NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Implementing a Business Intelligence Framework on Bike Sharing Systems

The GIRA case study

Tomás Conceição de Campos Cunha Louro

Project Work

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

Implementing a Business Intelligence Framework on Bike Sharing Systems

The GIRA case study

by

Tomás Conceição de Campos Cunha Louro

Project Work presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics.

Supervised by

Bruno Jardim, PhD, NOVA Information Management School

Duarte Rodrigues, MSc, NOVA Information Management School

July, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 5th July 2025

Tomás Conceição de Campos Cunha Louro

ACKNOWLEDGEMENTS

I would like to thank to all those who were part of my Master's journey and contributed to achieve another milestone in my academic journey.

A special acknowledgment to Professors Bruno Jardim and Duarte Rodrigues who guided me during the development of this project. Your experience, knowledge, continuous support and availability were crucial to achieve this outcome.

To my family, especially to my parents, a warm and heartfelt thank you for always ensuring I had everything I needed to pursue my studies. Your support have been fundamental throughout this journey. To my sister, thank you for supporting me from day one. To Sara, a warm and special thank you for being by my side, encouraging me to grow and strive to be the best version of myself. Your presence has been a source of strength and motivation.

To my friends, particularly Marta, Diogo and Guilherme, thank you for sharing this journey with me and for the relentless support during it. Thank you for making this experience not only successful but deeply enjoyable. To my colleagues at Marsh McLennan, thank you for the opportunity to take on this challenge and for your patience throughout this adventure. To Eduardo, thank you for taking the time to guide me, make me think outside the box and for your not-always clear but always right advices.

To all of you my sincerest thanks.

ABSTRACT

The fast paced evolution of modern cities and the urgent need for sustainable mobility solutions have positioned Bike-Sharing Systems as an essential tool to promote greener, healthier and more efficient urban transportation. This study focuses on GIRA, Lisbon's bike-sharing system, and aims to design and implement a Business Intelligence framework to support data-driven decision-making. Based in the Kimball Data Warehouse/Business Intelligence Lifecycle, this research provides a structured and robust solution that enables GIRA stakeholders to better understand and optimize the system's performance.

As scope of the study three key domains were identified: Usage Patterns, Stations, and Weather. Connected to these domains a set of research questions was formulated to explore user behavior across different calendar days, station-level performance and availability, and the influence of meteorological conditions on trip demand. Leveraging from the open data available a dimensional model was developed, comprising three fact tables – Trip, Occupancy and Weather, and three dimensional tables – Station, Date and Time. The final BI solution includes an interactive PowerBI report that visualizes past usage key performance indicators and trends, supporting operational and strategic decision-making.

Findings from the dashboard were compared with insights from a systematic literature review, revealing convergence. Results confirmed that proximity to transportation hubs and academic institutions is associated with higher demand for rides, aligning with international studies on the impact of built environment on BSS usage. Temporal patterns, including the usage fluctuation across different weekdays, weekends and holidays, were consistent with prior findings from other BSS implementations. Weather conditions were also found to affect the demand for rides, with increased ridership observed in cold and mild temperatures along with dry conditions.

This study is aligned with the Sustainable Development Goals 11 and 13 and contributes to the ongoing discourse on sustainable urban mobility. By transforming raw operational data into insights, the developed BI solution enhances the strategic planning and operational efficiency of GIRA.

KEYWORDS

Business Intelligence; Analytics; Bike Sharing Systems; Soft Mobility; Urban Planning

Sustainable Development Goals

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AHP	Analytic Hierarchy Process
API	Application Programming Interface
BS-EREM	Bike Share Emission Reduction Estimation Model
BSS	Bike Sharing System
CML	Câmara Municipal de Lisboa
CO_2	Carbon dioxide
CSV	Comma-Separated Values
DW	Data Warehouse
E	Extract
ETL	Extract, Transform and Load
GBM	Gradient Boosting Machines
GBT	Gradient Boosted Tree
GHG	Greenhouse Gas
GIS	Geographic Information System
GLM	Generalized Linear Models
KPI	Key Performance Indicator
L	Load
LR	Linear Regression
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MLP	Multi-Layer Perceptron
NN	Neural Networks
NO_X	Nitrogen oxides
OLS	Ordinary Least Squares

PICOC	Population, Intervention, Comparison, Outcome and Context
RF	Random Forest
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Logarithmic Error
SDG	Sustainable Development Goals
SLR	Systematic Literature Review
SM	Semantic Model
SQL	Structured Query Language
T	Transform
THI	Temperature-Humidity Index
TOPSIS	Technique for Order Preference by Similarity to the Ideal Solution
WoS	Web of Science

1. INTRODUCTION

The world is rapidly evolving, presenting numerous challenges and risks to the population across the globe. United Nations (UN) adopted in 2015 the 2030 Agenda for Sustainable Development, structured around seventeen Sustainable Development Goals (SDG) and serving as a global framework to promote peace and prosperity for people and the planet (United Nations, 2025a) . Among the key areas of action identified is sustainable transport, with world leaders recognizing that "Sustainable transportation can enhance economic growth and improve accessibility. (…) achieves better integration of the economy while respecting the environment, improving social equity, health, resilience of cities, urban-rural linkages and productivity of rural areas" (United Nations, 2025-a). Within this context, Bike Sharing Systems (BSS) play an important role in promoting more sustainable cities, with benefits on traffic, emissions and public health areas (Maleki et al., 2023).

BSS are defined as short-term urban bicycle rental programs that enable users to pick up a bicycle from one docking station and return it to another. These systems are particularly suited for point-to-point trips, offering sustainable and efficient mode of transportation in urban areas (Midgley, 2011). The following study addresses as a case study the BSS operating in the city of Lisbon – GIRA. The system operating since 2017, counts with 195 docking stations across the 24 parishes of the city, making almost 4.000 bike racks available to the users, as of May 2025 (Câmara Municipal de Lisboa, 2025).

The increasing popularity of micro and soft mobility solutions, such as bikes and scooters, has significantly transformed urban commuting in recent years. Lisbon is not an outlier of this transformation, where one of the sustainable means of transport available and used is GIRA, which has become a vital component of daily mobility for many people. This paradigm shift presents not only a response to growing concerns about the environment but also an alternative to the more and more challenging traffic congestion.

The present study is motivated by the need to support data-driven decision-making processes within GIRA's management, enabling more efficient and sustainable operations. Furthermore, the work aligns with the UN's SDGs, particularly SDG 11, Sustainable Cities and Communities and SDG 13, Climate Action, by promoting sustainable urban mobility and contributing to climate resilience through informed planning and resource allocation.

The present study has its primary objective the development of a Business Intelligence (BI) framework to support GIRA's managers and administrators in making data-driven decisions based on the analysis of past usage indicators. By exploring and analyzing the available data related to the GIRA BSS, the project seeks to address a series of research questions organized into three research domains: Usage Patterns, Stations and Weather. Within the Usage Patterns domain, the study investigates questions such as "How many unique users are recorded on a daily basis and how does this number vary based on different influencing factors?" and "How does bike usage fluctuate between business days, weekends and

holidays?". The Stations domain focuses on the operational aspects addressing questions like "Are there particular stations that consistently experience shortages of bikes?". Lastly, the Weather domain explores the relationship between weather conditions and BSS usage, addressing questions like "How do temperature, wind speed and precipitation levels affect bike-sharing demand?".

The research is structured in three distinct chapters. The first chapter is dedicated to a Systematic Literature Review (SLR), which examines existing research on BSSs in the fields of network design, demand factors, prediction of usage, environmental impact and the specific case of Lisbon's BSS. The second chapter focuses on the Kimball Data Warehouse/Business Intelligence (DW/BI) Lifecycle, the methodological framework guiding this research, originally introduced by Kimball and Ross in 1998 (Kimball & Ross, 2013). This chapter provides a thorough overview of the methodology's milestones. Finally, the third chapter presents the case study, detailing the development of the BI solution, the methodological steps undertaken, while the fourth chapter presents the resulting insights and outputs derived from this study.

2. LITERATURE REVIEW

2.1 THEORETICAL FRAMEWORK

This section presents a SLR on the BSS to provide an overview of existing research and to identify key findings and gaps in these fields. Delving into literature concerning a topic is an essential component of any research, as it helps to understand the state-of-art within the field of study approached. Among the several methods at researcher's disposal to conduct a literature review, the SLR stands out as a structured and rigorous approach. The SLR is defined as "a research methodology to collect, identify, and critically analyze the available research studies" (Carrera-Rivera et al., 2022, p.2).

In order to carry out a SLR the research topic should be defined and a preliminary review of related literature conducted. With this step completed, the SLR can be pursued with the purpose of finding more specific studies related to the subject (Carrera-Rivera et al., 2022).

According to Carrera-Rivera et al. (2022), the SLR process can be divided in two main stages: Planning and Conducting. In the Planning stage, the researcher should define a protocol that outlines the methodological framework for the review. This includes formulating research questions and employing criteria such as Population, Intervention, Comparison, Outcome and Context (PICOC) or a similar method. Moreover, keywords and their synonyms are identified to facilitate searches in digital libraries, such as Scopus or Web of Science (WoS). The researcher then establishes inclusion and exclusion criteria to refine the selection of studies, ensuring that only relevant and high-quality literature is included in the review. In the Conducting stage, the researcher follows the protocol previously established. This materializes in the building of queries to execute searches in digital libraries, gather studies and compile them in a database, "which is helpful for data extraction and quantitative and qualitative analysis" (Carrera-Rivera et al., 2022, p.6). Furthermore, during the study selection and refinement phase, duplicate studies are identified, and inclusion or exclusion criteria are applied. The final steps involve extracting data and performing critical analyses of the selected studies, which culminate in the reporting phase, where insights are documented and contextualized.

By systematically following these steps, the SLR ensures the provision of valuable insights into the current state of research on BSS. Therefore, this segment serves as a foundation for subsequent analyses and discussions, placing the research within a broader academic context.

2.2 RELATED WORK AND STATE OF ART ANALYSIS

2.2.1 KEYWORDS AND SEARCH QUERIES

To identify the most relevant research conducted in recent years, several queries were designed based on the following keywords defined for this research and respective synonyms:

- Bike Sharing Systems
- Business Intelligence
- Demand
- Environmental
- Prediction
- Data-Driven
- Decision-Making
- Empresa Municipal de Estacionamento e Mobilidade de Lisboa (EMEL)
- GIRA
- Shared Mobility
- Soft Mobility
- Sustainable Mobility
- Urban Mobility
- Urban Planning

Following this step, the queries were performed on Scopus and WoS digital repositories of scientific documents. Finally, to stream the results, the following filters were applied:

- Year range: 2018 to 2024
- Document type: Article or Conference paper
- Publication stage: Final
- Language: English

The queries identified approximately 200 documents, which were subsequently refined through a structured selection process based on predefined criteria. Initially, the titles and abstracts of the documents were reviewed to exclude unrelated studies that fitted in the performed queries. Following this, the number of citations for each article was analyzed, with preference given to highly cited works addressing similar topics.

2.2.2 BSS NETWORK DESIGN

The geographical placement of BSS docks needs to be carefully examined as it impacts the user's engagement and service efficiency (Bahadori et al., 2022). Recent studies have used spatial analysis tools like Geographic Information System (GIS) to determine optimal station locations, considering factors such as transit access, nearby attractions and infrastructure (Bahadori et al., 2021; Banerjee et al., 2020). Additionally, hybrid methods combining fuzzy logic and spatial analysis have been employed to improve location decisions, enhancing station exposure and service coverage (Eren & Katanalp, 2022).

Banerjee et al. (2020) proposed a GIS-based spatial methodology to identify optimal locations for BSS docking stations, leveraging a location-allocation spatial analysis tool. Their approach

tested the hypothesis that new station locations are influenced by proximity to transit, attractions, or food facilities. The study employed a modified Huff's gravity model (Huff, 1962), replacing traditional location size with a suitability score based on proximity to factors positively correlated with BSS usage and existing bike stations. Maximizing dock exposure to the population was a key objective of the model, as well as ensuring proximity to favorable factors while maintaining a minimum distance of 300 meters from existing stations.

A systematic review of station location techniques was conducted by Bahadori et al. (2021), emphasizing the critical role of station placement in the implementation and expansion of BSS. The study highlighted four key criteria necessary for the successful operation of a BSS: the bike network, operator, users and city's infrastructure. Additionally, the authors categorized modeling techniques into three groups: mathematical algorithms, Multi-Criteria Decision-Making (MCDM) and GIS. They advocated for the integration of GIS and MCDM, which enhances analytical precision and provides decision-makers with more accurate insights into optimal station locations.

Eren & Katanalp (2022) proposed a hybrid methodology for addressing the challenge of selecting optimal sites for BSS stations based on land-use types. The study incorporated fuzzy logic within a GIS framework to account for uncertainties associated with trip starting points. By combining this fuzzy-based GIS approach with the Analytic Hierarchy Process (AHP) for spatial analysis and employing VIKOR and Psychometric-VIKOR methods for evaluation, the authors developed a robust decision-making framework. The integration of these methods offers a valuable tool for decision-making to solve complex problems related to (new) BSS station placement.

2.2.3 DEMAND FACTORS IMPACTING BSS USAGE

The literature available in the field of study mentions several factors that impact on the demand of BSS trips. Researchers pointed out, amongst other factors, that weather (Eren & Uz, 2020), calendar days (Kim, 2018), built environment characteristics (Duran-Rodas et al., 2019) and socio-demographic variables (Wang et al., 2018), significantly shape the way users resort to BSS.

In 2018, Wang et al. delved into the ridership patterns of the Citi Bike system in New York City across five distinct age cohorts: "younger Millennials (born 1995 to 2000), mid Millennials (1989 to 1994), older Millennials (born 1979 to 1988), Generation Xers (born 1965 to 1978), and Baby Boomers (born 1946 to 1964)" (Wang et al., 2018, p. 1). The analysis revealed that older Millennials were the predominant users of the system, being this fact attributed to their higher likelihood of employment and subsequent commuting needs. This finding aligns with the study's broader conclusion weekday trip demand is influenced by employment density. Furthermore, variations in the influence of built environment factors across cohorts were observed. For instance, higher population density significantly encourages BSS usage among younger and mid Millennials but disclosed no notable impact on Generation X. Temporal

patterns further distinguished cohorts, with the users born between 1995 and 2000 generating a greater volume of trips during midday, in contrast to older groups that exhibited peak activity in the evenings. Moreover, weather conditions were found to have a minimal effect on trip demand among young Millennials, highlighting unique behavioral differences within this cohort.

Duran-Rodas et al. (2019) investigated the built environment factors influencing BSS ridership in multiple cities, focusing on two case studies: German cities and international cities. Using statistical and Machine Learning (ML) techniques, such as Ordinary Least Squares (OLS), Generalized Linear Models (GLM) with Lasso selection and Gradient Boosting Machines (GBM), the study assessed model performance through R^2 values, with all models showing similar results. Model performance was found to depend significantly on clustering rentals by their daily distribution, with temporal variations observed in statistically significant factors, "(…) parks, green areas, and bodies of water on weekends; banks in the mornings; (…) pubs, cinemas, and clubs at night (…)" (Duran-Rodas et al., 2019, p. 64). Moreover, the study identified city population, distance to city center, leisure-related establishments and transport-related infrastructure as key determinants of BSS ridership. Regarding the last mentioned category of features, the proximity to railway stations was significant in the German use case, while bus stations were more relevant internationally.

Eren & Uz (2020) developed a comprehensive review of the factors affecting BSS demand, particularly those concerning weather and built environment. The researchers concluded that adverse weather conditions significantly decrease the demand for BSS trip, highlighting that "if there is no precipitation at 20–30 °C temperatures, it is more likely to increase the number of bike-sharing trips." (Eren & Uz, 2020, p.9). Moreover, based on related studies, the authors classified factors such as Winter months, rainy and windy days, snow and heavy rain precipitation, high speed winds, temperatures ranging from negative Celsius degrees to positive 10 °C, and scorching heat as factors having a strongly negative correlation with trip demand. As for built environment factors impacting on the demand of trips, bike infrastructures, such as bike lanes or paths, land use for hotels, restaurants, commercial areas or educational facilities are positively correlated with the demand of rides, according to Eren & Uz (2020) previous reviewed works.

Furthermore, a previous study on the Tashu BSS (Kim, 2018), analyzed the effects deriving from weather conditions and temporal characteristics at station and system levels. In both levels of analysis clustering was performed, to group stations with similar properties. Additionally, the Temperature-Humidity Index (THI) and a heatwave indicator were introduced to assess the combined impact of temperature and humidity, as well as the effects of extreme heat. At the system level, the clusters analysis revealed that one cluster had a morning peak in usage, while the other two exhibited peaks in the evening. The effects of weather varied among clusters, appearing to correlate with the purpose of the trips. Regarding the impact of the variables introduced both showed a negative correlation with

daily bicycle demand, although the impact of temperature varied across different time periods.

Kim (2018) also investigated the influence of calendar days on trip demand. While no significant difference was observed between weekdays and weekends in the total number of rentals overall, when focusing on public holidays the number of rentals decreased. In a similar manner to the effect of high temperatures, the impact of calendar days was not uniform across all intraday time periods. To illustrate it, the number of weekend trips during rush hours was lower compared to weekdays; however, it was higher during daytime hours.

Following the analysis of the researches above concerning the demand factors influencing BSS usage, please refer to Table 1 to review the metrics used to quantify the analyzed variables.

Table 1 - Demand Factors Metrics

Category	Feature description	Unit	References
Seasons	Winter; Spring; Summer; Autumn	Months	Eren & Uz (2020)
Weather	Sunny; Cloudy; Rainy; Foggy	Days	Eren & Uz (2020)
Precipitation	Rain; Snow	cm	Eren & Uz (2020) & Kim (2018)
	Change of precipitation	%	Eren & Uz (2020)
Wind	-	km/h or m/s	Eren & Uz (2020) & Kim (2018)
Temperature	-	°C	Eren & Uz (2020) & Kim (2018)
Bike Infrastructures	Bike lane; Bike path; Off-road	km	Eren & Uz (2020) & Wang et al. (2018)
Land use	Parks and Leisure areas	%	Eren & Uz (2020)
	Food facilities	Count or distance	Eren & Uz (2020) & Duran-Rodas et al. (2019)
	Educational facilities	Count or distance	Eren & Uz (2020) & Wang et al. (2018)
Public Transports	Bus stops; Metro and Train stations	Count or distance	Eren & Uz (2020) & Wang et al. (2018)
Travel		Time or distance	Eren & Uz (2020)
Socio-demographic	Population	%	Eren & Uz (2020) & Wang et al. (2018)
	Gender	%	Eren & Uz (2020)
	Age	%	Eren & Uz (2020) & Wang et al. (2018)

Category	Feature description	Unit	References
Calendar	Weekdays; Weekends; Public holidays	Count	Kim (2018)

2.2.4 PREDICTION OF BSS USAGE

Accurate demand prediction is essential for the efficient management of BSS, particularly to address the common issue of system imbalance (Hulot et al., 2018), where bikes need to be redistributed to match user's demand. For this reason, researches using ML models and statistical inference tools have been developed, aiming to predict ridership demand at short intervals while providing real-time visualization to aid decision-making (Boufidis et al., 2020).

Hulot et al. (2018) investigated station-level demand prediction in BSS to address the common issue of system imbalance, where bikes must be redistributed throughout the day to match users' demand. Accuracy in the demand predictions is deemed by the authors to be a crucial factor to an effective redistribution of the bikes, although there is very few literature exploring it. The researchers aimed to predict trip demand and corresponding returns at dock level on an hourly basis, employing statistical inference and ML models, that take as inputs temporal and weather variables to predict the mean and variance of demand. The model resorts to K-means and Singular Value Decomposition for dimensionality reduction and then leverages predictive models, such Linear Regression (LR), Multi-Layer Perceptron (MLP), Gradient Boosted Tree (GBT) and Random Forest (RF). Concerning the evaluation of the predictions the R^2 and Root Mean Squared Error (RMSE) were the metrics chosen by Hulot et al. By focusing on the primary traffic patterns of each station, the study provided a structured framework for demand prediction in BSS.

The successful management for Boufidis et al. (2020) is closely linked to optimal distribution of bikes across the different docks, which needs an accurate demand forecasting. In this study, the authors developed a station-level demand prediction and visualization tool designed to support BSS operators. The tool predicts ridership demand at intervals of 1-, 2-, or 3-hours using ML models, including XGBoost, RF and Neural Networks (NN), with R^2 and Root Mean Squared Logarithmic Error (RMSLE) being employed as evaluation metrics to assess model performance. To enhance accessibility and usability, the tool was designed to automatically visualize predictions in real time, enabling operators to make data-driven decisions and manage their fleets and stations more effectively.

2.2.5 ENVIRONMENTAL IMPACT OF BSSS

BSSs offer a variety of benefits for societies and their users. Amongst those benefits are the environmental ones, contributing for a more sustainable urban mobility (Zhang & Mi, 2018).

To quantify the environmental benefits of BSS, Zhang & Mi (2018) developed a study using big data techniques to estimate the impacts on energy use and gas emissions in Shanghai during 2016. The researchers concluded that, from a spatial perspective, the benefits of BSS use were significantly greater in the districts with higher population density. From a temporal perspective, usage peaked in the morning and evening, with evening peaks being higher than those in the morning. The quantitative study concluded that BSS saved 8.359 tons of petrol and decreased Carbon dioxide (CO_2) and Nitrogen oxides (NO_X) emissions by 25.240 and 64 tons, respectively.

Kou et al. (2020) studied the impact of BSS usage on greenhouse gas (GHG) emissions by proposing a Bike Share Emission Reduction Estimation Model (BS-EREM), which was applied to eight cities in the United States. The model uses stochastic methods to estimate the transportation modes replaced by BSS trips, considering factors such as trip distance and duration. Their analysis revealed that total GHG emission reductions in 2016 ranged from 41 tons to 5.417 tons across the studied cities, while the reductions per trip varied between 283 and 581 grams of CO_2. The authors have also observed a linear relationship between GHG emission reductions and the number of trips, bicycles and docking stations. In addition, the study concluded that docks located in city centers contributed more to the total reductions of emissions, due to high trip volumes, while those in suburban areas showed higher reductions per trip due to longer travel distances. Despite these findings, the contribution of BSS to overall transportation sector emissions remains modest, accounting for less than 0,1% in the studied cities.

Concerning the above presented studies, the following table presents the metric used to measure the environmental impact of BSSs.

Table 2 - Environmental Metrics

Category	Feature description	Unit	References
Environmental	GHG emission	Tons	Zhang & Mi (2018)

2.2.6 LISBON'S BSS – GIRA

Regarding the BSS operating in Lisbon – GIRA, some research has been published in recent years, covering a variety of topics such as demand factors, hourly ridership prediction and the dock network design.

An integrated GIS-MCDM methodology to rank potential locations for expanding the GIRA BSS was proposed by Bahadori et al. (2022). The framework combines techniques such as the AHP, Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) and GIS to facilitate both planning and operational expansion of BSS. An AHP questionnaire conducted with GIRA's operational staff revealed as the most critical factors the city infrastructure, population density and slopes, whereas the bike network was deemed less important. Using ArcGIS

software, particularly its network analysis module, a suitability map was generated based on the geographic data for each prioritized criterion. The study identified 45 potential new dock locations, primarily concentrated in the city center, which was highlighted as the most suitable area for establishing additional stations.

Albuquerque et al. (2021) conducted an analysis of the spatiotemporal distribution patterns to understand GIRA's usage. By employing clustering techniques with travel distance, speed and duration as input features and correlating these with environmental factors, the study identified usage frequency and geographic patterns. Key findings indicate that most trips occur on weekday afternoons, suggesting a predominant use for commuting from work to home. Additionally, weather plays a significant role in trip demand, with the absence of rain and temperatures between 10 °C and 30 °C being two of the drivers associated to demand increase. The study also aligns Lisbon's BSS with usage patterns of medium-sized cities. The findings provide valuable insights for GIRA's stakeholders to optimize and improve operations.

Lastly, Lucas & Andrade (2021) developed statistical models to predict hourly ridership demand for GIRA, focusing on the origin-destination pair, and durations of trips. By employing GLM and zero-augmented models, the study explored the influence of factors such as weather and time periods on demand. The hurdle regression variant of the zero-augmented model proved effective to "capture spatial and temporal patterns of hourly origin-destination demand in BSS" (Lucas & Andrade, 2021, p. 11). Temporal analysis revealed peak usage during mornings, lunch hours and afternoons/evenings, with variations observed based on the day of the week.

2.3 KEY PERFORMANCE INDICATORS IDENTIFICATION

To correctly approach the variables used in their studies, authors must define Key Performance Indicators (KPI) that describe them and are able to quantify and measure them. As the related work serves as a foundation for this research, please refer to the following table to confer the KPIs related to BSS to be used in the hereby presented study.

Table 3 - Bike Sharing System's Key Performance Indicators Identification

Category	Feature description	Unit	References
BSS	Number of docks	Count	Eren & Uz (2020) & Wang et al. (2018)
	Number of racks per station	Count	Eren & Uz (2020) & Wang et al. (2018)
	Rentals	Count or %	Kim (2018) & Duran-Rodas et al. (2019)
	Travel	Time, distance or speed	Albuquerque et al. (2021)

2.4 RESEARCH GAPS

From the studies analyzed it is possible to outline several limitations, highlighting research gaps for the future. The major limitation outlined by the authors of the studies reviewed are the constraints to include socio-demographic information from the riders in the analyzes made, limiting the conclusions and outcomes (Lucas & Andrade, 2021; Wang et al., 2018). Moreover, Duran-Rodas et al. (2019) and Banerjee et al. (2020) referred that the non-inclusion of spatial and geographic features, namely longitude and latitude, impacted on the work conducted. Furthermore Banerjee et al. (2020) noted that the model developed was based solely on observed demand making the study lack in policy recommendations that need to be validated by evidence-based information. Therefore, the hereby developed research aims to develop a BI framework that leverages and manages data from GIRA usage, extracting insights and critical information for decision makers, while covering the gaps identified in previous works concerning BSS.

3. METHODOLOGY

3.1 RESEARCH DESIGN

In the context of the development of a BI framework aimed at managing the data generated by GIRA usage, the Kimball DW/BI Lifecycle introduced in 1998 (Kimball & Ross, 2013), is adopted as the guiding methodology. This framework provides a structured approach to DW/BI project implementation, as illustrated in Figure 3-1. The diagram describes the structure of tasks required for an effective DW/BI project, with each box marking a milestone in its development (Kimball et al., 2008). For a brief explanation of each task please resort to Appendix 1 – Kimball's DW/BI Lifecycle Milestones Description.

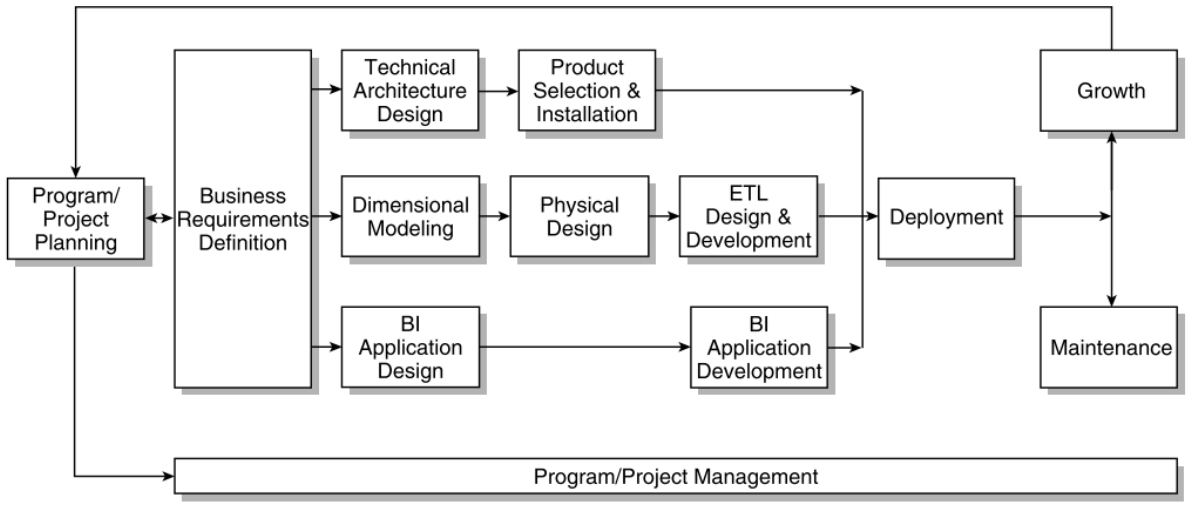

Figure 3-1 - The Kimball Lifecycle Diagram, in The Data Warehouse Lifecycle Toolkit, Kimball et al. (2008)

3.2PROGRAM/ PROJECT PLANNING AND MANAGEMENT

The herby developed study has its primary objective the development of a BI Framework, from which decision makers can extract insights and make data-based decisions. The subject of the project is the GIRA BSS that operates in Lisbon. Leveraging from EMEL open data portal (EMEL, 2020, 2021) and from the Câmara Municipal de Lisboa (CML) open data portal - (Câmara Municipal de Lisboa, 2018) ,the project will follow the Kimball DW/BI Lifecycle.

3.3 BUSINESS REQUIREMENTS DEFINITION

Lisbon is traditionally referred to as the city of the seven hills due to its distinctive topology. However, and contrasting with Lisbon's alias, 73% of its streets are either flat or have a slope below 5% (Câmara Municipal de Lisboa, 2025). This favorable landscape supports an extensive network of bike lanes, as illustrated in Figure 3-2, spanning across the city's 24 parishes. Furthermore, the municipality aims to expand the cycling paths up to 263 kilometers during 2025 (Câmara Municipal de Lisboa, 2025).

GIRA, Lisbon's BSS, has been operational since 2017 under the management of EMEL. With the recent expansion of its station network, GIRA now features 195 docking stations across the city, providing almost 4.000 bike racks (Câmara Municipal de Lisboa, 2025). The distribution of these stations throughout Lisbon is depicted in Figure 3-3 based on the CML geodata open portal.

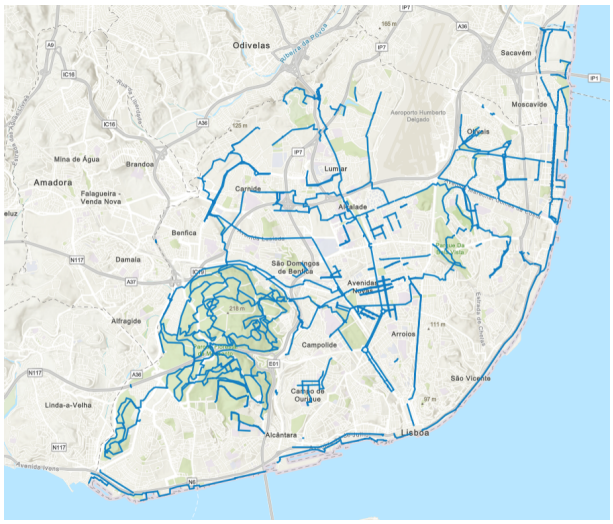

Figure 3-2 - Cycling Network, in (Câmara Municipal de Lisboa, 2019b)

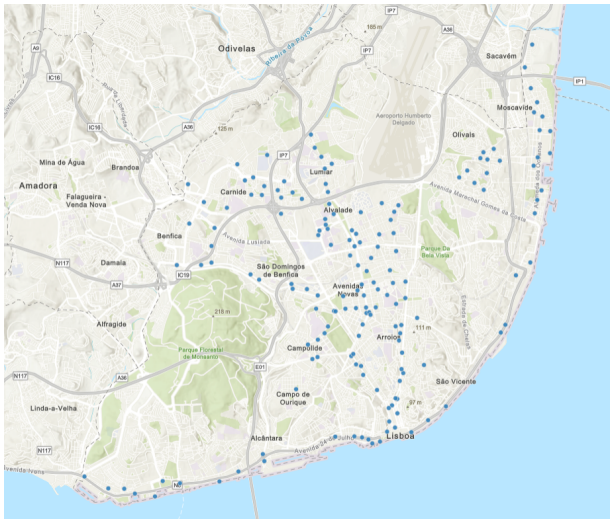

Figure 3-3 - GIRA Stations in Operation, in (Câmara Municipal de Lisboa, 2019a)

The Lisboa Aberta and EMEL's open data portals provide access to distinct data sets concerning Lisbon's BSS, encompassing information on stations, usage, trips, weather indicators from 2022 and 2023, and occupancy from 2022 to the first quarter of 2023.

The stations table provides detailed station-level information including station unique identifier, locations, geographical coordinates, zip codes, the number of bikes available, working schedules and other general attributes. The usage table aggregates daily statistics, including the total seconds of bike usage per day and the number of unique users, with each

row representing a single day. Similarly, the trips table focuses on daily trip metrics such as the average trip duration amongst other statistics. Moreover, the occupancy tables capture real-time station data such as the number of available bikes, total racks, coordinates and timestamped records, offering an overview of the operational status over time. Lastly, the weather table specifies the hourly evolution of several meteorologic variables, such as temperature, relative humidity, wind speed and precipitation. Resorting to Table 4 it is possible to check information regarding tables' shape and their timeframe, for detailed metadata concerning each table resort to the Appendixes 2 to 6.

Table 4 - Tables Characteristics

Table	Number of Columns	Number of Rows	Timeframe
Usage	4	1458	2019 to 2023
Trips	3	1792	2019 to 2024
Station	15	162	-
Weather	14	2067670	2022 and 2023
2022 Occupancy	6	3938292	full year
2023 Occupancy	6	933644	1st trimester

In addition to the occupancy tables from 2022 and 2023, data for 2020 and 2021 are also available. However, these datasets were excluded from the present study, as their corresponding timeframe coincides with the COVID-19 pandemic. Consequently, the associated metrics and data are likely to be biased potentially compromising the validity of the analysis.

To develop the Kimball's DW/BI Lifecycle thoroughly, the business requirements were reformulated into business questions as outlined in Table 5.

Table 5 - Business Questions

Business Scope	Business Question
Usage Patterns	1. What are the observed trends in bike usage for the years 2022 and 2023? 2. How many unique users are recorded on a daily basis and how does this number vary based on different influencing factors? 3. How does bike usage fluctuate between business days, weekends and holidays? 4. Are there specific days characterized by higher bike usage and are these fluctuations correlated with particular events or factors? 5. What is the average trip duration on a daily basis? 6. How do calendar days influence the peaks and troughs in trip frequencies?

Business Scope	Business Question
	7. How do the average trip durations differ between business days, weekends and holidays?
Stations	8. What is the average rate of bikes available at stations? 9. Are there particular stations that consistently experience shortages of bikes? 10. What is the average number of bikes available at each station during peak and off-peak hours? 11. How does bike availability differ between business days, weekends and holidays?
Weather	12. Is there any correlation between temperature and trips? 13. How do temperature, wind speed and precipitation levels affect bike-sharing demand?

3.4 TECHNOLOGY TRACK

Following the definition of business requirements in alignment with Kimball's DW/BI Lifecycle, the next milestones involve establishing the technical architecture design to follow, as well as select, install and test the software and hardware needed to meet the project's requirement.

3.4.1 TECHNICAL ARCHITECTURE DESIGN

Considering the previously identified business requirements of the project, the most appropriate architectural framework to adopt is Online Analytical Processing (OLAP). This approach aligns with the project's objectives, as OLAP is designed to manage large volumes of multidimensional data efficiently. Furthermore, OLAP serves as basis for strategic planning, decision support and insight discovery (Sharda et al., 2017).

The OLAP architecture is based on the concept of a cube, which represents the multidimensionality of data whose structure facilitates fast data analysis. The primary objective of this data structure is to overcome limitations of relational databases, by enabling users to navigate data and retrieve specific subsets. Moreover, this architectural design enables the user to perform various operations such as slices and drills. The slicing operation extracts a subset of data typically represented in two dimensions, while the drilling operation allows users to navigate across different levels of data granularity. Specifically, drill-up functionality provides a summarized view of the data, whereas drill-down enables a more detailed analysis, enhancing the depth of insights derived from the dataset (Sharda et al., 2017).

3.4.2 PRODUCT SELECTION & INSTALLATION

To conclude the Technology track of the Kimball DW/BI Lifecycle, the selection of development tools is a critical step. Considering the defined business requirements and the established technical architecture, Microsoft Fabric (Microsoft, 2025) and Microsoft Power BI (Microsoft, 2025) were identified as the most suitable solutions. Microsoft Fabric, a cloud

solution, provides a fully integrated and scalable ecosystem for data analytics, ensuring seamless data processing and management. Meanwhile, Microsoft Power BI serves as the front-end interface, enabling users to leverage the OLAP architecture effectively, transforming raw data into valuable insights. The following table details which applications inside the Microsoft's cloud solution will be used and for what purpose.

Table 6 - Tools Selected per Purpose

	Application	Function
Database	Fabric Lakehouse	A platform for the storage, management and analysis of data, offering the flexibility and scalability to handle large volumes of structured and unstructured data. (Microsoft, 2025)
ETL	Dataflow Gen2 and Data pipeline	Resorting to Dataflows and Data pipelines allows the users to ingest, prepare and transform data from different data sources (Microsoft, 2024e).
Staging Area	Data pipeline and Fabric Warehouse	Pipelines enable the automated execution of the dataflows created for the ETL phase, moving data from the Lakehouse to the Staging Area, where data integrity rules will be checked and ensured (Microsoft, 2024d).
Data Warehouse (DW)	Fabric Warehouse	A cloud-based scalable analytical DW that supports direct querying and pre-aggregated datasets on an enterprise grade, enabling seamless collaboration (Microsoft, 2024c).
Reporting	PowerBI	Front-end software that allows business users to autonomously get insights from data stored in several data sources, through dashboards and reports (Microsoft, 2024a).

3.5 DATA TRACK

3.5.1 DIMENSIONAL MODELING

The first step of the Kimball's DW/BI Lifecycle's data track is to develop the Dimensional Modeling of the project. Kimball et al. (2008) define Dimensional Modelling as a "logical design technique for structuring data so that it's intuitive to business users and delivers fast query performance." (Kimball et al., 2008, p.234). Moreover, according to the authors the logical design technique is composed of four distinct phases:

1. Choose the Business Process
2. Declare the Grain
3. Identify the Dimensions
4. Identify the Facts

3.5.1.1 CHOOSE THE BUSINESS PROCESS

Deepening each phase of the Kimball's methodology that encompasses four distinct steps, the demanded trips and the occupancy of GIRA stations around Lisbon's parishes were identified as the business processes to be modeled. The above mentioned processes were selected due to their relevance in capturing the dynamics of the BSS usage and their potential to support data-driven decision-making.

3.5.1.2 DECLARE THE GRAIN

The following step is to declare the grain, which consists in defining the meaning of each row of the fact table (Kimball et al., 2008). For this project each row of the Trip fact table characterizes the GIRA trips concerning a specific day, while each row of the Occupancy fact table describes a station at a specific date and hour. Lastly the Weather fact table contains the hourly historical register for several meteorological indicators.

Regarding the granularity derived from the dimensional tables, two principal hierarchies can be identified: one pertaining to the temporal dimension and the other to the geographical distribution of stations. The temporal granularity, built from the date dimensional table, has as its smallest grain the day, is structured at the daily level, thereby allowing for a detailed, day-by-day analysis of GIRA usage patterns – Figure 3-4. The geographical granularity, constructed through the Station dimension, enables the comparative analysis of GIRA both within and across parishes – Figure 3-5.

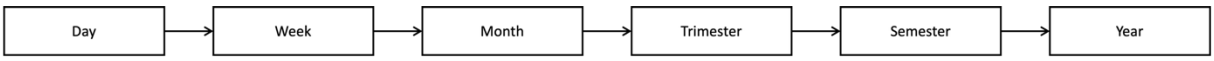

Figure 3-4 - Temporal Granularity

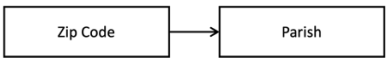

Figure 3-5 - Geographical Granularity

3.5.1.3 IDENTIFY THE DIMENSIONS

Following the grain declaration, the methodology states that the dimensional tables should be identified. The dimensional tables that compose this project are the Dim_Station, Dim_Date and Dim_Time tables.

The Dim_Station table contains information about the stations that compose the BSS, namely their ID, Name, Latitude, Longitude, Zip Code and Parish. The only column added to the original columns is the SK_Station that plays the Surrogate Key role. The datatypes of each column composing the Dim_Station table can be find below on Table 7.

Table 7 - Dim_Station structure

Variable	Datatype
SK_Station	INT
StationID	INT
Name	VARCHAR(250)
Latitude	DECIMAL(10,7)
Longitude	DECIMAL(10,7)
ZipCode	INT
Parish	VARCHAR(30)

The Dim_Date and Dim_Time tables are going to be created manually, maximizing the detail of analysis possible to perform and matching the needs of the proposed business questions. Similarly to the Dim_Station table, these dimensional tables have a column playing the Surrogate Key role, SK_Date and SK_Time, respectively. Information about their columns and datatypes can be found resorting to the following tables – Table 8 and Table 9.

Table 8 - Dim_Date structure

Variable	Datatype
SK_Date	INT
ProperDate	DATE
Day	INT
DayWeek	INT
DayName	VARCHAR(20)
Month	INT
MonthName	VARCHAR(20)
TrimesterNumber	INT
SemesterNumber	INT
Year	INT
WeekYear	INT
IsWeekend	VARCHAR(10)
IsHoliday	VARCHAR(10)
HolidayName	VARCHAR(50)

Table 9 - Dim_Time structure

Variable	Datatype
SK_Time	INT
Hour	INT
HourFormatted	TIME(0)

3.5.1.4 IDENTIFY THE FACTS

The final phase in the dimensional modeling process involves the identification and definition of fact tables. This project incorporates three distinct fact tables: Fact_Occupancy, Fact_Trip and Fact_Weather. In alignment with the approach used for dimension identification, comprehensive details regarding each fact table, including their respective columns and data types, are provided in Tables 10, 11 and 12.

Table 10 - Fact_Occupancy structure

Variable	Datatype
FK_Date	INT
FK_Station	INT
FK_Time	INT
NumberOfBikes	INT
NumberOfDocks	INT
StationStatus	VARCHAR(10)

Table 11 - Fact_Trip structure

Variable	Datatype
FK_Date	INT
AverageTripDuration	DECIMAL(18,6)
TotalSecondsPerDay	DECIMAL(18,6)
UserQuantityUnique	INT
AverageTripByUsers	DECIMAL(18,6)

Table 12 - Fact_Weather structure

Variable	Datatype
FK_Date	INT
FK_Time	INT
Day	DATE
Hour	TIME(0)
AvgTemperature	DECIMAL(18,6)
AvgRelativeHumidity	DECIMAL(18,6)
AvgWindSpeed	DECIMAL(18,6)
AvgWindDirection	DECIMAL(18,6)
AvgPrecipitation	DECIMAL(18,6)
AvgUVIndex	DECIMAL(18,6)
AvgAtmosphericPreassure	DECIMAL(18,6)

Following the Kimball's four steps approach, Figure 3-6 depicts the dimensional model developed for this project.

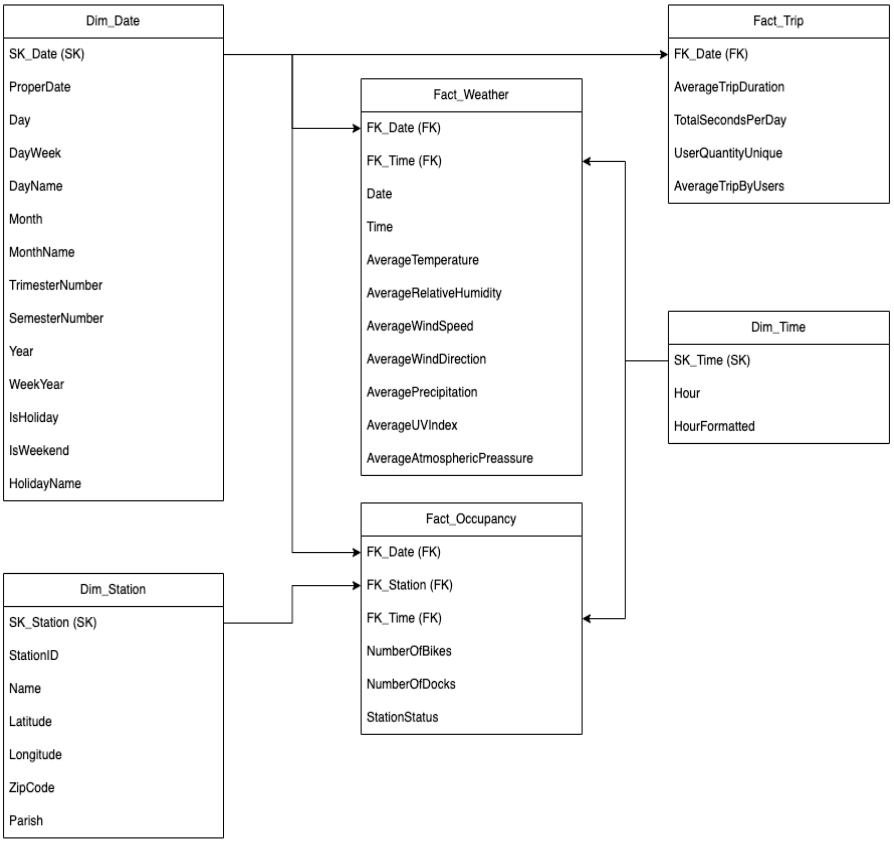

Figure 3-6 - Dimensional Modeling
Source: Based on (Kimball et al., 2008) and developed by the author

3.5.2 PHYSICAL DESIGN

The subsequent phase of the adopted methodology consists of the design of the physical model. For this milestone, the data warehousing capabilities of OneLake within the Microsoft Fabric cloud solution were employed. After the creation of a DW for the tables to be stored, a Structured Query Language (SQL) script was developed to mirror the created dimensional model within Fabric.

The SQL script begins by clearing pre-existing tables within the DW to prevent duplication. Following this, fact and dimension tables are created with each column name and data type explicitly defined. Furthermore, constraints are established to specify whether columns permit null values, ensuring data integrity and consistency within the warehouse.

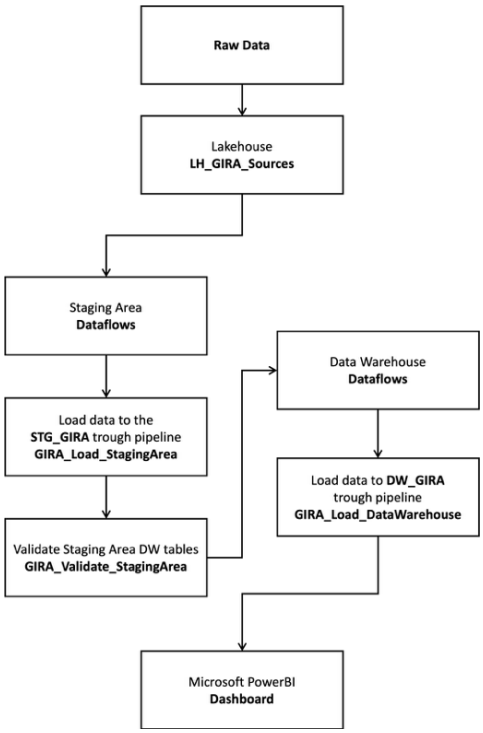

Figure 3-7 - Extract Transform and Load Architecture Diagram

3.5.3 ETL DESIGN & DEVELOPMENT

Following the creation of the physical model in OneLake's Data Warehouse the ETL process takes place. To facilitate the necessary data transformations and adjustments, the functionalities of Dataflow Gen2 and Data Pipeline within Microsoft Fabric were utilized. Dataflow Gen2 enables the ingestion, transformation and loading of data into OneLake, while Data Pipeline automates data movement across Fabric. Moreover, a Staging Area was created to ensure the compliance with data integrity rules.

In alignment with the project's scope, five distinct dataflows were developed, each corresponding to a fact or dimension within the physical model – Figure 3-8. Throughout all dataflows, various transformations were applied to ensure the consistency and correctness of the data.

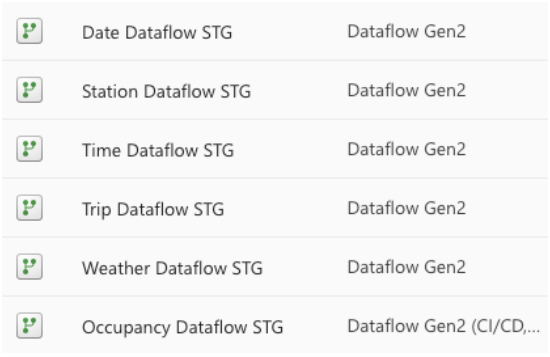

Figure 3-8 - Dataflows Gen2 created concerning the Extract Transform and Load and load to the Staging Area

The initial stage of dataflow creation focuses on the dimensions. The first dataflow developed as regards the date dimension which is derived from a list of dates ranging from January 1st, 2022, to December 31st, 2023. From that list, additional column such as year, month and day are extracted, while supplementary columns indicating the day of the week, day name, quarter, semester and week of the year are generated. Moreover, the table is merged with three other queries regarding 2022 and 2023 holiday's dates and names. Additional columns are introduced to identify whether each date corresponds to a weekend or a public holiday. To finalize the transformation of the date dimension table, an index column is incorporated and column names and data types are fixed.

Table 13 - Staging Area Date Dimension dataflow operations

Type of Operation	Description
List Dates	Generate dates between January 1st, 2022, to December 31st, 2023
Create columns	Create columns Year, Month, Month Name, Day, Day of Week, Day Name, Week of Year, Quarter and Semester based on date column
Merge	Merge existing columns with Holidays column
Create Columns	Create Is Holiday and Is Weekend columns
General column settings	Change data types, rename and remove columns

Regarding the station dataflow, the transformation steps involve appending three source files that store station-related information, followed by handling missing values in critical fields such as zip codes, non-relevant columns removal and column names and data type correction. Duplicate rows are also eliminated to ensure data consistency. A business key column is subsequently generated based on the station name which contains a unique numerical identifier for each station.

Table 14 - Staging Area Station Dimension dataflow operations

Type of Operation	Description
Handle null values	Replace null values on the ZipCode column
Append data	Append data source files
Handle missing values	Handle the missing values on the BK_Station column
Remove duplicates	Remove duplicate rows to have only one row per station
Merge	Merge existing columns with additional data to add station's parish
Handle missing values	Fill in the missing values regarding the Parish column with the corresponding value
General column settings	Change data types, rename and remove columns

The time dataflow is based on an integer number list from 0 to 23, with each line representing an hour of the day. From that a column with the start time of each hour of the day is created, an index column added and the data types and column names fixed.

Table 15 - Staging Area Time Dimension dataflow operations

Type of Operation	Description
List hours	Generate the hours of a day, from 0 to 23
Create columns	Create columns Hour of Day and Hour Formatted
Add index	Add a column with the index

Following the creation of each dimension's dataflow, the dataflows for the fact tables were constructed. The trip fact dataflow is based on multiple source files containing trip-related metrics, each corresponding to different time intervals within the years 2022 and 2023. Additionally, it incorporates the date dimension table. The transformation process begins with merging and appending the source files, followed by filtering rows, removing and renaming columns and eliminating duplicate records to ensure data consistency and accuracy.

Table 16 - Staging Area Fact Trip dataflow operations

Type of Operation	Description
Filter rows	Filter the rows from the source data comprehend only records from 2022 and 2023
Merge	Merge the existing columns with the date dimension to add the index column from the date table to the trip dataflow
Remove duplicates	Remove the duplicate rows to ensure data integrity
General column settings	Rename, reorder and remove columns

Similarly, the occupancy fact dataflow is resultant from source files that record real-time updates on station status – comprehending data such as the number of available bicycles, and from the date and station dimension tables. To achieve the required shape for the table, non-relevant columns are removed and date and time columns are generated from existing attributes thus enabling the establishment of a relationship with the respective dimensional tables.

Table 17 - Staging Area Fact Occupancy dataflow operations

Type of Operation	Description
Create columns	Add Date, Time and Start of Hour columns from the existing columns of the source data
Extract characters	Extract the first three characters of Station Name column to create the Station ID
Change data type	Adjust the data types from column Date and Time

Type of Operation	Description
Merge	Merge the existing columns with date, time and station dimension to add the index columns to the occupancy dataflow
Filter row	Filter rows to prevent lines with Station ID null from being in the table
General column settings	Reorder and remove columns

The weather dataflow uses four comma-separated values (CSV) files as source of data, each containing hourly meteorological data collected from different locations across Lisbon. To align the dataset with the physical model, non-relevant columns are removed and the column containing different meteorological indicators in its rows is pivoted, resulting in a structure where each metric is represented as a distinct column. Furthermore, inconsistent records are filtered out to grant data integrity and truthfulness. Lastly, the meteorological indicator's columns are grouped by date and time to present their average values, column names are refined and data types are properly assigned.

Table 18 - Staging Area Fact Weather dataflow operations

Type of Operation	Description
Pivot Columns	Pivot the Parameter and Value columns
Create column	Add Date and Time columns
Filter rows	Filter the rows to comprehend only records whose Temperature column values range from -7.5 and 47
Group rows	Group rows by Date and Time with average value of each meteorological parameter
Merge	Merge the existing columns with data and time dimensions
General column settings	Remove, rename and reorder columns

Finally, once all transformation steps are completed, the processed data is directed from its Dataflow Gen2 to the designated destination — namely the previously created tables in the Staging Area.

Following the development of the dataflows, a pipeline was implemented to automate the ingestion, transformation and loading of data from the Lakehouse into the Staging Area – Figure 3-9. The pipeline developed is composed of three major steps, having as initial stage two scripts designed to clear the existing data from the Staging Area fact and dimension tables, followed by a wait mechanism that prevents the next stage of the pipeline from beginning until the preceding step is successfully completed. Subsequently, the data is transformed and loaded into the dimension tables through the execution of the respective dataflows. Upon the successful completion of this stage, the data is then processed and loaded into the fact tables.

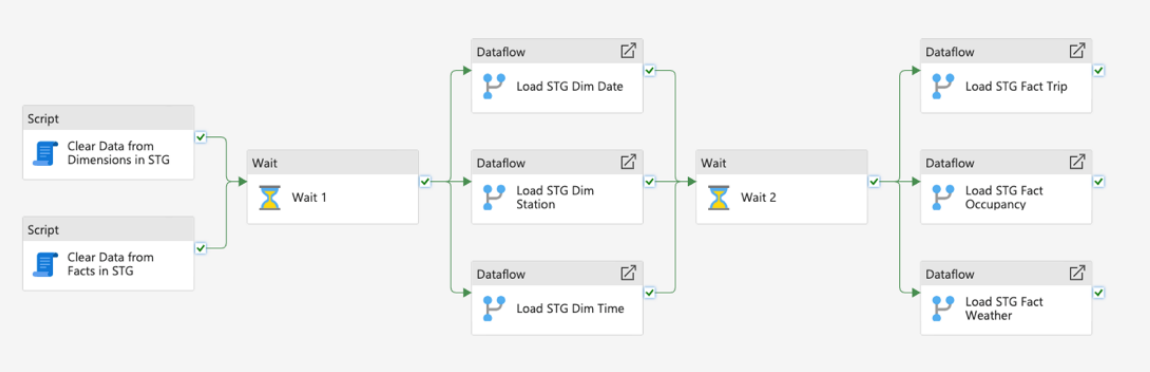

Figure 3-9 - Data Pipeline created regarding the load of data to the Staging Area

Following the loading of data into the Staging Area tables the contents of each table are subjected to validation procedures to ensure compliance with four data quality rules. The first and second rules, applied to all three dimensional tables, verify the integrity of business keys and the uniqueness of dimension attributes, respectively. The validation process proceeds with the fact tables. Specifically, the third rule ensures the integrity of foreign key combinations, while the fourth verifies the existence of each foreign key within the corresponding column of the associated parent dimension table. To process the above-mentioned data integrity rules a pipeline is created and executed – Figure 3-10.

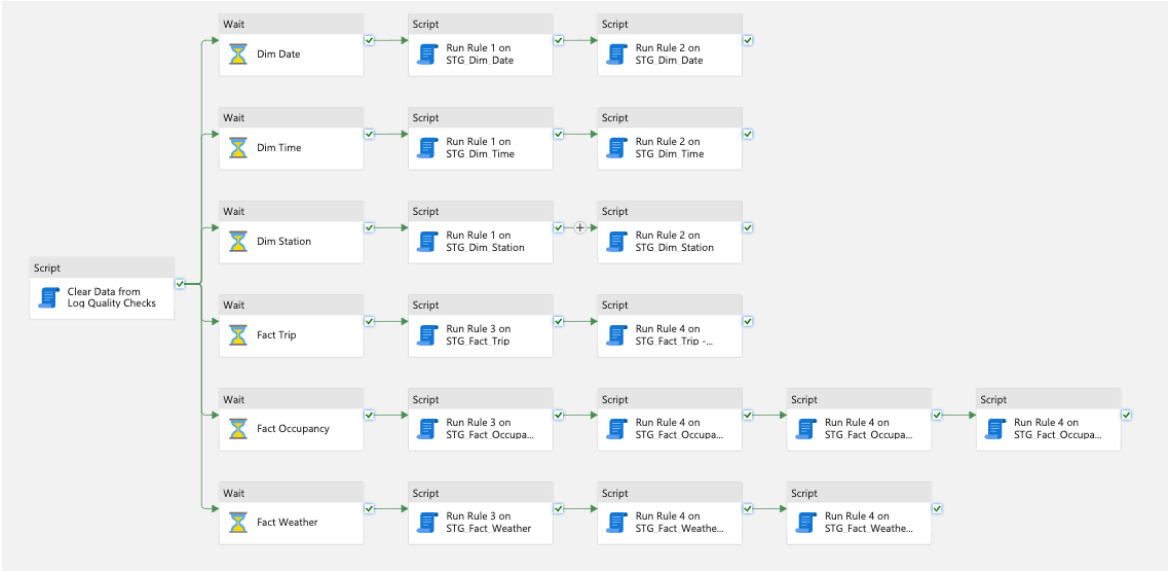

Figure 3-10 - Data Pipeline created to validate the data integrity

Table 19 - Validation Rules Scripts

Rule	Script
Rule # 1 **Check Integrity of Business Key**	INSERT INTO STG_GIRA.dbo.Log_Quality_Checks SELECT 1 AS ID_CHECK, 'Staging Area' AS ETL_PHASE, 'STG_Dim_Date' AS ETL_TABLE, 'check integrity of BK' AS ETL_CHECKTYPE, 'number of rows with repeated BK: ' + str(count(row_count)) AS DESCRIPTION_RESULT, CASE WHEN count(row_count) > 0 THEN 'FAIL' ELSE 'OK' END AS ETL_RESULT FROM (SELECT count(*) AS row_count FROM STG_GIRA.dbo.STG_Dim_Date GROUP BY SK_Date HAVING count(*) > 1) q;
Rule # 2 **Check Uniqueness of dimension attributes**	INSERT INTO STG_GIRA.dbo.Log_Quality_Checks SELECT 2 AS ID_CHECK, 'Staging Area' AS ETL_PHASE, 'STG_Dim_Date' AS ETL_TABLE, 'check uniqueness of all dim attributes' AS ETL_CHECKTYPE, 'number of rows NOT unique: ' + str(count(row_count)) AS DESCRIPTION_RESULT, CASE WHEN count(row_count) > 0 THEN 'FAIL' ELSE 'OK' END AS ETL_RESULT FROM (SELECT count(*) AS row_count FROM STG_GIRA.dbo.STG_Dim_Date GROUP BY ProperDate, Day, DayWeek, DayName, Month, MonthName, TrimesterNumber, SemesterNumber, Year, WeekYear, IsWeekend, IsHoliday, HolidayName HAVING count(*) > 1) q;
Rule # 3 **Check Integrity of Fact table's PK**	INSERT INTO STG_GIRA.dbo.Log_Quality_Checks SELECT 3 AS ID_CHECK, 'Staging Area' AS ETL_PHASE, 'STG_Fact_Trip' AS ETL_TABLE, 'check integrity of fact PK (combo all FKs)' AS ETL_CHECKTYPE, 'number of rows with repeated PK: ' + str(count(row_count)) AS DESCRIPTION_RESULT,

Rule	Script
	CASE WHEN count(row_count) > 0 THEN 'FAIL' ELSE 'OK' END AS ETL_RESULT FROM (SELECT count(*) AS row_count FROM STG_GIRA.dbo.STG_Fact_Trip GROUP BY FK_Date HAVING count(*) > 1) q;
Rule # 4 Check relationship of foreign key in Fact table	INSERT INTO STG_GIRA.dbo.Log_Quality_Checks SELECT 4 AS ID_CHECK, 'Staging Area' AS ETL_PHASE, 'STG_Fact_Trip' AS ETL_TABLE, 'check parent of FK for Date dimension' AS ETL_CHECKTYPE, 'number of rows without parent key: ' + str(max(row_count)) AS DESCRIPTION_RESULT, CASE WHEN max(row_count) > 0 THEN 'FAIL' ELSE 'OK' END AS ETL_RESULT FROM (SELECT count(*) AS row_count FROM STG_Fact_Trip LEFT JOIN STG_Dim_Date ON STG_Fact_Trip.FK_Date = STG_Dim_Date.SK_Date WHERE STG_Dim_Date.SK_Date IS NULL) q;

Upon the successful completion of the validation data pipeline, the data is considered ready for loading into the DW, which serves as final destination within the architecture.

Prior to transferring data to the DW, a dedicated Dataflow Gen2 is developed for each fact and dimension table. Within these dataflows, the original business keys and Surrogate keys are replaced by new Surrogate keys and foreign keys, respectively – Figure 3-11. This process ensures that the data utilized in the BI track remains decoupled from the source systems, thereby reinforcing the robustness and integrity of the solution's architecture.

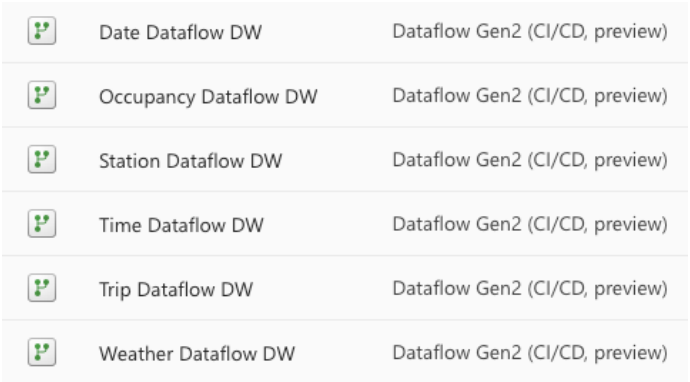

Figure 3-11 - Dataflows Gen2 created concerning the load of the Data Warehouse

Similar to the data transfer process from the Lakehouse to the Staging Area, the loading of data from the Staging Area to the DW is conducted through a dedicated pipeline – Figure 3-12.

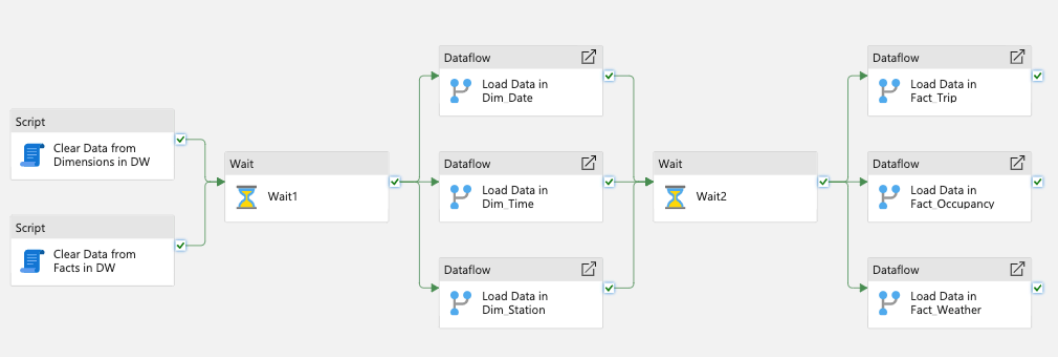

Figure 3-12 - Data pipeline created concerning the load of Data Warehouse

The development of the above mentioned ETL process was constrained by a significant limitation: the reliance on static data sources rather than establishing dynamic data connections through Application Programming Interfaces (API). Consequently, Excel files were utilized as the primary data sources instead of leveraging real-time data streams. Moreover, the data sources used prevented the establishment of regular and scheduled pipeline refreshes that would have enabled a continuous ETL processes, capturing and updating new data into the DW.

3.6 BI TRACK

The last path of the Kimball DW/BI Lifecycle is the BI Track, which encompasses the BI Application Design and BI Application Development stages. At this stage of the Kimball's methodology "the main (…) activities involve designing and building the initial set of reports and analyses called business intelligence (BI) applications" (Kimball et al., 2008, p.473).

Reflecting the subsequent milestone of the Kimball DW/BI Lifecycle, the development of the BI Application is initiated. Firstly, a Semantic Model (SM) is created – Figure 3-13. This model is directly connected to the DW, where the transformed and validated data is loaded. A SM is a logical representation of an analytical domain, typically structured in a star schema format, and serves as a foundation for deep data exploration. This model facilitates the execution of diverse analytical operations, including slicing, dicing, drilling, filtering and the computation of derived metrics (Microsoft, 2024b). Later on, a Power BI report is created having as data source the SM created, establishing a seamless interface with the ETL phase of the Kimball's Lifecycle and enabling the creation of dynamic and interactive visual analyses.

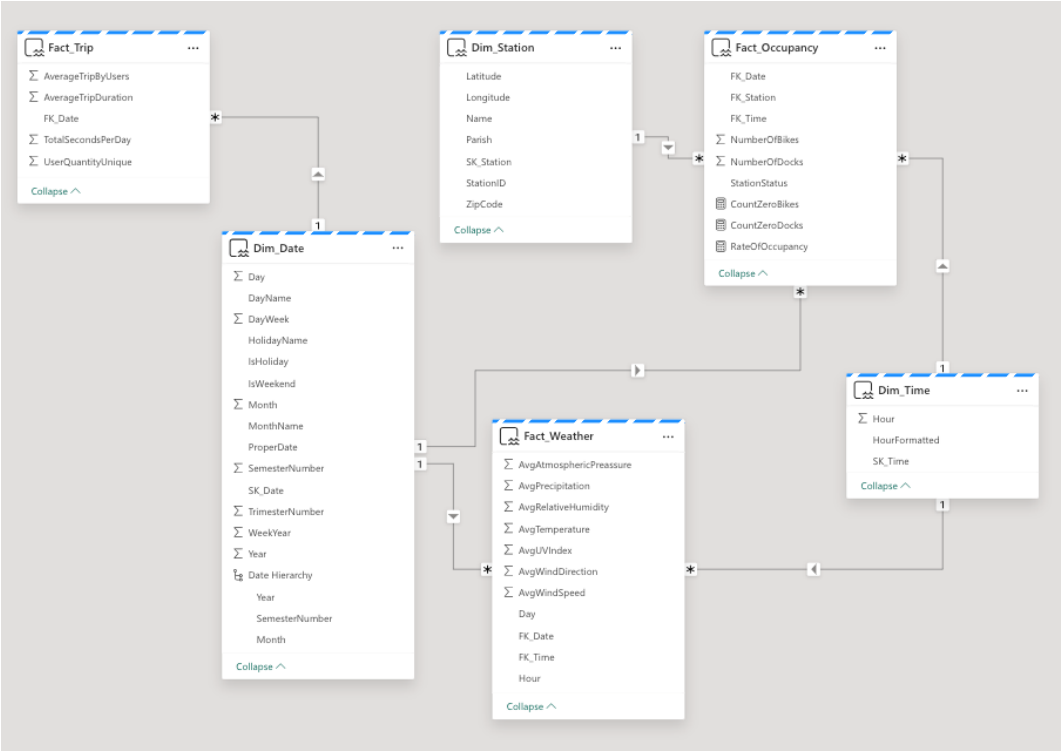

Figure 3-13 - GIRA Semantic Model

Additionally, based on the existing attributes within the fact and dimension tables, a set of calculated measures is developed. These measures, derived through logical and arithmetic operations on the original columns, are designed to support the resolution of the previously defined business questions, enable the generation of KPIs and facilitate the identification of patterns and trends within the data. A detailed description of the calculated measures is presented in Table 20.

Table 20 - Calculated Measures created for GIRA project

Calculated Measures	Measure Description	DAX Code
AvgUsagePerDay	Average number of hours of GIRA bicycles usage per day, in D HH:M format.	AvgUsagePerDay = Var AvgSeconds = AVERAGE('Fact_Trip'[TotalSecondsPerDay]) VAR Days = INT(AvgSeconds / 86400) VAR Hours = INT(MOD(AvgSeconds, 86400) / 3600) VAR Minutes = INT(MOD(AvgSeconds, 3600) / 60) RETURN FORMAT(Days, "0") & "D " & FORMAT(Hours, "00") & FORMAT(Minutes, "00")
AvgTimeFormatted	Average trip time, in MM:SS format.	AvgTimeFormatted = VAR AvgTime = AVERAGE('Fact_Trip'[AverageTripDurationMinutes]) VAR Minutes = INT(AvgTime)

Calculated Measures	Measure Description	DAX Code
		VAR Seconds = ROUND((AvgTime - Minutes) * 60, 0) RETURN FORMAT(TIME(0, Minutes, Seconds), "MM:SS")
CountZeroBikes	Counts the number of times a station experienced bicycle shortage.	CountZeroBikes = COUNTROWS(FILTER(Fact_Occupancy, Fact_Occupancy[NumberOfBikes] = 0))
CountZeroDocks	Counts the number of times a station experienced full dock usage.	CountZeroDocks = COUNTROWS(FILTER(Fact_Occupancy, Fact_Occupancy[NumberOfDocks] = 0))
RateOfOccupancy	Division between the number of bikes and the total number of docks available at a station.	RateOfOccupancy = DIVIDE(SUM('Fact_Occupancy'[NumberOfBikes]), SUM('Fact_Occupancy'[NumberOfDocks]))

The outcomes of the BI application design and development will be thoroughly detailed in the subsequent section, Results and Discussion.

3.7 Deployment, Growth and Maintenance

The Deployment, Growth and Maintenance are the last three milestones of the Kimball DW/BI Lifecycle. These comprehend activities to deliver the outputs to the end users, ensure their continued validity over time and expand their scope in future projects.

In the context of this specific project, rather than proceeding with the deployment of the solution to end users and establishing maintenance plans for its ongoing support, a set of recommendations for enhancing the existing solution will be proposed.

4. RESULTS & DISCUSSION

4.1 DASHBOARD OVERVIEW

To better analyze the usage of GIRA bicycles around the parishes of Lisbon a dedicated Power BI Report is created to approach the themes of Usage Patterns, Stations and Weather. Each report page enables a thorough visual exploration of the data and will be described in the following paragraphs.

The initial page of the report, depicted in Figure 4-1 offers a detailed overview of GIRA bicycle usage patterns for the years 2022 and 2023. The upper section of the report includes four Cards. On the left, the cards display the total number of users and the average number of unique users per day, while on the right, they show the average daily usage and the average trip duration in minutes. Positioned beneath the Cards is an Area chart that illustrates the monthly trend in average usage hours. Lastly, a Table highlights the days with the highest number of unique users.

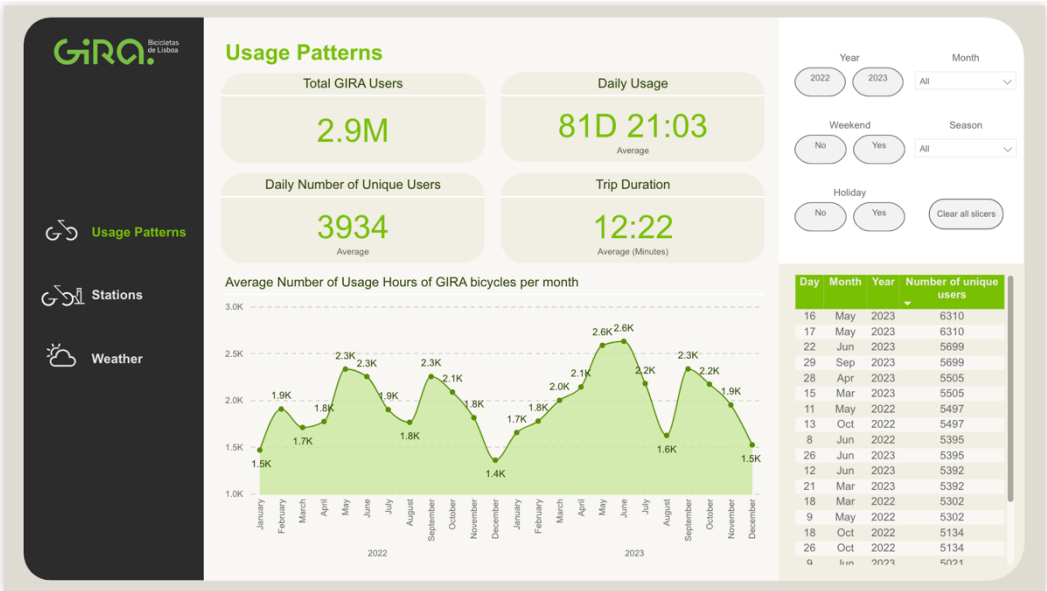

Figure 4-1 - Usage Patterns page report

The second page of the report comprises multiple visual components aimed at analyzing the GIRA docking stations – Figure 4-2. A Gauge visualization displays the average dock utilization rate, while a Clustered Column chart presents the average number of bicycles available over time. On the right, a Map illustrates the geographical distribution of stations across Lisbon's parishes. Additionally, a Table identifies the top ten stations most frequently experiencing bicycle shortages. Resorting to the Power BI bookmark functionality, two subpages are developed within the Stations analysis page to enable a focused evaluation of individual stations – Figures 4-3 and 4-4, addressing temporal and calendar aspects, respectively. Each of these subpages includes a Gauge visualization which displays the average utilization rate

for the selected station, similarly to the general Stations page. Additionally, three Cards are presented displaying the Parish in which the station is located, the Station's ID and its corresponding zip code. A Clustered Column Chart further complements the analysis by illustrating the average number of bicycles available at each hour of the day on the Station Hourly subpage and by month or day of the week on the Station Calendar subpage.

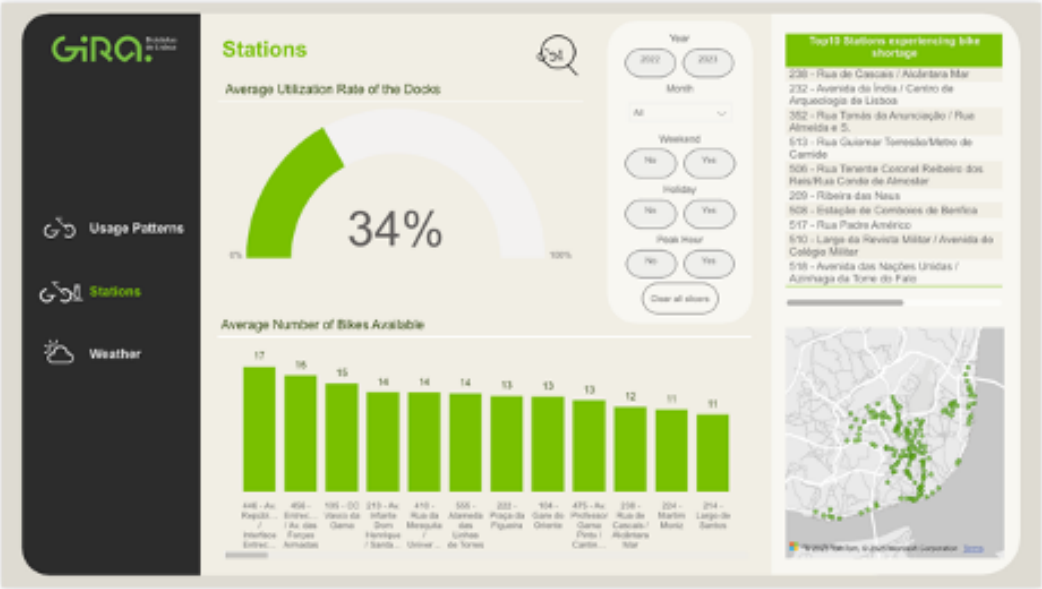

Figure 4-2 - Stations page report

Figure 4-3 - Stations Hourly page report

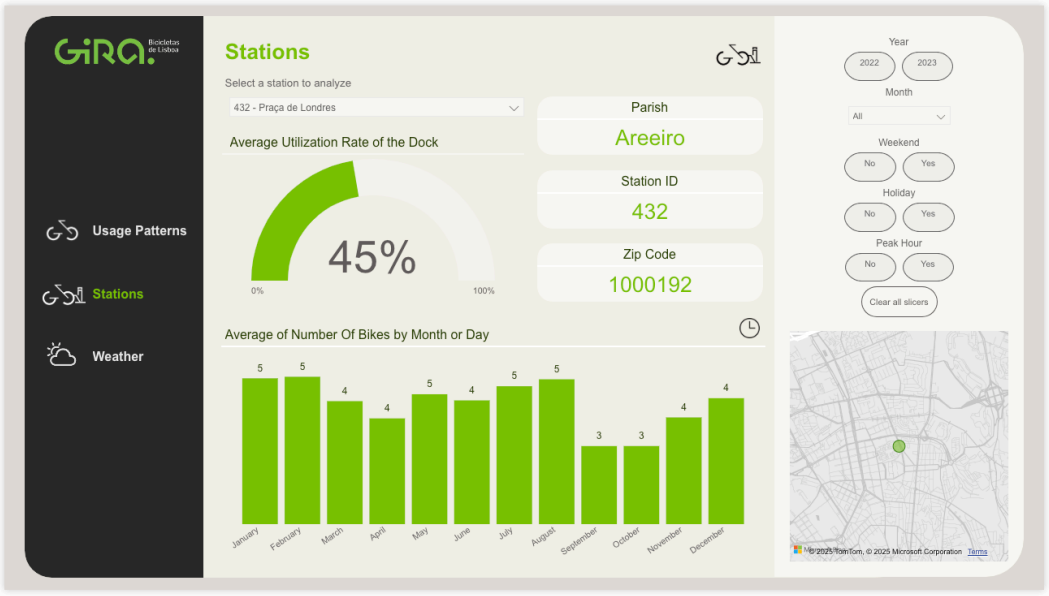

Figure 4-4 - Stations Calendar page report

The final page of the report explores the relation between GIRA bicycle usage and the weather conditions. It features two visualizations: a Column chart illustrating the distribution of usage hours across different temperature categories and a combined Column and Line chart that depicts the relationship between monthly average temperature and total hours of bicycle usage.

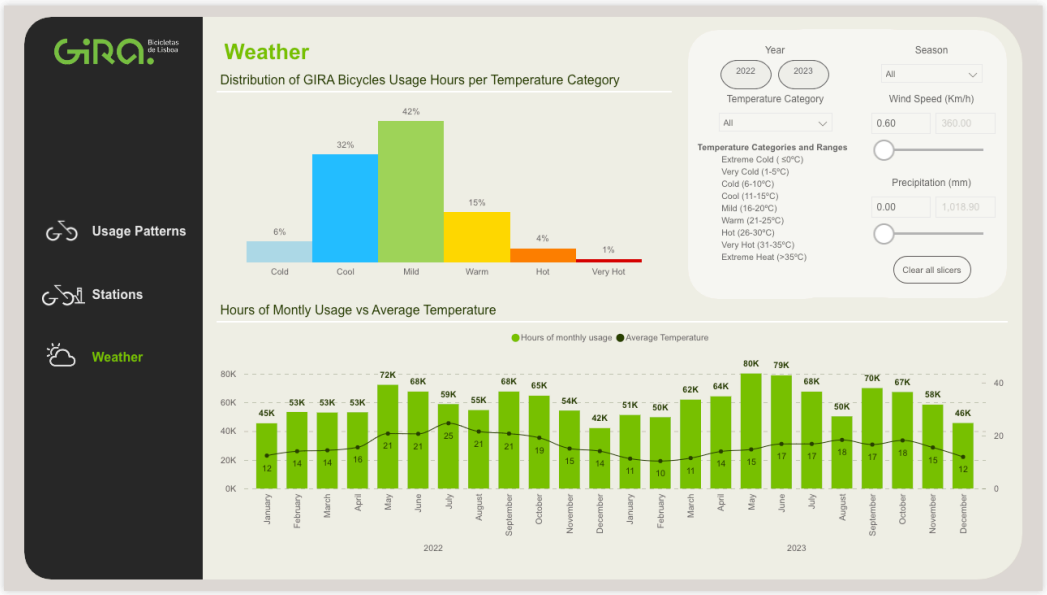

Figure 4-5 - Weather page report

Each page of the report includes interactive Slicers which allow users to filter the data dynamically performing more detailed analyses. These filters facilitate the identification of specific usage patterns and enhance the interpretability of the visualizations, thereby increasing the analytical depth and value of the dashboard. Filtering options include temporal

variables such as year, month and season, as well as contextual factors like peak hours, holidays, weekends, temperature categories, wind speed and precipitation levels.

4.2 DASHBOARD INSIGHTS

This subchapter aims at answering the formulated Business Questions in the different business scopes – Usage Patterns, Stations and Weather, regarding the utilization of GIRA bicycles around the 24 parishes of Lisbon.

Starting with the Usage Patterns scope, the following paragraphs address the business questions concerning the identification of overall trends in bicycle usage, assess the number of daily unique users and its variation across temporal and contextual factors and evaluate differences in usage between business days, weekends and holidays. Moreover, the analysis presents conclusions regarding the relationship between days with significantly higher usage levels and special events or external conditions. KPIs, such as average trip duration and its variation across different categories of calendar days, are also examined.

Regarding the Usage Patterns of the GIRA BSS throughout the years 2022 and 2023, a cumulative total of approximately 2.9 million users was recorded, corresponding to an average of 3.934 unique users per day. In terms of operational time, GIRA bicycles were used for an average of 81 days 21 hours and 3 minutes per day, which corresponds to around 1.965 hours of daily usage, with an average trip duration of 12 minutes and 22 seconds.

Analyzing the monthly variation in total usage hours reveals that May, June and September consistently registered the highest levels of usage across both years, with a combined monthly average of 2.400 hours. These months correspond to Spring and late Summer, typically associated with favorable weather conditions, ranging from mild to warm temperatures, which likely contribute to increased bicycle usage for commuting and recreational purposes. In contrast, the months of December, January and August recorded the lowest average usage durations. While this decline can be attributed to inclement weather during the Winter months (December and January) which are characterized by lower temperatures, higher precipitation and higher speed winds. In August the causes can be appointed to the traditional Summer holiday period, when the Portuguese population usually travels south or abroad.

Further insights can be drawn from the analysis of daily usage peaks. The dates with the highest number of unique users were the 16th and 17th of May 2023 and the 22nd of June 2023. These dates, all falling on weekdays, do not appear to coincide with any specific public events that might explain the high usage levels.

Focusing on variations between weekdays, weekends and public holidays, the analysis shows a decline in overall usage during non-working days compared to business days. The average number of unique users decreases from 4.430 on business days to 2.699 on weekends and to 2.515 on public holidays, a difference of almost two thousand daily users in both cases. In contrast, average trip durations increase slightly during these periods reaching 12 minutes and

35 seconds on weekends and 12 minutes and 50 seconds on holidays, a thirteen and twenty-eight second increase respectively.

When examining public holidays with the highest usage, the 13th of June 2023 ranks first followed by the 5th of October 2022 and the 10th of June 2023. The first and third dates coincide with the Santos Populares festivities, a traditional season during which residents of Lisbon spend more time outdoors and at historical neighborhoods of the city such as Alfama and Graça.

Shifting the focus to Stations scope, the next paragraphs aim at answering the business questions related to bicycle availability across the stations of the Lisbon's BSS infrastructure. The analysis investigates the average occupancy rate of bicycle docks, identifies stations that consistently experience shortages and assesses the average number of bicycles available during peak and off-peak hours. Additionally, it examines how bike availability varies across business days, weekends and holidays providing insights into spatial and temporal distribution patterns and potential pressure points within the system.

Starting with the average dock utilization rate of 34% was observed. This indicates that, on average, 34 out of every 100 docks had a bicycle parked and ready for use. This utilization rate remained relatively stable between 2022 and 2023, increasing marginally from 33% to 34%.

Resorting to the data presented in Table 21, it is possible to identify the stations that most frequently experience bicycle shortages. These stations consistently exhibit low availability during both peak and off-peak hours and have in common the fact that they are located in the parishes of Carnide or Benfica. Moreover, at least four of the highlighted stations, are close to major public transportation interfaces such as bus, metro or train stations.

Table 21 - Top 10 Stations experiencing bike shortages

Station Name	Number of Occurrences
238 – Rua de Cascais / Alcântara Mar	43696
232 – Avenida da Índia / Centro de Arqueologia de Lisboa	32313
352 – Rua Tomás da Anunciação / Rua Almeida e S.	24759
513 – Rua Guiomar Torresão / Metro de Carnide	20173
506 – Rua Tenente Coronel Reibeiro dos Reis / Rua Conde de Almoster	18376
209 – Ribeira das Naus	18183
508 – Estação de Comboios de Benfica	13551
517 – Rua Padre Américo	10491
510 – Largo da Revista Militar / Avenida do Colégio Militar	9659
518 – Avenida das Nações Unidas / Azinhaga da Torre do Fato	9407

Concerning the average number of bicycles available during peak and off-peak periods, the data reveals a stable pattern. During peak hours, 32% of docks are occupied, compared to 34% during off-peak hours. Several stations distinguish themselves for maintaining a higher-than-average availability of bicycles throughout the day. These include stations 446 and 456, located near *Estação de Entrecampos*; stations 104 and 105, near *Estação do Oriente*; and stations 410 and 475, situated close to *Universidade Nova de Lisboa* campus and *Cidade Universitária*, the *Universidade de Lisboa* campus. The sustained availability of bicycles at these stations may be attributed to their proximity to key activity centers, particularly those related to public transportation and academic institutions.

Finally, the average number of bicycles available on weekends and public holidays does not show significant variation when compared to regular weekdays, indicating consistent system performance across different types of calendar days.

The final scope to be analyzed is the Weather scope, which investigates the influence of meteorological conditions on the demand for GIRA rides. Specifically, the analysis examines the correlation between temperature and trip volume, as well as the impact of temperature, wind speed and precipitation on bicycle usage.

The analysis of the percentage distribution of GIRA bicycle usage across temperature categories reveals that approximately 75% of all trips occur on days classified as Cool (32%) or Mild (42%), with average daily temperatures ranging between 11°C and 20°C. A comparative analysis of data from 2022 and 2023 shows a marked decline in usage on Warm days (average daily temperatures ranging from 21°C to 25°C), with the proportion of usage hours falling from 20% in 2022 to 9% in 2023. This variation can likely be attributed to generally lower average temperatures observed in 2023.

Further examination of the monthly total hours of usage against the average daily temperature suggests a generally positive correlation between temperature and trip demand – higher temperatures are typically associated with increased usage and lower temperatures with reduced usage. Nevertheless, this trend does not hold consistency. For instance, in July 2022, although an average temperature was approximately 4°C higher than the adjacent months, the total hours of bicycle usage were significantly lower compared to both the preceding and subsequent two-month periods. This deviation may indicate the influence of additional contextual factors beyond temperature alone, such as heatwaves or holiday periods of schools and universities, with the latter being a factor previously seen as a propulsor of bike demand.

4.3 Comparison with Reviewed Literature

When comparing the insights derived from the developed dashboard with the findings reported in the reviewed literature, points of convergence can be identified. Duran-Rodas et al. (2019) and Eren & Uz (2020) studied the built environment influence in the demand for trips, identifying transport-related infrastructure and educational facilities as key

determinants of BSS ridership. Eren & Uz (2020) in particular, demonstrated a positive correlation between the presence of urban infrastructure and the volume of bike-sharing trips. This pattern is consistent with the results observed in the GIRA case study, where stations located near major train stations and academic institutions registered higher dock occupancy rates.

Furthermore, Eren & Uz (2020) emphasized that adverse weather conditions negatively affect bike-sharing demand, while moderate temperatures ranging between 20°C and 30°C in the absence of precipitation are considered optimal for trip volume. Although these conclusions are not directly observed on GIRA, it was observed that months normally associated with adverse weather conditions, such as December and January, had lower levels of bike demand, whereas months with favorable weather conditions, May and June, had higher levels of bike demand. When analyzing GIRA usage on days that fall within the optimal temperature rang identified by Eren & Uz, it was found that 22% of the trips occurred during such conditions.

Additionally, Kim (2018) in his analysis of the Tashu BSS, identified a reduction in bike usage on public holidays. This finding aligns with the GIRA data, where both the average number of unique users and the average daily usage decline on public holidays.

A deeper comparison with studies specifically focused on Lisbon's BSS further supports the validity of the insights derived from the current project. Albuquerque et al. (2021) reported that demand increases under weather conditions characterized by the absence of rain and temperatures between 10°C and 30°C, which aligns with the findings identified in the dashboard analysis, where approximately 93% of the trips covered fall within. Moreover, Lucas & Andrade (2021) demonstrated that BSS usage patterns vary according to the day of the week, which is also reflected in the GIRA data explored in this study.

5. CONCLUSIONS AND FUTURE RESEARCH

The present study aimed to develop a BI framework to support data-driven decision-making for decision makers regarding the existing BSSs. To ensure the rigor and correctness of each stage of the project performed, the methodology that Kimball DW/BI Lifecycle underlies was systematically applied. As a case study for the implementation of the proposed BI solution, Lisbon's BSS – GIRA, was selected, taking advantage of the availability of open data provided by EMEL and CML.

The Usage Patterns, the Stations and the relationship between the ride demand and the Weather conditions were selected as the main areas of analysis identified for the implementation of the developed framework. For each of the identified research scopes specific research questions were formulated such as "How does bike usage fluctuate between business days, weekends and holidays?", "Are there particular stations that consistently experience shortages of bikes?" and "Is there any correlation between temperature and trips?". These questions were then addressed and conclusions were drawn through the visualizations developed within the Power BI report.

Regarding the research scope focused on Usage Patterns, the study concluded that over the two-year period under study the months of May, June and September have registered higher number of average usage hours, indicating an increased demand for rides. Furthermore, when examining fluctuations across different calendar days, specifically among business days, weekends and public holidays, the data demonstrated a notable decrease of approximately 70% in the average number of unique users on weekends and holidays, when compared to the figures observed on business days. In contrast, the average trip duration increased on non-working days, weekends and holidays.

The Station analysis scope comprehended both aggregated and station-specific perspectives. When evaluating the average occupancy rate of docking stations across the two-year period, results indicated a slight increase from 33% in 2022 to 34% in 2023. This rate remained relatively stable on non-working days, showing only marginal variations of approximately 1%. One of the most interesting insights from the analysis was that among the ten stations most frequently experiencing bicycle shortages, four are located in close proximity to public transportation hubs. On the other hand, and contrastingly, GIRA stations situated near two of Lisbon's largest train stations, *Entrecampos* and *Oriente*, exhibited above-average bike availability throughout the day, alongside with the stations situated near *Universidade de Lisboa* and *Universidade Nova de Lisboa* campuses.

With regard to the relationship between weather conditions and the demand for GIRA rides, the analysis indicates a general trend of increased usage as temperatures rise. The data further reveals that bicycle usage is more pronounced during the Autumn and Spring seasons, contrasting with Winter and Summer, with the reduced levels of usage of the latter might be attributed to the traditional holiday season. Additionally, the analysis shows that

approximately 75% of all trips recorded during the study period occurred on days classified as Cool or Mild.

The development of this study was subject to certain limitations external to the adopted methodology. One of the constraints involved the reliance on static datasets as the primary data source for the ETL process as opposed to establishing a dynamic data connection via an API. This limitation arose due to the unavailability of real-time data feeds from the EMEL open data portal which was undergoing maintenance during the project development. Consequently, Excel files were employed as the main input for data extraction, which prevented the implementation of automated and scheduled pipeline refreshes, thereby blocking the establishment of a continuous ETL workflow capable of updating data regularly into the Data Warehouse. Another limitation encountered was the limited recency of the data sets related to stations, usage, trips, weather indicators and occupancy. The most recent data available corresponded to the year 2023. Aggravating the lack of recent data, in the specific case of the occupancy dataset the data did not cover the entire year. Although the datasets utilized were sufficient to validate the BI framework developed, their outdated nature limits their applicability for supporting current decision-making processes.

Future research and improvement plans for the BI solution can be outlined based on the limitations identified as well as on potential technical enhancements related to the ETL process. A reformulation of the ETL workflow could be achieved through the adoption of Python Notebooks in place of Dataflows Gen2 expected to enhance the transparency and manageability of error-handling procedures, while also optimizing memory usage for the processing of larger datasets. Additionally, the integration of ML models to predict the number of bikes available at each GIRA station in the hours following the most recent log could be explored. However, this improvement is subject to the availability of EMEL's open data portal, which is necessary to establish access via an API. Lastly, an analysis comparing the usage of the conventional versus the electric versions of GIRA's bike could be developed upon the availability of the necessary data.

BIBLIOGRAPHICAL REFERENCES

Albuquerque, V., Andrade, F., Ferreira, J. C., & Dias, M. S. (2021). *Understanding Spatiotemporal Station and Trip Activity Patterns in the Lisbon Bike-Sharing System*. https://doi.org/10.1007/978-3-030-71454-3_2

Bahadori, M. S., Gonçalves, A. B., & Moura, F. (2021). A systematic review of station location techniques for bicycle-sharing systems planning and operation. In *ISPRS International Journal of Geo-Information* (Vol. 10, Issue 8). MDPI. https://doi.org/10.3390/ijgi10080554

Bahadori, M. S., Gonçalves, A. B., & Moura, F. (2022). A GIS-MCDM Method for Ranking Potential Station Locations in the Expansion of Bike-Sharing Systems. *Axioms*, *11*(6). https://doi.org/10.3390/axioms11060263

Banerjee, S., Kabir, M. M., Khadem, N. K., & Chavis, C. (2020). Optimal locations for bikeshare stations: A new GIS based spatial approach. *Transportation Research Interdisciplinary Perspectives*, *4*. https://doi.org/10.1016/j.trip.2020.100101

Boufidis, N., Nikiforiadis, A., Chrysostomou, K., & Aifadopoulou, G. (2020). Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators. *Volume 47, Pages 51 - 58*, *47*, Barcelona. https://doi.org/10.1016/j.trpro.2020.03.072

Câmara Municipal de Lisboa. (n.d.). *Câmara Municipal de Lisboa, Mobilidade, Mobilidade Ciclável*. Retrieved January 26, 2025, from https://www.lisboa.pt/temas/mobilidade/modos-de-transportes/bicicleta

Câmara Municipal de Lisboa. (2018). *Lisboa Aberta*. https://lisboaaberta.cm-lisboa.pt/index.php/pt/

Câmara Municipal de Lisboa. (2019a, January 2). *Estações GIRA em Operação | Câmara Municipal Lisboa - Geodados*. https://geodados-cml.hub.arcgis.com/datasets/CML::esta%C3%A7%C3%B5es-gira-em-opera%C3%A7%C3%A3o/explore?location=38.735330%2C-9.158336%2C13.00

Câmara Municipal de Lisboa. (2019b, January 2). *Rede Ciclável | Câmara Municipal Lisboa - Geodados*. https://geodados-cml.hub.arcgis.com/datasets/CML::ciclovias-2/explore?layer=0&location=38.743502%2C-9.154065%2C13.00

Câmara Municipal de Lisboa. (2025, May 9). *Rede de bicicletas partilhadas de Lisboa continua a crescer - Informação Lisboa*. https://informacao.lisboa.pt/noticias/detalhe/rede-de-bicicletas-partilhadas-de-lisboa-continua-a-crescer

Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX*, *9*, 101895. https://doi.org/10.1016/J.MEX.2022.101895

Duran-Rodas, D., Chaniotakis, E., & Antoniou, C. (2019). Built Environment Factors Affecting Bike Sharing Ridership: Data-Driven Approach for Multiple Cities. *Transportation Research Record*, *2673*(12), 55–68. https://doi.org/10.1177/0361198119849908

EMEL. (2020, June 18). *GIRA - Bicicletas de Lisboa - Portal Dados Abertos*. https://dados.cm-lisboa.pt/dataset/giras-docas

EMEL. (2021, April 27). *GIRA - Bicicletas de Lisboa (Histórico)*. https://dados.cm-lisboa.pt/dataset/gira-bicicletas-de-lisboa-historico

Eren, E., & Katanalp, B. Y. (2022). Fuzzy-based GIS approach with new MCDM method for bike-sharing station site selection according to land-use types. *Volume 76*, *76*. https://doi.org/10.1016/j.scs.2021.103434

Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. In *Sustainable Cities and Society* (Vol. 54). Elsevier Ltd. https://doi.org/10.1016/j.scs.2019.101882

Huff, D. L. (1962). A Note on the Limitations of Intraurban Gravity Models. *Land Economics*, *38*(1), 64. https://doi.org/10.2307/3144725

Hulot, P., Aloise, D., & Jena, S. D. (2018). Towards station-level demand prediction for effective rebalancing in bike-sharing systems. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 378–386. https://doi.org/10.1145/3219819.3219873

Kim, K. (2018). Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Volume 66, Pages 309 - 320*, *66*, 309–320. https://doi.org/10.1016/j.jtrangeo.2018.01.001

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit* (3rd ed.). Wiley & Sons, Inc.

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). *The Data Warehouse Lifecycle Toolkit, Second Edition*.

Kou, Z., Wang, X., Chiu, S. F. A., & Cai, H. (2020). Quantifying greenhouse gas emissions reduction from bike share systems: a model considering real-world trips and transportation mode choice patterns. *Resources, Conservation and Recycling*, *153*. https://doi.org/10.1016/j.resconrec.2019.104534

Lucas, V., & Andrade, A. R. (2021). Predicting hourly origin–destination demand in bike sharing systems using hurdle models: Lisbon case study. *Volume 9, Issue 4, Pages 1836 - 1848*, *9*(4), 1836–1848. https://doi.org/10.1016/j.cstp.2021.10.003

Maleki, A., Nejati, E., Aghsami, A., & Jolai, F. (2023). Developing a supervised learning-based simulation method as a decision support tool for rebalancing problems in bike-sharing systems. *Expert Systems with Applications*, *233*, 120983. https://doi.org/10.1016/J.ESWA.2023.120983

Microsoft. (n.d.). *Microsoft Power Platform | Power BI - Data Visualization*. Retrieved May 1, 2025, from https://www.microsoft.com/en-us/power-platform/products/power-bi?market=pt

Microsoft. (2024a, March 22). *What is Power BI? - Power BI | Microsoft Learn*. https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

Microsoft. (2024b, June 8). *Default Power BI semantic models - Microsoft Fabric | Microsoft Learn*. https://learn.microsoft.com/en-us/fabric/data-warehouse/semantic-models

Microsoft. (2024c, August 22). *What is data warehousing in Microsoft Fabric? - Microsoft Fabric | Microsoft Learn*. https://learn.microsoft.com/en-us/fabric/data-warehouse/data-warehousing

Microsoft. (2024d, September 16). *Activity overview - Microsoft Fabric | Microsoft Learning*.

Microsoft. (2024e, December 18). *What is Data Factory - Microsoft Fabric | Microsoft Learn*. https://learn.microsoft.com/en-us/fabric/data-factory/data-factory-overview#dataflows

Midgley, P. (2011). *BICYCLE-SHARING SCHEMES: ENHANCING SUSTAINABLE MOBILITY IN URBAN AREAS*.

Sharda, R., Delen, D., & Turban, E. (2017). *Business Intelligence, Analytics, and Data Science _ A -- Ramesh Sharda, Dursun Delen, Efraim Turban -- 4, 2017-01-13 -- Pearson -- 9780134633282* (Pearson, Ed.; 4th ed.).

United Nations. (n.d.). *Sustainable transport | Department of Economic and Social Affairs*. Retrieved May 25, 2025, from https://sdgs.un.org/topics/sustainable-transport#description

Wang, K., Akar, G., & Chen, Y. J. (2018). Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. *Transportation Research Part A: Policy and Practice*, *116*, 1–14. https://doi.org/10.1016/J.TRA.2018.06.001

Zhang, Y., & Mi, Z. (2018a). Environmental benefits of bike sharing: A big data-based analysis. *Applied Energy*, *220*, 296–301. https://doi.org/10.1016/J.APENERGY.2018.03.101

Zhang, Y., & Mi, Z. (2018b). Environmental benefits of bike sharing: A big data-based analysis. *Applied Energy*, *220*, 296–301. https://doi.org/10.1016/J.APENERGY.2018.03.101

APPENDIX A

Appendix 1 - Kimball's DW/BI Lifecycle Milestones Description

Milestone	Description
Program/ Project Planning	The main activities of this milestone are defining the scope based on business requirements, resource staffing, task identification, assignment, duration estimation and sequencing.
Program/ Project Management	The management phase, which spans during the entire lifecycle, focus on overseeing and tracking the program/ project's progress.
Business Requirements Definition	At this stage, analysts develop a comprehensive understanding of the business's key drivers, enabling them to effectively translate these insights into the design of the DW/BI system.
Technology Track	Technical Architecture Design – this step of the Lifecycle focusses on establishing architectural framework and vision, considering the business requirements, current technical environment and technical strategic plans of the organization. Product Selection and Installation – this stage uses the output from the preceding to filter the software and hardware options available, which meet the requirements for the program/ project. Once selected the tools are installed and tested to ensure a seamless integration with the DW/BI.
Data Track	Dimensional Modeling – at this milestone a preliminary matrix representing the business processes is build, used then as blueprint for the data architecture. Physical Design – this step consists in defining the physical structures of the DW/BI, addressing topics such as the database environment and security procedures. Extract, Transform and Load (ETL) Design and Development – at this stage the data is extracted (E) from its source, it is transformed (T) and loaded (L) to tables, enabling users query it.
BI Application Track	BI Application Design – this milestone focuses on identifying BI applications as well as appropriate interfaces that meet the needs and requirements of the program/ project. "BI applications are the vehicle for delivering business value from the DW/BI solution" (Kimball et al., 2008, p.7) BI Application Development – at this stage the configuration of the metadata and tool infrastructure take place, alongside with the development of specific BI applications, either analytic or operational and the navigation portal.
Deployment	At this Lifecycle's phase the three tracks converge, requiring planning, testing and alignment to ensure a flawless implementation

Milestone	Description
Maintenance	This step focuses on monitor the usage of DW/BI, tune its performance, together with other technical operational tasks.
Growth	At the final step of the DW/BI Lifecycle the processes to be addressed should be prioritized and when expanding the existing system, the new objects should leverage from the existing foundations.

Appendix 2 - Station Table Metadata

Column	Data Type	Description
id_expl	INT	Station ID
estacaolocalizacao	VARCHAR	Station ID and address
latitude	DECIMAL	Station's latitude
longitude	DECIMAL	Station's longitude
dispbicicleta	INT	Number of bicycles available
horariofuncionamento	VARCHAR	Information about stations' working schedule
tarifario	VARCHAR	Information about usage tariffs
formaspagto	VARCHAR	Information about usage tariffs payment methods available
contactoservassistencia	VARCHAR	Information about support contacts
wifi	BOOL	Boolean variable for Wi-Fi availability at the station
aberturadt	DATETIME	Station's opening date and time
criacaodtt	DATETIME	Station's creation date and time
atualizacaodtt	DATETIME	Station's data update date and time
cp7	INT	Station's zip code
servicosextra	VARCHAR	Information about extra services provided by the station

Appendix 3 - Usage Table Metadata

Column	Data Type	Description
tripStartDate	DATE	Date concerning the register
totalSecondsperDay	INT	Total number of seconds of usage
userQtyUnique	INT	Total number of unique users
avgTripbyUserSeconds	DECIMAL	Average number of seconds of usage per user

Appendix 4 - Trips Table Metadata

Column	Data Type	Description
tripStartDate	DATE	Date concerning the register
avgTripSeconds	DECIMAL	Average trip time in seconds
avgTripSecondsRush	DECIMAL	-

Appendix 5 - Weather Table Metadata

Column	Data Type	Description
DTM_UTC	DATETIME	Universal Time Coordinated record's date and time
DTM_LOCAL	DATETIME	Local record's date and time of the
TEMATICA	VARCHAR	Meteorology / Weather
COD_PARAMETRO	VARCHAR	Meteorologic parameter code
PARAMETRO	VARCHAR	Meteorologic parameter
NR_ESTACAO	INT	Meteorologic station number
COD_SENSOR	VARCHAR	Sensor code
LOCAL	VARCHAR	Meteorologic station address
LATITUDE	DECIMAL	Meteorologic station latitude
LONGITUDE	DECIMAL	Meteorologic station longitude
UNIDADE	VARCHAR	Meteorologic parameter measure unit
ETIQUETA_NIVEL	VARCHAR	Category corresponding to the meteorologic parameter value
COR_NIVEL	VARCHAR	Color corresponding to the meteorologic parameter value
VALOR	DECIMAL	Meteorologic parameter value

Appendix 6 - df2022 and df2023 Metadata

Column	Data Type	Description
desigcomercial	VARCHAR	Station's commercial designation
numbicicletas	INT	Number of bicycles available
numdocas	INT	Station's number of docks
position	VARCHAR	Information about station's location – latitude and longitude
entity_ts	DATETIME	Update of occupancy date and time
estado	VARCHAR	Status of the station

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa