Abstract

DINA — “Data-driven Implicit Needs Analyzer " is a Minimum Viable Product supporting Smart Cities, which aims at displaying on an heatmap a set of indexes (security needs, lighting needs, connectivity needs) based on implicit citizen & tourist needs, inferred from Public Administration Open Data.

Citizens can play a crucial role in identifying or actively intervening in urban challenges, often providing new perspectives and solutions. Collaboration between citizens and Public Administration is an area where several “Smart Cities” actors are investing (fully in line with the United Nations Sustainable Development Goals). When it comes to Urban Planning, Public Lighting or Tourism Development, it is quite common to get feedback and needs explicitly from citizens and visitors via dedicated studies, portals or apps. We think that we can complement this understanding of citizens and visitors by inferring their needs based on Public Administration Open Data. By looking at accidents logs, Public Wi-fi usage, the location and capacity of touristic accommodations and claims/complaints (reports), we can deduct areas where people’s life could be improved.

We have built a Minimum Viable Product which is leveraging Open Data produced by the city of Rome, Italy from January 2019 to September 2020. Thanks to 5 datasets (71Mo / 350 000+ rows of data processed), we have been able to demonstrate that we can spot areas in Rome where there are higher needs over time in terms of security enforcement via cameras, public Internet connectivity and public lighting.

Keywords: *City Analytics, Smart Public Lighting, Residents, Tourists, Security, Comfort, Decision Support, Italy*

Contents

Solution: DINA — Data-driven Implicit Needs Analyzer

Introduction

This section summarizes our value proposition (targeted users, problems we are trying to solve, solution, benefits), the rationale and guiding principles. It is not a Business Model Canvas as such, but it will help the people interested in this project to quickly understand the business area that we are focusing on and the expected benefits for stakeholders. Details of the proposition will be documented in the following sections.

Targeted Users

Our **targeted users** are:

1. Chief Operating Officer and/or operations of the city's public lighting department
2. Analysts working in the Urban Planning department
3. Tourist Office officers (City's Destination Marketing Organizations)

Problems we are trying to solve

Citizens can play a crucial role in identifying or actively intervening in urban challenges, [i]often providing new perspectives and solutions.

The meaningful integration of citizens in urban governance processes is now valued more than ever.

Collaboration between citizens and Public Administration is an area where "Smart Cities" actors are investing (fully in line with the United Nations Sustainable Development Goals).

Some commercial solutions, like the Enel X's big data platforms, are designed to be able to integrate 3rd party data, like cities own systems or other innovative apps which enable a deeper involvement of citizens in the life of the city.

But gathering people's needs is not that straightforward

Today, a lot of data used by Smart City Analytics solutions relates to signals/facts (e.g. what a sensor managed to capture) or **explicit** citizens' needs (or issues) expressed by people via dedicated portals/apps.

When it comes to gathering citizens requirements, a city has:

- to implement, support and maintain additional tools/products
- to communicate about the tools accordingly
- to ensure adoption and keep users engaged over time
- to translate these explicit requirements into an action plan with concrete impacts on services (including public lighting, connectivity, security etc.)

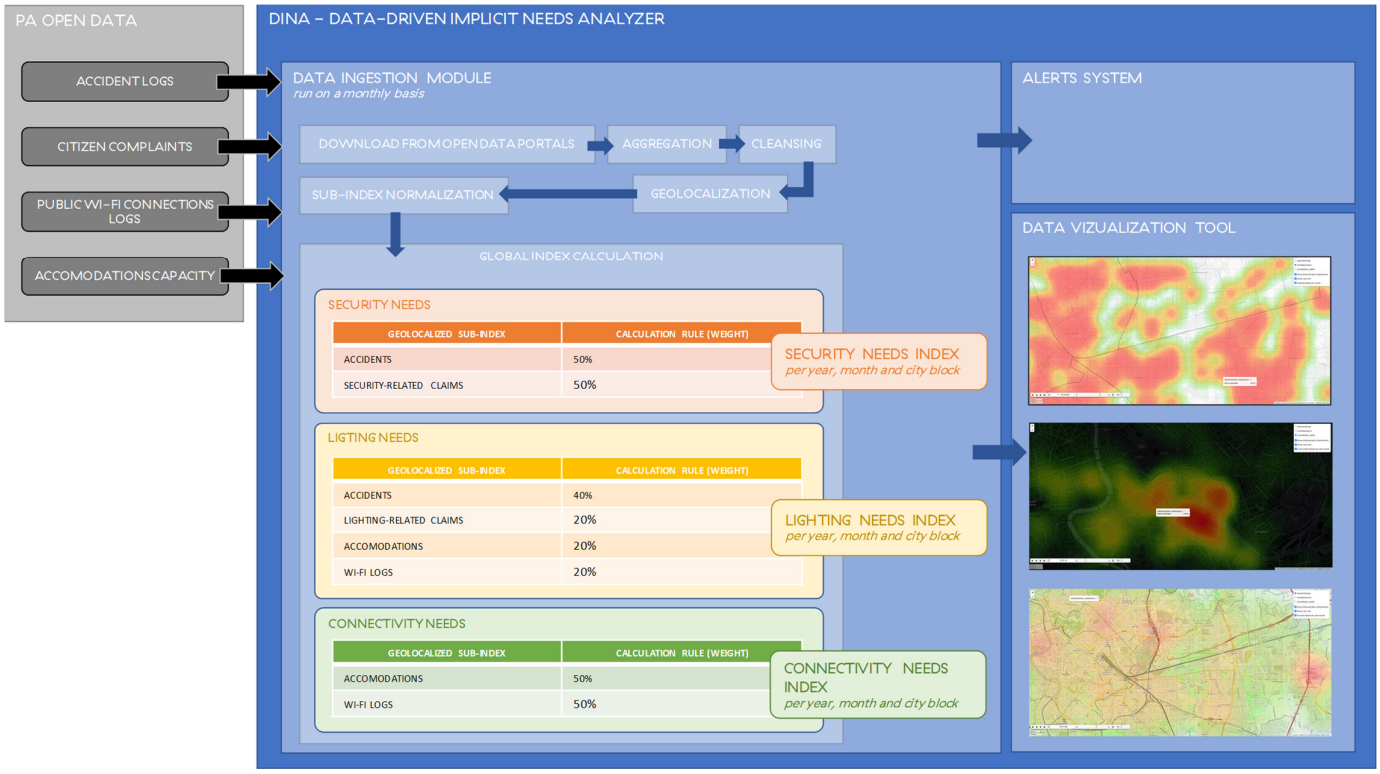
On top of that, based on our research (focused on the Italian market), most of the portals/apps which have been implemented to capture citizens' needs are in Italian, thus potentially refraining international tourists from expressing their own needs (or reporting issues) when visiting the country.

Capturing and assessing "explicit" needs is not enough

We make the assumption (to be validated with a living lab and focus groups) that citizens requirements (mainly residents and tourists) related to **security**, **connectivity** and **lighting** can be **implicit** and **inferred** from other signals. We want to complement traditional data sources used by the providers of "Smart Cities Analytics products", offer a solution to capture these "implicit needs" from Open Data and make them available to the end users via dedicated dashboards.

High level description of the solution

To answer the problems mentioned above, we designed and prototyped a modular solution called **DINA (Data-driven Implicit Needs Analyzer)**. It is made of 2 main components (DINA Open Data Ingestion Module and DINA Needs Index Visualization Tool) and can be complemented by optional components (like Alerts System).



High level view of the DINA components

1 – DINA Open Data ingestion module

This transversal component, supported by an extensible framework, gathers, cleans, enriches and normalizes Open Data (coming from Public Administrations or not) in order to provide **3 indexes** representing the needs of people living in or visiting the city:

- Security Needs

- Lighting Needs
- Connectivity Needs

These indexes will be delivered **per geographical location** (square of 300m x 300m in the proof of concept developed in the scope of this challenge) and per **month.**

Note: our vision is obviously to increase granularity (lower the size of the geographical areas and provide data per day or even per hour) when Public Administration Open Data will permit: this is why Smart Cities Solutions providers (the ones which could possibly leverage this work) have a key role to play to influence City's Open Data roadmaps.

2 — DINA "Needs Indexes" Visualization tool

The solution includes a data visualization tool so that users could act according to the insights generated by the first module: maps/heatmap (with the possibility to browse per time period – in our case: per *Month* or per *Year-Month* – and to filter by type of index).

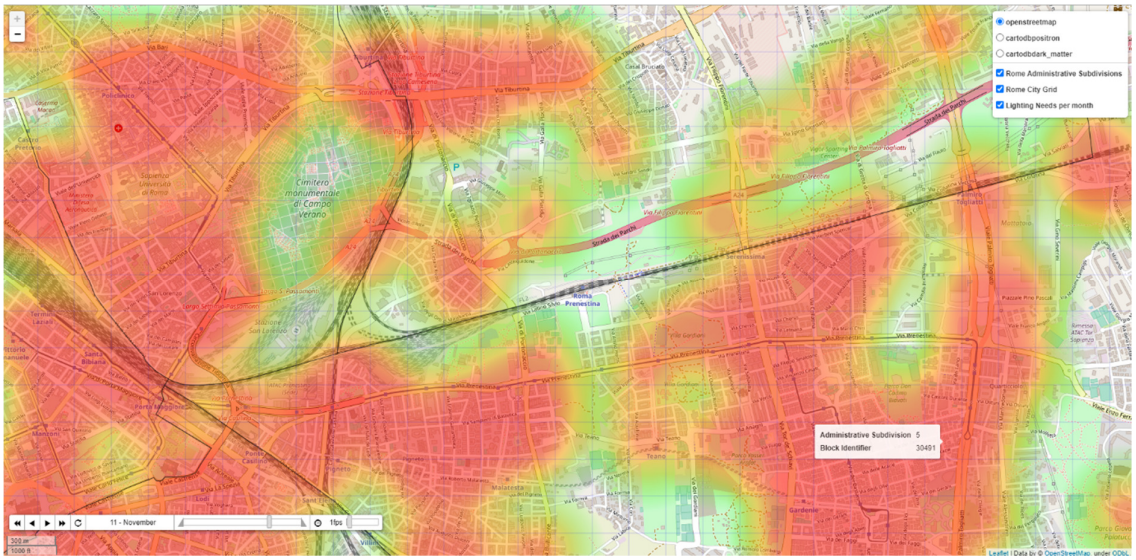
Which kinds of actions could be taken by DINA users, based on the insights?

Data visualization tool that we developed could answer this kind of questions (provided that it can also display areas where smart lamps, security cameras, wi-fi hotspots are installed — not implemented in the scope of this project) :

- *What are the city areas where we should invest more in smart lighting equipment?*
- *Do lighting needs evolve over time (depending on the month of the year) in some areas?*
  > If yes, I could adapt my smart lighting parameters to extend lighting time in such areas and possibly install new smart lamps
- *Are there areas in my country where security is a concern?*
  > If yes, I could determine the areas where I should install more security cameras
- *What are the areas where tourists may need a better wi-fi connectivity and where should I install more public Wi-fi hotspots?*

Below are 2 examples of heatmaps which has been generated by our Proof of Concept application, based on Rome's Open Data:

DINA Heatmap: Inferred 'Public Internet Connectivity Needs' in October 2019 in the Roma Municipio I (red means higher need)



DINA Heatmap: Inferred 'Improved Public Lighting Needs' in November (based on 2019-2020 historical data) in the Roma Municipio V (red means higher need)

Value Proposition

While the underlying technical solution of DINA is standard, the value proposition (formalized below using the *Geoff Moore's Value Positioning Statement*[ii]) delivered to the users might differ depending on their profile (see below). Main differences will be: the perceived benefits and channel to deliver the value proposition.

⊙1: Our Value Proposition for Chief Operating Officer and/or operations of the city's public lighting department

FOR Operators of the city's public lighting department and/or Chief Operations Officer

WILLING TO maintain and operate city's streetlight system while ensuring that the service fits with current and future needs of citizens and visitors

WE OFFER "DINA", a visual tool which displays on a map a set of indexes (security need index, lighting need index, connectivity need index) based on implicit needs assessment coming from various sources of data

SO THAT they could assess with a high level of granularity the city areas having special needs over time, adapt the lighting parameters and plan for future investments (like upgrading of connected lamps with cameras or wi-fi hotspots)

UNLIKE Other solutions leveraging explicit citizen needs or live video streams, we offer the possibility to capture implicit requirements, theoretically covering more population (including tourists) and allowing to predict what the needs will be in the coming months

Our Value Proposition for analysts in the Urban Planning department

FOR analysts working in the city's urban planning department

WILLING TO plan and manage city services in a most effective way and that can improve citizens quality of life, while ensuring that all citizen's needs are really considered

WE OFFER "DINA", a visual tool which displays on a map a set of indexes (security need index, lighting need index, connectivity need index) based on implicit needs assessment coming from various sources of data

SO THAT they could assess with a high level of granularity the city areas having special needs, plan for future investments and better target the zones for which deeper field analysis is requested (via on-the-field surveys for instance) while ensuring that a higher percentage of citizens is covered (and not only the ones who expressed explicitly their needs via dedicated channels)

UNLIKE Other solutions leveraging explicit citizen needs, we offer the possibility to capture implicit requirements, theoretically covering more population (including tourists) and allowing to predict what the needs will be in the coming months

Our Value Proposition for Tourist Offices (or DMOs)

FOR officers working in the city's tourist office or destination marketing organization

WILLING TO understand the needs of tourists and influence the city's development plans accordingly to provide a better experience to visitors

WE OFFER "DINA", a visual tool which displays on a map a set of indexes (security need index, lighting need index, connectivity need index) based on implicit needs assessment coming from various sources of data

SO THAT they could assess with a high level of granularity the city areas where tourists could have special needs (like charging stations, wi-fi connectivity), adapt and subsequently provide a better experience, thus attracting more visitors and positioning the city as the "best place to visit"

UNLIKE Other solutions leveraging explicit citizen needs, we offer the possibility to capture implicit requirements from tourists whose opinion is not always captured in an efficient way by traditional apps usually targeting residents (and not adapted to visitors, e.g. language)

The solution

Guiding principles and rationale for DINA

DINA (Data-driven Implicit Needs Analyzer):

1. Mainly focuses on **improving the life of citizens** (residents) and **tourists**
   Why?
   While our solution could also have a positive impact on a smarter use of energy, it's not our priority and our main goal is to **improve comfort and security**, to help cities to differentiate (e.g. for tourism, especially following the COVID-19 crisis which hit the tourism industry) and to focus on the United Nations SDG Goal 11.

2. Can be mainly considered as a **decision support tool** for cities, but also as a **business development tool** for "Smart Cities Solutions providers"
   Why
   We had a look at Open Data portals in the US, Italy, Spain and France and despite the obvious willingness of public administrations to publish more data: quality is quite low, update frequency can be improved and the lack of standardization (format, process) makes it difficult and even risky to use in an automation context (automate decisions based on Open Data), so we choose to focus on decision support, where human beings will always take the decision to act or not.

   By building such a dashboard (security, connectivity, lighting needs index), we also assume that Business Development teams of "Smart Cities Solution providers" could also leverage for internal purposes (to be validated of course), i.e. spot opportunities to deploy more smart lamps/smart urban furnitures or to upgrade the existing ones with innovative options.

3. Is based on Open Data which can be used to generate insights about implicit citizens' needs for: security, lighting and connectivity per location and period
   Why

Explicit citizens needs is already tackled by apps. We wanted to focus our effort on *implicit* needs, i.e. how to find a way to transform raw data into "requirements", in other words: interpret facts and their influence on citizens' needs.

The solution in detail

Summary of what has been implemented in the scope of this project

The basic principles of the DINA (Data-driven Implicit Needs Analyzer) solution have been implemented in a Proof of Concept. The later can be considered as a Minimum Viable Product (MVP) as it is using real data and producing real dashboards: the basic set of features has been implemented and is theoretically delivering its promised value (this value has however to be confirmed with a real customer, in our case: city of Rome, Italy).

Project has been open sourced and people are free to contribute (see section "Resources to access the Proof of Concept", screenshot of the GitHub repository and readme.md in the annexes).

For this RTP challenge, scope was:

- City: **Rome, Italy**
  > Rationale: Rome has an Open Data portal (https://dati.comune.roma.it) with a significant number of datasets compared to other cities in Italy

- Time period: **from January 2019 to September 2020**
  > Rationale: some relatively new open datasets have been produced and shared via the Open Data portal in 2020, not before. Then we considered that 2 years of historical data were good enough in the scope of this project

- Open Datasets: **5 datasets coming from Public Administration's Open Data which have been used to generate 3 indexes (total of 71Mo / 350 000+ rows of data processed)**
  1. Shapes of Administrative Subdivisions (Roma Municipi) coming from the Roma Urbanistica[iii] portal (used in our geolocalization process, see later)
  2. List of accommodation facilities in Rome Capital in 2019-2020[iv]
  3. List of road accidents that occurred in the territory of Roma Capitale[v] in the year 2019-2020
  4. Anonymized web browsing sessions found in the Rome WiFi system[vi]
  5. Single Reporting System of Roma Capitale. Reporting data for the year 2020[vii]
     The rationale for each dataset will be explained later in this document

General comments regarding the open data sources

Public Administration Open Data come with some challenges for an industrial usage. In the scope of this project we discovered several issues with city-related Open Data in general, and with Rome Open Data portal in particular:

1. Not a lot of datasets can provide data at city level in the *governments data portals*, which limits the usage of portals like dati.gov.it for Smart Cities use

cases, forcing us to leverage city's Open Data portals

2. Some datasets — as they are today – have different formats from one month to another which makes automated treatment difficult (data cleansing is key)
    a. Number and name of columns differ
    b. Number, Dates are using different formats (decimal separator etc.)

3. Some datasets are missing (not all months are covered for monthly/daily datasets, for no obvious reasons)

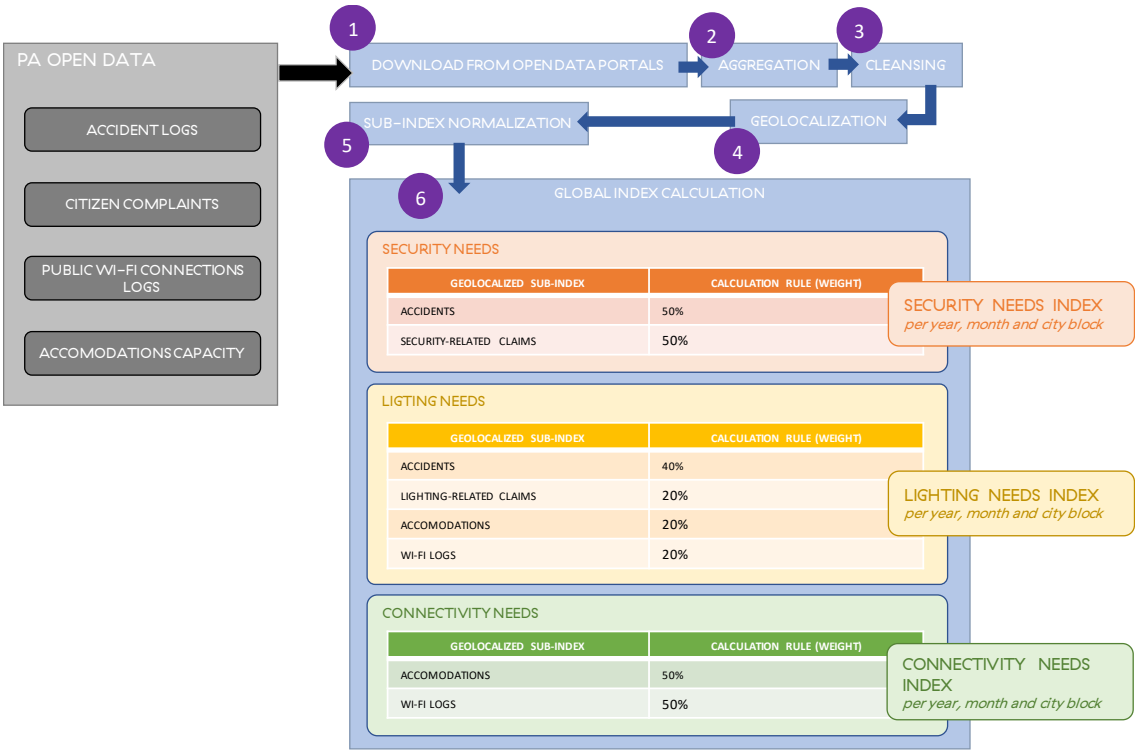
4. URLs to download datasets (at least in the Roma Open Data portal) do not have a common pattern and standardized format (like *<prefix>/DATA_YEAR_MONTH.csv*) which does not help data consumers to automate downloads and updates

5. Some datasets are in Italian, which is not a big issue as such but limits the adoption by non-Italian developers (low adoption implies low incentive for city's IT teams to improve the quality and quantity of data)

6. Datasets do not have the same granularity
    a. Location can be: Administrative Subdivision (Roma Municipi) > Street > Point of Interest > Latitude / Longitude
    b. Time can be: Year-Month > Date > Date and Time

One of the main challenges for people/organizations willing to leverage this project is to show the concrete value of Open Data in order to engage with the city's IT department in a data cleaning and industrialization exercise.

Technical description of the approach and Algorithms: focus on the data ingestion module



There is a 6 steps process in the DINA Data Ingestion Module

1) Download from Open Data Portals

We used as inputs some public datasets which contained enough data to be relevant and which could be found (using this format or another) in other cities' Open Data portal (to ensure that the solution is global and scalable).

Accidents logs

https://dati.comune.roma.it/catalog/dataset/d852

Description

List of road accidents that occurred in the territory of Roma Capitale[viii] in the year 2019-2020. The dataset contains all the road accidents in which a patrol of any Group of the Roma Capitale Local Police intervened. Therefore, only incidents in which the parties involved have reached a conciliation are excluded. The dataset does not include the accidents that occurred on the Grande Raccordo Anulare of Roma Capitale.

Rationale for using this dataset

We want to leverage this dataset to contribute to the Lighting Needs and Security Needs indexes.

In the dataset, for each row (i.e. a reported accident) there is a column called '*Illuminazione*' whose value equals '*Insufficiente*' when lighting has been considered by locale police as

not adapted. This information is precious to spot areas where a better lighting could have made the life of citizens better.

Note: in the future we could split this dataset into 2 datasets:

- Accidents caused by bad lighting (highly contributing to the lighting needs index)
- Accidents not caused by bad lighting (still relevant for the security needs index)

Anonymized web browsing sessions found in the Rome Wi-Fi system Jan-Oct 2020[ix]

https://dati.comune.roma.it/catalog/dataset/wifi2020

Description

The dataset describes the anonymized distribution of DigitRoma Wi-fi users, who carry out at the offices enabled for the Institutional Wi-Fi service: the data traffic in upload-download with the duration of the session, the date and the place of business. The information collected refers to the number of anonymized user sessions recorded on a daily basis.

Rationale for using this dataset

We want to leverage this dataset to contribute to the Lighting Needs and Connectivity Needs indexes.

What is the link between public Wi-fi and Lighting needs? Tourists visiting a country are more likely to visit the city by night than most of the citizens (or at least go back to their hotels later, thus needed extended public lighting times). City-wide wireless networks are known as being used by tourists who don't necessarily have a roaming plan. We use this dataset to spot areas where tourists have a higher probability to be located.

In other words: more public Wi-fi sessions => more tourists => more lighting needs.

Note: We do not filter the "language" used for the captive Wi-fi portal (Italian vs other language) nor the session start time (day time vs night time) but this could be considered as a further improvement to make the data even more relevant.

List of accommodation facilities in Rome Capital in 2019-2020[x]

https://dati.comune.roma.it/catalog/dataset/d865

Description

Exhaustive list of Hotels, B&B, etc. per year and month with the number of single/double/triple/... rooms

Rationale for using this dataset

We want to leverage this dataset to contribute to the Lighting Needs and Connectivity Needs indexes.

As mentioned before, presence of tourists in specific areas of the city may lead to bigger needs in terms of connectivity or lighting.

The maximum capacity of accommodations per area (number of rooms x capacity of the room) is a good proxy to determine areas where tourists are staying.

Single Reporting System of Roma Capitale. Reporting data for the year 2020[xi]

https://dati.comune.roma.it/catalog/dataset/sus1

Description

The Dataset aims at representing a framework for managing the flow of reports from citizens through the SUS (Single Reporting System https://www.comune.roma.it/web/it/di-la-tua-segnala.page). The reports present are geo-localized by municipality and by subject.

Rationale for using this dataset

We want to leverage this dataset to contribute to the Lighting Needs and Security Needs indexes. From this dataset we extract 2 sub-indexes per Year-Month and administrative subdivision:

- Lighting Claims (when people are complaining about public lighting issues)
- Security Claims (when people are complaining about various topics for which a security camera could help)

We are leveraging the content of a column called 'Argomento – codice' to spot reports worth taking into account in one of the 2 indexes:

Label	Argomento - codice	Descrizione area tematica	Sub-index
! SOSTA SELVAGGIA - RIMOZIONE VEICOLI IN INTRALCIO	296	MOBILITÀ E TRASPORTI	security_claim
! SOSTA SU ATTRAVERSAMENTO PEDONALE	298	SICUREZZA URBANA E PROTEZIONE CIVILE	security_claim
RIFIUTI ABBANDONATI	258	OPERE E MANUTENZIONE DELLA CITTÀ	security_claim
RIFIUTI ABBANDONATI - PERICOLO PER L'INCOLUMITA PUBBLICA	259	AMBIENTE	security_claim
SOSTA SU MARCIAPIEDE - GRAVE INTRALCIO	380	SICUREZZA URBANA E PROTEZIONE CIVILE	security_claim
ILLUMINAZIONE PUBBLICA - GUASTI	150	OPERE E MANUTENZIONE DELLA CITTÀ	lighting_claim
CASSONETTO INDIFFERENZIATO PIENO/RIBALTATO/DANNEGGIATO	48	OPERE E MANUTENZIONE DELLA CITTÀ	security_claim
SICUREZZA E PROTEZIONE CIVILE - PERICOLO PER L'INCOLUMITA PUBBLICA	288	SICUREZZA URBANA E PROTEZIONE CIVILE	security_claim
CASSONETTO CARTA PIENO/RIBALTATO/DANNEGGIATO	47	OPERE E MANUTENZIONE DELLA CITTÀ	security_claim
! SUONI/RUMORI/SCHIAMAZZI MOLESTI	319	SICUREZZA URBANA E PROTEZIONE CIVILE	security_claim
CASSONETTO PLASTICA PIENO/RIBALTATO/DANNEGGIATO	49	OPERE E MANUTENZIONE DELLA CITTÀ	security_claim
! SOSTA SU PASSO CARRABILE	300	SICUREZZA URBANA E PROTEZIONE CIVILE	security_claim
CASSONETTO SCARTI ORGANICI PIENO/RIBALTATO/DANNEGGIATO	50	OPERE E MANUTENZIONE DELLA CITTÀ	security_claim

2) Data Aggregation

For each dataset, we aggregate all the downloaded files to ease the processing.

3) Data Cleansing

For each aggregated dataset, we clean the data:

- Convert content to a format which will ease processing (e.g. date time, floats, integers)
- Remove useless columns (not used by DINA)
- Solve ad-hoc data quality issues (which depends on the dataset)

4) Geolocalization

First step is to generate a city grid which consists in splitting the city area (bottom-left to top-right coordinates) into 'city blocks' which correspond to areas of 300 meters x 300 meters by default.

Note: this value can be changed in our proof of concept but a smaller value will lead to a higher processing time.

Each block is tagged with the corresponding Administrative Subdivision (in our POC: Roma Municipi) and has a unique ID. For instance in the proof of concept here is the number of blocks created per Municipio:

Roma Municipio	**Number of 300mx300m city blocks**
1	406
2	390
3	1970
4	983
5	544
6	2283
7	913
8	940
9	3667
10	3011
11	1430
12	1464
13	1343
14	2673
15	3776
Grand Total	**25 793**

Blocks generated for Rome, Italy

Rationale for using blocks and not latitude/longitude:

Some datasets are using addresses, some are using points of interest, some are using latitude and longitude, some of them are using Municipio to attach a location. Aggerating per 300mx300m block is a good compromise between the need for granularity and the limitations of datasets.

Then, for each row in the dataset, we associate the corresponding block using different methods depending on the nature of the location data:

- Point of interest, address: we use a geolocation API (in our POC: Here! Freemium API) to get the Latitude/Longitude to retrieve the corresponding block ID
- Latitude/Longitude to retrieve the corresponding block ID
- Municipio: we retrieve all the blocks attached to the Municipio and assign the same value to all blocks

5) Sub-index normalization

For each data source such as accommodations or accidents, we create one or more tables containing a normalized sub-index per grid square (city block ID), per year and month.

To compute this sub-index, we extract from public datasets some values:

Dataset	Value per city block per year-month
Accidents logs	Number of accidents where lighting is considered as inadequate
Accommodations	Total capacity of all accommodations (number of rooms x rooms capacity, for each room type)
Public Wi-fi	Total amount of data downloaded during Internet sessions
Lighting Claims	Number of claims related to public lighting
Security Claims	Number of claims related to security or incivility

The sub-index is computed as the ratio between the value in the grid square in that month, divided by the largest value over the table, and normalized between 0 and 10. A value of 10 thus means that this city block and year-month have the maximum value of any grid square and year-month:

$$subindex_{jtk} = 10.\frac{value_{jtk}}{max_{m,l}(value_{jml})}$$

Where $subindex_{jk}$ is the value of sub-index corresponding to dataset j at the location k and year-month t and $value_{jtk}$ is the calculated value of dataset j (see table above) at the location k and year-month t (for example, the number of accidents in location k in march 2019).

6) Global Index Calculation

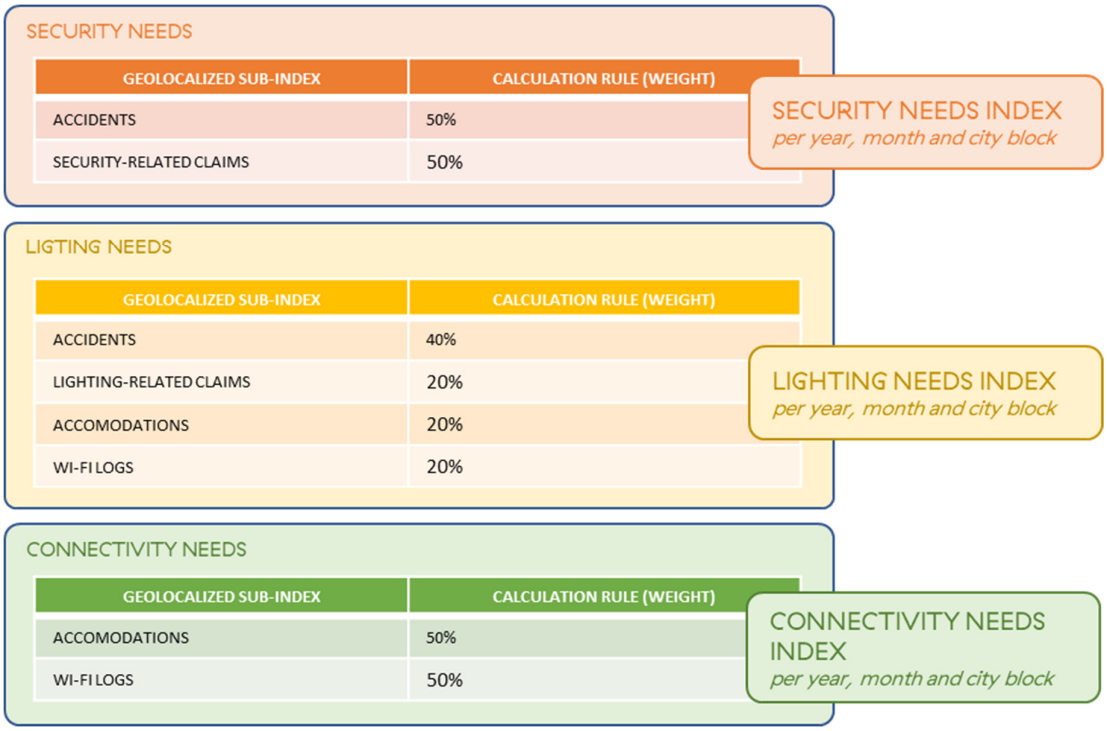
The Connectivity, Security and Lighting global indexes are computed from the tables generated for each sub-indexes. Each global index is defined as a weighted sum of different sub-indexes. The index number i is thus computed as follows:

$$Index_{itk} = \sum_{j \in Datasets_i} \alpha_{ij} \cdot value_{jtk}$$

Where $Index_{itk}$ is the value of the dataset I at the location k and month t, $Datasets_i$ is the list of datasets used to compute the index i, α_{ij} is the weight of dataset j for index i, and $value_{jk}$ is the value of dataset j (sub-index) at the location k and month t.

The weight can be considered as the contribution of the data (like number of accidents) to the inferred need.

In the scope of the proof of concept, we used the weights below to compute indexes but these values need to be determined by the customer (the city) based on an analysis of the Public Data, its quality and relevance.



For instance, we considered that the maximum capacity of accommodations has the same weight as the total duration of Wi-fi logs when it comes to "guessing" the needs for an improved Wi-fi connectivity in the city. If it is proven that the tourists are not using Public Wi-Fi that much for a given city, then the weight of 'Wi-fi logs' should be decreased.

Possible future improvements and known limitations

Known limitations

We acknowledge the fact that the contribution of the various datasets to the Connectivity, Security, Lighting needs indexes is subjective and the relevance of our hypothesis needs to be validated in the field with city officials.

Legitimate questions:

- *If someone is complaining about an issue with public lighting, does that mean that this person has more needs related to public lighting?*
- *Is public Wi-fi a good proxy to spot areas where tourists are located?*
- *If someone is reporting an issue related to security or incivility, does that mean that they would welcome the installation of a security camera?*

These indexes are an approximation and are mainly highlighting areas with a bigger need (relative comparison) and this is why we display the indexes using heatmaps (see at the end of this document).

Possible future improvements

Predict indexes

When displaying indexes per month, we are using historical data. A possible improvement could be to build predictive models (time series forecasting) to anticipate needs (i.e. predict needs indexes for the next months) but this implies the creation of one model per dataset (accidents, wifi, accommodations, lighting claims, security claims). We anticipate that building such models can be difficult knowing current data limitations (datasets size).

Demonstration

Resources to access the Proof of Concept

This MVP has been developed in Python 3.7.6. Its commented source code, input data, output data and dashboards are stored in a GitHub private repository which is available on demand (total size: 577Mo). We recommend to use an Anaconda distribution to browse the code, possibly update the parameters and run it.

Examples of what DINA can produce (output data and heatmaps) has been shared as a .zip file in Drobox and is accessible via this short URL: http://bit.ly/dinaoutput. This file contains:

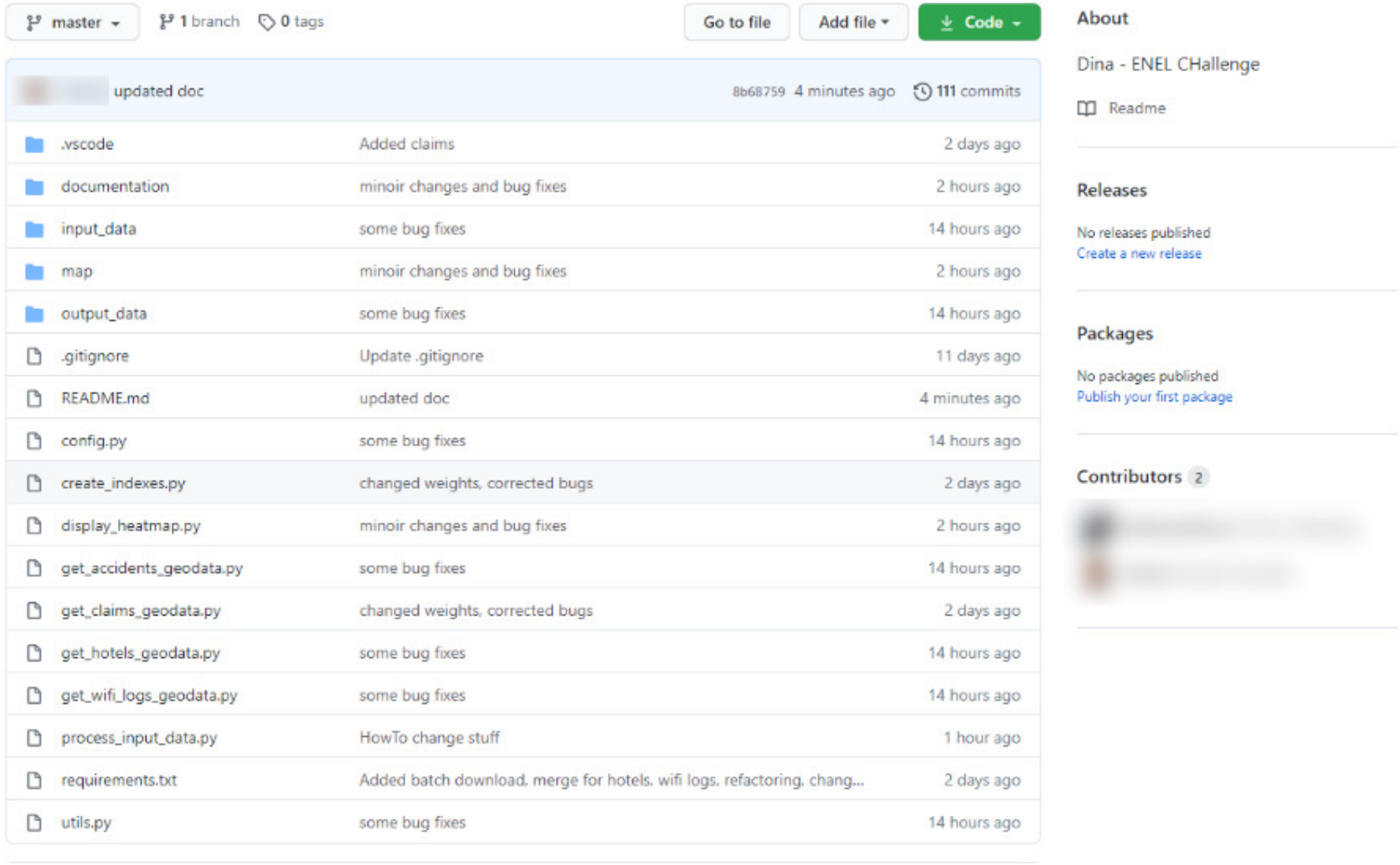
- Screenshots of the dashboard (including 3 animated gifs showing heatmaps variations over time)
- Html files which have been generated to display the heatmaps
    - Rome (Italy) – Connectivity Needs per month (html)
    - Rome (Italy) – Lighting Needs per month (html)
    - Rome (Italy) – Security Needs per month (html)
    - Rome (Italy) – Historical Connectivity Needs per year-month (html)
    - Rome (Italy) – Historical Lighting Needs per year-month (html)
    - Rome (Italy) – Historical Security Needs per year-month (html)
    - Rome (Italy) — List of Connectivity Needs Indexes (from 0 to 10) per month and city block (csv)
    - Rome (Italy) — List of Lighting Needs Indexes (from 0 to 10) per month and city block (csv)
    - Rome (Italy) — List of Security Needs Indexes (from 0 to 10) per month and city block (csv)

	A	B	C	D	E	F	G	H	I
1	block_ID	administr	Year	Month	accidents	hotels	wifi	lighting_claims	Index
2	25634	2	2019	1	0	0	0	0	0
3	25634	2	2019	2	0	0	0	0	0
4	25634	2	2019	3	0	0	0	0	0
5	25634	2	2019	4	0	0	0	0	0
6	25634	2	2019	6	0	0	0	0	0
7	25634	2	2019	7	0	0	0	0	0
8	25634	2	2019	8	0	0	0	0	0
9	25634	2	2019	9	0	0	0	0	0
10	25634	2	2019	11	0	0	0	0	0
11	25634	2	2020	1	0	0	0	8.75	1.75
12	25634	2	2020	2	0	0	0	1.875	0.375
13	28056	5	2019	1	0	0	0	0	0
14	28056	5	2019	2	0	0	0	0	0
15	28056	5	2019	3	0	0	0	0	0
16	28056	5	2019	5	0	0	0	0	0
17	28056	5	2019	7	0	0	0	0	0
18	28056	5	2019	8	0	0	0	0	0
19	28056	5	2019	10	0	0.027061	0	0	0.005412

Example of normalized indexes (Improved Lighting Needs Indexes) per city block ID, year, month with details about the corresponding administrative subdivision (Roma Municipio), and all sub-indexes which contributed to build the global index on the right (see calculation below)

Note: repository structure and readme.md files are provided at the end of this document

Content of the GitHub repository containing the commented code, input data, output files (data, html maps)



master · 1 branch · 0 tags · Go to file · Add file · Code

[redacted] updated doc — 8b68759 4 minutes ago — 111 commits

.vscode	Added claims	2 days ago
documentation	minoir changes and bug fixes	2 hours ago
input_data	some bug fixes	14 hours ago
map	minoir changes and bug fixes	2 hours ago
output_data	some bug fixes	14 hours ago
.gitignore	Update .gitignore	11 days ago
README.md	updated doc	4 minutes ago
config.py	some bug fixes	14 hours ago
create_indexes.py	changed weights, corrected bugs	2 days ago
display_heatmap.py	minoir changes and bug fixes	2 hours ago
get_accidents_geodata.py	some bug fixes	14 hours ago
get_claims_geodata.py	changed weights, corrected bugs	2 days ago
get_hotels_geodata.py	some bug fixes	14 hours ago
get_wifi_logs_geodata.py	some bug fixes	14 hours ago
process_input_data.py	HowTo change stuff	1 hour ago
requirements.txt	Added batch download, merge for hotels, wifi logs, refactoring, chang...	2 days ago
utils.py	some bug fixes	14 hours ago

About

Dina - ENEL CHallenge

Readme

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Contributors 2

Extract of the README.md in the GitHub repository

Getting Started

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes.
See deployment for notes on how to deploy the project on a live system.

Prerequisites

Things you need to install the software and how to install them.

Here API Key to benefit from Here! Rest APIs

1. Get an Here! Freemium Account (location APIs) here
2. Create an app and generate an API key
3. Store the API key in an environment variable called `here_api_key` (for Windows 10, follow this procedure)

Installing Python and pip

You need to install:

1. a recent version of Python (at least 3.7.6, see for instance this package for Windows)
2. the pip tool in order to install Python packages (see installation guide)

Once pip is installed, run the following command to download and install the requested python packages (folium for maps display, pandas for csv files processing) line and typing

```
$ pip install -r requirements.txt
```

Usage

Run the DINA Open Data Ingestion module

The first (optional) step is to process the input data from Open Data portals. This is done with the process_input_data block: `$ python process_input_data.py`
This will generate some data files in the output_data folder. Note: **this step is not required to start testing the solution** as some datasets requesting processing have already been generated and included in the repository. Indeed, some of them require heavy processing to be created. In particular, the hotel csv needs to call a geocoding API for each of the 90,000 hotels in Rome, which can take a long time.

From the files in the output_data, the indices can be generated. These are weighted averages of a subset of data indicators, as defined in the main method of create_indexes.py . To generate the indices, run: `$ python create_indexes.py`

Generate and display the DINA Heat Maps (Lighting Needs Index, Security Needs Index, Connectivity Needs Index)

Open a command line and go to the folder where DINA has been downloaded and commands: `$ python display_heatmap.py`

Examples of generated maps

It will generate 2 sets of 3 html files containing the maps in the sub-folder ./map/ and try to open them with a web browser: Indexes per month (whatever the year is, i.e. data is consolidated per month)

- DINA_Rome_Italy_300mX300m_Lighting_Needs_per_month_Heatmap.html
- DINA_Rome_Italy_300mX300m_Security_Needs_per_month_Heatmap.html
- DINA_Rome_Italy_300mX300m_Connectivity_Needs_per_month_Heatmap.html

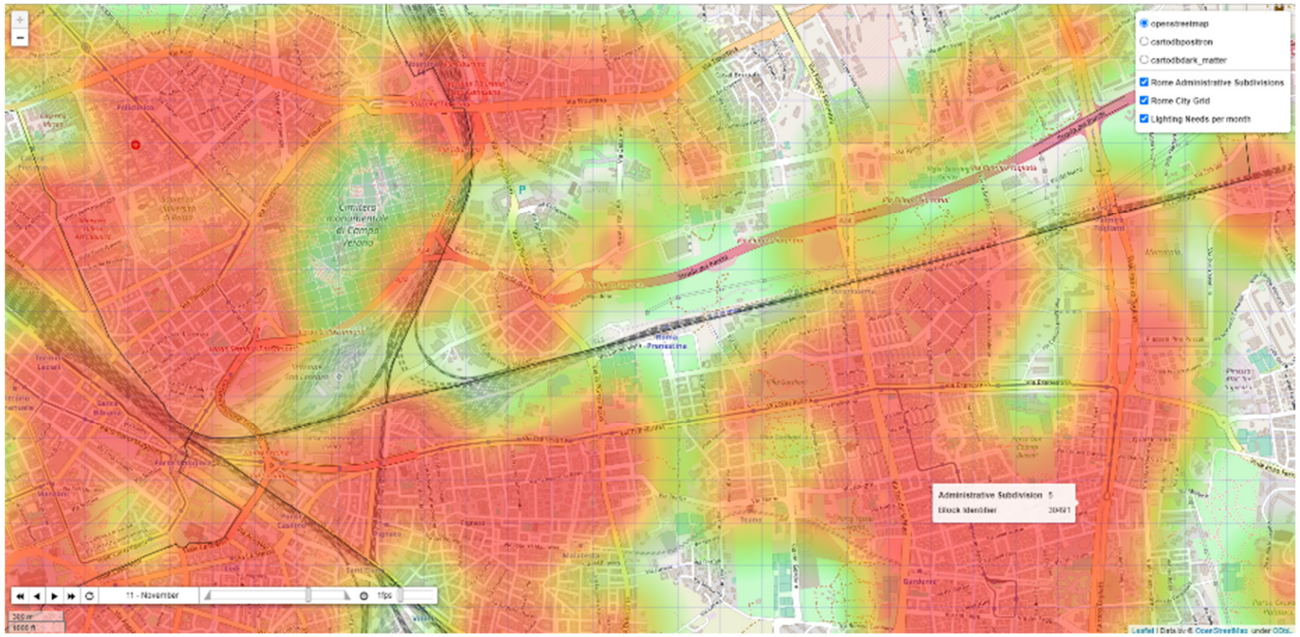
Indexes per year-month (better view for historical indexes per month, since 2019)

- DINA_Rome_Italy_300mX300m_Lighting_Needs_per_year-month_Heatmap.html
- DINA_Rome_Italy_300mX300m_Security_Needs_per_year-month_Heatmap.html
- DINA_Rome_Italy_300mX300m_Connectivity_Needs_per_year-month_Heatmap.html

Corresponding indexes files (csv) are available here:

- lighting_needs_index.csv
- security_needs_index.csv
- connectivity_needs_index.csv

Example of Heatmap for the city of Roma and Lighting Needs estimation in November (centred on Roma Municipio V):

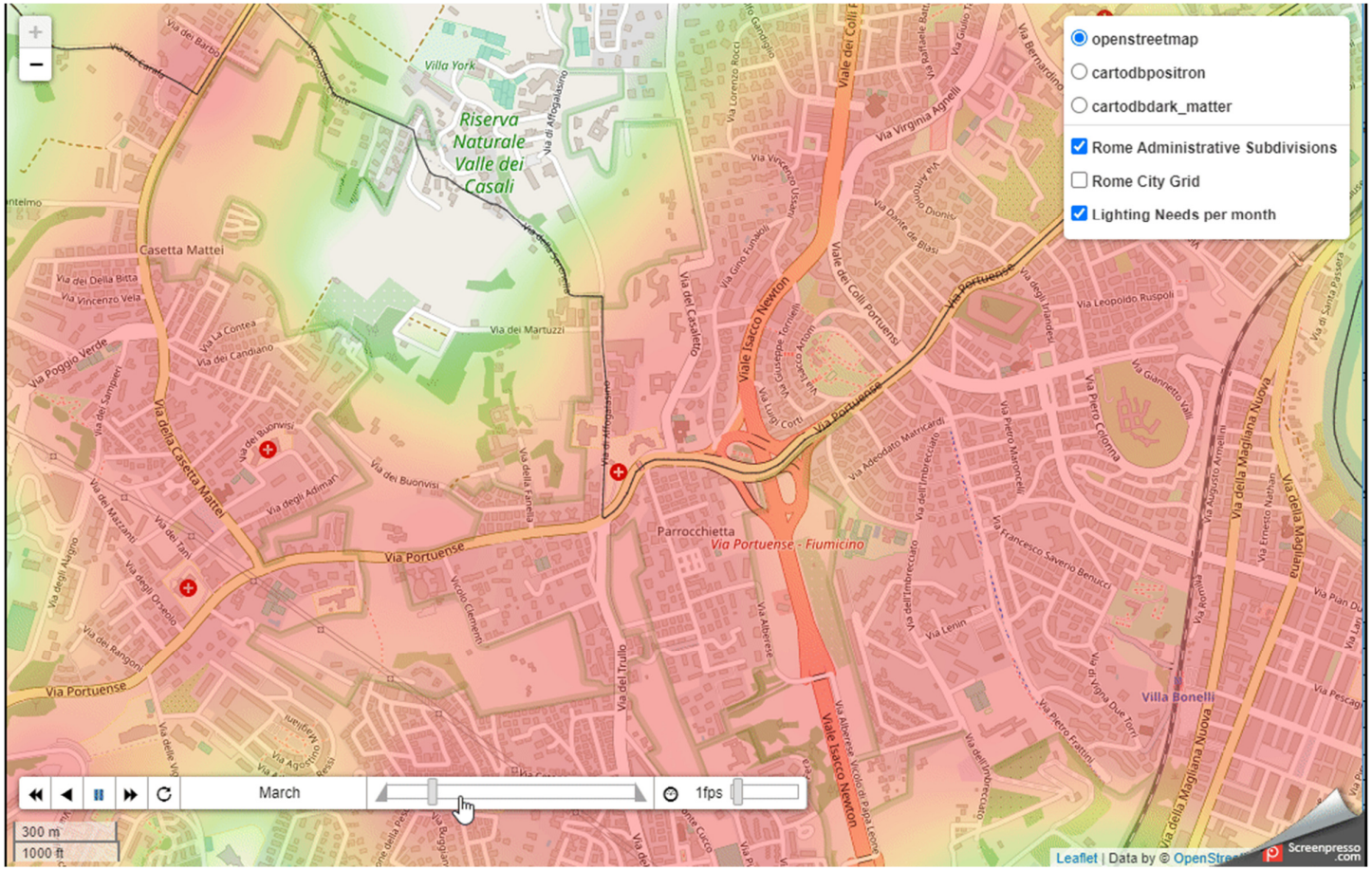


In the example above, we did not deduct any special lighting needs inside the "Cimitero monumentale di Campo Verano"... data is talking!

Example of Heatmap for the city of Roma and estimated historical Wi-fi connectivity Needs in October 2019 (centred on Roma Municipio I):

In the example above (using another type of display, closer to a 'Control Room'), the wi-fi connectivity need is probably influenced by a higher concentration of accommodations attracting tourists

Example of animated Heatmap for the city of Roma showing estimated lighting Needs estimation overtime from January to December:


openstreetmap
cartodbpositron
cartodbdark_matter
Rome Administrative Subdivisions
Rome City Grid
Lighting Needs per month
Riserva Naturale Valle dei Casali
Villa York
Casetta Mattei
Via Portuense
Via Portuense - Fiumicino
Parrocchietta
Viale Isacco Newton
Via della Magliana Nuova
Villa Bonelli
March
1fps
300 m
1000 ft
Leaflet | Data by © OpenStreetMap

Built Using

- Folium - Maps generation
- Pandas - CSV Files analysis
- Here Geolocation APIs - Geocoding of Hotels/Accommodation data

Datasets used in the scope of this proof of concept

Main source of data: Roma Open Data portal

Dataset Name	Description	Provider	Resource	Period Covered	Comments
Hotels/Accommodation Data	List of accommodation facilities in Rome Capital in 2019-2020.	Roma Open Data Portal	Roma Capitale accommodation facilities in 2020	01/2019 – 08/2020	The format of dataset changes from one month to another in 2019 so we voluntary skipped some months to make the process easier
Accidents Data	The dataset contains the list of road accidents that occurred in the territory of Roma Capitale in the year 2019. The dataset contains all the road accidents in which a patrol of any Group of the Roma Capitale Local Police intervened. Therefore, only incidents in which the parties involved have reached a conciliation are excluded. The dataset does not include the accidents that occurred on the Grande Raccordo Anulare of Roma Capitale.	Roma Open Data Portal	Road accidents in the territory of Roma Capitale – Year 2019-2020	01-2019 – 12/2019, 01/2020 – 02/2020	we merged monthly data files into 1 consolidated file

Dataset Name	Description	Provider	Resource	Period Covered	Comments
Public wifi Data	Anonymized web browsing sessions found in the Rome WiFi system	Roma Open Data Portal	Anonymized web browsing sessions found in the Rome WiFi system	04/01/2020 to 30/09/2020	
Shapes of Administrative Subdivisions	Shapes of Roma Municipi in geojson format (municipi.geojson)	Roma Urbanistica	https://romaurbanistica.carto.com/tables/municipi/public	N/A	
Single Reporting System of Roma Capitale. Reporting data for the year 2020	The Dataset aims to represent a framework for managing the flow of reports from citizens through the SUS (Single Reporting System https://www.comune.roma.it/web/it/di-la-tua-segnala.page), steps and actions that are taken from opening to closing the issue. The reports present are geo-localized by municipality and by subject - the data of the reports published are as they appear from the user reports, regardless of the veracity certified or not with respect to what is reported.	Roma Open Data Portal	Single Reporting System of Roma Capitale. Reporting data for the year 2020	01/2020 to 07/2020	We used a subset of claims types (aka 'Argomento - codice')

How to change important parameters

Grid size

The grid size governs the granularity of the indexes, and in turn of the heatmaps. The size of the blocks is set in two variables in `config.py`:

```
block_width = 300
block_height = 300
```

To change the grid size, change these values, and re-run the `process_input_data.process_input_data()` method to re-generate the indexes with the new grid.

Weight of the data sources in the indexes

The final three indexes (lighting, connectivity and security) are weighted sums of indexes over the different data sources. The weighing is set in the `indexes_calculation_parameters` dictionary in the `config.py` module. You can modify the values of the dictionary (float from 0.0 to 1.0) in order to change the weights of the data sources in each index and thus influence the global index, based on your analysis of which data is more relevant to infer implicit requirements from citizens and visitors.

```
indexes_calculation_parameters = {
    'lighting': {
        'accidents': 0.4,
        'hotels': 0.2,
        'wifi': 0.2,
        'lighting_claims': 0.2
    },
    'connectivity': {
        'hotels': 0.5,
        'wifi': 0.5
    },
    'security': {
        'accidents': 0.5,
        'security_claims': 0.5
    },
}
```

Python modules in this repository

config.py

Contains general configuration, location of input files, and the weighting of the different data sources for the three final indexes

create_indexes.py

The main method of this module generates the different indexes from the processed files in `output_data`. It uses the weights from `config.py`

display_heatmap.py

The main method of this module takes the generated indexes and generates the final heatmaps

get_accidents_geodata.py

The main method of this module processes the accidents data from the input_data folder, and counts the number of accidents that occurred with insufficient lighting per grid block and month. It is called from `process_input_data.py`

get_claims_geodata.py

The main method of this module processes the claims data from the `input_data` folder. As the claims data is only available at the administrative division level, the claims relative to the target categories (security and lighting) are counted by month and by administrative subdivision (e.g. municipio in the case of Rome). All the blocks in an administrative division are then given the score corresponding to that division. It is called from `process_input_data.py`

get_hotels_geodata.py

The main method of this module processes the hotel location data from the `input_data` folder. If the processed data is not geolocalized (i.e if the latitude and longitude columns are not present), the module calls the HERE api to geolocalize the hotel from its address, and outputs the input file with the added geolocalization columns as well (this helps saving extra calls). The main output is a count of rooms by grid block and month. It is called from `process_input_data.py`

get_wifi_logs_geodata.py

The main method of this module processes the wifi usage data from the `input_data` folder. If the processed data is not geolocalized (i.e if the latitude and longitude columns are not present), the module calls the HERE api to geolocalize the wifi hotspot from its address, and outputs the input file with the added geolocalization columns as well. Within the processing, coordinates for the wifi spots is also cached, as there can be many usage logs for a single wifi hostpot. The main output is the overall download data usage by grid block and month. It is called from `process_input_data.py`

process_input_data.py

The main method of this module processes the different data sources located in the `input_data` folder. It calls individual data processing modules for each of the datasets (accidents, hotels, wifi and claims)

utils.py

This module contains various methods that are used by other modules. In particular, it contains the method to call the geolocalization API (HERE), and different utilities for generating the geographical grid.

---

[i] https://urban.jrc.ec.europa.eu/thefutureofcities/the-citizens-city#the-chapter
[ii] https://the.gt/geoffrey-moore-positioning-statement/
[iii] https://romaurbanistica.carto.com/tables/municipi/public
[iv] https://dati.comune.roma.it/catalog/dataset/d865
[v] https://dati.comune.roma.it/catalog/dataset/d852
[vi] https://dati.comune.roma.it/catalog/dataset/wifi2020
[vii] https://dati.comune.roma.it/catalog/dataset/sus1
[viii] https://dati.comune.roma.it/catalog/dataset/d852
[ix] https://dati.comune.roma.it/catalog/dataset/wifi2020
[x] https://dati.comune.roma.it/catalog/dataset/d865
[xi] https://dati.comune.roma.it/catalog/dataset/sus1