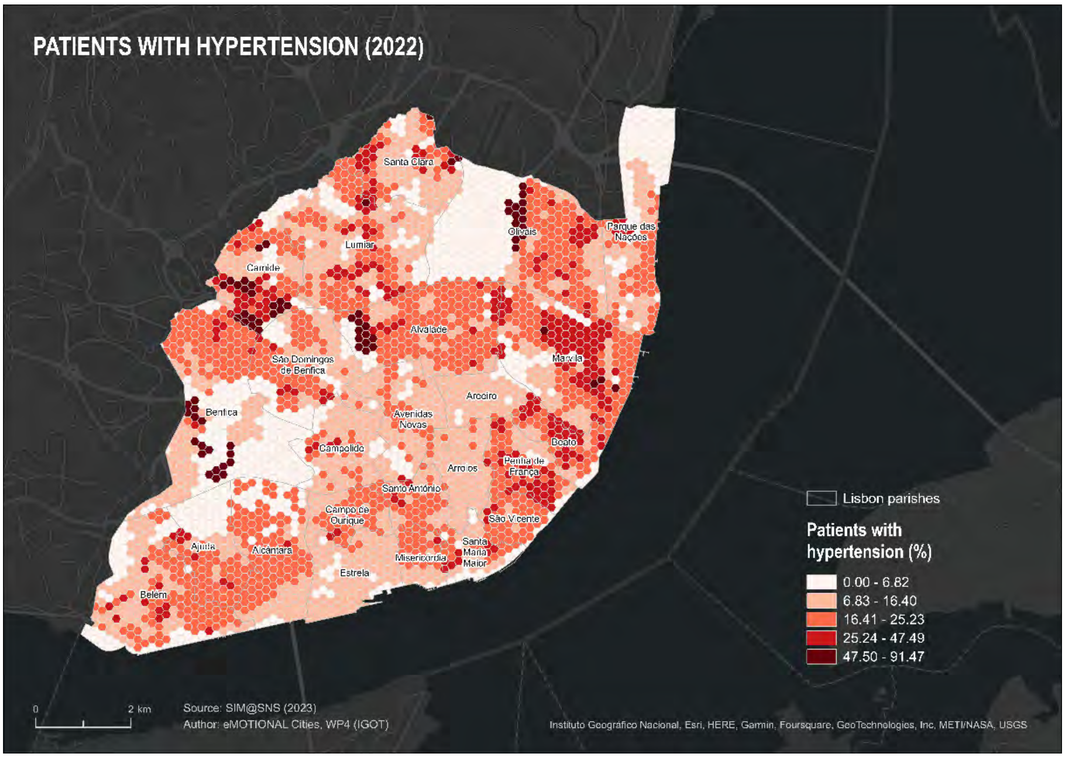


Figure 3.14. Patients with hypertension in Lisbon.

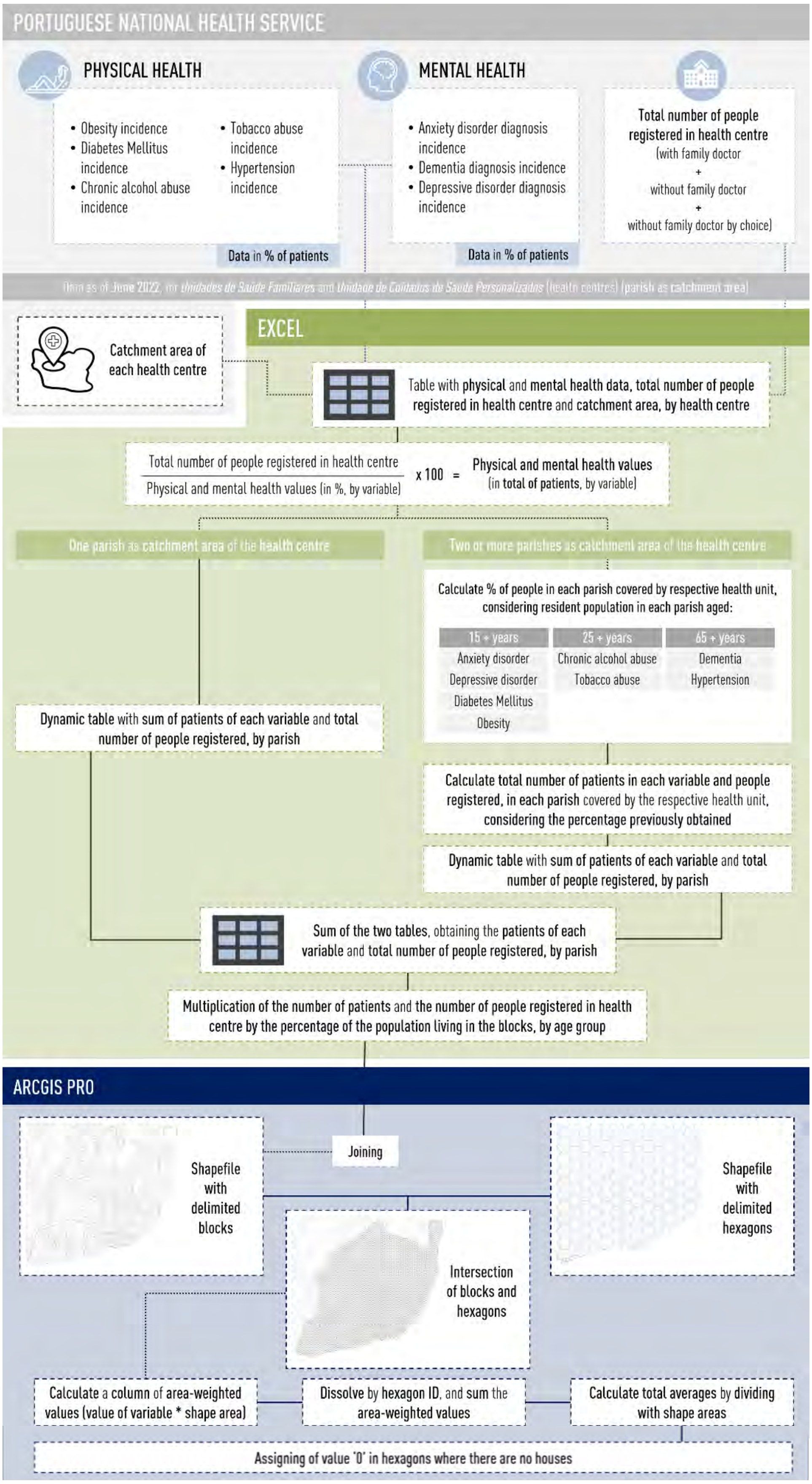


Figure 3.15. Methodological process to obtain physical and mental health variables from SIM@SNS to hexagon grid.

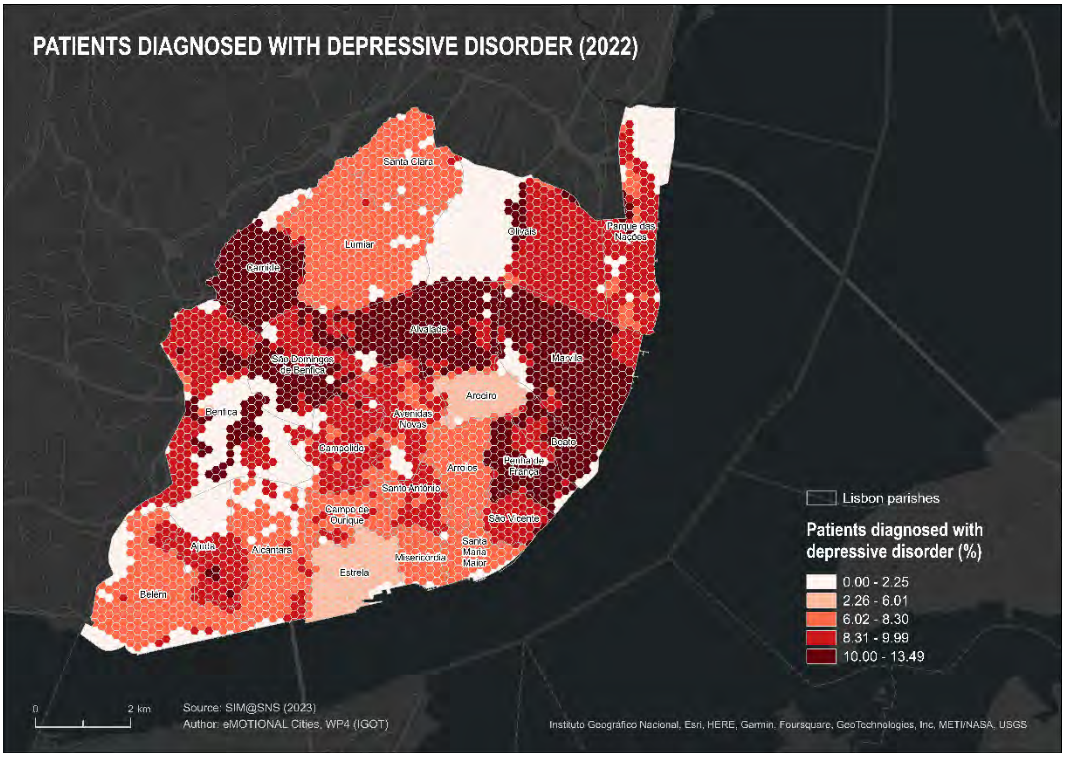


Figure 3.16. Patients diagnosed with depressive disorder in Lisbon.

3.3.2 Drug prescription

The data regarding prescribed drugs at pharmacy level gives an idea of the areas where there may be a greater prevalence of certain diseases that are being controlled by them. *Centro de Estudos e Avaliação em Saúde* (CEFAR) and *Associação Nacional de Farmácias* (ANF) provided, stratified by year (2018 to 2021), age group and gender, standardized values of prescribed drugs per subject at pharmacy level; in Lisbon, data for 189 pharmacies was provided.

In table format, the XY coordinates of each pharmacy were converted into a point feature in ArcGIS Pro, using the *XY Table to Point* tool in:

Geoprocessing > Data Management Tools

The catchment area of each pharmacy corresponds to a Voronoi polygon; each polygon contains only one pharmacy, and any location in it is closer to that pharmacy than to another (ESRI, https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/create-thiessen-polygons.htm). In order to obtain the catchment areas, the *Create Thiessen Polygons* tool was used, in:

Geoprocessing > Analysis Tools

Afterwards, the methodological process is similar to the one performed in **3.3.1**, in ArcGIS Pro section; however, in the last step, the value assigned to the hexagons where there are no houses corresponds to '-1.5', which is the lowest of the three variables; this value is justified by the fact that the data was originally provided with the values already standardized, with positive and negative values, so the value "0" cannot be assigned. The results of ‘drug prescription of antidepressants’ is represented in **Figure 3.17**; the other two variables are represented in **Appendix 2**.

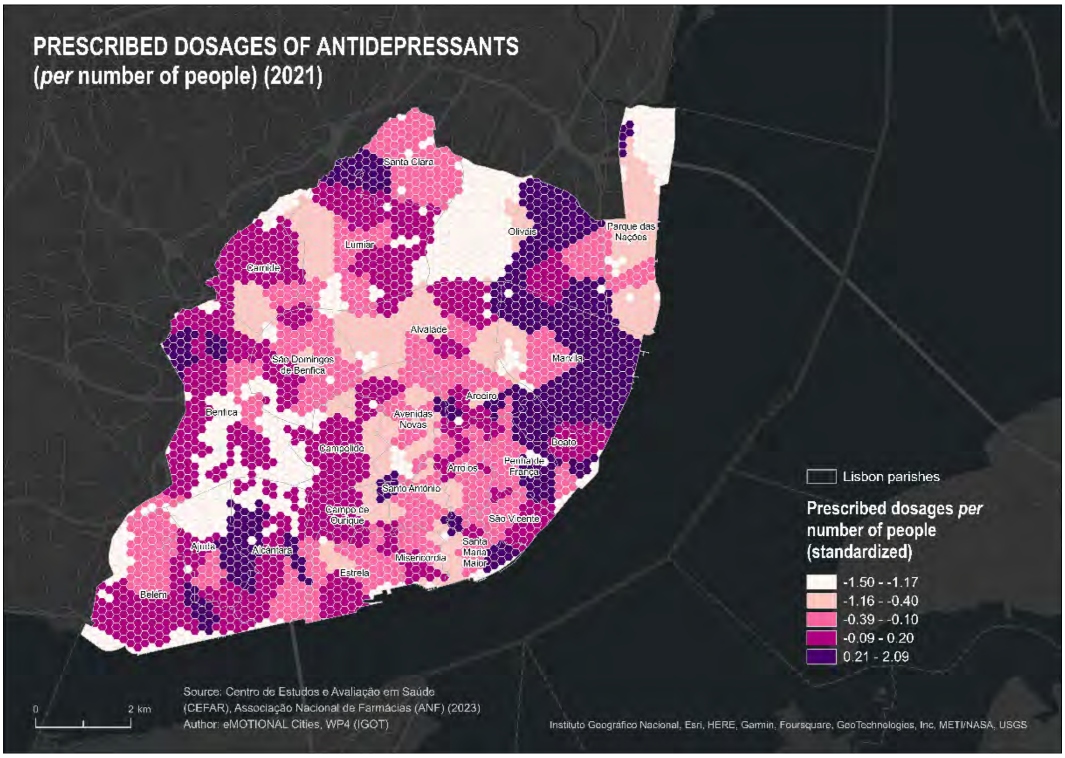


Figure 3.17. Prescribed dosages of antidepressants in Lisbon.

3.3.3 Buildings (average height and area ratio)

The data regarding the building inventory in 2017 — and respective height —, in Lisbon, was provided by the Lisbon City Council. However, since they were provided at building level, some pre-processing was required to convert the data to hexagon level.

Through ArcGIS Pro, to obtain the *‘average building height’* variable *per* hexagon (**Figure 3.18**), the *Summarize Within* tool was used, in:

> *Geoprocessing* > *Analysis Tools*, choosing the “mean” parameter as the desired output

To obtain the building area *per* hexagon (**Figure 3.19**), it is used the same tool of previous step, but without choosing the ‘mean’ parameter as the desired output. With a column with total area occupied by buildings in each hexagon, to calculate the *‘building area ratio’*, it was created, in generated feature from previous step, a column containing the area of each hexagon:

> *Geoprocessing* > *Data Management Tools* > *Add Field*, in which the field type is “Double”; then *Geoprocessing* > *Data Management Tools* > *Calculate Geometry Attributes*, in which the previously created field is selected, calculating the area in hectares

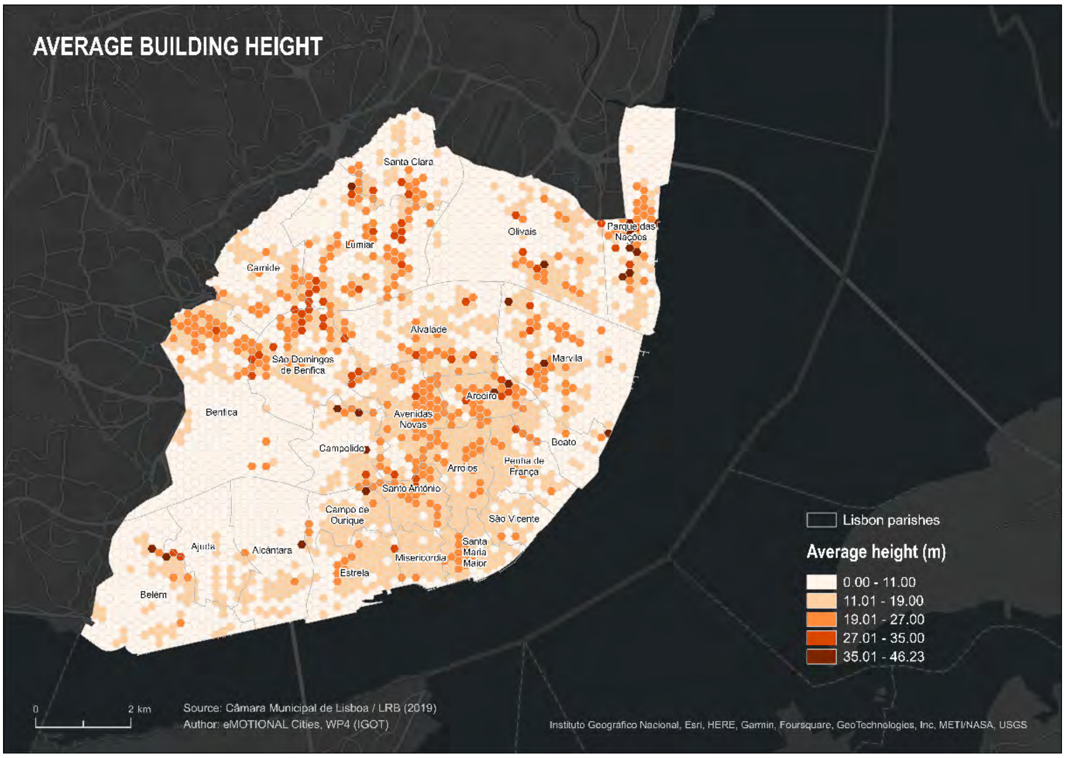


Figure 3.18. Average building height in Lisbon.

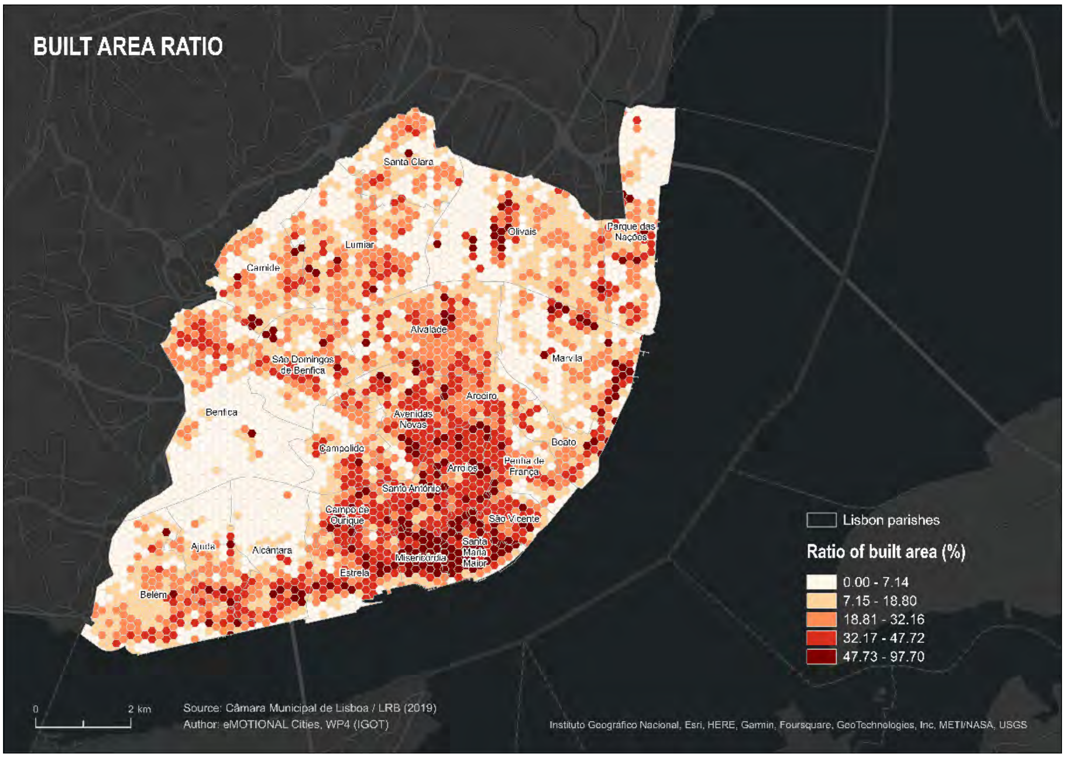


Figure 3.19. Ratio of built area in Lisbon.

and, finally, it was created a column where the percentage of buildings is calculated:

Geoprocessing > Data Management Tools > Add Field, in which the field type is “Double”; then *Geoprocessing > Data Management Tools > Calculate Field*, where the field with the total area of buildings is divided by the area obtained in the previous step, multiplied then by 100

3.3.4 Walkability index

The '*walkability index'* is calculated by using the following formula:

$$\frac{(2 * TD) + (-2 * MSS) + UV + \ IRS + AD}{7}$$

TD – total of intersections
MSS – mean street slope
UV – urban vibrancy
IRS – intensity of recreational spaces
AD – accommodation density

that was adapted from Pereira's (2017) formula:

$$\frac{(2 * ID) + (2 * MS) + LUM + AD}{6}$$

ID – intersection density
MS – mean slope
LUM – land use mix

Lisbon has a high continuous and discontinuous dense urban fabric area; in Pereira's (2017) case study area, this does not apply. Therefore, to better represent the diversity of land uses in Lisbon, the '*land use mix'* indicator was replaced by the indicators of *'urban vibrancy'* (diversity of Points of Interest – POIs) and *'intensity of recreational spaces'* (sport, leisure, and urban green spaces). The '*mean slope'* indicator was also replaced by the '*mean street slope'* because not every space is walkable; considering only the walkable roads, it is possible to obtain more realistic values. Moreover, the indicator *'intersection density'* was replaced by the indicator *'total of intersections'* due to the existence of hexagons with small dimensions in the study area boundaries, resulting in extremely high values in some hexagons, which would result in a significant bias.

The variables used in the index are presented in **Table 3.3**, and their treatment will be detailed in the following points. The methodological scheme is represented in **Figure 3.20**.

Table 3.3. Variables used in walkability index.

Variables	Data source	Datetime	Original coordinate system	Resolution
Total of intersections	NAVTEQ / ESRI (2016)	2016	WGS 1984	-
Mean street slope	*Instituto Geográfico do Exército* (n.d.)	-	Lisboa Hayford Gauss IGeoE	25x25 m
Urban vibrancy	GeoFabrik (2023)	29 March 2023	WGS 1984	-
Intensity of recreational spaces	European Environmental Agency / Copernicus (2020)	2018	ETRS 1989 LAEA	-
Accommodation density	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-

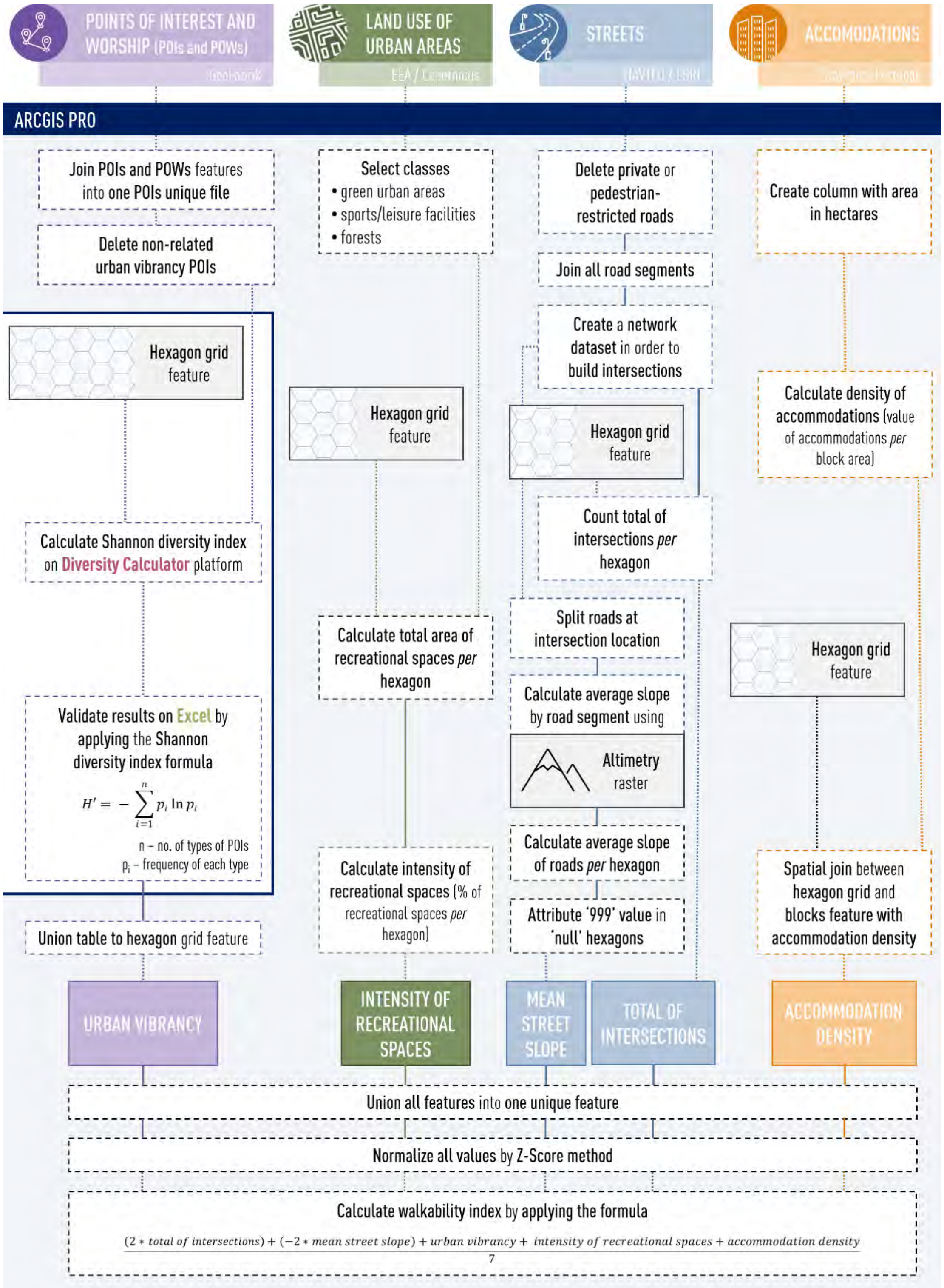


Figure 3.20. Methodological process to obtain walkability index.

Total of intersections

The *'total of intersections'* evaluates the connectivity between spaces. It was calculated based on the NAVTEQ / ESRI street network, with the street inventory for mainland Portugal. We selected and removed from the database the private roads, and pedestrian-restricted roads, in ArcGIS Pro:

Select by Attributes > SQL Expression where *PRIVATE = 'Y'* ***OR*** *AR_PEDEST = 'N'*

keeping only the roads where people can move freely by foot.

Due to the presence of errors in road topography (e.g., presence of intersections in places where there is no intersection of roads, due to the presence of non-joined road segments), it was necessary to conduct a database treatment. First, all segments were joined in ArcGIS Pro:

Geoprocessing > Data Management Tools > Dissolve

and then a Network Dataset was created containing the feature with all segments joined:

Catalog > New Feature Dataset; within this dataset *Catalog > Import Feature Class > 'feature with segments joined'*; within the same dataset *Catalog > New Network Dataset > 'feature imported in the previous step'*

Lastly, to get the correct intersections of all roads, from the Network Dataset created earlier:

Catalog > Build

With the point feature containing all intersections, it is possible to calculate the *'density of intersections'* in each analysis unit. For this, using ArcGIS Pro, it is used the *Spatial Join* tool, in:

Geoprocessing > Analysis Tools, in which the "target feature" corresponds to the feature with the delimitation of the hexagons, and the "join feature" corresponds to the intersections.

obtaining a column with the count of intersections *per* analysis unit.

Mean street slope

To obtain the *'average slope of roads'*, we used the previously obtained road network (with the road segments joined), the intersection network, and the altimetry; the latter is obtained through the contour lines network at a scale of 1:25 000, produced by the *Instituto Geográfico do Exército* (Portuguese Army Geospatial Information Centre) from which was obtained a raster file with a resolution of 25x25 meters.

The intersection network will be used to divide the road network into small segments where the network intersects; this allows us to obtain the average slope in each segment, making the results more detailed. To do this operation, we use the *Split Line at Point* tool from ArcGIS Pro:

Geoprocessing > Data Management Tools

Subsequently, the average slope *per* street segment is obtained through the *Add Surface Information* tool:

Geoprocessing > 3D Analyst Tools, in which the output corresponds to the average slope.

Lastly, to obtain the *'average slope of the roads'* within each hexagon, we use the *Summarize Within* tool:

> *Geoprocessing > Analysis Tools*, in which we obtain the average of the values in the field resulting from the previous step, with the average slope in each segment

With this step, some hexagons will have a "null" value (i.e., hexagons where there are no roads), subsequently influencing the calculations of walkability index. Because there are no roads marked, it is assumed that the walkability will be null; therefore, in these cases, the value of "999" is assigned in ArcGIS Pro. The "null" hexagons are selected through:

> *Select by Attributes* > SQL Expression where *'mean slope column' = NULL*, and through the *Calculate Field* tool (in *Geoprocessing > Data Management Tools*), the value "999" is assigned

Urban vibrancy

The *'urban vibrancy'* indicator consists of the POIs, available in the OpenStreetMap (OSM), which were obtained through GeoFabrik[1]. The methodology of Botta & Gutiérrez-Roig (2021) was adopted, where both POIs and Points of Worship (POWs) were used to calculate the Shannon diversity index, which corresponds to an entropy index. The index is calculated by the following formula:

$$H' = -\sum_{i=1}^{n} p_i \ln p_i$$

where 'n' corresponds to the total number of types of POIs and POWs, and 'p_i' the frequency of each type (Botta & Gutiérrez-Roig, 2021).

The data provided by GeoFabrik, at POIs and POWs level, are available either in points or in polygons, having, in both cases, POIs and POWs different from each other. So, and since it is necessary to obtain only one point feature with all the POIs and POWs (in future steps, we will call only 'POIs'), we initially converted the features from polygons to points, through the *Feature to Point* tool from ArcGIS Pro:

> *Geoprocessing > Data Management Tools*

Subsequently, the four-point features (2 pre-existing and 2 obtained in the previous step) were joined using the *Union* tool:

> *Geoprocessing > Analysis Tools*

To calculate the *'urban vibrancy'* indicator, it is important to do a distinction between the different POIs marked in the database. The concept of 'urban vibrancy' assumes that what brings vibrancy to the city are the points that attract people, popularity, and economic value (e.g., services, monuments, stores) (Jacobs, 1992). As such, POIs that

[1] A platform that provides the existing information in the OSM for any country, in vector format.

do not fit this scope, such as benches, trash cans, shelters, and surveillance cameras (the full list can be found in **Table 3.4**) were removed in ArcGIS Pro, by using *Select by Attributes* tool, where the SQL Expression is:

> *fclass = 'bench'* **OR** *fclass = 'camera_surveillance'* **OR** *fclass = 'camp_site'* **OR** *fclass = 'caravan_site'* **OR** *fclass = 'comms_tower'* **OR** *fclass = 'drinking_water'* **OR** *fclass = 'embassy'* **OR** *fclass = 'fort'* **OR** *fclass = 'golf_course'* **OR** *fclass = 'lighthouse'* **OR** *fclass = 'observation_tower'* **OR** *fclass = 'post_box'* **OR** *fclass = 'prison'* **OR** *fclass = 'recycling'* **OR** *fclass = 'recycling_clothes'* **OR** *fclass = 'recycling_glass'* **OR** *fclass = 'recycling_metal'* **OR** *fclass = 'recycling_paper'* **OR** *fclass = 'ruins'* **OR** *fclass = 'shelter'* **OR** *fclass = 'swimming_pool'* **OR** *fclass = 'telephone'* **OR** *fclass = 'tower'* **OR** *fclass = 'vending_parking'* **OR** *fclass = 'waste_basket'* **OR** *fclass = 'wastewater_plant'* **OR** *fclass = 'water_mill'* **OR** *fclass = 'water_tower'* **OR** *fclass = 'water_well'* **OR** *fclass = 'water_works'* **OR** *fclass = 'wayside_cross'* **OR** *fclass = 'windmill'.*

Table 3.4. POIs values removed from database.

bench	golf_course	recycling_metal	waste_basket
camera_surveillance	lighthouse	recycling_paper	wastewater_plant
camp_site	observation_tower	ruins	water_mill
caravan_site	post_box	shelter	water_tower
comms_tower	prison	swimming_pool	water_well
drinking_water	recycling	telephone	water_works
embassy	recycling_clothes	tower	wayside_cross
fort	recycling_glass	vending_parking	windmill

The POIs corresponding to swimming pools (*value = 'swimming_pool'*), although they may correspond to leisure, were removed due to the existence of an extremely high value of POIs corresponding to private pools, not accessible to the public. All POIs that fit the scope are mentioned in **Table 3.5**, separated by their thematic keyword only for comprehension purposes.

Table 3.5. POIs values included in database.

Key	Values
amenity	arts_centre, atm, bank, bar, bicycle_rental, biergarten, buddhist, cafe, car_rental, car_wash, christian, christian_anglican, christian_catholic, christian_evangelical, christian_lutheran, christian_orthodox, cinema, clinic, college, community_centre, courthouse, dentist, doctors, fast_food, fire_station, food_court, fountain, graveyard, hindu, hospital, jewish, kindergarten, library, market_place, muslim, nightclub, pharmacy, police, post_office, pub, public_building, restaurant, school, theatre, toilet, town_hall, university, vending_any, vending_machine, veterinary
historic	archaeological, artwork, attraction, castle, guesthouse, hostel, hotel, memorial, monument, museum, picnic_site, theme_park, tourist_info, viewpoint, wayside_shrine, zoo
leisure	dog_park, park, pitch, playground, sports_centre, stadium, track
shop	bakery, beauty_shop, beverages, bicycle_shop, bookshop, butcher, car_dealership, chemist, clothes, computer_shop, convenience, department_store, doityourself. florist, furniture_shop, garden_centre, general, gift_shop, greengrocer, hairdresser, jeweller, kiosk, laundry, mall, mobile_phone_shop. newsagent, optician, outdoor_shop, shoe_shop, sports_shop, stationery, supermarket, toy_shop, travel_agent

The Shannon diversity index was calculated by using the ‘Diversity Calculator’ (https://millermountain.com/diversity/), an online platform that calculates this metric using vector files. By importing the features corresponding to the delimitation of the analysis units (hexagons) and the POIs, using correspondingly the fields indicating the ID of each hexagon and the ‘value’ of each POI, diversity indicators are calculated in each hexagon, including the Shannon diversity index (in the platform it corresponds to «H’»). The values were subsequently exported to Excel format, and the results obtained were validated by manually applying the formula in some hexagons.

Using the feature with the delimitation and identification of each hexagon in the study area, and in order to import the results obtained into GIS environment, the table downloaded in the previous step was joined through the *Join Field* tool in ArcGIS Pro:

> *Geoprocessing* > *Data Management Tools*, using as a common field the identifier of each hexagon

Recreational spaces intensity

The *‘intensity of recreational spaces’* was obtained using the Urban Atlas database, produced by the European Environmental Agency and by Copernicus. In their 2018 version, it delimits the land use typology with a focus on interurban areas, being the best option for city-level studies (such as the case of Lisbon).

According to the report *“Urban Green Spaces and Health”* of WHO (2016), the green spaces indicator can be obtained through the Urban Atlas database, considering the classes ‘green urban areas’ (code 14100), ‘sports and leisure facilities’ (code 14200), ‘agricultural areas, semi-natural areas, and wetlands’ (code 20000 and 40000), and ‘forests’ (code 31000). For Lisbon’s case study, only the class corresponding to ‘agricultural areas, semi-natural areas and wetlands’ was excluded or not considered, because it was agreed that they do not promote the walkability of spaces. In the case of ‘forests’ class, Lisbon includes the Monsanto Forest Park (Parque Florestal de Monsanto in Portuguese), a place with leisure and sport spaces; according to the Urban Atlas database, it is characterized as a forest.

In ArcGIS Pro, the classes mentioned in the previous paragraph were isolated:

> *Select by Attributes* > *SQL Expression* where *code_2018 = ‘14100’* ***OR*** *code_2018 = ‘14200’* ***OR*** *code_2018 = ‘31000’*

and, by using the *Summarize Within* tool:

> *Geoprocessing* > *Analysis Tools*

the total area corresponding to the sum of the three classes, in hectares, was obtained.

To calculate the *‘intensity of recreational spaces’*, it was created, in generated feature from previous step, a column containing the area of each hexagon:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Geometry Attributes*, in which the previously created field is selected, calculating the area in hectares

and, finally, it was created a column where the percentage of recreational spaces is calculated:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Field*, where the field with the total area of recreational spaces is divided by the area obtained in the previous step, multiplied then by 100

Accommodation density

The *'accommodation density'* was obtained by using data from the statistical survey of Census 2021. For this indicator, the gross residential density was considered, obtained through the variable 'total number of accommodations".

The calculation of this indicator, due to the nature of the variable itself and the existence of hexagons with small dimensions in the study area boundaries, results in extremely high values in some hexagons, as explained in **3.1**. So, to obtain the density of accommodations *per* hexagon (the Census data are available up to block level scale), it is first necessary to create a column in the feature with the data at block level, in ArcGIS Pro, containing the area of each block:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Geometry Attributes*, in which the previously created field is selected, calculating the area in hectares

Subsequently, it is created a column where the *'density of accommodations' per* hectare is calculated:

Geoprocessing > *Data Management Tools* > *Add Field*, in which the field type is "Double"; then *Geoprocessing* > *Data Management Tools* > *Calculate Field*, where the field with the accommodation count is divided by the area obtained in the previous step

Finally, to transform the variable into hexagons, it was used the *Spatial Join* tool, in:

Geoprocessing > *Analysis Tools*, where the value in each hexagon will correspond to the value of each parish / block that has the highest percentage of area in each hexagon

Walkability index calculation

To calculate the *'walkability index'*, and to be able to compare all variables under analysis, the results of each indicator were normalized by applying the Z-Score method. After the union of all features with the final results of each variable from ArcGIS Pro, in:

Geoprocessing > Analysis Tools > Union

the values of the respective columns were normalized:

Geoprocessing > Data Management Tools > Standardize Field

and a column was created where the formula for calculating the walkability (**Figure 3.21)**, previously defined, will be applied:

Geoprocessing > Data Management Tools > Add Field, in which the field type is "Double"; then *Geoprocessing > Data Management Tools > Calculate Field*, where it is calculated the *'walkability index'*

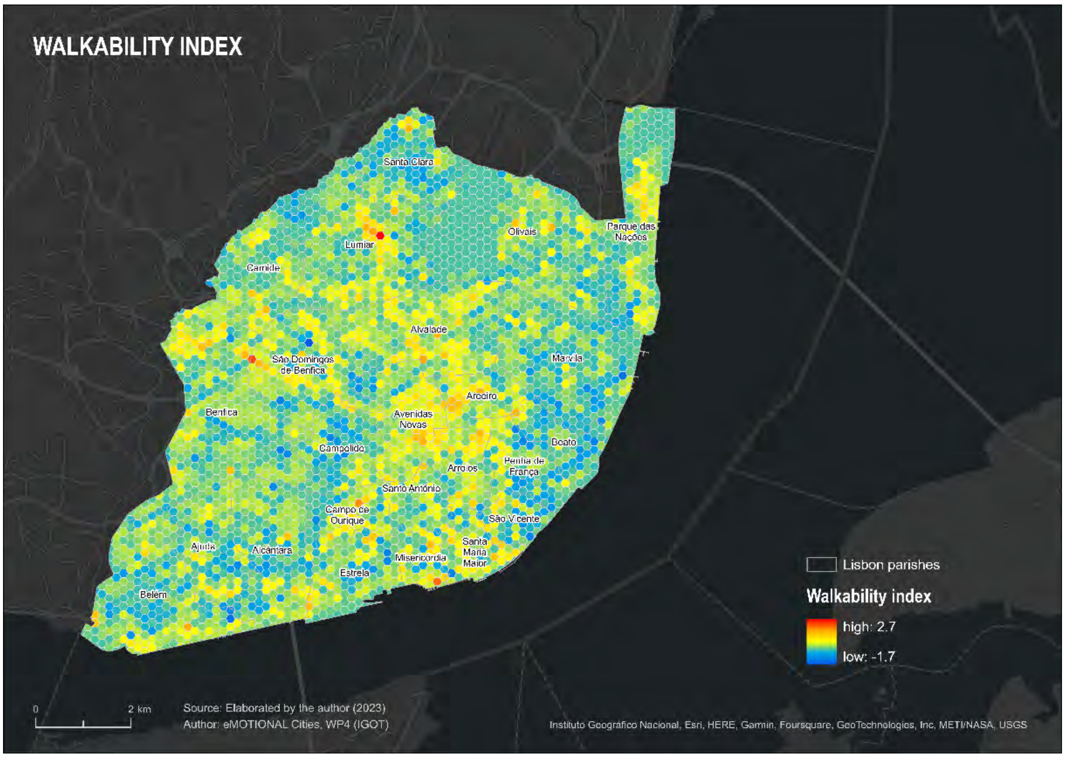


Figure 3.21. Walkability index in Lisbon.

3.3.5 Distance to green spaces

According to WHO (2016), everyone should have, within a 300-meter Euclidean distance, a green space that they can enjoy. For this indicator, and using the delimitation of land use classes from the Urban Atlas database, it was adopted the guidelines defined by the WHO (2016) regarding the classes to be considered to define a green space, mentioned in the previous point (**3.3.4 – subpoint 'Recreational spaces intensity'**); however, in this indicator, the class *'sports and leisure facilities'* was removed due to the existence of a considerable number of spaces that did not correspond to green spaces that the population could benefit from.

In ArcGIS Pro, the chosen classes were isolated in:

Select by Attributes > SQL Expression where *code_2018 = '14100'* ***OR*** *code_2018 = '31000'*

Afterwards, the distance to the green spaces was calculated using the Euclidean distance, which calculates the distance between two points (in this case, between green spaces). To do this, it was used the *Euclidean Distance* tool, in:

Geoprocessing > Analysis Tools, where the boundaries of the study area are defined as the spatial limit for calculating the Euclidean distance

The distances were then weighted from 0 (corresponding to areas where the distance to green spaces is greater than 300 meters) to 1 (corresponding to areas where the distance to green spaces is 0 meters) by applying a formula in *Raster Calculator* tool, in:

Geoprocessing > Analysis Tools, in which the conditional statement to perform the analysis was *("green space Euclidean distance" > 0, -0.0033 * "green space Euclidean distance" + 1*, 1); a linear equation is contained within it, where if the distance to green spaces is greater than 0, the equation is $y = -0.0033x + 1$; otherwise, the value is 1.

However, as the resulting raster contains values below 0, the following conditional statement was applied again in the same tool:

Con("weight distance" < 0, 0,"weight distance"), where if the weighting is less than 0, the value 0 is set; otherwise, the weightings of the raster itself are set.

From the resulting raster and the feature with hexagon delimitation, by using the *Zonal Statistics as Table* tool, in:

Geoprocessing > Image Analyst Tools

it was obtained the average weighted distance in each hexagon. Subsequently, the table obtained in the previous step was joined through the *Join Field* tool in:

Geoprocessing > Data Management Tools, using as a common field the identifier of each hexagon

The results of *'distance to green spaces'* is represented in **Figure 3.22**.

Figure 3.22. Distance to green spaces in Lisbon.

4. Statistical and spatial analysis

After the collection and pre-processing of all the variables included in our study, we proceeded with statistical analysis to identify spatial patterns, trends and relationships within our dataset.

We will outline the steps taken to conduct preliminary statistical analysis of the data, and afterwards we will go into detail, about the process through which we were able to identify critical urban areas that display the greatest negative impacts on the investigated health outcomes. The methodological process is represented in **Figure 4.1**.

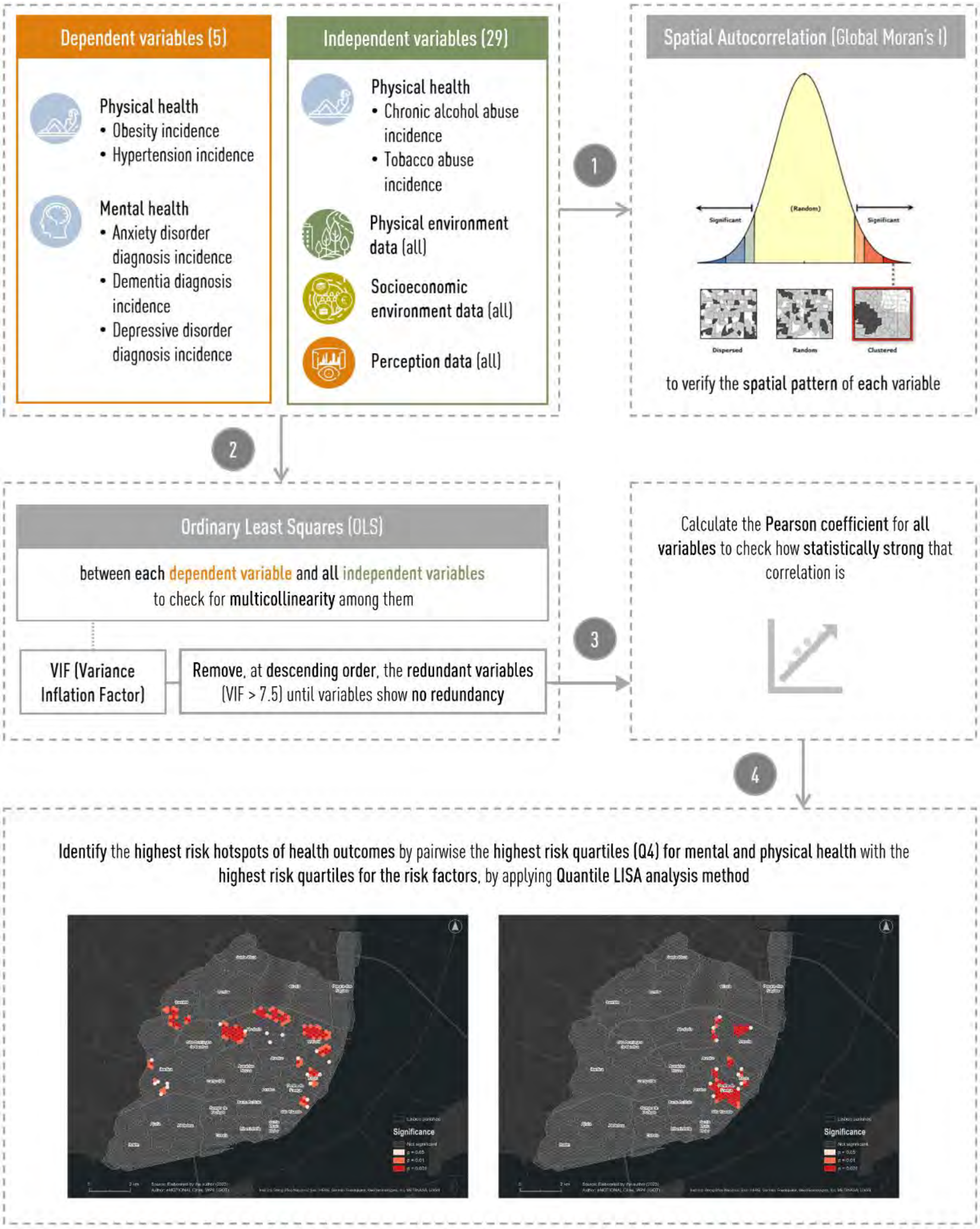


Figure 4.1. Methodological process to obtain higher risk areas of physical and mental health diseases.

Moran's Autocorrelation

The first step was to perform a spatial autocorrelation to all variables, using the *Spatial Autocorrelation (Global Moran's I)* tool, in ArcGIS Pro:

> *Geoprocessing* > *Spatial Statistics Tools*, in which is generated a report

This procedure determines whether the pattern of the data is 'clustered', 'scattered', or 'random', according to a set of features and an associated attribute. The 'null hypothesis' being tested states that the data are randomly distributed among the features in the study region. This tool computes *'Moran's Index value'*, *'variance'*, *'z-*

score' and *'p-value'*; in this case, when the *'p-value'* is statistically significant, the 'null hypothesis' can be rejected if, simultaneously:

The *'z-score'* is positive — the distribution is more spatially clustered than would be expected, if the relationships were absolutely random;

The *'z-score'* is negative — the distribution is more spatially dispersed than would be expected, if the relationships were random (a spatially dispersed pattern reflects a competitive process, i.e., a feature with a high value repels other features with high values; the opposite is also valid).

Variance Inflation Factor (VIF) analysis

Multicollinearity between independent variables in a study indicates the existence of variables that bring redundancy and similar results to each other, biasing the final results obtained. To investigate the existing correlation between all the independent variables, and through an exploratory regression of the Ordinary Least Squares (OLS)[2] model, it is possible to obtain this parameter; to do so, we applied the *Ordinary Least Squares (OLS)* tool, in ArcGIS Pro:

Geoprocessing > *Spatial Statistics Tools*, in which is generated a report

The OLS tool outputs statistical results and diagnostics that provide information on coefficients (*'r-squared'*), *'standard error'*, *'p-values'* and *'Variance Inflation Factor (VIF)'* of the data, where the last is the desired parameter.

Pearson correlation

This indicator determines the linear correlation between all the variables under analysis (dependent and independent) — and how statistically strong that correlation is —, if the distribution of all the values for each variable follows a normal trend. The correlation was calculated in SPSS, using the *Bivariate* tool at:

Analyze > *Correlate*, where the Pearson coefficient calculation was selected

Quantile LISA analysis

Based on the methodology developed by the University of Cambridge for London (in **deliverable 4.3.** *'Mapping of cities based on cognitive aspects and emotional responses triggered by the built environment'*), the Quantile LISA analysis method was also adopted for Lisbon in order to mapping the hotspots; this indicator consists of a

[2] Linear regression is a statistical method used to estimate the linear relationships between a dependent variable and one or more independent variables, determining a single linear equation that fits the data distribution and is used to predict future outcomes. The most well-known linear regression approach is the OLS, which minimizes the sum of squares of residuals between the observed and the predicted values (i.e., variance) of a dependent variable.

bivariate or multivariate analysis between two or more variables, performing a linear spatial autocorrelation between quantiles (GeoDa, https://geodacenter.github.io/workbook/6d_local_discrete/lab6d.html#quantile-lisa).

The tool for conducting this analysis (*Multivariate Quantile LISA*) is implemented in GeoDa software. Once the file with all variables under study has been imported, the tool can be found in the *Space* toolbox menu. Subsequently, it is necessary to create a spatial weights file; inside the tool, in the *Select Spatial Weights* option, a matrix of weights is generated by using the *'queen contiguity'* method. Lastly, the number of quantiles is selected (four in this study) and the respective quartile under analysis (in this study, it will be the extremes — Q1 and Q4).

Results

For the final maps, data regarding the prescription of drugs (anxiolytics, antidepressants and antidementia drugs) was not considered due to the absence of this type of data for the physical health thematic; therefore, data regarding incidence of diseases (depression, dementia, and anxiety) was used. However, the drug prescription data (and its results) will be used to identify new hotspots in different areas of Lisbon.

The results (**Table 4.1**) showed that the spatial distribution of the dependent variables exhibited a significantly clustered pattern, which suggests that there may be other factors influencing this spatial distribution. This calls for further analysis of the spatial relationships between the variables and their geographic context, to gain a deeper understanding of the underlying factors driving the observed patterns. These results were satisfactory and in line with our expectations.

Table 4.1. Results obtained in Spatial Autocorrelation (Global Moran's I) analysis.

Thematic	Dependent variables	**Spatial Autocorrelation** (Global Moran's I)	
		Moran's Index	z-score
Physical Health	Obesity incidence	0.687	68.247
	Hypertension incidence	0.542	53.843
Mental Health	Anxiety disorder diagnosis incidence	0.676	67.156
	Dementia diagnosis incidence	0.535	53.179
	Depressive disorder diagnosis incidence	0.687	68.252

Using the previous results, we checked the possible existence of multicollinearity between the independent variables; from a total of 29 variables, four indicated high VIF values — higher than 7.5 (**Table 4.2**). In descending order, the variables that showed redundancy were removed from the model one by one; after removing the *'patients with tobacco abuse'* variable, only the *'unemployed people ratio'* variable continued to show a high VIF, and so was removed.

Table 4.2. VIF values for the different independent variables.
(shaded cells – variables where VIF > 7.5).

Dimensions	Metrics	VIF
Physical Environment Data	Average age of buildings	4.265
	Buildings with repair needs ratio	1.564
	Average building height	1.799
	Building area ratio	2.368
	Walkability index	1.365
	Altimetry	3.454
	Beds / customers in tourist accommodations	2.357
	Density of fast-food outlets	2.326
	NDVI	2.524
	Distance to green spaces	1.563
	Noise level	1.638
	$PM_{2.5}$	1.625
	NO_2	2.526
	Mean temperature	3.539
	Vulnerability to excessive heat index	3.654
	Vulnerability to flash floods index	2.862

Dimensions	Metrics	VIF
Socioeconomic Environment Data	Purchasing power	2.307
	Unemployed people ratio	**9.642**
	People with low literacy level ratio	4.978
	Population density	1.559
	Gender ratio	1.352
	Youth people ratio	3.432
	Elderly people ratio	3.889
Urban Health Data	Life births rate	2.370
	Mortality rate	1.812
	Patients with Diabetes *Mellitus*	**10.225**
	Patients with chronic alcohol abuse	**9.389**
	Patients with tobacco abuse	**18.152**
Perception data	Density of positive tweets	1.215

Following the removal of redundant variables, and to be able to select the variables for analysis using the Quantile LISA method, the Pearson coefficient between all the variables was calculated (**Table 4.3**). Considering not only the conclusions from **deliverable 2.2**. *'Conceptual framework'* and **deliverable 4.3**., but also the results of the Pearson correlations and the empirical knowledge of the city of Lisbon itself, 12 variables (six from **Physical Environment Data** dimension, five from **Socioeconomic Environment Data** dimension and one from **Perception Data** dimension) were selected.

Table 4.3. Pearson coefficients, by health outcomes and risk factors.
(shaded cells – variables chosen for final analysis)

Dimensions	Metrics	Mental health			Physical health	
		Depression	Dementia	Anxiety	Hypertension	Obesity
Urban Health Data	Life births rate	0.007	,049**	,082**	,045**	,037*
	Mortality rate	,083**	,181**	0.025	,096**	,059**
	Patients with Diabetes *Mellitus*	,960**	,609**	,958**	,658**	,969**
	Patients with chronic alcohol abuse	,856**	,574**	,829**	,644**	,894**
Physical Environment Data	Average age of buildings	,605**	,324**	,625**	,343**	,591**
	Buildings with repair needs ratio	,211**	,137**	,179**	,133**	,225**
	Average building height	,377**	,204**	,404**	,215**	,357**
	Building area ratio	,246**	,147**	,278**	,171**	,259**
	Walkability index	**,075****	**,113****	**,088****	**,075****	**,065****
	Altimetry	-0,063	0,047**	-0,081**	-0.004	-,088**
	Beds / customers in tourist accommodations	-,064**	-,053**	-0.031	-,078**	-0.027
	Density of fast-food outlets	,127**	,134**	,147**	,109**	,100**
	NDVI	**-,134****	**-0.017**	**-,142****	**-,054****	**-,140****
	Distance to green spaces	**-,125****	**-0.014**	**-,122****	**-,050****	**-,139****
	Noise level	**-,167****	**-,094****	**-,188****	**-,128****	**-,223****
	$PM_{2.5}$	**,066****	**,040***	**,062****	**,048****	**,127****
	NO_2	**,138****	**,037***	**,151****	**0.000**	**0.006**
	Mean temperature	**,266****	**,081****	**,289****	**,132****	**,275****
	Vulnerability to excessive heat index	0.032	,052**	,056**	-,035*	-0.028
	Vulnerability to flash floods index	,135**	-0.020	,208**	-0.007	,060**
Socioeconomic Environment Data	Purchasing power	**-,081****	**-0.029**	**-0.023**	**-,068****	**-,154****
	People with low literacy level ratio	**,384****	**,358****	**,351****	**,371****	**,404****
	Population density	**,249****	**,150****	**,270****	**,149****	**,260****
	Gender ratio	**,193****	**0.011**	**,207****	**0.021**	**,207****
	Youth people ratio	,299**	-,061**	,322**	-,036*	,290**
	Elderly people ratio	**,415****	**,779****	**,392****	**,777****	**,392****
Perception data	Density of positive tweets	**-0.003**	**-0.019**	**0.015**	**-0.004**	**0.029**

With the Quantile LISA analysis method, only the highest risk hotspots were identified, i.e., when considering the variables *'depressive disorder diagnosis incidence'* and *'NDVI'*, the Q4 and Q1 quantiles were selected, respectively, where the incidence of depression will be the highest and the green spaces will be non-existent.

Table 4.4 represents the results of the analysis carried out by using the Quantile LISA analysis method, after obtaining the maps for the 12 independent variables mentioned above and selecting the most spatially relevant variables (in **Appendix 3** contains all maps). It is divided by health outcome thematic (mental health and physical health), and selected risk factors; at the end, a map is obtained for each risk factor combined by each health outcome, for each health outcome thematic (combining the different maps of outcomes), and a final map (combining the two maps of health outcome thematic).

Based on the results obtained when pairwise the highest risk quartiles (Q4) for mental and physical health with the highest risk quartiles for the selected independent variables (**Figure 4.2** and **Appendix 4**), there is a higher concentration of high risk hotspots mainly in the eastern Lisbon area (Penha de França-Beato-Marvila/Braço de Prata axis), but also at some points along the 2ª Circular (Alvalade and the old area of Carnide) and in the Bairro da Boavista. These areas will therefore be potential areas of interest for on-site verification of the presence of risk factors and evaluating their influence on people.

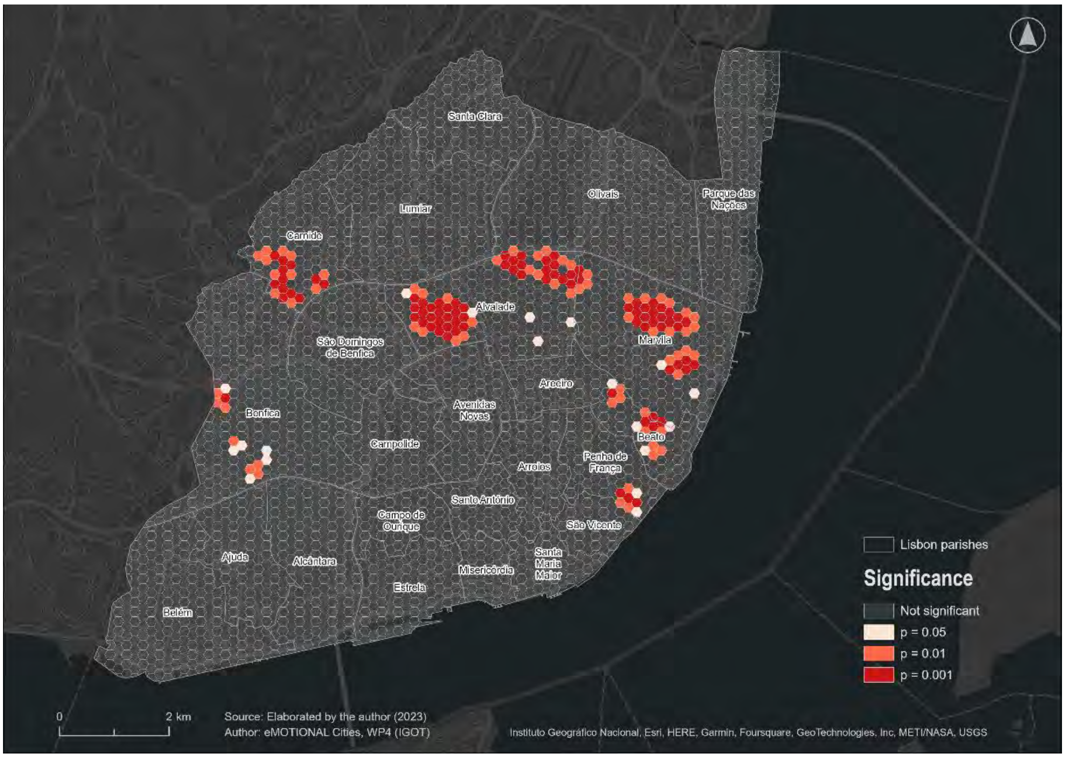


Figure 4.2. High mental and physical health risk associated with high ratio of elderly people in Lisbon.

Table 4.4. Results of Quantile LISA analysis, by health outcomes and selected risk factors.

5. Conclusions

This report presents an overview of the procedures developed so far in the research conducted at IGOT, emphasising the methodology used to treat and analyse our data.

Our study focuses on four categories of urban health-related variables: urban physical environment, health-related variables, socioeconomic-related variables, and perception-related variables. We began by describing the processes used for data collection and variable selection, which was sustained on the theoretical framework on urban health and wellbeing. Then, we explored the techniques performed to transform the initially raw data into meaningful and measurable variables for further analysis.

Following the pre-processing stages, statistical analysis was performed to uncover spatial patterns, trends, and relationships between variables. First, we conducted spatial autocorrelation diagnosis to our dependent variables, using Moran's autocorrelation method. The results revealed that the variables' spatial distribution was significantly clustered, meaning they were not distributed randomly, but rather showed a distinct spatial pattern. This suggests the existence of other variables with heterogeneous geographic distribution that potentially affect the negative health outcomes, as expected.

In order to investigate the correlation between independent and dependent variables, and estimate future outcomes, we first applied OLS regression analysis to obtain an indicator that verifies the collinearity of the variables — VIF —, and then applied Pearson correlation to see how statistically strong correlation between variables is. The Quantile LISA results enabled us to identify crucial urban regions with the highest negative impacts on the assessed health outcomes, by pairwise the highest risk quartiles (Q4) for mental and physical health variables with the highest risk quartiles for the final nine independent variables.

We hope that this report brings some insights, and that it can work and a guidebook on the methodologies applied to gain a deeper understanding of the complex relationships between urban environments and negative health outcomes. Our team will continue to explore and improve different approaches to data spatial analysis, to provide additional knowledge about the underlying determinants of urban health, in the city of Lisbon.

6. References

Aldegunde, J. A. Á., Sánchez, A. F., Saba, M., Bolaños, E. Q., & Palenque, J. Ú. (2022). Analysis of PM2.5 and meteorological variables using enhanced geospatial techniques in developing countries: a case study of Cartagena de Indias City (Colombia). *Atmosphere*, *13*(4). https://doi.org/10.3390/atmos13040506

Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., ... & Lamsal, L. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *The Lancet Planetary Health*, *6*(1), 49-58. https://doi.org/10.1016/S2542-5196(21)00255-2

Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, *206*(3-4), 347-359. https://doi.org/10.1016/j.ecolmodel.2007.03.041

Botta, F., & Gutiérrez-Roig, M. (2021). Modelling urban vibrancy with mobile phone and OpenStreetMap data. *PLOS ONE*, *16*(6). https://doi.org/10.1371/journal.pone.0252015

Jacobs, J. (1992). *The Death and Life of Great American Cities*. Vintage Books. ISBN: 0-679-74195-X

Mejía, D. C., Alvarez, H., Zalakeviciute, R., Macancela, D., Sanchez, C., & Bonilla, S. (2023). Sentinel satellite data monitoring of air pollutants with interpolation methods in Guayaquil, Ecuador. *Remote Sensing Applications: Society and Environment*, *31*. https://doi.org/10.1016/j.rsase.2023.100990

Morillo, M. C., Martínez-Cuevas, S., García-Aranda, C., Molina, I., Querol, J. J., & Martínez, E. (2022). Spatial analysis of the particulate matter (PM10) an assessment of air pollution in the region of Madrid (Spain): spatial interpolation comparisons and results. *International Journal of Environmental Studies*. https://doi.org/10.1080/00207233.2022.2072585

Pereira, S. (2017). *Ambiente Construído - Atividade Física: Uma Equação Para a Saúde. Perspetiva Interdisciplinar Sobre a Construção da Cidade Saudável*. [Built Environment - Physical Activity: An Equation for Health. Interdisciplinary Perspective on the Construction of the Healthy City] [PhD Thesis]. Instituto Superior Técnico da Universidade de Lisboa.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution*, *20*(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall. ISBN: 0-412-24620-1

Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., ... & Martin, R. V. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, *55*(22). https://doi.org/10.1021/acs.est.1c05309

WHO. (2021). *WHO global air quality guidelines. Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization, https://apps.who.int/iris/handle/10665/345329

WHO Regional Office for Europe. (2016). *Urban green spaces and health*. WHO Regional Office for Europe. https://apps.who.int/iris/handle/10665/345751

Appendix 1. Urban health data for spatial analysis (in Lisbon)

Dimensions	Aspects	Metrics	Methods	Data Source	Datetime	Original coordinate system	Resolution
Urban Health Data	Physical health	Life births rate	*Per* 1 000 inhabitants	Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Mortality rate	*Per* 1 000 inhabitants	Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Patients with Diabetes *Mellitus*	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients with chronic alcohol abuse	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients with tobacco abuse	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients with obesity	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients with hypertension	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
	Mental health	Patients diagnosed with dementia	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients diagnosed with anxiety disorder	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Patients diagnosed with depressive disorder	In percentage (%)	*Sistema de Informação e Monitorização do SNS* (SIM@SNS) (2023)	June 2022	ETRS 1989 Portugal TM06	-
		Drug prescription of anxiolytics (N05B)	Normalized value of dosage data per number of people, *per* pharmacy	*Centro de Estudos e Avaliação em Saúde* (CEFAR) / *Associação Nacional de Farmácias* (ANF) (2023)	2021	ETRS 1989 Portugal TM06	-
		Drug prescription of antidepressants (N06A)	Normalized value of dosage data per number of people, *per* pharmacy	*Centro de Estudos e Avaliação em Saúde* (CEFAR) / *Associação Nacional de Farmácias* (ANF) (2023)	2021	ETRS 1989 Portugal TM06	-
		Drug prescription of antidementia (N06D)	Normalized value of dosage data per number of people, *per* pharmacy	*Centro de Estudos e Avaliação em Saúde* (CEFAR) / *Associação Nacional de Farmácias* (ANF) (2023)	2021	ETRS 1989 Portugal TM06	-
Physical Environment Data	Buildings	Average age of buildings	Constructed between 1919 and 2021	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Buildings with repair needs ratio	In percentage (%)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Average building height	In meters (m)	*Câmara Municipal de Lisboa* / LRB (2019)	2019	ETRS 1989 Portugal TM06	-
		Building area ratio	In percentage (%)	*Câmara Municipal de Lisboa* / LRB (2019)	2019	ETRS 1989 Portugal TM06	-
	Streets	Walkability index	Total of intersections	NAVTEQ / ESRI (2016)	2016	WGS 1984	-
			Mean slope of streets, in meters (m)	*Instituto Geográfico do Exército* (n.d.)	-	Lisboa Hayford Gauss IGeoE	25x25 m
			Urban vibrancy, by Points of Interest (POI) diversity	GeoFabrik (2023)	2023	WGS 1984	-
			Sports, leisure, and green urban spaces ratio, in percentage (%)	European Environmental Agency / Copernicus (2020)	2018	ETRS 1989 LAEA	-
			Accommodation density, *per* hectare (ha)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
	Land use	Altimetry	In meters (m)	*Instituto Geográfico do Exército* (n.d.)	-	Lisboa Hayford Gauss IGeoE	25x25 m
		Beds / customers in tourist accommodations	*Per* 1 000 inhabitants	*Turismo de Portugal* (2021)	2021	ETRS 1989 Portugal TM06	-
		Density of fast-food outlets	Kernal density of data from 100 Montaditos, Burger King, Burger Ranch, Domino's Pizza, McDonald's, Pizza Hut, Telepizza, KFC, Taco Bell, Subway, Pans & Company, and Papa John's chains	Elaborated by the authors (2022)	17 to 21 March 2022	ETRS 1989 Portugal TM06	10x10 m
		Normalized Difference Vegetation Index (NDVI)	Annual mean	European Union / ESA / Copernicus (2022)	2021	WGS 1984	10x10 m
		Distance to green spaces	Green spaces up to 300m linear distance	European Environmental Agency / Copernicus (2020)	2018	ETRS 1989 LAEA	-
	Environmental	Noise level	Lden noise map (day, afternoon, and night), in weighted decibel [dB(A)]	*Câmara Municipal de Lisboa* (2021)	2020	ETRS 1989 Portugal TM06	-
		Particulate Matter ($PM_{2.5}$)	Annual mean, in micrograms *per* cubic meter air ($\mu g/m^3$)	Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., ... & Martin, R. V. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, 55(22)	2021	WGS 1984	0.01° × 0.01°
		Nitrogen Dioxide (NO_2)	Annual mean, in parts *per* billion by volume (ppbV)	Anenberg, S. C., Mohegh, A., Goldberg, D. L., Kerr, G. H., Brauer, M., Burkart, K., ... & Lamsal, L. (2022). Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *The Lancet Planetary Health*, 6(1), 49-58	2020	WGS 1984	0.0083 ° x 0.0083 °
		Mean temperature	In the warmest months (June and August), in Celsius degrees (ºC)	Copernicus Climate Change Service (2019)	2017	ETRS 1989 LAEA	100x100 m
		Vulnerability to excessive heat index	From "inexistent" risk to "high" risk	PMAAC-AML (2018)	Actual vulnerability	ETRS 1989 Portugal TM06	-
		Vulnerability to flash floods index	From "very low" risk to "high" risk	PMAAC-AML (2018)	Actual vulnerability	ETRS 1989 Portugal TM06	-
Socioeconomic Environment Data	Socioeconomics	Purchasing power	*Per* capita, in euros (€)	Esri, Michael Bauer Research GmbH (2022)	2021	ETRS 1989 Portugal TM06	-
		Unemployed people ratio	In percentage (%)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		People with low literacy level ratio	Education level up to 9th grade, in percentage (%)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
	Demographics	Population density	*Per* hectare (ha)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Gender ratio	Number of males *per* 100 females	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Youth people ratio	0 to 24 years, in percentage (%)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
		Elderly people ratio	65 years or above, in percentage (%)	2021 Census, Statistics Portugal (2022)	2021	ETRS 1989 Portugal TM06	-
Perception data	Sentiment analysis	Density of positive tweets	Kernel density of tweets	Twitter (2022)	2018 to 2021	WGS 1984	-

Appendix 2: Urban health maps for spatial analysis (in Lisbon)

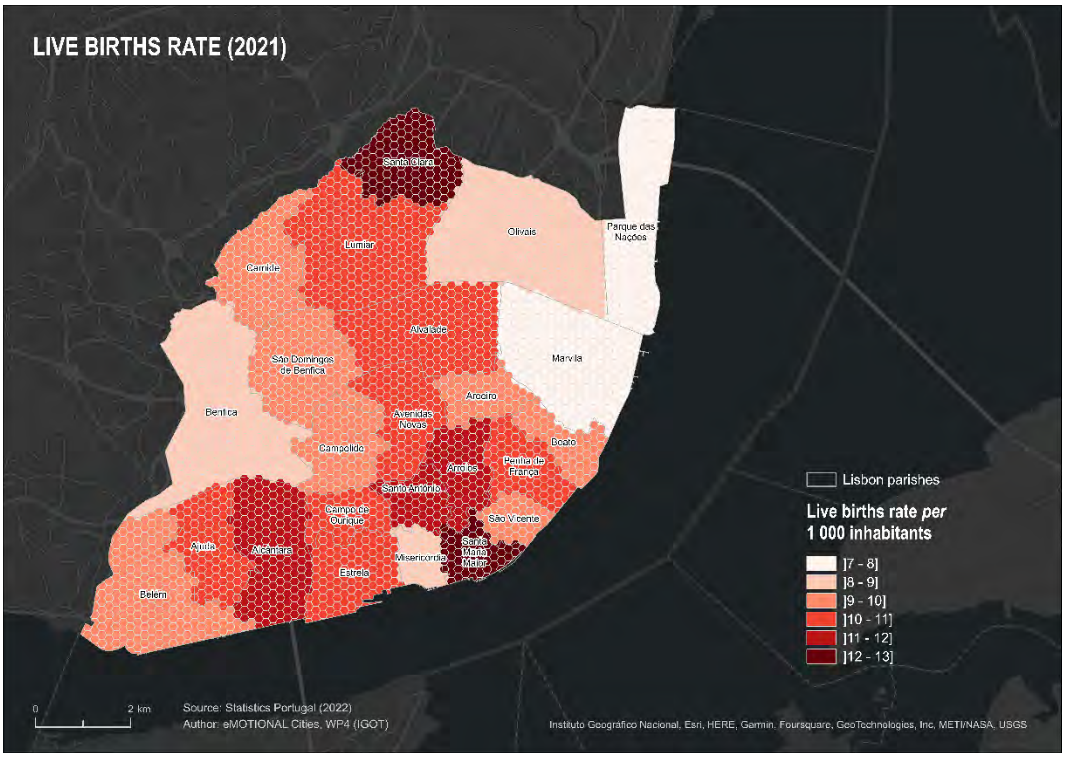


Figure A2.1. Live births rate in Lisbon.

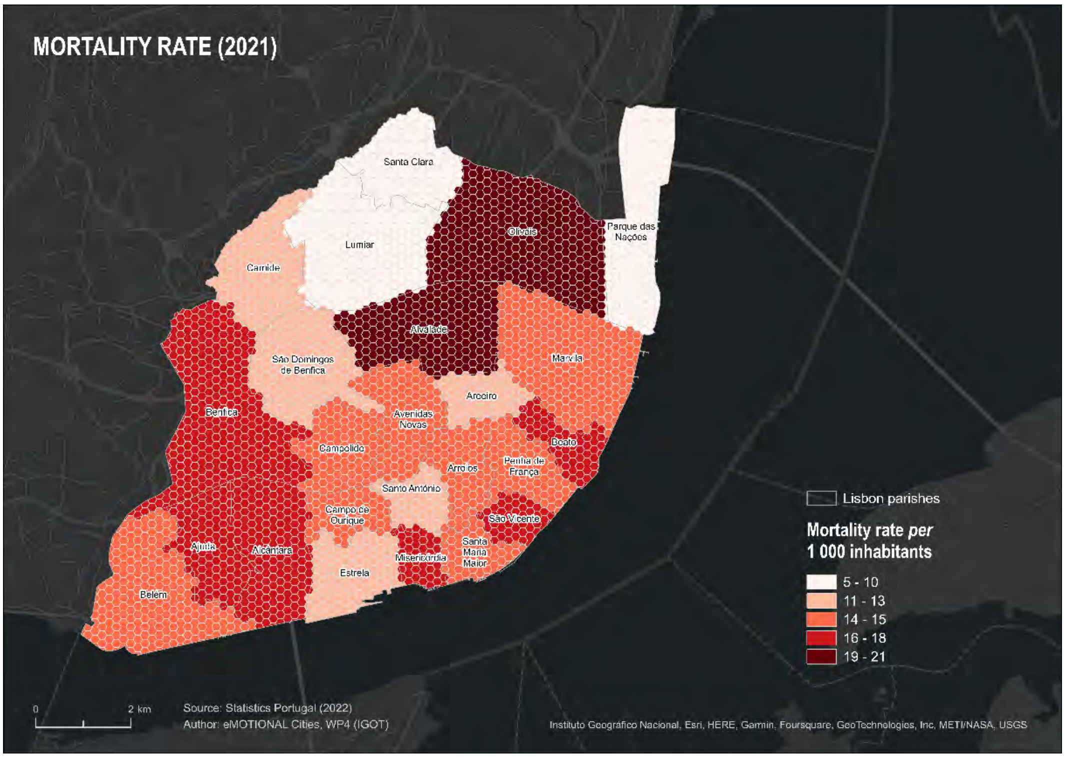


Figure A2.2. Mortality rate in Lisbon.

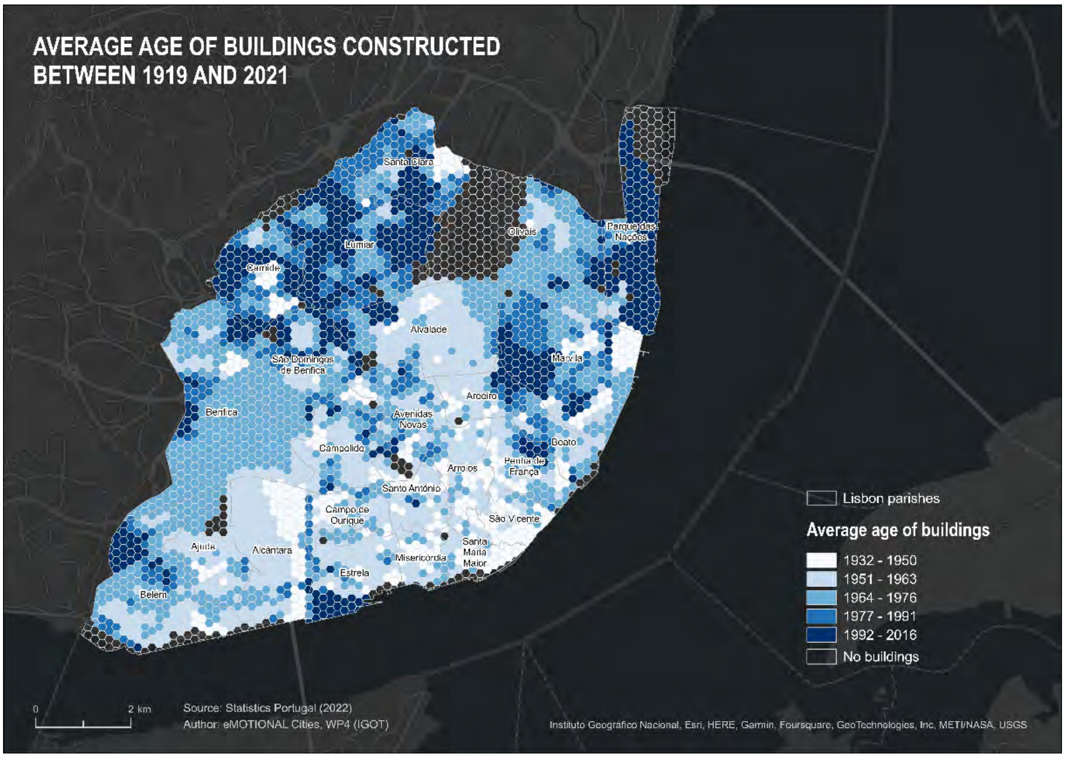


Figure A2.3. Average age of buildings in Lisbon.

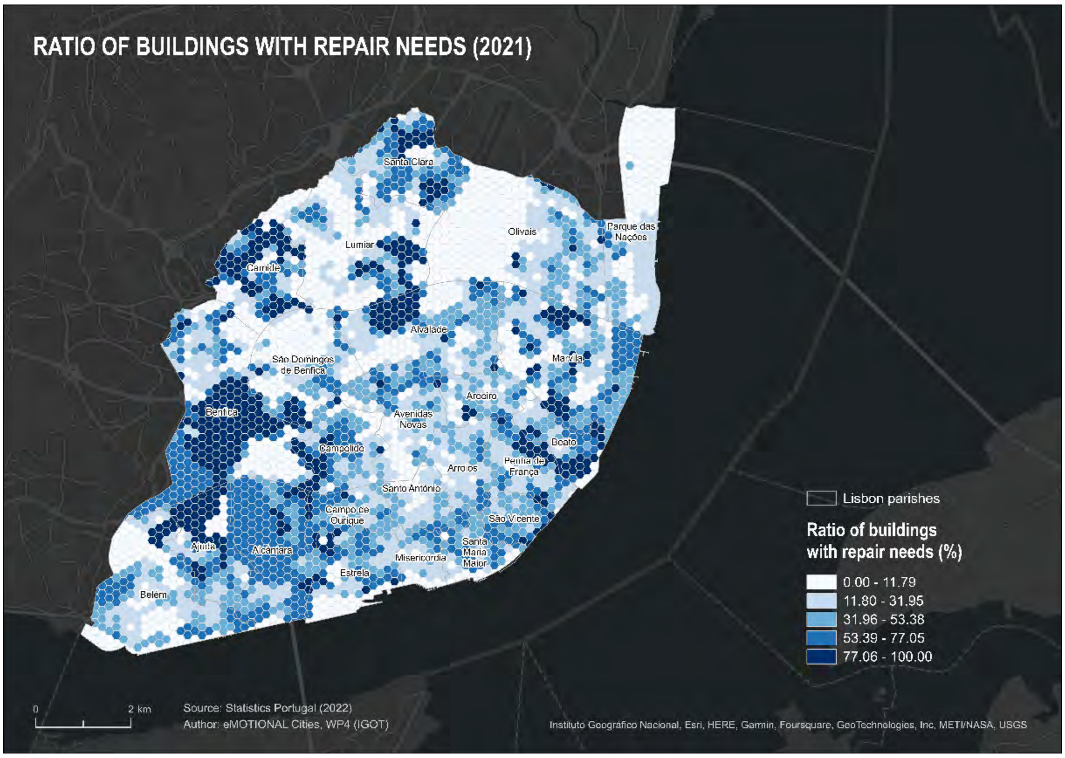


Figure A2.4. Ratio of buildings with repair needs in Lisbon.

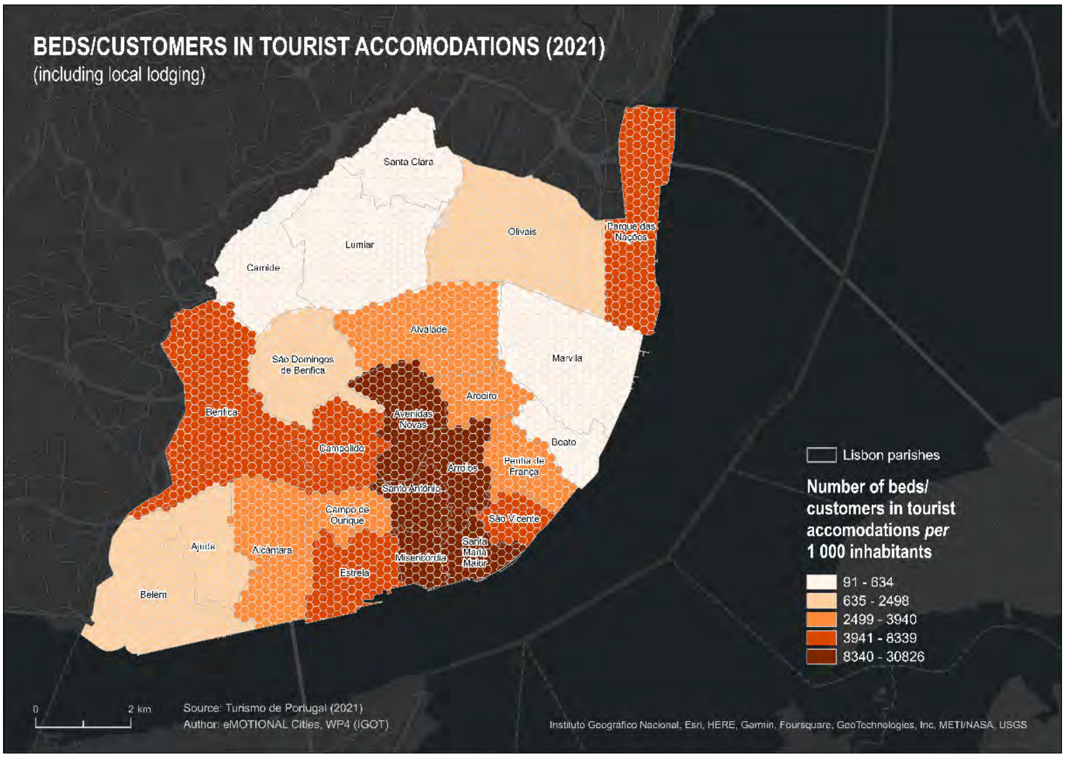


Figure A2.5. Beds / customers in tourist accommodations in Lisbon.

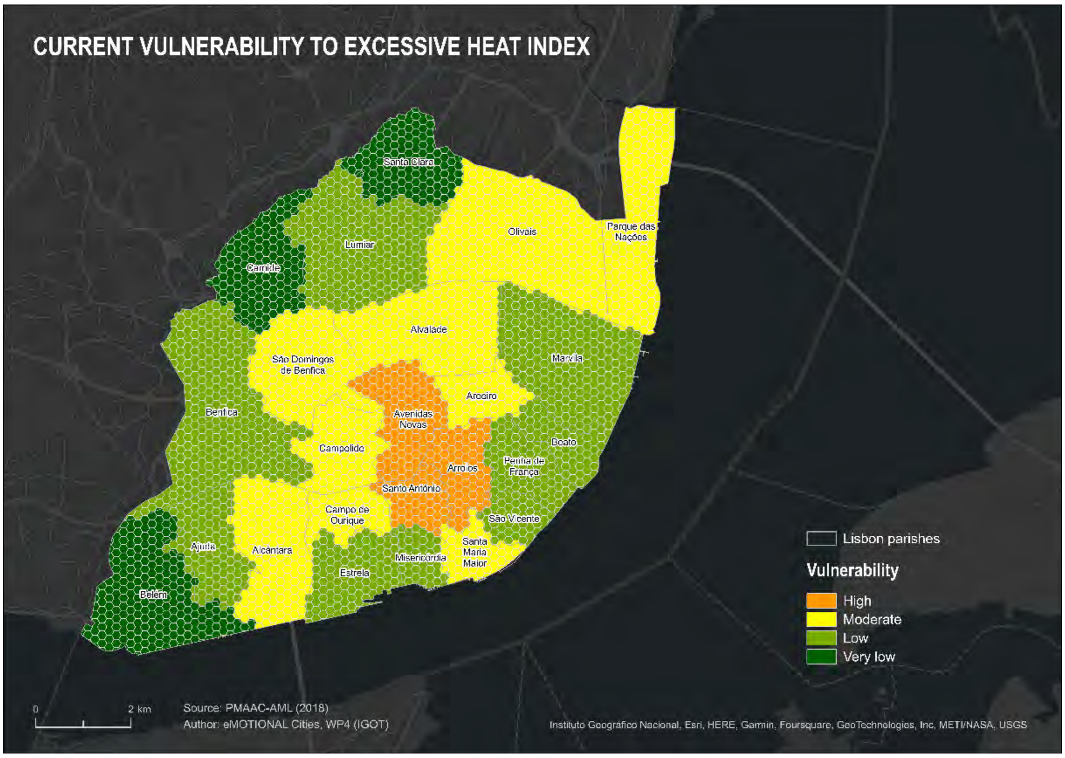


Figure A2.6. Vulnerability to excessive heat index in Lisbon.

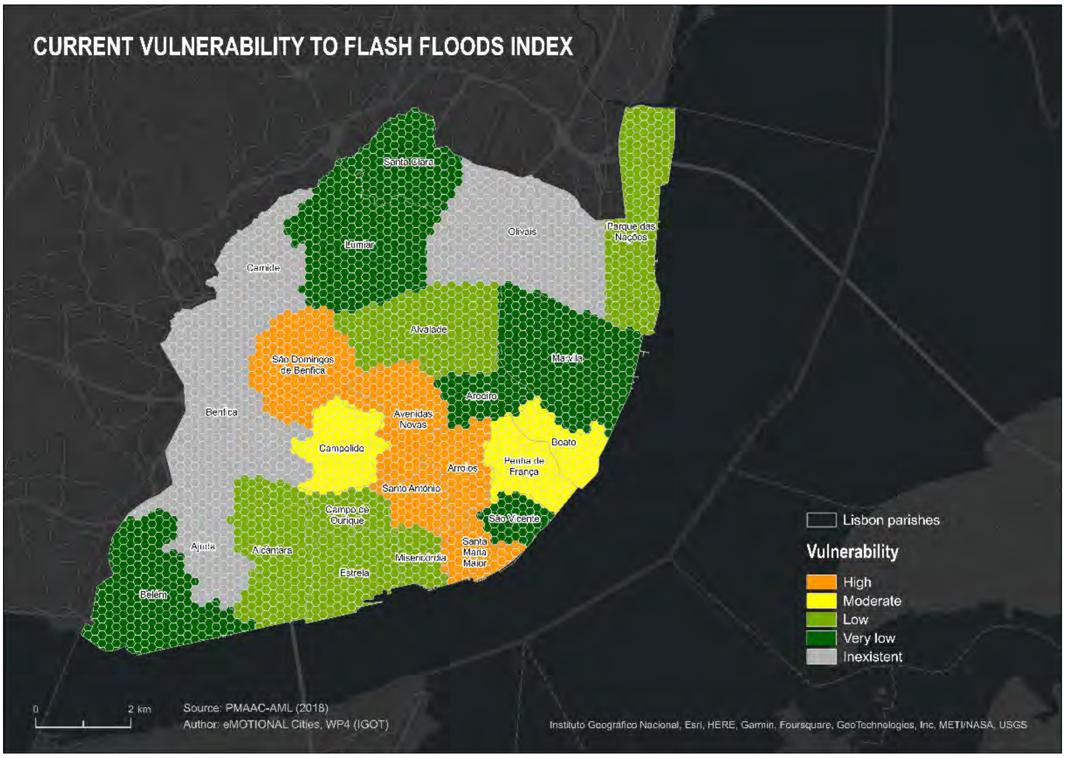


Figure A2.7. Vulnerability to flash floods index in Lisbon.

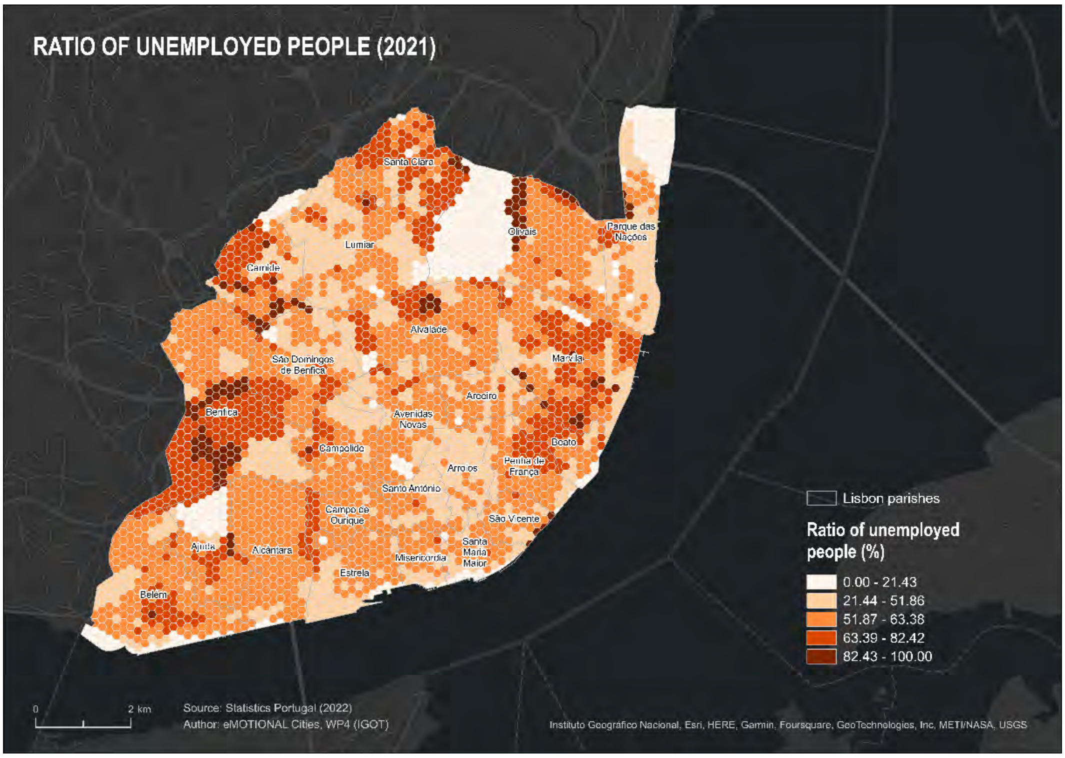


Figure A2.8. Unemployed people ratio in Lisbon.

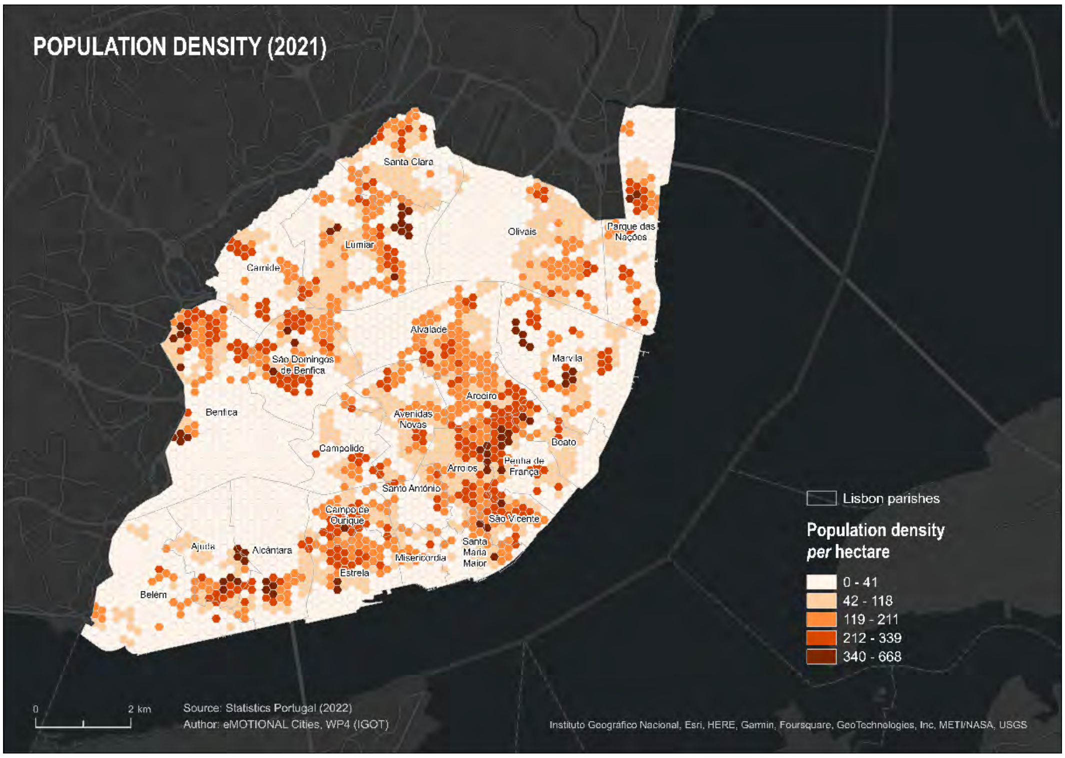


Figure A2.9. Population density in Lisbon.

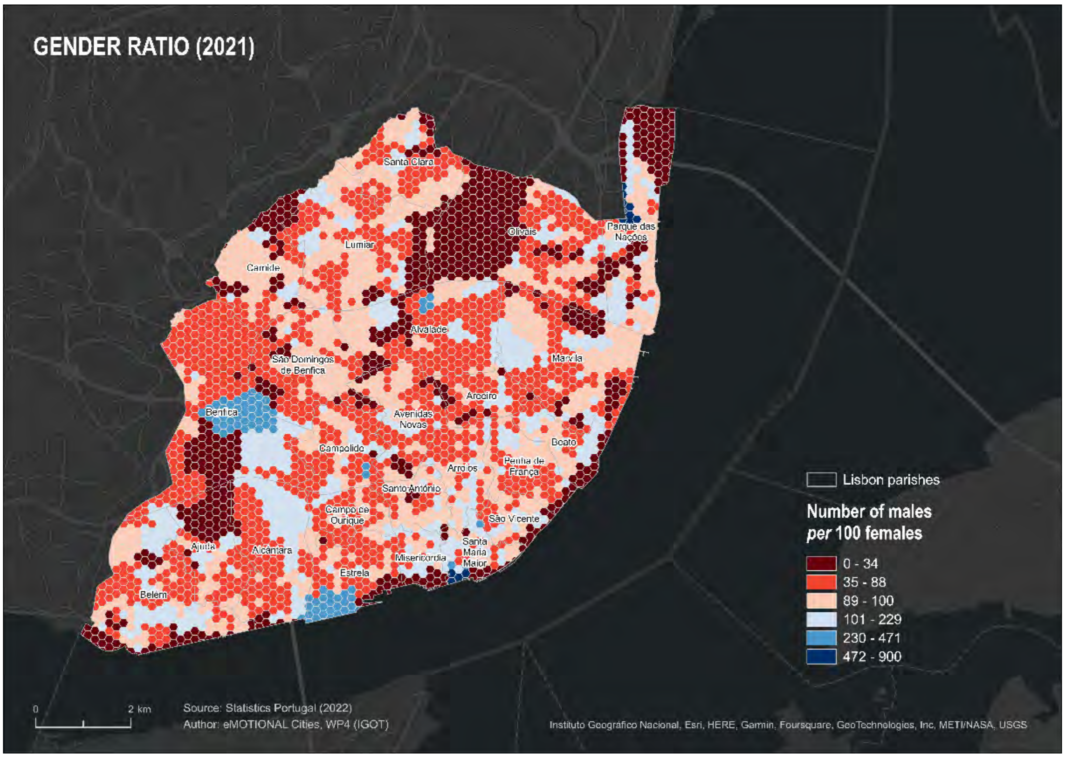


Figure A2.10. Gender ratio in Lisbon.

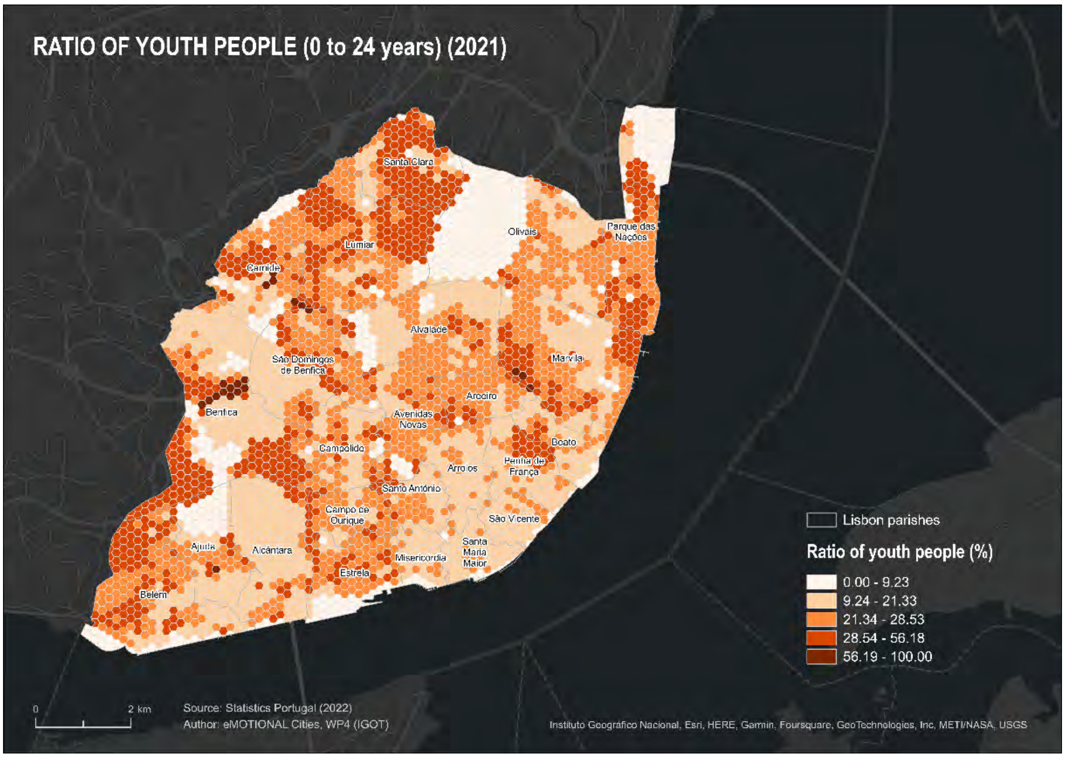


Figure A2.11. Ratio of youth people in Lisbon.

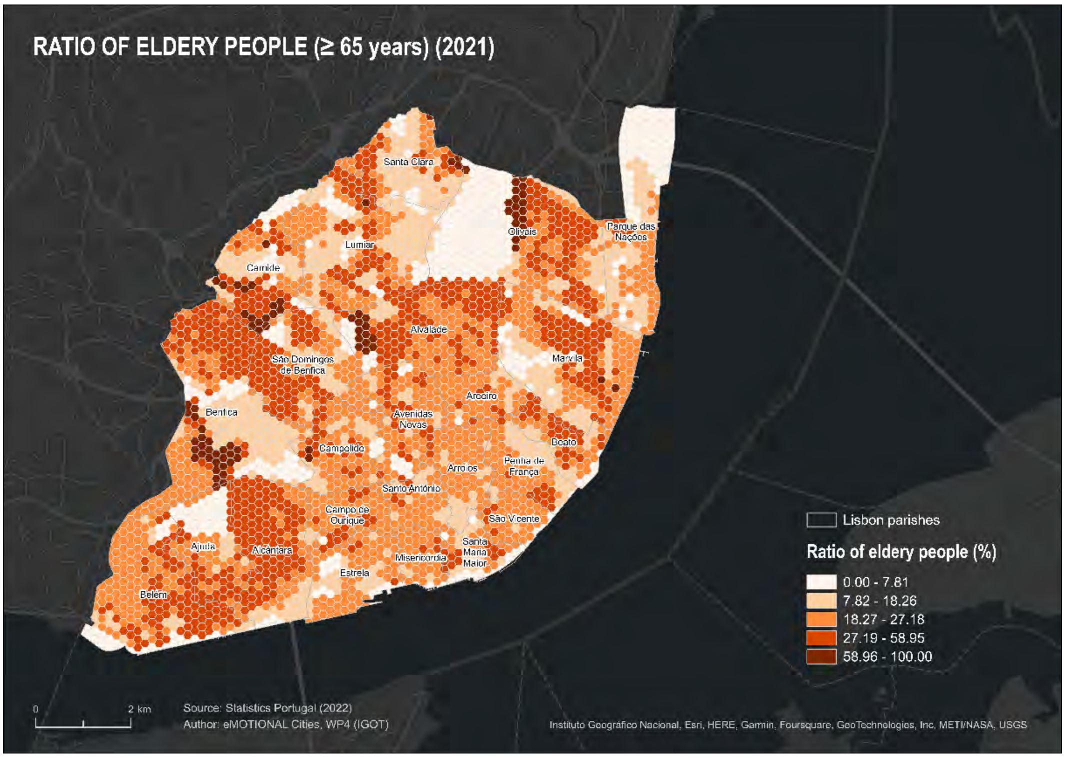


Figure A2.12. Ratio of elderly people in Lisbon.

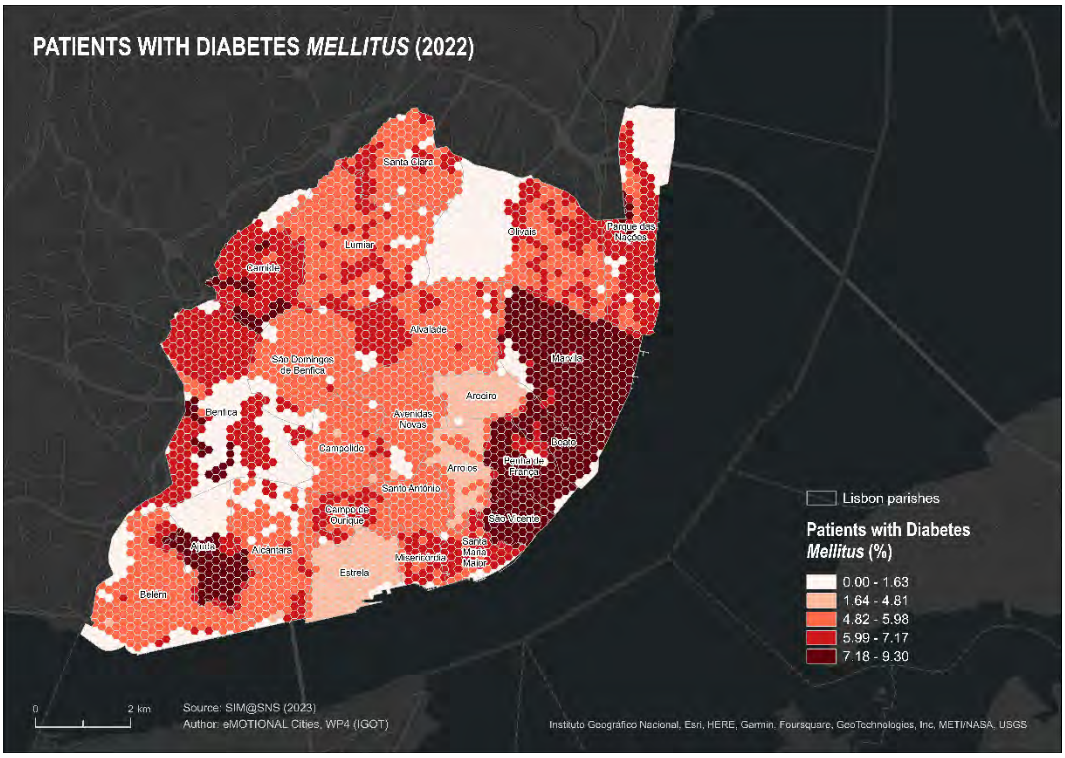


Figure A2.13. Patients with Diabetes *Mellitus* in Lisbon.

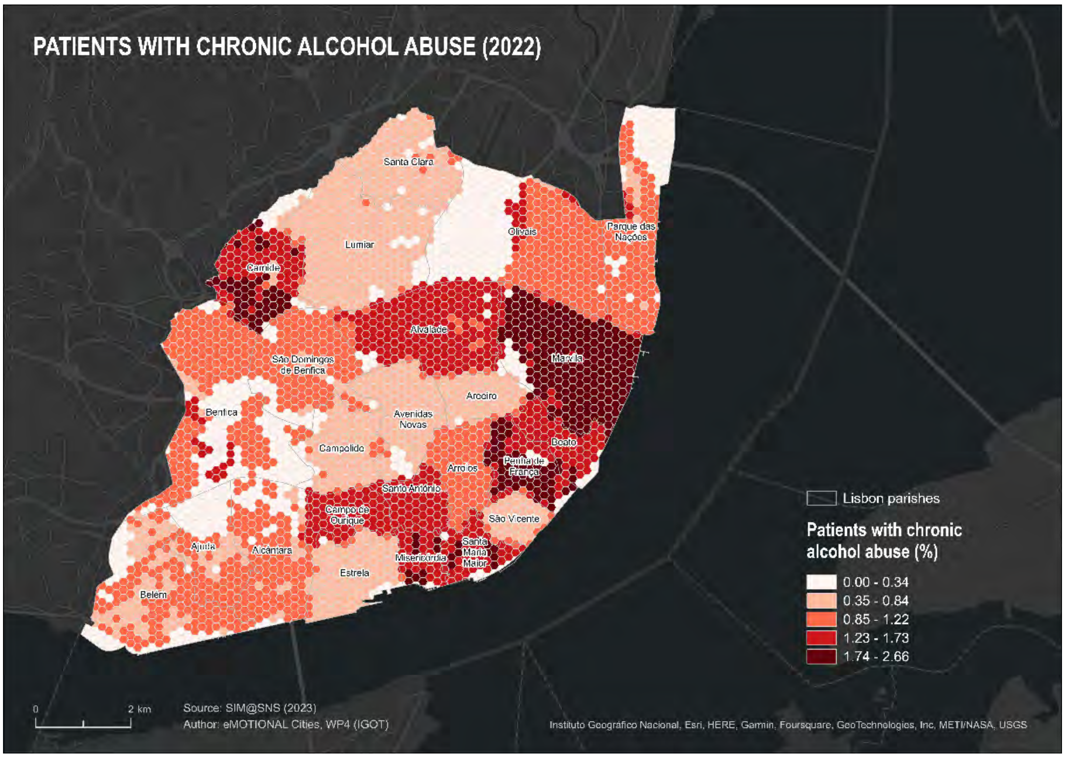


Figure A2.14. Patients with chronic alcohol abuse in Lisbon.

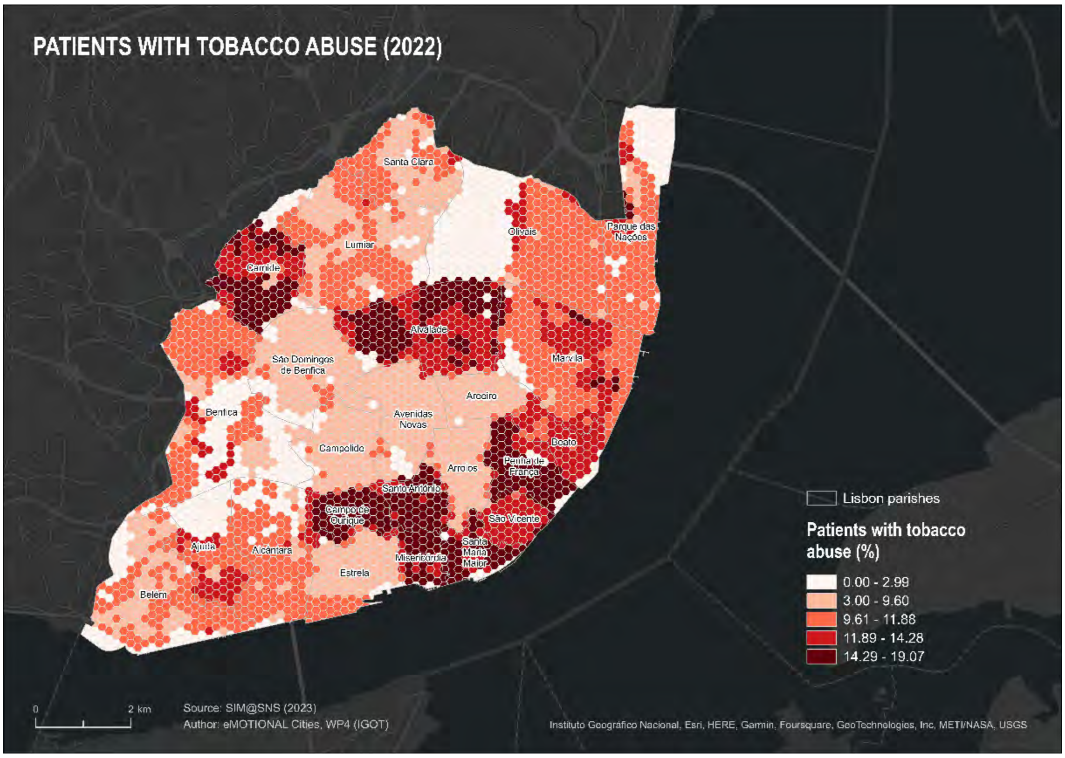


Figure A2.15. Patients with tobacco abuse in Lisbon.

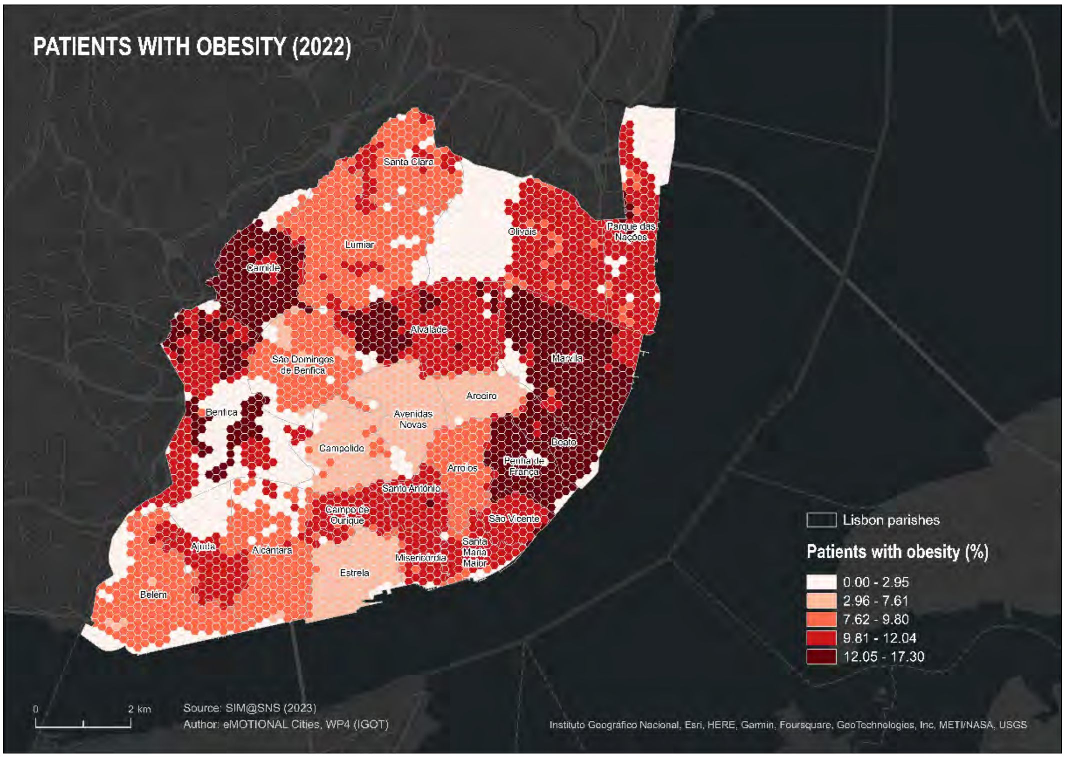


Figure A2.16. Patients with obesity in Lisbon.

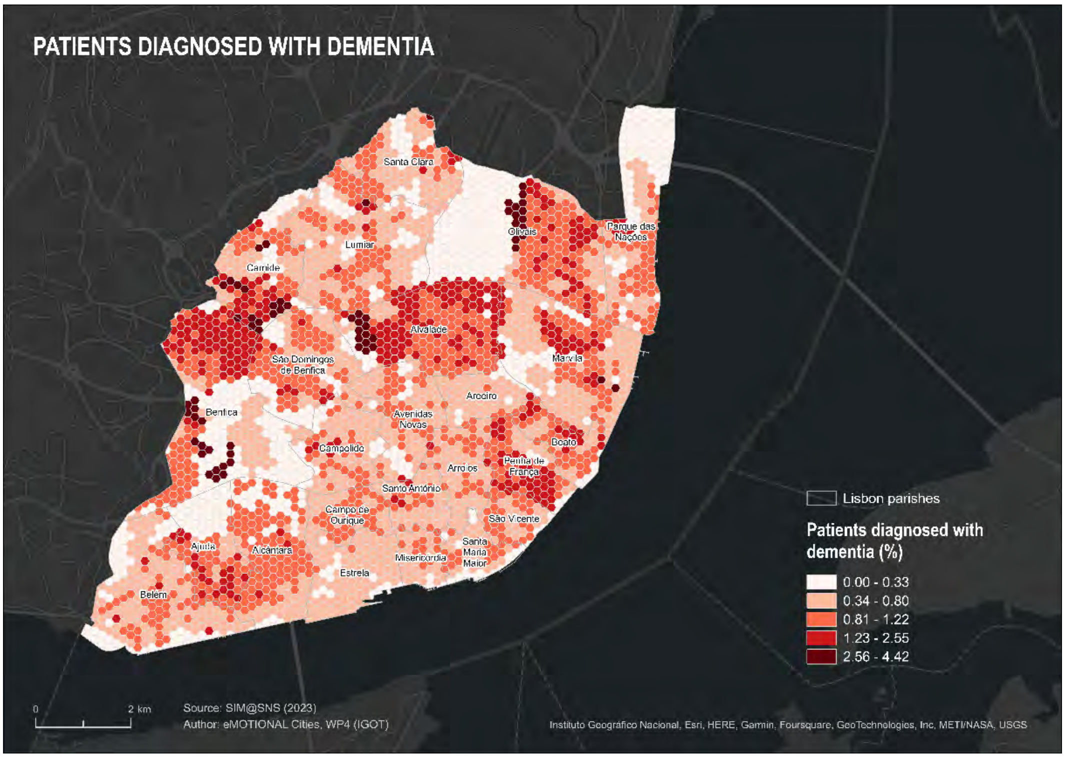


Figure A2.17. Patients diagnosed with dementia in Lisbon.

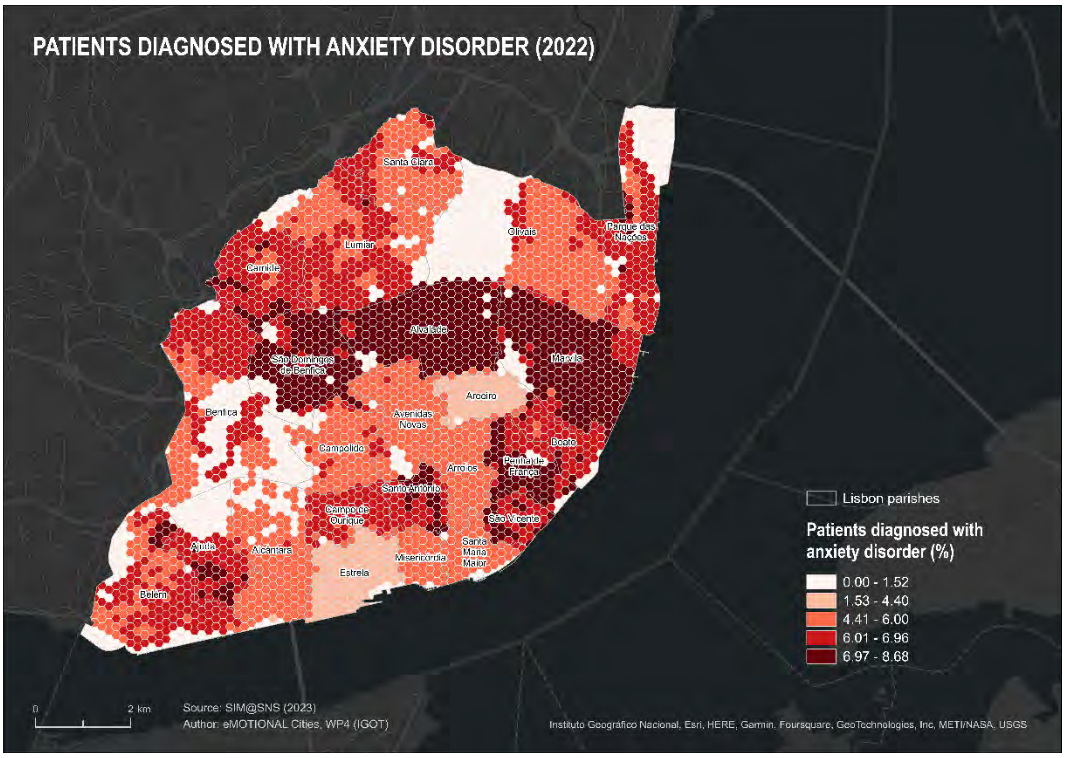


Figure A2.18. Patients diagnosed with anxiety disorder in Lisbon.

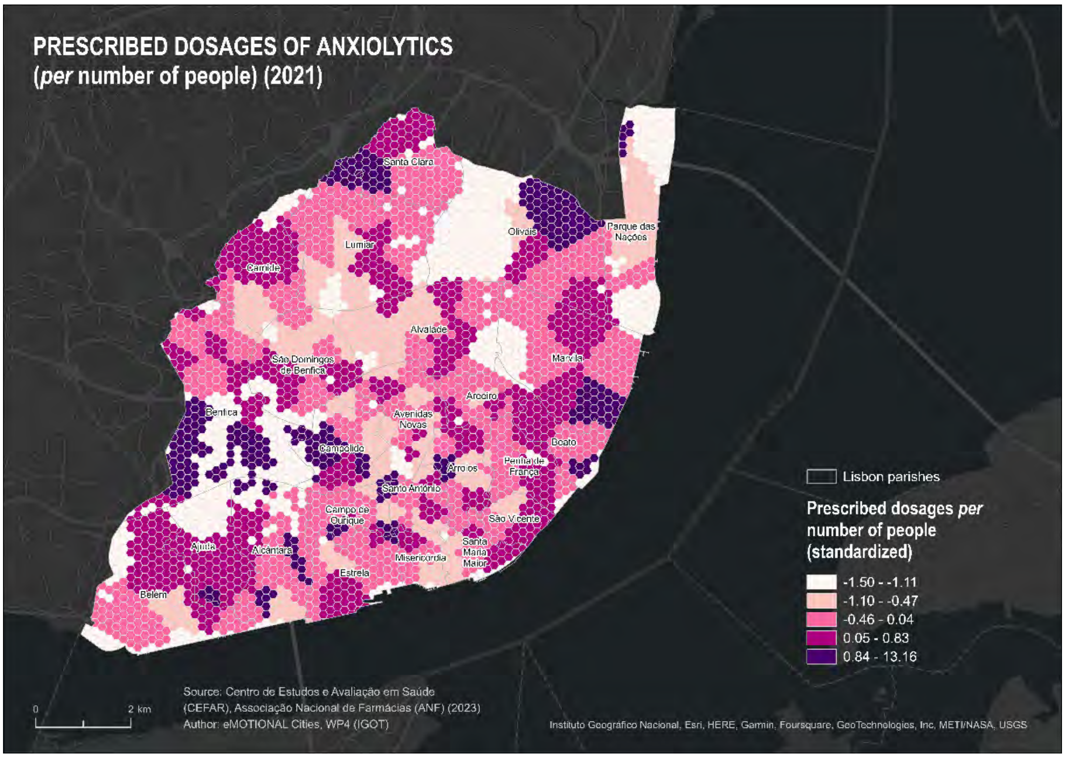


Figure A2.19. Prescribed dosages of anxiolytics in Lisbon.

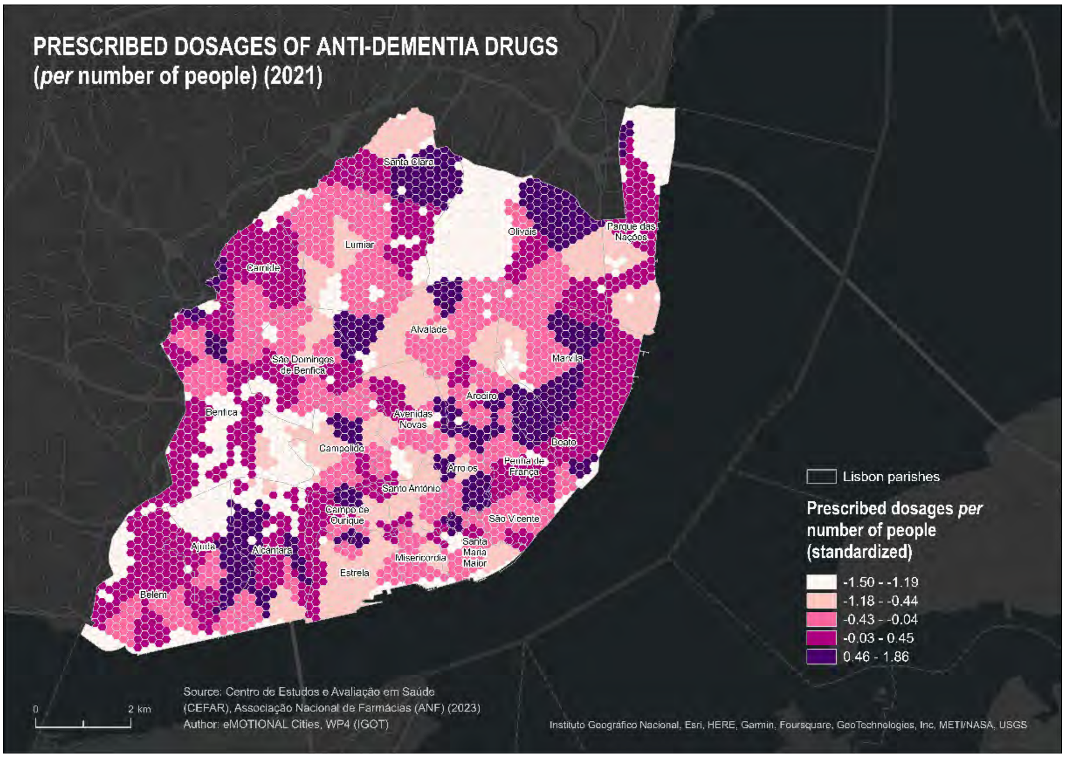
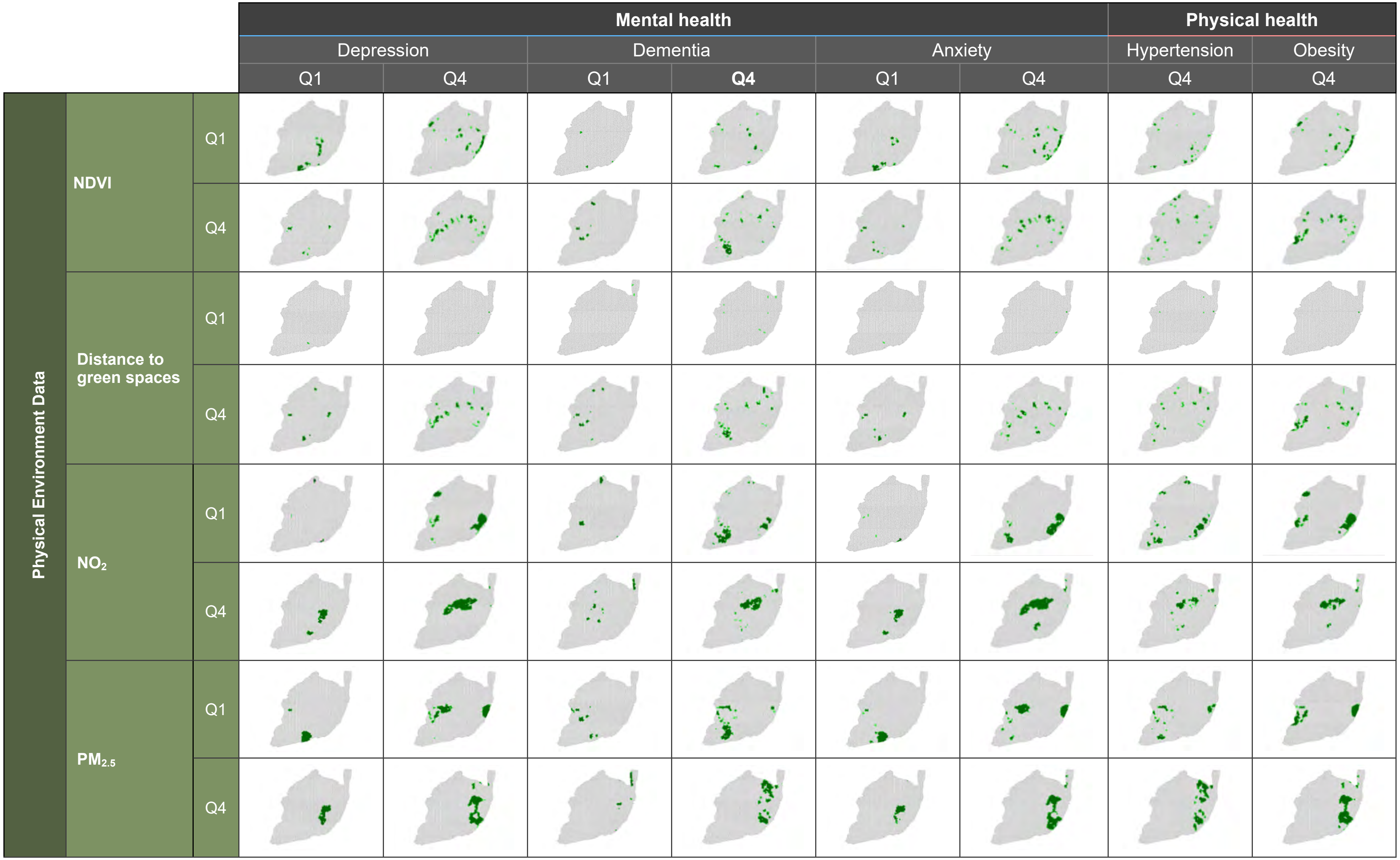
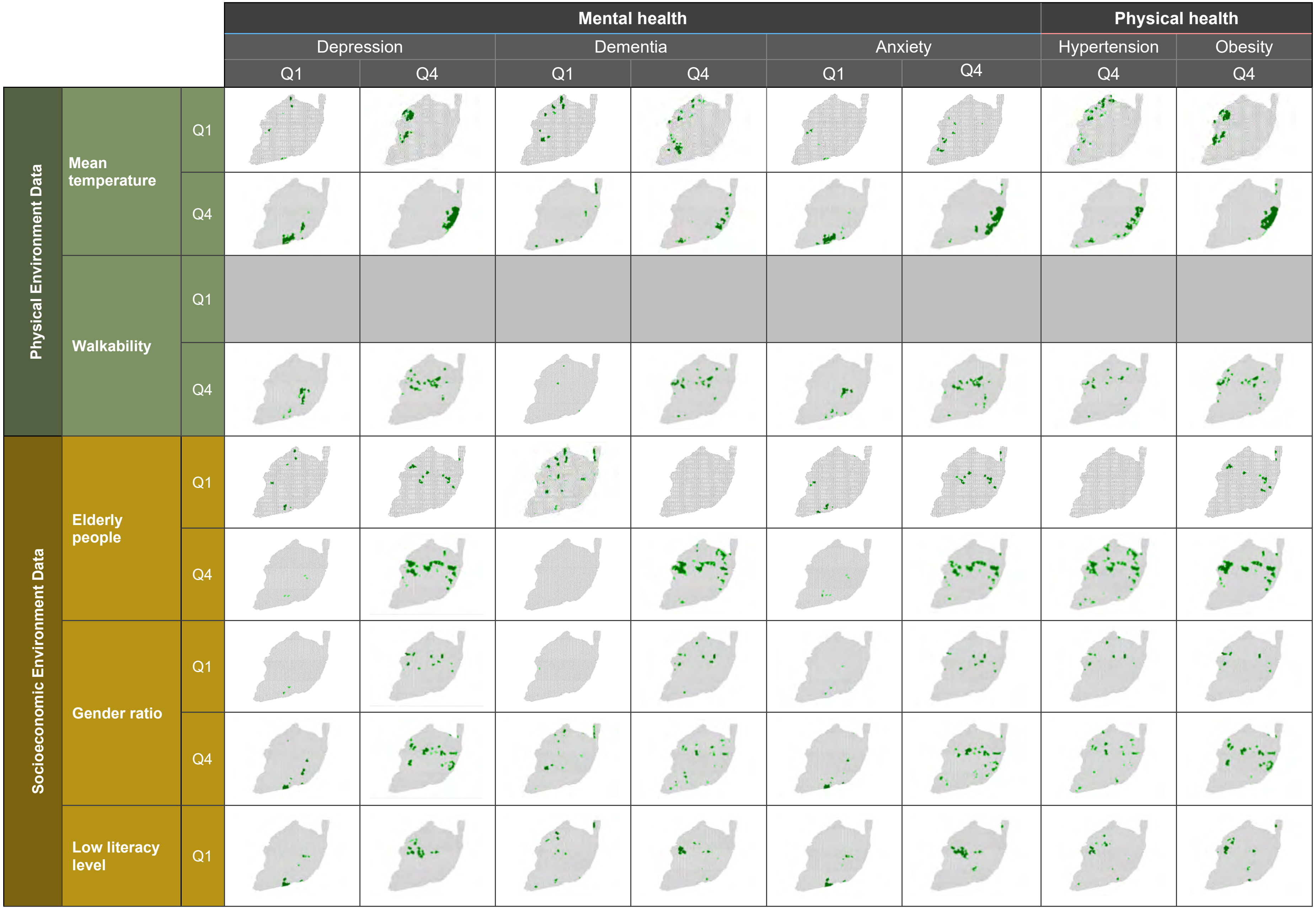


Figure A2.20. Prescribed dosages of anti-dementia drugs in Lisbon.

Appendix 3. Quantile LISA analysis for spatial analysis (in Lisbon)




Mental health
Physical health
Depression
Dementia
Anxiety
Hypertension
Obesity
Q1
Q4
Physical Environment Data
Mean temperature
Walkability
Socioeconomic Environment Data
Elderly people
Gender ratio
Low literacy level


Mental health
Physical health
Depression
Dementia
Anxiety
Hypertension
Obesity
Q1
Q4
Socioeconomic Environment Data
Low literacy level
Purchasing power
Population density
Perception data
Density of positive tweets

Appendix 4. Hotspots of health outcomes (in Lisbon)

Figure A4.1. High mental health risk associated with low NDVI in Lisbon.

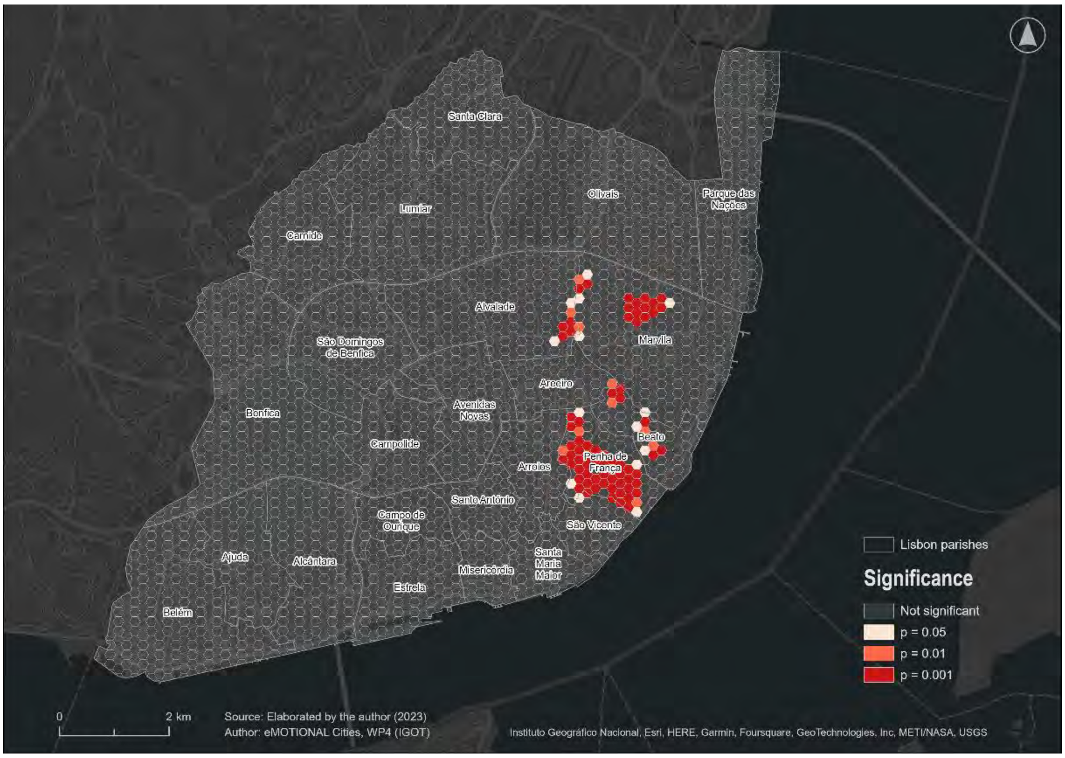


Figure A4.2. High mental health risk associated with high $PM_{2.5}$ in Lisbon.

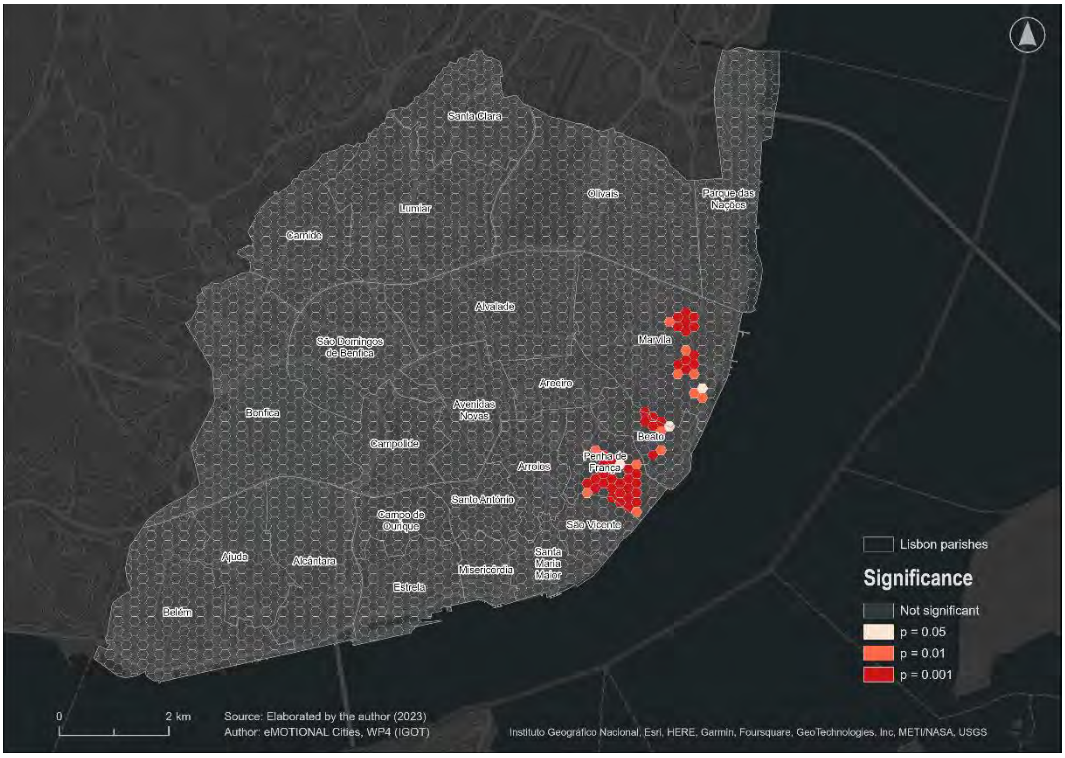


Figure A4.3. High mental health risk associated with high temperature in Lisbon.

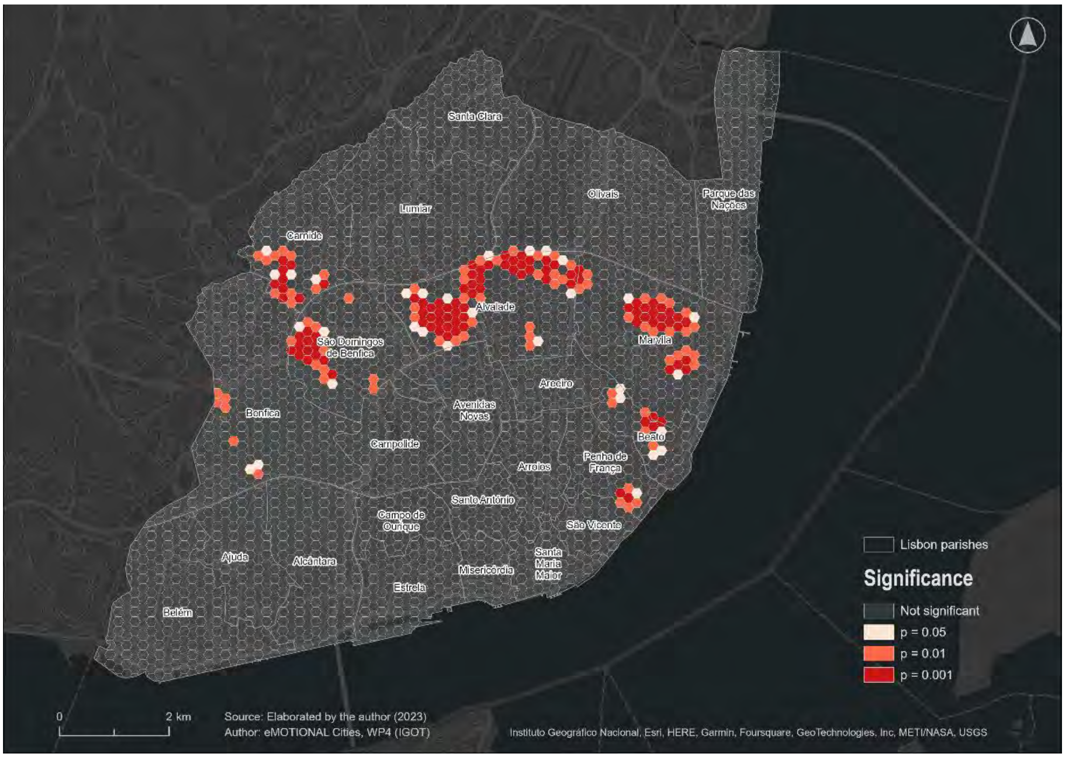


Figure A4.4. High mental health risk associated with high ratio of elderly people in Lisbon.

Figure A4.5. High mental health risk associated with low ratio of gender in Lisbon.

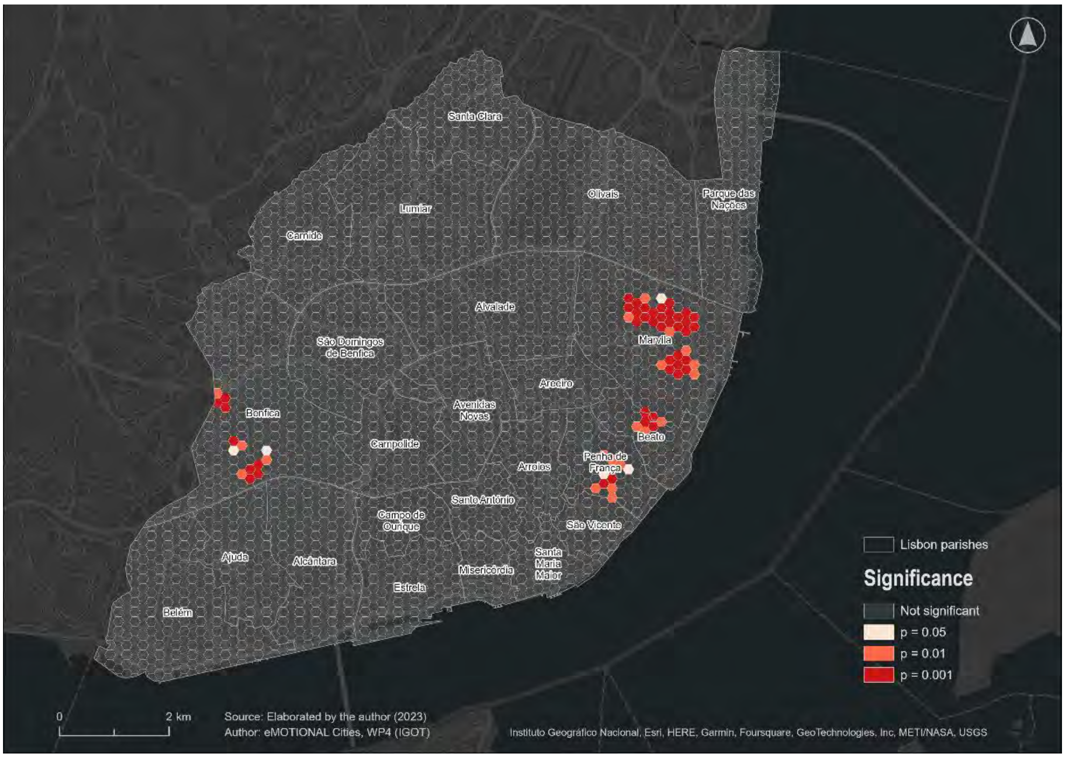


Figure A4.6. High mental health risk associated with low socioeconomic level in Lisbon.

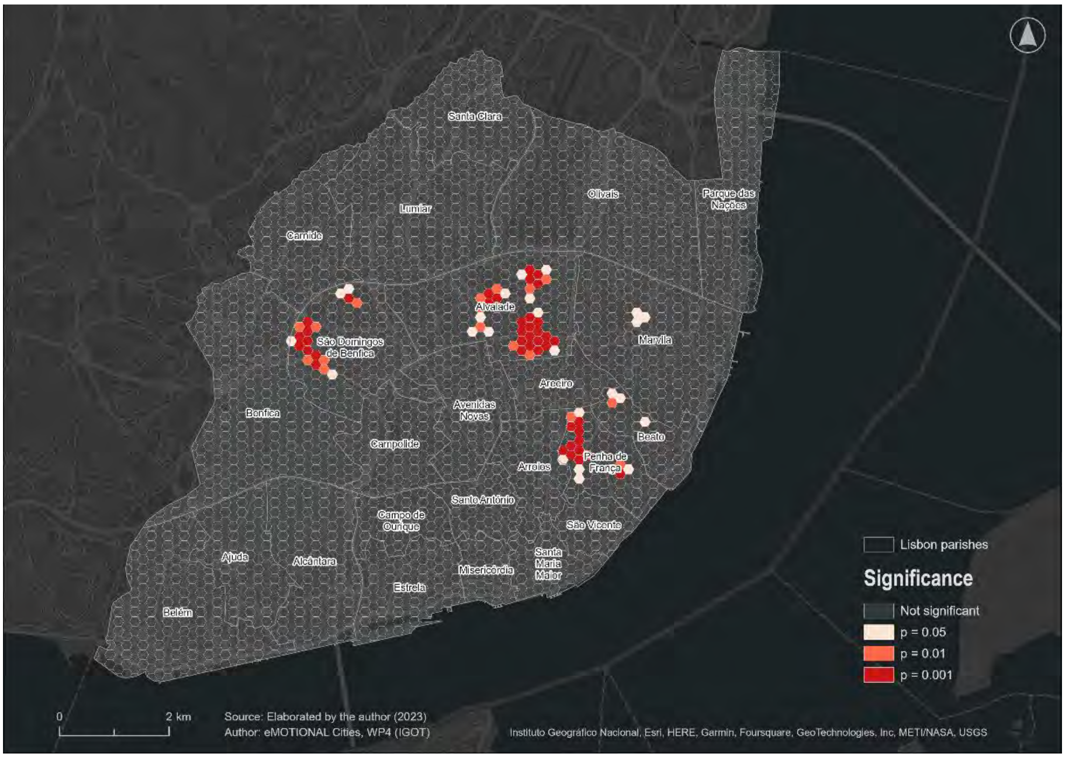


Figure A4.7. High mental health risk associated with high population density in Lisbon.

Figure A4.8. High mental health risk associated with low density of positive tweets in Lisbon.

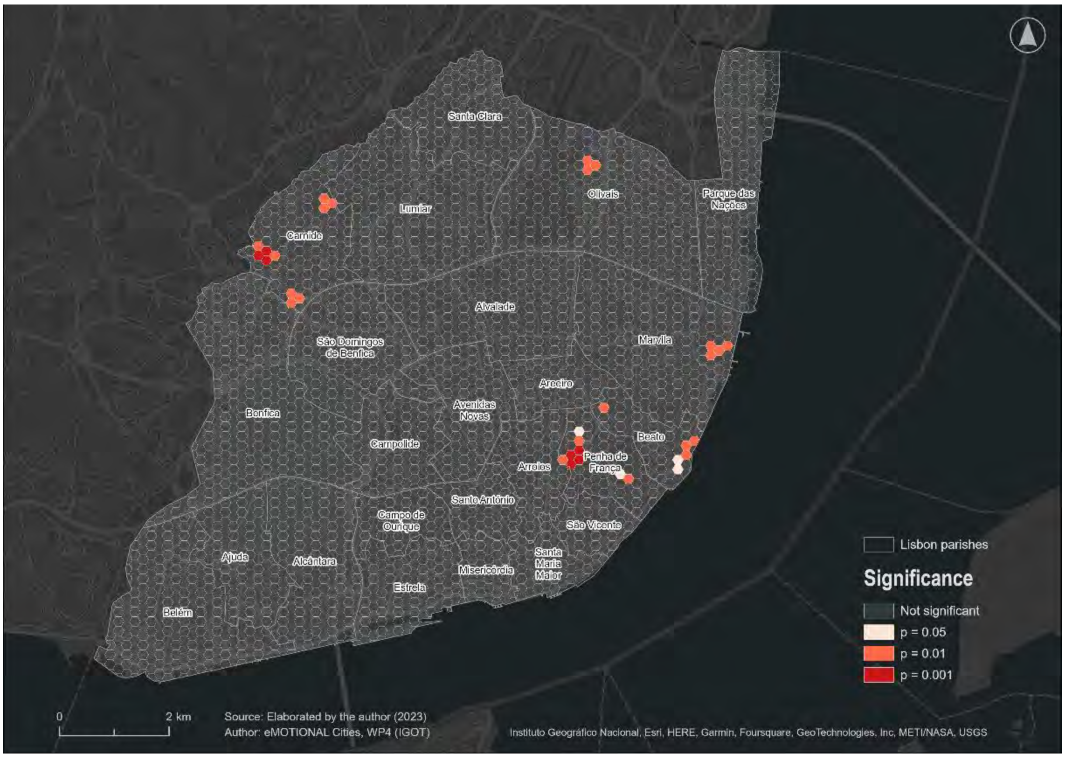


Figure A4.9. High physical health risk associated with low NDVI in Lisbon.

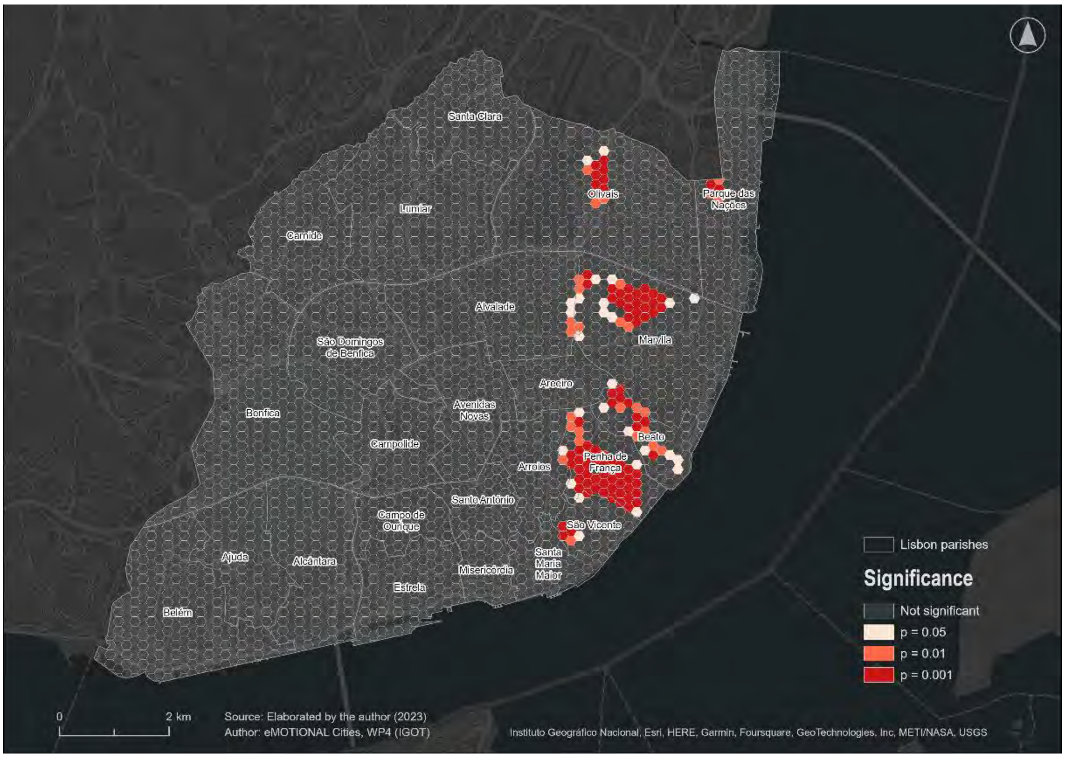


Figure A4.10. High physical health risk associated with high $PM_{2.5}$ in Lisbon.

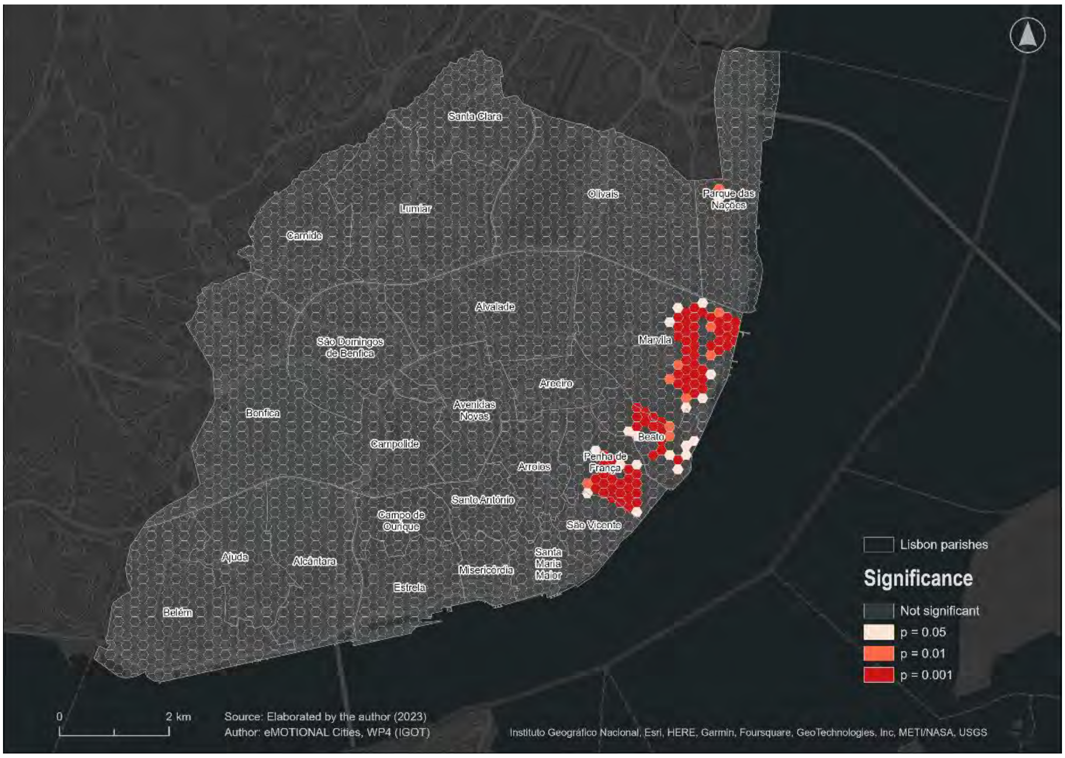


Figure A4.11. High physical health risk associated with high temperature in Lisbon.

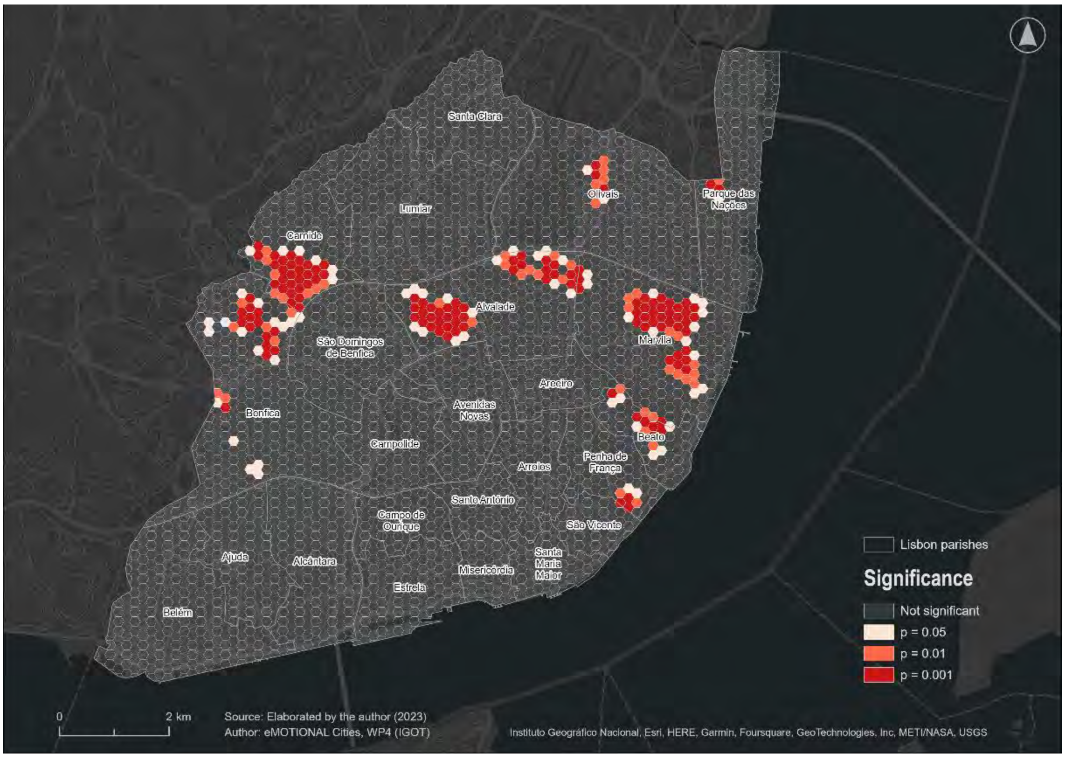


Figure A4.12. High physical health risk associated with high ratio of elderly people in Lisbon.

Figure A4.13. High physical health risk associated with low ratio of gender in Lisbon.

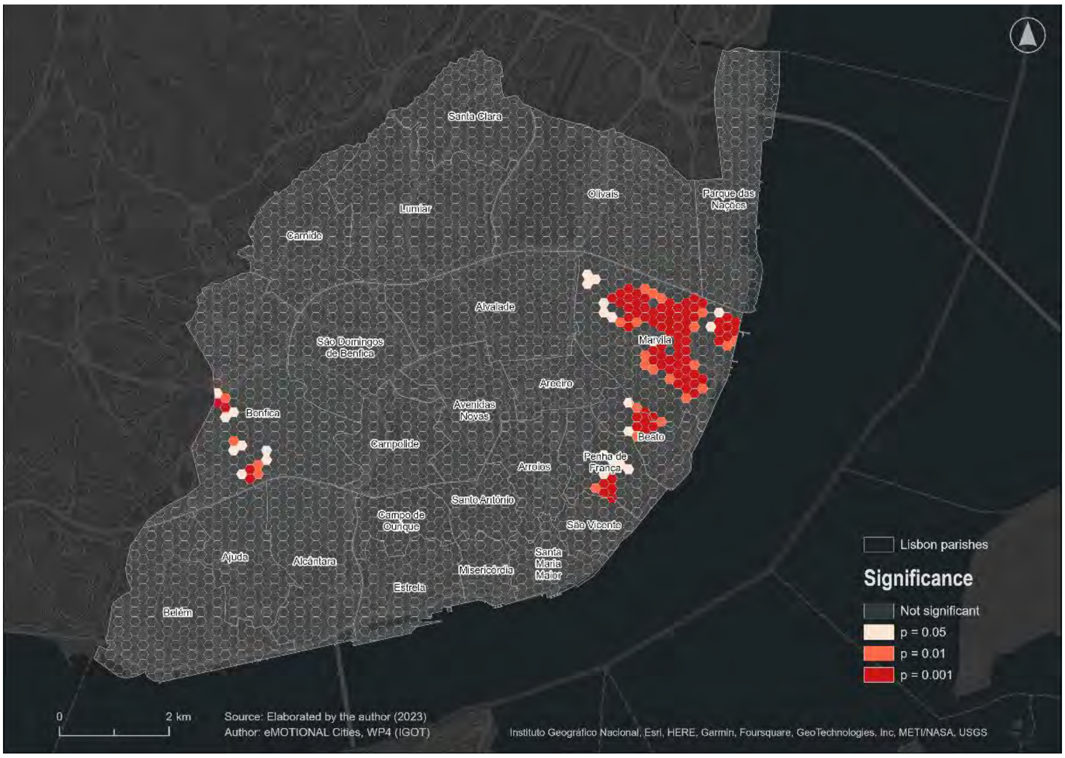


Figure A4.14. High physical health risk associated with low socioeconomic level in Lisbon.

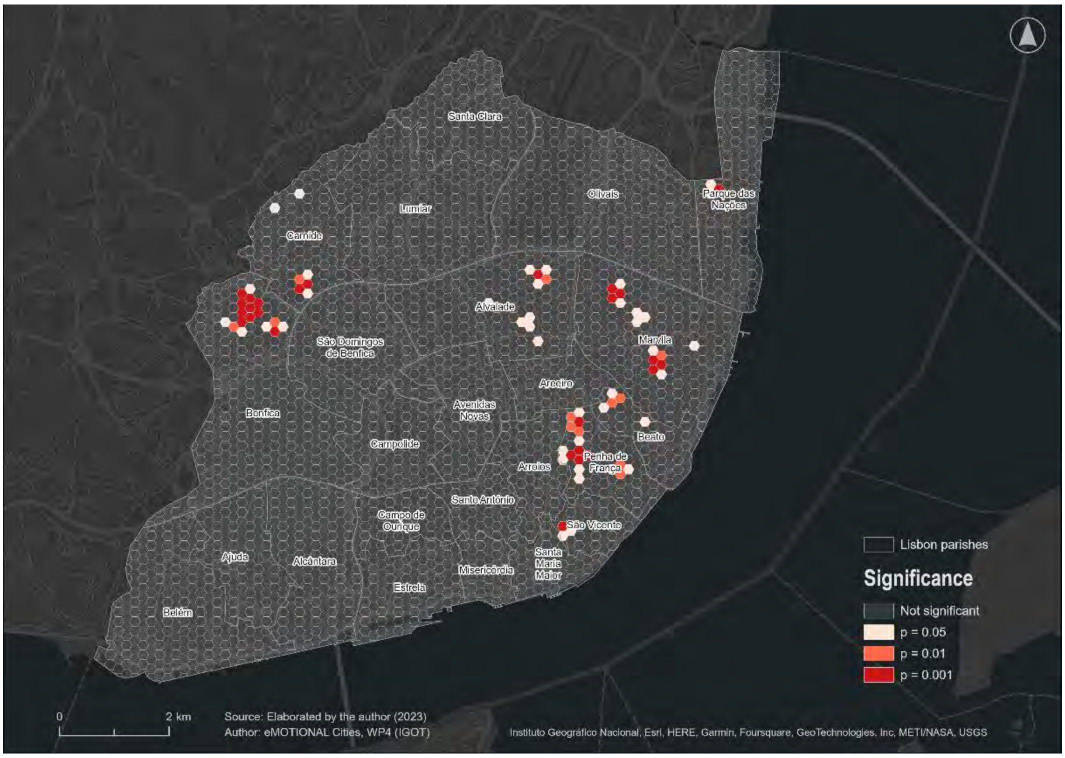


Figure A4.15. High physical health risk associated with high population density in Lisbon.

Figure A4.16. High physical health risk associated with low density of positive tweets in Lisbon.

Figure A4.17. High mental and physical health risk associated with low NDVI in Lisbon.

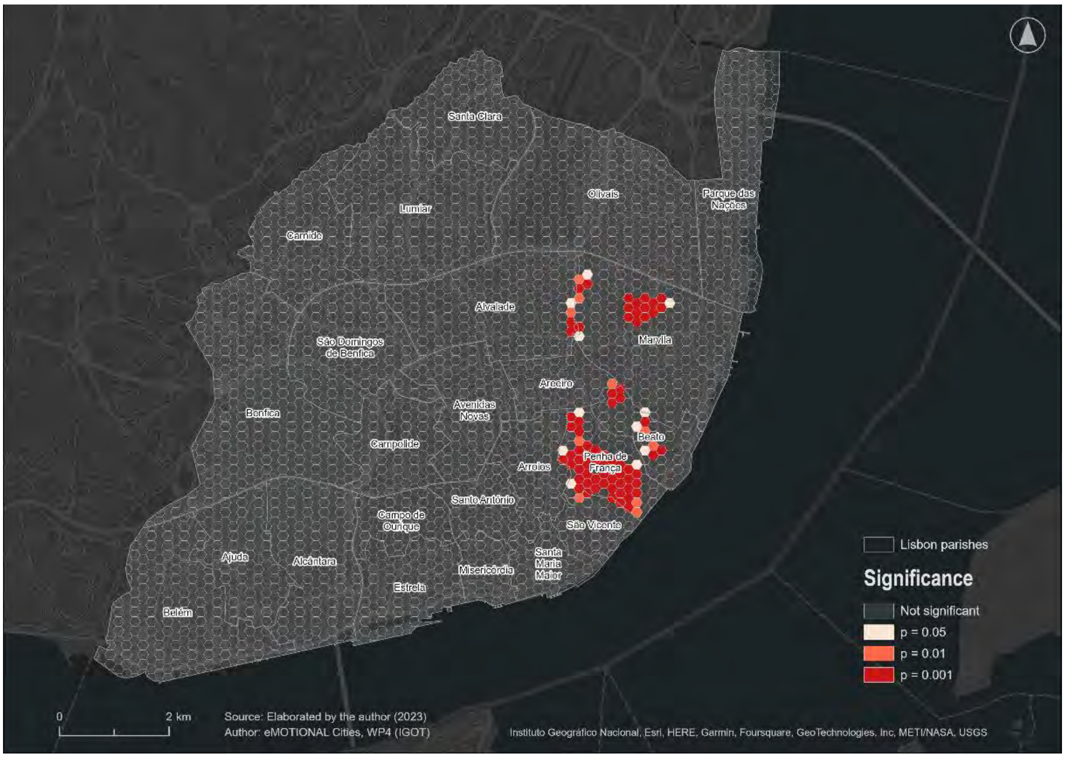


Figure A4.18. High mental and physical health risk associated with high $PM_{2.5}$ in Lisbon.

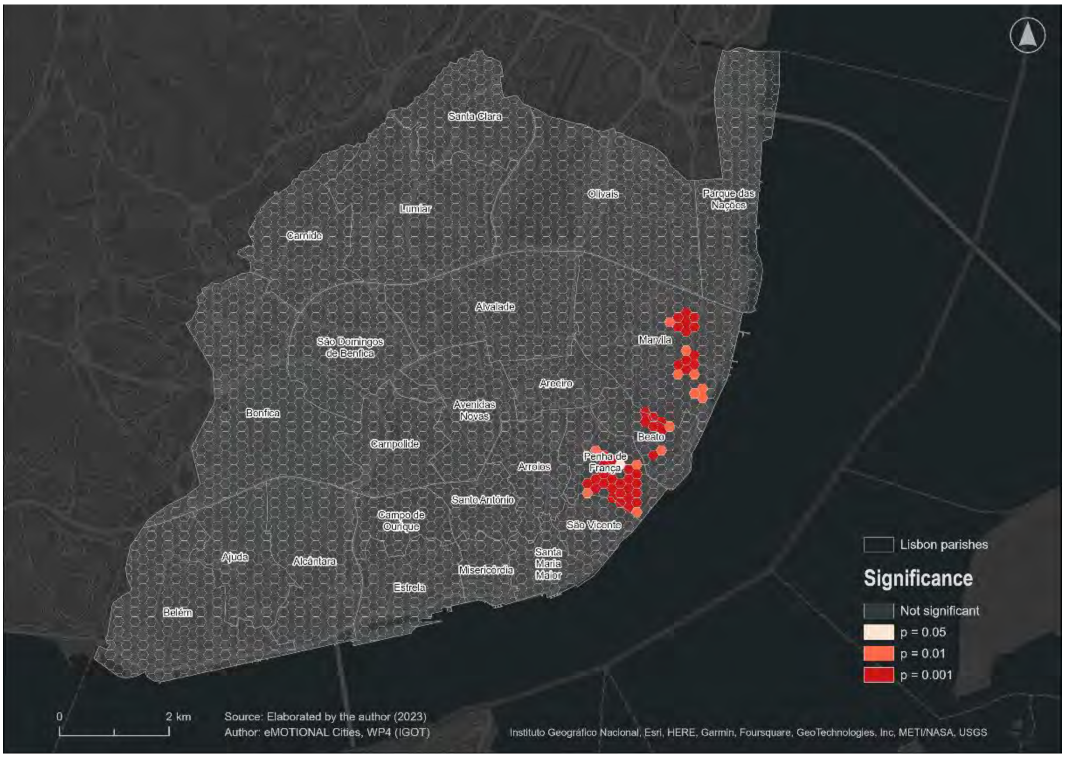


Figure A4.19. High mental and physical health risk associated with high temperature in Lisbon.

Figure A4.20. High mental and physical health risk associated with low ratio of gender in Lisbon.

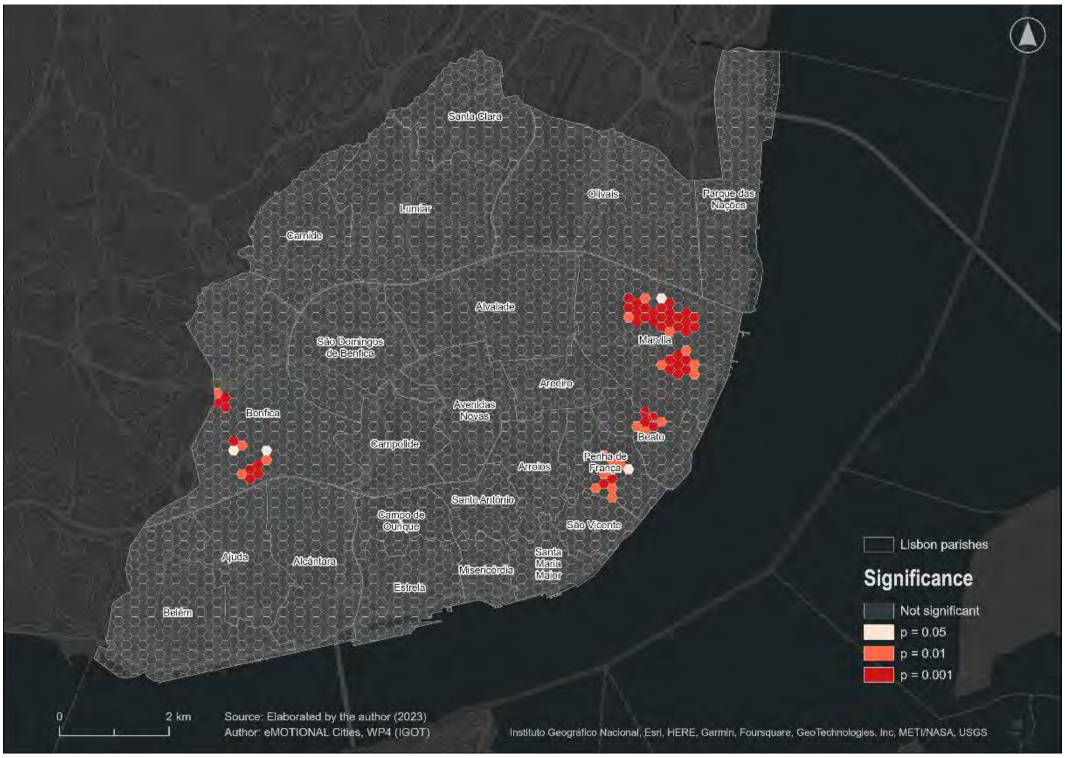


Figure A4.21. High mental and physical health risk associated with low socioeconomic level in Lisbon.

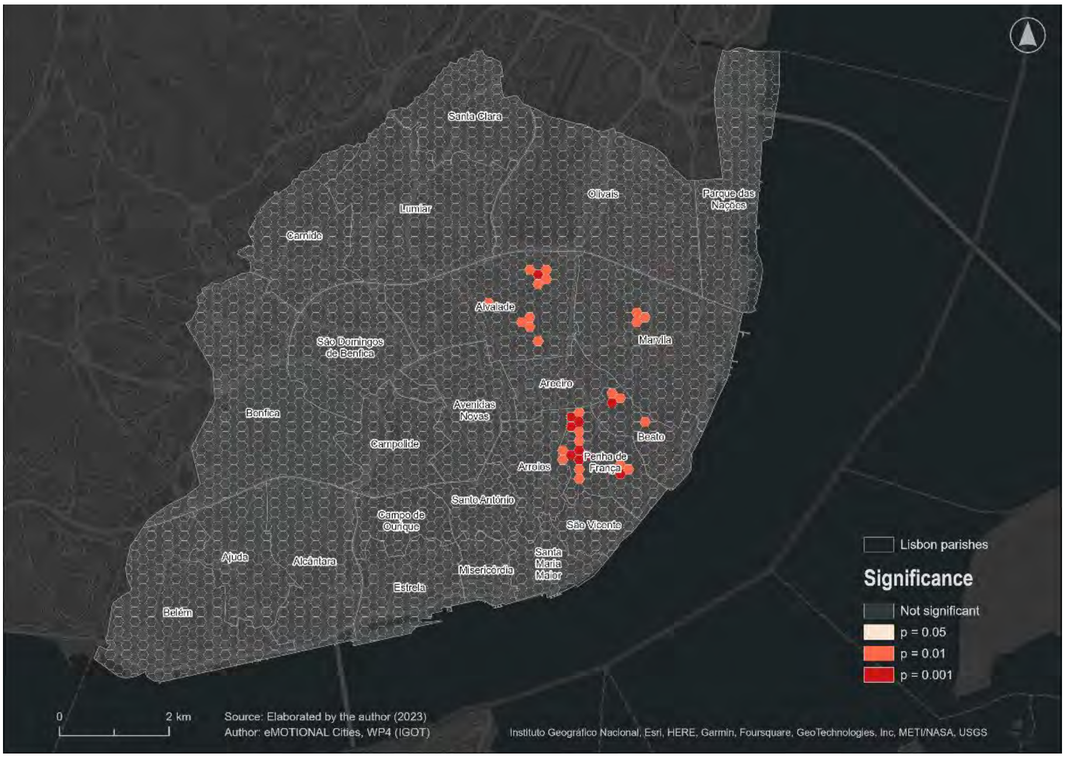


Figure A4.22. High mental and physical health risk associated with high population density in Lisbon.

Figure A4.23. High mental and physical health risk associated with low density of positive tweets in Lisbon.