Abstract

DINA — “Data-driven Implicit Needs Analyzer " complements the Enel X Control Room and City Analytics products, by 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 Enel X is 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 (like YouCity). 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, JuiceLamp, 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 positioning of our proposal vs existing Enel X products, the rationale and guiding principles. It is not a Business Model Canvas as such but it will help the seeker (Enel X) 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 a subset of the users traditionally targeted by Enel X City Analytics and Control Room products:

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

Enel X is a data driven company

Enel X has a proven experience in developing and deploying solutions (decision support tools, operations management) which consume various sources of "almost static" and "live" data in order to provide insights (Enel X City Analytics) or semi-automation (Switch-on/Switch-off/Dimming of lighting units from the Enel X Control Room).

The sources of information leveraged by Enel X platforms are already very diverse: Enel X sensors, remotely controlled Smart Lighting Devices, Mobile Apps like Enel X YuUrban[i].

Enel X acknowledges that the voice of citizens matters to improve life in the cities

Citizens can play a crucial role in identifying or actively intervening in urban challenges, [ii]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 Enel X is investing (fully in line with the United Nations Sustainable Development Goals).

Indeed, on top of YuUrban, 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 (for instance, Enel X platform could integrate YouCity[iii], an Italian start-up with a social vocation that aims to promote and facilitate

collaboration between citizens, public bodies and associations for the realization of projects with a public and social impact).

But gathering people's needs is not that straightforward

Today, a lot of data used by Enel X 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 like YuUrban, or YouCity, if implemented.

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.)

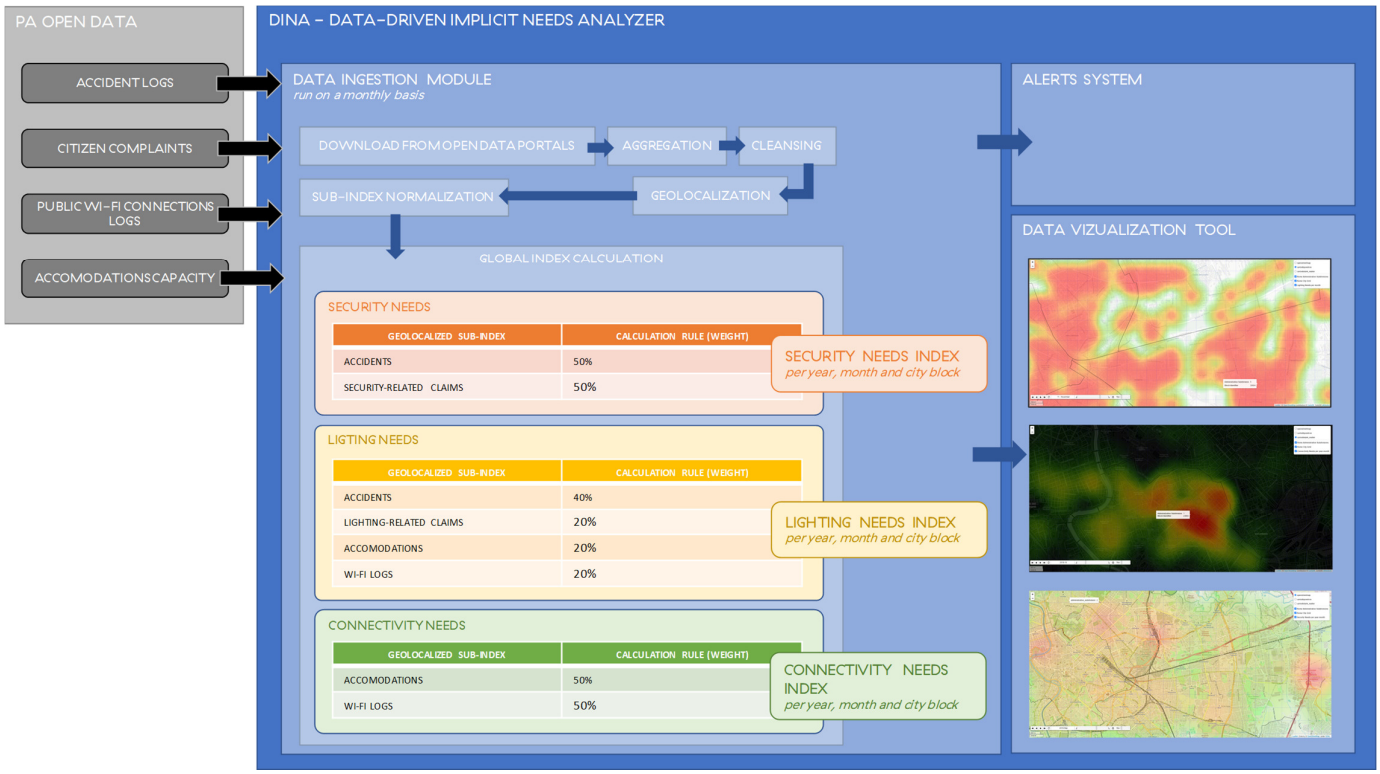
On top of that, based on our research (focused on the Italian market), most of the portals/apps which have been implemented to capture citizen's needs are in Italian, thus potentially refraining international tourists from expressing their own needs (or reporting issues) when visiting the country.

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 Enel X data sources and 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 Enel X Smart City solutions with **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 Enel X has 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) :

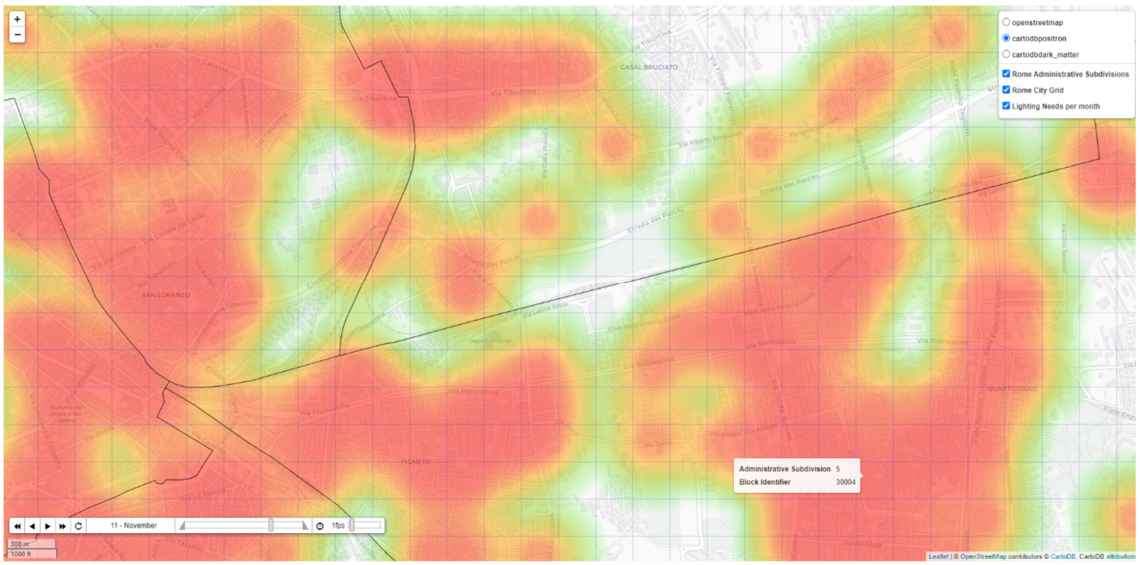
- *What are the city areas where we should invest more in smart lighting equipment?*
- *Do lighting needs evolve over time (depending on the month of the year) in some areas?*
  > If yes, I could adapt my smart lighting parameters to extend lighting time in such areas and possibly install new smart lamps
- *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?*

For Enel X sales teams, this tool could also be used to seize the opportunity to install and promote JuiceLamps (areas where connectivity, lighting and security needs are all significant and where installing 1 single equipment could make sense).

Note: In the scope of this challenge, the visualization solution will be delivered as a standalone User Interface, but the vision is to integrate it in City Analytics and Control Room directly, to complement existing visualization tools (e.g. adding a new layer to existing maps) ; see 'Positioning vs Enel X offer' section.

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)

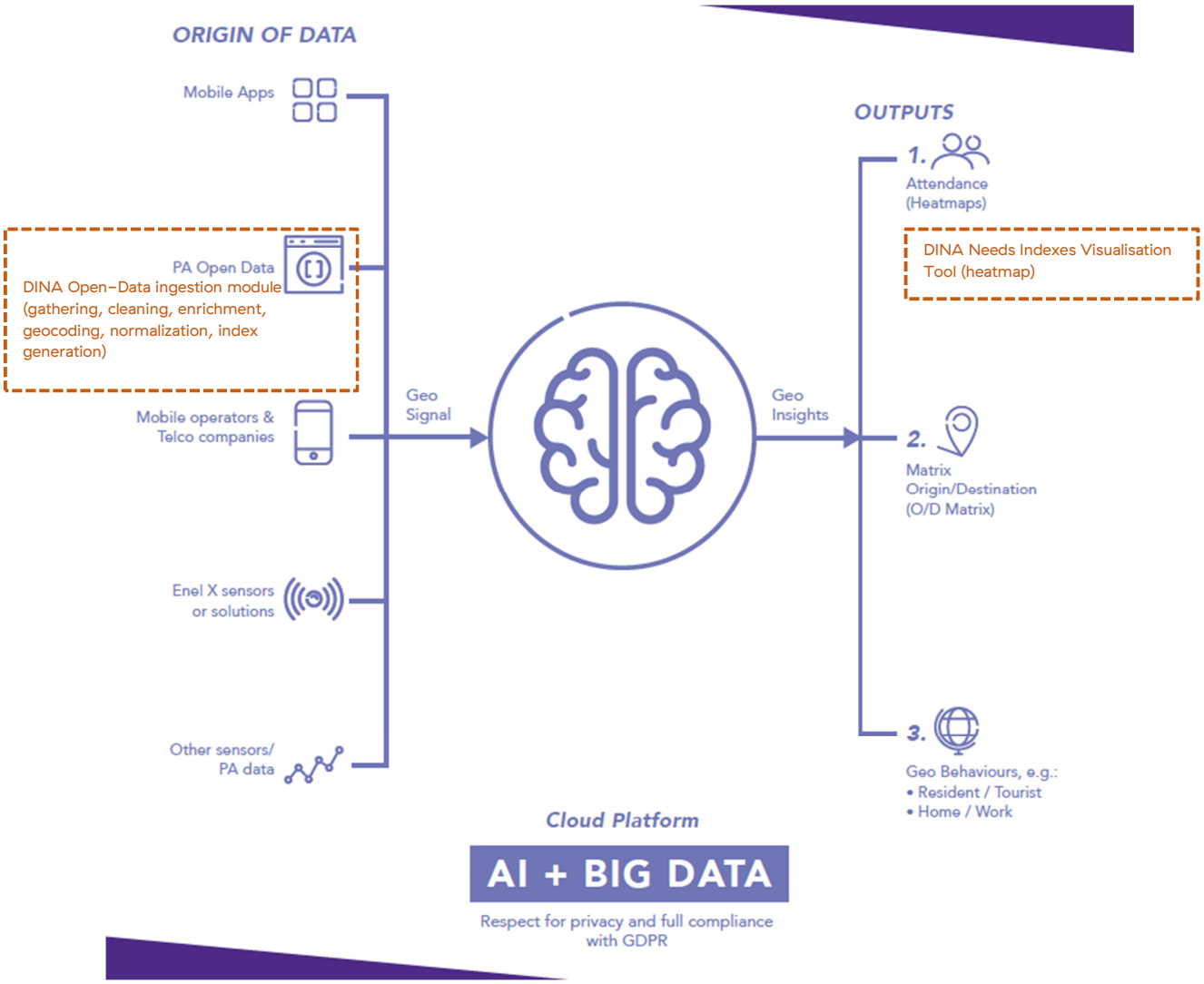


DINA Heatmap: Inferred 'Improved Public Lighting Needs' in November (based on 2019-2020 historical data) in the Roma Municipio V (red means higher need)

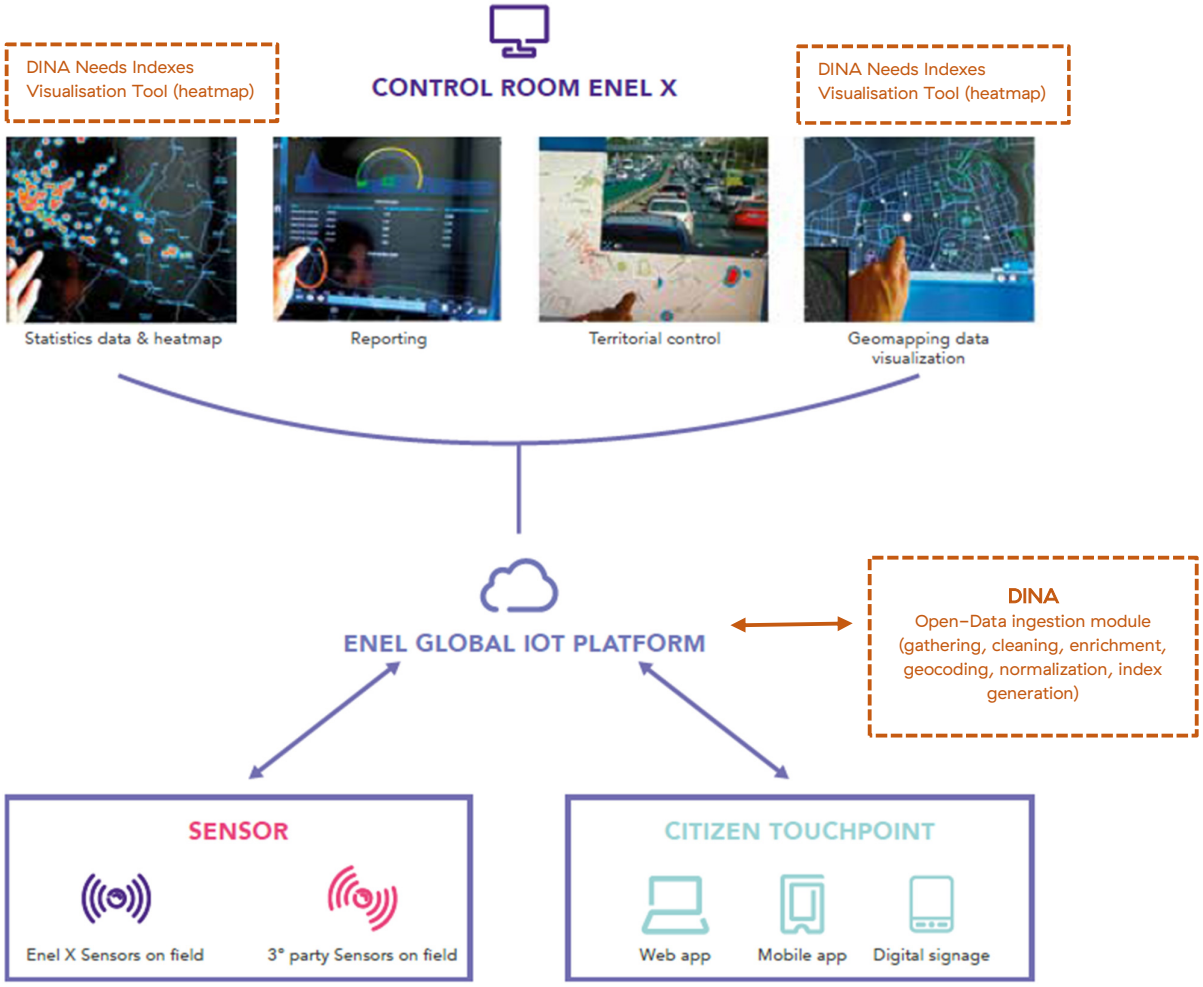
Positioning vs Enel X offer

DINA does not aim at becoming a standalone offer but more an add-on to Enel X Decision Support products:

- Enel X City Analytics (Heatmap, User Profiling)
- Enel X Control Room for Smart Lighting (even if we foresee some benefits for the JuiceLamp product deployment as well)



Positioning of DINA vs City Analytics



Positioning of DINA vs Control Room

Value Proposition

While the underlying technical solution of DINA is standard, the value proposition (formalized below using the *Geoff Moore's Value Positioning Statement*[iv]) 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 dashboard which complement the Enel X Control Room Statistical Data/Heatmap and Geomapping data visualization tools, by displaying 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, like *Enel X JuiceLamps*)

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 dashboard which complement the Enel X City Analytics by displaying 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 dashboard which complement the Enel X City Analytics by displaying 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

Part of the Enel X's vision[v] is to transform cities to improve citizens quality of life by developing, together with cities & authorities, the right solutions for lighting and smarter energy use, in addition to offering connectivity and mobility solutions that will improve the lives of residents.

DINA (Data-driven Implicit Needs Analyzer):

1. Relates to the Enel X **Smart Public Lighting** offer (and specifically the Control Room solution[vi] and possibly the new Enel X Juice Lamp[vii]) and to the Enel X **City Analytics**[viii] solution.
   Why?
   This is where Enel X seems to build a unique value proposition (c.f. recent launch of JuiceLamp) compared to competition.

2. 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.

3. Can be mainly considered as a **decision support tool** for cities, but also as a **business development tool** for Enel X
   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 Enel X Business Development teams could also leverage for internal purposes (to be validated of course), i.e. spot opportunities to deploy more smart

lamps or to upgrade the existing ones with innovative options (c.f. JuiceLamps with addons which directly contribute to improving lighting, security via cameras and connectivity via Wi-Fi hotspots).

4. 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 like YuUrban. 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 citizen's needs.

The solution in detail

Summary of what has been implemented in the scope of this Reduction-To-Practice

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).

Enel X can claim the source code and make some amendments if needed (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 (Enel X's current markets). On top of that, Enel's headquarters being located in Rome, we assumed that it would be easier for Enel X to engage with Rome's officials if the solution needs to be presented/tested/evaluated

- 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[ix] portal (used in our geolocalization process, see later)
  2. List of accommodation facilities in Rome Capital in 2019-2020[x]
  3. List of road accidents that occurred in the territory of Roma Capitale[xi] in the year 2019-2020

4. Anonymized web browsing sessions found in the Rome WiFi system[xii]
5. Single Reporting System of Roma Capitale. Reporting data for the year 2020[xiii]

   The rationale for each dataset will be explained later in this document

General comments regarding the open data sources

As Enel X probably noticed, 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)

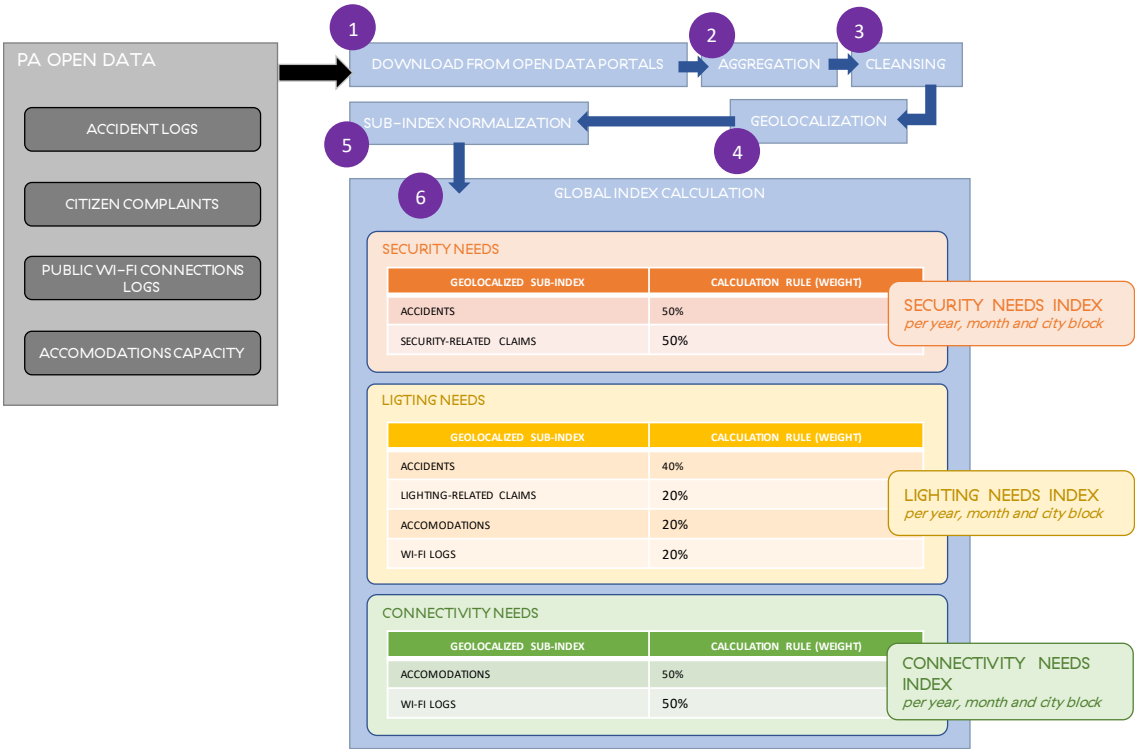
4. URLs to download datasets (at least in the Roma Open Data portal) do not have a common pattern and standardized format (like *<prefix>/DATA_YEAR_MONTH.csv*) which does not help data consumers to automate downloads and updates

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 Enel X is to show the 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[xiv] 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 cause by bad lighting (still relevant for the security needs index)

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

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? Tourist 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[xvi]

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[xvii]

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 point 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

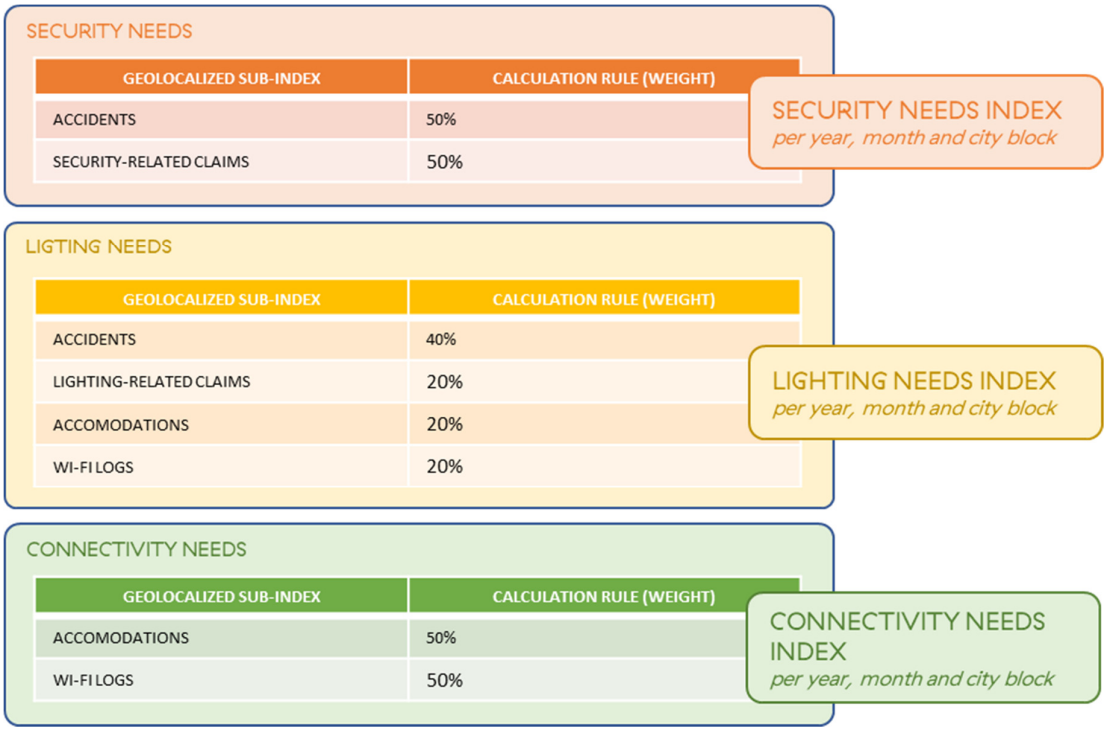
The Connectivity, Security and Lighting global indexes are computed from the tables generated for each sub-indexes. Each global index is defined as a weighted sum of different sub-indexes. The index number i is thus computed as follows:

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

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).

Combine indexes

If Enel X wants to spot opportunities to deploy JuiceLamps (c.f. value proposition for Enel X Business Development teams), then it could be interesting to combine the 3 indexes (connectivity, lighting, security) and to display them on a single map. The code that we will deliver to Enel X can be easily adapted to do so.

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:

- link to the repository is not provided as part of this written proposal because it contains elements (PIIs) which could be used to identify the 2 solvers who contributed to this solution (which is breaking Innocentive rules). Solvers understand that they may be asked to provide access to the code if this submission is shortlisted
- repository structure and readme.md files are provided at the end of this document

Fit with the challenge's criteria

This section describes shortly how the current proposal fits with the challenge's criteria[xviii], to ease the evaluation.

Solution Requirements

1. The innovative solution must be able to be applied to established Enel X activities such as those listed above or equivalent

Answer: YES

Justification: As described in the "Positioning" section, the solution complements Enel X City Analytics and Control Room offers. The choice of "needs indexes" (security, lighting, connectivity) is fully in line with the value proposition of the Enel X JuiceLamps and we can assume that our solution can also be used as a Business Development tool for Enel X, which highlights the need to deploy JuiceLamps in a city

2. The innovative solution must demonstrate an added value to the current Public Services Offering, leveraging Public Administration open data

Answer: YES

Justification: The solution is using Public Administration data (see list of datasets used for the proof of concept), as a way to complement other data sources. We tried to overcome the challenges linked to this data (lack of standard format between cities, different levels of geolocation data, lack of fresh data etc.) and propose a modular solution

3. The innovative solution, fully working with Open/Public data only, can achieve higher values/impacts or grant additional features/insights when combined with Enel X solutions

Answer: YES

Justification: We did not have to detailed documentation of Enel X solutions, so we guess that our solution brings value to existing dashboards/heatmaps because we offer a "current and future needs" view with a method to transform data into a "need". By combining our geolocalized data to Enel X's one (like location of smart lamps, Wi-Fi hotspots), the visualization tool becomes even more relevant.

4. Any best practices have to be verified and documented (not just an "idea")

Answer: YES

Justification: Details have been provided in the "Solution/Demonstration" of this document. We also checked manually that our indexes are credible (like: lighting needs not connectivity needs in cemeteries)

5. The innovative solution must demonstrate "on field" its capabilities ("reduction to practice" needed)

Answer: YES

Justification: A proof of concept has been developed. Code is available on demand. Modules that have been developed are shortly described at the end of this document.

6. Solutions that focus on the following geographical areas (Europe, United States, South America)

Answer: YES

Justification: Solution aims at being global. For the Reduction to Practice challenge, we focused on Italy (Enel X main market where Enel X will be in a better position to test the concept with its innovation-friendly customers) and more specifically to Rome, Italy. We initially wanted to work with Bologna data but it appears that the Rome's Open Data portal offered more relevant and fresh data (at least, to support our specific use cases).

7. All solutions must be safe and comply with any local regulations and be sustainable environmentally and economically

Answer: YES

Justification: We don't foresee any problem: data is Open Data coming from Public Administration (e.g. data is already anonymized to be GDPR compliant) and we use it to feed a decision support tool. We do not expect that our module will be used one day to automate processes without any human intervention or validation, which simplifies the implementation.

8. Solutions should support Enel's commitment on SDGs (Sustainable Development Goals) with specific emphasis on SDG9 and SDG11.

Answer: YES

Justification: The solution is mainly supporting the SDG Goal 11: *Make cities and human settlements inclusive, safe, resilient and sustainable* and specifically target 11.3:

By 2030, enhance inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries

By extension, we can also support targets:

- *11.1: By 2030, ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums*
- *11.2: By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons*

9. The proposed technology should offer Enel "freedom to practice" or be available for potential licensing. There should be no third-party patent art preventing the use of specific equipment and materials for their commercial application.

Answer: YES

Justification: We did a quick state-of-the-art review and did not see any obvious protected prior art which could prevent Enel X from using the solution. Note: for this exercise, and based on our experience, a real patent engineer or patent attorney are requested to be 100% sure.

10. Desired: Easy installation

Answer: YES

Justification: Details are described at the end of this document in the readme.md section ("Getting Started")

11. Desired: Reasonable investment costs

Answer: YES

Justification:

- **Initial cost:** initial development and implementation cost should be limited (we assume that Enel X data processing capabilities can manage this additional source of information, including hosting and data pipelines ingestion)
- **Ad-hoc implementation cost:** every time a new city is implemented with Enel X solutions, we assume that integrators are already dealing with similar activities (integration of 3rd party apps, Public Administration own data sources etc.) in our case it's a matter of adapting connectors and data processing modules to the new data source which should be easy thanks to Open Data portals and the availability of data: quick and easy by nature). Main challenge is Data Cleansing and normalization
- **Running cost:** very small compared to live data coming from cameras. Depending on the granularity of the PA Open Data, we can expect one refresh per month which has a very low cost

12. Desired: Existing commercial solution (or minimally at a pilot/demonstration stage)

Answer: YES

Justification: No commercial solution. A proof of concept has been developed in the scope of this challenge

13. Desired: Applicable in a global context (that is, they would be transferable to different parts of the world)

Answer: YES

Justification: As Enel X Innovation teams have probably noticed, the issue with Open Data, and in particular "Public Administration Open Data" is:

- Data at *country level* are not granular enough to be used efficiently in a smart-city context
- Data at *city level* can have a good-enough granularity but they are suffering from low refresh rate and a lack of standardization (Italy is one example but we noticed the same with France and US)
  > in other words, if you code a data parser for one city, it might not be re-used for another city

Still, we can overcome these challenges as the general principles of our solution are:

- to build an extensible and modular platform (fully in line with the principle of the Enel X Big Data platforms)
- to provide, as outputs, 3 **normalized** indexes per geolocation (300mx300m) and per month, which is an approximation that should fit with most of the cities' open

data (whatever the granularity is): outcome will always look the same in Enel X frontends. What will change is the way we calculate, depending on the city.

- to let each city define how the index is calculated
  > For instance Rome will leverage, among other things, the car crash reports to build the “security needs index” (because this info is already public) whereas Bologna might use another data source if car crash reports are not publicly available. In the US crime records are more systematically available and could complement the security index

Project Deliverables

1. Detailed technical description of an approach/algorithm that can meet the above Solution Requirements

See section “Technical description of the approach and Algorithms: focus on the data ingestion module” and readme.md at the end of this document

2. Rationale for the proposed technology/best practice, including direct cost

See paragraphs “Guiding principles and rationale for DINA” page 9.

3. If a best practice, provide documentation of its proven record and savings

Not applicable (our solution is not about savings)

4. If available, drawings for any devices, schematics for installation and rough estimations of costs (e.g. investment, operational, maintenance) and power consumption

Not applicable

5. Provide data and elaborate on any known limitations

Known limitations are described in the document.

6. The Solver has to provide as deliverable within the idea of the model the Input Data and Output Data. The Solver will be asked to supply the algorithms, the mathematical models and related scripts using standard languages (Java, Python, R...). It will be also required to release the visualization layer/ open source tool (including source code) to enable the Seeker to compile, execute the algorithm, and validate the method using the relevant data sets.

See section “Resources to access the Proof of Concept” page 20.

Other criteria

Overall scientific and technical feasibility of the proposed solution;

Proposed solution is feasible, and skills required to implement it in-real-life (in production) can be acquired easily (Enel X already own these skills). We used Python 3.7.8 and standard libraries (Pandas, GeoPandas, Folium) to develop the solution.

Economic potential of concept (e.g. Total Cost of Ownership);

Economic Potential is hard to quantify because the solvers are missing economic data from Enel X (like traction model for the Smart Lighting product line, revenue model, cost structure). However, the initial assumption is that cities are getting smarter and smarter, digitalization trend is real, cities will always require public lighting so this solution will contribute to the revenues of Enel X smart lighting products.

Business potential for Enel;

See above. Beyond the direct revenues that could be generated by such a solution (if the value perceived by cities is big enough and if cities demonstrate willingness to pay for an additional chargeable module of the Control Room... which needs to be validated), we can consider indirect-revenues and business value for Enel X like:

- Adding stickiness to Enel X products like City Analytics, JuiceLamp or the Control Room and enable up-sell opportunities
  > "DINA" will highlight the value of deploying more lamps, or upgrading (or replacing) the existing ones with additional add-ons like wi-fi or video surveillance cameras
- Supporting cities in their planning exercise (e.g. plan investment in smart lighting solutions thanks to the decision support tool) and position at the same time Enel X as the right provider
  > therefore, help Enel X to anticipate needs (delivery, roadmap, production, commercial guidelines etc.)
- Engaging even more with city officials and their IT teams (critical when it comes to the digitalization of processes)
  > solution requires clean and updated Open Data to work ; Enel X will not only influence the people who are accountable for Open Data at city level, it will position itself as an enabler of data openness and a key partner ("*we consume your open data, and help you to improve their quality*")

Novelty and creativity;

Novelty lies in the use of Open Data as proxies to guess/predict needs. We are proposing a process to translate facts into requirements.

Potential for proprietary position (i.e., is the technology novel or protectable);

Technology as such is not new and might not be protected via a patent in the European Union. Business Processes behind the scene might be protected via:

- at least defensive publications, which could give Enel X the freedom to operate

- possibly patents in some regions of the world (e.g. USPTO[xix], even patentability criteria have changed in the last 5 years)

User's capabilities and related experience

The 2 solvers who elaborated this proposal have a strong experience in IT, Data Science, Innovation Management, and the Travel & Tourism Industry in an international context. They already contributed to the design, development and deployment of business intelligence products targeting authorities (like Destination Marketing Organizations, tourists' offices etc.)

Realism of the proposed solution;

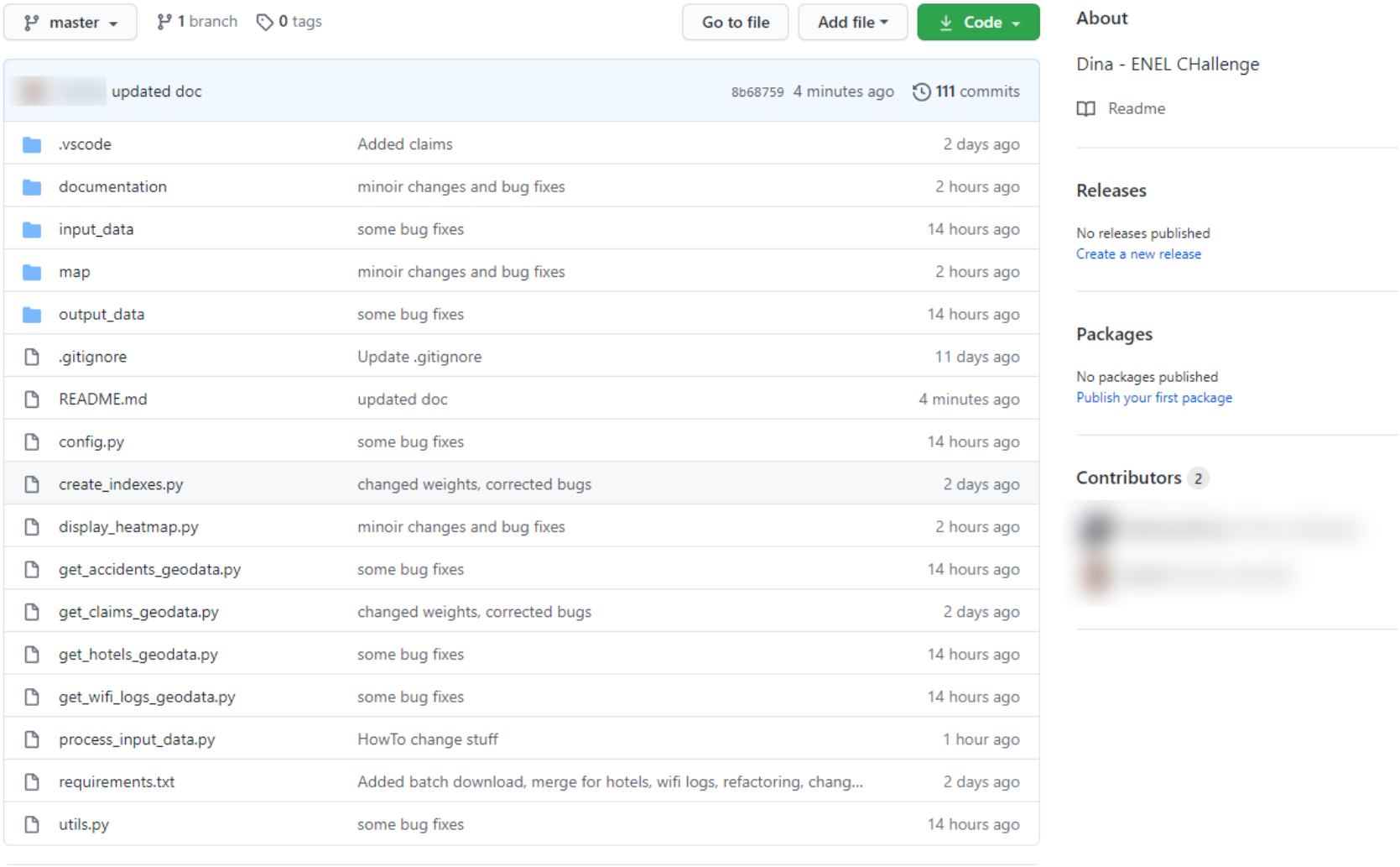
The proposed solution is simple, pragmatic and opens new opportunities (more complex and uncertain but probably more impactful).

Maturity level of the proposal.

Let's be honest: Maturity is medium. As stated in the document, we made a number of assumptions that need to be validated in the field with Enel X and city officials.

Annexes

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

[redacted]

Content of README.md

DINA – Data-driven Implicit Needs Analyzer

A tool which leverages Public Administration Open Data to provide cities with actionable insights about what citizens implicitly need in terms of security, lighting and connectivity

Table of Contents

About

This project is a proposal submitted in the scope of an Innocentive Challenge New Smart City Solutions Enabled by Open Data 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 Destination Marketing Organizations)

The solution prototyped in the scope of this project is called DINA (Data-driven Implicit Needs Analyzer) and is made of 2 main components:

1. DINA Open Data ingestion module It aims at becoming a transversal component, supported by an extensible framework, which gathers, cleans, enriches and normalizes Open Data (coming from Public Administrations or not) in order to provide Smart City analytics solutions with

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 20m x 20m) and per month.

Note: our vision is obviously to increase granularity (lower the size of the geographical areas, provide data per day or even per hour) when Public Administration Open Data will permit and this is why Enel X has a key role to play to influence City's Open Data roadmaps.

2. DINA "Needs Indexes" Visualization tool The solution includes a data visualization module 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: month- and to filter by type of index). Note: In the scope of this challenge, the visualization solution will be delivered as a standalone User Interface, but the vision is to integrate it in City Analytics and Control Room directly, to complement existing visualization tools (e.g. adding a new layer to existing maps)

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

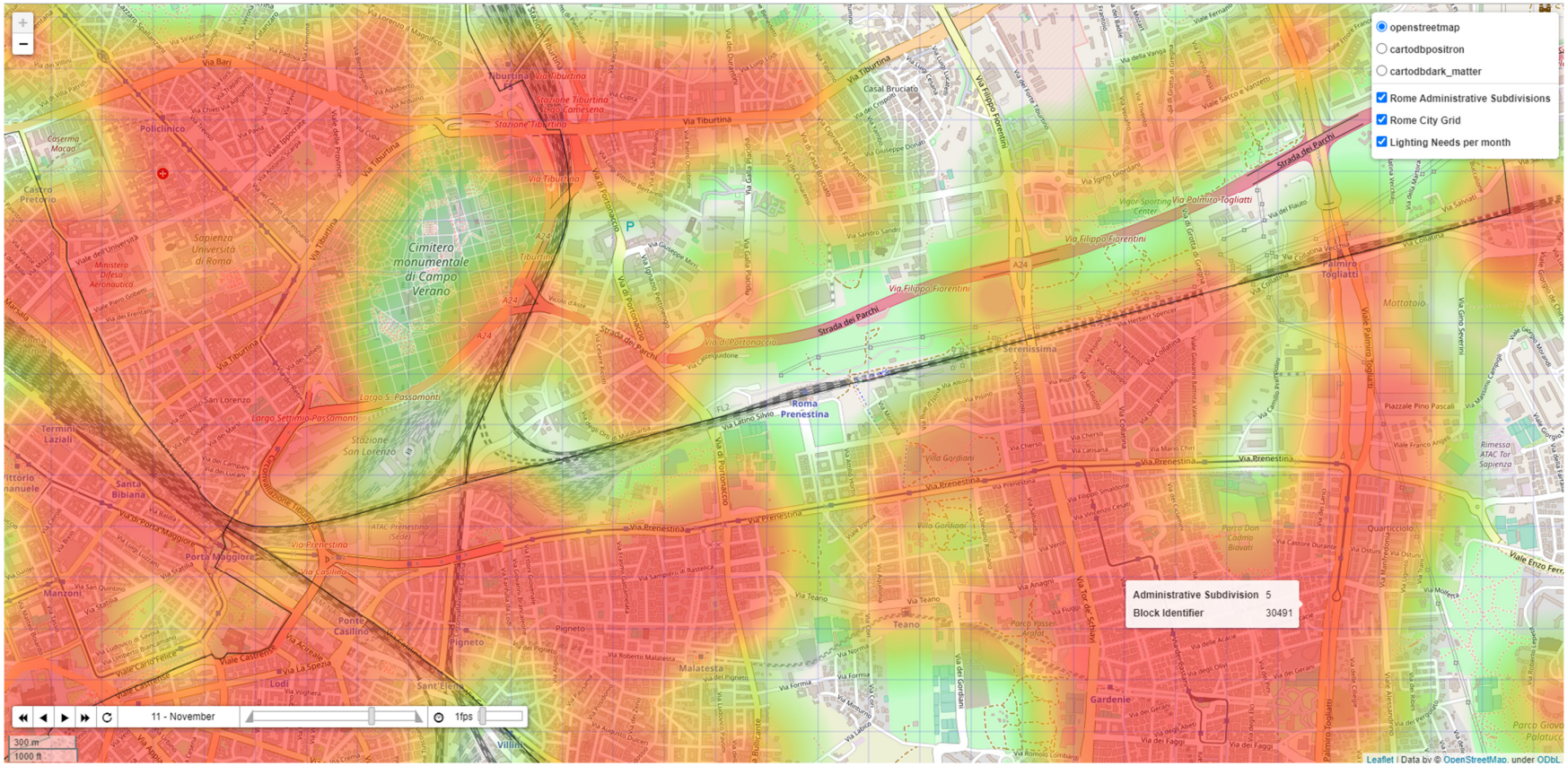
Indexes per year-month (better view for historical indexes per month, since 2019)

- 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):

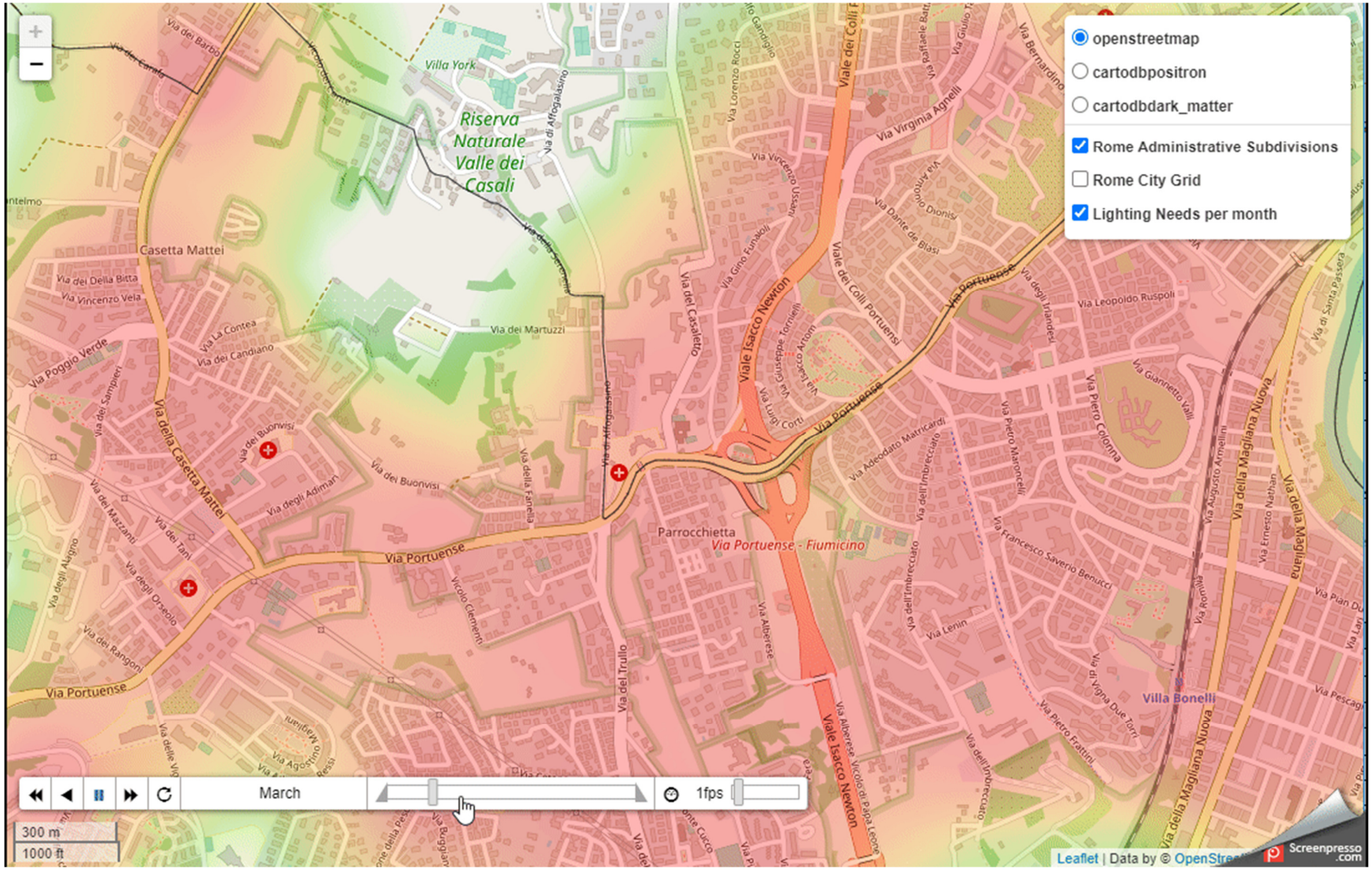


In the example above, we did not deduct any special lighting needs inside the "Cimitero monumentale di Campo Verano"... data is talking!

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 della Casetta Mattei
Viale Isacco Newton
Viale dei Colli Portuensi
Parrocchietta
Via Portuense - Fiumicino
Villa Bonelli
March
1fps
300 m
1000 ft
Leaflet | Data by © OpenStreetMap
Screenpresso.com

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	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
	does not include the accidents that occurred on the Grande Raccordo Anulare of Roma Capitale.				
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://www.enelx.com/it/en/smart-cities/solutions/public-lighting/yourban-app
[ii] https://urban.jrc.ec.europa.eu/thefutureofcities/the-citizens-city#the-chapter
[iii] https://www.youcity.org/
[iv] https://the.gt/geoffrey-moore-positioning-statement/
[v] https://www.Enel X.com/en/who-we-are/vision
[vi] https://www.Enel X.com/content/dam/enel-x/ecity/pdf/en/Brochure-PUBLIC-LIGHTING-EN.pdf
[vii] https://www.Enel X.com/it/en/smart-cities/solutions/public-lighting/juicelamp
[viii] https://www.Enel X.com/it/en/smart-cities/solutions/smart-services/city-analytics
[ix] https://romaurbanistica.carto.com/tables/municipi/public
[x] https://dati.comune.roma.it/catalog/dataset/d865
[xi] https://dati.comune.roma.it/catalog/dataset/d852
[xii] https://dati.comune.roma.it/catalog/dataset/wifi2020
[xiii] https://dati.comune.roma.it/catalog/dataset/sus1
[xiv] https://dati.comune.roma.it/catalog/dataset/d852
[xv] https://dati.comune.roma.it/catalog/dataset/wifi2020
[xvi] https://dati.comune.roma.it/catalog/dataset/d865
[xvii] https://dati.comune.roma.it/catalog/dataset/sus1
[xviii] https://openinnovability.enel.com/projects/New-smart-city-solutions-enabled-by-open-data

[xix] https://www.uspto.gov/