DEPARTMENT OF REGIONAL AND URBAN STUDIES AND PLANNING

MASTER'S DEGREE PROGRAMME IN

DIGITAL SKILLS FOR SUSTAINABLE SOCIETAL TRANSITIONS

Artificial Intelligence Impact on Multi-Criteria Spatial Decision Support System (MC-SDSS) for Developing Affordable Smart Cities Solutions

Master Thesis by

Dana Al Mamlouk

Advisor

Prof. Dr. Sara Torabi Moghadam

Co-Advisor

Prof. Dr. Patrizia Lombardi

Prof. Gianvito Urgese

Academic Year 2023/2024

Degree	Master of Science in Digital Skills for Sustainable Societal Transitions
Title	Artificial Intelligence Impact on Multi-Criteria Spatial Decision Support System (MC-SDSS) for Developing Affordable Smart Cities Solutions
Author	Dana Al Mamlouk
University	Politecnico di Torino
Advisor	Prof. Dr. Sara Torabi Moghadam DIST - Interuniversity Department of Urban and Regional Studies and Planning
Co-advisor	Prof. Dr. Patrizia Lombardi Prof. Gianvito Urgese DIST - Interuniversity Department of Urban and Regional Studies and Planning
Internship	Planet Smart City
Visiting Period	From 13/04/2022 until 07/06/2023. This research is under the R10 Universe and Sei Milano Project, entitled by Planet Smart City Company.

PLANET smart city
Places that matter

AIDDA ETS
FARE IMPRESA AL FEMMINILE
Associazione Imprenditrici e Donne Dirigenti di Azienda
Ente del Terzo Settore

AIDDA AWARD

Associazione Imprenditrici e Donne Dirigenti d'Azienda

AIDDA association is founded in 1961 in Turin and is the first Italian association to provide support for female entrepreneurship through the role of women. It's been also providing tools and services of excellence for women who have roles in Italian economic structures. It has a huge network out of 800 members, works with 1300 companies has 13 delegations and 35000 employees.

Granted ***2000 Euros*** *for this thesis project with title:*

Artificial Intelligence Impact on Multi-Criteria Spatial Decision Support System (MC-SDSS) for Developing Affordable Smart Cities Solutions

Date: March 2023


AIDDA
ASSOCIAZIONE IMPRENDITRICI E DONNE DIRIGENTI DI AZIENDA
Affiliata alle FEMMES CHEFS D'ENTREPRISES MONDIALES

Borsa di Studio AIDDA Piemonte Valle d'Aosta 2023

€ 2.000 - duemila euro

a Dana **AL MAMOUK**
per la sua tesi di Laurea
"A.I evolution impact on developing countries through application of multi-criteria spatial decision support system for creating affordable smart cities."
Laurea Magistrale in Digital Skills for Sustainable Societal Transitions
Politecnico di Torino

La Presidente di delegazione
VALERIA PANINI

AIDDA Delegazione Piemonte Valle d'Aosta c/o Unione Industriale di Torino – Via Fanti 17, 10128 Torino | delegazionepiemonte@aidda.org |www.aidda.org

Abstract
Artificial intelligence can do tasks that usually require human intelligence through smart software and hardware capabilities. Artificial intelligence in a smart city includes systems that are used to mine huge amount of data such as smart grids, smart sensors, and internet of things (IoT). Digital transitions and innovation will be able to establish economic opportunities, enhance services, and enable citizen engagement in smart cities which will also have a huge impact on society. It will be able to solve several issues we're facing right now in urban areas to let us reach a digital sustainable innovative city. The constant innovation is re-shaping smart-city initiatives through self-driving cars, robots, city brains and management of city infrastructure. Urban services are being monitored by artificially intelligent entities and operating in an autonomous manner. Therefore, the aim of the thesis is to study the impact of artificial intelligence on multi-criteria spatial decision support system (MC-SDSS) for developing affordable smart cities solutions. The methodology for this research follows a qualitative and quantitative methods of research based on experimental method where it analyses smart city projects in Pune, India and in Milano, Italy through Planet Smart City company in Torino, Italy. Based on this thesis one of the conclusions we can reach is that digital transitions and implementation of artificial intelligence technology will be able to assess which indicators are better to be implemented in a smart city project to design an affordable, smart, sustainable, and healthy community.

Keywords: Digital transitions, Digital sustainability, Artificial Intelligence, Digital Twins, and Affordable Smart Cities.

Content

List of Figures

List of Tables

01

Introduction

1. Chapter 1: Introduction

1.1 Background and problem statement

Two-thirds of population will live in cities by 2050 while providing investment opportunities for tech development companies. Cities worldwide are implementing advanced infrastructural facilities and society management systems through spending in 2021 more than $130 billion to make the cities more sustainable and smarter so that residents can have higher quality of life. Cities are the place where digital innovators, data and infrastructure exists. (Popelka, 2023)

Urban living has several opportunities to be developed through urban big data available within the usage of internet of things, artificial intelligence, and machine learning which can be useful to make optimal decisions and through the optimal usage of resources. Human settlements in cities are being complex, especially after the industrial revolution that led the humankind to enter Anthropocene phase which is the reason why human activities are having a huge impact on environmental aspects. When the science of cities started to flourish in 1960s, this complexity started to get the needed attention from researchers. Thus, many studies started to introduce the importance of analysing the relationship between human activity and urban spaces. In addition, these urban spaces are dynamic and have several factors to be studied, this was the reason to shed the light into mining big data to be able to do these analyses between human activities and urban spaces. Moreover, urban strategies and plans can be analysed and guided through the usage of this big data mined from different implemented systems on different scales. Thus, designing liveable urban spaces that are more connected, efficient, and economically viable will need the usage of real-time data that is processed through artificial intelligence which will be designed based on human needs and activities. (Kamrowska-Załuska, 2021)

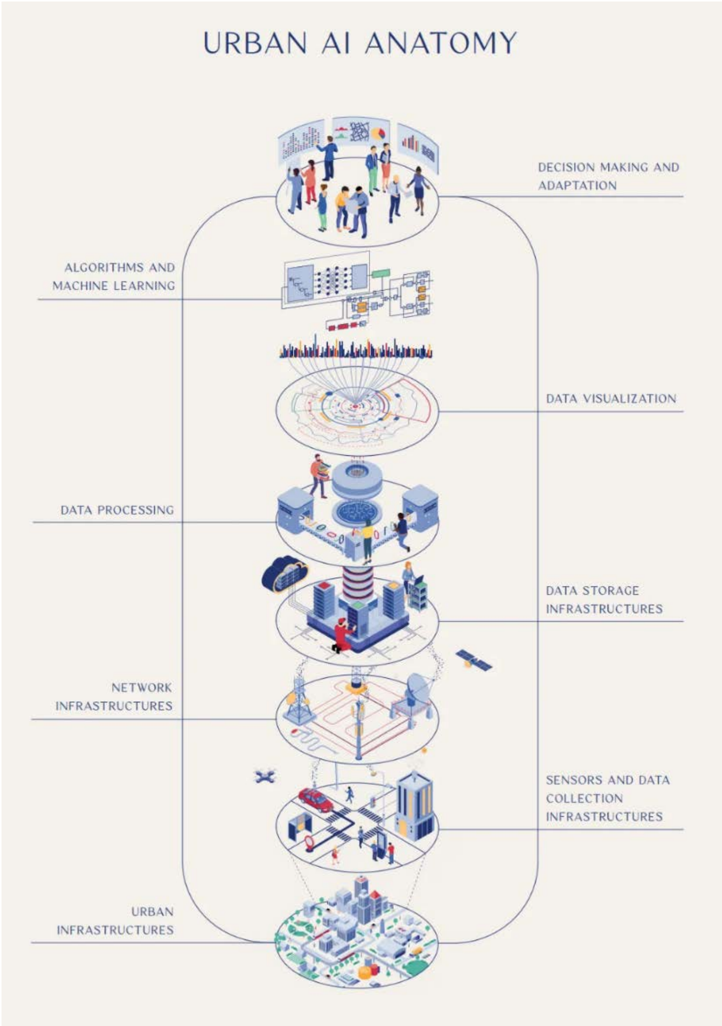


Figure 1.1 The Urban AI Anatomy detailed in the Urban AI Guide. Source: Popelka, S., Narvaez Zertuche, L., Beroche, H. (2023). Urban AI Guide. Urban AI. DOI: 10.5281/zenodo.770883

Figure 1.1 above shows the urban artificial intelligence system anatomy, which shows the process to solve urban complex problems and how algorithms and machine learning, data visualization, data processing, data storage, network infrastructure, sensors and data collection infrastructure, and urban infrastructure is a complex system that's connected to help urban technologists make decisions to solve issues that are related to human comfort. (Popelka, 2023)

Usage of artificial intelligence has effect on several scales, including the urban scale. Artificial intelligence can gather all the geo-data that exist, analyse it, and use it to help cities be a better place for this number of residents. Thus, urban spaces will be developed through usage of these new technologies to provide higher quality of life. (Allam & Dhunny, 2019)

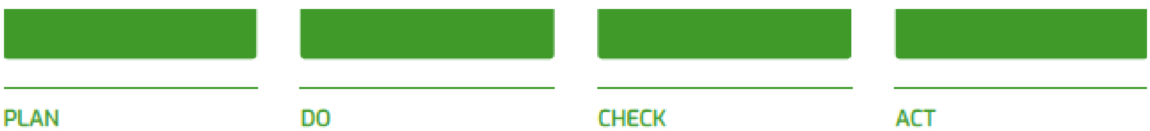
There are several issues that we have in our urban spaces that needs to be monitored, such as climate change, overpopulation, and usage of resources. Usage of artificial intelligence will help to address these challenges and achieve the sustainable development goals for a better future for a city. (Ortega-Fernández et al., 2020)

1. 1 Sustainable development goals - Source: https://cose-eu.org/2021/10/11/sustainable-development-in-developing-countries/

1.2Planet company context


Figure 1.3. Planet Company

Planet company's is a real estate company that have several projects tested with success in North America, South America, Europe, and Asia. Their goal is to integrate sustainability in project development as a primary objective through building safe, accessible, and durable homes for residents. They also create future-proof housing units that are sustainable through innovative smart solutions including digital services and social innovation programmes based on needs of residents and the community. They create projects through implementing sustainable urban planning practices. They also use technological solutions that provide residents a higher quality of life and create sustainable places for everyone.



Planet take into consideration sustainability strategies throughout all project development stages based on Plan-Do-Check-Act model by ISO 14004:2016 and they also take responsibility of resource management throughout all the process of the project. They also develop and design affordable smart cities in developing countries to satisfy the needs of the residents that live there. The vision of the company is to enhance the lives of residents by integrating the latest technological solutions in networks, home automation, air quality control, and security in an affordable smart city.

They have several projects in different countries, and all these projects are based on four key pillars: planning and architecture, technological systems, social innovation, and environment; that created more than 200 smart solutions inside planet's catalogue.

1.3Research objectives

The aim of this thesis is to study the impact of artificial intelligence on multi-criteria spatial decision support system (MC-SDSS) for developing affordable smart cities solutions. Through development of smart solutions for energy consumption and urban heat island indicators through two case studies to solve this issue through implementation of artificial intelligence.

Considering multi-actors and multi-criteria aspects to create sustainable and inclusive decision making, the thesis follows an interdisciplinary mixed methodology. The main objective of this thesis is to fine tune the GIS-based interactive impact assessment tool for smart cities which is the IIAT dashboard. Therefore, to achieve the main objective, the following sub-objectives were an essential path:

1. Collect and analyse data to improve the predictions and forecasting of futuristic data.
2. Propose theoretical impact assessment framework based on artificial intelligence A.I
3. Elaborate recommendations for continuous improvement, allowing companies to adapt and evolve its operations in response to changing factors.

Although the thesis has as a study site as R10 Universe project in Pune (India) and another study site as Sei Milano project in Milano (Italy), it can be also applied for future projects in different countries.

1.4Thesis structure context

This thesis consists in five chapters and the contents are organized to achieve the objectives discussed in the previous Section.

Chapter 2 describes the literature review on the theme. It starts with an overview of the main problems regarding the digital transitions nowadays and how they are affecting smart cities. Afterwards, it defines artificial intelligence and its usage in smart cities in addition to its historical background. Later, overall case studies are reviewed to give a perception about how this topic is being researched these days and how Urban A.I is impacting the society. After the discussion about digital transitions, the last Section discusses briefly what are the main challenges for artificial intelligence, and how it will affect the community.

Chapter 3 firstly illustrates the proposed methodology, giving a schematic flowchart of the methodological approach, with its phases, objectives, and steps. There is two case studies in this research yet they have the same methodological framework approach. The three-phases of the thesis follows "a mixed methodology" that combines qualitative and quantitative approaches. The chapter is later divided into the three main phases of the thesis' methodology, to explained in detail how each one was accomplished. Phase 1) Analyse and collect data, with selection of indicators, and collecting data from sensors on site construction. Phase 2) Assessment and Evaluation, with the prediction of futuristic energy data of the first case study in Pune and with indicator impact assessment framework proposal for the second case study in Milano. Recommendations, with the preparation of recommendations for continuous improvement that will allow companies to evolve. Chapter 3 also provides a comprehensive description of the study site, with the context and the main features of the two case studies in Pune, India and in Milano, Italy.

Chapter 4 reports the results obtained. According to the proposed three phases framework, it first discusses the procedure into the indicator's selection. It shows the best method to predict futuristic energy data and to propose a theoretical framework based on artificial intelligence for urban heat island enhanced results.

Chapter 5 sums up the conclusions and discussions. It gives a general idea of all the thesis procedure and highlights some limitations and proposals for each Phase of the methodology. In the end, it addresses future developments for further research on the topic.

02

Literature Review

2. Chapter 2: Literature Review

This chapter briefly explores the literature theory of this thesis going through the digital transitions in cities, approaching Artificial Intelligence (A.I) and Artificial Intelligence issues. Through this chapter, digital transition will be discussed and explained into details about what is artificial intelligence, understanding of artificial intelligence, types, principles, categorizations, and applications of artificial intelligence and how it will affect and be implemented in cities.

2.1 Digital Transitions in Cities

2.1.1 Digital Transition definition

There are two concepts for digital transitions: Digitization and digitalization. Digitisation is a process of transforming all paper documents into digital formats. To create new business processes and new customer experiences through using digital technologies. Digital transition will make engagement with customers different and let you be flexible since change is easy to be made on digital platforms.

Digitalization is different from digitisation because it changes how business is done since it changes how industries and businesses should enable better decision making through this process. For example, it gave Netflix the ability to make customized ads based on customer's need which enhanced the viewing experience on that online streaming platform (salesForce, 2022).

2.1.2 Digital Transition in cities

As population constantly grows in a city, having around 180,000 residents moving into a city to live every day. The Organisation for Economic Cooperation and Development (OECD) agreed that the world population will be 9 billion residents by 2050 and 70% will be living in urban centres. While meanwhile 75% of energy production is already consumed by cities and 60% of greenhouse gas emissions are also emitted by cities, cities started to develop strategies to transform digitally to solve global challenges (Iberdrola, 2023).

Figure 2.1 What makes a city smart? source: https://www.iberdola.com/innovation/smart-cities

Cities depend on Information Communication Technologies (ICT) to become smart cities or cities 4.0 through usage of big data to manage and optimize all services in the city sustainably. This will eventually result in reducing energy consumption, CO2 emissions, and increase the well-being of their residents. It's approved by the UN that the environmental commitment of a city is one of the main pillars in the SDGs. Thus, all the data collected and the real measures concerning measuring air quality through sensors on streetlights, PVC on traffic lights, improving waste collection through monitoring digitally the containers, smart water, and electricity meters, and reducing consumption of natural resources through electric vehicles and cycling. For example, there is a totally sustainable eco-city without cars in Masdar in Dubai. Another example is for Dongtan near Shanghai that use only renewable energy and all waste that is generated will be recycled. All these improvements in cities and transforming them into smart cities, based on various studies, there will be a turnover with around 1.5 trillion dollars worldwide by 2020 because of all these innovations and cost saving these digital transitions creating (Iberdrola, 2023).

All these cities need the tools to be able to make this digital transformation, the most important tool is the usage of artificial intelligence and internet of things (IoT). Artificial intelligence will be able to support decision making through utilising the data that are collected by the smart sensors in all services and infrastructure (European Parliament, 2023)

2.1.3 Digital Twins in urban planning

A physical object, person, or proves that is represented digitally in a digital version of its environment. Digital twins can help in decision-making processes since organizations can simulate real situations and outcomes. It's one of top 12 strategic technology trends for 2022 by Gartner which is recognized for city planning and management as an effective tool (McKinsey Explainers, 2023).

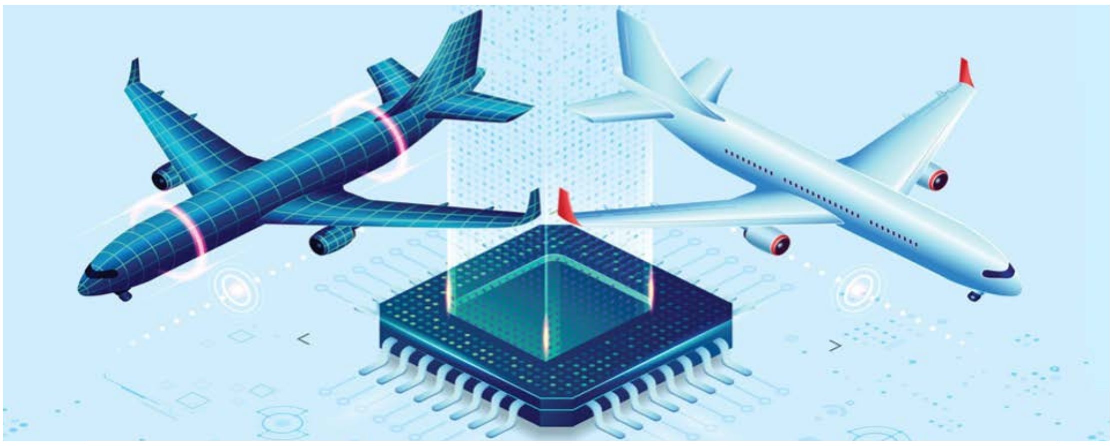


Figure 2.2 Digital Twins technology - source: McKinsey Company

Physical objects will be linked to a virtual real-time mode through connecting its real-time data through digital representation programmed by mathematical models through artificial intelligence and pattern recognition (Vessali et al., 2022).

This replica will be constantly updated which makes it a living object that will let the user analyse data, monitor systems, and make simulations the same way as doing it with a physical object. This will be applied also on smart cities which will be using sensors, drones, and mobile devices to collect data from built environments including data for vehicles, buildings, infrastructure, and individuals. This will be through devices like Internet of Things (IoT) and augmented by usage of artificial intelligence through advanced analysis (Vessali et al., 2022).

Digital twins will help decision-makers to design more resilient and sustainable designs which be through choosing the optimal planning operations and will also help in cost savings. This is needed in new whole cities that are being developed also in Middle East where new five urban areas are going to be developed in Dubai by 2040. Saudi Arabia also have new major cities as Neom, Al Ula, and Amaala will be developed by 2030 (McKinsey Explainers, 2023).

These countries such as UAE and KSA for example needs to be planning of time for all scenarios and constantly developing their new cities due to the high increase in population.

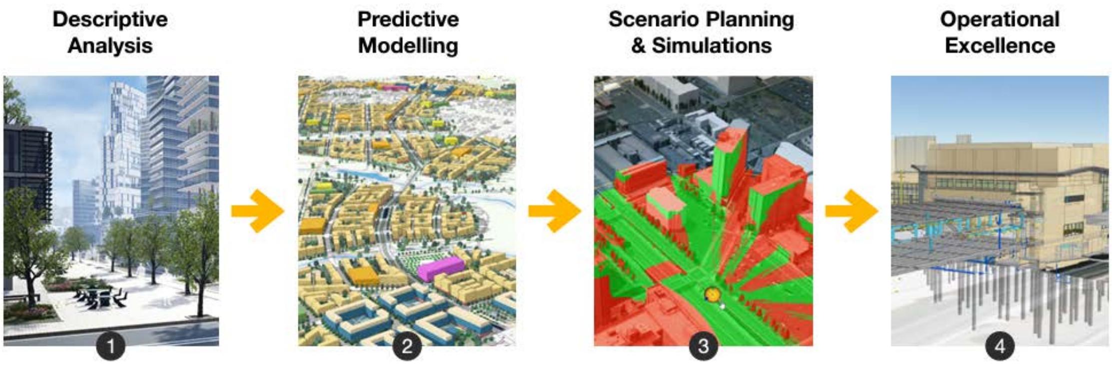


Figure 2.3 Digital Twin solutions go through various levels

Digital Twins will be going through several steps such as, (i) descriptive analysis, (ii) predictive modelling, (iii) scenario planning and simulations and finally (iv) operational excellence. This will require dynamic and static data analysis while needing also 2D and 3D representations of situations in these cities. Then the model will be through augmented reality (AR) virtual reality (VR) and mixed reality (MR) (Vessali et al., 2022).

The figure below shows a clear vision how digital twin will be linked later to metaverse. This will have a huge impact on all private and public stakeholders because they will be using all these data and processes through digital twins to understand the decisions through a clear vision (Vessali et al., 2022).

Figure 2.4 Digital twin and Metaverse source: pwc

One of major differences that digital twins will be showing is through traffic planning due to high amount of data that is available concerning this field. This will help-out in finding more parking spaces and provide deep analysis for revenues generated in cities. This will also be studying deeply through data provided the expected trips in cities from users. Moreover, they can be used lso in energy reduction since they can estimate building's energy usage and provide solutions to property owners so that they make the building net-zero as possible. This will also increase citizen's engagement in the city which makes the society more aware to the technology that is around and make the educational level higher (Robare, 2023).

This struggle in urban cities due to rapid growth in population is highly also increasing the amount of data that are being mined. Thus, this will be through combining traditional tools in urban planning such as 2D and 3D with dynamic virtual models of the cities in real-time with huge amounts of data. This will make digital twins having a direct impact on multi-criteria spatial decision support system (MC-SDSS) also since it will affect the process while making decisions in smart cities (Yi, 2022).

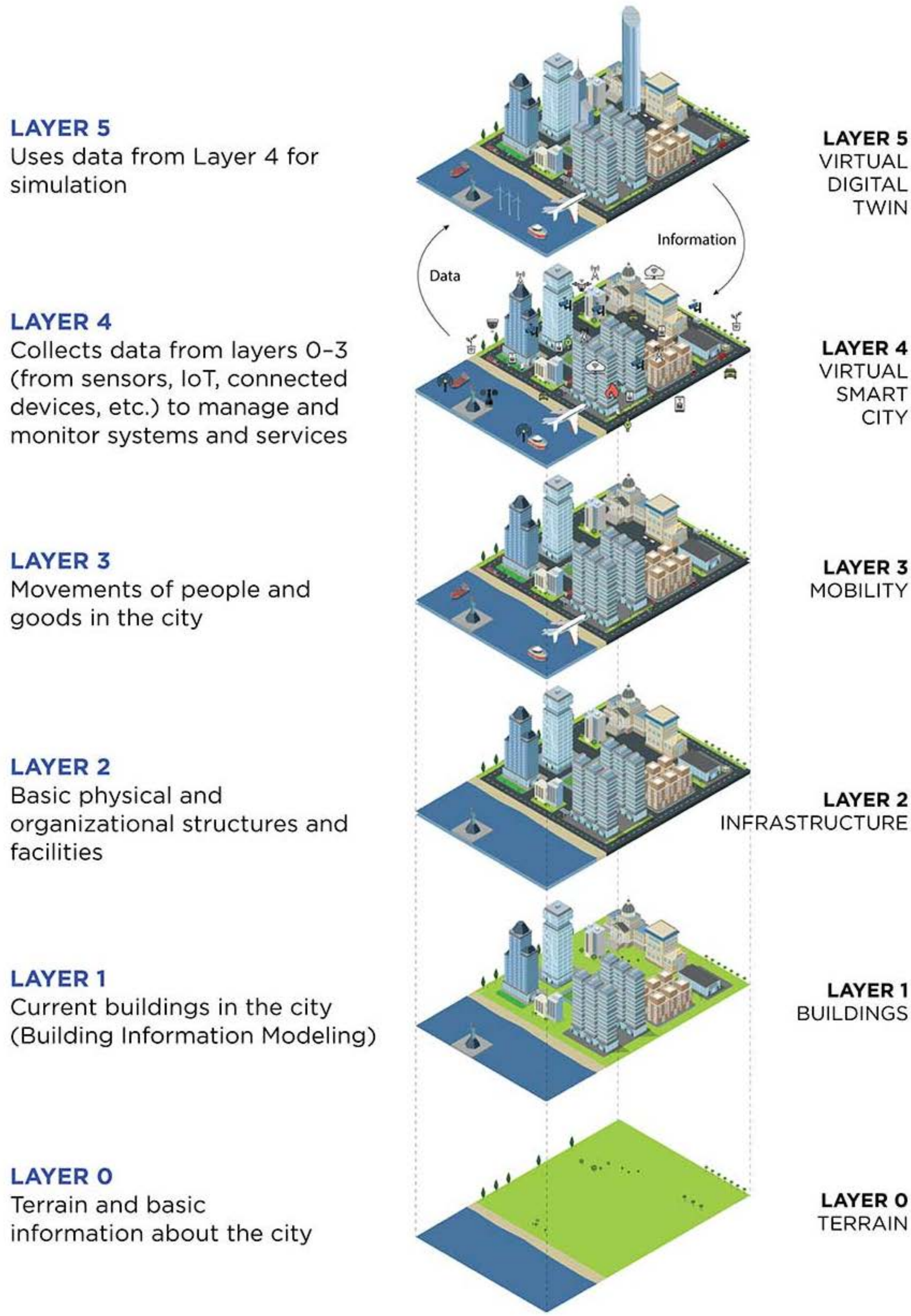


Figure 2.5 A digital twin smart city for citizen Feedback - source: https://www.planning.org/planning/2021/spring/smart-city-digital-twins-are-a-new-tool-for-scenario-planning/

In Figure 2.5 it is shown how they use different layers to build this digital twin city through using terrain and basic, then buildings, then infrastructure, mobility which will allow them to have virtual smart city and then based on these sensors that are available, they will be able to build virtual digital twin for the city (Hurtado & Gomez, 2021).

2.2 Artificial Intelligence in Cities

2.2.1 Definition of Artificial Intelligence

Figure 2.6 A.I robotic arms Source:https://www.fundcalibre.com/artificial-intelligence-making-the-most-of-lockdown

Artificial intelligence (AI) is also known as machine learning because of its sophisticated methods for analysing vast volumes of data to learn a specific task. By resolving issues that were previously exclusive to human intellect, artificial intelligence (AI) research can make a technology intelligent. Artificial Intelligence is used in many scientific fields and in our daily lives. For example, radiologists use it to calculate the volume and shape of tumours, astronomers use it to determine the positions of planets in solar systems, and we use it daily as a powerful search engine for the most basic questions we have. AI offers countless possibilities, including the ability to create fresh strategies for combating climate change. Moreover, AI is a science with multiple approaches which let the AI be created through machine learning and symbolic learning that creates a shift in every sector of tech industry (Laskowski, 2022).

2.2.2 How Artificial Intelligence is Used?

AI is based on four different approaches that the AI is defined by them:

a. Humane Thinking
b. Rational Thinking
c. Humane Acting
d. Rational Acting

AI is the loop between thinking, action, and perception that create the above-mentioned approaches which are defined the first two as reasoning and processes while the second two deals with the behaviour. (Laskowski, 2022)

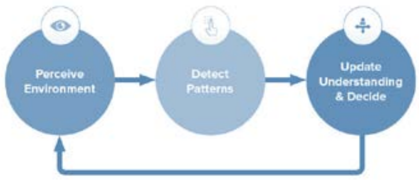


Figure 2.7 How A.I is used? Source:https://su.org/resources/exponential-guides/the-exponential-guide-to-artificial-intelligence-2/

2.2.3 How AI works?

AI works through two main system approaches: Machine Learning and Symbolic Learning which is basically the reason how AI exist.

a. Machine Learning:

It is the ability to let a machine improve from experience without being programmed to that specific action. That is a direct application of AI which focuses on development of computer programs. Observations and inserted data are all needed for the process of learning to begin, direct experience and instruction will be able to let the machine look for the pattern to make in the future better decisions. This will allow the computer to learn automatically without human intervention(Expert.ai Team, 2022).

b. Symbolic Learning: It is based on enhancing a mental blueprint of a coding system to what should be done so that an action is completed. The human knowledge and behaviour will be embedded into a computer program. To define symbols and manipulate them, they developed tools for that. The rules are clear cut and can be easily obtained as an input of symbols. Giving the program info and it will be able to compare between the data you gave it once and the data you want to get as an output is the symbolic learning of the AI. That is linked directly to the deep learning where multiple abstraction layers make a neural connection together to have pattern recognition and the enormous data will be classified correctly (Dickson, 2019).

2.2.4 History of Artificial Intelligence

AI started in 1943 as Warren McCullough published "A Logical Calculus of Ideas Immanent in Nervous Activity". It was a proposal for the first mathematical model to build a neural network. After several years, Donald Hebb published in 1949: "A Neuropsychological Theory" which discussed that neuron are strongly connected and the more they are used the stronger they are connected. The main problem they were facing is that computer in that stage could not store data to have a memory so that it remembers what it did (Alyssa Schroer, 2024)

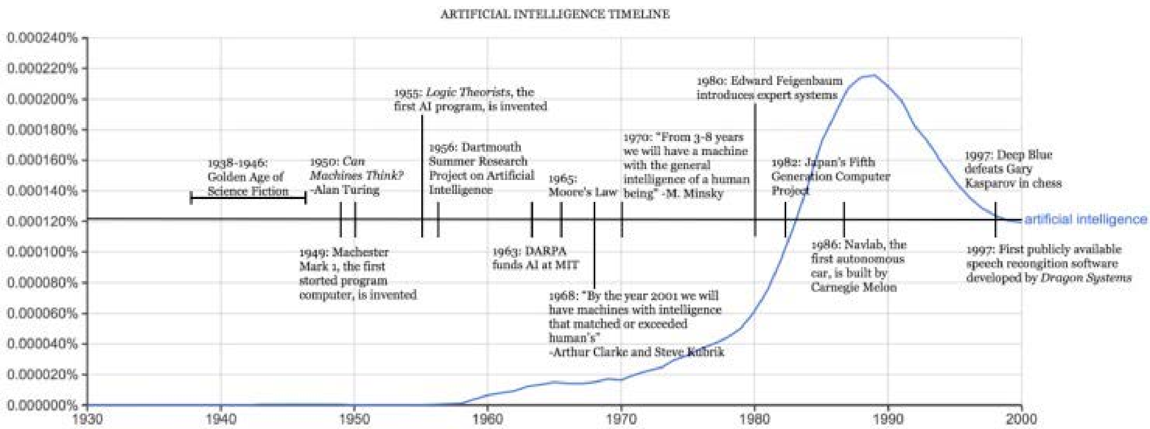


Figure 2.8 Historical timeline of AI Source: http://trends.e-strategyblog.com/2017/06/10/history-robotics-artificial-intelligence-infographic-2/28454

After that period from 1957 to 1974, AI started to flourish more. People started to distinguish which algorithms they should be applying to solve their issue. Newell and Simon's started to develop the spoken language interpreting. They even convinced the government to start funding the AI research because they were highly interested in a machine that can translate spoken language. After that in 1970, Marvin Minsk was interpreting that after 3 to 8 years they will have a machine that is able to get a general AI of a normal human intelligence. Yet as McCarthy agreed on that the computers at this stage are still very weak to be able to exhibit intelligence (Anyoha, 2017).

In 1980, the expansion in toolkit of algorithms and the fund boost AI was reignited. David Rumelhart and John Hopfield started to let the computers use "deep learning" by inserting the experience to it. In addition, Edwar Feigenbaum started to introduce the decision-making capability as a human expert through expert systems which was able to respond based on a given situation that were used mostly in industries. Japan government started to fund the expert system as the "Fifth Generation Computer Project". 400$ million dollars were invested in 1982-1990 to be able to revolutionize the computer process and implement in it the logic programming, and to improve the artificial intelligence. After that, from 1990 to 2000s the AI was able to thrive without government funding. An IBM's Deep Blue defeated Gary Kasparov as the chess master in playing program. They implement Windows as a software at the same year which was an extra reward for the spoken language interpretation (Kistler, 2019).

Now we live in an information age which is based on data. We no longer have a problem in the computer that makes it unable to memories its actions and what they suffered from 30 years ago. Which was able to let the artificial intelligence become super intelligent. Right now, it is no longer hard to have two different languages translated on the spot through AI. Driverless cars are also a technique they are studying. AI in the future will be capable to exist through all the industries so that they will be able to transform any device to a smart machine that is able to interpret and give solution and perform repetitive tasks (Anyoha, 2017).

2.2.5 Artificial intelligence implemented in cities (Urban A.I)

There are certainly innovations that are affecting the way we live such as artificial intelligence and Internet of Things (IoT). Cities will enhance the quality of lives for residents through its alteration to dynamic and responsive entity. IoT sensors can be employed into energy usage and smart meters to enable continual monitoring and analysis which will provide optimized energy consumption and will lead to reduced costs. For example, in Barcelona, they implemented around 20,000 smart meters to control irrigation and water levels which led to 25% increase in water conservation and saved around 555,000$ per year (Ly, 2023).

Smart city innovations can be seen also through smart poles which offer lighting, wireless connectivity, and environmental monitoring through diversity of sensors and cameras that can have data and transmit it in real time through high-speed internet access. This will improve how city administrators will be able to make decisions based on real data. Moreover, artificial intelligence will also have a huge impact on transportation field through having smart infrastructure and achieving autonomous vehicles, this will require advanced sensors and other smart city devices to create an interconnected framework so that the vehicles will perceive their environment accurately (Ly, 2023).

The usage of information to improve quality of life and reduce the operational costs for cities is the concept of urban artificial intelligence. This term is found by Hubert Beroche in 2021 and is a Paris-based think tank while he explored 12 cities and try to understand the impact of artificial intelligence on cities. Water management, urban logistics, and infrastructure maintenance will be optimized through artificial intelligence to make more efficient cities (Popelka, 2023).

Urban A.I will affect the whole community as researchers, public servants, and all stakeholders in urban community. Artificial intelligence models are highly effective for future urban environments to predict which infrastructure failures will happen before they occur. This makes if safer for urban residents and save costs for infrastructure owners. Thus, urban mobility for example will have a better communication between mobility providers, governments, and citizens because of all the collected data as shown in Figure 2.5 the heat map that shows the most used areas by micro mobility vehicles (Popelka, 2023).

Figure 2.9 Micro mobility usage areas heatmap Source: http://trends.e-strategyblog.com/2017/06/10/history-robotics-artificial-intelligence-infographic-2/28454

They have used smart city technologies and A.I to promote "service-oriented cities" as a bottom-up model. Citizens will be serving as customers and co-producers since they have impact on systems that are constructed to serve their needs as producing data for A.I to collect them and analyse them to provide these services (Popelka, 2023).

There are several components to develop a smart city as shown in Figure 2.6. A.I will help smart cities to apply 30% of its applications by 2025, these applications are through seven dimensions.

2.2.6 Applications of Artificial Intelligence:

a. A.I for governance which uses machine learning in legal frameworks and related to justice. Algorithms will be used to predict patterns and be able to increase accountability. Yet governance is seeking to assess the safety of A.I, increase transparency and define rules to control personal data without harming individuals.

b. AI for safety, security, and healthcare will be used in healthcare to predict medications needed for patients and be able to make accurate decisions based on each case separately.

c. AI for education and citizen participation will make users interact with A.I platforms that can have quick answers for their questions such as Chat-GPT that will help in education. In addition to involving people in learning about their own consumption in energy sectors for example and make them participate in creating a sustainable community.

d. AI for economy due to the cost saving that is led by the monitoring of energy consumption and production. These monitoring of real time data will help us save costs and let governments invest in other projects that will benefit the whole community.

e. AI for mobility and logistics due to having autonomous vehicles due to the smart infrastructure that can be making fast smart connections between vehicle-to-vehicle and vehicle-to-streets. This will increase the safety measures in driving with less percentage of making errors and accidents. Which will also decrease the death rates because of car accidents.

f. AI for infrastructure to collect all data through fast internet networks. This will make collecting real time data available and companies will be able to collect, analyse data and predict future scenarios, which will help decision making process for governments.

g. AI for the environment due to all the smart systems that can be implemented to create a zero-carbon emission community through the sensors that monitor the consumption of data. Reducing CO2 and greenhouse gas emissions will benefit the environment and achieve all the sustainable development goals that should be achieved within 2030.

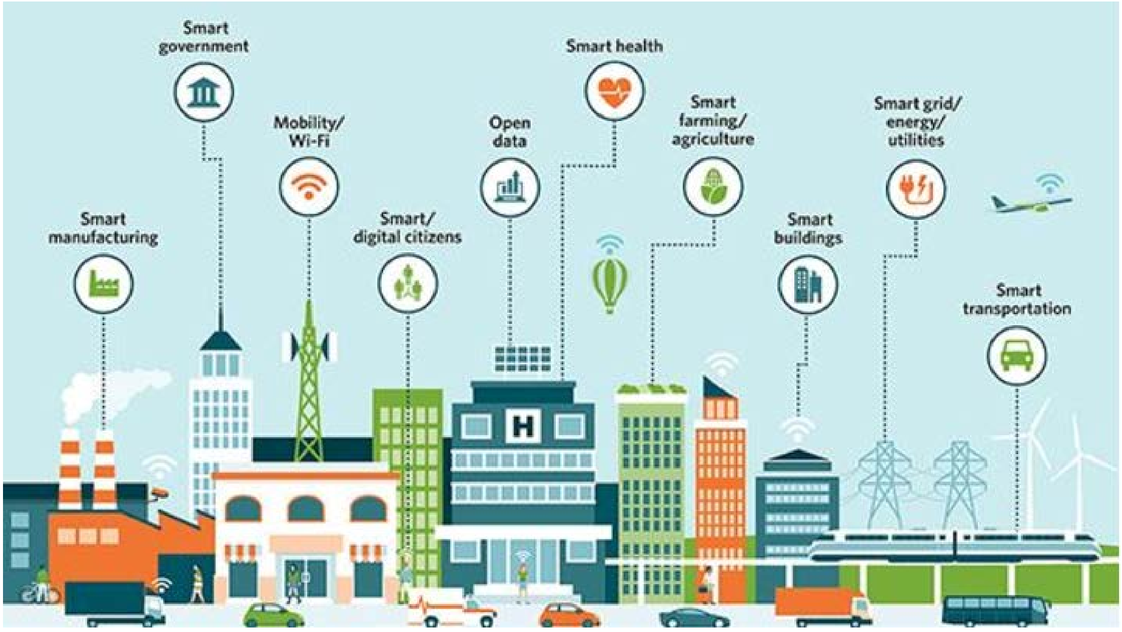


Figure 2.10 Smart city components Source: http://trends.e-strategyblog.com/2017/06/10/history-robotics-artificial-intelligence-infographic-2/28454

To transform conventional networks and services, smart cities will be able to use this innovative technology. For example, A.I can enhance smart urban lighting and smart waste management since EU cities have up to 90 million streetlights and 75% are older than 25 years. Through the usage of A.I, the cities were able to use public lighting via remote monitoring and control data on these streetlights to know the air quality, temperature, wind speed, pedestrian flows, and humidity. These innovations can be found in cities as Barcelona, Munich, and Copenhagen. Through COVID 19 pandemic, these sensors helped residents to know where they can spend their time while respecting social distance measures (European Parliament, 2023).

Additionally, usage of sensors on containers will be able to detect the filling rate and this will develop a Smart waste management system. To stimulate efficient waste production and discarding, patterns of how, when and where waste is discarded can be detected by A.I. Rotterdam city have this innovative solution for its residents to make it also efficient for autonomous vehicles that are used for waste collection. A.I can be also used to separate waste collection (European Parliament, 2023).

2.2.7 Artificial intelligence transparency issues

There is another perspective around using artificial intelligence which is developing trust between residents and the government because it will affect the whole community on social level. Citizen's trust may be able to increase towards the governments and government's decisions within the usage of artificial intelligence or decrease it. Due to a violation of citizen's privacy or lack of fairness in A.I for services this trust may decrease. Additionally, because of lack of transparency of black-box systems challenges may also arise through having unclear responsibility and accountability in decision-making. Usage of A.I with lack of transparency may have a strong negative impact for governments and society which will lead to unhealthy communities (Zuiderwijk et al., 2021).

Based on several forefront legal discussions on artificial intelligence, the lack of transparency has been a serious issue that should be addressed. Governments are putting pressure to mount the design and govern A.I to be transparent, fair, and accountable (Rodrigues, 2020). Mitigating possible unwanted results from automated decision-making is the reason why they are trying to achieve transparency. There are a lot of people who were denied jobs, refused loans, people who are denied flying anymore without knowing what really happened while the algorithms took that decision without justifications (Desai & Kroll, 2017).

Based on Evert Haasdijk, a senior manager Forensic at Deloitte, people are afraid of A.I because it's hard to be explained how Artificial intelligence works and that's the reason it's not transparent, especially data-driven technology like machine learning which is used in public sectors since it's the relation between input and output data which can cause imaginations to run wild. Thus, usage of A.I in public sector based on several studies may be able to threatens the transparency of public sectors which will be able to also decrease the trust of people in the governments (Deloitte, 2019).

A human designer of a program may be able to have bad intentions so he will pursue to differentiate, suppress speech, or engage in other illegal acts through the interference of A.I in data value cycle in Figure 2.7. Thus, transparency will be able to police if a developer might want to have bad intentions and differentiate suppress speech or engage illegal acts through behaviours in algorithms (Desai & Kroll, 2017).

When developers want to justify how A.I work, they have a lot of issues explaining it due to its complex system. The system is a black box which makes the developers don't know how machine learning algorithms gets from point A to B. They trust the results based on historical accuracy and previous probabilities not because of a verified formula that could be audited. Thus, most developers know that transparency in A.I is critical, yet they still hope to achieve transparency (Mind AI, 2018).

The black box in artificial system is due to its complexity, difficulty of developing explainable solutions, and risk concerns. Even if all the parameters are known, the operation in artificial systems are based on neural networks which depend on complicated interactions which will make it not possible to comprehend how the task is done. It's also difficult to create a user experience which can understand how the procedure of the task happened. This might make it impossible for the user to know how the results are reached. In addition, the simpler the system, the easier the hackers can attack it, cause the system to malfunction and come up with unwanted results (Holland Michel, 2020).

Artificial intelligence usage in public sectors will threatens the transparency because if the algorithm is easy to read and simple without sophistications so that they can know how the system work, the hackers may be able to hack the system easily and all the data will be accessible illegally, thus all the data of citizens will be at risk. If the public sectors want to create a strong system that hackers can't access, they should use sophisticated algorithms which are hard to be explained to citizens since it's a black-box system, though the accuracy of its predictions may rise (Zerilli et al., 2022).

Due to high difficulties in translating algorithmically derived concept into understandable human concept, this will make it hard for the public sectors to explain why the system reached that result. Moreover, publishing the exact algorithm may not be the solution because the users will not understand the algorithms, and that would harm the whole system due to making it possible for the hackers to access the algorithms and access the data for all citizens. Thus, using artificial intelligence will threatens the transparency because it will also threaten the data of the users which will decrease the trust in the governments (Holland Michel, 2020).

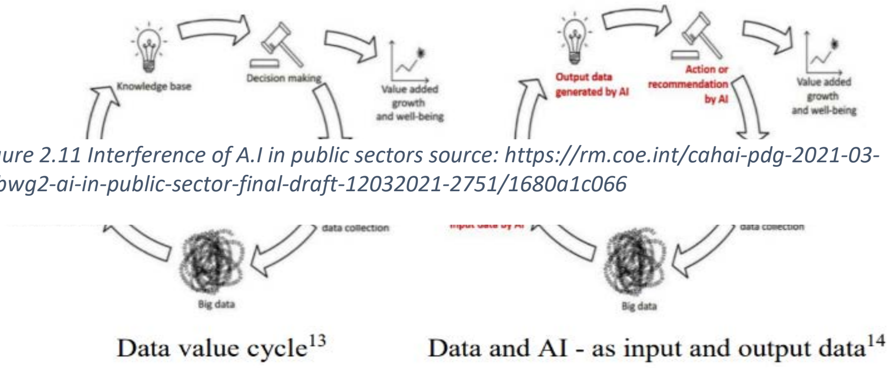


Figure 2.11 Interference of A.I in public sectors source: https://rm.coe.int/cahai-pdg-2021-03-subwg2-ai-in-public-sector-final-draft-12032021-2751/1680a1c066

03

Methodology

3. Chapter 3: Methodology

Through this chapter, it will show the design and methodology of the thesis research. It is important to use artificial intelligence in cities as mentioned in chapter 2 through several methodologies while trying to achieve different tasks. The aim is to implement the usage of artificial intelligence to predict futuristic scenarios and choose the most optimized scenario through energy consumption datasets.

3.1Methodology Framework

In this research there are two case studies. The first one is R10 universe project in Pune, India, while the other one is Sei Milano project in Milano, Italy. Both projects' have the same methodological framework approach that consists of 3 different phases and different approaches to use artificial intelligence. For each case study there are three main phases to achieve three main objectives explained in section 1.3. The **first phase (1)** is the "Collect and Analyse Data", which includes two steps for each case study. The selection of indicators through indicators importance and collection and analysis of data for both case studies. Both indicators' selection in both case studies is based on indicators' importance yet the collection and analysis of data for the first case study in Pune is from sensors on site construction but for the second case study in Milano is from database.

The **second phase (2)** is the "Assessment and Evaluation", which includes one step for each case study. For the R10 universe project is to predict futuristic energy data based on artificial intelligence, while for the second case study sei Milano project is to create indicator impact assessment framework proposal through artificial intelligence too.

After that, the **last phase (3)** is the "Recommendations". It has a qualitative approach to elaborate the final recommendations. This final phase gives recommendations for companies in order stay in continuous improvement allowing them to adapt and evolve its operations in response to changing factors around them.

It is important to break down the study into different phases and steps to understand the research process for each case study. Figure 3.1 shows a schematic flowchart of the methodological framework of this research based on these two different case studies.

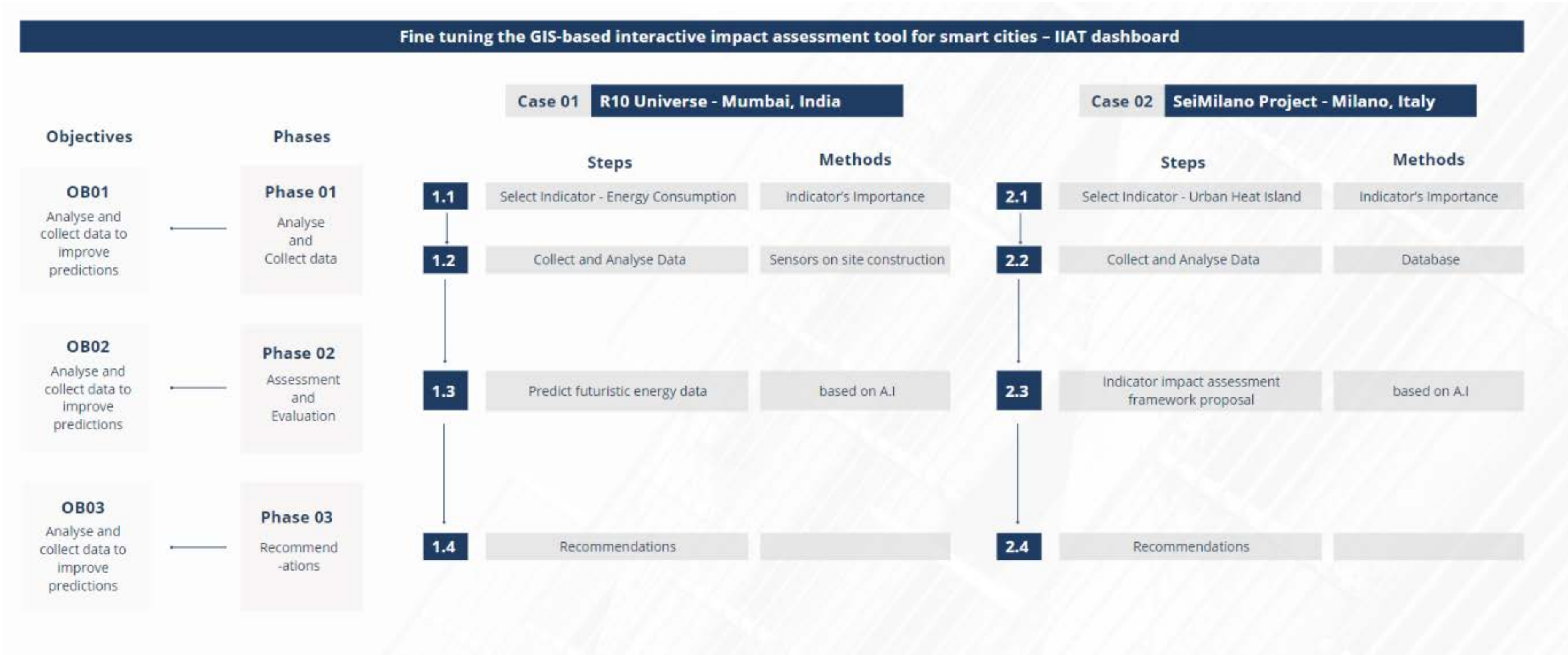


Figure 3.1. Methodological Framework – Source: by author

3.1.1 First Phase: Collect and Analyse Data

The aim of the first phase is to collect high-quality data for R10 Universe and SeiMilano Projects, after selecting the indicator needed to be studied on. There were several indicators as shown in Table 3.1. that are previously chosen by Politecnico di Torino and Planet company in a previous project based on multi-criteria spatial decision support system (MC-SDSS). Firstly, to achieve the goal and select the best indicator, documents from Planet company concerning the projects and company was analyzed. They provided the characteristics of the indicators and how they can be calculated. Then based on indicators' importance due to Planet company as a main stakeholder for this study, the indicators were chosen for each case study separately (STEP 1.2 and STEP 2.1).

Table 3.1. Selection of indicators – Politecnico di Torino

Category	Topic - Indicator
1. Energy & Climate change	**Embodied Energy of residential buildings**
	Primary Energy demand of construction materials and components (A1-A3)
	Primary Energy demand of materials transport (A4)
	Primary Energy demand of construction and installation (A5)
	Embodied Renewability Index
	Embodied Carbon of residential buildings
	CO_{2eq} Emissions of construction materials and components (A1-A3)
	CO_{2eq} Emissions of materials transport (A4)
	CO_{2eq} Emissions of construction and installation (A5)
	CO_{2eq} Emissions compensation or Credits Purchased
	Operational Energy of Residential buildings
	Primary energy consumption of residential buildings
	Thermal comfort
	Operational Carbon of Residential buildings
	CO_{2eq} Emissions of residential buildings
	Renewable Energy
	Renewable energy produced on site
	Renewable energy exported
	Urban Heat Island
	Urban Heat Island mitigation
	Smart Energy Management
	Energy management (construction)
	Energy management (operation)
2. Water	**Water consumption (embodied)**
	Water consumption of materials and components (A1-A3)
	Water consumption of construction phase (A4-A5)
	Water consumption (operational)
	Water consumption for domestic use
	Outdoor water consumption
	Potable water consumption for non-potable use
	Water consumption of non-residential areas
	Water recycling
	Rainwater harvesting
	Greywater recycling
	Smart Water Management
	Water Management (construction)
	Water Management (operation)
3. Habitat & Biodiversity	**Site and Land use**
	Land Evaluation and Site analysis
	Net Land consumption (artificial surfaces)
	Urban Greening Factor
	Green infrastructure quality and Environmental regeneration
	Sustainable Drainage Systems (SuDS)
	Canopy cover
	Plant diversity
	Smart Environmental Management
	Environmental management (construction)
4. Resources & Waste	**Sustainable procurement**
	Sustainable Materials and products
	Suppliers and Contractors
	Waste Production
	Construction waste
	Recovery and Recycling
	Recycled materials use
	Smart waste management
	Design for disassembly
	Waste management (construction)

Secondly, after selection of indicators, it was possible now to start collecting data for both case studies yet in different methods. For the first case study which is R10 universe, the collection of data was from sensors on site construction then they were analysed in a proper way. Then another project was chosen as a benchmark that is named PJME project in United States. The data for this project was having the same process for R10 Universe project to see the difference in results at the end. To reach the goal for this step, it was important to choose which method should be used to analyse the data collected properly so that they are ready to be used for the next phase (STEP 1.2).

There were several methods that can be chosen for analysing the collected data:

1. **Time series analysis:** used to analyse the temporal patterns of data and trends in data too. It understands the seasonality, trends, and patterns through time. Which is important to make predictions, detect anomalies and optimize the data performance.
2. **Regression analysis:** used to identify the relationship between several variables such as temperature and humidity.
3. **Monte Carlo simulation:** Used to get models of possible outcomes and probability distribution through a computerized technique, it is usually used when making decisions.
4. **Factor analysis:** Decrease number of factors if there is large number of variables through multiple separates to group multiple related factors together.
5. **Cohort analysis:** Analysts can track their behaviours based on trends and patterns after grouping data based on common characteristics.
6. **Cluster Analysis:** Used also for identifying patterns and segments in data due to studying similar points from cluster analysis groups.
7. **Sentiment analysis:** it is linked to textual data; it sorts and understand emotions from textual data. It is separated into three main types of analysis which are fine-grained sentiment analysis, emotion detection, and aspect-based sentiment analysis. (Stevens, 2023)

Based on the data collected and based on the quality of data provided, the proper method was chosen to analyse the data. The data was analysed through python programming language and using anaconda Jupyter notebook that have all the necessary libraries to analyse this data.

Meanwhile, for the second case study which is Sei Milano Project, the data is collected after the indicator's selection from the database of Planet company, then the site was studied, and smart artificial intelligence solutions was provided, as well was stakeholders were studied concerning this project. Then they were linked to SDGs and their effect (STEP 2.2).

3.1.2 Second Phase: Assessment and Evaluation

3.1.2.1 Pune, India: R10 Universe Project

The aim of the second phase is to use the collected data for R10 Universe project and predict the futuristic energy data through usage of artificial intelligence. To reach this goal it was important first to choose which method is proper to create a theoretical framework to predict futuristic energy data later. There are several methods available to predict futuristic data that was collected previously in phase (1):

1. **Exponential Smoothing State Space Models (ETS):** based on error, trend, and seasonality and useful for capturing patterns in time series data and forecasting. Used efficiently for a simple model.
2. **Long Short-Term Memory (LSTM) Networks (Deep Learning):** effective for complex temporal data and usually requires a lot of complex data. They can be powerful for predicting data, but they can be computationally expensive. There might be a risk for overfitting in case of limited amount of data.
3. **Seasonal-Trend decomposition using LOESS (STL):** it also uses trend, seasonality and remainder components and used for predicting futuristic patterns but need huge amounts of data and not good in non-linear data.
4. **Theta method:** it uses average of past observations and is usually a simple technique to be used but might not capture complexity in data due to its simplicity.
5. **Prophet:** is developed by Facebook to predict data over time and requires minimal data preprocessing. It uses trends, seasonality, and holidays to forecast data.
6. **Neural Prophet:** is an extended model of prophet and designed by Facebook but it needs huge datasets to make the predictions and forecasts efficiently and might not be accurate with limited data.
7. **ARIMA (Auto-Regressive Integrated Moving Average):** deals with past values over time and suggest future values based on that. Usually needs stationary data to work well.
8. **SARIMA (Seasonal ARIMA):** is an advanced model of ARIMA but needs larger data points to estimate the predictions in high accuracy.

After studying all these methods to predict collected data and achieve the goal for phase 2, two models were chosen from this list based on data characteristics. Then the results of these two models will be compared to know which model is best to achieve this phase at its best. The data was predicted through python programming language and using Anaconda Jupyter notebook that have all the necessary libraries to analyse this data.

3.1.2.2 Milano, Italy: Sei Milano project

The aim of the second phase is to use the collected data for Sei Milano project and create an indicator impact assessment framework through usage of artificial intelligence. To reach this goal it was important first to create the theoretical framework for this project for futuristic projects in Planet to work on. Firstly, the indicator's equation was studied, the research included all needed data to achieve this equation. Then a literature review part was needed as a qualitative approach and seek to find smart solutions. Then the theoretical framework was developed based on how these smart solutions can be implemented in further projects.

3.1.3 Third Phase: Recommendations

The aim for the third phase was to prepare recommendations for both case studies so that Planet company can take advantage of these recommendations and use them in further projects. These recommendations help to achieve the main objective of this research study based on the outcome from previous phases. This phase is connected to 2030 Agenda for Sustainable Development Goals that must be achieved by companies for the development of developing countries around the world.

Thus, the recommendations are designed based on these three points elaborated below:

1. Name of recommendation
2. Linkage with 17 Sustainable Development Goals (SDGs) set by United Nations General Assembly

3. Description of recommendation

3.2Case Studies

The Study Site of this thesis is R10 Universe project in Pune India and the second case study is Sei Milano project in Milano, Italy. R10 Universe project is in Pune, India which is a city with a huge population in western Indian state of Maharashtra as it is considered the second-largest city after Mumbai in Maharashtra. It is highly dense and it's an important financial centre in India since it's the largest city there. Its area is 331 km2 and its population is now estimated to be 7.5 million people in 2023. The main transportation in Pune is public transportation that has buses and suburban railway services, and there is an under-construction metro rail system in 2022. While the other project is in a metropolis in northern Lombardy region, and a global hub for fashion and design. It has the gothic duomo which is a node in that area and an important landmark for all people who visit Milano. The population is over 1.352 million and has 181.8 km2 area.

There are several factors that classify if a country is a developing or developed country such as economy, society, and education. There is also the gross domestic product (GDP), human development index (HDI) and healthcare access are all important factors to be considered while classifying.

Italy is an innovative country with its 26th rank out of 132 countries worldwide in global innovation index 2023. In addition, Italy has several reasons why it is a developed country: (i) Historical buildings and innovations, (ii) Design and fashion, (iii) Automotive and manufacturing, (iv) Food innovation, (v) Startups, (vi) green and sustainable practices, (vii) cultural art, (viii) tourism. All these aspects are reasons why Italy is an innovative and developed country with a balanced system of economic, social, and environmental developments.
Meanwhile, in Pune, India is considered a populous area. It has improved in several areas, yet it still faces challenges such as: (i) Economy, (ii) infrastructure, (iii) basic services issues, (iv) social development issues like gender in-equality and child labour, (v) income disparities, and (vi) healthcare services are not easily accessible.

3.2.1 Pune, India: R10 Universe Project

This smart city is a new multi-functional neighbourhood. It has 1,775 homes, 160 commercial units, and it is expected to have above 5400 residents. This smart city project has 50 smart solutions implemented by Planet smart city company and has app-enabled security through digital services and social innovation programs. These enable residents to have access to smart solutions through applications on their smart phones such as smart locks, which makes the engagement of citizens in this project is higher.

Figure 3.2. R10 Universe Project in Pune, India

Figure 3.3 R10 Universe Project in Pune, India

This society will be soon updated to be having also intelligent lighting system, smart water management, and an optimised irrigation system and air quality monitoring system which is an infrastructure with IoT-based smart solutions.

Figure 3.4 R10 Universe Project in Pune, India ground floors and shops plan

Figure 3.5 Metro lines that are under-construction in Pune, India - souce: by author

This Figure 3.5 shows the future metro lines that are under-construction in Pune, India which shows the future connections between the area where R10 universe project exists and the central of the city which provides innovative vision for the city which will help in developing more smart solutions due to higher economical raise within that area. The other Figure 3.6 shows the railway map of Maharashtra which also shows the connection between Pune central city and R10 universe project.

B.G. 1.676 m	M.G.	N.G.	TOTAL
5686.02	105.00	407.61	6198.63

DIVISION	COLOUR	B.G. 1.676 m	M.G.	N.G.	TOTAL
MUMBAI (CR)		192.25	—	20.29	212.54
MUMBAI (WR)		351.89	—	—	351.89
PUNE		531.15	—	—	531.15
SOLAPUR		968.57	—	—	968.57
BHUSAVAL		800.46	—	243.90	1044.36
NAGPUR (CR)		958.70	—	34.42	993.12
NAGPUR (SEC)		512.00	—	109.00	621.00
NANDED		810.00	105.00	—	915.00
SECUNDERABAD		132.00	—	—	132.00
HYDERABAD		55.00	—	—	55.00
KONKAN RAILWAY		374.00	—	—	374.00

Figure 3.6 Railway map of Maharashtra - source: https://www.maharail.com/railway-map-of-maharashtra.php

3.2.2 Milano, Italy: Sei Milano Project

This smart city is a new mixed-use neighbourhood in south-west of Milano, Italy. It has 160,000 m2 squared of parkland, 10,000 m2 of shops, 500 new homes, and it is 16 minutes away from piazza del duomo which is close to the centre of Milano.

Figure 3.7 Sei Milano Project – Source: https://www.mcarchitects.it/en/projects/seimilano

The project is dedicated to sport and leisure through its master plan as it to be part of the city. It has good quality of life is provided through integrating a superior infrastructure and smart technologies for residents. The main theme of the project is "garden city" which is the integration of architecture and landscape. Moreover, the concept of the project is to create a relationship between the buildings, the offices, and the landscape which creates a harmony between spaces. Thus, this will be as a huge added value for residents in that place due to taking into consideration their mental health as much as another smart solutions.

Figure 3.8Sei Milano Project – Source: https://www.mcarchitects.it/en/projects/seimilano

As seen in the following figures, it is shown in the perspectives and the masterplan how the landscape is taking most of the site while emphasizing on the arched river that passes through the project and how the buildings are only at the peripheries of the project. Moreover, there are different facilities in this project and most importantly the metro station "Bisceglie" which is only 10 mins away through walking distance and it takes us to the main Duomo di Milano within 16 minutes only which provides an important location for this project as shown in Figure 3.9.

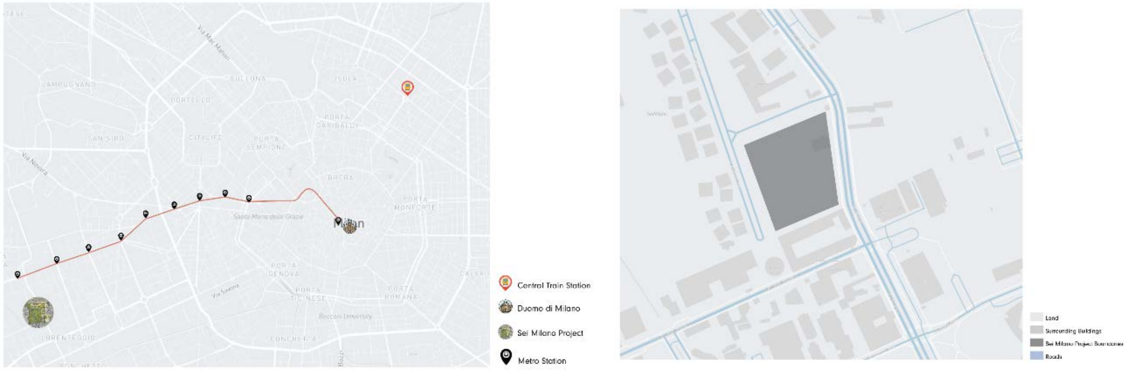


Figure 3.9 Sei Milano Site Analysis - Source: by author

Within the site analysis diagrams shown in Figure 3.9, it shows the different facilities around the project in addition to the main roads which makes the project connected to the major parts of the city.

This project has 216 smart solutions implemented by Planet Smart City company and they are studying more smart solutions that can be implemented through this project to make the resident's life better.

04

Results

4. Chapter 4: Results

Through this chapter, it will show the results that are presented following the main Phases of the methodology illustrated in Chapter 3, namely Collect and Analyse data; Assessment and Evaluation; and Recommendations.

4.1R10 Universe Project in Pune, Italy

Starting from the site analysis of the project knowing that it is in a developing country and after studying the surrounding of it, this section serves the purpose of fine tuning the GIS-based IIAT dashboard through usage of artificial intelligence. This section is divided into 3 sections that shows the whole process for this case study.

4.1.1 Collect and analyse data

The collection and analysis of data is first based on the selection of indicator as STEP (1.1), this step is done through choosing the most important indicator that must be studied in the project for Planet Company based on their preferences and availability of data. Thus, the selected indicator among all the indicators was the energy consumption indicator from energy and climate change section as shown in methodology section 3. In addition, energy consumption is the utilization of energy resources in different activities among the site and these activities are different based on the project's phase and based on the users inside the site.

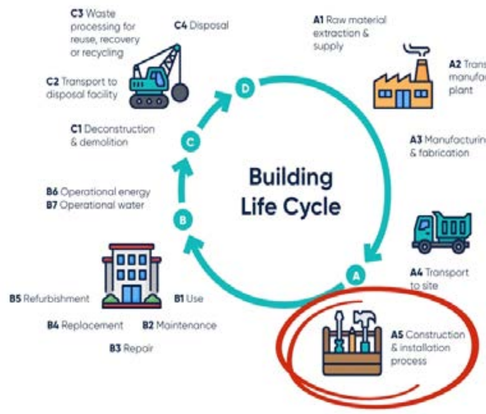


Figure 4.1 Building life cycle – www.naylorlove.co.nz/life-cycle-assessments

As shown in Figure 4.1, the project is still in phase A5 which is the construction and installation process. Thus, the project is still under construction and all the activities for energy consumption in this site is related to this phase.

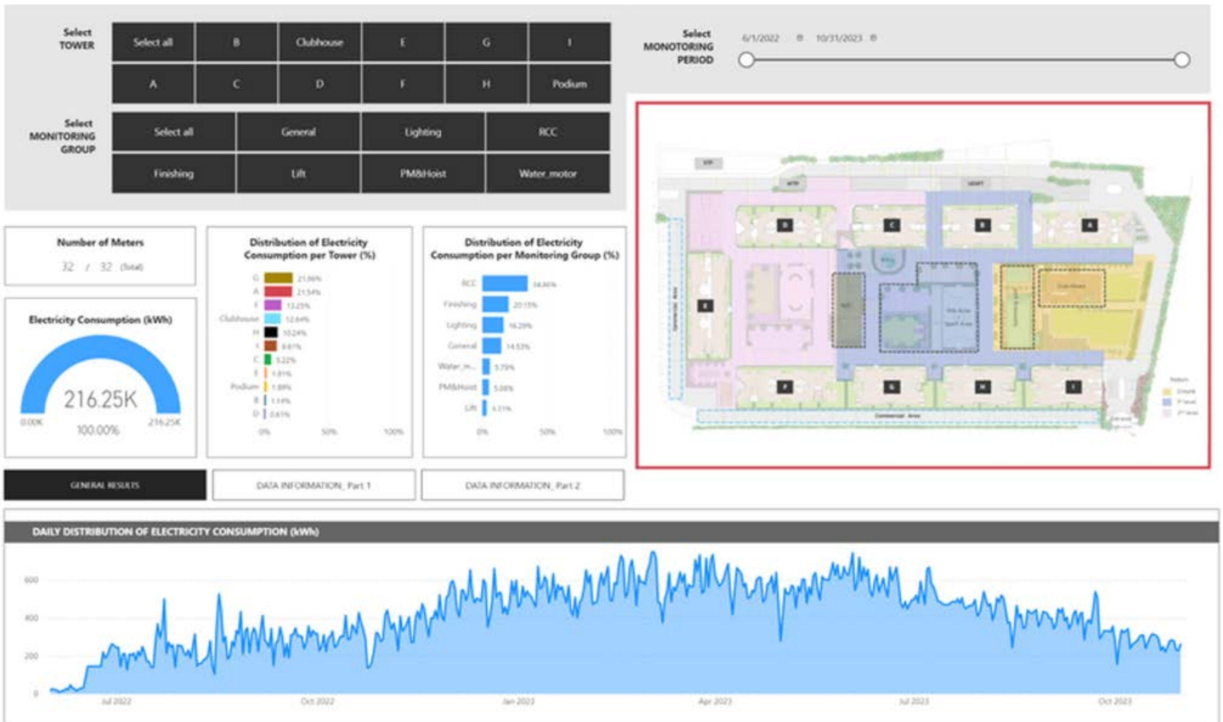


Figure 4.2 Dashboard of R10 energy consumption – source: Planet smart city

Through analysing the collected data from sensors on site construction, the following Figure 4.2 and table 4.1 shows how the data is divided into the site.

Table 4.1. Raw data of R10 energy consumption – source: author

Data	Raw Data	Unit
Tower's Names	A,B,C,D,E,F,G,H,I, Club house, Podium	
Activities	Finishing, General, Lift, Lighting, PM & Hoist, Water Motor, RCC	
Consumption	Energy data of site construction activities	/kWh
Date.Time	Daily and Hourly Data	Hourly

The following Table 4.1 and Figure 4.2 shows that the energy consumption in this project is divided into 7 different towers with names: A, B, C, D, E, F, G, H, I, Club house, and Podium. The activities in this site construction are related to the construction phase which are: (i) Finishing, (ii) General, (iii) lift, (iv) lighting, (v) pm & hoist, (vi) water motor, and finally (vii) RCC. Energy consumption of these activities on these towers are calculated hourly and gathered daily by /kWh.

Moreover, to be able to use artificial intelligence for energy consumption we should be able to predict energy consumption usage through this project within the construction phase which will be a theoretical framework approach for further projects in developing countries and will be explained in next section 4.1.2.

The collected data for R10 universe project is only available for one year (2022 – 2023) which makes it impossible to have great accuracy in the results, thus it is important to have another project as a benchmark for all the results. The reference as a benchmark will be a project named PJMW in U.S which is PJM interconnection regional transmission organization (RTO) for 14 different states in the United States. This project has been collecting energy data for different buildings in these states for the last 16 years from 2002 until mid-2018 which makes it a great benchmark for predicting future energy consumption data as a theoretical framework.

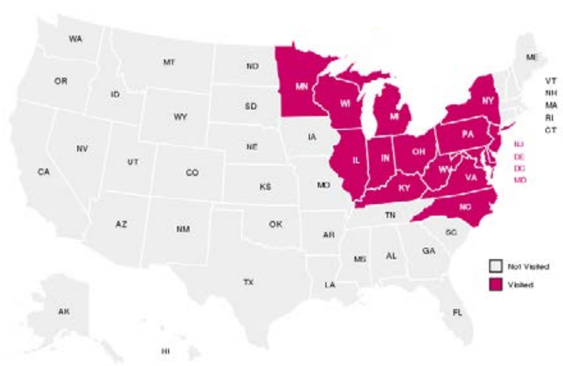


Figure 4.3. PJMW hourly consumption data – Source: PJMW

In the following Figure 4.3 and Figure 4.4, it is shown the 14 states included in this energy collection and the data consumption analysed throughout all the 16 years. Moreover, to be able to analyse all this data as mentioned before in section 3 Methodology, the most appropriate method to be used is Time Series Analysis out of all the mentioned methods above in section 3 since both projects have only one variable which is energy consumption in respect to time (hours), thus it is the most appropriate method to study this type of projects. Both projects also can be normalized from 0 to 1 which will make the comparisons' scale more similar.

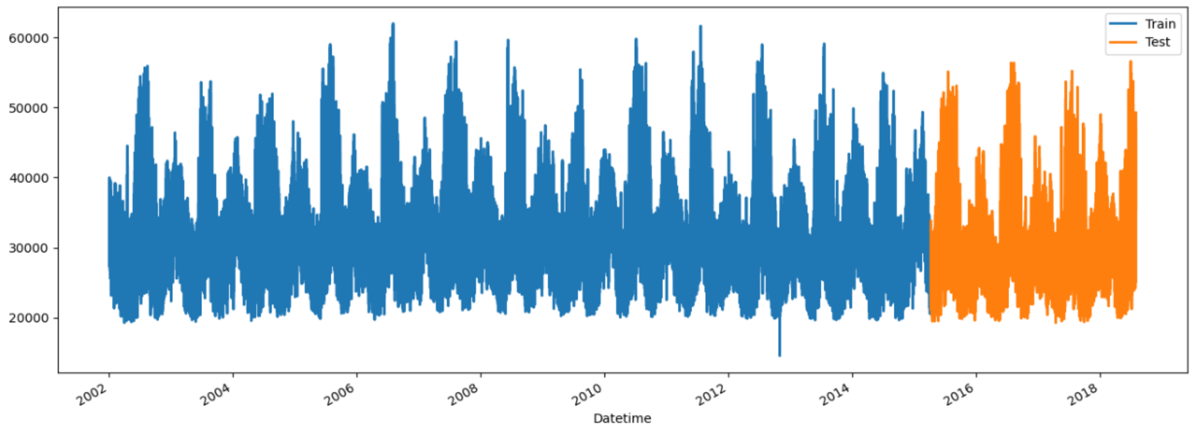


Figure 4.4. PJMW analysis in U.S – Source: author

Time series analysis is a statistical subset method of artificial intelligence (A.I) and is used to study data points that are collected over time as shown in Figure 4.5. It analyses patterns, trends, and behaviours within the collected energy data to make further predictions and forecasting. It is used mainly for forecasting data based on previous data, it is also used to help in decision-making for projects and it makes the MC-SDSS decisions easier to make due to having further predictions on energy consumption for the project. Moreover, it is useful in pattern recognition of energy consumption for users which makes it useful to know what is the possible enhancements that should be made for optimizing this usage. Finally, it is also important for risk management studies to be able to analyse the cost of energy consumption for a project in any building phase it might be.

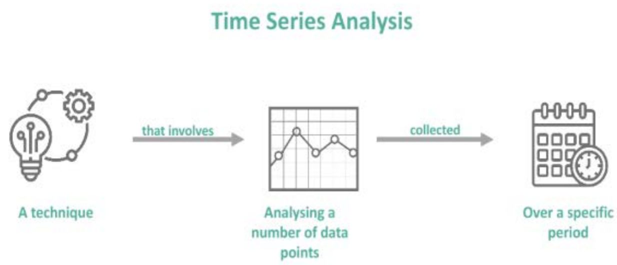


Figure 4.5 Time series analysis – source: KNIME official website

The following figures shows the data collected for both projects R10 Universe and PJMW project after they have been analysed.

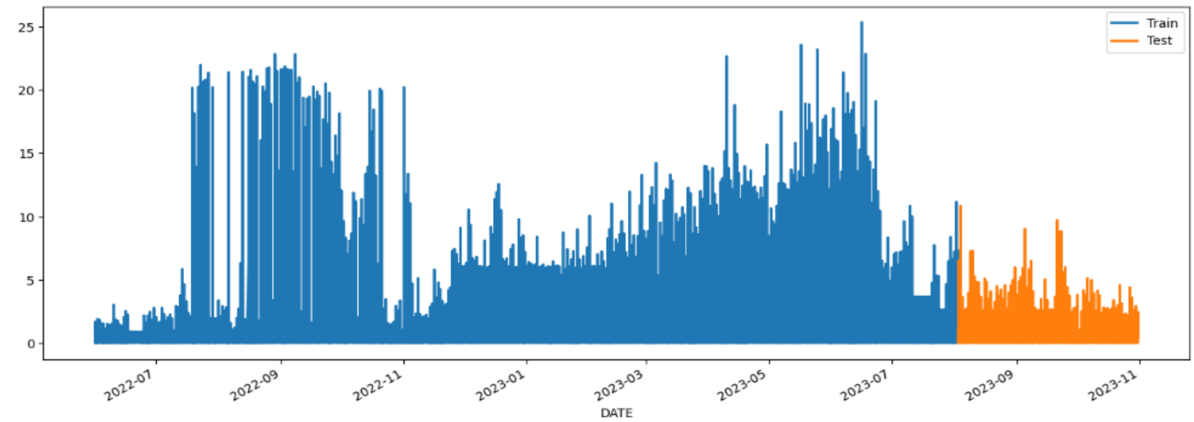


Figure 4.6 Planet all data – Source: author

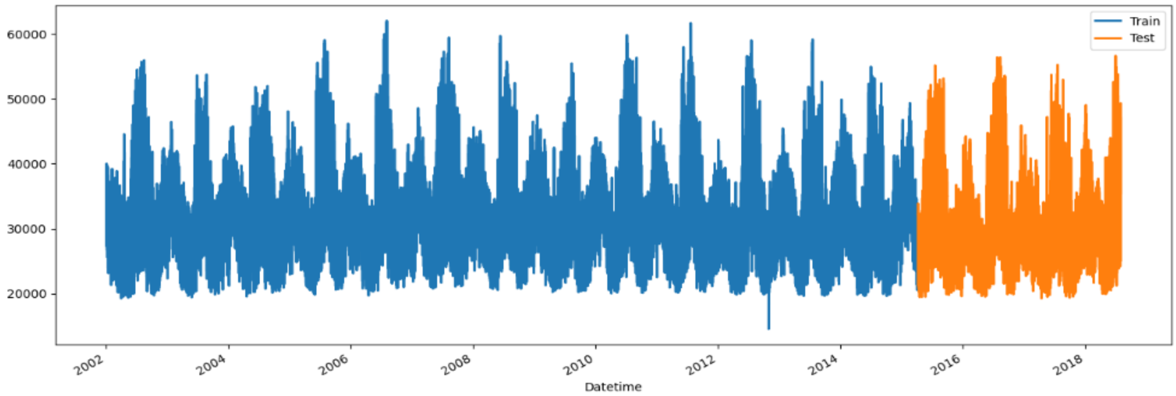


Figure 4.7 PJMW all data – Source: author

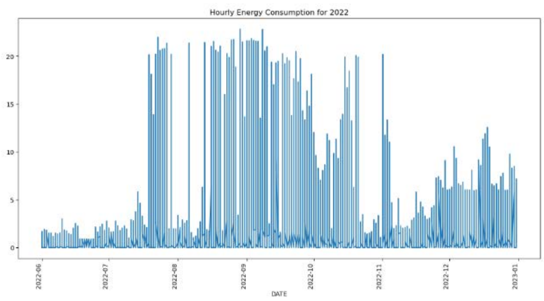


Figure 4.9 Planet for one year – Source: by author

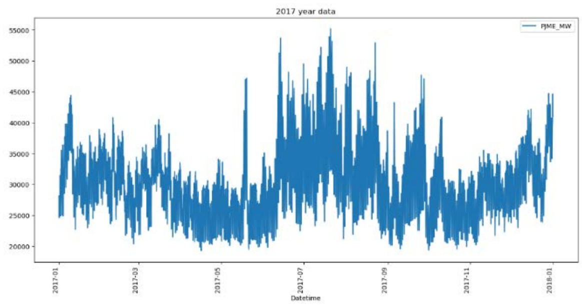


Figure 4.8 PJMW for one year – source: author

Throughout these figures, in Figure 4.6 and Figure 4.7 it is shown all the data available in the database for both projects, R10 and PJMW, in addition to Figure 4.8 and Figure 4.9 that shows the data collected within one year. This difference in amount of data, is not only showing the difference in number of years, but also in the quality of data available for one year for each project. This comparison result will be also shown in the following figures showing the data analysis through time series analysis within one week data, monthly data, several boxplots, daily trends, and finally yearly trends.

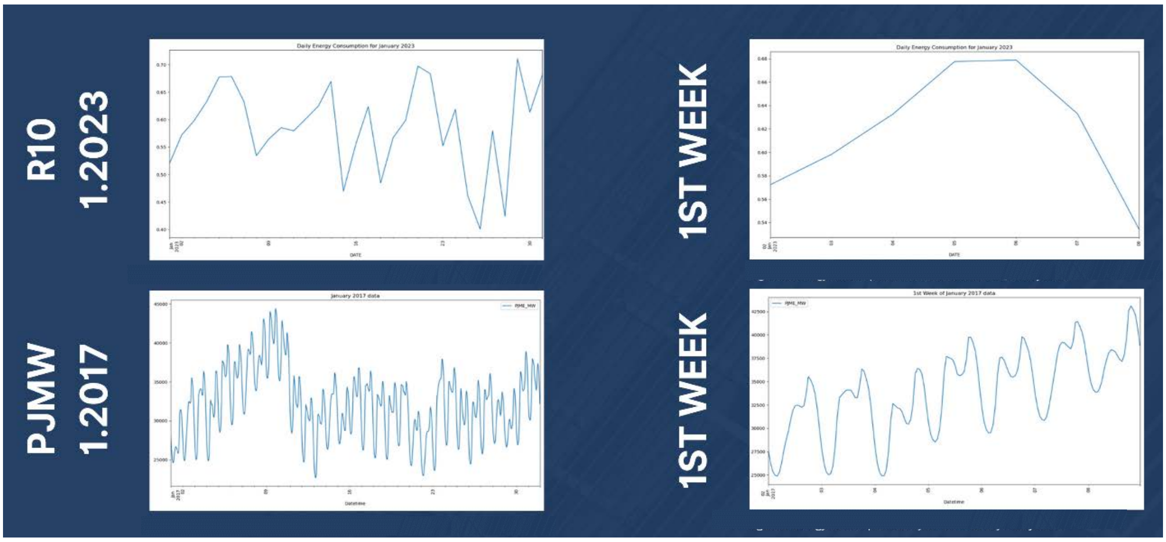


Figure 4.10 R10 and PJMW data for January month and first week of January – source: author

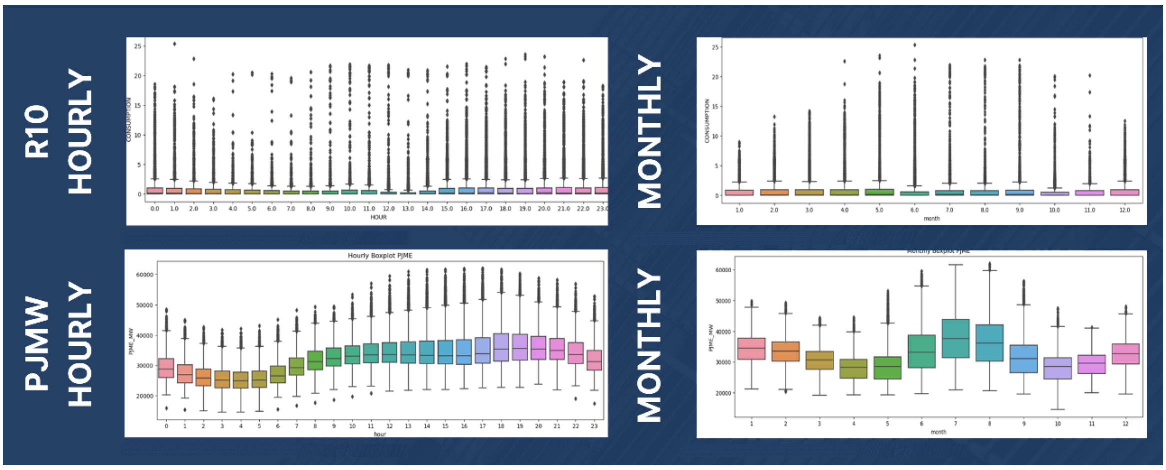


Figure 4.11 R10 and PJMW data hourly and monthly boxplots – source: author

The above Figure 4.10, Figure 4.11, and Figure 4.12 shows the importance of collecting data throughout several years due to making it capable of analysing the trends, patterns, and behaviours of energy consumption. This will make it possible to detect the high and low peaks. In addition, the quality of energy consumption data collected is as important too as the number of years, because in R10 it is obvious that there is a huge number of missing data due to low percentage of data found in monthly and weekly data as shown in figure 4.1.10.

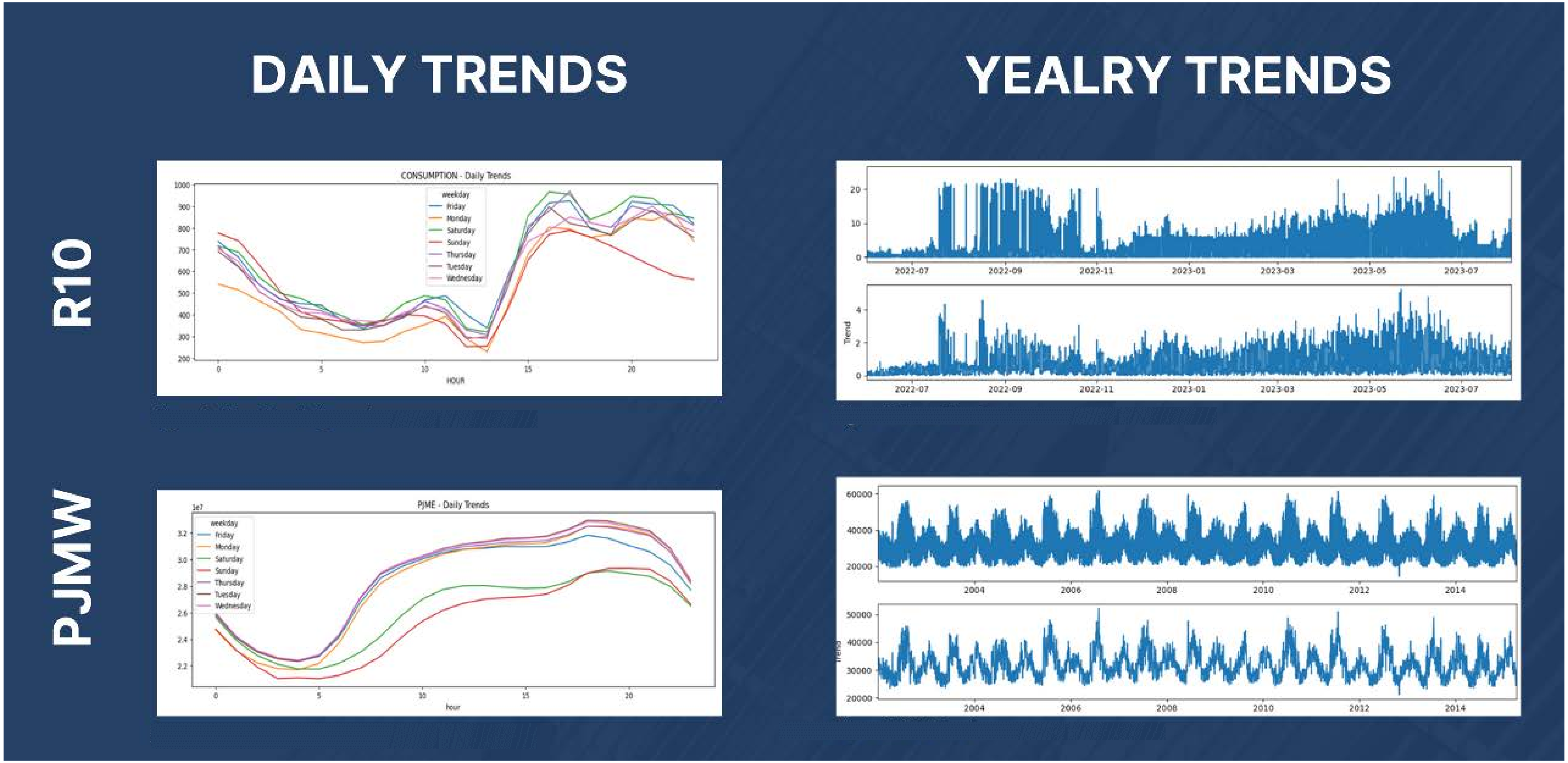


Figure 4.12 R10 and PJMW data daily and yearly trends – source: author

4.1.2 Predict futuristic energy data

The prediction of futuristic energy data is through usage of artificial intelligence (A.I) as STEP (1.2), this step is done through choosing the most suitable method in time series analysis as mentioned above in section 3 Methodology. Thus, the chosen methods to be used for this type of predictions are: (i) ARIMA, and (ii) Prophet model. In ARIMA, it is Auto-Regressive Integrated Moving Average which is a powerful statistical method for forecasting time series analysis. This model is chosen due to the minimal amount of historical data needed which is suitable for R10 universe project as seen previously in phase 1 collected data and works well with data having one variable which is time. (i) AR, which captures the relationship between variable and past values, (ii) I, which is differencing to make the series stationary, and finally (iii) MA, which is to detect the relationship between current value and residual error from past predictions. ARIMA has several parameters that is the order selection. (i) p, which is the Auto-Regressive order and is used to decide how much past observations to make, (ii) d, which is integrated order and is used to decide within it the differencing number to be made so that the data will be stationary, and finally (iii) q, which is the Moving Average order and is used to decide the past forecast errors to take into consideration. After that, the model ARIMA will be trained on previous collected historical energy consumption data, and it will be used to make further forecasting. In figure Figure 4.13 it is shown an example of ARIMA model in time series analysis.

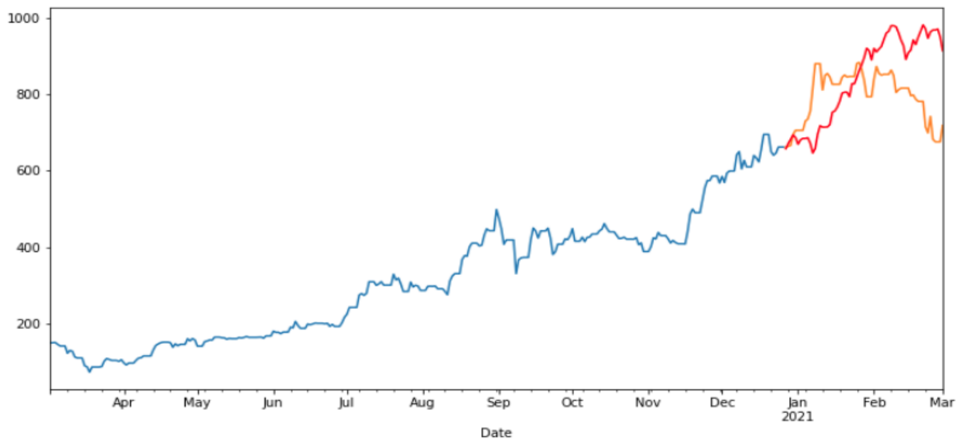


Figure 4.13 Example of ARIMA model – source: https://towardsdatascience.com/an-introduction-to-time-series-analysis-with-arima-a8b9c9a961fb

The second model is the prophet model which is created and developed by Facebook. This model is designed to study the trends, seasonality and holidays for time series analysis and require also minimal parameters and historical data. In comparison to other models such as neural Prophet model, it requires less cost, less complexity, and less data which is also suitable for R10 universe project. Prophet deconstructs the data into trend, seasonality, and holidays. After that it uses these data to make forecasting for future data values and works with irregular patterns. Figure 4.14 shows an example of Prophet model forecasting for future data values.

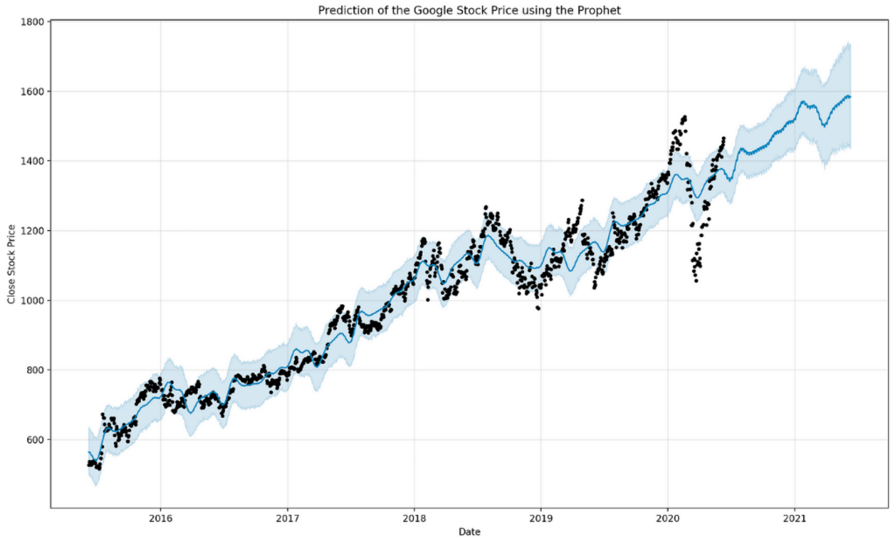


Figure 4.14 Example of prophet model – source: https://towardsdatascience.com/an-introduction-to-time-series-analysis-with-arima-a8b9c9a961fb

Within the results seen in the following figures, it is shown the different results that came out of both models, ARIMA and Prophet for both projects R10 and PJMW.

In Figure 4.15, the results of the ARIMA model shows the difference results that came from different quantity and quality of data. Showing how the predictions are more stable after training the data within the PJMW project while taking the moving average. The following model for ARIMA model has the parameters that control the AR and MA components are (max_p =2 and max_q=1) meaning that the AR can be maximum 2 and the MA can be maximum at 1 while the d is usually considered by auto-arima function internally to make the data stationary.

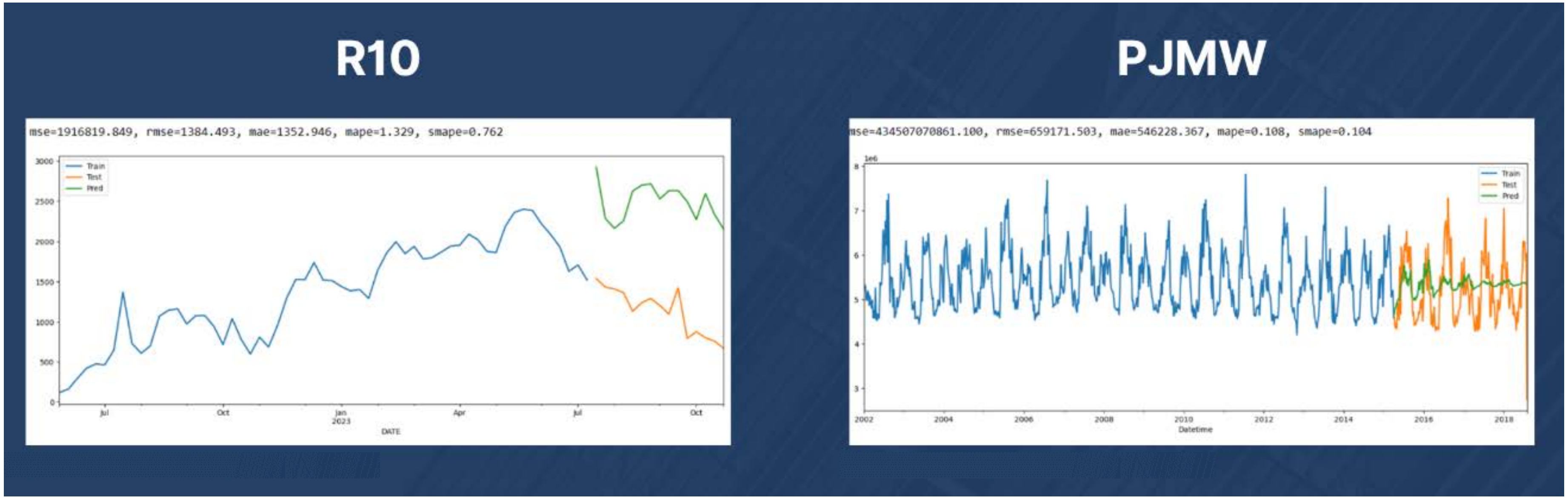


Figure 4.16 ARIMA model results – source: author

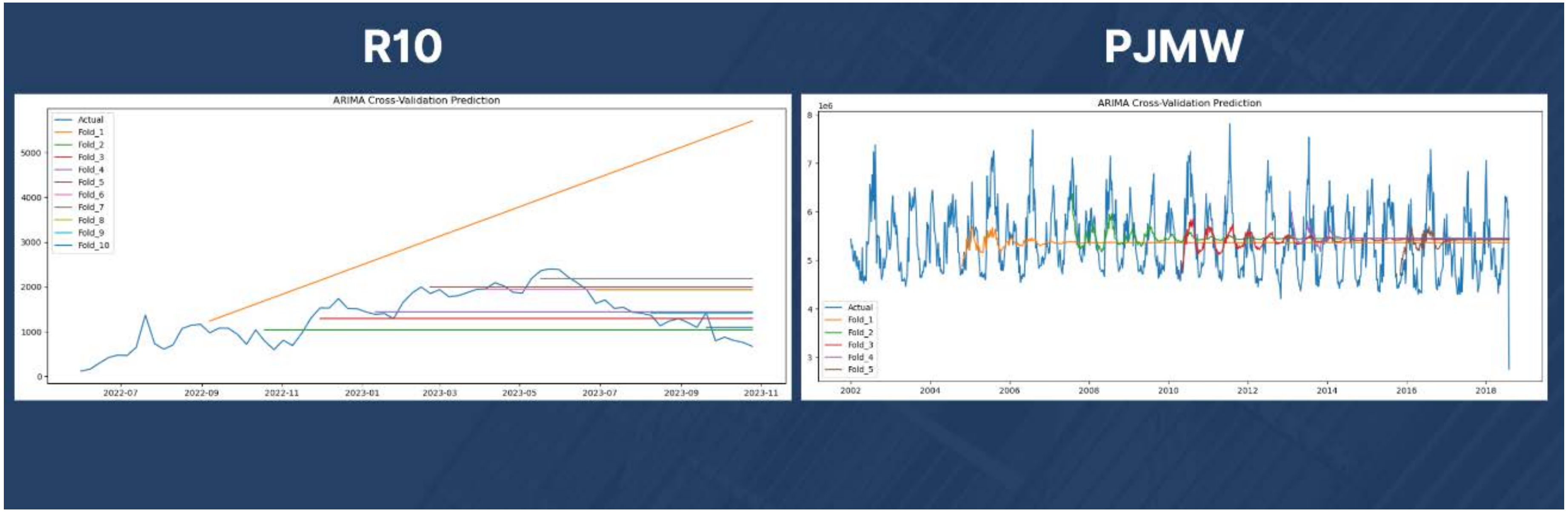


Figure 4.15 ARIMA cross-validation – source: author

Figure 4.16 shows the results of ARIMA cross-validations that are done for each project to train the historical data more and be able to achieve the best model of parameters. Due to the lack of quality and quantity of data in R10 project, in ARIMA model it was not able to reach good results of cross validation after 10 folds, which the project PJMW was able to achieve good results after cross-validation with 5 folds. These cross-validation results were stored in a list and then used to make the forecast after that for both projects as shown in Figure 4.17. The following results in ARIMA model for both projects shows the importance of data quantity and quality after all to be able to make the final forecasting. The second model is the Prophet model, and the following Figure 4.18 shows the results for this model for both projects R10 and PJMW after historical data being trained.

After these predictions, it is shown that because the data for R10 project is only for one year, the prophet model failed to predict the data more than 3 months and started to give negative numbers of energy consumption which is irrational. While since there is 16 years of data in PJMW, the prophet model was able to predict data in more consistency. After that in the following Figure 4.19 it is showing the cross validation with fold 5 that is created for this prophet model. The model started to divide the data into 5 separate data sets and started to train its model on them. The following figure shows the importance of training the data through cross validation due to the stability in predictions that happens because of that.

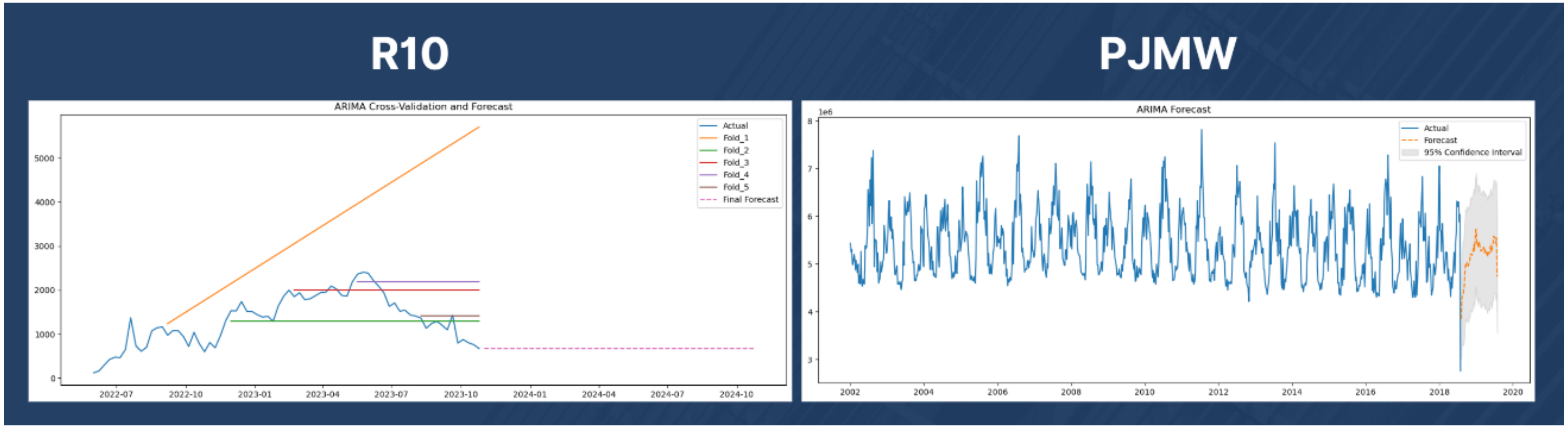


Figure 4.17 ARIMA forecasting results – source: author

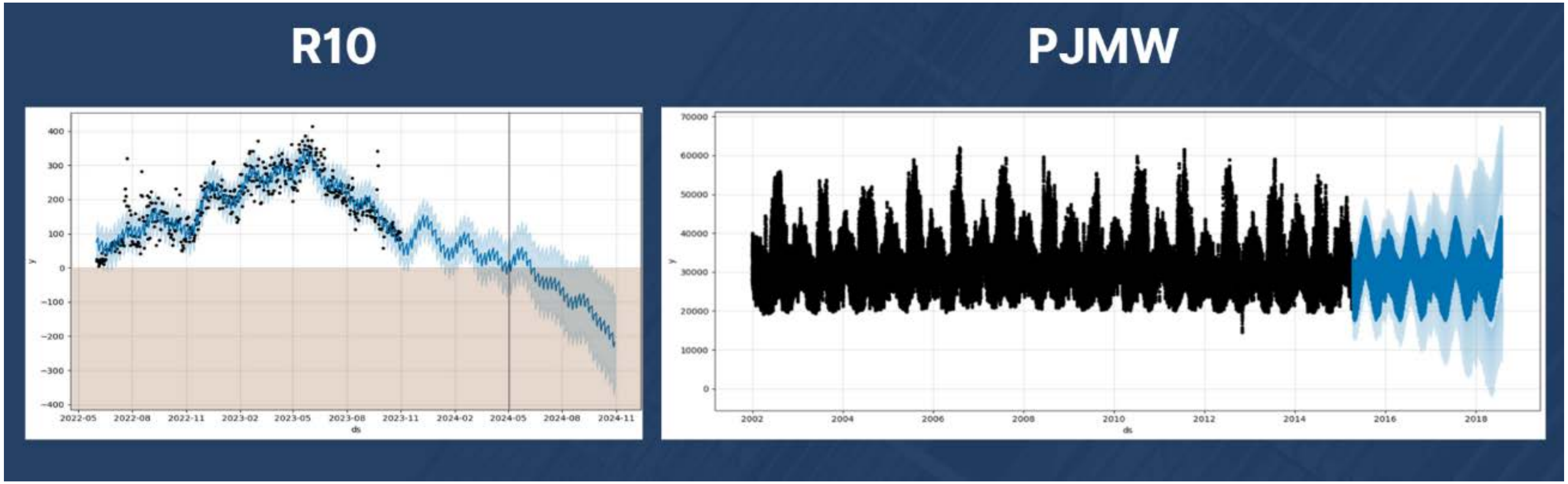


Figure 4.18 Prophet forecasting results – source: author

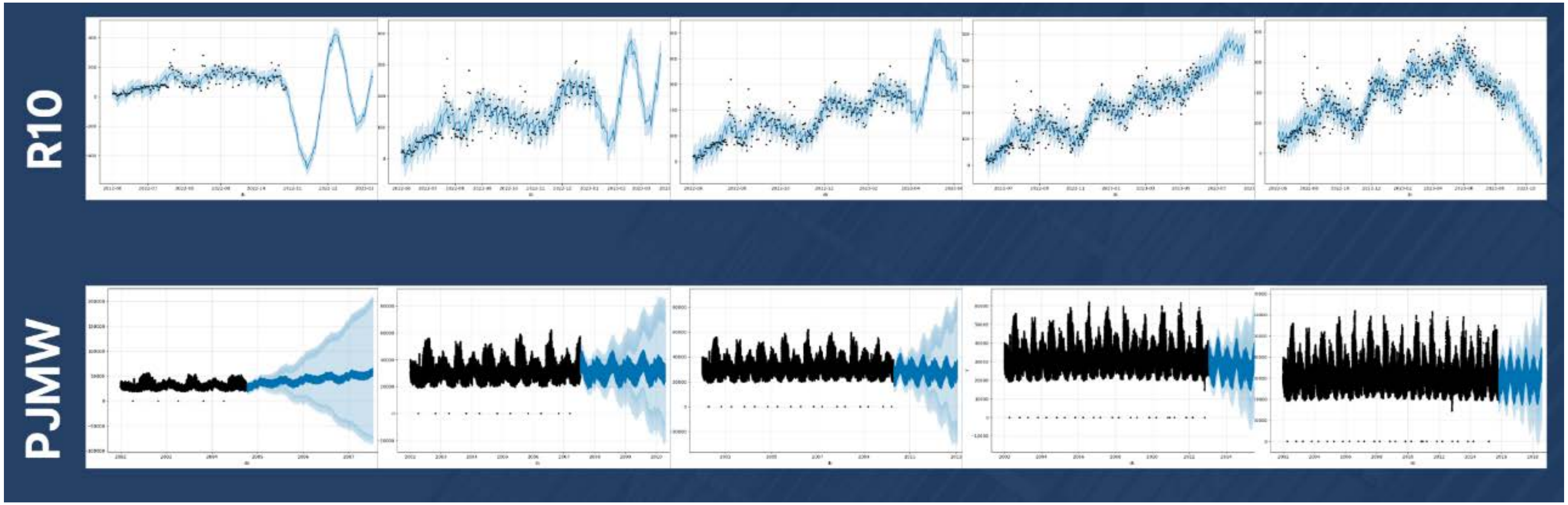


Figure 4.19 Prophet cross validations results – source: author

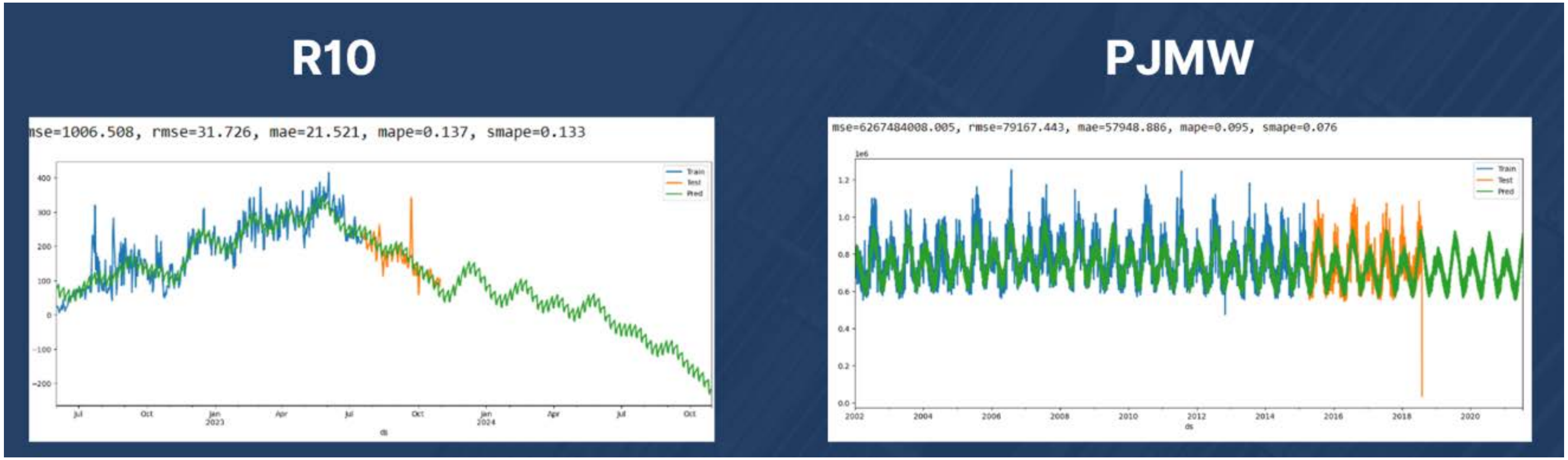
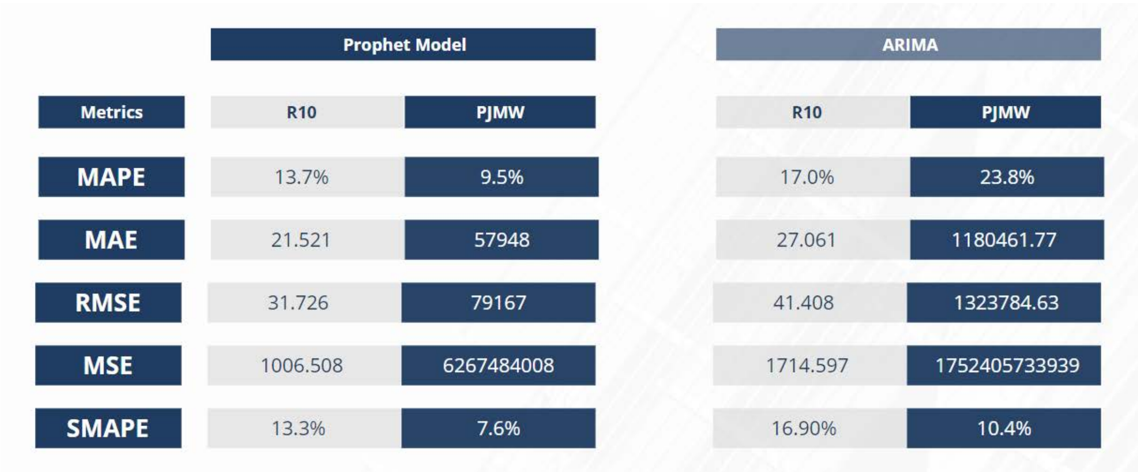


Figure 4.20 Prophet forecasting – source: author

After making the cross validation, the results are saved and used to make after that the prophet forecasting as shown in Figure 4.20.

After predicting the futuristic energy consumption for both projects, it is important to study and analyse the errors. The errors in these models to be able to analyse what model works better are (i) Mean squared error (MSE) which it calculates the squared difference between actual and predicted values then it gives the average of it, (ii) Mean absolute error (MAE) which it calculates the difference between predicted and actual values, (iii) Mean absolute percentage error (MAPE) which it calculate the average percentage between actual and predicted values, (iv) Symmetric mean absolute percentage error, where it compares the relative error compared to the absolute size of the actual and predicted values, and finally (v) Root mean squared error (RMSE) which is the square root of MSE. These metrics will be able to check how accurate the model and let you understand more about how your model is working. These metrics for both projects R10 and PJMW with the forecasts of both models Prophet and ARIMA are shown below in Figure 4.21. These results show that prophet model is better to use as a method for predicting futuristic energy consumption due to having less errors in both projects. The numbers of errors in PJMW within the MAE, RMSE, and MSE were huge due to the huge amount of data in PJMW project.



	Prophet Model		ARIMA	
Metrics	R10	PJMW	R10	PJMW
MAPE	13.7%	9.5%	17.0%	23.8%
MAE	21.521	57948	27.061	1180461.77
RMSE	31.726	79167	41.408	1323784.63
MSE	1006.508	6267484008	1714.597	1752405733939
SMAPE	13.3%	7.6%	16.90%	10.4%

Figure 4.21 Prophet forecasting – source: author

4.1.3 Recommendations for R10 Universe

After collecting, analysing, and predicting energy consumption data in R10 universe project through the outcomes of two above phases, 4 recommendations were elaborated (STEP 1.4). They aim to fine tune the GIS-based IIAT dashboard for continuous improvement, allowing companies to adapt and evolve its operations in response to changing factors. The four recommendations are:

1. Usage of prophet model instead of ARIMA model due to lower percentage of errors
2. Gather energy consumption for several years in developing countries project.
3. Place more energy consumption sensors for several activities after site construction is built.
4. Avoid having missing data and focus on quality of data too not only on quantity.

Therefore, based on the structure of the recommendations explained in section 3.1.4 Methodology, the four recommendations are detailed below through:

1. Name of recommendation
2. Linkage with 17 sustainable development goals set by UN general assembly

3. Description of recommendation

The following Figure 4.22,Figure 4.23, Figure 4.24, and Figure 4.25 will be showing the recommendations detailed as mentioned above.

Parameters	Recommendation 1
Title	Usage of prophet model instead of ARIMA model due to lower percentage of errors
Linkage with SDGs	
Description	The usage of prophet instead of ARIMA is to enhance accuracy for further projects, since prophet is designed to be user-friendly, have less errors in the previous phases, this will make it more valuable for decision-making and risk management.

Figure 4.22 Recommendation 1 – source: author

Parameters	Recommendation 2
Title	Gather energy consumption for several years in developing countries project
Linkage with SDGs	
Description	To enhance energy monitoring in developing countries and being able to use A.I within these countries. They will be able to promote sustainable practices, environmental impact assessment and community engagement and awareness which is socially important.

Figure 4.23 Recommendation 2 – source: author

Parameters	Recommendation 3
Title	Place more energy consumption sensors for several activities after site construction is built
Linkage with SDGs	
Description	This will let the citizens being able to monitor there energy consumption while using the project to fine-tune energy efficiency, optimize consumption, enhancing sustainable practices, and affecting the environmental impact

Figure 4.24 Recommendation 3 – source: author

Parameters	Recommendation 4
Title	Avoid having missing data and focus on quality of data too not only on quantity
Linkage with SDGs	
Description	The quality of data will enhance the optimization of energy consumption which will enhance the technological integration of real time-monitoring and being able to use this data for better sustainable practices.

Figure 4.25 Recommendation 4 – source: author

4.2 Sei Milano Project in Milano, Italy

Starting from the site analysis of the project knowing that it is in Italy and after studying the surrounding of it, this section serves the purpose of fine tuning the GIS-based IIAT dashboard through usage of artificial intelligence. This case study was research-oriented, aimed at identifying suitable sensors and strategies to mitigate Urban Heat Island (UHI) effect. This section is divided into 3 sections that shows the whole process for this case study.

4.2.1 Collect and analyse data

The collection and analysis of data is first based on the selection of indicator as STEP (2.1), this step is done through choosing the most important indicator that must be studied in the project for Planet Company based on their preferences and availability of data. Thus, the selected indicator among all the indicators was the urban heat island indicator from energy and climate change section as shown in methodology section 3. In addition, urban heat island is where a project in the city might tackle due to having higher temperature than rural areas due to human activities and urbanization as shown in Figure 4.26. This might happen due to usage of more concrete, less green areas, increased energy consumption and heat-absorbing materials.

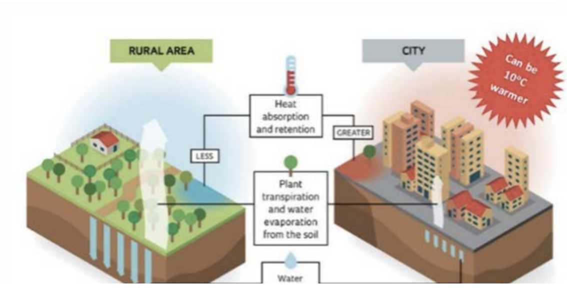


Figure 4.26 Urban Heat Island – source:

Hot environments according to NASA are places that lacks water, plants and trees which have hot environments, and they might also be characterized by dark colours. This analysis can be made even by naked eye on Google satellites and predict which place is hotter than the other one. (Google, 2023)

> *"If you're looking at a Satellite View image that's mostly green, you understand that there are trees there that can provide shade," he says. "Compare that to an image that's mostly gray – lots of concrete and asphalt. It's easy to guess which one will feel hotter." - Yael Maguire, Vice President and General Manager of Geo Sustainability at Google*

The following Figure 4.27 shows the reason why some cities are hotter than others, and how high population is affecting global temperatures which leads governments and organizations to work on reducing greenhouse gas emissions. Thus, Google started to implement artificial intelligence as a solution to analyse this imagery through image-recognition process to focus on determining the tree-planting efforts, and how this can be costing less than doing it using drones which might cost also to five to eight years. Which shows how artificial intelligence interference will cost less money and time. (Google, 2023)

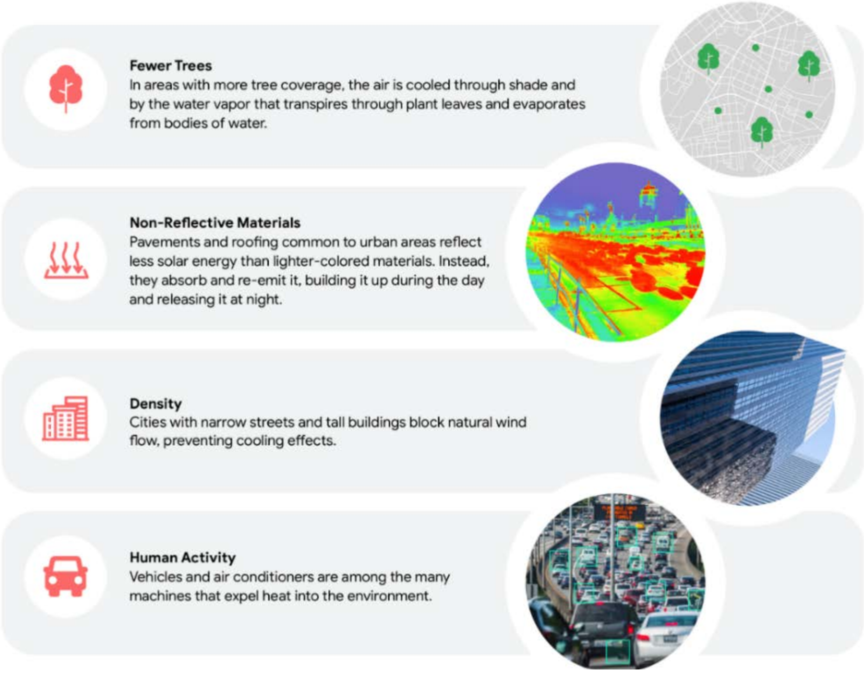


Figure 4.27 How cities get Marooned on "Heat Islands" - source: https://sponsored.bloomberg.com/article/google-sustainability/can-ai-help-cool-the-worlds-hottest-cities

Sei Milano project is still in phase A5 which is the construction and installation process. Thus, the project is still under construction and all the activities for energy consumption in this site is related to this phase.

Through analysing the collected data from planet database, Figure 4.28 shows the analysed collected data for this indicator based on MC-SDSS process, this includes smart solutions as mentioned below, (i) mapping tree coverage and cool-rooftops, (ii) smart irrigation and water management, (iii) building energy management and (iv) smart heat warning systems.

Then collecting data for the activities that must be handled through (i) Data collection, which is accurate data is essential for AI to analyse urban heat island solutions, (ii) Machine learning algorithms, which are used to analyse data and identify patterns that can inform urban heat island solutions, (iii) Predictive models, are used to forecast the effects of urban heat island solutions, and finally (iv) Visualization tools, can be used to communicate the results of AI analysis to city officials, urban planners, and the public. In addition, through this figure it is shown that the main stakeholders for this project are Planet smart city company, public administrators, and EU & UN. Moreover, the resources will be from Planet smart city and EU funds. The dissemination will be to insert sensors for data collection and to analyse and predict these data. Thus, the main impact of this project on SDGs is through SDG (13) which is to reduce carbon emissions, SDG (11) that improves air quality and finally SDG (3) which reduces number of heat-related illnesses and deaths.

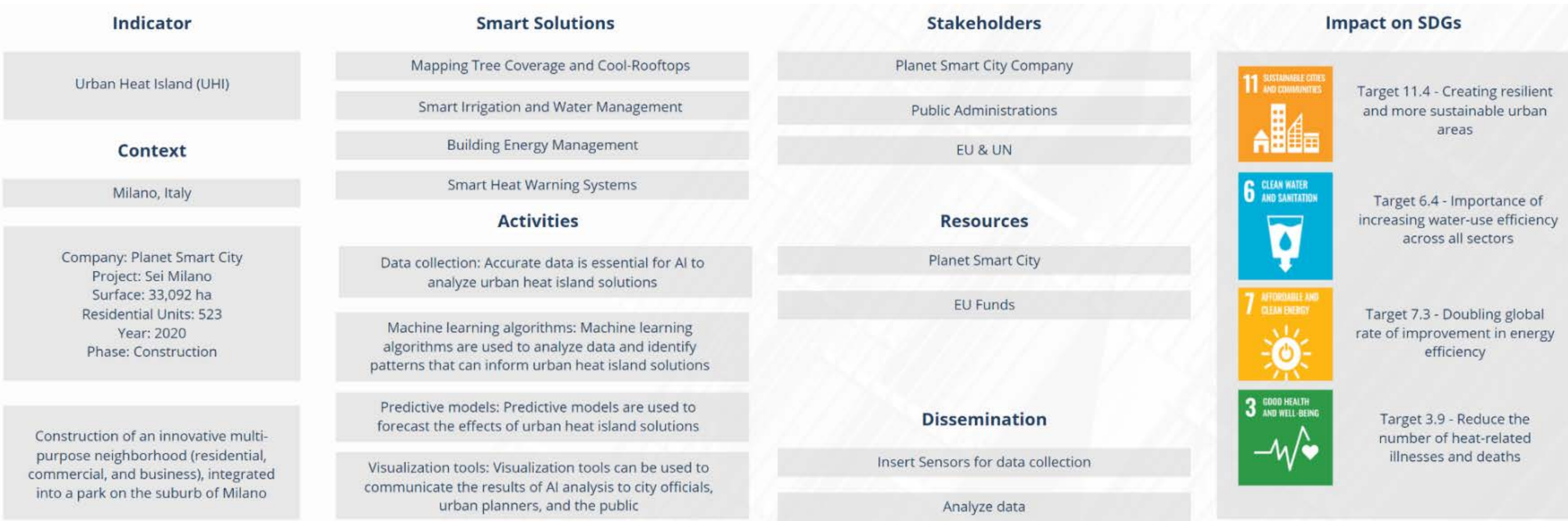


Figure 4.28 Indicator impact assessment framework – source: author

4.2.2 Indicator impact assessment framework proposal

After the collection of all these data and analysing the urban heat island (UHI) indicator in phase 1, this section will be creating the impact assessment framework proposal for further projects that want to take the urban heat island effect into consideration.

The first solution is usage of artificial intelligence to analyse cool roofs/pavements mapping through satellites. This solution is developed by Google which is using artificial intelligence to reduce hot environments in urban areas. The application is through Google's Environmental Insights Explorer (EIE) which have data around building types, trees canopy and data around environments in urban areas. This same solution can be applied to smart city projects to analyse the number of trees, plants and materials needed for the project through artificial intelligence analysation for the current situation.

In Figure 4.29 below as Google explains, the company must start mapping and sampling numerous numbers of aerial images for the city which shows it in different situations, different weather conditions and different places. Then the team must start labelling each region in the built environment the images that contains trees and plants. This is the way where artificial intelligence will train itself to recognize the trees from aerial images. Then this will train the deep neural network to start segmenting trees in new imagery. This process will let the decision-making process to know where to plant trees and put water elements in the project easier based on artificial intelligence analysis to previous examples and dataset.

This process will be also repeated for cool-roof mappings as shown in Figure 4.30 below where the albedo which is the solar reflectivity will be measured from aerial imagery and satellite reflectivity measurements which will also create a dataset and analyse where do cool-rooftops appear and where do they lack. This will also make it easier for decision-makers to change materiality of rooftops to make the environment cooler.

Figure 4.29 Google Enviornmental Insights EIE - source: https://sponsored.bloomberg.com/article/google-sustainability/can-ai-help-cool-the-worlds-hottest-cities

Figure 4.30 Cool-Roof Mapping from Google - source: https://sponsored.bloomberg.com/article/google-sustainability/can-ai-help-cool-the-worlds-hottest-cities

The second solution is smart irrigation and water management which will help analyse weather forecasts, soil moisture levels, and plant water requirements in real-time based on artificial intelligence. This will help improve several sustainable development goals for the smart city that is being built, including clean water and sanitation, affordable and clean energy, sustainable cities and communities, responsible consumption and productions, climate actions, life below water and life on land.

Figure 4.31 Smart Irrigation system - source: www.senraco.com

This will schedule the water distribution and can help maintaining healthy vegetation while conserving water. This will help to cool the surrounding environment by keeping healthy vegetation.

This solution is already implemented in Cartagena city council by FIWARE company. As Figure 4.32 shows, the smart irrigation management system elements needed to be implemented are wireless management system, water meter & leaks detector, soil moisture sensors, electro valve control, and weather station with sensors to detect temperature, humidity, rain, wind, UV radiation, and evapotranspiration. These sensors are all implemented in one system to gather real-time data so that artificial intelligence will be analysing the needs of vegetation to be irrigated (FIWARE, 2021).

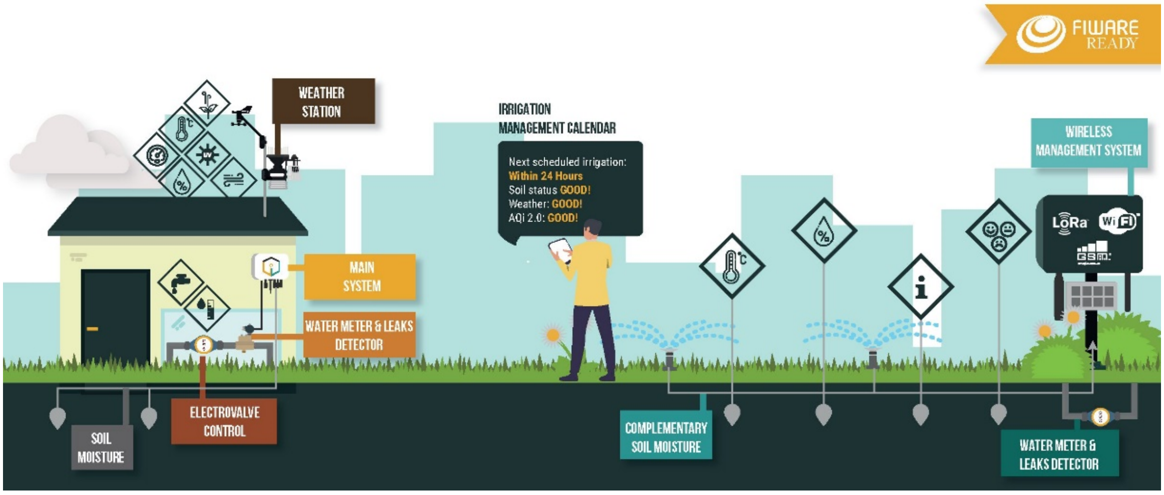


Figure 4.32 Smart Water Management System - Source: https://fiware-foundation.medium.com/smart-irrigation-system-implemented-in-cartagenas-city-db49c3067ef0

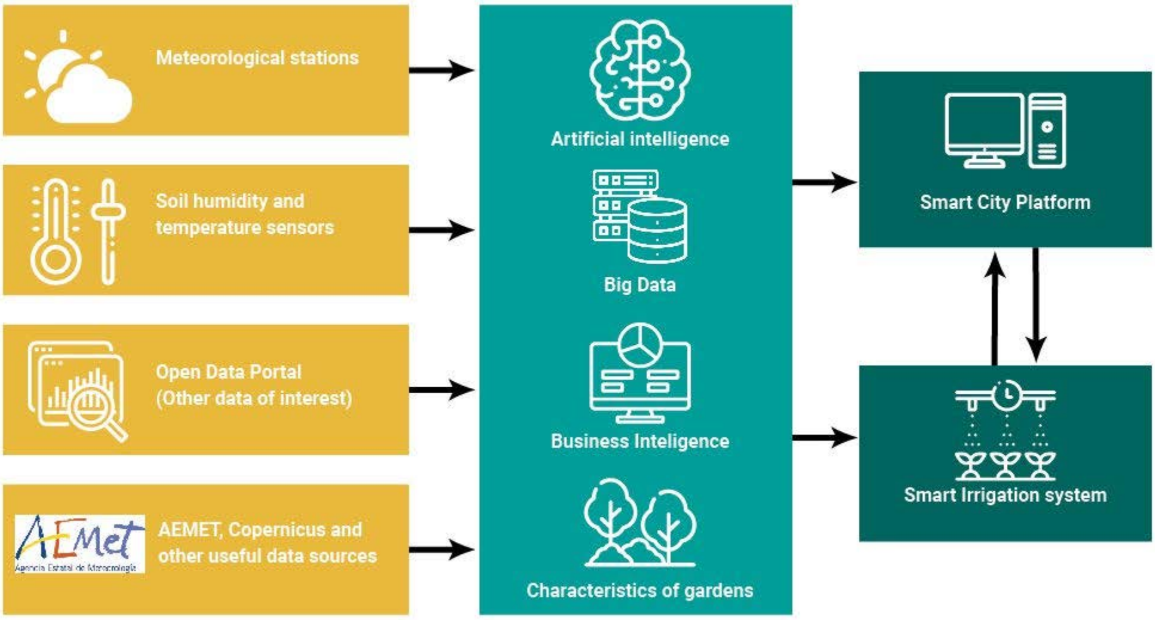


Figure 4.33 Smart Water Management - source: https://fiware-foundation.medium.com/smart-irrigation-system-implemented-in-cartagenas-city-db49c3067ef0

Figure 4.33 shows how the system works through an Internet of Things' (IoT) device, it shows the connections between the sensors in meteorological stations, soil humidity and temperature sensors, open data portal and other data sources in the system, all these datasets should be linked to artificial intelligence, big data, business intelligence and characteristics of gardens to be linked later into smart city platforms and smart irrigation system. Then all these data are gathered as a result in user interface as shown below. This user-interface shows the recommendations by artificial intelligence based on meteorological data and nature of type of soil collected (FIWARE, 2021).

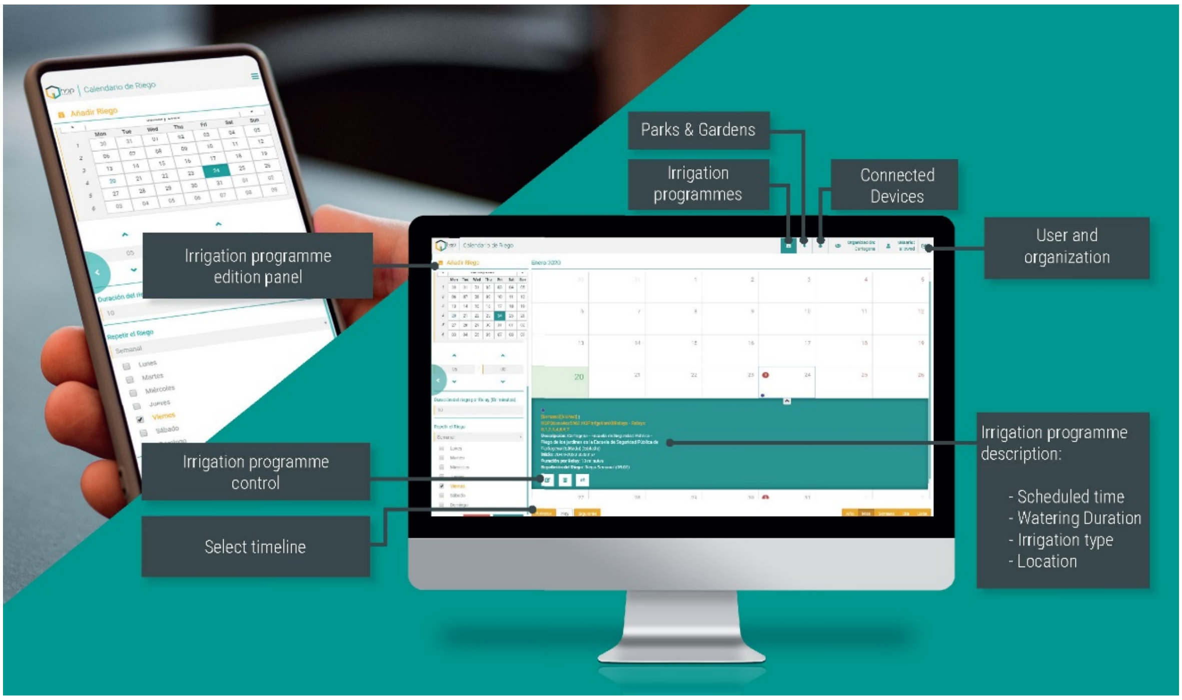


Figure 4.34 User-friendly application for smart water management system - source: https://fiware-foundation.medium.com/smart-irrigation-system-implemented-in-cartagenas-city-db49c3067ef0

The algorithms are developed to focus on size of green areas, water consumption used, rainfall and after that develop predictions to allow the responsible organization (in this case Planet company) to collect, analyse and visualize these data in a simplified manner. This solution will conserve green areas, decrease water loss, and automates irrigation.

The third solution is developing a building energy management through usage of artificial intelligence too. Optimizing energy usage in buildings will enhance energy consumption and indoor comfort levels in different buildings. Adjusting these heat loss and gain within the building through energy consumption optimization will reduce energy usage and enhance thermal comfort which will also enhance the environment (Grid Edge, 2019).

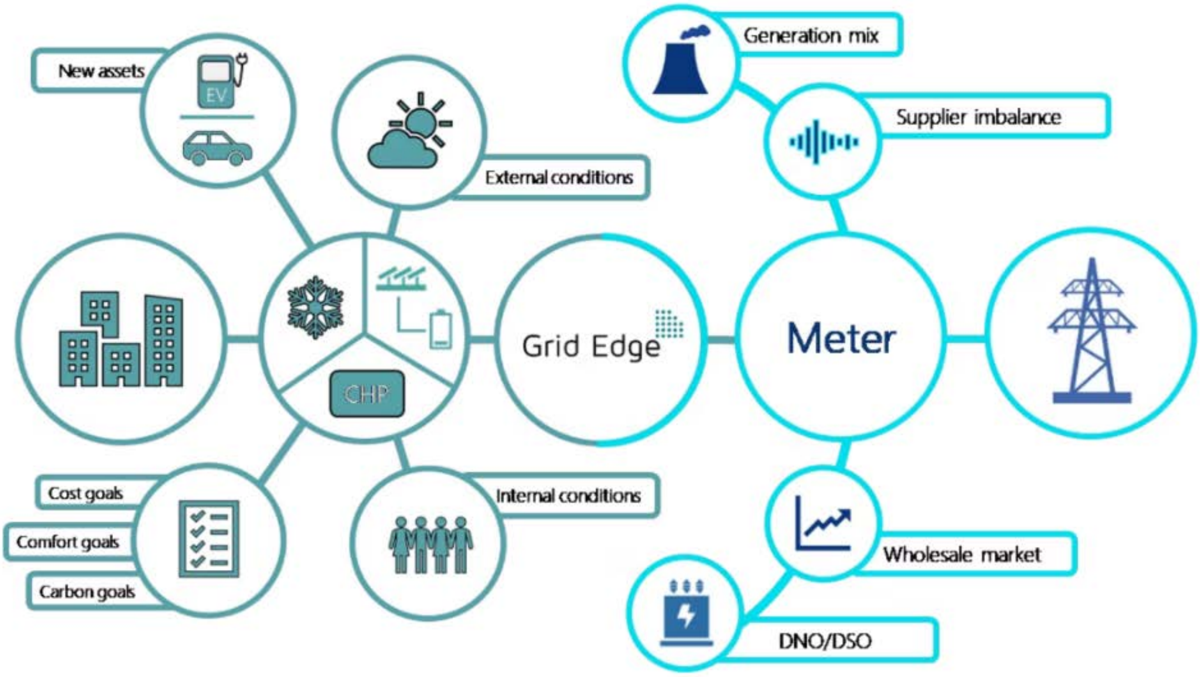


Figure 4.35 Grid edge systm for building energy management system - source: https://www.iea.org/articles/case-study-artificial-intelligence-for-building-energy-management-systems

Figure 4.35 above shows how artificial intelligence is applied into building energy management systems and how different elements are connected. The system is developed by Grid Edge company that is developed by UK which works by gathering data from building's existing energy management systems and other sensors as weather conditions stations. These gathered data will be optimized by artificial intelligence to enhance energy efficiency performance in buildings (Grid Edge, 2019).

The needed elements are electricity meters, and wider electricity networks which will also monitor energy prices and optimize the electricity load of the buildings based on cost and energy consumption. This will enhance cost savings, revenues of annual on-site energy costs, and most importantly carbon emissions which will affect urban heat island directly (Grid Edge, 2019).

This will be connected to a user-friendly interface that will show Planet company all the energy consumption data, in addition to recommendations based on artificial intelligence analysis on how to manage human activities inside the buildings to optimize energy consumption which will lead into citizen's engagement in this process and enhance the community on social level since they will be learning about new technologies.

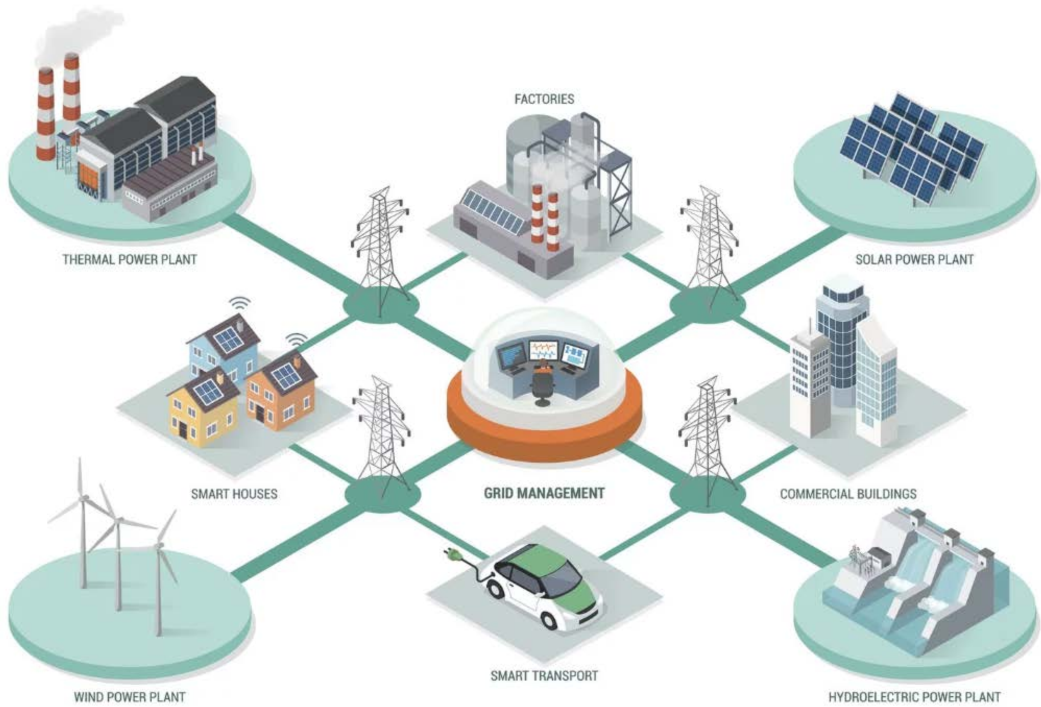


Figure 4.36 Grid management system - source: https://www.iea.org/articles/case-study-artificial-intelligence-for-building-energy-management-systems

The fourth solution is through a qualitative approach too while choosing the sensors for the smart heat warning systems. This is through choosing different sensors through studying their usage, place of implementation, calculations and the references needed to measure the needed sensors. Figure 4.37 below is all the sensors that are chosen with all the details mentioned above.

The chosen sensors were all needed to collect data within the project to be able to avoid heat waves in future scenarios. The sensors are (i) Temperature, which measure temperature of air and surfaces and can be placed on buildings, pavements, or park. This sensor will be needed for each 1000 m2 and that is based on ISO and Italian Energy Efficiency Fund, (ii) Humidity, which measures the amount of moisture in air and can be places on green roofs or parks. This will be needed every 2000 m2 and that is by National Institute of Metrological Research and ENEA (iii) Solar Radiation, which measure the solar radiation received at site and is placed on rooftops with unobstructed views of the sky. This will be needed every 1000 m2 from Associazione Italiana Energia Solare, AIES, (iv) Thermal imaging, which is mounted on drones and will be needed every 10,000 m2 by Ministero dello Sviluppo Economico, (v) Soil moisture, which will measure the moisture in soil mix and shows the efficiency of green infrastructure that are found in site. This will be needed evert 500 m2 by Consiglio Nazionale delle Ricerche, CNR and Associazione Italiana di Ingegneria Agraria AIIA, and finally (vi) Air Quality, which measures the pollutants found in air such as NO2 and can be placed on buildings and parks. It will be needed for every 1000 m2 and this is by Agenzia regionale per la protezione dell’ambiente della Lombardia, ARPA Lombardia.

Sensor Names	Temperature	Humidity	Solar Radiation	Thermal Imaging	Soil Moisture	Air Quality
Description	Measures temperature of air or surfaces	Measures the amount of moisture in air	Measures the amount of solar radiation received at the site	To develop visual thermal maps in cities, to check the UHI	Know the effectiveness of green infrastructure solutions	Measures level of pollutants,for example: Nitrogen dioxide
Place of Implementation	Buildings, Pavements, or parks	Green roofs or parks	Rooftops with unobstructed views of the sky	Mounted on drones	Soil at parks or other green areas	Buildings and parks
Calculations	One sensor for each 1,000 m2	One sensor for each 2,000 m2	One sensor for each 1,000 m2	One sensor for each 10,000 m2	One sensor for each 500 m2	One sensor for each 1,000 m2
References	International Organization for Standardization (ISO) and the Italian Energy Efficiency Fund	National Institute of Metrological Research (Istituto Nazionale di Ricerca Metrologica, INRIM) and ENEA	Associazione Italiana Energia Solare, AIES	Ministero dello Sviluppo Economico	Consiglio Nazionale delle Ricerche, CNR and Associazione Italiana di Ingegneria Agraria, AIIA	Agenzia regionale per la protezione dell'ambiente della Lombardia, ARPA Lombardia

Figure 4.37 Sensor's chosen through A.I - source: by author

Steps	Description	Application
Data Collection	Gather relevant data related to temperature, humidity, wind conditions, and other factors that influence heat levels	Tableau Software
Data preprocessing	Collected data needs to be processed and cleaned to remove any outliers or inconsistencies	
Feature Extraction	Relevant features or variables are extracted from the collected data	
Historical Analysis	The system analyzes historical data to identify patterns and trends in heatwave occurrences	KNIME Software
Model Development	Machine learning or deep learning, can be used to develop a predictive model	
Training the Model	The model is trained using historical data, where it learns to recognize patterns and make predictions based on the input features	Python Language
Real-Time Monitoring	The system continuously monitors real-time weather conditions by collecting up-to-date data from various sources	
Prediction & Risk Assessment	The trained AI model is used to predict the likelihood and severity of heatwaves based on the current and forecasted weather conditions	
Alert Generation	When the system detects a high probability of a heatwave or extreme heat conditions, it generates alerts or warnings	Through applications
Adaption & Cont. Improvement	The system can learn from feedback and user responses to improve its predictions and warning accuracy over time	

Figure 4.38 Theoretical framework for data processing - source: by author

These sensors will be collecting data within several time of the year continuously, which will lead to predictions and risk assessments as shown within these steps in Figure 4.38. This figure will show how to use artificial intelligence in monitoring the urban heat island (UHI) effect within sei Milano project and it is used as a theoretical framework.

The figure above shows the processing of data that should be handled after gathering data from the sensors for the smart heat warning systems and which applications must be used for it. The process has several steps starting from (i) Data collection which is gathering relevant data related to temperature, humidity, wind conditions, and other factors that influence heat levels, and this can be handled through Tableau software to visualize the data. Then we have (ii) Data pre-processing which is processing and cleaning data to remove any outliers or inconsistencies, (iii) Feature extraction which is extracting the needed variables from the collected data, (iv) Historical analysis which is where the system analysis the historical data to identify patterns and trends for previous heatwaves, (v) Model development which is machine or deep learning (ML or DL) which is used to train the model, (vi) Training the model which is recognizing patterns and training the data to let the system learn this historical data, (vii) Real-Time Monitoring which is collecting data continuously by real-time data, for example weather conditions, to keep the model updated with recent data, (viii) Prediction and Risk assessment which is forecasting based on historical data the further data that will happen, and all these steps from (ii) till (viii) can be applied by python programming languages and KNIME. After that, (xi) Alert Generation which is letting the system inform the governments and citizens the high probability of heatwaves that might happen through mobile applications, and finally (x) Adaption and Continuous Improvement, which letting the system to learn from previous data and feedback to improve its predictions and warning accuracy over time.

4.2.3 Recommendations for Sei Milano Project

After collecting, analysing, and creating the theoretical framework through the outcomes of two above phases, 2 recommendations were elaborated (STEP 2.4). They aim to fine tune the GIS-based IIAT dashboard for continuous improvement, allowing companies to adapt and evolve its operations in response to changing factors. The four recommendations are:

1. Further data collection needed on actual materials and textures used.
2. Place more sensors to be able to detect heat waves before they arrive to protect users.

Therefore, based on the structure of the recommendations explained in section 3.1.4 Methodology, the four recommendations are detailed below through:

1. Name of recommendation
2. Linkage with 17 sustainable development goals set by UN general assembly

3. Description of recommendation

The following Figure 4.39 and Figure 4.40 will be showing the recommendations detailed as mentioned above.

Parameters	Recommendation 1
Title	Further data collection needed on actual materials and textures used to calculate UHI
Linkage with SDGs	
Description	By understanding the materials and textures on site that absorb, retain or reflect heat, urban planners, policymakers, and researchers can be able to calculate the UHI effect and this can enhance urban planning, climate mitigation and energy efficiency.

Figure 4.39 . Recommendation 1 - Source: by author

Parameters	Recommendation 2
Title	Place more sensors to be able to detect heat waves before they arrive to protect users
Linkage with SDGs	
Description	Detecting heatwaves will be able to save lives in countries where are heatwaves especially in developing countries for further projects and this will raise awareness and improve citizen's engagements which will have economical, environmental and social impact.

Figure 4.40 . Recommendation 2 - by author

05

Conclusions and Future Developments

Chapter 5: Conclusions and future developments

Technological solutions in a smart city are to improve citizen's quality of life, and that is the most important aspect of digital transitions in cities. Digital transition is a powerful tool for all stakeholders since it will enhance the performance in decision-making (Petrova-Antonova & Ilieva, 2019). These digital transitions will be important towards a sustainable future and transform how we interact with each other and with the environment around us, since within all these decisions that are being taken automatedly, the citizens must be able to get full clarification from the governments why and how they are taking these decisions and they should be able to explain to them what algorithms they are working on and how they are working on it, but that would need high citizen's engagement and high education for citizens to be able to understand these theories or else they will feel lost and unheard which will create a huge gap in the community between the governments and citizens. This shows the importance of transparency while using artificial intelligence in several sectors that is connected to citizen's life (Garrido, 2023).

Developing countries have until now within this era a lot of problems concerning basic life requirements such as income, rate of growth, sustainability, consumption, and a lot of issues that have economic, environmental, and most importantly social impact. Citizen's need to be educated well to be able to handle this digital transition that is happening around the world in these developing countries first to be able to handle all these changes and adapt with them too. Digital platforms are socio-technical transitions for developing countries since they will have impact on several factors across the country but needs evolutionary stages in the current state.

Conclusive Remarks

The thesis focusses on two case studies, R10 universe project which is in a developing country (India) and the other is Sei Milano project in a developed country (Italy) and highly developed city (Milano). The methodology was divided into three phases, integrating quantitative and qualitative approaches. The phases and objectives of the thesis were:

Phase 01 – Collect and Analyse Data

Objective: Collect and analyse data to improve the predictions and forecasting of futuristic data, this was within both projects in R10 universe which is through selecting energy consumption as indicator based on indicator's importance then through collecting and analysing data based on sensors on site construction. After that it was through selecting the indicator for the second case study Sei Milano project in Italy which was Urban heat island (UHI) based also on indicator's importance and then to collect and analyse data from planet company database.

Phase 02 – Assessment and Evaluation

Objective: Propose theoretical impact assessment framework based on artificial intelligence, for the first case study in Pune, India the step was to predict futuristic energy data through artificial intelligence subsets (Time series analysis, ARIMA, and Prophet) since data was available in phase 01 for this project. Concerning the second case study in Milano, Italy, it was creating indicator impact assessment framework proposals through artificial intelligence, but it had a research-oriented approach due to absence of data in phase 01.

Phase 03 – Recommendations

Objective: Elaborate recommendations for continuous improvement, allowing companies to adapt and evolve its operations in response to changing factors for both R10 and Sei Milano case studies based on the outcomes from phase 01 and 02.

Limitations

First Phase

In the first phase, for R10 universe, the data was collected from energy consumption sensors that are on site construction, and since the project was still in construction phase, the data that is collected is connected to construction activities on site. There is also another suggested project which is PJMW that is as a benchmark for the analysis in phases 01 and 02 for this case study. For Sei Milano project, the data was collected from the database of Planet's company to analyse the site, the buildings, and the surrounding of the project. This thesis focused on the Phase 01 in two different countries of India and Italy which shows different potentials in each project, but it also had several limitations.

Considering all these, it is important to highlight some limitations and proposals from the first phase:

- Data collection for both projects was weak since the data was collected for the first project for only one year which is not enough for this analysis.
- It is important to have quality data not only quantity, which is due to having missing data within this one-year collection of energy consumption.
- The database should have specific data for what type of materials used in Sei Milano project and what are the SRI for each material need to be calculated.

Second Phase

In the second phase, for both projects, the purpose was to predict futuristic energy consumption data for R10 universe and create a theoretical framework on what data would be needed for Sei Milano project and what type of sensors needed to be placed so that they can make the future predictions to handle urban heat island (UHI). Since there was weakness in data availability in phase 01, there was several limitations and proposals due to that:

- There was huge amount of error percentage in ARIMA and Prophet models while calculating the predictions for R10 universe project and that is shown due to the comparison with the benchmark PJMW.
- Usage of prophet model would be better due to less percentage of errors in both projects (R10 and PJMW)
- Gathering energy consumption for several years are needed to be able to make these predictions later.
- More sensors are needed to be placed after the construction phase is completed for both projects concerning energy and urban heat island so after that they will be able to track citizen's consumption which will be having more benefits for such development.
- Avoid having missing data within the period of collecting data from the sensors to avoid having huge percentage of errors in future.

Third Phase

The last phase was the conclusion of the thesis. By joining some important remarks from previous phases and based on literature review, recommendations were elaborated for companies' improvements later concerning this type of case studies. Furthermore, the final framework was able to create several recommendations within both projects and the four recommendations for the first case study are:

1. Usage of prophet model instead of ARIMA model due to lower percentage of errors.
2. Gather energy consumption for several years in developing countries projects.
3. Place more energy consumption sensors for several activities after site construction is built.
4. Avoid having missing data and focus on quality of data too not only on quantity.

Moreover, the recommendations for the second project Sei Milano, are:

1. Gather further data that is needed on actual materials and textures used to calculate UHI.
2. Place more sensors to be able to detect heat waves before they arrive to protect users.

Future developments

The methodology used for the first project, R10 universe, was able to proof that developing countries can benefit too from existence of artificial intelligence (A.I) and the digital transition revolution that is happening across smart cities across the world but that would be applicable after gathering and analysing data for several years to be able to predict energy consumption and it will help to achieve the sustainable development goals (SDGs). The methodology used for the second project, Sei Milano, was able to provide that smart heat warning system will decrease the heat issues and that they should specify materials for each usage. Moreover, these studies must be implemented from the early phases of a project, so that all materials that are being applied will be having a reasonable range of solar reflective index (SRI) to handle urban heat island on earlier stages.

Thus, this thesis hopes to improve the needed infrastructure knowledge and information so that a developing country can benefit from artificial intelligence evolution through having several affordable smart city projects around these cities. This will be having impact not only on economic and environmental development, but also on social development in communities due to spreading awareness and high education across citizens how to interact and engage with these digital platforms and transitions. An educational infrastructure will be needed for this type of projects to be able to be achieved in correct manners across several developing countries.

Within this framework, some future developments can be proposed to further encourage the thesis extension:

- For developing countries to benefit from artificial intelligence (A.I) they should implement several sensors across their projects to collect data for several years.
- Data must be collected with high quality of development to avoid having missing data and errors.
- Studies concerning materiality and SRI values must be taken into consideration in early phases of a project.
- Educational development and awareness must be spread across developing countries to be able to engage more citizens into the process.

These future developments will help to expand research in digital transitions in developing countries and enhance the performance of smart cities. These future developments will have huge impact on community and will be driving the city forward toward sustainability. These issues are essential to strengthen a common vision of digital transitions and usage of artificial intelligence, accompanied by future research.

Moreover, to fine tune the GIS-based IIAT dashboard through usage of artificial intelligence, it is important to be doing the process for all available indicators in the project which will help in:

- Improving accuracy through usage of artificial intelligence algorithms which will increase data analytics of historical energy consumption, weather data, building's characteristics, air quality, and traffic. Usage of machine learning algorithms will develop complex pattern's that might not be able to be analysed through traditional methods.
- Dynamic adaptation and optimization through adjusting urban systems based on different conditions due to real-time situations within the context. Smart cities will be able to be more efficient due within usage of these algorithms in GIS-based IIAT dashboard.
- Continuous learning since artificial intelligence can learn constantly within feedback and new information provided at the IIAT dashboard which makes the predictive models improving the decision-making algorithms based on urban dynamics.
- Automating the GIS-based IIAT dashboard while providing real-time data from sensors, IoT devices and citizen inputs within all the indicators in the smart city project which will help also in having personalized services and enhance citizen's engagement in the project. This will be also provided through dynamic visualization that is available in IIAT dashboard within real-time data such as dynamic filters based on AI-generated insights. Moreover, this will be helpful through usage of IIAT

parameterization within artificial intelligence predictions will allow creation of more accurate what-if scenarios within the IIAT dashboard.

List of References

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80–91. https://doi.org/10.1016/j.cities.2019.01.032

Alyssa Schroer. (2024, January 3). *What Is Artificial Intelligence?* Built In.

Anyoha, R. (2017, August 28). *The history of Artificial Intelligence* . Harvard University .

Deloitte. (2019). Transparency and Responsibility in Artificial Intelligence. *Deloitte*.

Desai, D. R., & Kroll, J. A. (2017). Trust But Verify: A Guide to Algorithms and the Law. *Harvard Journal of Law & Technology, Forthcoming*, 17–19.

Dickson, B. (2019, November 18). *What is symbolic artificial intelligence?* TechTalks .

European Parliament. (2023, December 19). *EU AI Act: first regulation on artificial intelligence*. European Parliament.

Expert.ai Team. (2022, March 14). *What is Machine Learning? A Definition.* Expert.Ai.

FIWARE. (2021, April 20). *Smart Irrigation System implemented in Cartagena's city*. FIWARE.

Garrido, L. (2023, September 28). *What does the green and digital transition mean for cities?* EURO CITIES.

Google. (2023). *Can AI Help Cool the World's Hottest Cities?* Bloomberg.

Grid Edge. (2019). *Case Study: Artificial Intelligence for Building Energy Management Systems*.

Holland Michel, A. (2020). *The Black Box, Unlocked: Predictability and Understandability in Military AI*. https://doi.org/10.37559/SecTec/20/AI1

Hurtado, P., & Gomez, A. (2021). Smart City Digital Twins Are a New Tool for Scenario Planning. *American Planning Association*.

Iberdrola. (2023). *Smart cities: the technological revolution reaches the cities*. Iberdrola.

Kamrowska-Załuska, D. (2021). Impact of AI-Based Tools and Urban Big Data Analytics on the Design and Planning of Cities. *Land*, *10*(11), 1209. https://doi.org/10.3390/land10111209

Kistler. (2019). *Timeline of AI*. AIArtists.

Laskowski, N. (2022). *What is Artificial Intelligence and How Does AI Work?* TechTarget.

Ly, D. (2023, April 7). *On The Horizon For Smart Cities: How AI And IoT Are Transforming Urban Living*. Forbes.

McKinsey Explainers. (2023). What is digital-twin technology? *McKinsey & Company*.

Mind AI. (2018). Lack of transparency could be AI's fatal flaw. *Medium*.

Ortega-Fernández, A., Martín-Rojas, R., & García-Morales, V. J. (2020). Artificial Intelligence in the Urban Environment: Smart Cities as Models for Developing Innovation and Sustainability. *Sustainability*, *12*(19), 7860. https://doi.org/10.3390/su12197860

Petrova-Antonova, D., & Ilieva, S. (2019). Methodological Framework for Digital Transition and Performance Assessment of Smart Cities. *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, 1–6. https://doi.org/10.23919/SpliTech.2019.8783170

Popelka, S. (2023). A Guide to Urban Artificial Intelligence. *Urban AI*.

Robare, M. (2023). Digital Twins used in Urban Planning and Infrastructure. *Urban SDK*.

Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, 100005. https://doi.org/10.1016/j.jrt.2020.100005

salesForce. (2022, May). *What Is Digital Transformation?* SalesForce.

Stevens, E. (2023, May 10). *The 7 Most Useful Data Analysis Methods and Techniques*. CareerFoundry.

Vessali, K., Galal, H., & Dr. Nowson, S. (2022). *How digital twins can make smart cities better*.

Yi, W. (2022). *Planning Urban Cities Smartly With Digital Twins*.

Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns*, *3*(4), 100455. https://doi.org/10.1016/j.patter.2022.100455

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

APPENDIX A

R10 UNIVERSE Project for Predicting energy consumption.

The R10 universe project energy consumption predictions through usage of time series analysis through SRIMA and Prophet forecasting methods are done through anaconda software – jupyter notebook and python programming language, the codes below in appendix A shows the prediction for energy consumption codes.

In [1]:
```
pip install statsmodels
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: statsmodels in c:\programdata\anaconda3\lib\site-packages (0.13.2)
Requirement already satisfied: numpy>=1.17 in c:\users\asus\appdata\roaming\python\python39\site-packages (from statsmodels) (1.23.5)
Requirement already satisfied: scipy>=1.3 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (1.9.1)
Requirement already satisfied: pandas>=0.25 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (1.4.4)
Requirement already satisfied: patsy>=0.5.2 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (0.5.2)
Requirement already satisfied: packaging>=21.3 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\programdata\anaconda3\lib\site-packages (from packaging>=21.3->statsmodels) (3.0.9)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->statsmodels) (2022.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->statsmodels) (2.8.2)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
pip install pbr
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pbr in c:\users\asus\appdata\roaming\python\python39\site-packages (6.0.0)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
pip install prophet
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: prophet in c:\users\asus\appdata\roaming\python\python39\site-packages (1.1.5)
Requirement already satisfied: matplotlib>=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from prophet) (3.5.2)
Requirement already satisfied: numpy>=1.15.4 in c:\users\asus\appdata\roaming\python\python39\site-packages (from prophet) (1.23.5)
Requirement already satisfied: pandas>=1.0.4 in c:\programdata\anaconda3\lib\site-packages (from prophet) (1.4.4)
Requirement already satisfied: tqdm>=4.36.1 in c:\programdata\anaconda3\lib\site-packages (from prophet) (4.64.1)
Requirement already satisfied: holidays>=0.25 in c:\users\asus\appdata\roaming\python\python39\site-packages (from prophet) (0.38)
Requirement already satisfied: importlib-resources in c:\users\asus\appdata\roaming\python\python39\site-packages (from prophet) (6.1.1)
Requirement already satisfied: cmdstanpy>=1.0.4 in c:\users\asus\appdata\roaming\python\python39\site-packages (from prophet) (1.2.0)
Requirement already satisfied: stanio~=0.3.0 in c:\users\asus\appdata\roaming\python\python39\site-packages (from cmdstanpy>=1.0.4->prophet) (0.3.0)
Requirement already satisfied: python-dateutil in c:\programdata\anaconda3\lib\site-packages (from holidays>=0.25->prophet) (2.8.2)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (1.4.2)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (0.11.0)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (21.3)
Requirement already satisfied: pyparsing>=2.2.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (3.0.9)
Requirement already satisfied: pillow>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.0.0->prophet) (9.2.0)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=1.0.4->prophet) (2022.1)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-packages (from tqdm>=4.36.1->prophet) (0.4.5)
Requirement already satisfied: zipp>=3.1.0 in c:\programdata\anaconda3\lib\site-packages (from importlib-resources->prophet) (3.8.0)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil->holidays>=0.25->prophet) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [4]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import os
import sys
import warnings
warnings.filterwarnings('ignore')
import matplotlib.pyplot as plt
from prophet import Prophet
```

In [5]:
```python
import os
import sys

module_path = os.path.abspath(os.path.join('../../'))
if module_path not in sys.path:
    sys.path.append(module_path)

# Define project directories
project_dir = os.path.join(module_path, "Timeseries")
data_dir = os.path.join(project_dir, 'data')
cur_dir = os.path.join(project_dir, 'notebooks')
image_dir = os.path.join(project_dir, 'images')

# Check and create directories if they don't exist
for directory in [project_dir, data_dir, cur_dir, image_dir]:
    if not os.path.exists(directory):
        os.makedirs(directory)
        print(f"Directory created: {directory}")
    else:
        print(f"Directory already exists: {directory}")

# List files in the data directory
if os.path.exists(data_dir):
    list_files = os.listdir(data_dir)
    # Use list_files or perform operations on the files within 'data_dir'
    # Example: for file in list_files: load_data(os.path.join(data_dir, file))
    # Uncomment the line below to print the list of files
    # print(list_files)
else:
    print("Data directory does not exist:", data_dir)
```

```
Directory already exists: C:\Timeseries
Directory already exists: C:\Timeseries\data
Directory already exists: C:\Timeseries\notebooks
Directory already exists: C:\Timeseries\images
```

In [6]:
```python
# Read the CSV file and perform operations
if os.path.exists(data_dir):
    # File path for the CSV file
    csv_file_path = os.path.join(data_dir, "Planet.csv")

    # Check if the CSV file exists
    if os.path.exists(csv_file_path):
        # Read CSV into DataFrame
        df = pd.read_csv(csv_file_path)

        # Process DataFrame
        df.index = pd.to_datetime(df['DATE'])
        df.drop(columns='DATE', inplace=True)
        df.head()
    else:
        print("CSV file does not exist:", csv_file_path)
else:
    print("Data directory does not exist:", data_dir)

df.head()
```

Out[6]:

DATE	TOWER	MONITORING	METER	HOUR	CONSUMPTION
2022-06-01	A	Finishing	Tower_A_Finishing	0.0	0.0
2022-06-01	A	Finishing	Tower_A_Finishing	1.0	0.0
2022-06-01	A	Finishing	Tower_A_Finishing	2.0	0.0
2022-06-01	A	Finishing	Tower_A_Finishing	3.0	0.0
2022-06-01	A	Finishing	Tower_A_Finishing	4.0	0.0

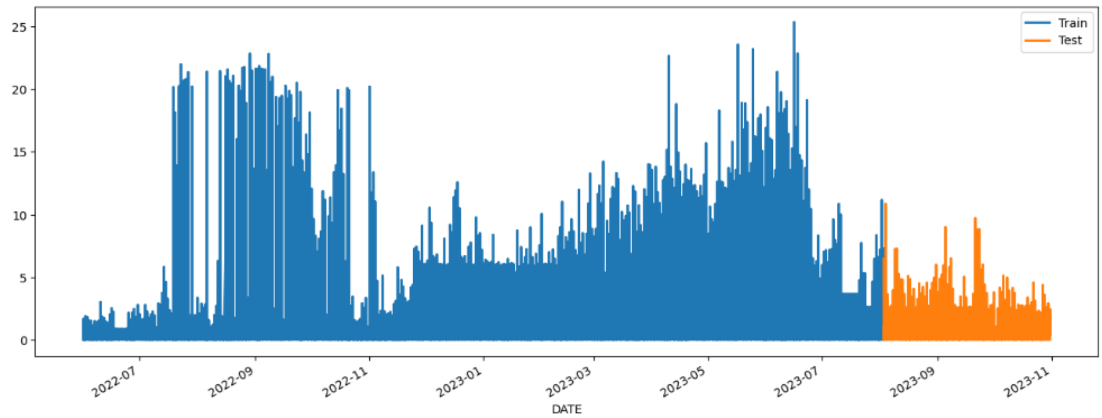
In [7]:
```python
df.describe()
df.columns
```

Out[7]:
```
Index(['TOWER', 'MONITORING', 'METER', 'HOUR', 'CONSUMPTION'], dtype='object')
```

In [8]:
```python
df.index
```

Out[8]:
```
DatetimeIndex(['2022-06-01', '2022-06-01', '2022-06-01', '2022-06-01',
               '2022-06-01', '2022-06-01', '2022-06-01', '2022-06-01',
               '2022-06-01', '2022-06-01',
               ...
               '2022-06-13', '2022-06-13', '2022-06-13', '2022-06-13',
               '2022-06-13', '2022-06-13', '2022-06-13', '2022-06-13',
                      'NaT',        'NaT'],
              dtype='datetime64[ns]', name='DATE', length=150002, freq=None)
```

In [9]:
```python
import os
import pandas as pd

# Rest of your code that creates directories and other operations

# Function to read CSV and perform operations
def read_csv_and_operations():
    # File path for the CSV file
    csv_file_path = os.path.join(data_dir, 'Planet.csv')

    # Check if the CSV file exists
    if os.path.exists(csv_file_path):
        # Read CSV into DataFrame
        df = pd.read_csv(csv_file_path)
        df.index = pd.to_datetime(df['DATE'])
        df.drop(columns='DATE', inplace=True)
        return df
    else:
        print("CSV file does not exist:", csv_file_path)
        return None

# Function to convert DataFrame to Prophet format
def convert_to_prophet(df):
    df = df.reset_index()
    df.columns = ['ds', 'y']
    df.index = df.ds
    df.index.names = ['DATEIndex']
    return df

# Function to get train and test data
def get_train_test_generic(for_prophet=False, freq=None):
    df = read_csv_and_operations()
    if df is not None:
        df.sort_index(inplace=True)
        if freq is not None:
            df = df[['CONSUMPTION']].resample(freq).sum()
        train_len = int(0.8 * len(df))
        train_df = df.iloc[:train_len].copy()
        test_df = df.iloc[train_len:].copy()  # Define test_df

        if for_prophet:
            train_df = convert_to_prophet(train_df)
            test_df = convert_to_prophet(test_df)

        return train_df, test_df, df

    else:
        # Return None for train, test, df if CSV reading failed
        return None, None, None

# Call the function to get train and test data
tr_df, te_df, df = get_train_test_generic()

# Continue with plotting or further analysis
if tr_df is not None:
    from pylab import rcParams
    import matplotlib.pyplot as plt

    rcParams['figure.figsize'] = 16, 6
    tr_df['CONSUMPTION'].plot(label='Train', legend=True, linewidth=2)
    te_df['CONSUMPTION'].plot(label='Test', legend=True, linewidth=2)
    plt.show()
```



In [10]:
```
tr_df, te_df, df =get_train_test_generic()
display(tr_df)
display(te_df)
```

	TOWER	MONITORING	METER	HOUR	CONSUMPTION
DATE					
2022-06-01	A	Finishing	Tower_A_Finishing	0.0	0.000000
2022-06-01	F	Lift	Tower_F_Lift	2.0	0.000000
2022-06-01	F	Lift	Tower_F_Lift	3.0	0.000000
2022-06-01	F	Lift	Tower_F_Lift	4.0	0.000000
2022-06-01	F	Lift	Tower_F_Lift	5.0	0.000000
...	...	...	...	...	...
2023-08-03	Clubhouse	General	ClubhousGeneral	20.0	3.826667
2023-08-03	Clubhouse	General	ClubhousGeneral	19.0	3.826667
2023-08-03	Clubhouse	General	ClubhousGeneral	18.0	3.826667
2023-08-03	Clubhouse	General	ClubhousGeneral	17.0	7.290000
2023-08-03	Clubhouse	General	ClubhousGeneral	16.0	5.840000

120001 rows × 5 columns

	TOWER	MONITORING	METER	HOUR	CONSUMPTION
DATE					
2023-08-03	Clubhouse	General	ClubhousGeneral	15.0	6.46
2023-08-03	Clubhouse	General	ClubhousGeneral	14.0	5.65
2023-08-03	Clubhouse	General	ClubhousGeneral	12.0	1.39
2023-08-03	F	Finishing	Tower_F_Finishing	11.0	0.00
2023-08-03	B	RCC	Tower_B_RCC	3.0	3.86
...	...	...	...	...	...
2023-10-31	A	PM&Hoist	Tower_A_PM&Hoist	4.0	0.02
2023-10-31	A	PM&Hoist	Tower_A_PM&Hoist	2.0	0.01
2023-10-31	A	RCC	Tower_A_RCC	14.0	0.42
NaT	NaN	NaN	NaN	NaN	NaN
NaT	NaN	NaN	NaN	NaN	NaN

30001 rows × 5 columns

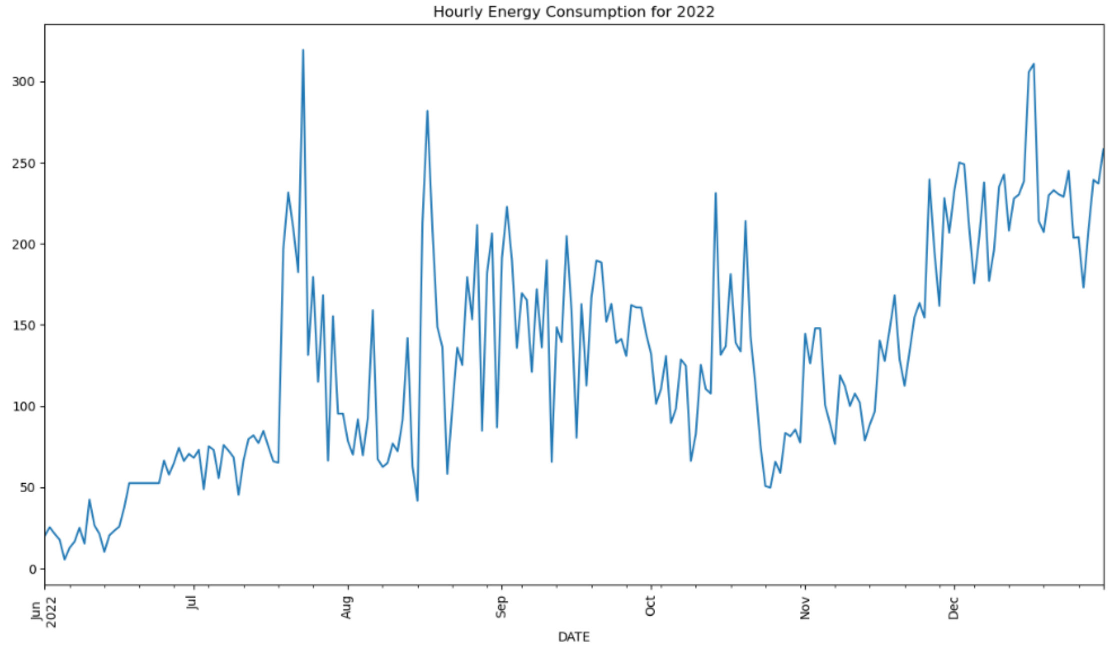
In [11]:
```
tr_df, te_df, df =get_train_test_generic(freq='1D')
tr_df.head()
```

Out[11]:

	CONSUMPTION
DATE	
2022-06-01	20.1200
2022-06-02	25.3800
2022-06-03	21.3100
2022-06-04	17.6700
2022-06-05	5.4964

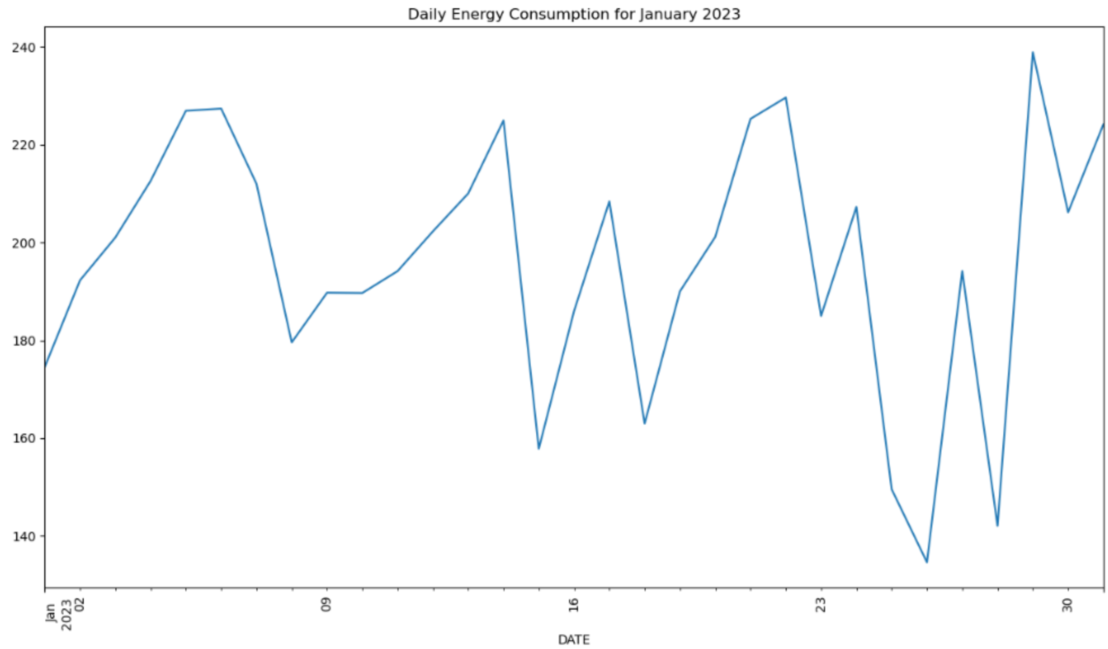
In [12]:
```
#Hourly energy consumption across year 2017
df[df.index.year == 2022]['CONSUMPTION'].plot(figsize=(15, 8), title='Hourly Energy Consumption for 2022', rot=90)
```

Out[12]: <AxesSubplot:title={'center':'Hourly Energy Consumption for 2022'}, xlabel='DATE'>



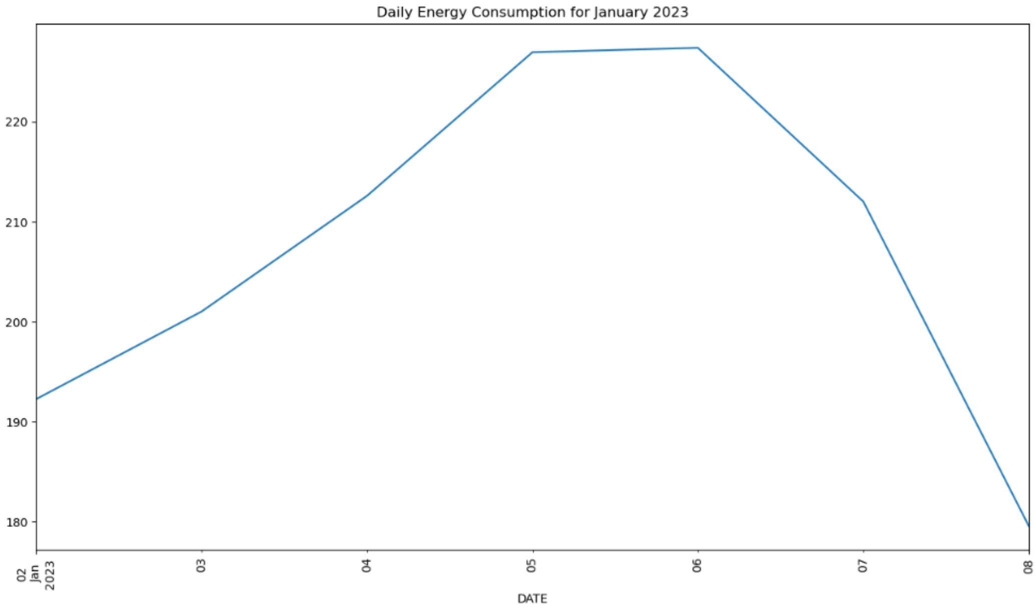
In [13]:
```
df[(df.index.year == 2023) & (df.index.month == 1)]['CONSUMPTION'].resample('D').mean().plot(figsize=(15, 8), title='Daily Energ
```

Out[13]: <AxesSubplot:title={'center':'Daily Energy Consumption for January 2023'}, xlabel='DATE'>



In [14]:
```
df[(df.index.year == 2023) & (df.index.month == 1)& (df.index.week==1)]['CONSUMPTION'].resample('D').mean().plot(figsize=(15, 8),
```

Out[14]: <AxesSubplot:title={'center':'Daily Energy Consumption for January 2023'}, xlabel='DATE'>

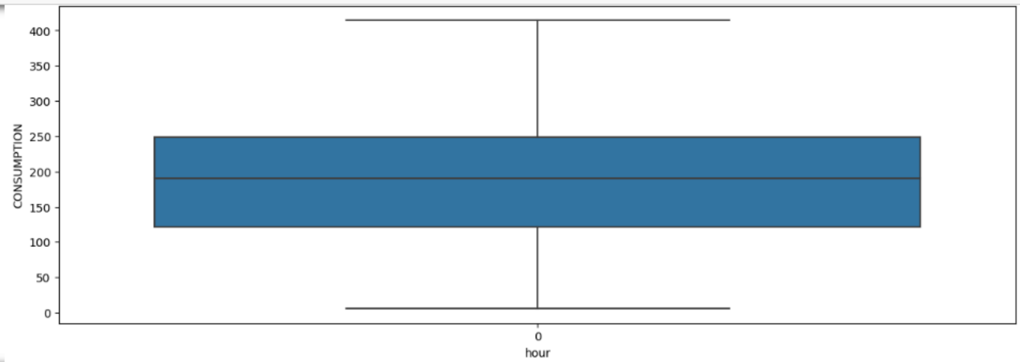


In [15]:
```
#Visualization by Group
df['date'] = df.index.date
df['year'] = df.index.year
df['quarter'] = df.index.quarter
df['month'] = df.index.month
df['dom'] = df.index.day #day of month
df['dow'] = df.index.dayofweek
df['doy'] = df.index.dayofyear
df['weekday'] = df.index.day_name()
df['hour'] = df.index.hour
df.head()
```

Out[15]:

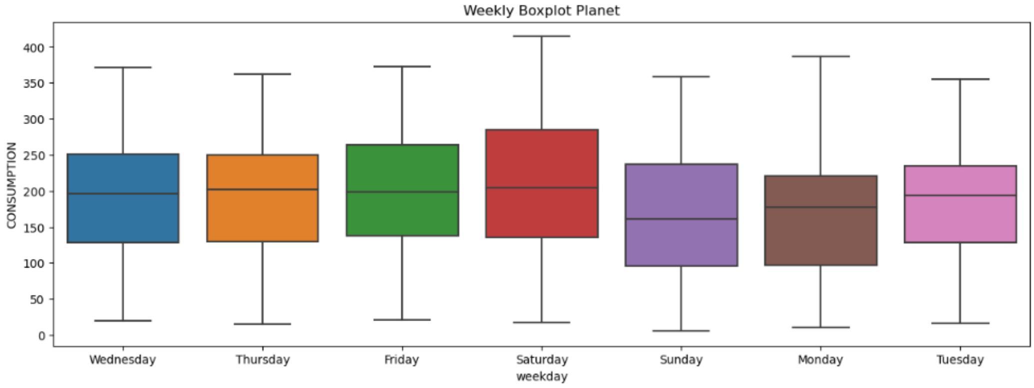
DATE	CONSUMPTION	date	year	quarter	month	dom	dow	doy	weekday	hour
2022-06-01	20.1200	2022-06-01	2022	2	6	1	2	152	Wednesday	0
2022-06-02	25.3800	2022-06-02	2022	2	6	2	3	153	Thursday	0
2022-06-03	21.3100	2022-06-03	2022	2	6	3	4	154	Friday	0
2022-06-04	17.6700	2022-06-04	2022	2	6	4	5	155	Saturday	0
2022-06-05	5.4964	2022-06-05	2022	2	6	5	6	156	Sunday	0

In [16]:
```
#Hourly energy consumption - Low at morning and some peak times around 6 PM
import seaborn as sns
import matplotlib.pyplot as plt
fig, ax=plt.subplots(figsize=(15,5))
sns.boxplot(df.hour, df.CONSUMPTION)
ax.set_title('Hourly Boxplot PJME')
```



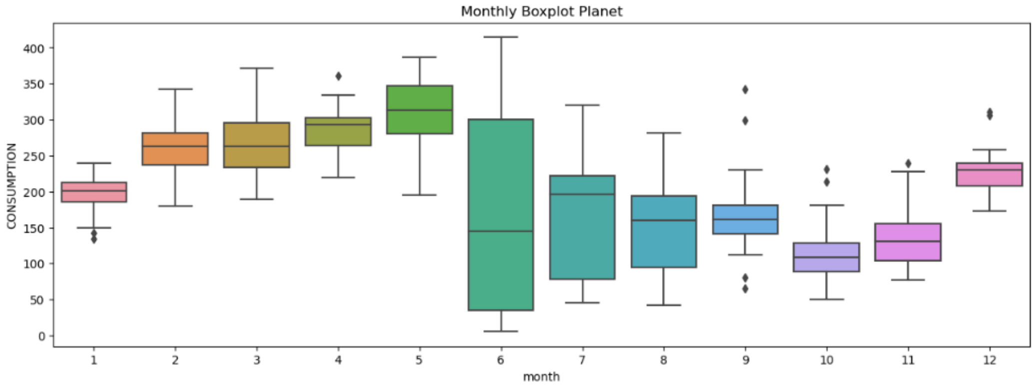
In [17]:
```
#Hourly energy consumption - Low at morning and some peak times around 6 PM
import seaborn as sns
import matplotlib.pyplot as plt
fig, ax=plt.subplots(figsize=(15,5))
sns.boxplot(df.weekday, df.CONSUMPTION)
ax.set_title('Weekly Boxplot Planet')
```

Out[17]: Text(0.5, 1.0, 'Weekly Boxplot Planet')



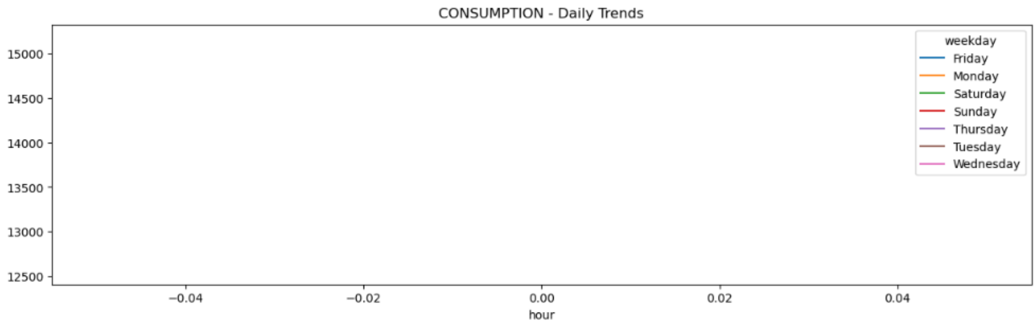
In [18]:
```
#Hourly energy consumption - Low at morning and some peak times around 6 PM
import seaborn as sns
import matplotlib.pyplot as plt
fig, ax=plt.subplots(figsize=(15,5))
sns.boxplot(df.month, df.CONSUMPTION)
ax.set_title('Monthly Boxplot Planet')
```

Out[18]: Text(0.5, 1.0, 'Monthly Boxplot Planet')

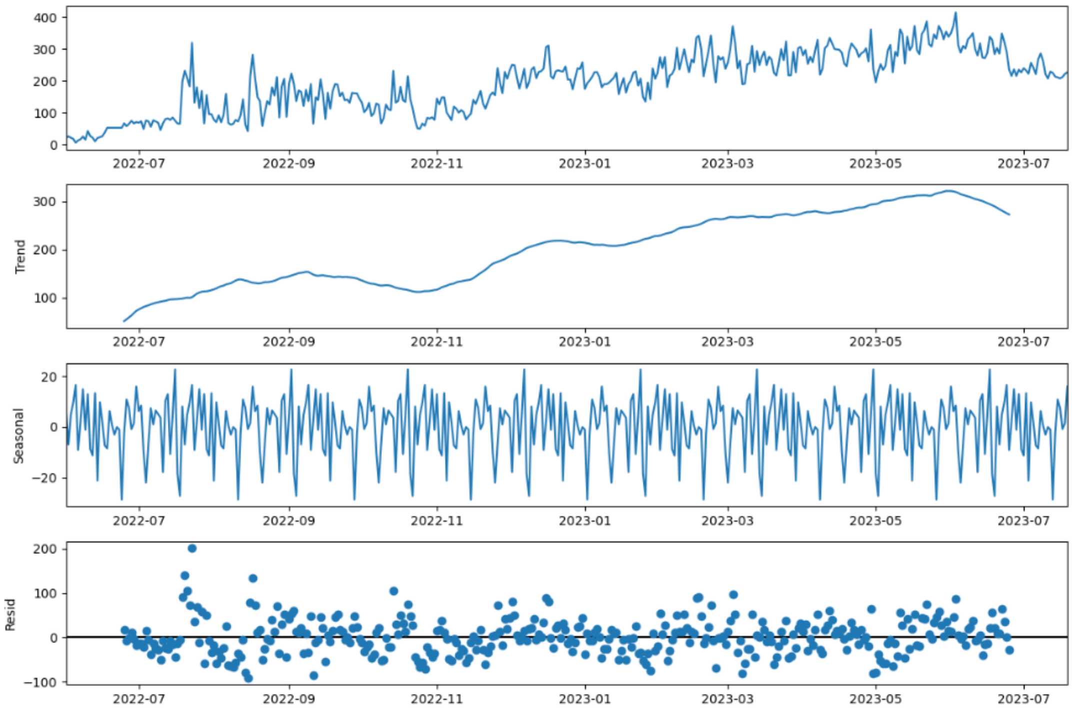


In [19]:
```
#Weekday and Hour energy consumption
df2= df.pivot_table(index=df['hour'], columns='weekday', values='CONSUMPTION', aggfunc='sum')
df2.head()
df2.plot(figsize=(15,4), title='CONSUMPTION - Daily Trends')
#Weekends behavior is different from weekdays behavior - useful info in modeling
```

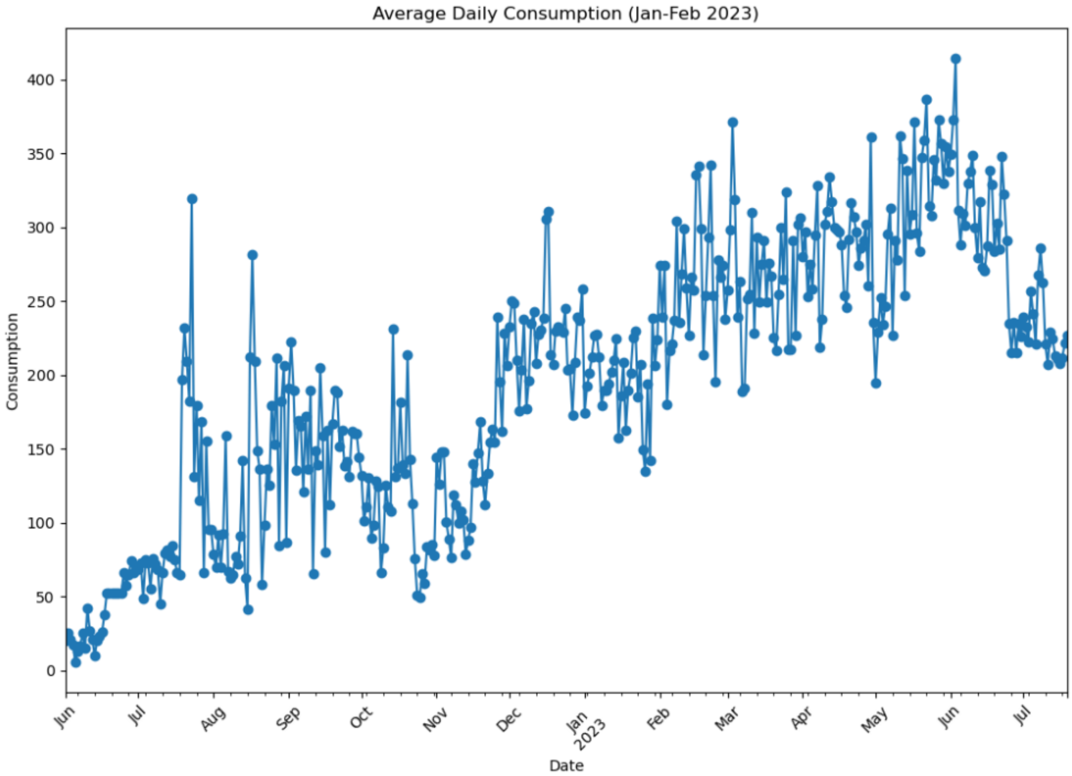
Out[19]: <AxesSubplot:title={'center':'CONSUMPTION - Daily Trends'}, xlabel='hour'>



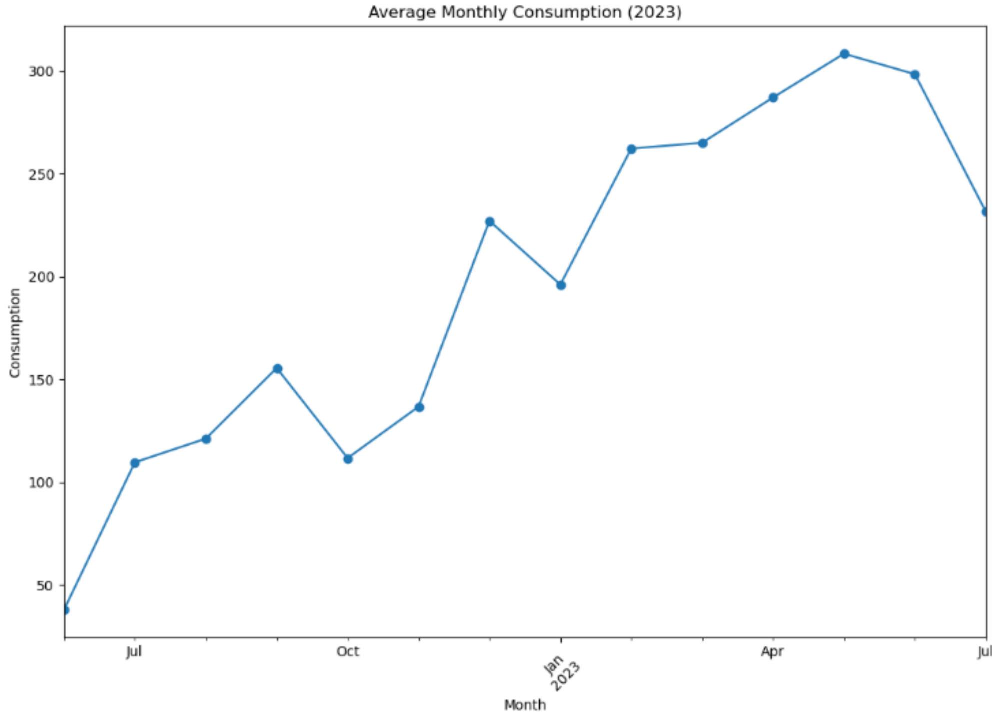
In [20]:
```
#Create data functions
#ETS Decomposition
from statsmodels.tsa.seasonal import seasonal_decompose
result = seasonal_decompose(tr_df[['CONSUMPTION']], period = 48)
from pylab import rcParams
rcParams['figure.figsize']=12,8
fig = result.plot()

#seasonality is very dense ,so we take a small period and we look at it as in the next step (daily seasonality)
```



In [21]:
```
# Assuming 'tr_sub' contains your subset DataFrame with hourly data
import matplotlib.pyplot as plt

tr_df['Date'] = tr_df.index.date  # Extracting only the date
consumption_by_date = tr_df.groupby('DATE')['CONSUMPTION'].mean()  # Aggregate by date

# Plotting consumption without hour
plt.figure(figsize=(12, 8))
consumption_by_date.plot(title='Average Daily Consumption (Jan-Feb 2023)', marker='o')
plt.xlabel('Date')
plt.ylabel('Consumption')
plt.xticks(rotation=45)
plt.show()
```



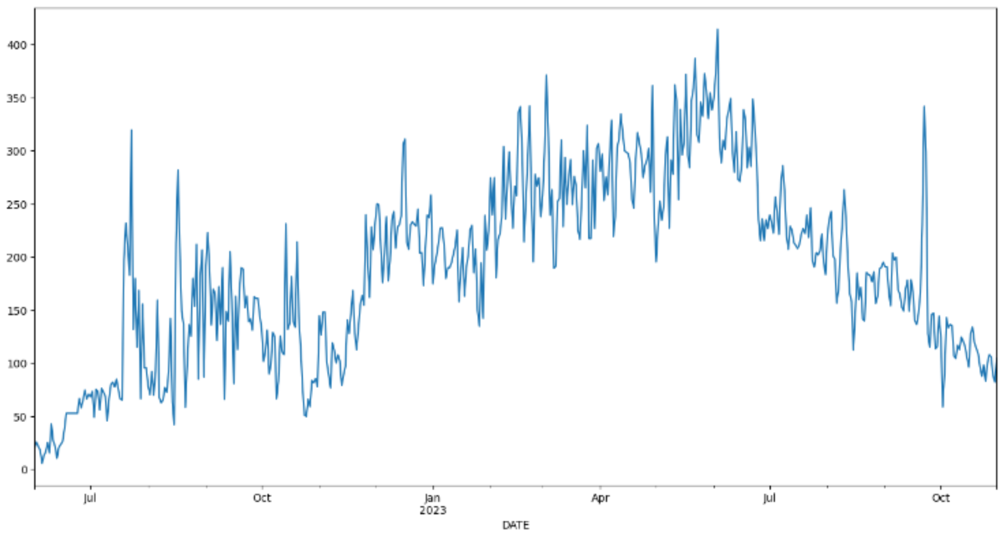
In [22]:
```
# Assuming 'tr_df' contains your entire dataset
import matplotlib.pyplot as plt

tr_df['Month'] = tr_df.index.to_period('M')  # Extracting the month
monthly_consumption = tr_df.groupby('Month')['CONSUMPTION'].mean()  # Aggregate by month

# Plotting monthly consumption
plt.figure(figsize=(12, 8))
monthly_consumption.plot(title='Average Monthly Consumption (2023)', marker='o')
plt.xlabel('Month')
plt.ylabel('Consumption')
plt.xticks(rotation=45)
plt.show()
```



In [23]:
```
# resampled view of data at aggregated level
# daily, weekly and monthly

daily_df = df['CONSUMPTION'].resample('1D').mean()
display(daily_df.head())
rcParams['figure.figsize'] = 16,8
daily_df.loc[daily_df.index>'2022'].plot()
```

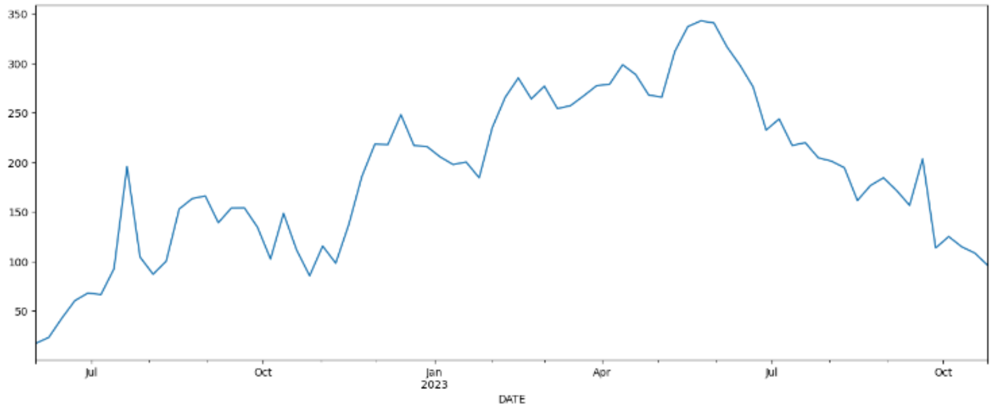
```
DATE
2022-06-01    20.1200
2022-06-02    25.3800
2022-06-03    21.3100
2022-06-04    17.6700
2022-06-05     5.4964
Freq: D, Name: CONSUMPTION, dtype: float64
```

Out[23]: <AxesSubplot:xlabel='DATE'>



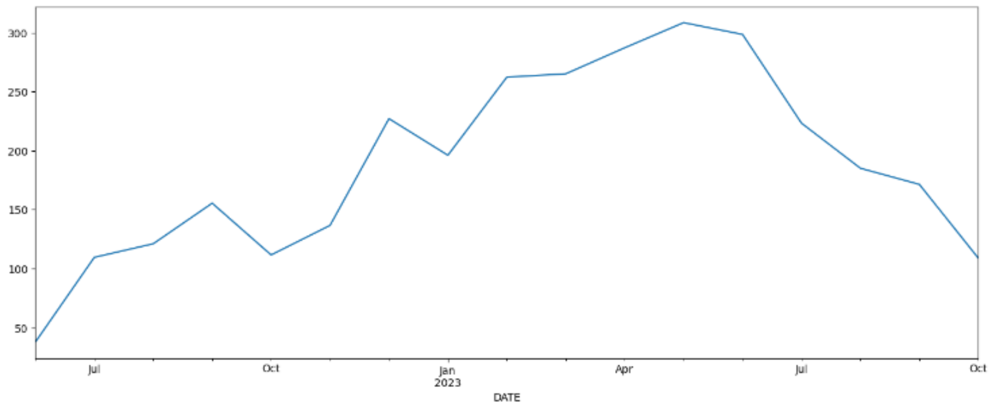
In [24]:
```
daily_df = df['CONSUMPTION'].resample('7D').mean()
rcParams['figure.figsize']=16,6
daily_df.loc[daily_df.index> '2022'].plot()
#resampling at 7 days (7D)
```

Out[24]: <AxesSubplot:xlabel='DATE'>



In [25]:
```
daily_df = df['CONSUMPTION'].resample('1M').mean()
rcParams['figure.figsize']=16,6
daily_df.loc[daily_df.index> '2022'].plot()
#resampling at monthly level
```

Out[25]: <AxesSubplot:xlabel='DATE'>



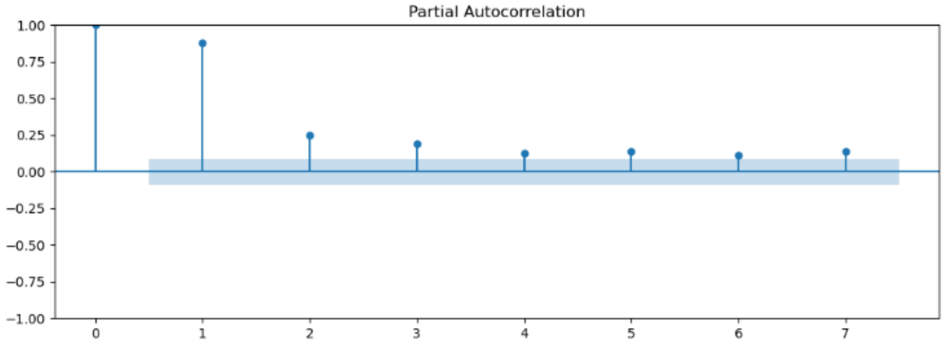
In [26]:
```
#Useful validation metrics and plots
from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_absolute_percentage_error
def validate (test_df, pred_df, var):
    pred_res= pred_df[var]
    test= test_df[var]
    mse= mean_squared_error(pred_res,test)
    rmse = mean_squared_error(pred_res,test, squared=False)
    mae = mean_absolute_error(pred_res, test)
    mape= mean_absolute_percentage_error(pred_res,test)
    print("mse={:.3f}, rmse={:.3f}, mae={:.3f}, mape={:.3f}".format(mse, rmse, mae, mape))
    return mse, rmse, mae, mape

#make sure the index is datetime variable
def plot_and_validate(train_df, test_df, pred_df, var):
    rcParams['figure.figsize'] = 16,6
    train_df[var].plot(label='Train', legend=True, linewidth=2)
    test_df[var].plot(label='Test', legend=True, linewidth=2)
    pred_df[var].plot(label='Pred', legend=True, linewidth=2)
    validate(test_df, pred_df, var)

def smape_func(a,f):
    return 1/len(a) * np.sum(2*np.abs(f-a)/(np.abs(a)+np.abs(f)))

def validate(test_df, pred_df, te_var, pred_var):
    merged = test_df.merge(pred_df, left_index=True, right_index=True)
    pred_res = merged[te_var]
    test = merged[pred_var]
    mse = mean_squared_error(pred_res, test)
    rmse = mean_squared_error(pred_res, test, squared=False)
    mae=mean_absolute_error(pred_res, test)
    mape = mean_absolute_percentage_error(pred_res, test)
    smape = smape_func(test, pred_res)
    print("mse={:.3f}, rmse={:.3f}, mae={:.3f}, mape={:.3f}, smape={:.3f}".format(mse, rmse, mae, mape, smape))
    return mse, rmse, mae, mape, smape

def plot_and_validate(train_df, test_df, pred_df, te_var, pred_var):
    rcParams['figure.figsize'] = 16, 6
    train_df[te_var].plot(label='Train', legend=True, linewidth=2)
    test_df[te_var].plot(label='Test', legend=True, linewidth=2)
    pred_df[pred_var].plot(label='Pred', legend=True, linewidth=2)
    validate(test_df, pred_df, te_var, pred_var)
```

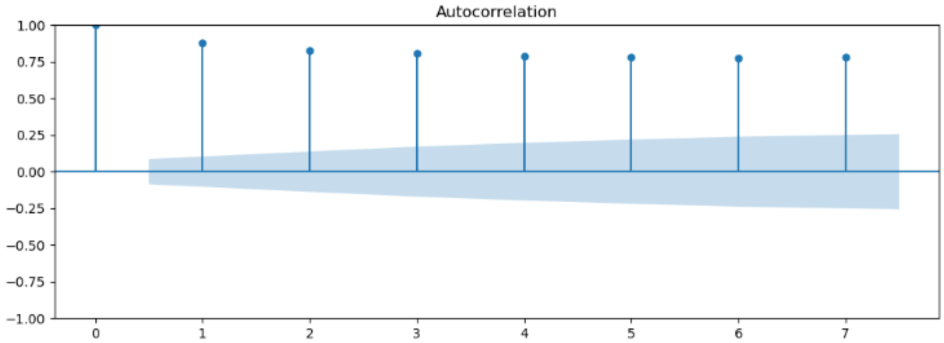
In [27]:

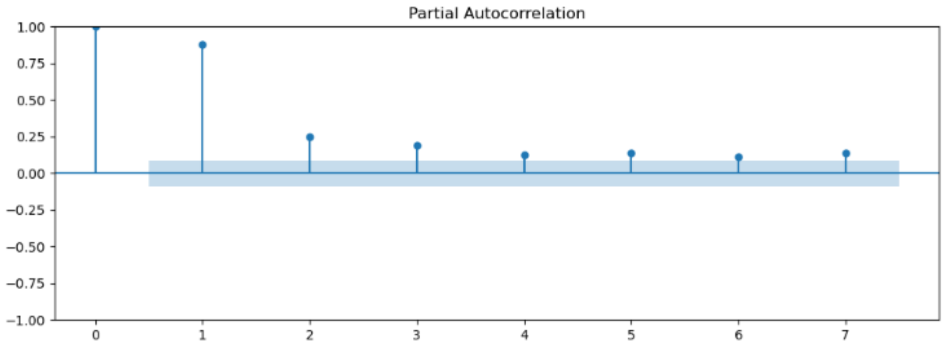
```python
# Autocorrelation plots
from statsmodels.tsa.stattools import acf, pacf
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
print(acf(df['CONSUMPTION']))
rcParams['figure.figsize'] = 12, 4
plot_acf(df['CONSUMPTION'], lags=7)
plot_pacf(df['CONSUMPTION'], lags=7)
```

```
[1.         0.87978912 0.82912929 0.80539097 0.7866191  0.77983821
 0.7750167  0.77857717 0.74937186 0.72589817 0.72621873 0.70393886
 0.70365074 0.70602223 0.71349155 0.6934315  0.67248355 0.65626429
 0.64362543 0.64178048 0.64236266 0.64318725 0.62232194 0.60082889
 0.59631172 0.58596233 0.57506506 0.58959998]
```

Out[27]:







In [28]:

```
pip install pmdarima
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pmdarima in c:\users\asus\appdata\roaming\python\python39\site-packages (2.0.4)
Requirement already satisfied: pandas>=0.19 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (1.4.4)
Requirement already satisfied: numpy>=1.21.2 in c:\users\asus\appdata\roaming\python\python39\site-packages (from pmdarima) (1.
23.5)
Requirement already satisfied: statsmodels>=0.13.2 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (0.13.2)
Requirement already satisfied: joblib>=0.11 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (1.1.0)
Requirement already satisfied: scipy>=1.3.2 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (1.9.1)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (63.4.
1)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in c:\programdata\anaconda3\lib\site-packages (from pmdarima)
(0.29.32)
Requirement already satisfied: scikit-learn>=0.22 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (1.0.2)
Requirement already satisfied: packaging>=17.1 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (21.3)
Requirement already satisfied: urllib3 in c:\programdata\anaconda3\lib\site-packages (from pmdarima) (1.26.11)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\programdata\anaconda3\lib\site-packages (from packaging>=17.1->pm
darima) (3.0.9)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.19->pmdarima) (2022.
1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.19->pmdarim
a) (2.8.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn>=0.22->pmd
arima) (2.2.0)
Requirement already satisfied: patsy>=0.5.2 in c:\programdata\anaconda3\lib\site-packages (from statsmodels>=0.13.2->pmdarima)
(0.5.2)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels>=0.13.2->pmdar
ima) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [29]:
```
print(tr_df.info())  # Check the information about the DataFrame
print(tr_df.head())  # Display the first few rows of the DataFrame
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 414 entries, 2022-06-01 to 2023-07-19
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   CONSUMPTION  414 non-null    float64
 1   Date         414 non-null    object
 2   Month        414 non-null    period[M]
dtypes: float64(1), object(1), period[M](1)
memory usage: 12.9+ KB
None
            CONSUMPTION        Date    Month
DATE
2022-06-01      20.1200  2022-06-01  2022-06
2022-06-02      25.3800  2022-06-02  2022-06
2022-06-03      21.3100  2022-06-03  2022-06
2022-06-04      17.6700  2022-06-04  2022-06
2022-06-05       5.4964  2022-06-05  2022-06
```

In [30]:
```
#Auto ARIMA
import time
from pmdarima import auto_arima
start = time.time()
tr_df, te_df, df = get_train_test_generic(freq='1D')
display(tr_df.head())
model = auto_arima(tr_df, m=12, trace=True, njobs=-1) #calling auto_arima
print(model)
print("Total time taken= {}".format(time.time()-start))
#the results are the best model
```

DATE	CONSUMPTION
2022-06-01	20.1200
2022-06-02	25.3800
2022-06-03	21.3100
2022-06-04	17.6700
2022-06-05	5.4964

```
Performing stepwise search to minimize aic
 ARIMA(2,1,2)(1,0,1)[12] intercept   : AIC=4199.038, Time=1.19 sec
 ARIMA(0,1,0)(0,0,0)[12] intercept   : AIC=4300.365, Time=0.02 sec
 ARIMA(1,1,0)(1,0,0)[12] intercept   : AIC=4248.196, Time=0.18 sec
 ARIMA(0,1,1)(0,0,1)[12] intercept   : AIC=4204.710, Time=0.23 sec
 ARIMA(0,1,0)(0,0,0)[12]             : AIC=4298.419, Time=0.02 sec
 ARIMA(2,1,2)(0,0,1)[12] intercept   : AIC=4197.403, Time=0.72 sec
 ARIMA(2,1,2)(0,0,0)[12] intercept   : AIC=4195.726, Time=0.37 sec
 ARIMA(2,1,2)(1,0,0)[12] intercept   : AIC=4197.403, Time=0.79 sec
 ARIMA(1,1,2)(0,0,0)[12] intercept   : AIC=4193.678, Time=0.21 sec
 ARIMA(1,1,2)(1,0,0)[12] intercept   : AIC=4195.270, Time=0.41 sec
 ARIMA(1,1,2)(0,0,1)[12] intercept   : AIC=4195.271, Time=0.47 sec
 ARIMA(1,1,2)(1,0,1)[12] intercept   : AIC=4196.852, Time=0.75 sec
 ARIMA(0,1,2)(0,0,0)[12] intercept   : AIC=4193.376, Time=0.14 sec
 ARIMA(0,1,2)(1,0,0)[12] intercept   : AIC=4195.072, Time=0.26 sec
 ARIMA(0,1,2)(0,0,1)[12] intercept   : AIC=4195.069, Time=0.28 sec
 ARIMA(0,1,2)(1,0,1)[12] intercept   : AIC=4196.822, Time=0.58 sec
 ARIMA(0,1,1)(0,0,0)[12] intercept   : AIC=4203.724, Time=0.09 sec
 ARIMA(0,1,3)(0,0,0)[12] intercept   : AIC=4194.201, Time=0.19 sec
 ARIMA(1,1,1)(0,0,0)[12] intercept   : AIC=4191.831, Time=0.14 sec
 ARIMA(1,1,1)(1,0,0)[12] intercept   : AIC=4193.467, Time=0.30 sec
 ARIMA(1,1,1)(0,0,1)[12] intercept   : AIC=4193.467, Time=0.34 sec
 ARIMA(1,1,1)(1,0,1)[12] intercept   : AIC=4195.073, Time=0.55 sec
 ARIMA(1,1,0)(0,0,0)[12] intercept   : AIC=4247.831, Time=0.06 sec
 ARIMA(2,1,1)(0,0,0)[12] intercept   : AIC=4193.690, Time=0.18 sec
 ARIMA(2,1,0)(0,0,0)[12] intercept   : AIC=4225.874, Time=0.10 sec
 ARIMA(1,1,1)(0,0,0)[12]             : AIC=4190.861, Time=0.07 sec
 ARIMA(1,1,1)(1,0,0)[12]             : AIC=4192.539, Time=0.15 sec
 ARIMA(1,1,1)(0,0,1)[12]             : AIC=4192.541, Time=0.18 sec
 ARIMA(1,1,1)(1,0,1)[12]             : AIC=4194.132, Time=0.34 sec
 ARIMA(0,1,1)(0,0,0)[12]             : AIC=4202.214, Time=0.04 sec
 ARIMA(1,1,0)(0,0,0)[12]             : AIC=4245.940, Time=0.04 sec
 ARIMA(2,1,1)(0,0,0)[12]             : AIC=4192.764, Time=0.10 sec
 ARIMA(1,1,2)(0,0,0)[12]             : AIC=4192.758, Time=0.11 sec
 ARIMA(0,1,2)(0,0,0)[12]             : AIC=4192.234, Time=0.06 sec
 ARIMA(2,1,0)(0,0,0)[12]             : AIC=4224.047, Time=0.05 sec
 ARIMA(2,1,2)(0,0,0)[12]             : AIC=4194.740, Time=0.21 sec

Best model:  ARIMA(1,1,1)(0,0,0)[12]
Total fit time: 10.028 seconds
 ARIMA(1,1,1)(0,0,0)[12]
Total time taken= 10.232279777526855
```

In [31]:
```
#auto_arima for weekly level
import time
start = time.time()
tr_df, te_df, df = get_train_test_generic(freq='7D')
display(tr_df)
from pmdarima import auto_arima
model  = auto_arima(tr_df, max_p = 2, start_q = 0, max_q = 1,
                   max_P = 2, max_Q = 1,
                   m=52, trace=True, njobs=-1, maxiter=10)
print(model)
print("Total time taken = {}".format(time.time() - start))
```

DATE	CONSUMPTION
2022-06-01	119.142609
2022-06-08	161.193427
2022-06-15	296.653166
2022-06-22	421.126834
2022-06-29	474.810000
2022-07-06	464.160000
2022-07-13	646.480000
2022-07-20	1368.950000
2022-07-27	729.204621
2022-08-03	607.994626

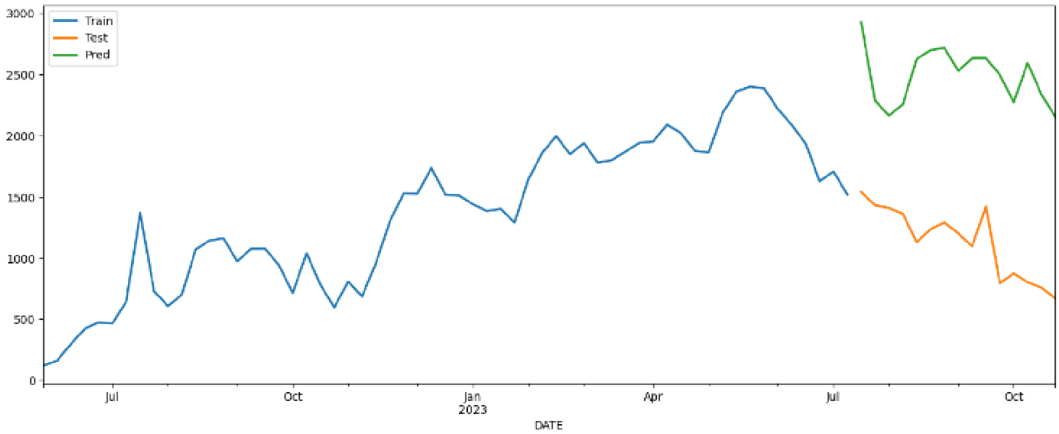
In [32]:
```
tr_df.head()
```

Out[32]:

DATE	CONSUMPTION
2022-06-01	119.142609
2022-06-08	161.193427
2022-06-15	296.653166
2022-06-22	421.126834
2022-06-29	474.810000

In [33]:
```
#Auto Arima predictions
predictions = model.predict(n_periods = len(te_df))
pred_df = pd.DataFrame(predictions, columns=['PRED'])
pred_df.index = te_df.index
plot_and_validate(tr_df, te_df, pred_df, 'CONSUMPTION', 'PRED')
#the forecast is in an average value , not doing a very good job
```

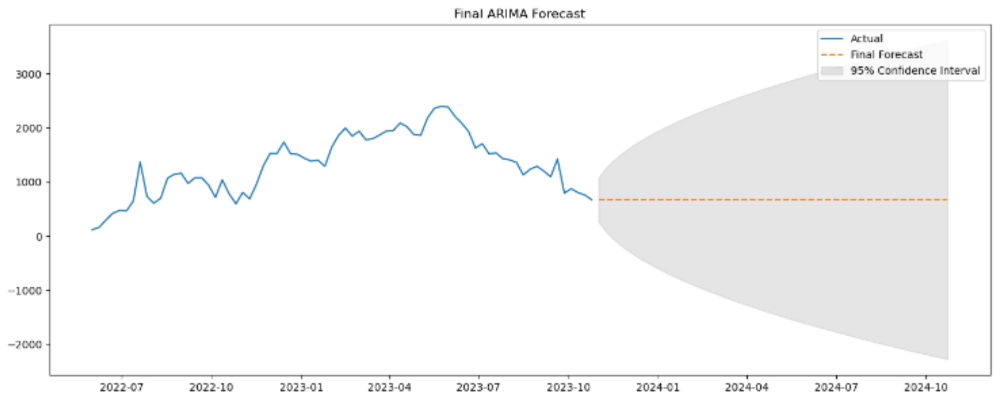
```
mse=1916819.849, rmse=1384.493, mae=1352.946, mape=1.329, smape=0.762
```



In [55]:

```python
import pandas as pd
import numpy as np
from pmdarima import auto_arima
from sklearn.metrics import mean_squared_error, mean_absolute_error
import matplotlib.pyplot as plt

# Assuming the functions read_csv_and_operations, convert_to_prophet, and get_train_test_generic are already defined

# Call the function to get train and test data with weekly frequency
tr_df, te_df, df = get_train_test_generic(freq='7D')

# Assuming you have already run cross-validation and stored the results in forecast_df
# This is just an example; you should adjust the actual results accordingly
# forecast_df = ...

# Concatenate the training and test data for building the final model
train_and_test_df = pd.concat([tr_df, te_df])

# Fit ARIMA model using auto_arima on the entire dataset
final_model = auto_arima(train_and_test_df['CONSUMPTION'], d=1, start_p=1, start_q=1,
                         max_p=2, max_q=2, m=52, seasonal=True, D=0,
                         trace=True, njobs=-1, maxiter=10)

# Make predictions on future dates (assuming 'n_periods' periods into the future)
n_periods = 52
final_forecast, conf_int = final_model.predict(n_periods=n_periods, return_conf_int=True)

# Create a DataFrame with the final forecast results
forecast_index = pd.date_range(df.index[-1], periods=n_periods + 1, freq='7D')[1:]
final_forecast_df = pd.DataFrame(index=forecast_index)
final_forecast_df['Forecast'] = final_forecast
final_forecast_df['Lower Bound'] = conf_int[:, 0]
final_forecast_df['Upper Bound'] = conf_int[:, 1]

# Extract actual values and predicted values only for the test period
actual_values = te_df['CONSUMPTION'].values
predicted_values = final_forecast_df['Forecast'].values[:len(actual_values)]

# Calculate metrics
mse = mean_squared_error(actual_values, predicted_values)
mae = mean_absolute_error(actual_values, predicted_values)
mape = np.mean(np.abs((actual_values - predicted_values) / actual_values)) * 100
smape = np.mean(2 * np.abs(predicted_values - actual_values) / (np.abs(actual_values) + np.abs(predicted_values))) * 100
rmse = np.sqrt(mse)

# Display the metrics for the final forecast
print(f'Final Mean Squared Error (MSE): {mse:.2f}')
print(f'Final Mean Absolute Error (MAE): {mae:.2f}')
print(f'Final Mean Absolute Percentage Error (MAPE): {mape:.2f}%')
print(f'Final Symmetric Mean Absolute Percentage Error (SMAPE): {smape:.2f}%')
print(f'Final Root Mean Squared Error (RMSE): {rmse:.2f}')

# Plot the original data, the final forecast, and the 95% confidence interval
plt.plot(train_and_test_df.index, train_and_test_df['CONSUMPTION'], label='Actual')
plt.plot(final_forecast_df.index, final_forecast_df['Forecast'], label='Final Forecast', linestyle='--')
plt.fill_between(final_forecast_df.index, final_forecast_df['Lower Bound'], final_forecast_df['Upper Bound'], color='gray', alpha
plt.title('Final ARIMA Forecast')
plt.legend()
plt.show()
```

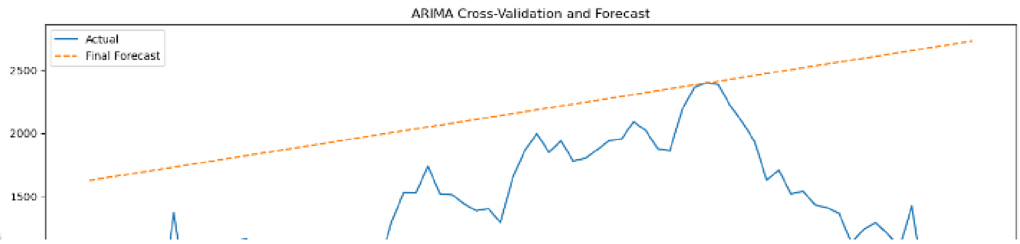
```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(1,0,1)[52] intercept   : AIC=997.440, Time=0.77 sec
 ARIMA(0,1,0)(0,0,0)[52] intercept   : AIC=990.681, Time=0.01 sec
 ARIMA(1,1,0)(1,0,0)[52] intercept   : AIC=993.456, Time=0.32 sec
 ARIMA(0,1,1)(0,0,1)[52] intercept   : AIC=993.419, Time=0.48 sec
 ARIMA(0,1,0)(0,0,0)[52]             : AIC=988.776, Time=0.01 sec
 ARIMA(0,1,0)(1,0,0)[52] intercept   : AIC=992.681, Time=0.19 sec
 ARIMA(0,1,0)(0,0,1)[52] intercept   : AIC=992.681, Time=0.19 sec
 ARIMA(0,1,0)(1,0,1)[52] intercept   : AIC=994.681, Time=0.24 sec
 ARIMA(1,1,0)(0,0,0)[52] intercept   : AIC=991.453, Time=0.04 sec
 ARIMA(0,1,1)(0,0,0)[52] intercept   : AIC=991.424, Time=0.04 sec
 ARIMA(1,1,1)(0,0,0)[52] intercept   : AIC=993.453, Time=0.04 sec

Best model:  ARIMA(0,1,0)(0,0,0)[52]
Total fit time: 2.343 seconds
Final Mean Squared Error (MSE): 292018.99
Final Mean Absolute Error (MAE): 463.83
Final Mean Absolute Percentage Error (MAPE): 36.67%
Final Symmetric Mean Absolute Percentage Error (SMAPE): 47.60%
Final Root Mean Squared Error (RMSE): 540.39
```



In [53]:

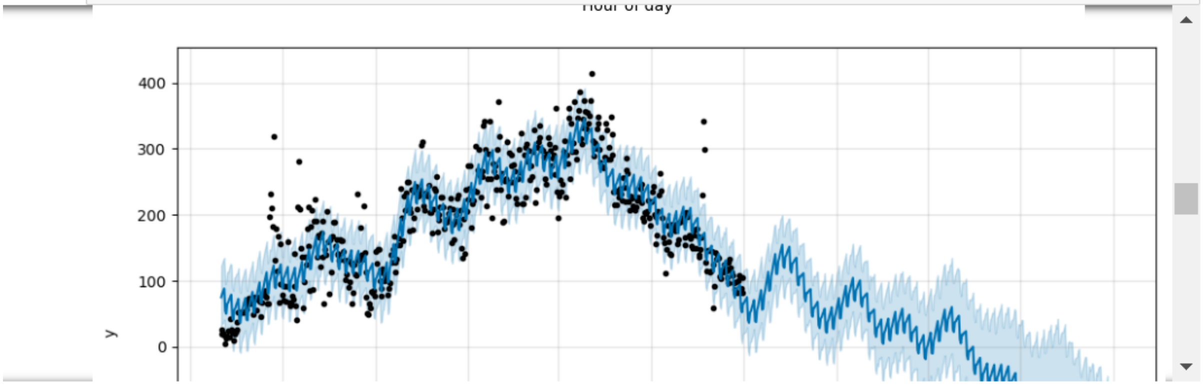
```python
import pandas as pd
from pmdarima import auto_arima
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error, mean_absolute_error
import numpy as np
import matplotlib.pyplot as plt

# Assuming the functions read_csv_and_operations, convert_to_prophet, and get_train_test_generic are already defined

# Call the function to get train and test data with weekly frequency
tr_df, _, df = get_train_test_generic(freq='7D')  # Use _ for the unused test_df

# Perform ARIMA cross-validation
n_splits = 5  # Number of folds for cross-validation
tscv = TimeSeriesSplit(n_splits=n_splits)

# Initialize lists to store Mean Squared Errors, Mean Absolute Errors, and predictions for each fold
mse_list = []
mae_list = []
predictions_list = []

for train_index, test_index in tscv.split(df):
    train, test = df.iloc[train_index], df.iloc[test_index]

    # Manually set d and D parameters to address the seasonal differencing issue
    model = auto_arima(train['CONSUMPTION'], d=1, start_p=1, start_q=1,
                       max_p=2, max_q=2, m=52, seasonal=True, D=0,
                       trace=True, njobs=-1, maxiter=10)

    # Make predictions on the entire dataset (including unseen dates)
    predictions = model.predict(n_periods=len(df), return_conf_int=False)

    # Evaluate the model performance on the test set
    mse = mean_squared_error(test['CONSUMPTION'], predictions[test_index])
    mae = mean_absolute_error(test['CONSUMPTION'], predictions[test_index])

    mse_list.append(mse)
    mae_list.append(mae)
    predictions_list.append(predictions)

# Calculate and print the average Mean Squared Error across all folds
avg_mse = sum(mse_list) / len(mse_list)
print(f'Average Mean Squared Error across {n_splits} folds: {avg_mse}')

# Calculate and print the average Mean Absolute Error across all folds
avg_mae = sum(mae_list) / len(mae_list)
print(f'Average Mean Absolute Error across {n_splits} folds: {avg_mae}')

# Combine predictions from all folds for final forecast
all_predictions = np.mean(predictions_list, axis=0)

# Evaluate the model performance on the entire dataset
final_mse = mean_squared_error(df['CONSUMPTION'], all_predictions)
final_mae = mean_absolute_error(df['CONSUMPTION'], all_predictions)

# Calculate and print additional metrics for the final forecast
mape = np.mean(np.abs((df['CONSUMPTION'] - all_predictions) / df['CONSUMPTION'])) * 100
rmse = np.sqrt(final_mse)

# Print final forecast metrics
print(f'Final Mean Squared Error: {final_mse}')
print(f'Final Mean Absolute Error: {final_mae}')
print(f'Final Mean Absolute Percentage Error (MAPE): {mape:.2f}%')
print(f'Root Mean Squared Error (RMSE): {rmse:.2f}')

# Plot the actual vs predicted values for the entire dataset
plt.plot(df.index, df['CONSUMPTION'], label='Actual')
plt.plot(df.index, all_predictions, label='Final Forecast', linestyle='--')

plt.title('ARIMA Cross-Validation and Forecast')
plt.legend()
plt.show()
```

```
Average Mean Squared Error across 5 folds: 780114.3391772133
Average Mean Absolute Error across 5 folds: 590.1533477358184
Final Mean Squared Error: 983625.6911960907
Final Mean Absolute Error: 858.6482140001256
Final Mean Absolute Percentage Error (MAPE): 119.87%
Root Mean Squared Error (RMSE): 991.78
```



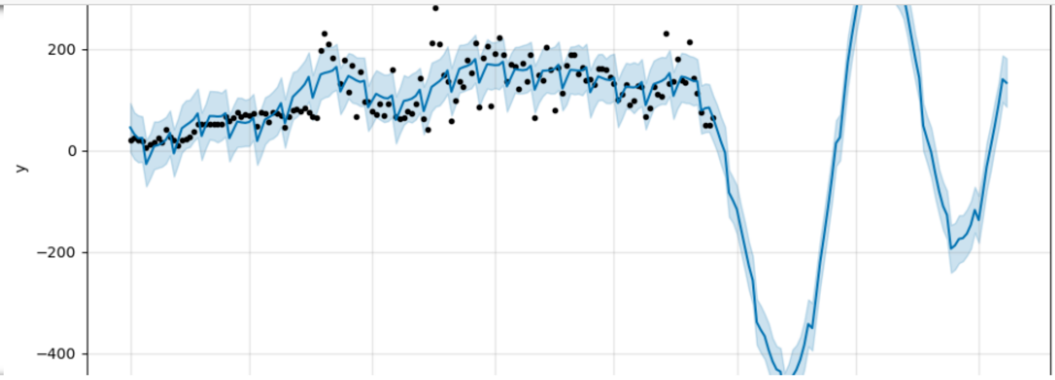
In [40]:

```
from prophet import Prophet

# Assuming you have a DataFrame df
# Ensure your DataFrame 'df' has a datetime column and a target variable column

# Display the head of your DataFrame to confirm its structure
display(df.head())

# Function to convert DataFrame to Prophet format
def convert_to_prophet(df):
    # Reset index and rename columns
    df = df.reset_index()
    df.columns = ['ds', 'y']
    # Set 'ds' column as datetime
    df['ds'] = pd.to_datetime(df['ds'])
    return df

# Convert the DataFrame to Prophet format
prophet_df = convert_to_prophet(df)

# Display the head of the converted DataFrame to check its structure
display(prophet_df.head())

# Initialize and fit Prophet model
m = Prophet(daily_seasonality=True, weekly_seasonality=True, yearly_seasonality=True)
m.fit(prophet_df)

# Create a future DataFrame for forecasting
future = m.make_future_dataframe(periods=365, freq='D')  # Example: Forecasting for 1 year (365 days)

# Make predictions
forecast = m.predict(future)

# Plot the forecast and its components
fig = m.plot(forecast)
m.plot_components(forecast)
```



In [41]:

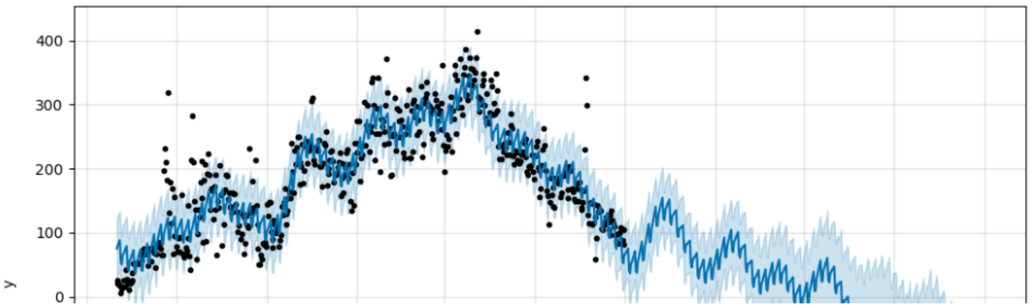
```python
import pandas as pd
from prophet import Prophet
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Assuming the functions read_csv_and_operations, convert_to_prophet, and get_train_test_generic are already defined

# Function to convert DataFrame to Prophet format
def convert_to_prophet(df):
    df = df.reset_index()
    df.columns = ['ds', 'y']
    df['ds'] = pd.to_datetime(df['ds'])
    return df

# Function to perform Prophet cross-validation and save results
def prophet_cross_validation(df, n_splits=6):
    tscv = TimeSeriesSplit(n_splits=n_splits)
    mse_list = []

    forecasts = pd.DataFrame()  # DataFrame to store forecast results

    for i, (train_index, test_index) in enumerate(tscv.split(df)):
        train, test = df.iloc[train_index], df.iloc[test_index]

        # Initialize and fit Prophet model
        m = Prophet(daily_seasonality=True, weekly_seasonality=True, yearly_seasonality=True)
        m.fit(train)

        # Create a future DataFrame for forecasting
        future = m.make_future_dataframe(periods=len(test), freq='D')

        # Make predictions
        forecast = m.predict(future)

        # Evaluate the model on the test set
        mse = mean_squared_error(test['y'], forecast['yhat'][:len(test)])
        mse_list.append(mse)

        # Plot the forecast and its components for each fold
        fig = m.plot(forecast)
        m.plot_components(forecast)
        plt.show()

        # Store forecast results in the DataFrame
        forecasts[f'Fold_{i + 1}'] = forecast['yhat']

    # Calculate and print the average Mean Squared Error across all folds
    avg_mse = sum(mse_list) / len(mse_list)
    print(f'Average Mean Squared Error across {n_splits} folds: {avg_mse}')

    # Save forecast results to a CSV file
    forecasts.to_csv('prophet_cross_validation_results.csv', index=False)

# Call the function to get train and test data with daily frequency
tr_df, te_df, df = get_train_test_generic(freq='1D')

# Convert the DataFrame to Prophet format
prophet_df = convert_to_prophet(df)

# Perform Prophet cross-validation and save results
prophet_cross_validation(prophet_df)
```



In [50]:

```python
import pandas as pd
from prophet import Prophet
import matplotlib.pyplot as plt

# Assuming the functions read_csv_and_operations, convert_to_prophet, and get_train_test_generic are already defined

# Function to convert DataFrame to Prophet format
def convert_to_prophet(df):
    df = df.reset_index()
    df.columns = ['ds', 'y']
    df['ds'] = pd.to_datetime(df['ds'])
    return df

# Function to train Prophet model and show forecasts until 2024
def prophet_forecast_until_2024(df):
    # Initialize and fit Prophet model
    m = Prophet(daily_seasonality=True, weekly_seasonality=True, yearly_seasonality=True)
    m.fit(df)

    # Create a future DataFrame for forecasting until 2024
    future = m.make_future_dataframe(periods=365 * 1, freq='D')  # Forecast for 3 years

    # Make predictions
    forecast = m.predict(future)

    # Plot the forecast and its components
    fig = m.plot(forecast)
    m.plot_components(forecast)
    plt.show()

# Call the function to get train and test data with daily frequency
tr_df, te_df, df = get_train_test_generic(freq='1D')

# Convert the DataFrame to Prophet format
prophet_df = convert_to_prophet(df)

# Train Prophet model and show forecasts until 2024
prophet_forecast_until_2024(prophet_df)
```

```
18:02:55 - cmdstanpy - INFO - Chain [1] start processing
18:02:55 - cmdstanpy - INFO - Chain [1] done processing
```



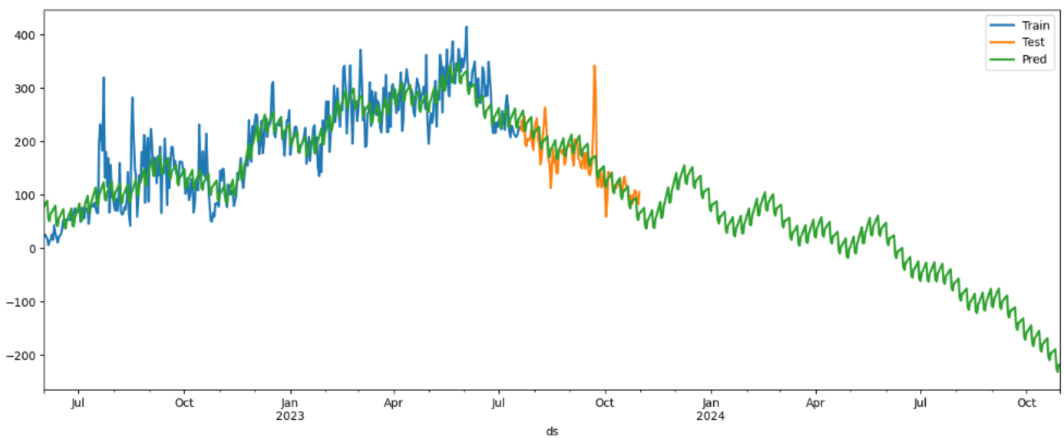
In [43]:

```python
forecast.head()
forecast.index = forecast.ds
```

In [44]:

```python
forecast.index = forecast.ds
plot_and_validate(tr_df, te_df, forecast,'CONSUMPTION','yhat')
```

```
mse=1006.508, rmse=31.726, mae=21.521, mape=0.137, smape=0.133
```

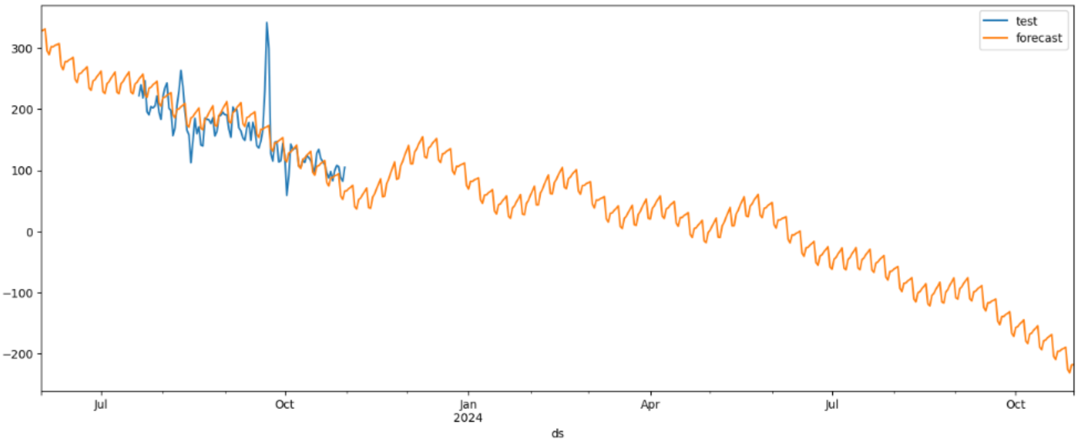


In [46]:
```
print(te_df.columns)
print(forecast.columns)
```

```
Index(['CONSUMPTION'], dtype='object')
Index(['ds', 'trend', 'yhat_lower', 'yhat_upper', 'trend_lower', 'trend_upper',
       'additive_terms', 'additive_terms_lower', 'additive_terms_upper',
       'daily', 'daily_lower', 'daily_upper', 'weekly', 'weekly_lower',
       'weekly_upper', 'yearly', 'yearly_lower', 'yearly_upper',
       'multiplicative_terms', 'multiplicative_terms_lower',
       'multiplicative_terms_upper', 'yhat'],
      dtype='object')
```

In [47]:
```
te_df.loc['2023-06-01':]['CONSUMPTION'].plot(legend=True, label='test')
forecast.loc['2023-06-01':]['yhat'].plot(legend=True, label='forecast')
```

Out[47]: <AxesSubplot:xlabel='ds'>



In [103]:
```
# Assuming your DataFrame is named te_df
print(te_df.columns)
```

```
Index(['CONSUMPTION'], dtype='object')
```

In [49]:
```
#Trend change points
from prophet.plot import plot_yearly, add_changepoints_to_plot
fig = m.plot(forecast, figsize = (14,6))
a = add_changepoints_to_plot(fig.gca(), m,forecast)
```