Polytechnic of Torino

MASTER DEGREE DIGITAL SKILLS FOR SUSTAINABLE SOCIETAL TRANSITIONS

The use of mobile phone data to characterise the mobility patterns: challenges and limits

University supervisor :

Prof. Cristina Pronello

Company supervisors:

Mr. Luigi Galieni, Ferrovie dello Stato Italiane

Mrs. Sara Nourbakhsh, Ferrovie dello Stato Italiane

Candidate: Seyedeh Atefeh Asad

December 2024

INDEX

Table of Figures

Acknowledgements

This thesis represents not just my work but the support, kindness, and encouragement of many wonderful people who have been part of my journey.

First, my deepest thanks go to my amazing professor. She is an incredible person who, even with her busy schedule, always found time to guide me and provide thoughtful feedback. Her support has been a pillar of strength throughout this process, and I am truly grateful.

I also want to express my heartfelt gratitude to my dear supervisors at FS, Luigi and Sara. They were not only my mentors but also my greatest supporters and friends. Their efforts, kindness, and dedication went far beyond what I could have hoped for. Words cannot fully express how much their guidance has meant to me.

To my family—my mother, father, and beloved brother—thank you for being my foundation. You gave me the wings to fly and the courage to chase my dreams. Your love and belief in me have been my greatest source of strength.

Lastly, I want to thank myself. There were moments when I was tough on myself, but I stayed strong and kept moving forward. To that version of me, I say: thank you for not giving up. I am proud of the resilience and determination that brought me here.

Grazie tutti .

Abstract

Reducing carbon emissions in the transport sector is essential for the EU to reach its climate targets. EU policies focus on promoting sustainable mobility while ensuring that regions across Europe remain well-connected. To effectively shape these sustainable transport systems, it is essential to understand the mobility patterns and the motivations behind them. Travel behaviour – which includes choices about when, where, and how individuals travel – directly influences the flow of people and goods, and is shaped by factors like transport alternatives, personal preferences, and geographic location. The rapid evolution of data collection methods has opened new possibilities for analysing and modelling these behaviours, offering deeper insights to support efficient and sustainable transport systems. In this context, this research aims to examine and compare the potential of two different data collection methods for understanding travel behaviour. Specifically, a survey-based approach and mobile phone data analytics are compared, focusing on Vodafone's mobile network data and the Audimob Observatory survey. Those databases are provided by Ferrovie dello Stato Italiane (FSI), the national Railways, which is the primary partner of this research. The thesis followed a methodology based on four steps.

The first phase provides a comprehensive literature review that examines existing research on methods used to analyse travel behaviour, focusing on mobile phone data and traditional survey techniques. In the second step, Tableau software was used to analyse travel data in both Vodafone and Audimob datasets to identify mobility patterns. This phase explored which aspects of travel behaviour each dataset was able to capture. After using Tableau, the third step compared the two approaches to investigate and understand their unique features; to this end, a SWOT analysis was conducted to consider the strengths, weaknesses, opportunities, and threats of both methods considered. Finally, the benefits of each approach were assessed to determine their suitability for different research objectives and their value for planning purposes. This comparison ultimately offers insights into the advantages and disadvantages of big data versus traditional surveys in understanding travel behaviour. The results of the comparison of the two datasets show how each of them has unique features but using them individually is to properly describe the mobility patterns. Big data provides a large volume of information, which is particularly useful for analysing trips across various times and zones. However, it lacks the precision needed to identify specific details – such as the mode of transport – which can be more accurately captured using datasets like Audimob.

Another issue is that despite the valuable data provided by the data sources (Vodafone), there is still a lack of transparency regarding the algorithms and methods used to generate it. This uncertainty makes it challenging to fully understand and interpret the dataset, thereby limiting the depth and reliability of the analysis. Additionally, when comparing the two datasets over the same time, significant discrepancies are often observed, further complicating the reliability of the information provided by big data alone. Therefore, developing a method that integrates both datasets is crucial. By combining them, the complementary strengths of each can be fully leveraged, resulting in more comprehensive and accurate insights.

1. Introduction

Travel behaviour refers to the patterns, preferences, and decisions individuals or groups make when moving from one place to another. It encompasses various aspects of travel, including how people choose to travel, where they go, and why and when they travel. Understanding travel behaviour involves analysing these factors to build a comprehensive picture of people's movement in daily life, whether it is a commute to work, a grocery store, or a vacation. Travel behaviour studies can focus on short trips within a city or on long-distance travel across countries, each with distinct patterns and influencing factors.

A key component of travel behaviour is the mode of transport that people choose. This can include walking, cycling, driving, taking public transport and other modes. Mode selection is influenced by several factors like distance, cost, convenience, and individual preferences, making it one of the most important elements of travel behaviour. Another important aspect is the purpose of the trip, each purpose shapes travel patterns, influencing aspects such as timing, frequency, and even the choice of travel mode. Timing and frequency also play essential roles in travel behaviour. When and how often people travel impacts traffic flow, public transport demand, and even economic activities in specific areas. For example, commuting patterns create peak hours during morning and evening rush periods, while seasonal tourism peaks influence travel demand in popular tourist destinations. People’s route choices further contribute to understanding travel behaviour, as routes are often selected based on convenience, traffic conditions, and safety considerations, which can vary significantly based on individual circumstances.

Studying travel behaviour is vital for governments, businesses, and urban planners, as it helps them design more efficient, accessible, and sustainable transport systems. By understanding travel behaviour, policymakers can work proactively to improve infrastructure, reduce congestion, and promote sustainable mobility. Travel behaviour insights also allow planners to design cities that better support residents’ mobility needs, enhance the quality of life, and reduce environmental impacts associated with transport systems.

However, gathering and interpreting data on travel behaviour poses several challenges. One major obstacle is the diversity in travel habits among different demographic groups which makes it difficult to obtain a realistic representative display and may require different data collection methods. People's travel patterns vary based on factors such as age, income, and purpose of travel. This diversity requires delicate data collection methods and sophisticated analytical tools to ensure accurate representation across the population. Additionally, privacy concerns present a significant hurdle. Collecting detailed travel data, especially with advanced technologies like mobile tracking, raises ethical concerns about personal privacy. Governments and organisations must handle data responsibly to maintain public trust while acquiring valuable insights.

The unpredictability of travel patterns also complicates data collection, especially with the rise of flexible, technology-driven options like ride-sharing and e-scooters. Today's travellers have more choices than ever, making travel decisions highly adaptable and often changing. Additionally, unforeseen events—such as road closures, weather changes, or tourism influxes—can significantly alter travel behaviour. Capturing these spontaneous changes requires more real-time data collection, which is resource-intensive and logistically complex. Ensuring data accuracy and reliability remains a significant hurdle. While commonly used, self-reported surveys are prone to errors like recall bias or respondents' unwillingness to disclose certain details. Even technologically gathered data can contain inconsistencies due to incomplete data sets, technical issues, or limited sampling. These inaccuracies can distort findings and lead to transport solutions that do not fully address public needs.

Addressing these challenges requires robust data validation and multiple data sources to create a more accurate and holistic view of travel behaviour. The FS Research Centre – the FS Group's in house high-skill centre for advanced studies and research on mobility and relative issues – gets data for analysing travel behaviour from different data sources. One of them is the Audimob survey, a survey carried out by Isfort, the High Institute for Transport Education and Research. FS Research Centre has expanded its approach by integrating big data analysis through a partnership with This thesis consists of four main chapters. Chapter two, "State of the Art," covers various papers and studies related to travel behaviour and mobility. The chapter aims to provide a comprehensive overview of existing work in these fields, enhancing familiarity with the relevant research. Chapter three outlines the main objectives of the thesis, followed by the methodology used to achieve them.

The methodology chapter is divided into three parts: first, comparing different datasets to highlight their unique features and deepen the understanding of their respective contributions to analysing travel behaviour; second, conducting a SWOT analysis for each dataset to assess the strengths, weaknesses, opportunities, and threats of each; and third, visualising the data using Tableau software. In chapter four, the results from chapter three are analysed to explore the differences between big data (Vodafone) and the Audimob survey findings. Finally, chapter five serves as the conclusion, providing an analysis of the results obtained and discussing their implications.

2. State of the art

Understanding travel behaviour is crucial for urban planning, transport policy, and environmental sustainability. To this aim, analysing the methods used to study travel behaviour comprehensively is essential. This section delves into this by reviewing various articles and research studies focused on travel behaviour. This section scrutinises the various methodologies employed in these studies, ranging from traditional surveys and observational techniques to advanced data analytics and business intelligent (BI) models, specifically Tableau software. Examining these previous studies can give insights into their effectiveness, limitations, and potential for application in different contexts. In figure 1 the concepts covered in the state of art of this thesis are presented.

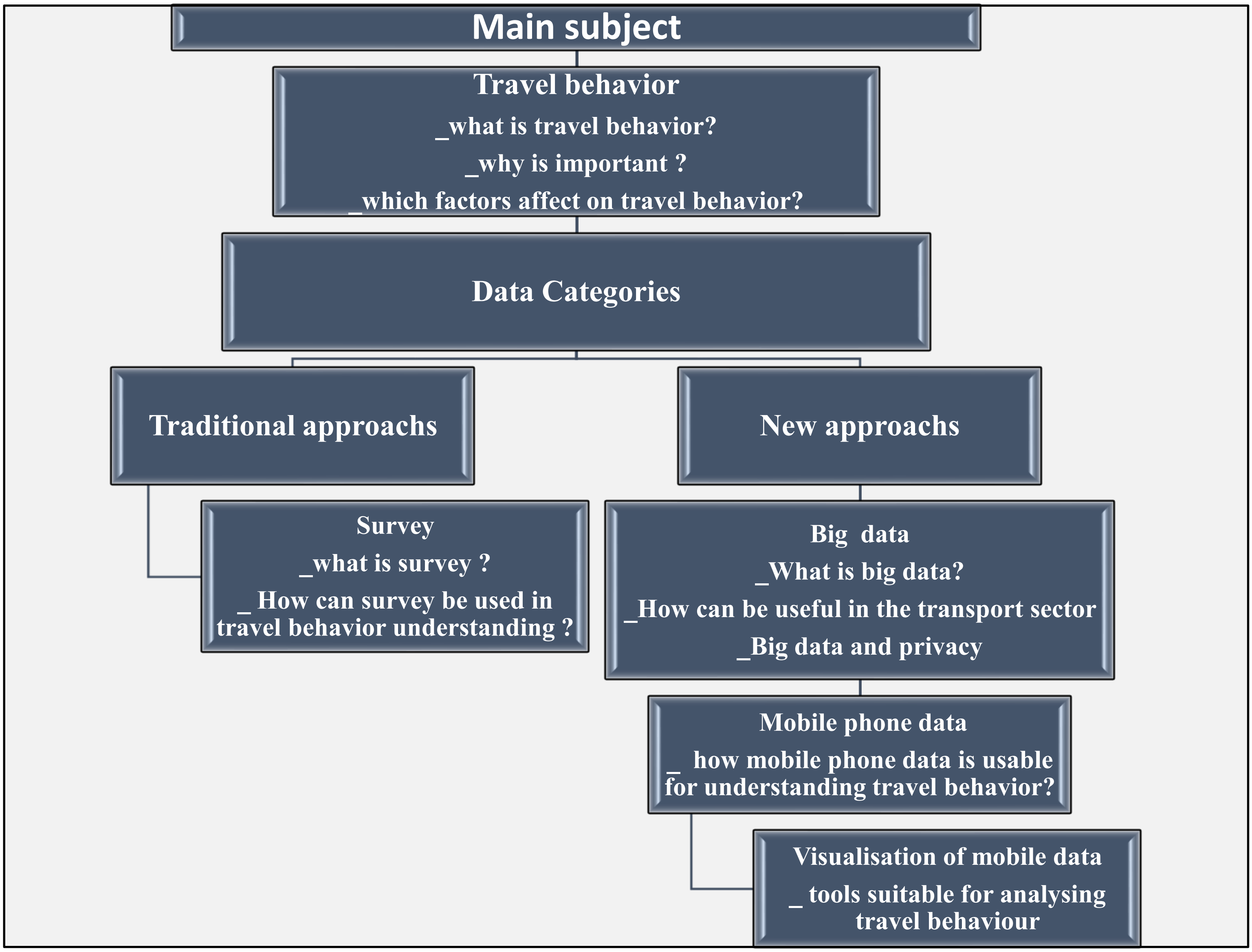


Figure 1 _ subjects of the state of arts

2.1 Travel Behaviour Studies

Travel behaviour studies require large amounts of data but often lack sufficient resources. The rise of mobile phone data, however, has created new possibilities for research in this field (Wang et al., 2018). Research on travel behaviour examines how people move outside their home locations for a variety of reasons. It examines factors such as how many trips people make, where they travel, what modes of transportation they use, who they travel with, which route they take, and how they decide on those decisions. (Axhausen and Zürich, 2007) .

The study of travel behaviour primarily focuses on analysing and modelling travel demand, drawing upon theories and analytical approaches from various scientific fields. The rapid economic growth in major cities within developing countries is generating significant employment opportunities, making these urban areas key destinations for those seeking better livelihoods. As a result, there is a steady migration from rural to urban areas, which places continuous pressure on these cities to meet the growing demands for social services, economic opportunities, and infrastructure. This trend of increasing urbanization and the accompanying socio-economic shifts are expected to impact travel behaviour and add layers of complexity to it (Subbarao. 2020). The primary aim of analysing travel behaviour is to understand how to enhance commuters' mobility while developing safe, environmentally friendly, and sustainable transportation systems. (Mwale et al., 2022). And also, achieving the ability to predict how people will respond to changes in their travel environments. (Subbarao et al, 2020).

Travel behaviour research uses a range of methods, including economic approaches like optimization, modelling, simulation, and forecasting, as well as behavioural and sociological techniques like surveys, interviews, and case studies. (DARIDO, 2012). Various methodological approaches have been employed to explore the relationship between travel behaviour and its influencing factors. Mathematical models serve as the main analytical tool, used to assess the elasticity of different service-level attributes and identify key factors influencing travel demand. (Mwale et al., 2022). The reviewed studies utilized a range of approaches, including ordinary least squares (OLS), Poisson regression (PR), geographically weighted regression models (GWRM), multinomial logit (MNL), nested logit (NL), structural equation models (SEM), and qualitative methods (QM)(Mwale et al., 2022). One of the traditional frameworks for analysing travel demand is the 'four-step model,' which examines four dimensions of travel: trip generation, trip distribution, mode choice, and trip assignment. However, a key limitation of this model is its inability to represent the underlying motivations for travel accurately. It also fails to adequately reflect the derived nature of travel demand. To address these shortcomings, more advanced frameworks such as activity-based models and tour-based models have been developed. These newer models better capture travel demand as a derived demand, emphasizing the motivations behind travel. Additionally, they shift focus from “mobility,” which centres on expanding transport services and infrastructure, to “accessibility,” which prioritizes enabling access to destinations. (Mwale et al., 2022).

2.1.1 Factors Affecting Travel Behaviour

A trip chain refers to a sequence of linked travel purposes that involve making several stops. Person-trip surveys conducted in developing countries reveal a strong inclination for car and motorcycle trip chains, as they efficiently address various travel needs such as household errands, shopping, personal business, social activities, and recreation. This indicates that trip chaining has become a necessary aspect of daily commuting, particularly for households that own vehicles in these regions. In this context. Travel decisions, including mode choice, trip frequency, and route selection, are influenced by various factors such as location, socio-demographics, psychology, and culture. The probability of owning and using personal vehicles tends to rise with factors like income, household size, education, and residential area. Additionally, private cars are often viewed as a status symbol and a sign of wealth. (Dissanayake et al,.2002).

Policymakers, investors, and planners need to understand how land use and travel demand interact, as both transportation infrastructure and real estate development require significant financial investment. Leck (Leck,2006) examines the impact of urban form on travel behaviour through meta-analysis, revealing several key insights. Such as emphasizing the significant impact of population density and employment density on travel behaviour, even after accounting for social and demographic factors such as income and age, and, in addition, the impact of land use diversity on travel behaviour.. Urban areas with a mix of offices, shops, and public facilities greatly influence mode choice, leading residents in these diverse environments to prefer commuting by transit or slower modes. (Mwale et al., 2022).

To understand travel behaviour concerning decision-making and travel choices, it is essential to examine factors such as travel frequency, trip purpose, mode of transport, and chosen routes. It is also important to understand the effects of latent factors such as safety, convenience, and comfort on travel demand (Mokhtarian and Cao, 2008). In addition, it is important to consider that research on travel behaviour is intricately linked to the analysis and modelling of travel demand, drawing from various theoretical and methodological frameworks across different disciplines(Firdausi et al., 2023). Figure 2 summarizes the most important factors that have effects on travel behaviour in different authors' articles according to their results.

No	Author	Variable
1	Roya Etminani-Ghasrodashti, Mahyar Ardeshiri	Lifestyle and land use
2	sunta di vaio,, ana lopezojeda´, casiano manrique-de-lara-penate, lourdes trujillo	Characteristics of tourists, use of modes of transport, frequency of visits, travel behaviour of tourists, level of satisfaction.
3	Xinyi Tang, Dianhai Wang, Yilin Sun, Mengwei Chen, E. Owen D. ,Waygood	Age, gender, type of companion, car ownership, length of vacation and purpose of travel
4	Chanittha Chansuk, Tosporn Arreeras, Chalailuk Chiangboon, Canteen Phonmakham, Nattawat Chotikool, Ratthanan Buddee, Sorrawich Pumjampa, Thanmit Yanasoi ,Saharat Arreeras	Gender, age, education, occupation, income, vehicle ownership, number of family members, status and residence
5	Berhanu Esubalew Bayih,, Apar Singh	The level of tourist satisfaction, the frequency of domestic tourist visits, the motivation of domestic tourists to visit again and the willingness of tourists to recommend.

6	Ali Mahdi , Jamil Hamadneh , Domokos Esztergár-Kiss	Socio-demographics, travel time, travel expenses, age, gender, income and car ownership
7	Ricky Yao Nutsugbodo, Edem Kwesi Amenumey, Collins Adjei Mensahc	Factors of affordability, accessibility, mode availability, sociodemographic safety and comfort of tourists and mode preferences
8	Bartosz Exchange, Markus Mailer, Kay W. Axhausen	Travel time, travel expenses, the composition of the travel group, purpose of the trip, fitness level of respondents, their knowledge of long-distance travel to the destination and mobility options at the destination and selected weather elements.

Figure 2_Factors influence on travel behaviour(source :Firdausi et al., 2023, p. 2)

In their paper, Shariff and Shah examine the factors that influence travel behaviour and propose practical solutions for reducing traffic congestion(Shariff et al., 2008). The researchers performed an in-depth review of existing literature and data to identify key factors influencing travel behaviour, including location, land use, transportation availability, costs, traffic congestion, environmental impacts, and social factors. They then analysed current transportation planning strategies and suggested improvements like transit-oriented development, telecommuting, staggered working hours, new urbanism, and better public transport services.

2.1.2 Travel behaviour-mining by technology

As previously mentioned, there are several ways of extracting necessary data to understand traveller behaviour. Servizi et al. (2021), examine the limitations and opportunities in using smartphone-based travel surveys (SBTS) for transport behaviour analysis. It focuses on identifying the main machine-learning algorithms used for processing data from such surveys and understanding the interplay between these methods and the accuracy of the data collected (ground truth). The methodologies discussed include the use of various machine-learning techniques for the automatic generation of travel diaries, mode detection, and activity inference. The review analyses different datasets collected globally, considering their representativeness and the quality of ground truth. The process involves a detailed examination of technologies such as assisted global positioning systems (AGPS), inertial navigation systems (INS), geographic information systems (GIS) and Internet-of-Things (IoT), which support data collection and validation in SBTS.

The findings reveal significant physical limitations of smartphones as data collection devices, methodological challenges in generating accurate travel diaries and issues with data validation through user interaction. Despite these challenges, the review highlights the potential of SBTS to reduce the costs associated with traditional travel surveys and enhance the resolution of transport behaviour data. The article underscores the importance of standardization in performance evaluation and the need for comprehensive datasets to improve the reliability of machine-learning methods in this field. Overall, the review concludes that while SBTS has not yet replaced traditional travel surveys, they offer promising tools for understanding transport behaviour through advanced analytics and machine learning (Servizi et al., 2021).

Graells-Garrido et al. (2023) aim to create a flexible Data Fusion (DF) framework to improve area-level estimates of travel modes in urban transport systems. The framework combines traditional data sources, like travel surveys and manual counts, with new digital data, such as mobile network data, to give a clearer picture of urban transport patterns. The objective is to overcome the limitations of traditional data, which are often expensive, infrequent and subject to data latency, by leveraging the strengths of real-time and granular mobile data to support more resilient and sustainable urban transport systems. The methodology proposed in this study involves three main stages:

- First, the Initial modal split estimation: this stage uses official sources and domain knowledge to build a candidate modal split for all administrative units. This involves data from the National Institute of Statistics (INE) on smart card transactions and car permits, population projections and expert assumptions on changes in transportation modes due to factors like ride-hailing app usage and social unrest.

- Second, Data fusion, using the matrix factorization, stage generates a latent representation of multiple datasets, combining traditional and digital sources related to mobility patterns, socio-demographic characteristics and urban infrastructure. Matrix factorization techniques, particularly Non-negative Matrix Factorization (NMF), are used to decompose and reconstruct matrices representing different data relationships.

- Third, revised modal split estimation: A revised modal split is generated using the information from the latent representations obtained earlier. This new estimation incorporates all available data to enhance the initial mode split calculation.

The results of this study demonstrate the effectiveness of the Data Fusion framework in providing more accurate and updated estimates of urban transport mode splits. By integrating diverse data sources, the framework can capture complex mobility patterns that are not apparent from individual datasets alone. The case study focus of the paper (in Santiago, Chile) shows significant improvements in estimating mode splits by incorporating real-time data from mobile networks and adjusting for socio-demographic and urban infrastructure factors. This integrated approach offers a robust tool for urban transport planning, capable of adapting to rapid changes and providing valuable insights for sustainable mobility interventions (Graells-Garrido et al., 2023).

2.2 Data collection methods for understanding travel behaviour

This stage is crucial for analysing the data to answer research questions. Effective data collection helps ensure accurate results by reducing potential errors in the research process (Taherdoost,2021) but, before understanding and finding the suitable method is important to know the type of the data (Kabir, 2016). Qualitative data consists of non-numerical information. This type of data helps answer "how" and "why" questions in research and is often collected through unstructured methods like interviews. On the other hand, quantitative data refers to numerical information that is calculated and analysed mathematically (Taherdoost,2021).

In research, there are two major types of data collection, primary and secondary. Primary data is first-hand information collected by researchers for a specific study like surveys, interviews, and questionnaires, this type of data offers high validity and reliability. Secondary data, on the other hand, is information that has already been collected for different purposes like journals, and websites but may lack the specificity and accuracy that some research requires (Taherdoost,2021). Travel behaviour research has at least five distinct data collection methods. The first, and most prevalent, method has evolved from the traditional household-based questionnaire and diary, which records daily activities and trips. This approach has been widely utilised to estimate revealed preference discrete-choice models and the majority of activity-based models. It has become the national standard for monitoring behavioural changes through a repeated cross-sectional methodology. Modern travel-behaviour surveys are increasingly interactive, leveraging internet-based scheduling and geospatial technologies to create complex, dynamic interviews. The use of smartphones has notably enhanced the richness of the collected data.

The second method is the panel survey, designed to provide deeper insights into changes in behaviour over time. The third approach involves asking respondents hypothetical questions, varying the attributes of choices to explore options that cannot be tested in real life.The fourth method is real-life interactive experimentation, which aims to influence behaviour by giving participants feedback on their past behaviour. Finally, the fifth approach employs qualitative research methods, applicable to any behavioural issue. Additionally, online sources like social media (e.g., Facebook, Foursquare, Twitter) are used to gather data on travel satisfaction, weekly activities, traditional origin-destination patterns, and place-specific sentiments. (Goulias, 2016) In this thesis, the first approach will be more prominent, highlighting traditional surveys and big data extracted from mobile phones.

2.2.1 Survey-based data collection

Household travel surveys (HTS) have been a key tool for gathering essential data for regional transportation planning. They have proven highly effective in collecting consistent and high-quality information on passenger travel patterns, providing valuable insights for transportation planning efforts (Miller, et al,.2018).

In the past, it was assumed that travel behaviours remained fairly consistent over time, leading to data collection intervals every eight to fifteen years. (Lawson, et al 2023). Regional and state transportation planners depend on detailed travel information to guide their planning efforts. This information encompasses who is travelling, their travel times, destinations, purposes, and modes of transportation. Household travel surveys (HTS) are instrumental in gathering this data, allowing planners to gain insights into travel patterns within a given area. While travel demand models often utilize HTS data for development and calibration, planners can also use the information to generate descriptive statistics about regional travel and analyse trends over time. Users of household travel survey (HTS) data can extract a range of valuable metrics, such as trip rates per household and individual, travel mode distribution, vehicle occupancy rates, and geographic travel patterns including origin-destination matrices, all enriched with detailed demographic information. (City of Seattle & Puget Sound Regional Council, 2019)

2.2.2 Big Data based data collection

Understanding travel behaviour and forecasting future travel demand has long been crucial for researchers and planners aiming to develop precise transportation projects. This necessity drives the continuous collection of data. Traditionally, data collection has been conducted actively, such as through travel surveys where participants self-report their activities and travel via paper forms, web interfaces, or phone interviews. Another common method involves combining travel surveys with GPS loggers, requiring participants to both fill out questionnaires and carry GPS devices. However, the rapid advancement and widespread adoption of mobile technologies have facilitated the collection of vast amounts of passive data, or big data, leading to a significant increase in studies focused on human movement patterns. (Chen et al., 2016)

Mobile big data presents an opportunity to generate significant social impact, aligning with the United Nations Sustainable Development Goals.(GSM Association, 2019). Big data is increasingly recognized for its potential to enhance urban environments by uncovering and explaining patterns in urban development and dynamics. This capability allows city planners and residents to make more informed decisions and improve overall urban living conditions. In the realm of transportation, big data has advanced the development of intelligent transportation systems by offering deeper insights into travel patterns, including the timing, locations, and modes of transportation used by individuals (Wang et al., 2018). Over time, the architectures of big data have evolved significantly, with innovations in large-scale data processing leading to the development of alternatives like the Lambda and Kappa architectures (Kolodziej , et al 2019). However, one of the primary challenges in big data remains the management and storage of vast amounts of data (Nayak et al., 2023). To address these challenges, several important platforms have been developed over the past decades that support Kappa or Lambda architectures and can implement a variety of tools and technologies for different system components, such as Hadoop, Spark, Kafka, and Samza (Kolodziej, et al ,. 2019). Working with big data involves the potential to face various challenges, including data collection, storage, analysis, search, sharing, transfer, visualization, querying, updating, privacy, and management of data sources (Arunkumar , et al, 2021).

In the book "Ethics of Big Data: Balancing Risk and Innovation" (Kord Davis,2012), the main subject is to delve into the ethical dimensions of big data, aiming to balance the innovation benefits with the risks associated with increased information sharing. The authors intend to equip organisations with tools and methodologies to conduct explicit ethical inquiries, thereby improving ethical practices in the context of big data. The book underscores the complexity of ethical issues in big data, emphasizing the necessity for transparent and explicit discussions on these topics. It identifies key ethical concerns, such as privacy and data ownership, and suggests that organisations develop the ability to engage in ethical inquiry and implement policies to address these challenges (Davis et al, 2012).

The GSMA (Groupe Special Mobile Association) It is an industry association that advocates for the interests of mobile network operators globally. The primary goal of the GSMA report titled "Mobile Privacy and Big Data Analytics" is to explore how big data analytics can be utilized responsibly in the mobile industry while maintaining and enhancing consumer trust and privacy. The report emphasizes the potential benefits of big data analytics in creating a connected world through the Internet of Things (IoT). The overarching aim is to ensure that the economic and societal benefits of big data are realized in a manner that respects established privacy principles. The methodology employed in this report involves a detailed examination of various scenarios where big data analytics is applied. The report concludes that big data analytics can significantly enhance digital life by providing actionable insights that benefit society. However, to achieve these benefits, adopting good data privacy practices and considering the legal, ethical, and practical implications of data usage is crucial. The report suggests implementing privacy-by-design principles, conducting data privacy impact assessments, and ensuring transparency and accountability in data handling processes. By doing so, organisations can foster an environment of trust and maximize the positive impact of big data analytics (GSMA, 2017).

Big data allows the integration of diverse data types, including spatial, temporal, and contextual information, to create a comprehensive picture of travel behaviour. Advanced analytical techniques such as machine learning and data mining can uncover patterns and trends that are not immediately obvious, aiding in the development of predictive models and decision-making tools. These models can simulate the effects of different transportation scenarios, helping planners and policymakers to design more efficient and user-centric transportation systems. By leveraging big data, the field of travel behaviour research can move towards more adaptive, responsive, and sustainable urban mobility solutions (He, 2017). The article "Applying Mobile Phone Data to Travel Behaviour Research: A Literature Review" provides an in-depth analysis of how mobile phone data has been utilised in understanding and predicting travel behaviour. The paper's main goal is to examine the current state of research in this area, highlighting the progress made so far, the potential benefits, and the challenges that must be addressed. The methodology involves a comprehensive literature review of existing studies that have employed mobile phone data to analyse various aspects of travel behaviour. This includes identifying travel patterns, exploring influencing factors, and modelling and predicting travel behaviour. The paper discusses the different types of mobile phone data, such as data from cellular networks, GPS, and Wi-Fi, and their applications in travel behaviour research. Despite the advancements, the authors emphasize that there is still considerable potential for mobile phone data to further enhance the understanding of human mobility, provided the challenges related to data privacy, accuracy, and integration with other data sources are effectively managed. Compared to traditional travel surveys, mobile phone data offers unique advantages, Mobile phone data offers unique advantages, such as covering large populations and areas, providing continuous data collection, and offering precise information on location and movement. While previous studies have identified daily, weekly, and seasonal travel patterns, monthly and seasonal patterns have been less studied due to the limitations of traditional surveys. The ability to track data continuously through mobile phones can fill this gap, enabling a deeper understanding of longer-term travel patterns. (Wang, et al,. 2018).

Peng et al. (2021) aim to address the limitations of traditional travel surveys and rule-based methods by leveraging long-period mobile phone call data (CDRs) to identify urban travel patterns and modes. The study focuses on Wuhan, a major city in China, using data from March 2016, which includes over one billion records from 6.78 million users, representing about 60% of the city's population. The methodology involves several key steps: pre-processing the CDR data to aggregate it over a 31-day period to mitigate sparsity issues. The results indicate that the proposed method can effectively recognize travel behaviour when users have a call frequency of over 30 times per month, achieving high recognition accuracy. This approach provides a comprehensive understanding of urban travel patterns, which can significantly aid urban planners in improving public transportation systems, road conditions, and shared travel options. (Peng et al., 2021)

Essadeq and Janik (2021) explores the application of mobile phone data to enhance transport modelling in France. The primary purpose of the study was to evaluate the potential of mobile phone tracking to capture long-distance trips and improve the accuracy of transport demand models. This approach aims to offer a cost-efficient alternative to traditional data collection methods, leveraging the widespread use of mobile phones to gather extensive mobility data. In their research, the authors employed data from a French mobile telecommunication company, focusing on the year 2018. They developed and calibrated a method to process this data, identifying trips, differentiating them from stationary activities, and inferring transport modes. The paper compared the results derived from mobile phone data with traditional data sources such as SNCF Réseau traffic forecasts, air traffic data, and rail traffic data. The findings demonstrated that mobile phone data could provide reliable trip information, thus supporting the hypothesis that such data can significantly enhance passenger transport planning. The authors concluded that mobile phone data, while complex and requiring careful interpretation, can significantly enhance transport modelling (Essadeq et al,.2021).

Deutsch et a.l(2012), aim to explore the potential of smartphones equipped with various sensors to collect travel behaviour data. The methodology discussed in the paper includes an overview of possible sensors in smartphones, such as GPS and accelerometers, and their application in data collection. By leveraging the sensors in smartphones, researchers can collect data passively with minimal respondent burden or interactively through periodic prompts. This approach not only reduces survey costs but also increases the likelihood of data collection by utilizing devices that respondents are already familiar with and likely to carry. The results and outcomes anticipated from this research include improved data accuracy and richness, as well as higher response rates due to reduced participant burden. The ability to collect detailed geographic and temporal data enables more nuanced analysis of travel behaviour. The paper concludes that integrating smartphone technology into travel behaviour research offers substantial benefits, providing a more detailed and less intrusive means of data collection compared to traditional methods (Deutsch et al., 2012).

in the paper "Mobile Phone Data Analytics Against the COVID-19 Epidemics in Italy" Producers and collectors of this article uses mobile phone data to analyse human mobility patterns during the COVID-19 pandemic. The methodology involves using Call Detail Records (CDRs) and Extended Detail Records (XDRs) provided by the mobile operator WINDTRE. The analysis covers the period from February 3rd, 2020, to March 28th, 2020, focusing on various provinces. The results indicate a significant reduction in mobility diversity during the national lockdown (WINDTRE, 2020).Yan Sun et al.(2021) provide a comprehensive synthesis of the current state of research at the intersection of mobile technology and transport behaviour. It seeks to identify successful implementations of mobile technology in transport behaviour studies, propose an integrated research model, and outline a future research agenda. By reviewing scholarly sources from databases such as Web of Science, JSTOR, and SAGE, the study categorizes the applications of mobile technology in transport into three main areas: 1) smartphone apps for sustainable transport and travel planning, 2) cellular signalling data for analytical models, 3) the use of CDRs, Wi-Fi, and GPS data for transport behaviour studies. The methods used in this article follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines. The study applied qualitative analysis to these selected articles, examining the system designs and research topics of mobile technologies used in transport studies. The proposed integrated research model reflects how previous studies have achieved behavioural outcomes, emphasizing the influence of app design elements on mobility and the interaction between technology acceptance and transport behaviours. The results indicate that mobile technology can significantly enhance our understanding of transport behaviours, providing real-time data and insights that inform better decision-making and policy development in transport planning. The proposed integrated research model highlights how design elements of mobile apps influence user mobility patterns understanding and emphasizes the necessity for continued investigation into the synergistic effects of technology and transport behaviours (Sun, Liu, & Zhang, 2021).

Nils Breyer's,(2019) enhance the understanding and utilization of cellular network data for traffic analysis. The primary goal is to demonstrate the potential and limitations of this data source in providing comprehensive insights into travel patterns and traffic demand, which are crucial for effective traffic planning and management. The thesis is particularly significant as traditional methods of gathering travel information, such as travel surveys and traffic counts, are expensive and provide limited insights. In contrast, cellular network data offers a more comprehensive and potentially real-time view of travel patterns, making it a promising alternative for traffic analysis.

The methodology presented in the thesis involves a data-driven traffic estimation pipeline that consists of several key steps;

1) trip extraction; which identifies periods of movement and standstill in raw cellular network data to define trips based on the start and end positions marked by cell antennas to identify the modes of travel used for the detected trips.

2) travel demand estimation; where the detected trips are used to estimate overall travel demand, which is then scaled to represent the entire population.

3) route estimation which estimates possible routes taken for each origin-destination pair based on the identified trips.

4) flow distribution; assigns the estimated travel demand to the transportation network, determining the flow on each network link, akin to the network assignment step in traditional traffic models.

The results present three algorithms for trip extraction from cellular network data and analyse the effects of different types of cellular network data on this process. Breyer's research highlights the substantial potential of cellular network data for improving traffic analysis, despite the inherent challenges. It underscores the necessity for new processing methods and evaluation techniques to effectively leverage this data source for both long-term traffic planning and short-term traffic management. (Breyer, 2019).

2.3 Data visualisation tools

The term "intelligence" on which Business Intelligence (BI) is based, typically refers to the gathering, analysis, and dissemination of information. It can also pertain to secret interventions in the political or economic matters of other countries, an activity often referred to as "covert action." (Amoako, 2013). The concept of "intelligence" has been employed by artificial intelligence researchers since the 1950s. It became popular in the business and IT sectors in the 1990s. By the late 2000s, the term "business analytics" was introduced to highlight the analytical aspect of business intelligence (Davenport 2006). Since the early 2000s, the growth of the Internet and the Web has created new opportunities for data collection and analytical research, especially with HTTP-based Web 1.0 systems. exemplified by Google and Yahoo and e-commerce platforms like Amazon and eBay, enabled organisations to establish an online presence and directly engage with customers. Later, with the advent of with the advent of Web 2.0, businesses were able to quickly collect large amounts of real-time feedback and opinions from a wide range of customers. Over the past twenty years, business intelligence and analytics (BI&A), along with big data analytics, have grown in importance in both academic and business communities(Chen, Chiang, & Storey, 2012).

Data visualisation is the pictorial representation of data, designed to enhance the user's understanding. Today, data visualisation plays a crucial role in conveying the meaning of data by presenting it visually through charts, maps, or graphs. This visual approach makes the data more intuitive and easier for the human mind to comprehend, helping users identify patterns and trends within large datasets. Presenting information in an aesthetically pleasing way can be considered an art form (Siddiqui, 2021). Big data visualisation typically refers to visualizing a vast number of data points (items and attributes) within a large space (Zheng, 2023). In the modern business intelligence landscape, the rise of Big Data Analytics has become a fundamental element for organisational success (Bhara et al., 2023). To analyse data effectively within the context of business intelligence, having the right tools is essential to achieving the most accurate results. Data visualisation tools can be categorised into three main types: spreadsheets, data visualisation software, and programming libraries.

- Spreadsheets, such as Microsoft Excel and Google Sheets, are widely used across various fields for data visualisation.
- Data visualisation software is specifically designed for visualising and analysing data, with tools like Tableau, QlikView, and Power BI offering advanced capabilities such as interactive dashboards, heat maps, and network diagrams.
- Programming libraries such as Matplotlib, ggplot2, and D3.js provide a more flexible and customisable way to visualize data, but they demand a greater level of technical skill. (Srivastava, 2023).

Devadharshini L. and Boomika D.M. (2022), evaluate and compare the effectiveness of three data visualisation tools: Tableau, Power BI, and R programming. The authors aim to analyse the strengths and weaknesses of each tool in terms of speed, usability, and cost, thereby providing insights into their practical applications in the field of data science. This comparison is intended to guide users in selecting the most appropriate tool for their specific needs in data visualisation and analysis. The methodology of the study involves a comparative analysis based on specific parameters such as speed, ease of use, and cost. The authors have reviewed and assessed each tool by exploring their working mechanisms, features, and functionalities. This involved:

- Tableau: They explored Tableau's capabilities in connecting and extracting data from various sources, its dashboard creation process, and its user interface, which facilitates non-technical users to create interactive graphics.
- Power BI: They investigated how Power BI integrates with various data sources, its step-by-step process for creating business intelligence dashboards, and the specific features that enable data manipulation and visualization.
- R Programming: They examined R's application in statistical computation and graphics, its workflow for data analysis, and the process of data importing and manipulation using R packages.

The article highlights that the choice of a data visualisation tool depends on the specific needs of the user, such as the speed of processing, ease of use, and cost considerations. Tableau is recommended for extensive business data applications, Power BI for cost-effective and user-friendly business analytics, and R programming for in-depth statistical analysis in research environments (Devadharshini & Boomika, 2022).

Partha (2023) focuses on the in-depth comparative analysis of two leading data visualisation tools – Power BI and Tableau. The study evaluates different options for users, from free versions to enterprise-level packages, in terms of pricing structure. The tool compares features such as ease of use, customization options, and the ability to create interactive dashboards, as well as the user interface and visualisation capabilities. Furthermore, the research analyses the integration options for each tool, such as how well they integrate with other applications and data sources, and it assesses the data modelling and ETL (Extract, Transform, Load) capabilities that allow users to prepare and transform their data. As a result, Powered BI offers a wide range of pricing options and a strong integration with the Microsoft ecosystem, making it a versatile option for businesses already using Microsoft products. Tableau is a powerful tool for users who prioritise advanced visualisations and data analysis because of its ease of use and extensive exploration capabilities(Parthe, 2023).

3. Objectives and Methodology

The objectives of this thesis focus on analysing the specific data sources provided by the FSI Group (key partner of this research), which are actively used by the company in their operations to understand the mobility patterns of people involved. These data sources include Vodafone big data, representing mobile network data, and the Isfort Audimob database, which offers insights into travel behaviours through surveys.

The *primary objective* of this study is to delve deeply into the potential of Vodafone's mobile data sources (big data) as a revolutionary tool for analysing transport systems that can be a way to work on information that can transform our understanding of travel behaviour. By exploring these data sources, the study aims to uncover how they can provide information about various aspects of mobility patterns. This includes tracking travel patterns, understanding the flow, and examining the effects of various dimensions on travel patterns. The study also seeks to address the challenges and limitations associated with using mobile data for mobility analysis. This exploration will pave the way for innovative approaches to transport planning, ultimately leading to improved mobility and quality of life for all citizens.

The *second key objective* of this study is to develop a comprehensive method for comparing mobile data and traditional survey data within the context of travel behaviour. It is crucial to identify an approach that allows for a thorough comparison, one that fully captures all relevant Variables of both data sources. This will enable a complete understanding of their differences and strengths.

The *third key objective* aims to provide an in-depth understanding of the weaknesses and strengths of both mobile data and traditional survey data. This involves a critical analysis of the accuracy, reliability, and applicability of each data type in different transport scenarios. By identifying these aspects, the research seeks to determine methods and features for collecting and analysing mobility data, ultimately improving the accuracy and reliability of travel behaviour studies.

After reviewing the literature and defining the thesis goals according to the data provided by FSI, the appropriate methodology is followed to address the objectives. Figure 3 summarises the methodological approach developed to achieve the research objectives using a mixed-methods approach, integrating both qualitative and quantitative techniques, providing two steps: data collection and data analysis.

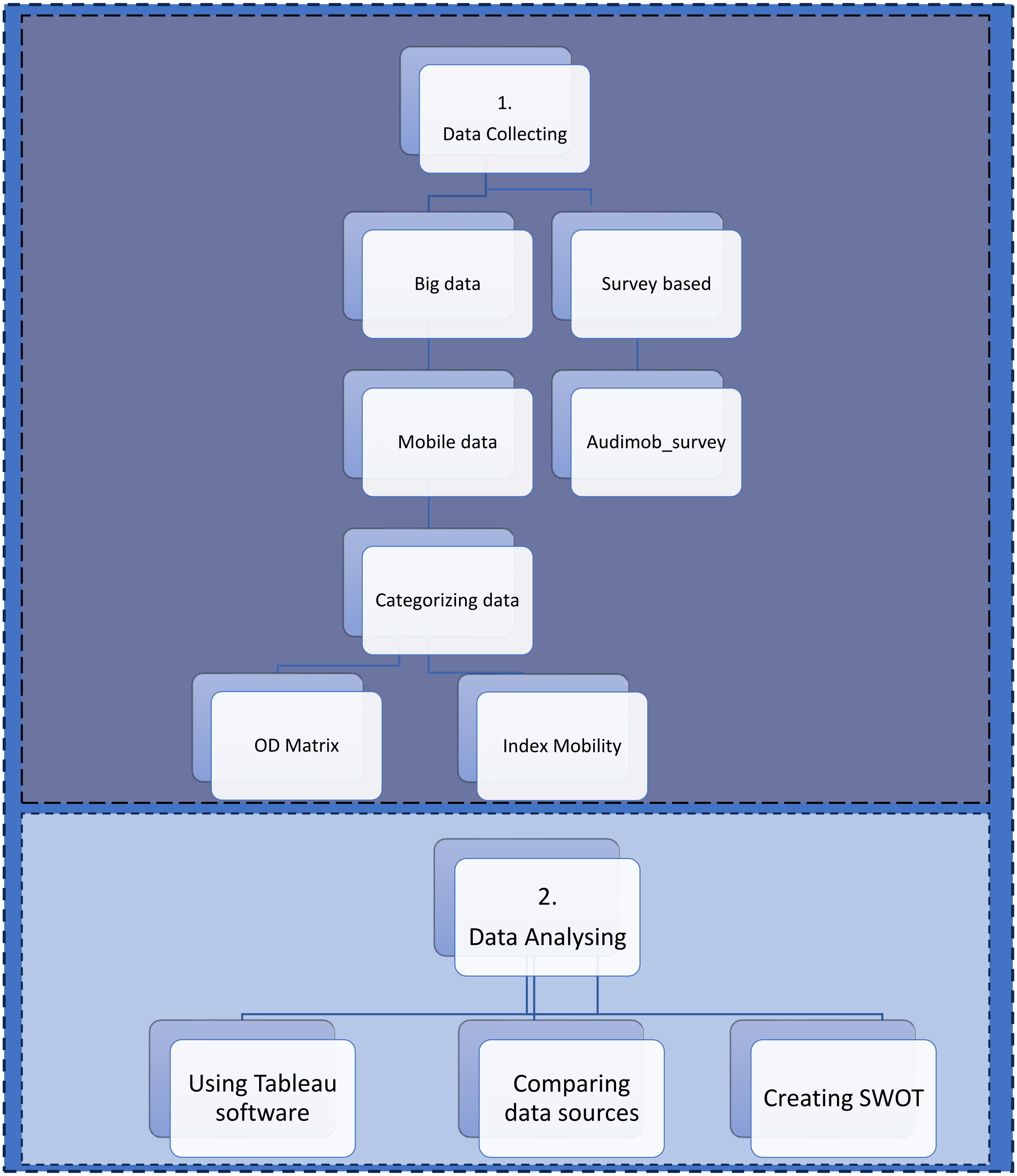


Figure 3_ flow chart of the process

3.1 Data collection

With the increasing number of trips and a growing population, FSI has decided to leverage big data to enhance the efficiency of understanding travel patterns by purchasing travel data over consecutive years.

In this thesis, primary data collection is the main one that is used and these methods can be tailored to the specific objectives and context of the study. It is essential to plan and follow data collection protocols carefully to ensure that the results are reliable and valid. FSI Group acquires pre-processed and categorised mobile data from Vodafone on an annual or monthly basis, according to specific needs. Additionally, ISFORT provides FSI Group with data from the Audimob survey, further enhancing the dataset with in-depth insights into travel behaviour.

3.1.1 Big data: Vodafone data

Vodafone has a substantial presence in the Italian market, evidenced by the number of active SIM cards, which ranges between 22 and 23 million. The network infrastructure includes over 200,000 mobile network cells, ensuring robust connectivity and service availability across a wide geographic area. Vodafone's population coverage stands at nearly 99% for 4G and close to 100% for 2G, trying to cover both urban and rural areas. Vodafone's network infrastructure is designed to ensure high spatial granularity, especially in densely populated areas where cell diameter can be reduced to as little as 100 metres. Regularly installing dedicated coverage cells enhances spatial precision within specific perimeters, such as indoor spaces like museums, theatres, and stores.

Vodafone handles big volumes of daily network activity, with up to 30 billion positional data points collected daily from interactions with approximately 23 million SIM cards. This translates to around 2,000 events per SIM card per day. Vodafone's sampling covers approximately 30% of the market, and over 40% of foreigners are present in Italy. The network collects positional data at a very high frequency, multiple times per minute, allowing user presence and movement tracking. This high-frequency sampling is crucial for detailed spatial analysis, far exceeding the approximately 10 daily events recorded by traditional Circuit Switch Data (CSD) methods. Vodafone maintains an extensive data history, storing information for up to 18 months. This comprehensive temporal coverage allows for trend analysis over complete seasonal cycles and helps identify year-on-year differences and long-term trends (Vodafone Italia S.p.A., 2019).

The data's spatial granularity is further enhanced by additional data sources such as applications on devices, providing geolocation accuracy to within a few metres using AGPS data from approximately 10 million devices. Vodafone Analytics transforms anonymised, aggregated, and projected data from Vodafone's 4G and 4.5G networks into information for business needs and the management of local and central administrations. The service provides insights into presence, movement, origin, and behavioural profiles of both the resident and foreign population in Italy, leveraging big data technologies and artificial intelligence. Adhering to GDPR regulations, Vodafone ensures that data used in Vodafone Analytics is anonymised and aggregated irreversibly, protecting customer privacy. Customers are informed through Vodafone's privacy notice about the usage of their data in anonymised and aggregated forms. The analysis results are extrapolated to reflect the entire population, not just Vodafone users, using proprietary algorithms and expansion techniques.

Vodafone Analytics offers a high degree of flexibility and customisation, providing KPI personalization based on the analysis framework and client requirements. The service can be beneficial for public administrations and businesses, enabling data-driven decision-making and improving service quality through easy-to-use and accessible data insights. Data are made accessible through interactive graphical interfaces, massive file transfers, or standard APIs, facilitating the integration into existing systems. The Vodafone Analytics platform processes tens of billions of raw mobile network records daily, analysing data using proprietary algorithms. The platform's architecture is thought to be scalable and modular, trying to support custom developments. The entire data processing workflow complies with privacy regulations, incorporating mandatory security measures and minimising data protection risks. Vodafone Analytics adheres to several quality and security standards, including:

- UNI EN ISO 9001:2015 – Quality management systems
- ISO/IEC 27000-1:2011 – Information security management
- OHSAS 18001 – Occupational health and safety management systems

These standards should ensure that the services provided are reliable and secure, according Vodafone declarations about their analytics offerings (Vodafone Italia S.p.A., 2019).

Vodafone data are analysed by two datasets, produced for FSI, which aggregate raw data using different approaches: first OD matrices and second Mobility indexes.

OD Matrices dataset

During the analysis of the OD matrices provided by Vodafone, FSI initially hypothesised that a trip would be segmented if there was a break in movement of four hours or more. However, further evaluation revealed that this approach led to the loss of many trips, as numerous shorter trips were not captured due to the extended time threshold. To address this, FSI refined the methodology by introducing a one-hour threshold. This threshold defines the period a person remains stationary before any subsequent movement is considered part of a new trip. However, this is not a fixed or universally applicable measure because its optimal value can vary depending on the characteristics of the data and the specific context of the analysis. For example, in this thesis, while a one-hour threshold was effective for detecting shorter trips, a four-hour threshold was ultimately chosen as it proved more effective in accurately capturing the overall trip behaviour under the given circumstances. The Italian territory is divided into 3,000 zones (figure *4*), with trips being recorded from one zone (O) to another (D), although mobility within each individual zone is not tracked. The zones are sized according to the level of municipalities, using a higher level of detail for larger and more populated cities, and grouping together smaller and less populated ones.

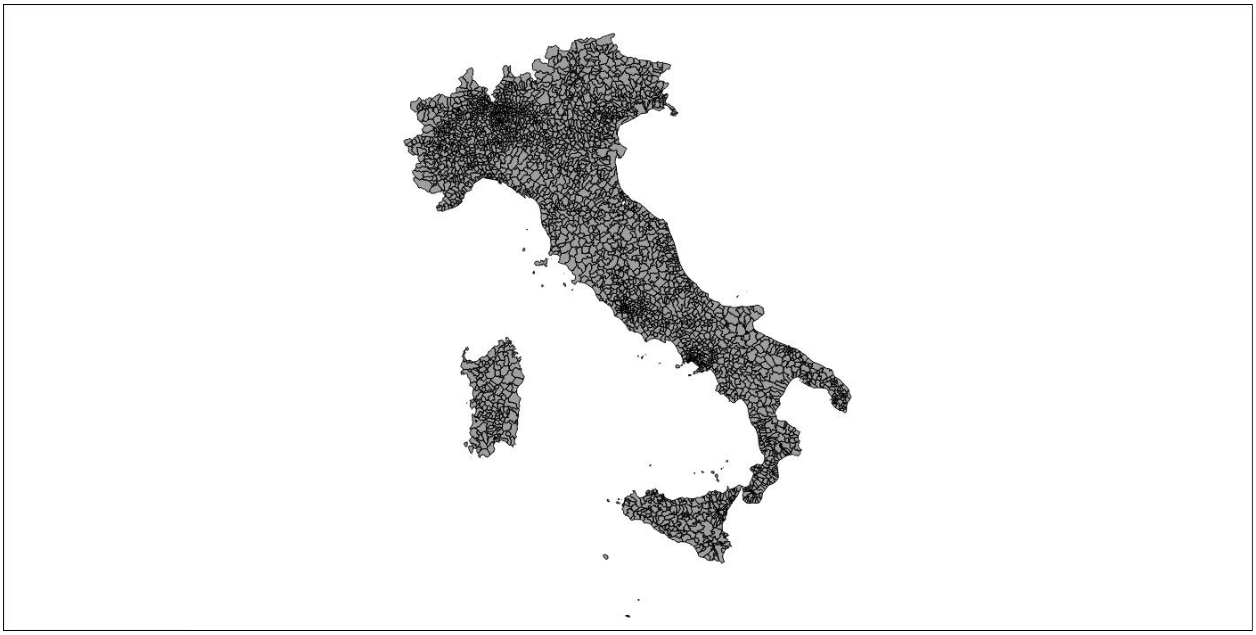


Figure 4_Map of Italy divided by zones

Mobility Indexes

Mobility indexes are another dataset used from data provided by Vodafone. In this dataset, the day is divided into one-hour intervals. At each interval, the travelled points are recorded based on information collected from the antennas. If movement occurs between different locations and connections are made to various antennas during the hour, the position is calculated based on the centre point between those antennas, known as the central core. The weighted average distance from the antennas and the central core is then calculated, based on time. If that distance is less than 800 metres the user is considered “stationary” for that hour, otherwise, he’s considered “travelling”. As a result, each hour is analysed by considering the positions occupied within that timeframe. Therefore, a trip is defined as moving from the central core of a stationary hour to the central core of the next stationary hour. The distance travelled is determined by connecting the central core of the two stationary hours, passing through all central cores for “travelling” hours in between. Figure 5 shows a graphical model illustrating the functioning of mobility indexes function in 2023.

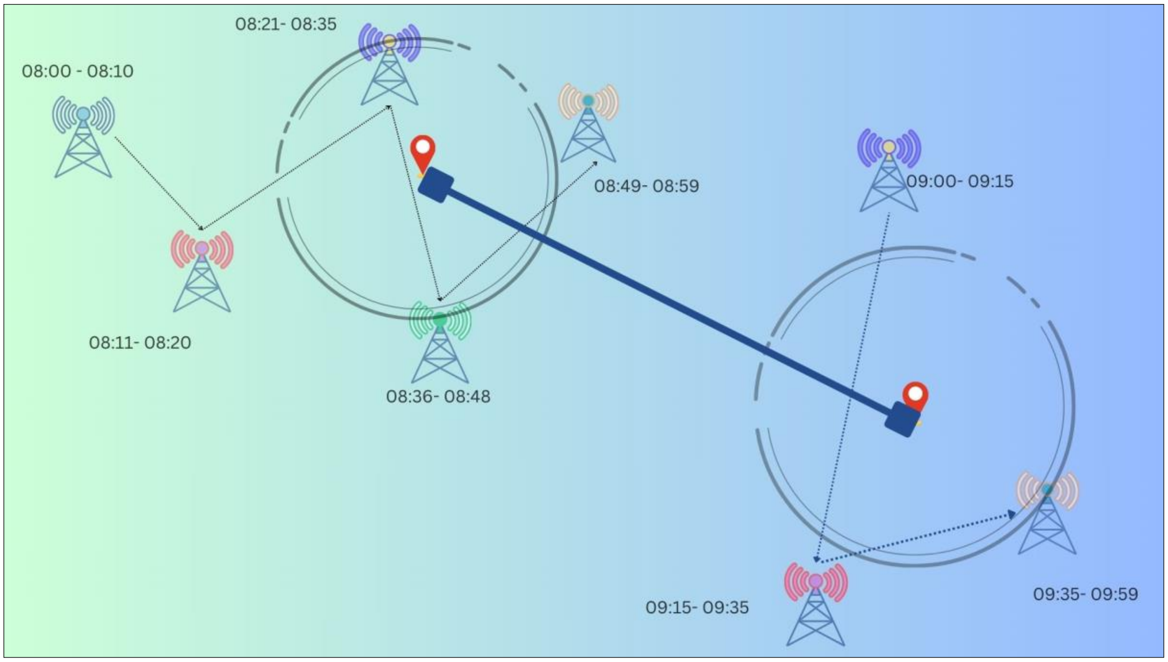


Figure 5_ Calculation of travelled distance for mobility index definition in 2023

In 2024, Vodafone updated its algorithm. As shown in Figure 6, the new approach now involves more than just two centres. It also considers the last antenna from the previous hour, the first antenna of the next hour, and both centres.

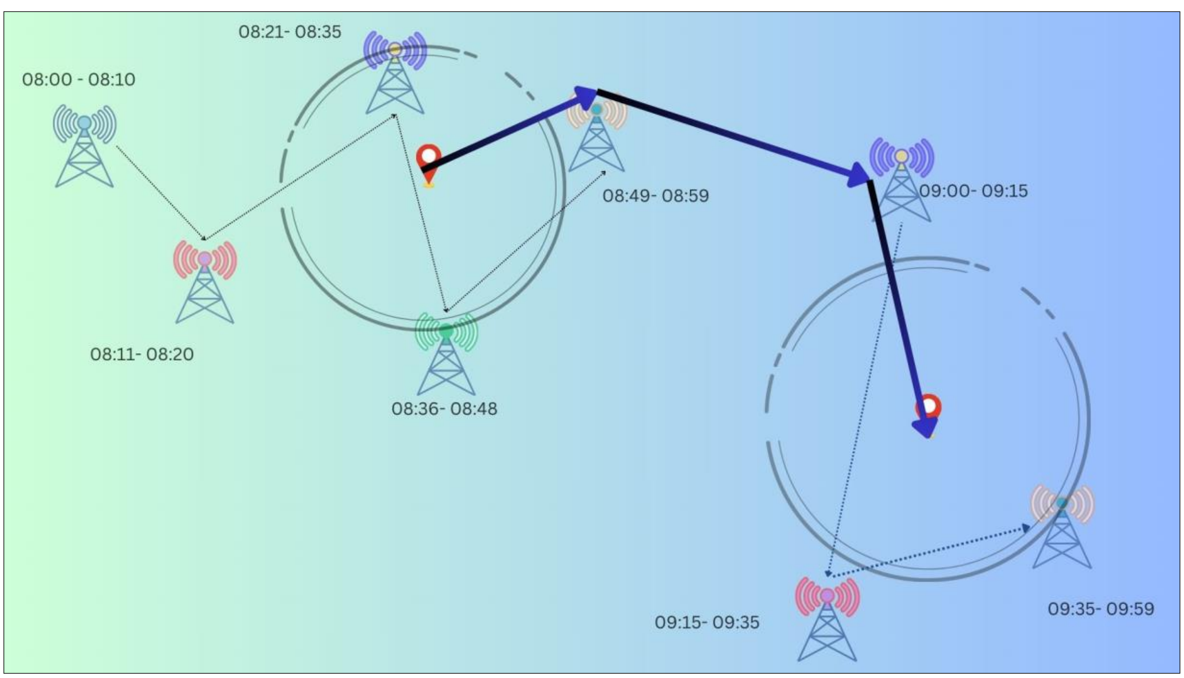


Figure 6_Calculation of travelled distance for mobility index definition2024

3.1.2 Audimob Observatory

The Audimob Observatory allows to monitor mobility in Italy. The report “Survey on mobility styles and behaviour of residents in Italy” (Isfort) shows the results of the annual survey on local and urban mobility. The territorial level is the Country as a whole, with a focus on both urban and national mobility. The time reference is annual (Methodological elements for the analysis of passenger mobility using Big Data, Ferrovie dello Stato Italiane 2023)

Figure 7 illustrates all the categories and their details within the Audimob framework.

Category	Details
Reference Methodology	• Adapted to Eurostat Guidelines.
Reference Population	• Residents aged 14-84 years.
Sample	• A quota sample of over 16,000 individuals, optimized for gender, age, and region (oversampling metropolitan cities).
Observation Method	• Observing all journeys made the day before the interview (both weekdays and holidays). • Completion of a log detailing each journey. • Recording socio-economic characteristics and vehicle availability.
Main Variables	• Time, duration, length, origin, destination, purpose of the journey. • Means of travel (by modal stage). • Fuel source (if using a private car). • Accompanying persons.
Survey Technique	• 70% Computer-Assisted Telephone Interviewing (CATI). • 30% Computer-Assisted Web Interviewing (CAWI).
Handling non-response	• Replacing units with quotas.
Indicators Published	• Annual time reference and national territorial reference.

Journey Characteristics	• Classes of journey length (% distribution). • Classes of daily journey frequency (% distribution, cluster criterion). • Purpose of the journey (% distribution) • Mode of transport (% distribution).
Sustainable Mobility	• Sustainable mobility rate (% distribution by mode of transport).
Time Slots	• Time slots (% distribution).
Journey Systematic Nature	• Systematic journeys (% distribution).
Mobility Types	• Urban, extra-urban, medium-/long-haul, and total mobility.
Gender-Based Metrics	• Mobility rate (% by gender).
Age Groups	• Mobility rate (% by age).
Occupational Status	• Mobility rate (% by occupational status).
Geographical Breakdown	• Mobility rate (% by geographical region).
Municipality Demographics	• Mobility rate (% by municipality size).
Municipality Classification	• Mobility rate (% by SNAI classification).
Smart Working Metrics	• Average number, length, and duration of journeys.

Figure 7_ category and details of Audimob survey

3.2 Data analysis

Data analysis involved a three-stage process:

- first, a data analysis using Business Intelligence (BI) tools, specifically Tableau for visualisation;
- second, a SWOT analysis was used to characterise the pros and cons of the two data sets and their potential;
- third, a comparison of data set features to understand if and how to use those data to have a comprehensive and reliable picture of mobility patterns.

3.2.1 Data analysis using Business Intelligence

Visualising the raw data allows for a clearer understanding of key differences and the identification of the main features within the dataset. These insights gained through visualisation are critical for guiding further analysis and drawing meaningful conclusions from the data.

Tableau is a business intelligence and data visualisation tool that enables users to transform data into interactive visualisations such as charts, maps, and dashboards. This allows for real-time analysis by breaking down complex data into easily understandable components (Andry et al., 2021). Tableau enables users to connect to various cloud-based data sources, including big data, SQL databases, spreadsheets, Google Analytics, and Salesforce. It allows for the creation of calculations, reference lines, forecasts, and statistical analyses such as trend analyses, regressions, and correlations. Without the need for coding, Tableau also gives advanced users the ability to pivot, split, and manage metadata to enhance and optimise data sources (Salgador, 2018). Tableau is utilised for analysing and visualising datasets from Vodafone and Isfort, helping to understand the main trends and changes.

Vodafone data analysis

The analysis of the Vodafone dataset utilises two distinct datasets: OD matrices and mobility indexes. These datasets originate from the same raw data source provided by Vodafone. However, as Vodafone does not provide direct access to raw data, the analysis is based on a semi-processed dataset that enables customised examination.

The OD matrices focus on specific origin-destination flows, while the mobility indexes are processed to capture overall mobility trends. These datasets are compiled using different algorithms for data collection and processing. Through Tableau, various measures and dimensions are applied to analyse and interpret the data effectively. Vodafone datasets include detailed measures and dimensions that are shown in Figure 8.

Measure	• Trips
	• Passengers* kilometres
	• Time
	• Total distance
	• Age class

Dimensions	• Destination
	• Destination Province
	• Destination Region
	• Day of the week
	• Distance km
	• Distance Group
	• Time slot
	• Modality
	• months
	• Nationality
	• Origin of trip
	• Province of origin
	• Region of origin
	• Residence class
	• Systematic and none systematic trip
	• Type of the day
	• Year
	• Unique User

Figure 8_features of big data sets

The two key datasets, **OD matrices** and **mobility indexes**, enable in-depth analysis across different periods and spatial distributions, providing valuable insights into travel behaviour and trends through:

1. **comparison of different periods:** one of the strengths of these datasets is their ability to compare travel behaviours across distinct periods, even when faced with unpredictable changes such as the COVID-19 pandemic. Each period represents the number of trips made during specific months from 2019 to 2022 in two periods of six months, before and after COVID-19. These two periods are:
    - **Period 1:** August 2019 to January 2020 (pre-pandemic travel behaviour)
    - **Period 2:** August 2021 to January 2022 (post-pandemic travel behaviour).

This comparison highlights shifts in mobility patterns before and after the onset of the pandemic, capturing changes in trends and travel behaviour effectively;

2. **in-depth temporal analysis:** for a more granular look into mobility patterns, a specific month is analysed in detail. **October 2021** is selected as a representative "normal" month, allowing for deep temporal analysis of daily and hourly travel data. This approach helps identifying routine mobility behaviours and potential anomalies within a typical month;

3. **use of mobility index data:** in addition to OD matrices, the mobility index dataset in 2021 is leveraged to explore different aspects of mobility. This dataset provides insights into various measures, such as total distance, which help quantifying the total distance travelled by passengers. This dataset is selected because it captures overall mobility, excluding trips shorter than 800 metres. In contrast, OD matrices represent a smaller sample of total mobility and do not allow for the calculation of total kilometres travelled. With this dataset, it is possible to accurately sum up the overall kilometres using mobility indexes.

Furthermore, the dataset is well-suited for comparison with Audimob, as it aligns with the period, measurement structure, and dimensions used in Audimob analysis.

3.2.2 Comparison of Datasets

The methodology for comparing the datasets involves a structured approach to identify, analyse, and extract relevant factors and details from both datasets. Specifically, all dimensions and measures present in both datasets are carefully identified and compared. In the context of Business Intelligence (BI), dimensions are descriptive attributes or categories that allow data to be grouped and analysed, such as time periods, geographic locations, or product categories. Measures, on the other hand, are numerical metrics or values that quantify data, such as sales figures, revenue, or customer counts. Measures are analysed within the context of dimensions to derive actionable insights.

The comparison method focuses on identifying all dimensions available in both datasets and evaluating their level of detail. This involves determining which dimensions and measures are shared by both datasets, which are unique to only one dataset, and the level of granularity or detail at which they are recorded. By systematically mapping dimensions and measures across the two datasets, similarities, differences, and any gaps in coverage are highlighted. This approach ensures a comprehensive comparison of the datasets and their respective data points.

.

3.2.3 SWOT analysis

For the analysis of the two datasets, Big Data (Vodafone) and Survey (Audimob) SWOT analysis was adopted. This approach enabled a comprehensive evaluation by identifying internal factors—such as strengths and weaknesses—and external factors, like opportunities and threats, for each dataset. While internal factors are within our control, external factors, shaped by market and environmental influences, are largely beyond it. This structured method provided a strategic overview, guiding us from broad observations to more detailed analyses tailored to each dataset's context.

4. Results

This chapter presents the results, organised into three sections:

1. visualisation of the results using Tableau with Vodafone and ISFORT Data. This section details the insights gained from visualising the Vodafone and ISFORT datasets in Tableau, emphasising key trends and patterns in different periods;
2. comparison of the features of the two datasets. This section analyses and compares the features of the two datasets to identify similarities, differences, and unique characteristics, providing a comprehensive overview of their compatibility and individual value;
3. SWOT Analysis of each dataset. In the final section, SWOT tables are created for both datasets, summarising their strengths, weaknesses, opportunities, and threats to support strategic insights.

4.1 Analysis of OD matrices

OD (origin-destination) matrices, captured via Vodafone SIM data, cover 3,000 zones across Italy, excluding trips within the same zone.

The dataset provides valuable insights across multiple dimensions, including age, distance groups (in kilometres), and travel modes. The dataset includes specific day-level details, indicating whether a day is a working day or a holiday, along with data for each day of the week and hourly time groups for the trips. The data distinguishes between systematic and non-systematic trips and offers information on the residency status of travellers, identifying Italians and foreigners. The available measures for analysing these dimensions include the number of trips and passenger kilometres, enabling comprehensive mobility analysis.

4.1.1 Trips by transport mode before and after COVID

Figure 9 shows the number of trips taken in two periods of time that were before explained, segmented by different modes of mobility. The chart categorises trips into different modes: “Altro” (Other devices, including private car, bicycle, etc.)"Ferrovia" (Railway), "Mascherato" (unknown), and "Via Aerea" (Air travel).

Figure 9_comparing the number of trips taken in different months across the years

Figure 10 compares two periods based on train travel data before and after the COVID-19 pandemic. For every month in the post-COVID period, the number of train journeys is lower than in the pre-COVID period, with January showing a particularly significant drop. This indicates that train travel has not fully recovered to pre-pandemic levels.

Figure 11 shows, that the train travels were impacted significantly by COVID-19 and, even months later, faced challenges in fully recovering. In December and January, there was another decline, especially in January 2022, with 34% fewer trips compared to the normal period.

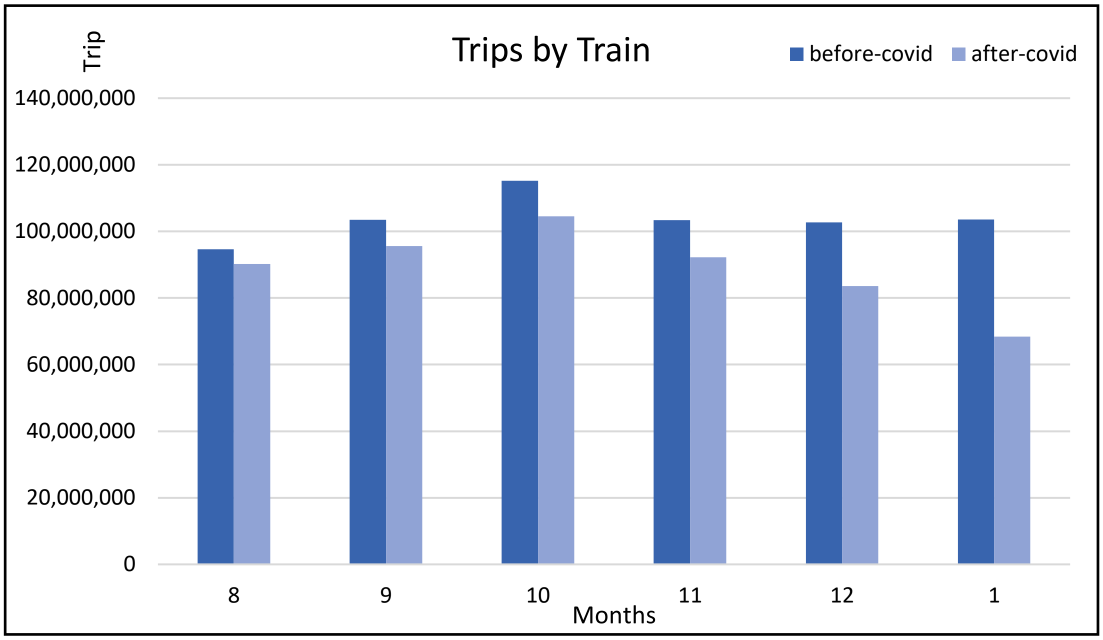


Figure 10_ comparing travels by train in two periods

First 6Months	Trips	Second 6Months	Trips	difference	%
Train					
8	94,620,371	8	90,166,402	-4,453,969	-4.7
9	103,509,747	9	95,578,951	-7,930,796	-7.7
10	115,210,522	10	104,521,140	-10,689,382	-9.3
11	103,368,283	11	92,230,753	-11,137,530	-10.8
12	102,698,177	12	83,574,156	-19,124,021	-18.6
1	103,536,002	1	68,376,280	-35,159,722	-34.0

Figure 11_difference of the number of trips in two periods

According to Figure 12, a comparison of trips with other vehicles before and after COVID-19 across several months. It reveals that trips before COVID-19 (represented by darker blue) generally remained steady with minor fluctuations, while trips after COVID-19 (represented by lighter blue) show a slight decrease, particularly in August and December. The months following COVID-19 (October to January) exhibit a more consistent trend in both before and after periods, with trips in the post-COVID period being slightly lower but remaining relatively close to pre-COVID levels.

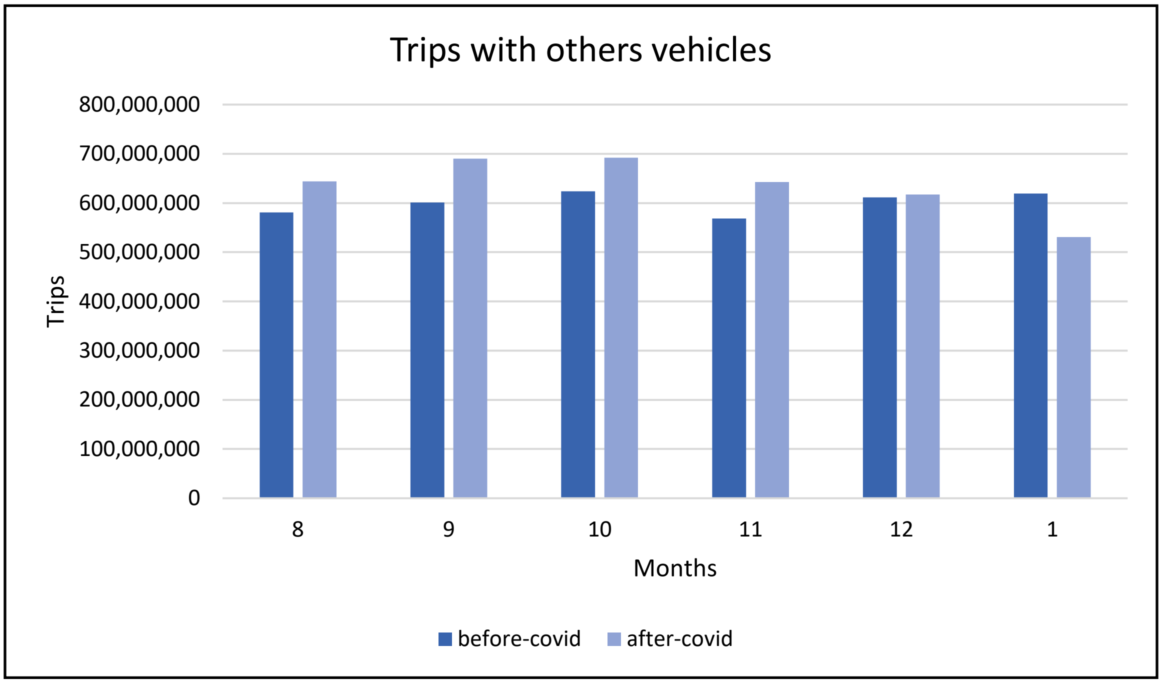


Figure 12_Comparing trips by other vehicles in two periods

Figure 13 shows that unlike trains, which saw a decline, travel by other transport modes increased in the post-COVID period. However, even in this mode, there was a decrease in January 2022 by 14% due to the new wave of COVID-19 in this month.

first 6Months	Trips	Second 6Months-	Trips	difference	%
Altro(other vehicles)					
8	580,874,227	8	644,043,312	63,169,085	10.9
9	601,293,398	9	690,362,955	89,069,557	14.8
10	624,096,958	10	691,965,773	67,868,815	10.9

11	568,772,014	11	642,673,635	73,901,621	13.0
12	611,256,729	12	617,206,182	5,949,453	1.0
1	618,897,815	1	530,616,237	-88,281,578	-14.3

Figure 13_difference in the number of trips in two periods

Figure 14 highlights a noticeable difference in the number of trips by aeroplane between the two periods across all months, illustrating a clear contrast in the data before and after the pandemic.

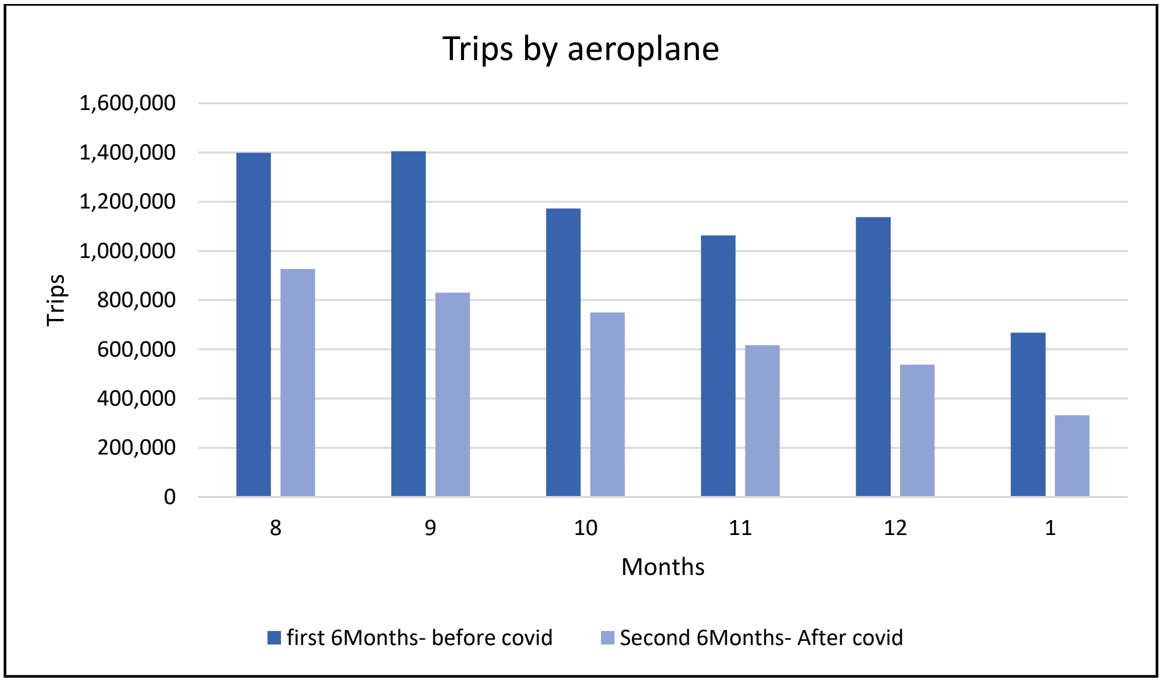


Figure 14 _ comparing trips in two periods

In Figure 15 air travel is shown as one of the sectors most severely impacted by the pandemic. The recovery has been slow, with the number of trips staying below one million in the post-COVID period. The figure highlights a significant decline in the number of aeroplane trips during the second 6-month period compared to the first. The percentage change consistently shows a decrease across all months, with the most dramatic reduction occurring in December (52.7%) and January (50.1%). In general, the reduction in trips ranged from around 33.7% in August to 41.9% in November, indicating a steady decline as the months progressed.

First 6Months	Trips	Second 6Months	Trips	difference	%
Aeroplane					
8	1,398,829	8	926,947	-471,882	-33.7
9	1,404,960	9	830,088	-574,872	-40.9
10	1,172,139	10	749,372	-422,767	-36.1
11	1,063,585	11	617,817	-445,768	-41.9
12	1,137,193	12	538,172	-599,021	-52.7
1	667,602	1	332,983	-334,619	-50.1

Figure 15_difference in the number of trips in two periods

4.1.2 Passenger-km and trips in two periods

Figure 16 illustrates the relationship between passenger kilometres and the number of trips across different months in two six-month periods. In this figure, the passenger kilometre metric, which measures the total distance travelled by all passengers (calculated by multiplying the number of passengers by the distance travelled in kilometres), is depicted by the height of the bars, while the number of trips is represented by an orange line.

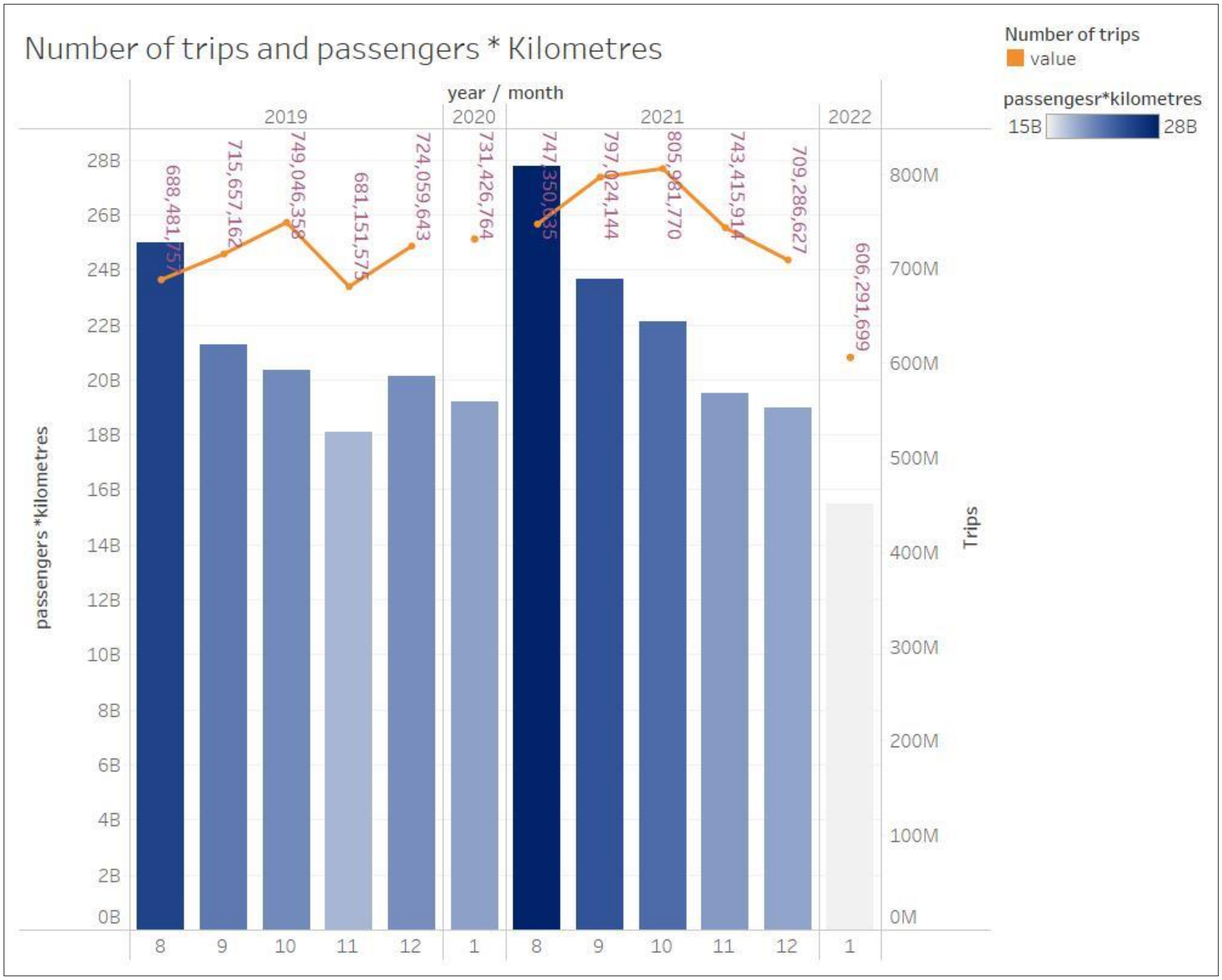


*Figure 16_Trips with considering passenger*km*

According to figure 16, both passenger kilometres and total trips increased during the second period, except December – and especially January – where a decline occurred due to a new wave of COVID-19.

Figure 17 displays the number of trips in various distance groups across different months and years. The distance groups are divided into four intervals: 0-2 km, 2-10 km, 10-50 km, and more than 50 km. Each bar within these groups represents the number of trips made during specific months from 2019 to 2022.

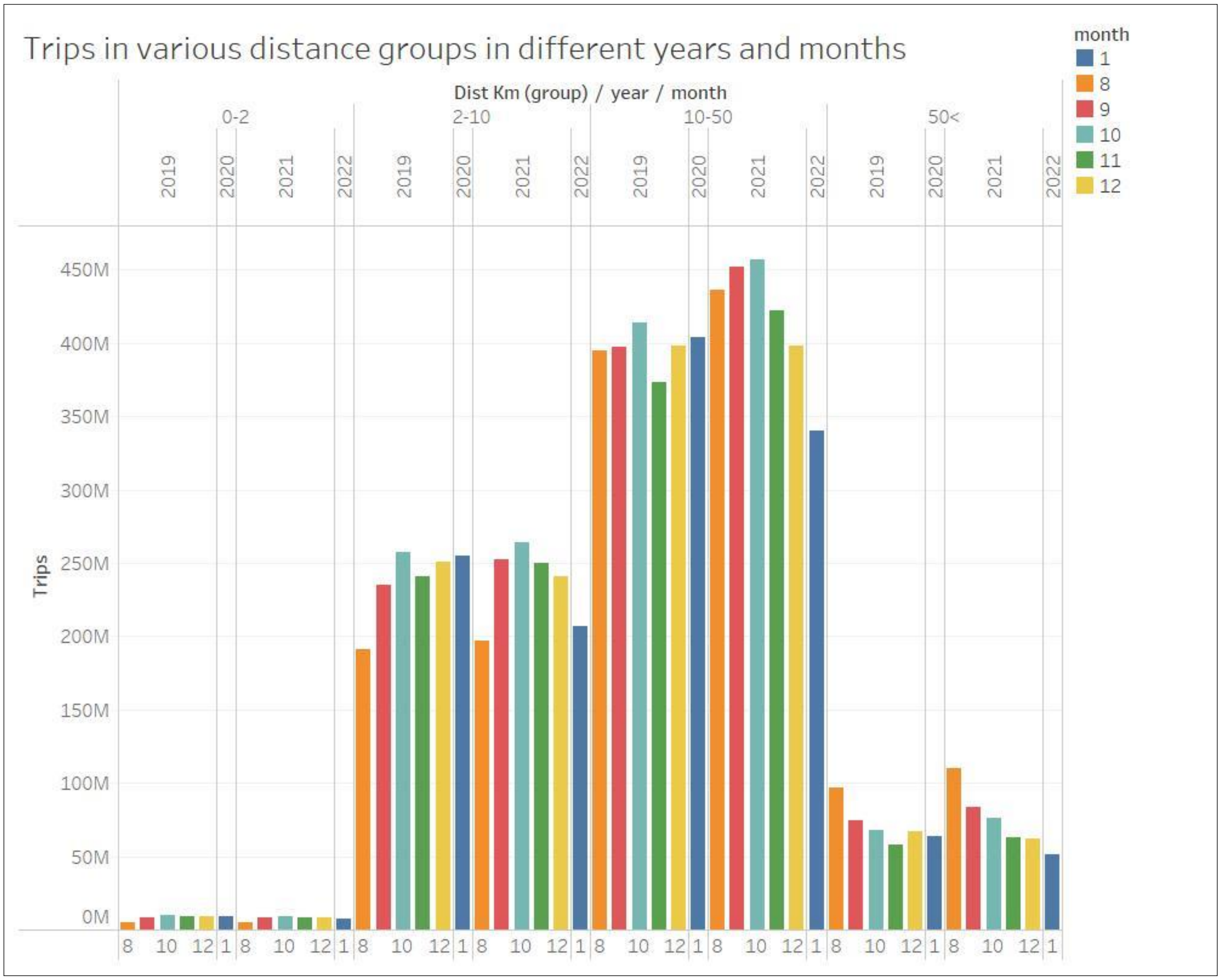


Figure 17_Comparing travel according to the distance

Overall, in the Vodafone matrices, most trips are recorded within the 10–50-kilometre range in both periods. This distance category has shown a positive trend, indicating a stronger recovery after COVID-19 than before the pandemic. Additionally, there were not many changes in the other distance groups between the two periods. However, January saw the most significant changes across all groups, with a decrease in trip numbers.

Figure 18 illustrates the number of trips taken that are less than 2 km. The chart covers data for both periods before and after COVID-19. It focuses on trips with distances less than 2 kilometres and uses the passengers*kilometres (p*km) .

Figure 18_comparing travels according to the distance

In none of the months analysed did the trips reach pre-pandemic levels. In the second period, however, October remained the peak month with the highest number of trips while there was a decline of approximately two million trips in January.

Figure 19 depicts the number of trips form 2 to 10 kilometres, where the highest number of trips occur in the first four months of the second period, till to December. However, in December trips start to decrease and in January a decrease of approximately 50 million trips can be observed.

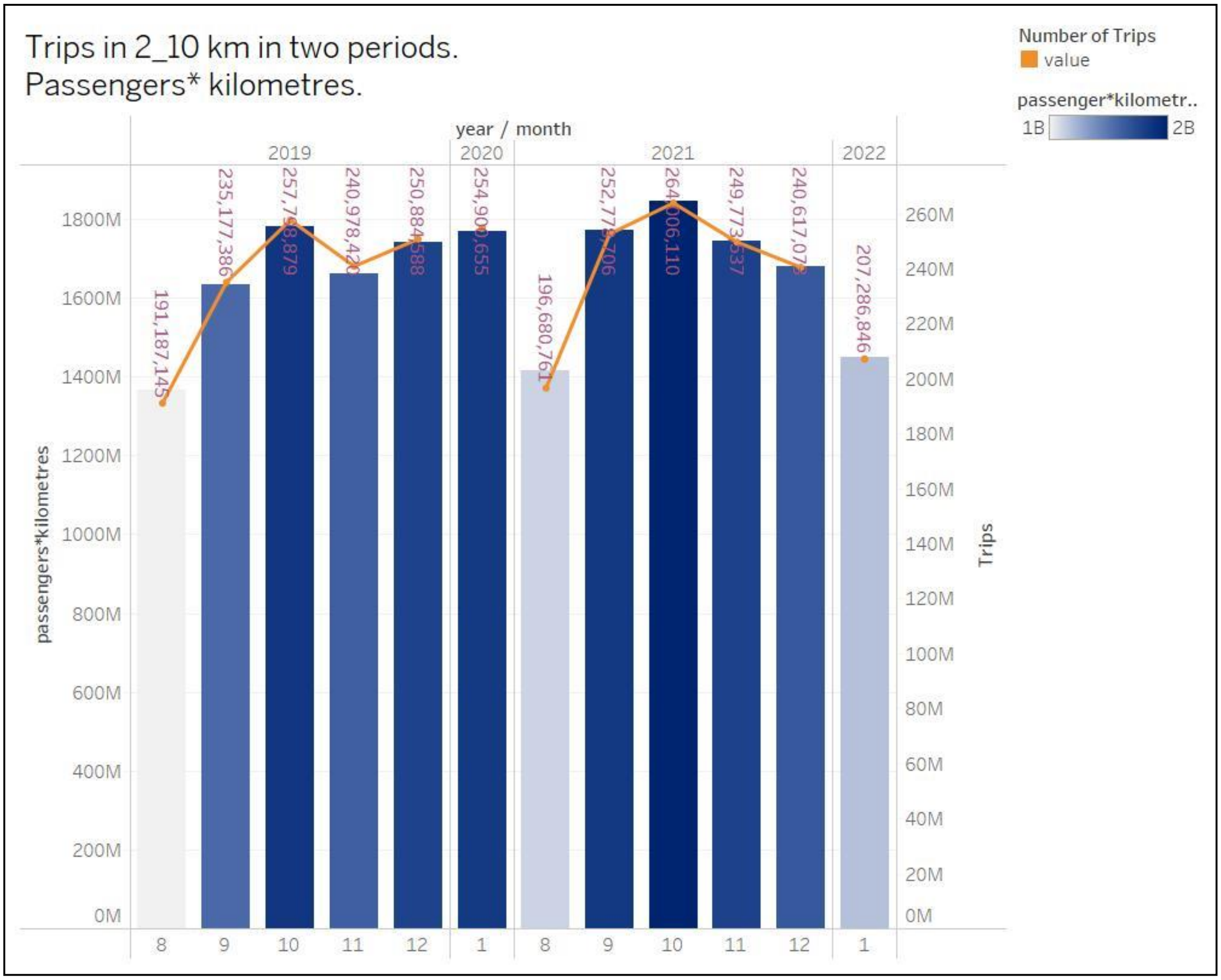


Figure 19_ trips between 2-10 km in two periods

Figure 20 illustrates the number of trips from 10 to 50 kilometres that show a significant increase in all months of the second period (except January), fully recovering and even surpassing normal trends. In fact, in the first three months of the second period, trips increased by about 50 million each month, unlike trips of other length intervals, which only showed a peak in October. However, similarly to other distance intervals, January saw a decline of around 60 million trips.

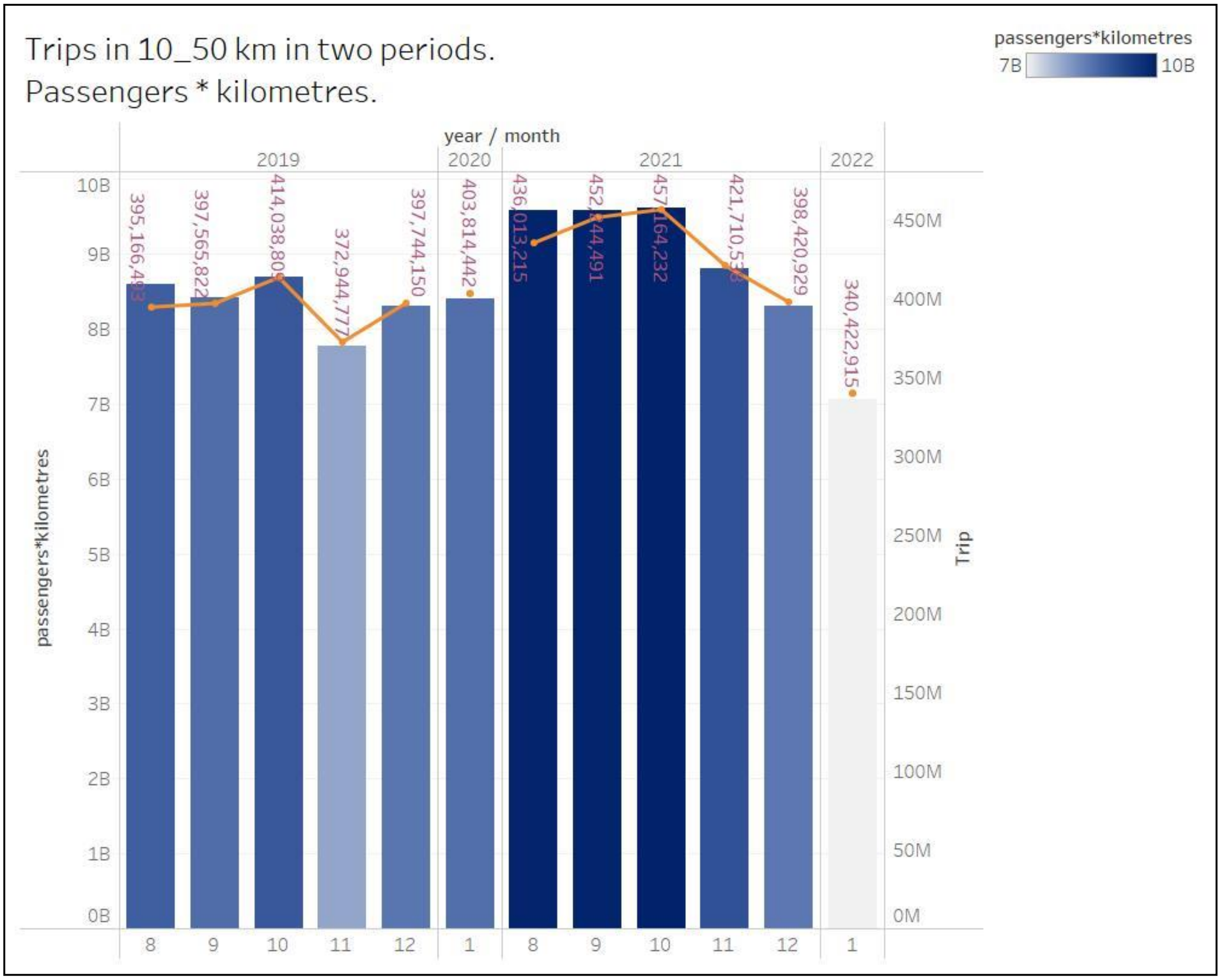


Figure 20_Trips between 10-50 km in two periods

Figure 21 focuses on trips over 50 km, showing a unique pattern in mobility for longer distances. These trips occur more frequently during the summer holidays, with 2021 surpassing 2019 levels, suggesting a full recovery from the pandemic and reaching over 100 million trips. However, in January, trips decreased again compared to pre-COVID levels.

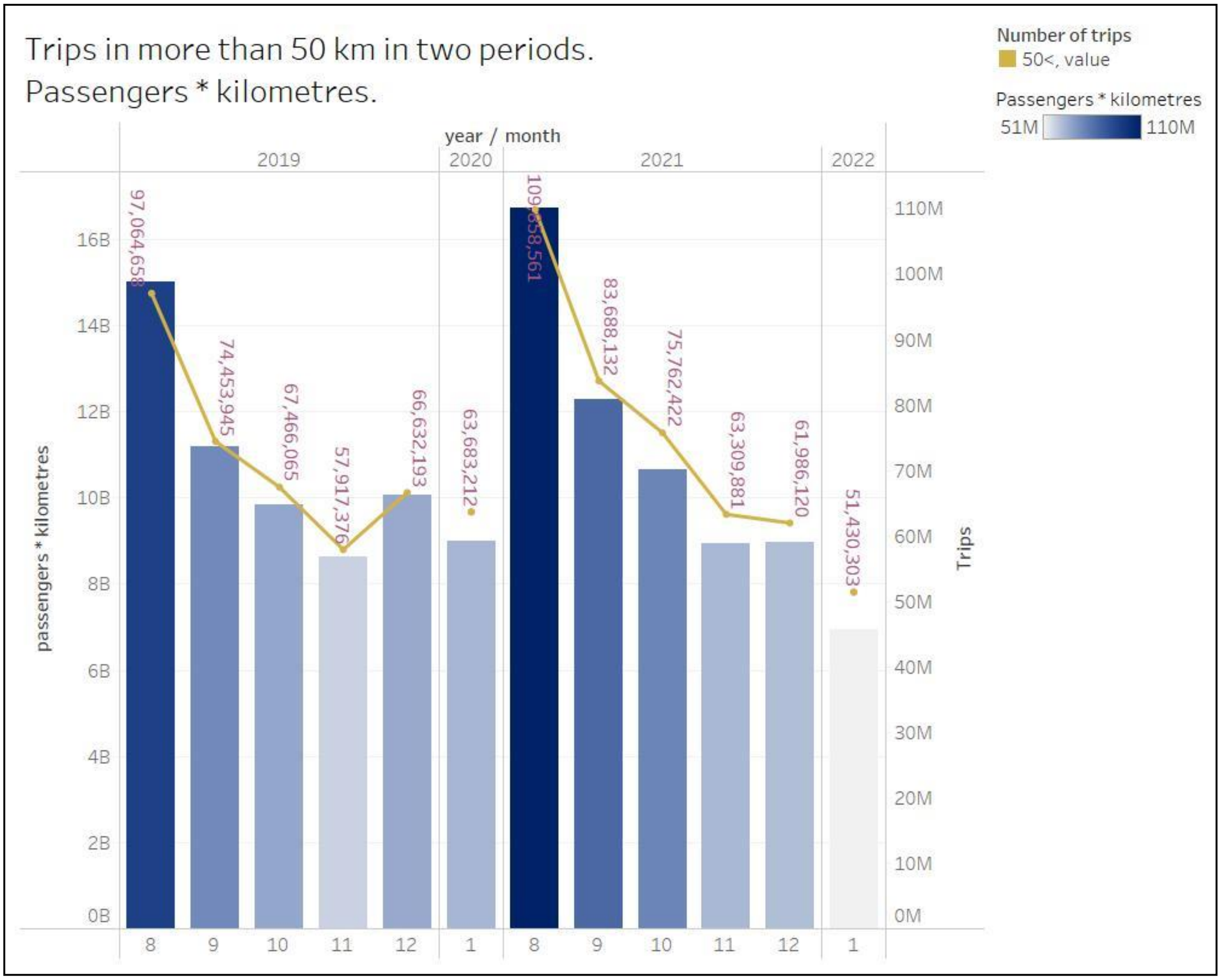


Figure 21_Trips in more than 50 km in two periods

4.1.3 OD matrices monthly data analysis

In the following figures, the focus shifts to a single month in 2021 – October – which is often considered a typical month for travel analysis. This marks the second phase of the analysis, demonstrating how effectively big data (mobile phone data) can be used to delve into detailed temporal patterns.

Age classes

In Figure 22, different groups are presented based on various age classes, showing a comparison between age-related categories.

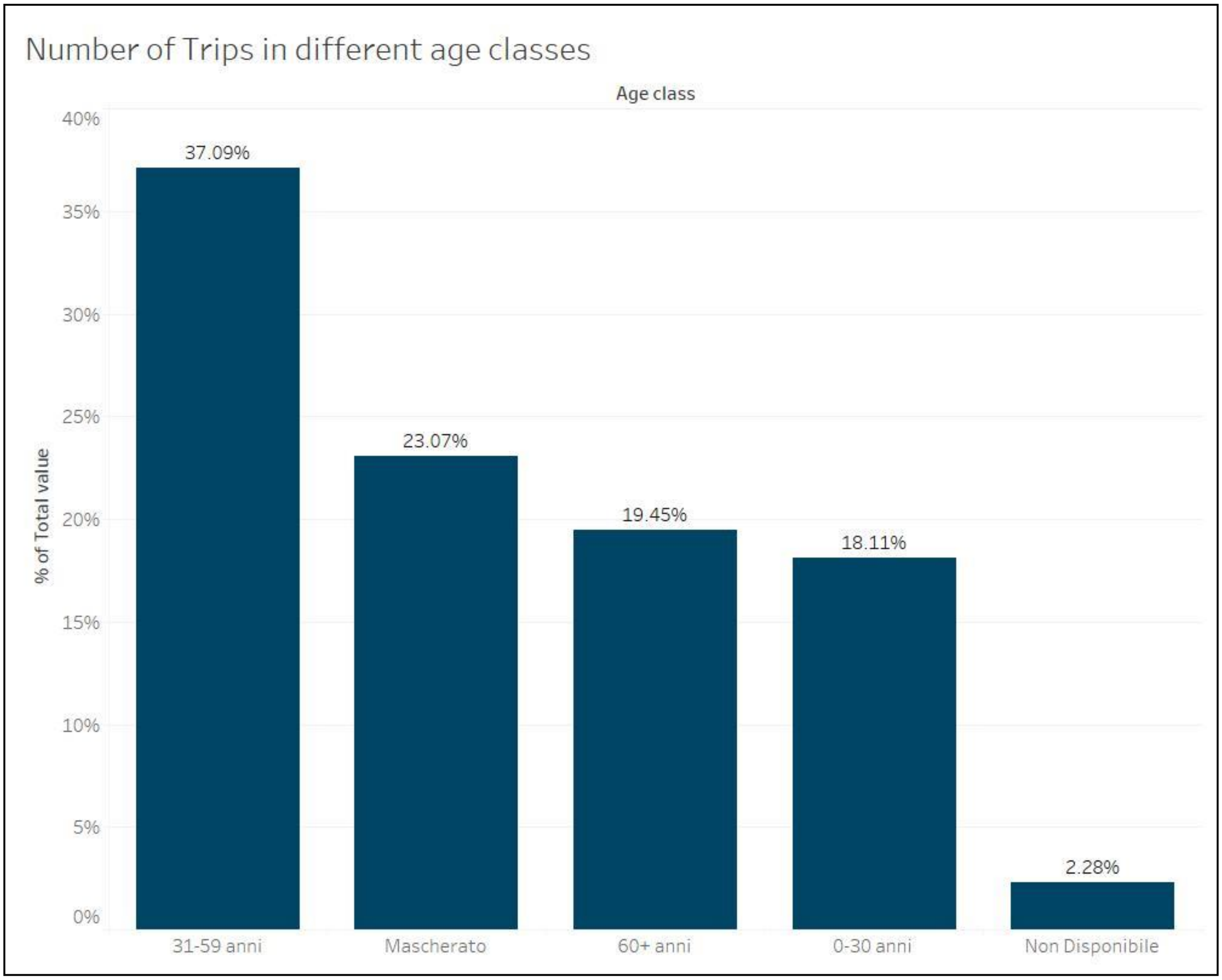


Figure 22: Number of trips in different age classes

The age-related data are divided into five intervals. The majority of trips, about 37%, were made by individuals in the 31-59 age interval. The second largest group, labelled as "Mascherato" (hidden), includes cases where some information is obscured due to privacy regulations, accounting for 23% of total trips. Individuals aged more than 60 represent 19% of the trips, while the youngest age group, up to 30 years old, cover 18%. Lastly, there is a group of trips where age data are unavailable.

Trips on all days of the week

According to Figure 23, there is not much difference in the distribution of trips from Monday to Saturday, being Tuesday the day of the week with the highest number of trips. On the contrary, Sunday records a sharp drop in the average daily trips.

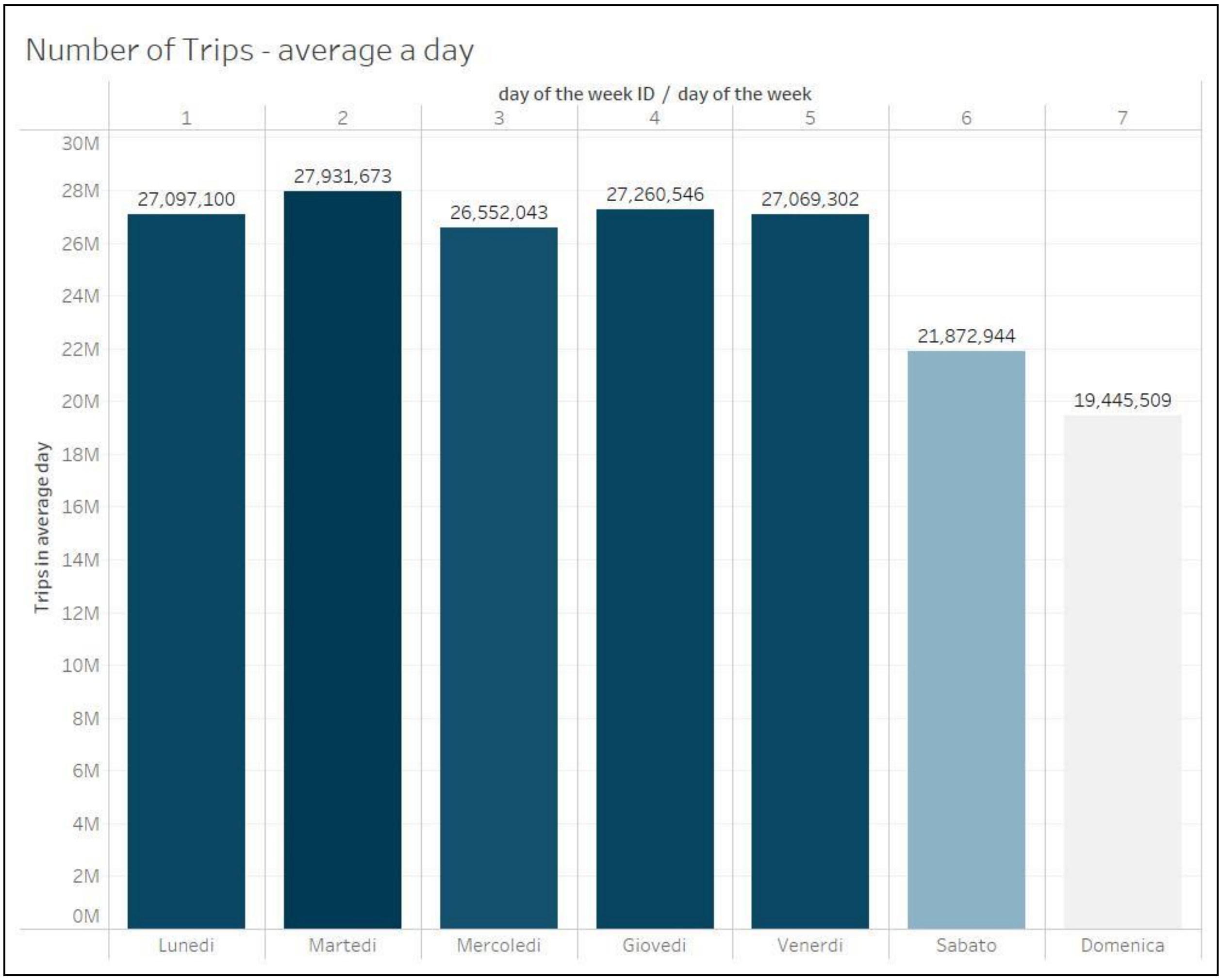


Figure 23 : Number of trips on an average day of the week

The figure 23 shows that the average number of trips during weekdays remains relatively consistent, while the number of trips on weekends is noticeably lower compared to weekdays.

Trips in different time periods

Figure 24 shows travel values classified according to different periods of the day. These categories include morning trips (typically corresponding to the start of office and school periods) or trips in the afternoon (coinciding with the return to home from work and school).

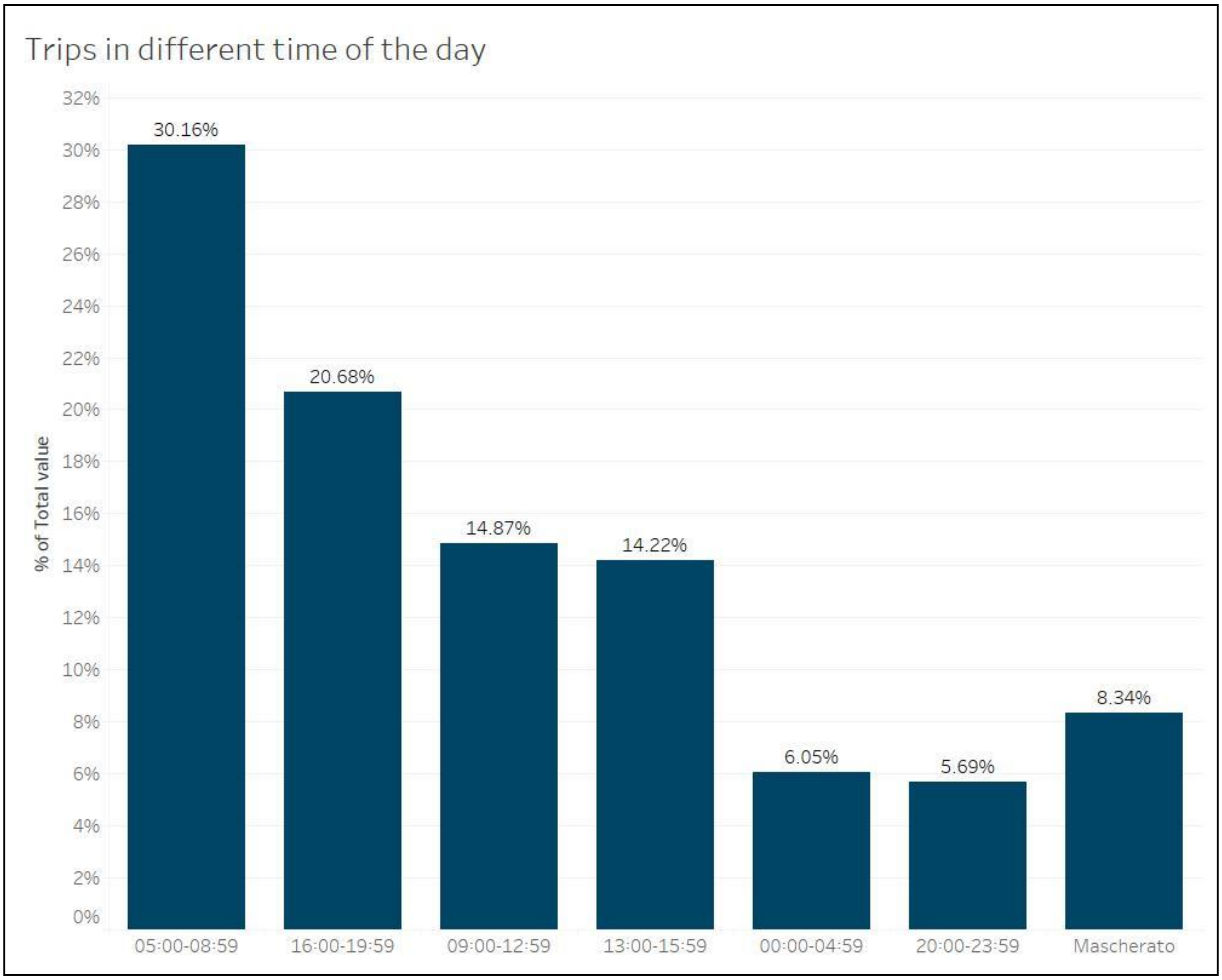


Figure 24_Comparing trips in different Time slots

The period between 05:00 and 08:59 records the highest number of trips, around 30%, being the peak interval for travelling. The second busiest period is in the afternoon, from 16:00 to 19:59, with approximately 21%. These two peak periods likely correspond to typical commute times for work, school, and returning home in the afternoon.

Systematic and non-systematic trips

Systematic mobility is formed by trips occurring repeatedly on the same origin-destination (O/D) and/or within the same time slots. The comparison of systematic versus non-systematic trips is illustrated in Figure 25.

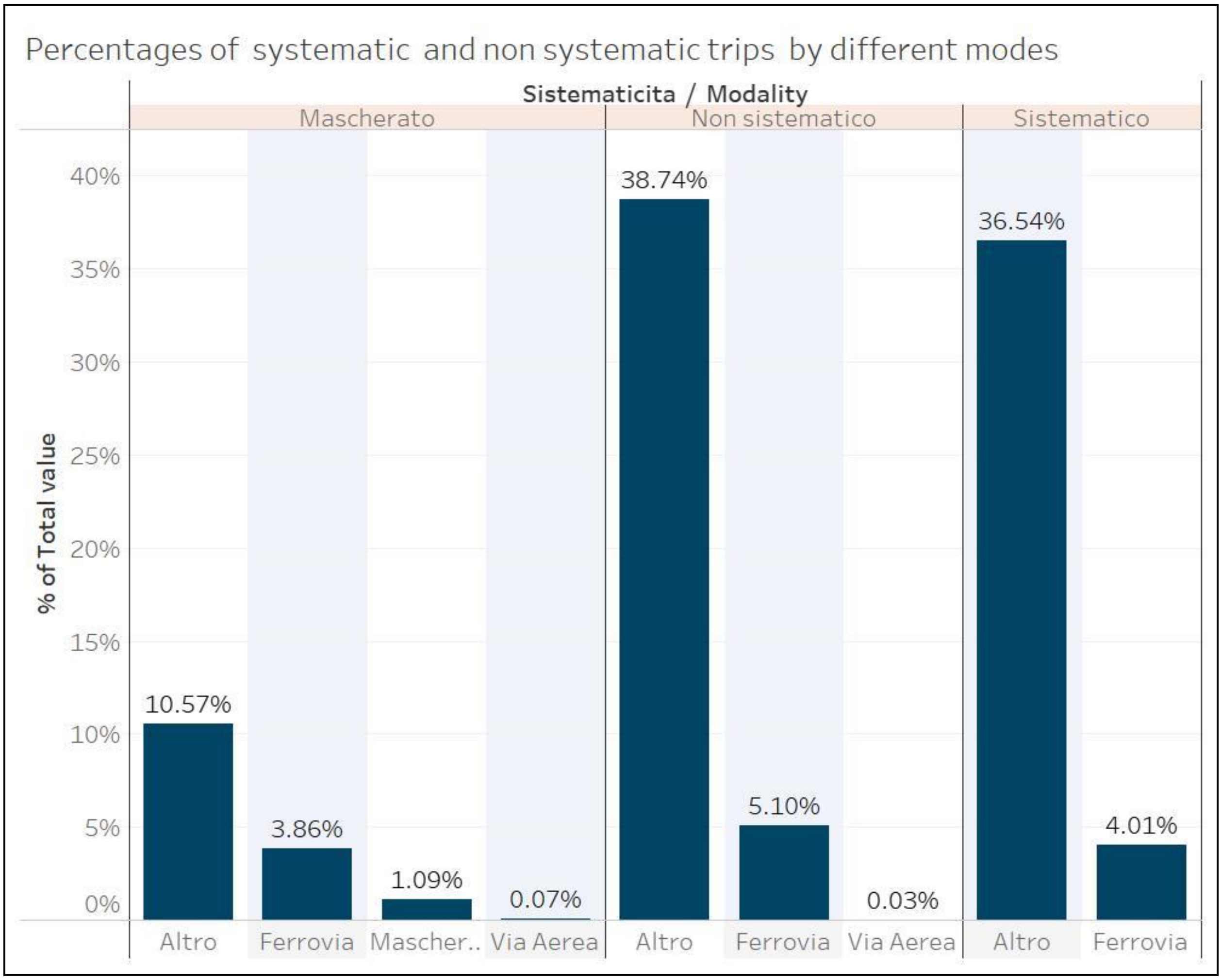


Figure 25_ Systematic and non-systematic trips

Figure 25 shows the percentage distribution of systematic (Sistematico) and non-systematic (Non sistematico) trips across various transport modes. In non-systematic trips, the "Altro" category dominates with 38.74%, while rail (Ferrovia) accounts for 5.10%, and air travel (Via Aerea) is negligible at 0.03%. For systematic trips, "Altro" also leads with 36.54%, followed by rail at 4.01% and air travel at 0.07%. The "Mascherato" group, representing hidden data, shows the following percentages: 10.57% for "Altro," 3.86% for rail, and only 1.09% for the hidden category, with air travel at 0.07%.

Trips by transport mode

Figure 26 presents trips made using three transport modes: Altro (Other), Ferrovia (Railway), and Via Aerea (Airway) and hidden data. It focuses on trips that occurred in October 2021.

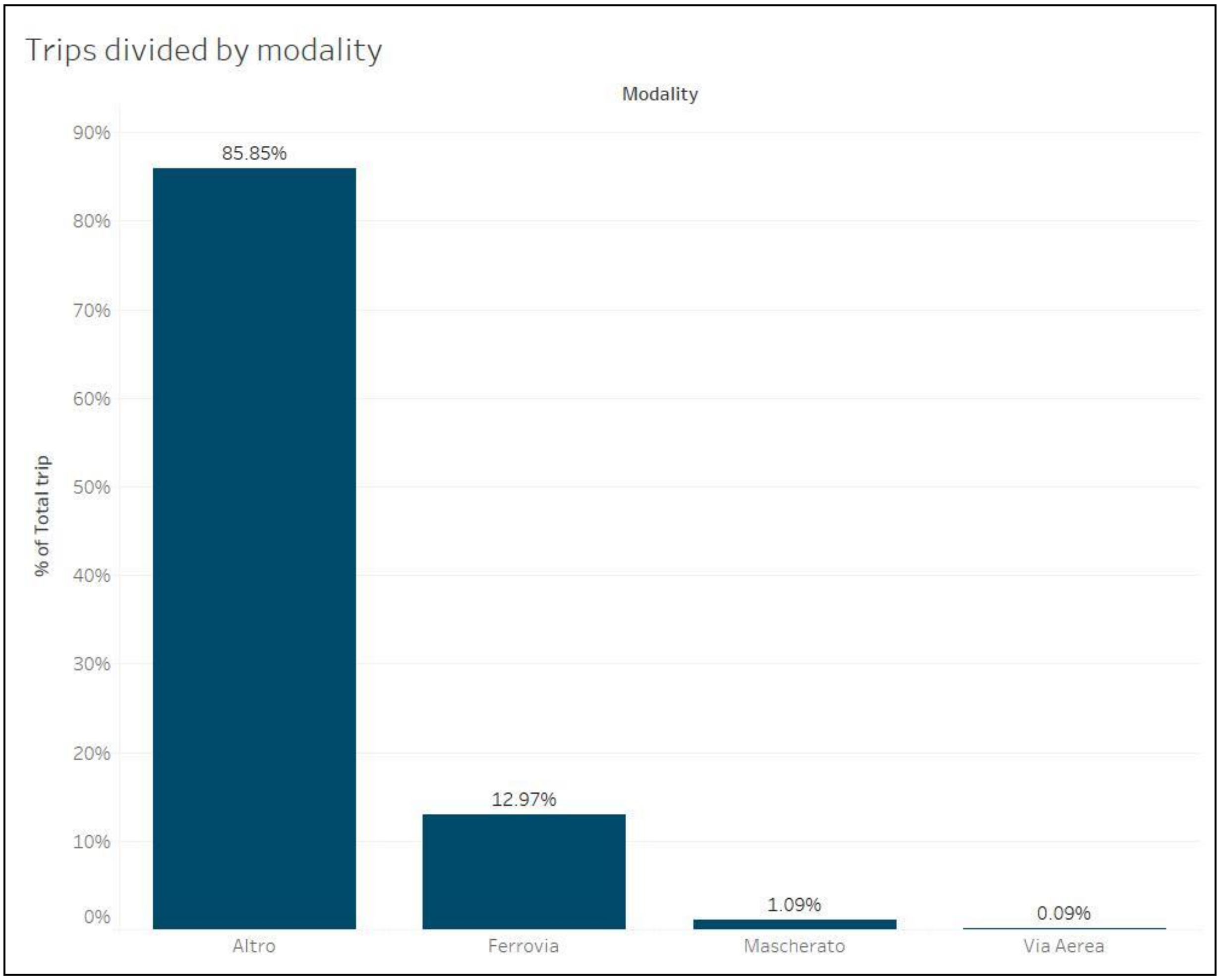


Figure 26_Trips divided by modality

The "Altro" category is the most dominant one, making up 85.85% of all trips. Rail transport (Ferrovia) follows with 12.97%. The "Mascherato" group contains hidden data and accounts for 1.09% of the trips. Air travel (Via Aerea) is the least used mode, representing 0.09% of total trips.

Trips by Region

Figure 27 shows the number of trips generated in each region. In October 2021, the majority of trips originated from the regions of Lombardy, Lazio, and Campania. At the opposite end of the chart, Basilicata, Molise, and Valle d'Aosta recorded the lowest number of trips. Among these regions, Lombardia shows the highest number of trips, which will be analysed in greater detail in Figure 28 which illustrates the destinations of trips originating from the Lombardia region. As shown in Figure 28, out of approximately 95% of trips originate from Lombardy while the remaining trips are primarily directed to neighbouring regions such as Piedmont, Emilia-Romagna, and Veneto.

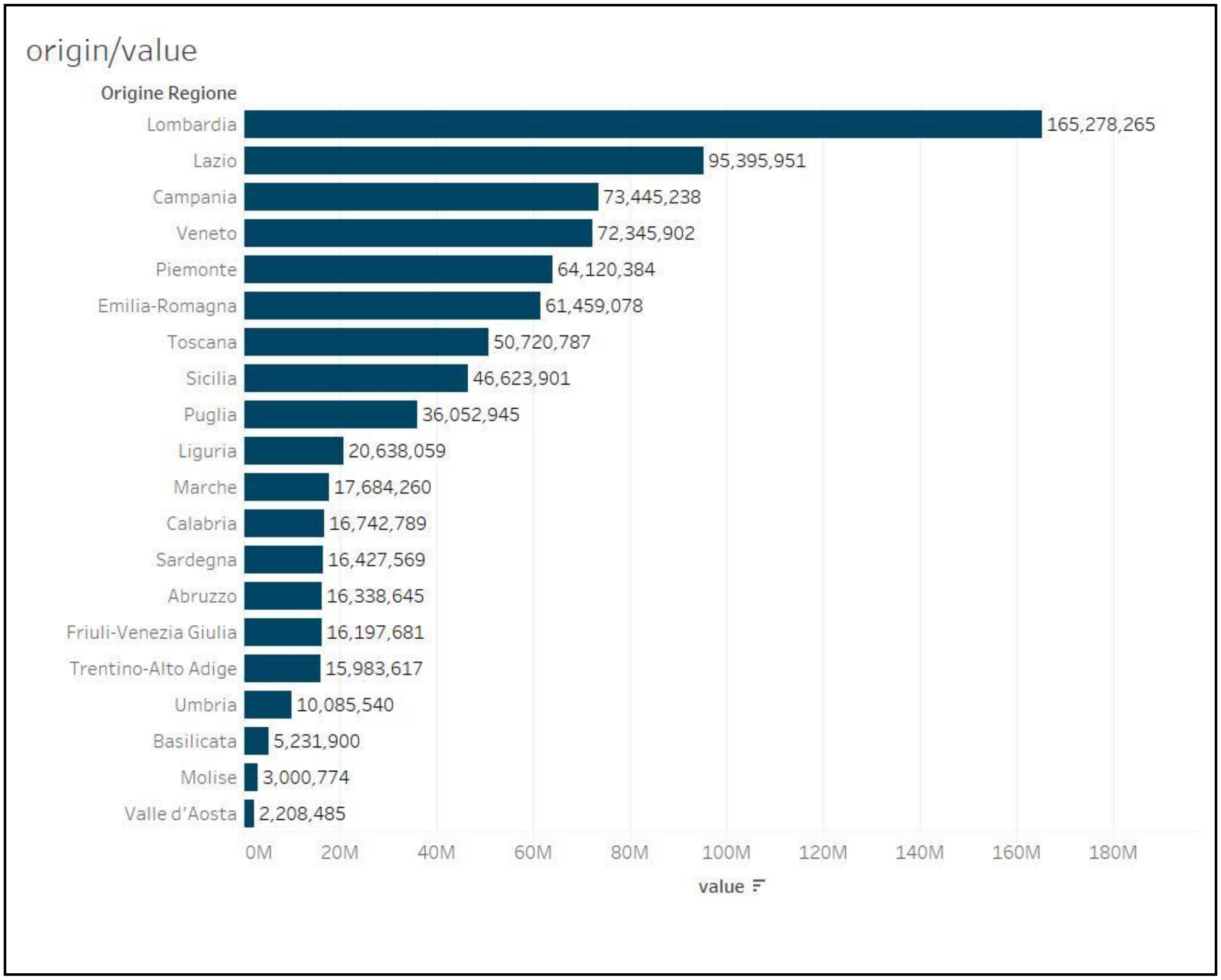


Figure 27_Comparing amounts of trips by origin regions

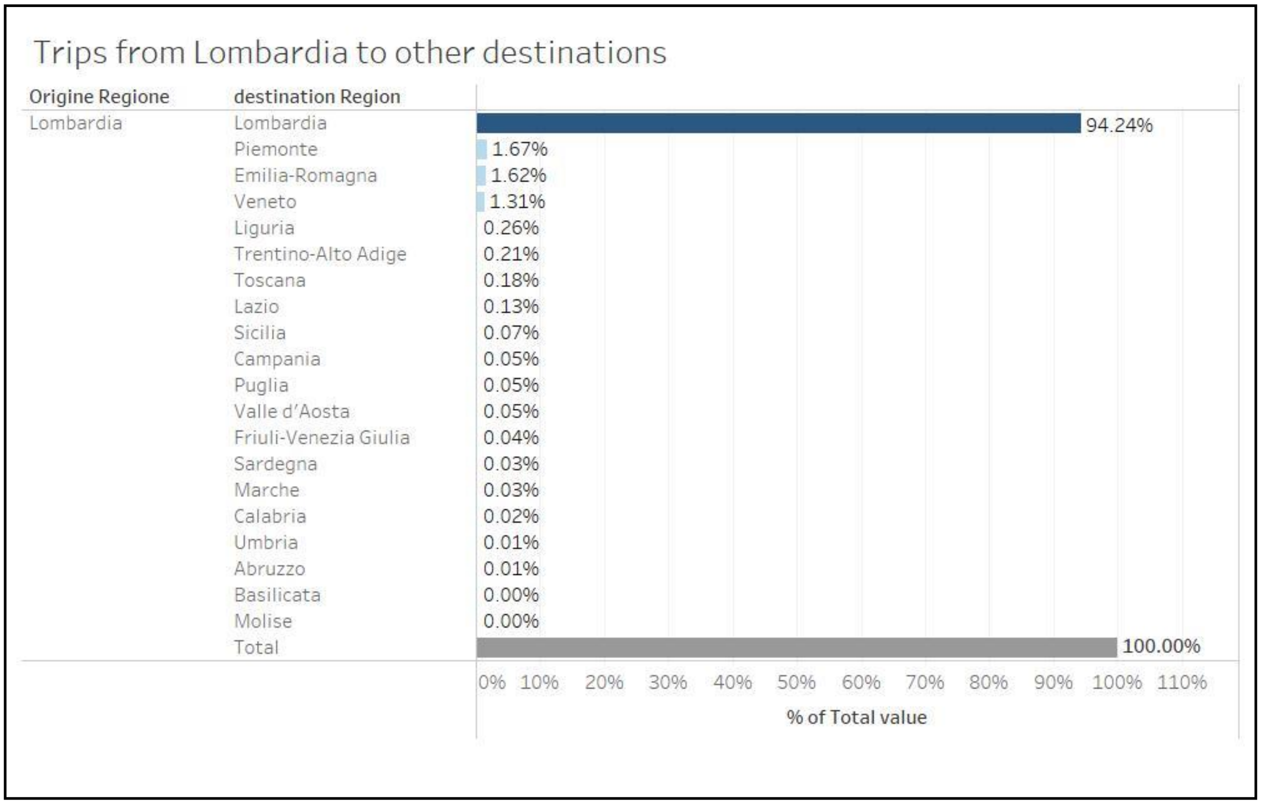


Figure 28_ Trips from Lombardi to destinations

4.1.4 Analysis of Mobility Indexes Data

The mobility indexes dataset in 2023 in October was used for analysis with Tableau. So, after considering some features and measures, charts were created for a deeper analysis of Vodafone datasets.

Total distance by transport modes

Figure 29 shows that the majority of kilometres, exceeding 45 billion, were made using modes classified as "other." In contrast, the fewest trips, amounting to less than 10 billion, were made by aeroplanes.

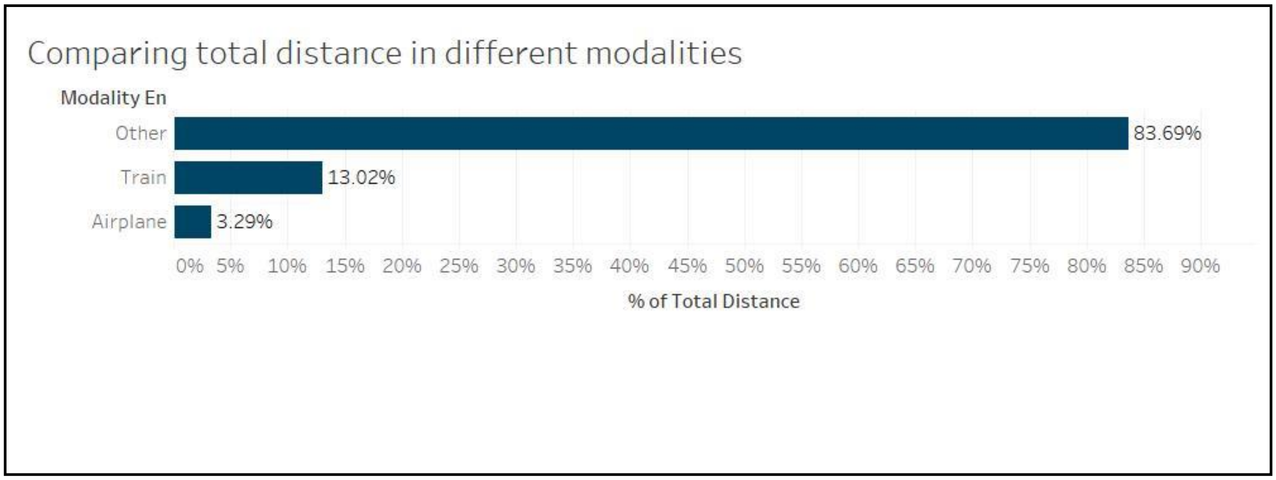


Figure 29_comparing total distance by transport modalities

Travellers and users

Figure 30 shows two groups of people, who are not travelling and travellers which means understanding among people using phones how many travel and how many do not travel. Figure 30 divides data into two categories of mobile phone users, based on their mobility: *stationary* and *travellers*. This classification comes from the mobility index dataset, where users are classified as either "0" (stationary individuals who do not travel) or "1" (individuals who travel). The chart reveals a significant disparity between the two groups. Travellers constitute a much larger number of unique users being 75% of the total. In contrast, the stationary group accounts for around 25% of unique users, showing that travellers make up a considerably larger portion of the unique users.

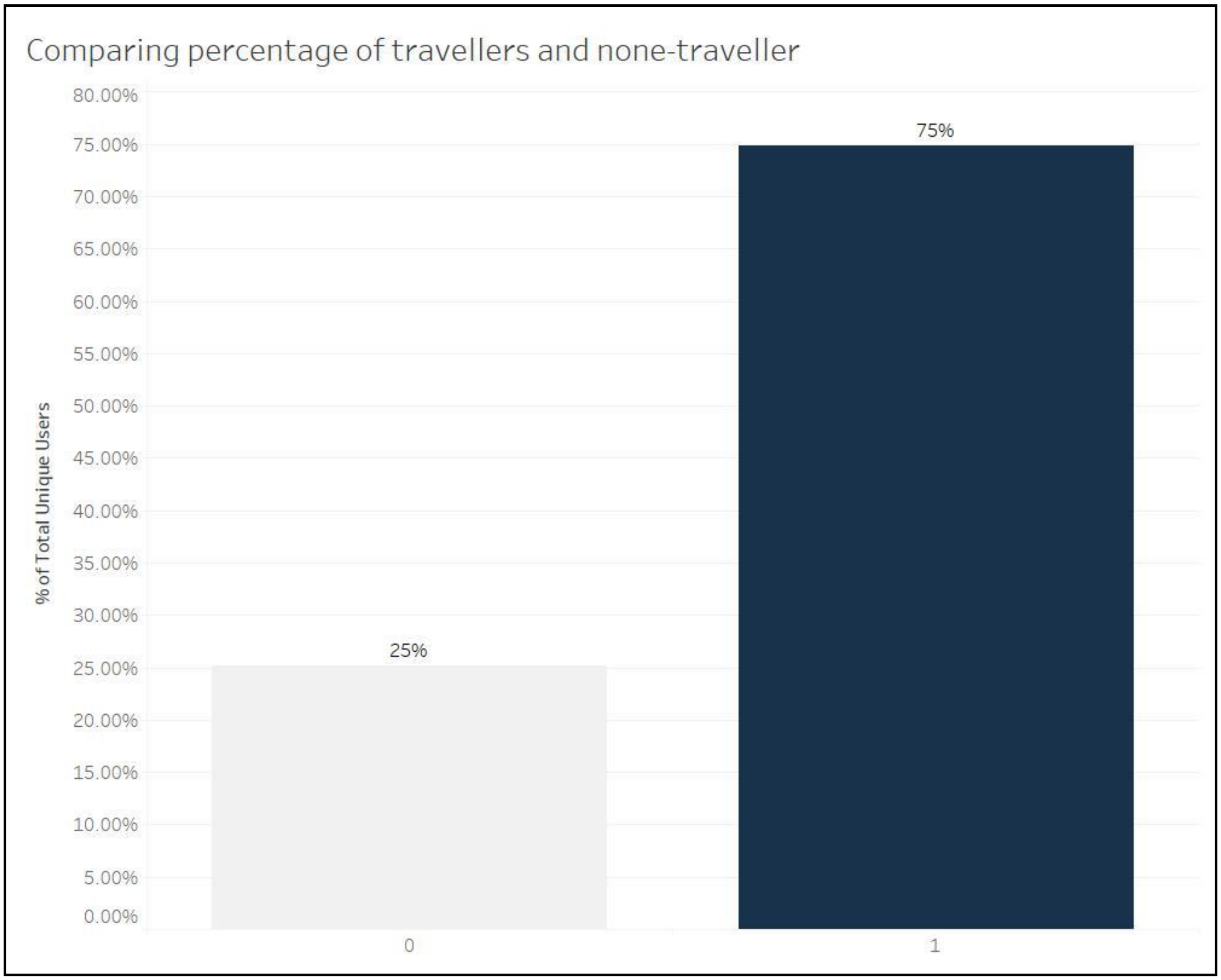


Figure 30 _ travellers and not travellers

Unique users by province

Figure 31 displays the percentage of travellers across various provinces, classified in two groups: travellers (1) and non-travellers (0). Data reveal that provinces such as Bergamo, Brescia, Caserta, Padova, and Salerno record the highest percentage of travellers, with values ranging from 78% to 81%. In contrast, Bari (68%) and Palermo (69%) show relatively lower proportions of travellers. Non-travellers make up the remaining percentages, illustrating regional variations in travel activity.

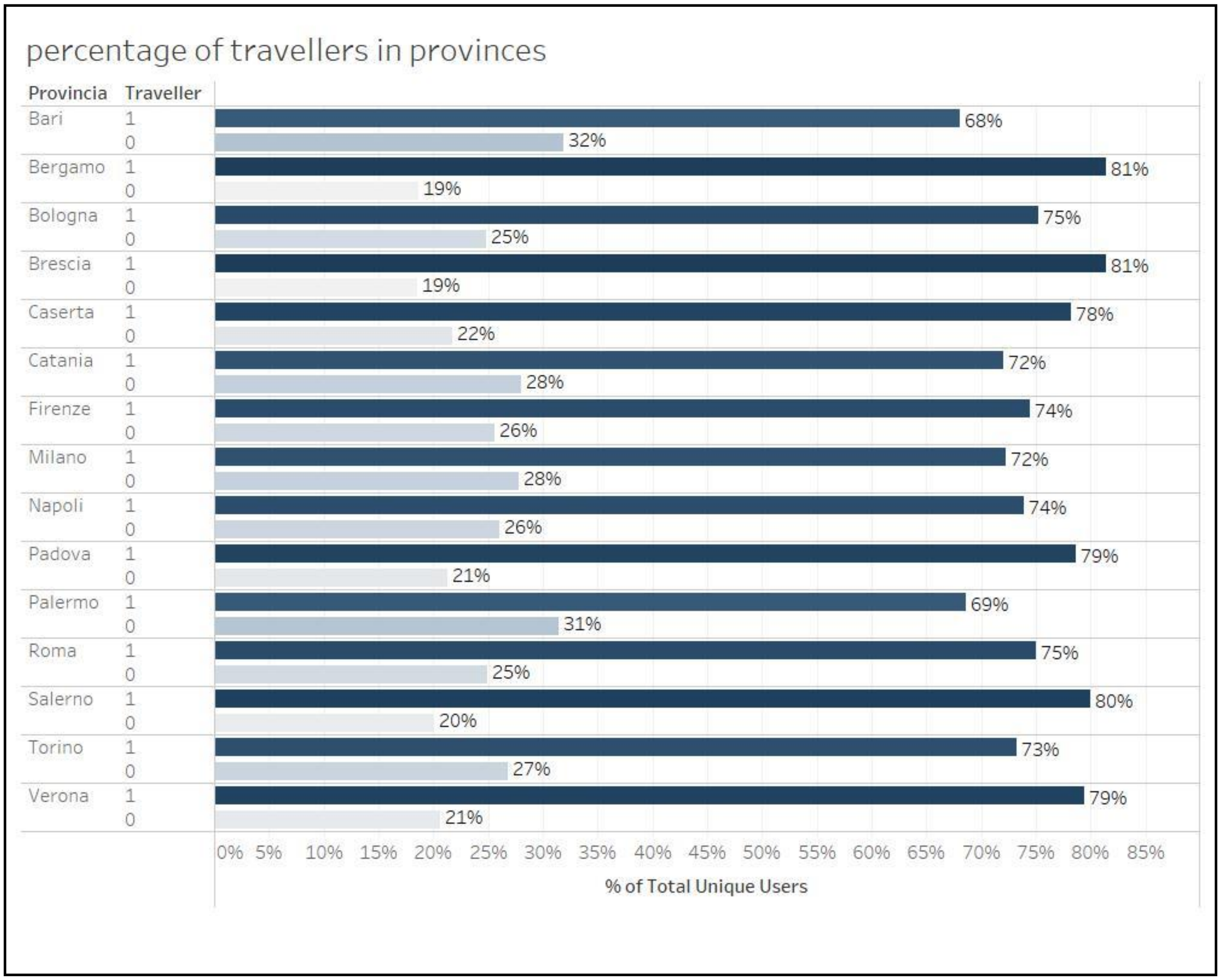


Figure 31_ Travellers and none travellers in different provinces

Total distance travelled in different provinces

Figure 32 shows the total distance travelled in different provinces; the distances are measured in kilometres. Roma, Milano and Napoli are the first three provinces with the highest amount of total distance travelled.

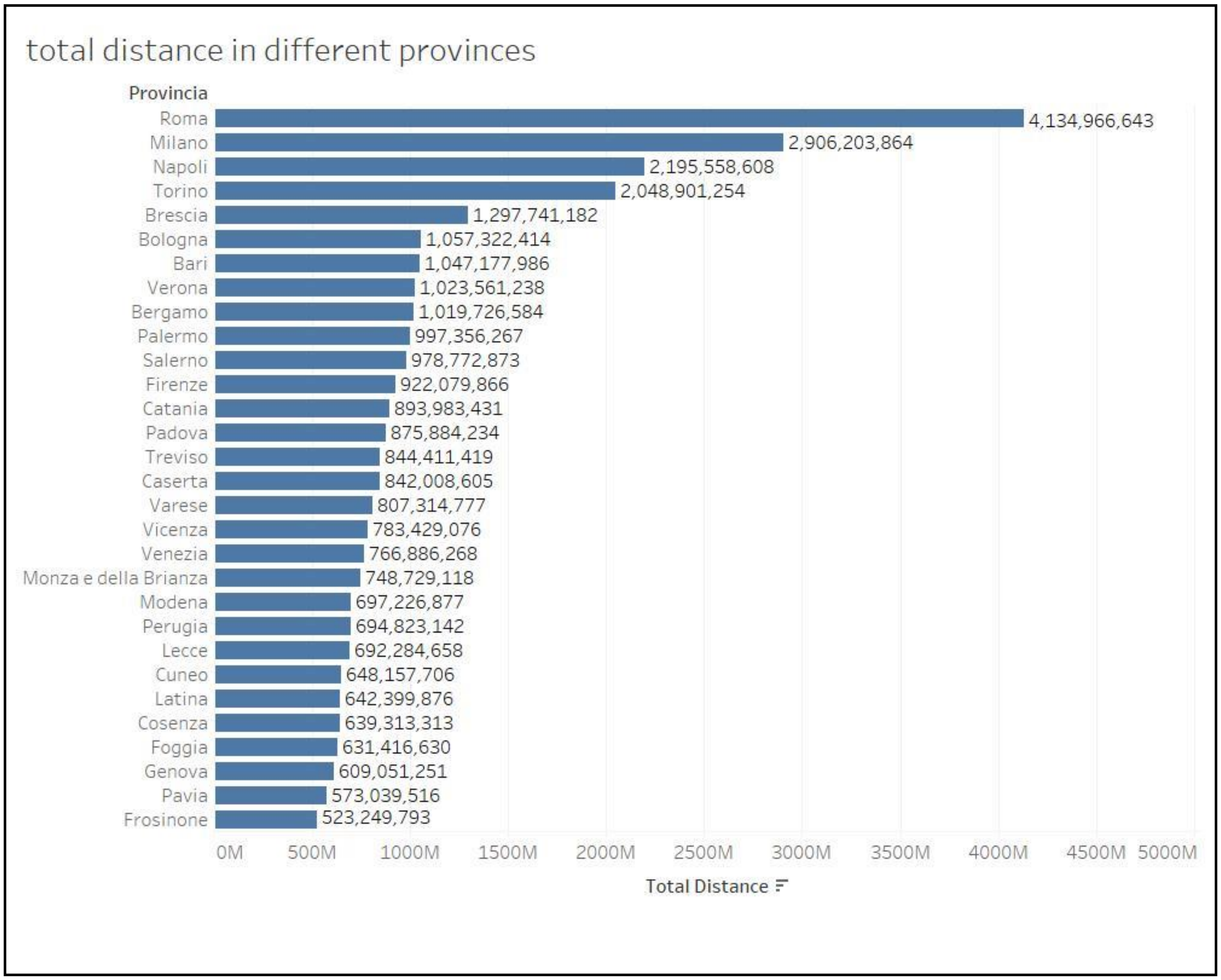


Figure 32_ Total distance in provinces

Total distance, trips and unique users

Figure 33 compares the metrics of total distance, trips, and unique users across three transport modes: air aeroplanes, train, and other. The data reveal distinct patterns in using these modes of transport. This visualisation compares three metrics – total distance travelled, total trips, and total unique users – across three transport modes: "Other," "Train," and "Aeroplane." The "Other" category dominates across all metrics, accounting for 88.88% of the total distance, 98.06% of the trips, and 97.78% of the unique users. Train travel contributes 8.93% of the total distance, 1.88% of trips, and 2.11% of unique users, while aeroplane travels are minimal, with 2.18% of the distance covered, 0.06% of trips, and 0.11% of unique users. These results show that most travel activities are concentrated in the "Other" category, with trains and aeroplanes playing much smaller roles.

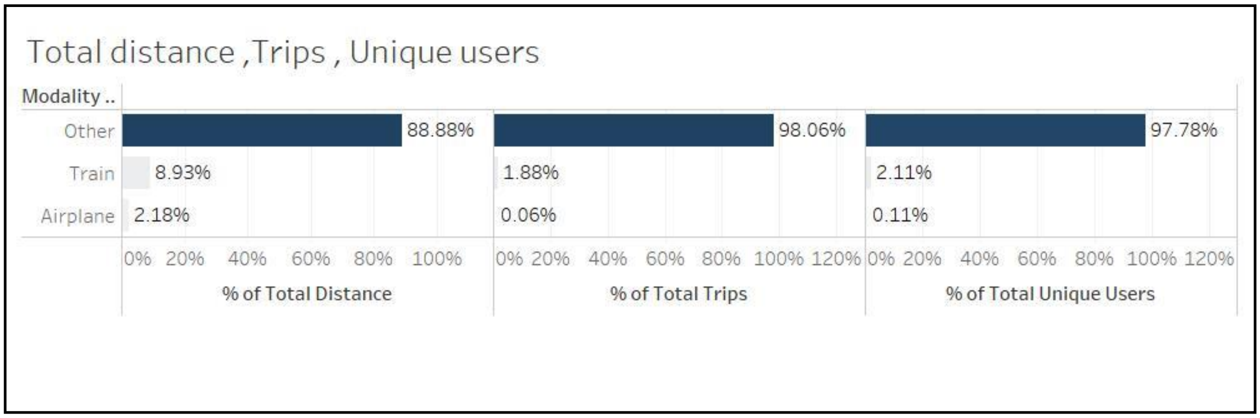


Figure 33_comparing three measures in different modalities.

These insights highlight the overwhelming dominance of the "Other" category across all three measures. This category includes a wide range of transport modes used very frequently. The train category, although significantly more utilised than aeroplanes, still falls far behind the "Other" category in all aspects. Meanwhile, air travels register the lowest figures in terms of distance, trips, and unique users, emphasising its relatively minor role in comparison.

4.2 Analysis of Audimob Survey Data

This section aims to visualise the work conducted by ISFORT through the Audimob survey using 2023 data (the Italian acronym for the High Institute for Transport Education and Research).

Considering all the features of Audimob's dimensions and measures, a selection of key aspects has been visualised in Tableau. These visualisations aim to produce comprehensive insights, with a sample of them that is presented in the following section.

Number of trips disaggregated by time

Figure 34 shows a clear variation in the number of trips across different times of the day. The highest percentage of trips occurs between 9:01 AM and 2:00 PM, accounting for about 33% of the total trips throughout the day. Following this, the next highest number of trips takes place between 2:00 PM and 6:00 PM, making up approximately 27% of the total. In the early morning, from 5:00 AM to 9:00 AM, the number of trips drops to around 23%, indicating a decrease compared to the afternoon period. The lowest percentage of trips is observed after 6:00 PM, with a significant decline, comprising only 16% of the total trips of the day.

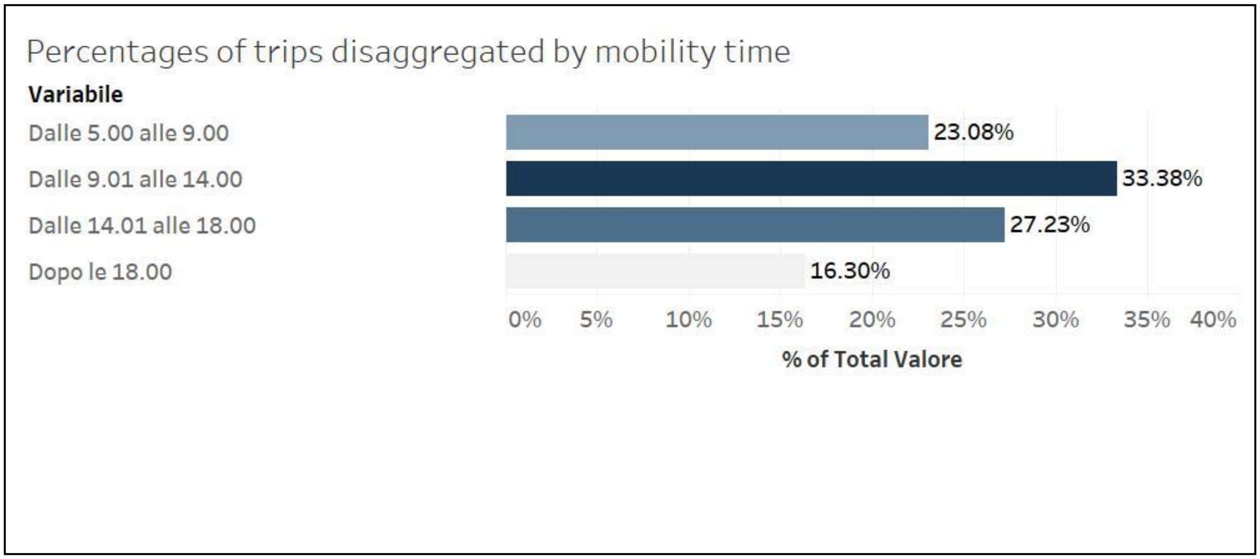


Figure 34_ Number of trips by mobility time

Number of Trips by distance

Figure 35 shows the percentage of trips disaggregated by distance class in four groups. The majority of trips are characterised by a short distance, between 2 km and 10 km, 45% of all the trips. Then there were trips less than 2 km, with 30%. Trips more than 10 km are less popular. There were 21% trips from 10 to 50 km and around just 2% of trips more than 50 km.

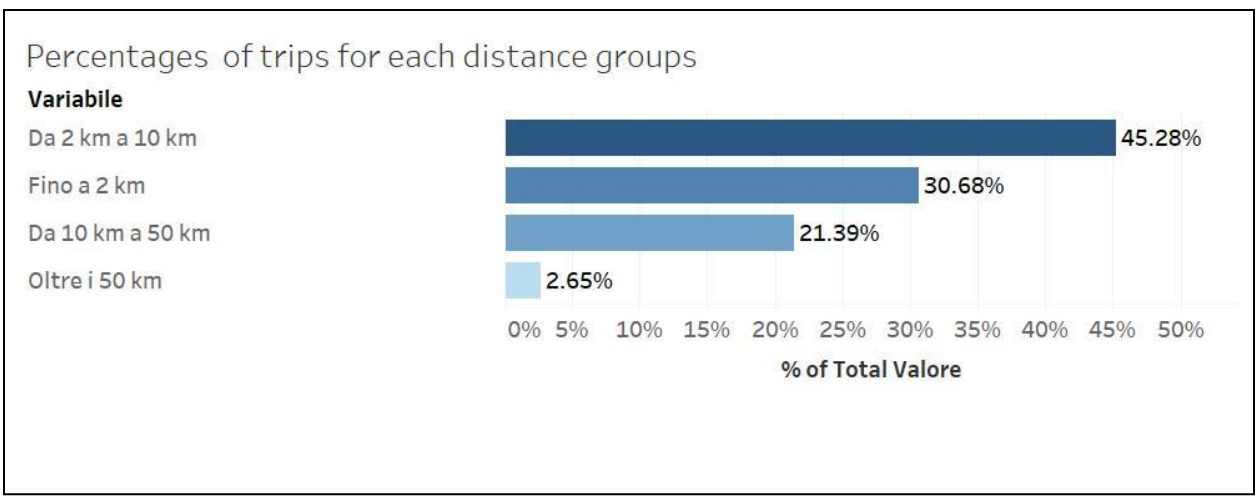


Figure 35_Number of trips disaggregated by distance

Trips by purpose

One key factor influencing travel behaviour is the purpose of the trip. In Figure 36, we can observe how the most common reason for travelling is "returning home," which is approximately 46 % of all trips. Other reasons, such as family management, work, and leisure, have a similar number of trips, with only slight variations among them. In contrast, the purpose of “study” is significantly lower than the others, with around 17%**.**

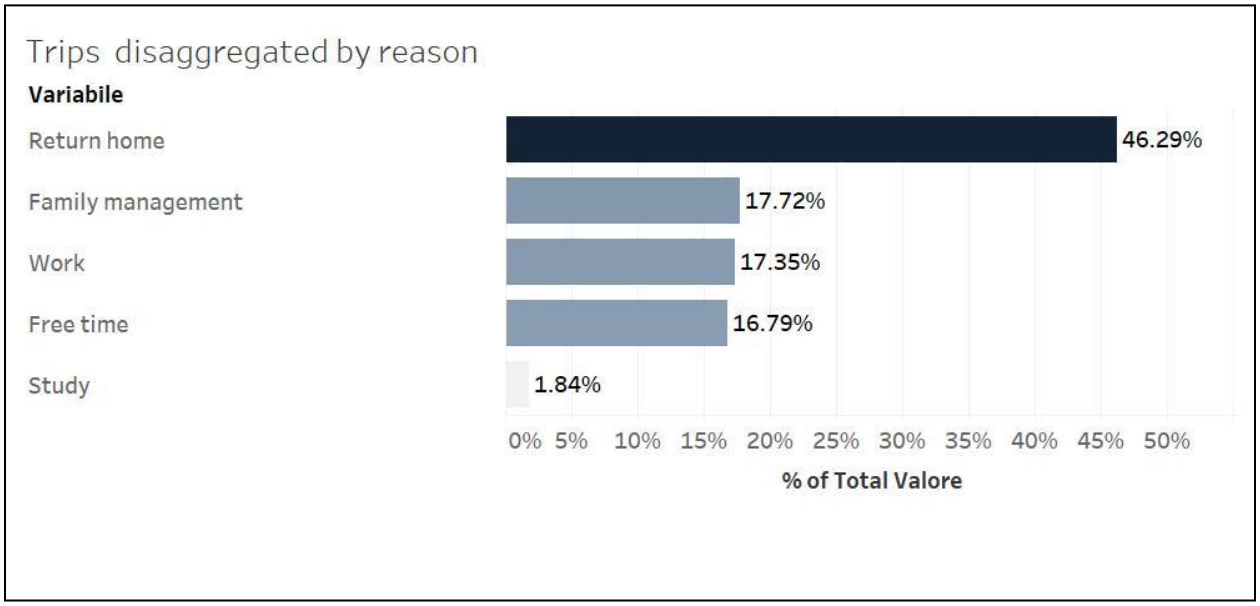


Figure 36_ Trips disaggregated by reasons

Percentage of trips disaggregated by modality

Figure 37 displays the percentage of total trips made using different types of transport modes. The dominant mode is car/other private vehicles, which accounts for 64.52% of all trips, indicating a heavy reliance on private cars. Walking follows as the second most common mode, making up 18.57% of trips. Other modes, such as urban public transport and motorcycles, each represent around 4% of trips, while bicycles are slightly lower at 3.92%. The remaining modes, including public transport combinations, trains, and long-distance travel, account for very small portions, with most of them contributing less than 2%. Long-distance trains are the least used, with only 0.01% of trips. This breakdown highlights the overwhelming preference for private vehicle use.

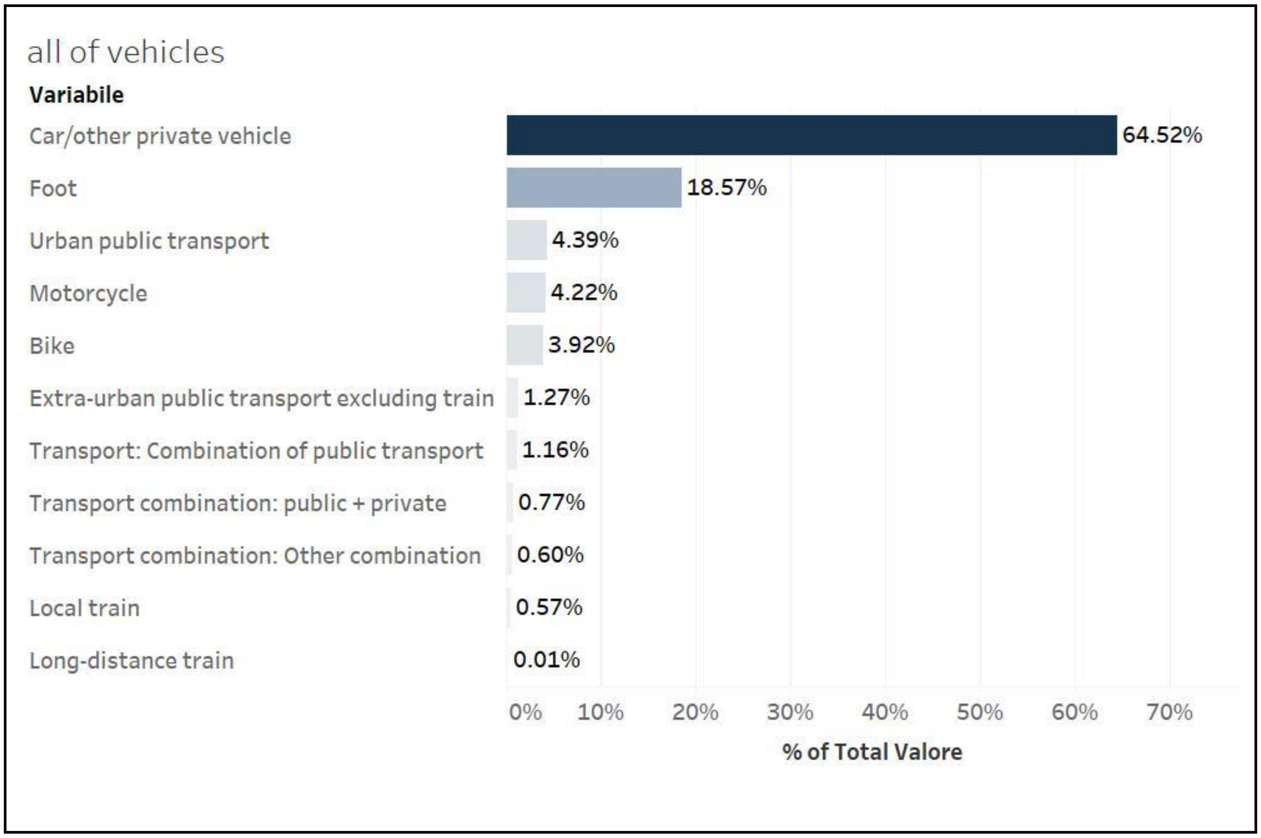


Figure 37_ percentages of each travel mode

Percentage of trips by transport mode and gender

Figure 38 highlights preferences by gender of different transport modes. Both men and women prefer private vehicles, with around 64.5% of trips made by car or other private modes, showing a high reliance on personal car. Walking is the second most common mode, particularly among women (21.4% compared to men's 15.7%), suggesting women may opt for short-distance travel more often. Men, however, are more inclined to use motorcycles and bicycles(6.3% and 4.8% for men, versus 2.1% and 3.1% for women, respectively). Public transport usage is low overall, but women more commonly use urban public transport than men (4.9% versus 3.9%). Less popular modes, like extra-urban public transport, transport combinations, and trains, account for less than 2% of travel across genders.

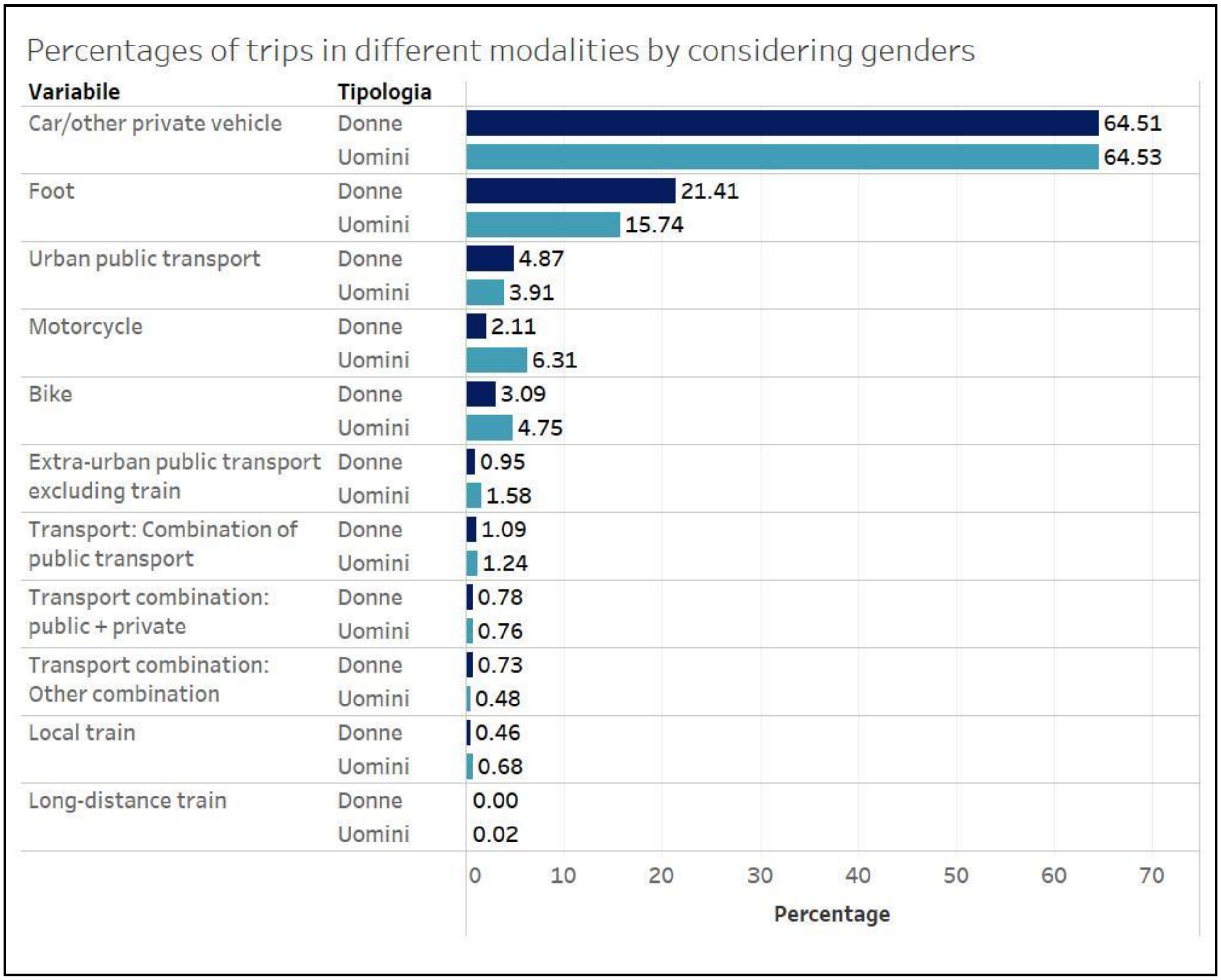


Figure 38_ Percentages of trips in different modalities by considering gender

4.3 Comparison between Vodafone and Audimob data

The two datasets differ in their sources and types of information, providing complementary insights into transport dynamics. Two datasets for the year 2023 will be compared to have a more comprehensive understanding of travel trends and mobility patterns. Figure 39 summarises the different dimensions used for comparing the two data sets in more details.

dimensions	Vodafone data	Audimob data
Age classes	• 14-30 • 31-59 • 60+ • hidden	• 14-29 • 30-45 • 46-60 • 61-84
Day of the week	• All days of the week	
Days	• Every day of the month	• Average working days
Destination	• 3000 zones	• Urban trips • Extra urban trips
Level of detail in the country	• North • South • Centre • Island • Provinces • Regions	• South and island • Centre • Northwest • North-East
Distance	• All kilometres (in distance intervals) • Hidden	• 0-2 • 2-10 • 10-50 • 50<
Travel Time	• 00:00-04:59 • 05:00-08:59 • 09:00-12:59 • 16:00-19:59 • 20:00-23:59 • Hidden	• 5:00-9:00 • 09:00-14:00 • 14:00-18:00 • 18:00<
Mode	• Train • Aeroplane	• Foot • Bike

	- Other - Hidden	- Motorcycle - Car/Car/another private vehicle - Local train - Long-distance train - Urban public transport - Extra-urban public transport excluding train - Transport combination 1: combination of public transport only - Transport combination 2: combination of public transport + private transport - Transport combination 3: other combination
Monthly data	All of the months	_
Nationality	- Italian - Strangers - hidden	_
Residential information	- Residence elsewhere - Residence in the destination - Residence in origin - Hidden	_
Trip frequency for each mode	- Systematic - Non systematic - Hidden	- All days, 3-4 days of the week - Sometimes in the month, year - Never - Systematic, 3-4 days in the week

		• Non-systematic, 1-2 times a week or more occasionally
Type of the day	• Working days • None working days • Hidden	• _
Unique traveller in comparison of whole population	• Travellers • Stationary	• Share % of mobile population
Duration of trip	• Time	• >15mins • 16-30 mins • 31-60 mins • >60mins
Purpose of travel destination	_	• Work • Study • Free time
User satisfaction	_	• User satisfaction information
User opinion	_	• Willing to change of modality
User information	_	• Type of work (smart work, office work) • How many people own a licence in the household

		• Household members • Other trips, future trips • Willing to continue using the same transport mode
Unit of measurement of sustainability	_	• Sustainable mobility • Fuel type
Professional condition	_	• Worker • Student • Unemployed, Housewife, Retired, etc.
Number of household members	_	• <2 • 3 • >3
Professional position	_	• Managers and executives • Other employees • Entrepreneurs, freelancers, self-employed, traders and others on their own account
Educational level	_	• None/Elementary/Middle school • High school diploma • University diploma-degree, post-university
Gender	_	• Male • Female

Population size	-_	• <5000 • 5001-20.000 • 20.001-50.000 • 50.001-250.000 • >250.000
measures	• Total distance • Number of trips • Time • Unique user	• Number of trips • car occupancy coefficient • Distance

Figure 39_Comparison of the parameters included in the datasets

Analysis of transport data uses various data sources to understand mobility patterns and travel behaviours. Both datasets, Vodafone and Audimob, offer complementary insights into transport dynamics. Audimob data are particularly detailed in capturing trip characteristics during working days. They provide insights into the distinction between weekdays and weekends, and macro time trends and classify trips by their distance and time interval, which helps understand the distribution of trip lengths and durations. In contrast, Vodafone data is deeper in the time factors like yearly, monthly, weeks and days. This allows for a more continuous monitoring of travel behaviours across different times.

The unit of measurement differs significantly between the two datasets. Audimob data focuses on the number of trips, the percentage of trips by distance, and trips by time classes. This approach offers a detailed view of the distribution of trips based on their characteristics. On the other hand, Vodafone data includes the number of trips, trips by distance, and passenger kilometres, providing a quantitative measure of travel volumes and distances covered.

Motivation for trips is another area where Audimob data provides richer information because classifies trips based on their purpose, distinguishing between systematic and non-systematic trips. Vodafone data, however, do not explicitly include trip purpose. Concerning transport mode, Audimob data include a wide range of transport modes such as walking, biking, motorbiking, cars, and public transport as well as trip chains. This diversity allows for a detailed analysis of different modes of transport. In contrast, Vodafone data focuses more on trains, aeroplanes, and other modes (bus, private car...) which may offer insights into long-distance and intercity travel patterns.

Sustainability metrics are explicitly addressed in Audimob data, evaluating sustainable mobility through percentages and fuel types. This is crucial for understanding the environmental impact of travel behaviours while Vodafone data do not specifically address sustainability. Customer opinions are another strength of Audimob that collects detailed information on customer satisfaction, safety from contagion risks, willingness to change modes and personal life aspects. This qualitative data can provide insights into user experiences and preferences. Vodafone data do not incorporate customer opinions directly, limiting its ability to provide such nuanced insights.

Both datasets provide information on the area of travels. Audimob data classify trips by distance classes (urban, large, medium, long distance).Vodafone data specifies the origin, destination, and code zones, extending over 3000 zones. This granularity in Vodafone data allows for a detailed spatial analysis of travel patterns.

Customer demographic information is another area of difference. Audimob data include age classes, nationality, gender, and number of household members, how many household members own driver's licenses. These demographic data are useful for understanding the travel behaviours of different population segments. Vodafone data include age groups and nationality, offering a less detailed but still useful demographic perspective.

In the analysis, Vodafone mobility index data divide the day into one-hour intervals, recording travelled points based on information from antennas. The central core, or the centre of these points, is calculated for each hour. By connecting the central core of one hour to the central core of the next, the distance travelled is determined. Trips are defined as distances greater than 800 metres within an hour. This method offers significant advantages, primarily due to its ability to provide real-time insights. The capacity to collect and analyse data in real-time enables timely insights into travel behaviour, making it a valuable tool for understanding and responding to mobility patterns. However, the approach is not without its drawbacks. One major concern is privacy. Traditional surveys can pose privacy issues because they involve interviews with individual respondents, but these rea managed for a long time not causing any problems. Instead, the use of mobile phone data introduces more complex privacy challenges. Even when these data are anonymised, they can still raise significant concerns due to the potential for hidden patterns or sensitive information to be revealed unintentionally. Additionally, in the big data resource, the data processing required to analyse large volumes of data from multiple sources is complex and IT resource intensive. The reliance on continuous data collection also means that signal issues and data availability can affect the reliability of the data.

Audimob on the other hand, employs traditional, well-established survey methods with a track record of reliability. One of the key advantages of this approach is that it generally poses fewer privacy risks compared to mobile data collection. However, there are notable disadvantages as well. For instance, an Isfort dataset requires a full year of manual processing for the interviews, making it quite costly. Due to the high expenses, the number of interviews is limited to around 16,000, as the cost increases with the number of respondents. Consequently, an Isfort dataset tends to be more expensive than that obtained from Vodafone. Additionally, traditional survey methods may not capture travel routes and times as accurately as mobile data, resulting in less precise insights. Moreover, these methods may not provide real-time data, resulting in potential delays in data analysis and insights. Conducting surveys can also be resource-intensive, requiring significant human resources and time. Despite these challenges, traditional survey methods remain reliable for collecting travel behaviour data, particularly in contexts where privacy concerns and cost considerations are paramount.

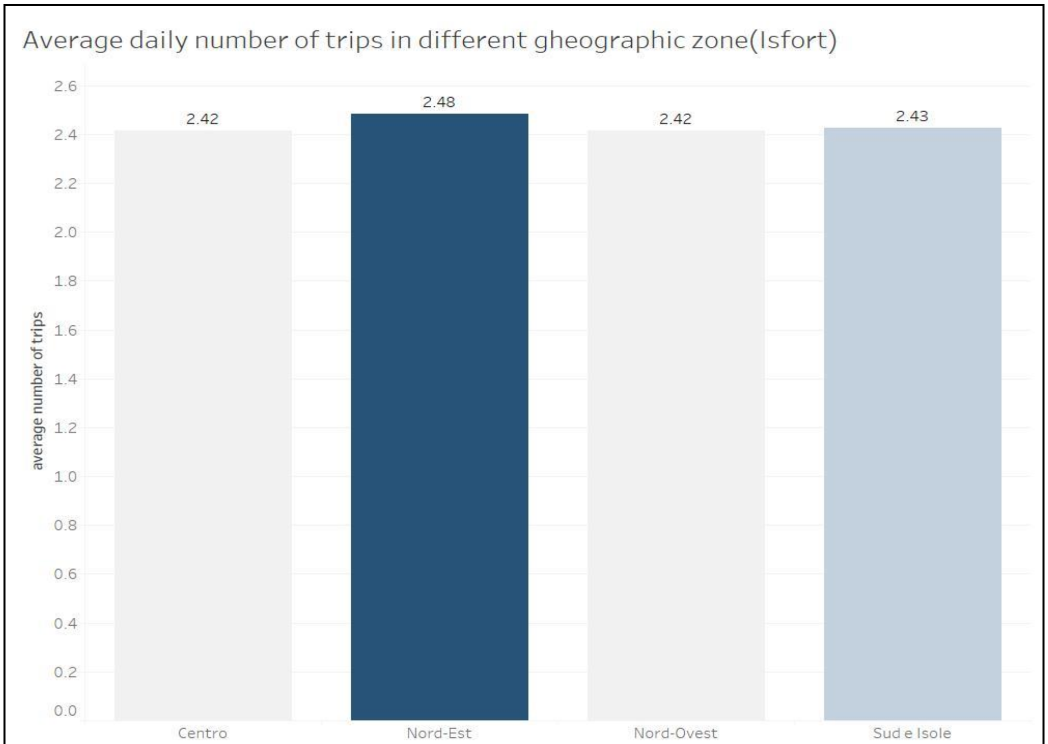
Both Audimob and Vodafone datasets offer valuable insights into transport dynamics, each with its unique strengths. Audimob data provide detailed qualitative insights and sustainability metrics, while Vodafone data offer extensive quantitative measures and spatial granularity. On the other hand, it is important to highlight that there are some differences between data provided by these two data sources, as shown in Figure 40 which shows the comparison in terms of distance per person and total number of travellers. Vodafone users travel an average of 47.04 km daily, almost twice than Isfort's (25.6 km). Additionally, 74.87% of Vodafone's unique users are travellers, compared to 81.81% in the Isfort dataset, suggesting a higher proportion of travellers in the Isfort database.

dimensions (considering 2023)	Vodafone data	Isfort
Distance average per person	47.04	25.6
Total travellers out of total population	74.87%	81.81%

Figure 40_Comparison of two data sets

Average Number of trips in different zones

Figure 41 reports average daily trips across the same four geographical zones considering both the different data sources (Isfort versus Vodafone). According to the figure, they slightly differ in values and distribution.



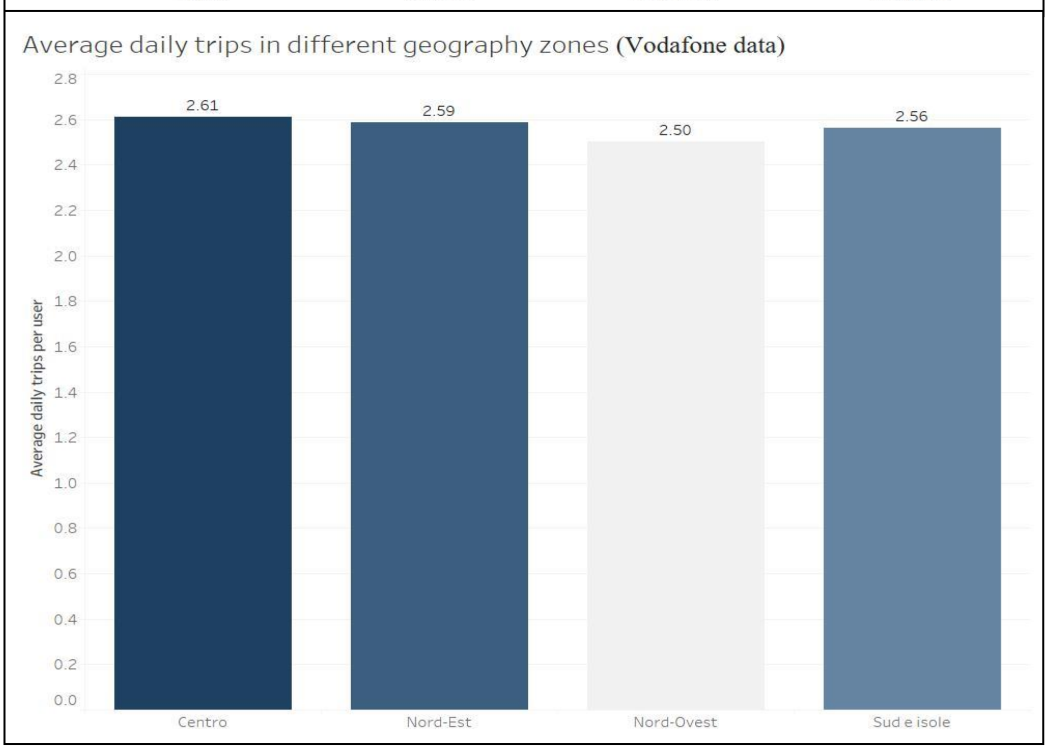


Figure 41_ Comparing the average number of trips in different geographic zones

In the Isfort data (left chart), the range of average trips is narrower, with all regions clustered around 2.4 trips per day. Nord-Est has a slight lead at 2.48 trips, but overall, there is a minimal regional variation. In contrast, the Vodafone data (right chart) shows slightly higher average trip counts overall, with Centro reaching 2.61 daily trips, followed by Nord-Est at 2.59. The spread across regions is also slightly wider here, indicating more noticeable regional differences in daily travel frequency compared to the Isfort data. Figure 42 presents a map illustrating the average daily trip across all regions of Italy.

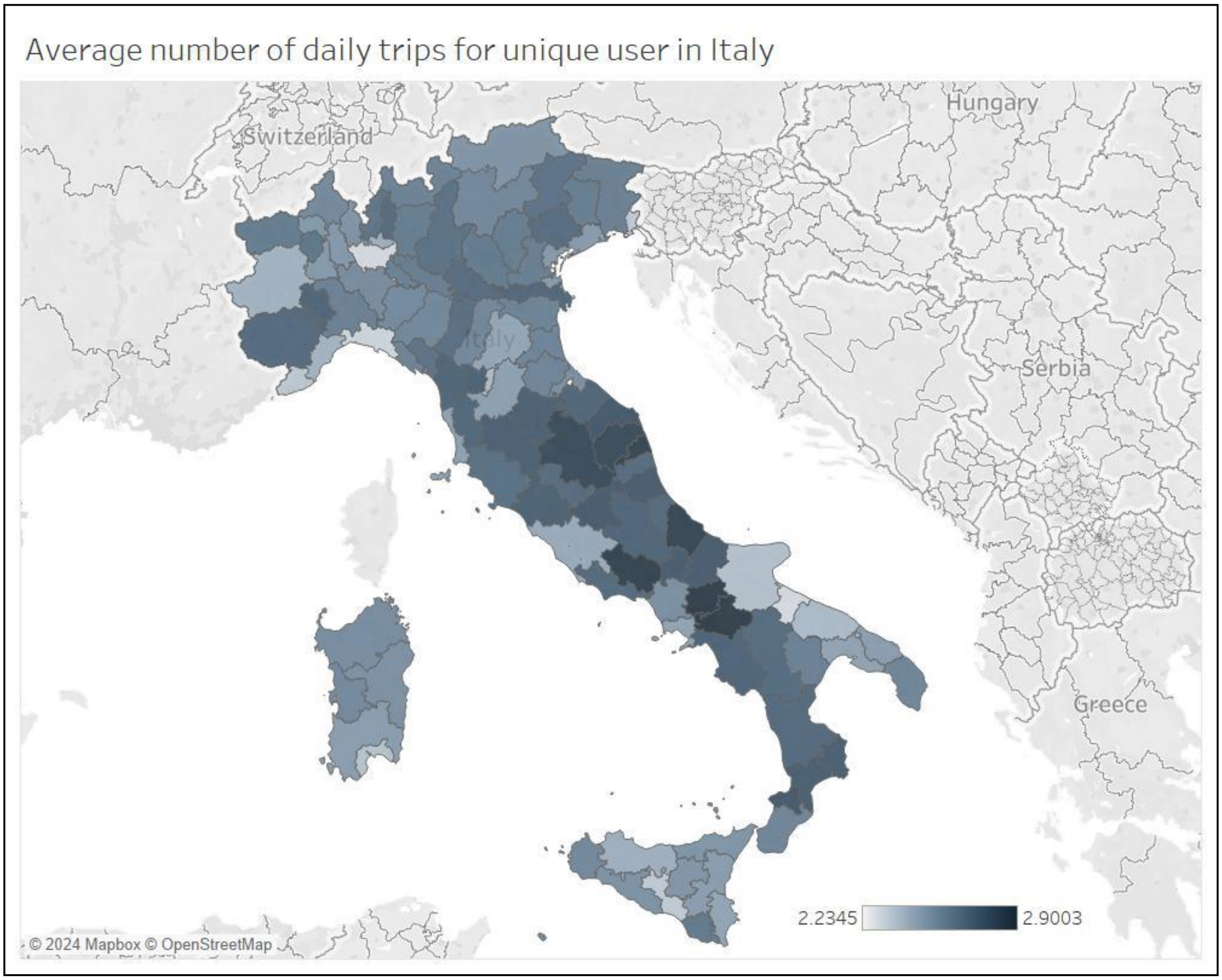
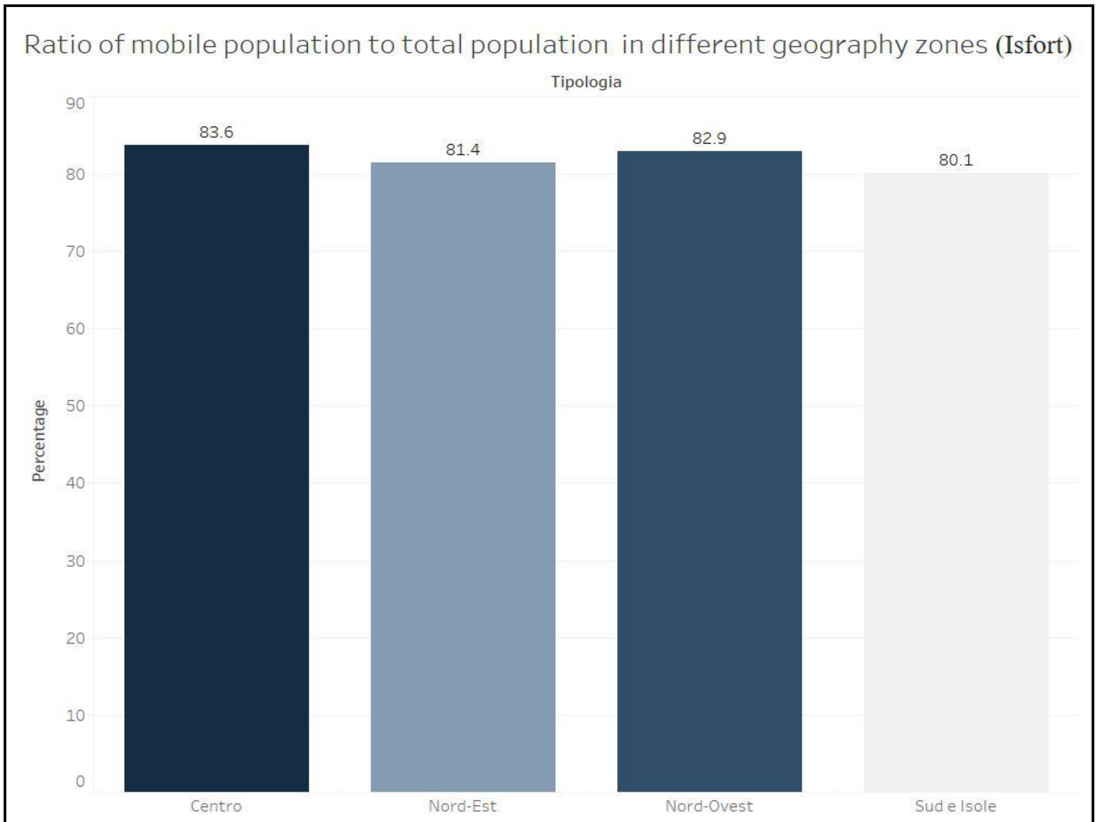


Figure 42_ Average number of daily trips for unique user in Italy 2023

As shown on the map, the majority of daily trips take place in central Italy, followed by northern Italy.

Geography and trip percentages

Figure 43 shows the trip distribution across Italian geographical zones considering both data from Isfort and Vodafone.



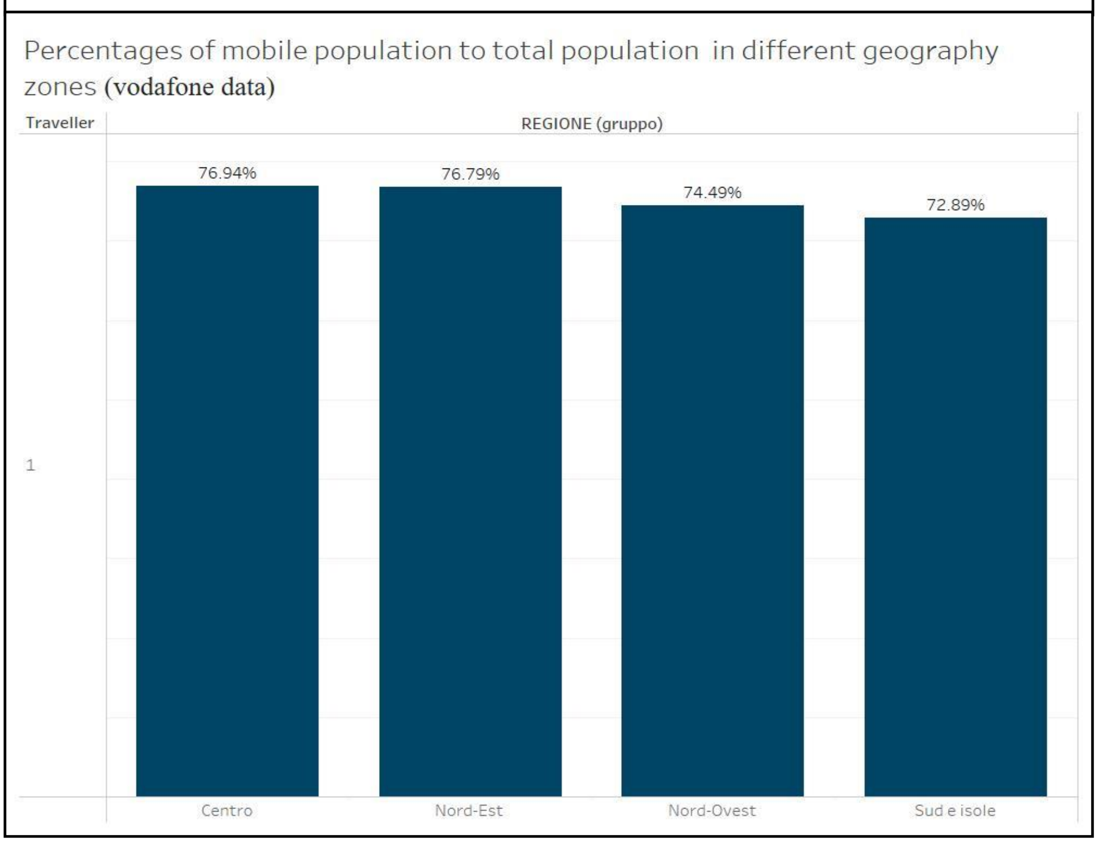


Figure 43: comparing trip distribution across geographical zones in both datasets.

On the left (Isfort data), "Centro" has the highest percentage of trips at 83.6%. The differences among the zones are minimal, it is a fairly balanced trip distribution. The right chart (Vodafone data) "Centro" again leading at 76.9% and "Sud e Isole" lowest at 72.9%. These percentages are slightly lower than in the Isfort data but still show a similar trend of even distribution. Overall, both datasets highlight Centro as the most active travel zone and Sud e Isole as the least active, suggesting similar regional mobility patterns.

Number of trips disaggregated by distance group

Figure 44 shows distance groupings with slight differences in labelling and distribution. It compares the number of trips by distance groups, based on data from Vodafone and Isfort in 2023.

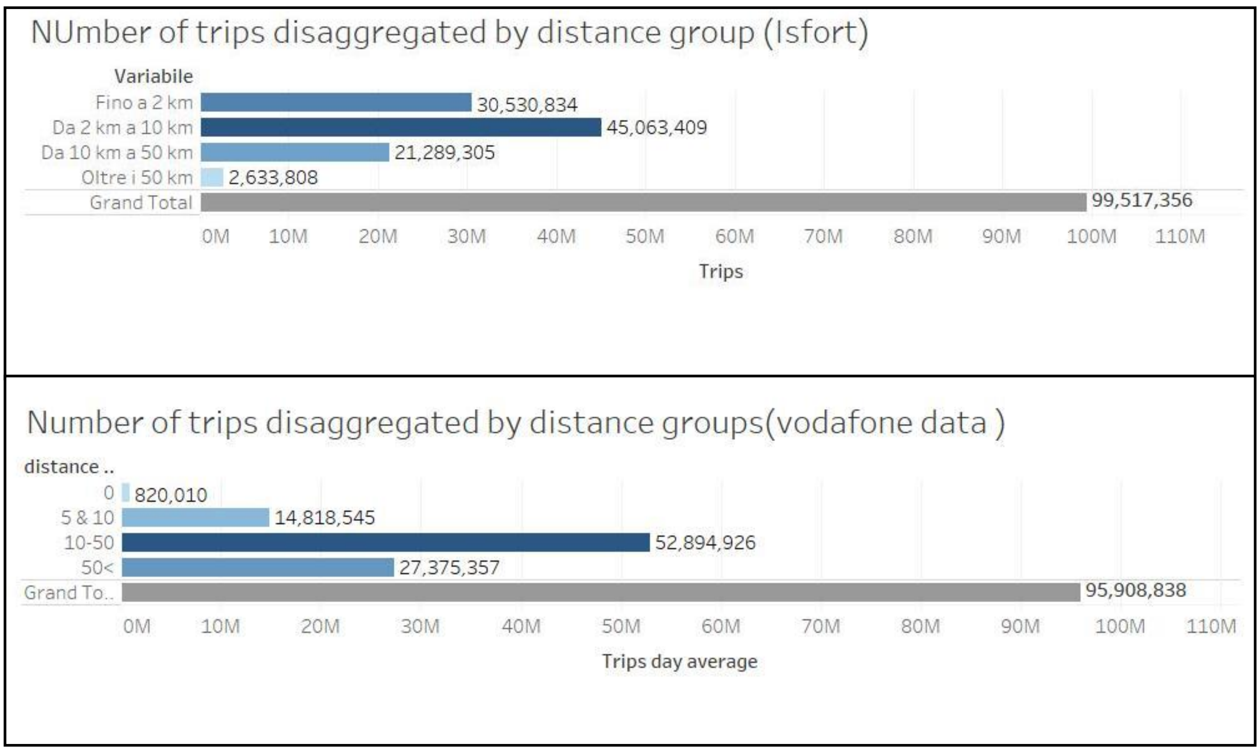


Figure 44: Number of trips disaggregated by distance group in both datasets

. When comparing the distribution of trips across these distance groups, the Isfort data report significantly more short-distance trips, with 30,530,834 trips recorded up to 2 km and 45,063,409 trips between 2 and 10 km. In contrast, the Vodafone data have a combined total of only 15,638,555 trips for distances up to 10 km. For the 10-50 km range, Vodafone recorded 52,894,926 trips and Isfort 21,289,305 trips. This group has the highest number of trips in the Vodafone data set. However, there is a big difference in the long-distance category (over 50 km): Vodafone reports

27,375,357 trips, while Isfort only records 2,633,808 trips. This disparity suggests that Vodafone data include more long-distance trips, whereas Isfort data emphasises shorter trips.

4.4 SWOT for both sources

The SWOT table presented in Figure 45 outlines the characteristics of Big Data and Audimob data across four categories: strengths, weaknesses, opportunities, and threats.

	Strengths	**Weakness**
Big data based data collecting	• Capturing large volumes of data. • Monitoring real-time data. • Improving the accuracy of data in certain types. • Using periodic large amounts of data to predict and identify trends in travel behaviour. • Reducing the need for extensive field surveys and manual data collection. . • Providing monthly updates with recent data and minimal delay. • Covering extensive areas and hard-to-reach places where manual data collection is difficult. • Providing data specific to each day of the week. • Covering the entire territory of the country at the same time. • Continuously monitoring human movement.	• Raising data privacy concerns and security problems. • Incurring high costs for preparing and buying data from mobile operators. • Requiring advanced technical expertise for data analysis and management. • Dealing with overwhelming amounts of data. • Experiencing a lack of data specifically for each and all transport modes. • Using vast storage space for data retention. • Working with both useful and irrelevant data to extract valuable information. • Incurring high costs for maintaining data. • Demanding considerable time for data categorisation and cleaning.

	Opportunities	Threats
	• Developing new, data-driven transportation models and services. • Providing more user-friendly services. • Shaping public transportation policies and infrastructures.	• Impacting data collection through rules in data privacy and security. • Making data useless due to data privacy concerns and large amounts of hidden data. • Harming policy-making and planning due to inaccurate and imprecise data. • Enforcing strict data protection policies, such as Directive 95/46/EC in Europe, on personal data processing and free movement of such data.
	Strengths	**Weakness**
survey-based data collecting	• Establishing direct relationships with users. • Providing detailed information. • Providing precise information. • Providing trustworthy data. • collecting all transportation accurately. • Targeting specific questions and objectives. • Gathering both quantitative and qualitative data.	• Providing a smaller amount of data. • Being time-consuming for data collection. • Being money-consuming for data collection. • Affecting response based on self-attitude. • Achieving all answers to survey questions is not possible if participants refuse to answer. • Experiencing low participation and non-response bias. • Requiring significant manual effort for data entry and analysis.

		• Detecting tourists' travel behaviour during their short stay is generally unreachable. • Failing to capture rapid changes with these methods.
	Opportunities	**Threats**
	• Advancing survey methods to enhance data quality and reach. • Combining survey data with other data sources to provide comprehensive insights. • Using surveys as a tool to engage with the community and stakeholders, fostering a participatory approach.	• Becoming useless as technology continues to enhance. • Being affected by economic factors that impact the continuation of these approaches.

Figure 45: SWOT table for both datasets

5. Discussion and conclusions

This thesis aimed to analyse and compare the characteristics of two datasets provided by FSI to identify mobility patterns. The first data set was bought by FSI from Vodafone; the large-scale data set provided metrics such as distance, time, geographic zones, and temporal details like months, weeks and days, etc. The second data set was obtained by FSI from ISFORT (Audimob survey); it included measures with more details compared to Vodafone one, like gender, user satisfaction, specific geographic distance, reason of trips, etc. The datasets were analysed to highlight their unique attributes and the results were visualised using Tableau software. Some limitations affected both data sets. One of the main issues related to big data (Vodafone) was "user privacy" related to the non-availability of disclosing some details by Mobile Network Operators, app aggregators, and data analytics companies. The primary concept is to integrate this protection by design, ensuring that individual data are never provided in a manner that could be traced back to a particular person. Due to privacy regulations such as GDPR, a significant amount of data is concealed. For instance, with Vodafone's big data in October 2021, details about 8,745,485 trips that will cover 1.09% of total trips were hidden based on the mode of transport. Additionally, 67,222,937 trips were hidden information on the timing of the trips, that is 8.34% of total trips; 125,660,544 trips (15,59%) were hidden concerning whether they were systematic or not. Although filters may hide a lot of data, big data still provides a vast amount of usable information for decision-making. However, while hidden data can be a significant issue, it can be mitigated by reducing the number of dimensions and categories in the data. The likelihood of obscuring data is reduced by simplifying the data and avoiding excessive detail. Vodafone Analytics enforces a privacy rule that prohibits displaying estimates below 15 units in analyses. Analysing multiple dimensions simultaneously can complicate results, as excessive data splitting may mask important units.

The inherent characteristics of mobile phone data present significant challenges in accurately identifying different modes of transport, such as distinguishing between cars, buses, as well as bicycles and walking in densely populated urban settings. The sheer volume of data adds another layer of complexity, making tracing the exact routes difficult. Furthermore, short trips – those less than 800 metres or lasting less than one hour – and trips within the same geographic area are particularly hard to differentiate. These constraints can introduce inaccuracies and affect the overall reliability of the results. In addition, FSI faces significant challenges due to the high costs of purchasing comprehensive datasets every month. As a result, there are noticeable gaps in continuous data availability, which poses a challenge for thorough analysis. Without consistent data across all months, conducting reliable and accurate assessments or tracking trends over time becomes difficult.

A significant challenge when working with big data was understanding the underlying algorithms used for data processing. Although mobile phone data providers offered useful descriptions and information about data categorisation and cleaning, many details remained proprietary to protect their business interests. This lack of transparency complicated data analysis, as it limited a full understanding of the processing methods, making it difficult to draw reliable conclusions. As an example, the "transport mode" variable is considered less reliable because it is not derived directly from raw data but is instead inferred using algorithms. These algorithms are continuously evolving, meaning that criteria and methods for identifying transport modes, such as trains and aeroplanes, can change over time. This dynamic nature introduces variability, which can affect the consistency and accuracy of the data, making comparisons and trend analyses less reliable.

On the other hand, working with survey-based data, such as Audimob, also posed challenges, particularly in obtaining detailed temporal information, like data specific to certain months or individual days. Surveys typically do not capture this level of detail, which limits their capacity to provide in-depth, time-sensitive insights.

For instance, Audimob cannot offer data granularity because integrates detailed information at the monthly level with distance travelled. Audimob classifies distances into fixed groups, whereas data from Vodafone allows for flexible grouping of distances. This flexibility enables us to analyse changes in travel patterns across specific months, an ability that Vodafone data provides but is not possible with traditional survey methods like Audimob. As previously mentioned, both datasets had their own advantages and disadvantages. However, when comparing these two primary datasets, it became clear how significantly they could yield different results. The discrepancy in data granularity means that even if the volume of data collected from each source differs significantly, aligning these data sets by the same time periods is problematic. Consequently, integrating data from different operators to gain comprehensive insights was often impractical and could result in inaccurate findings.

The lack of alignment in both the amount and the specific time periods of the data makes it difficult to compare the two methods straightforwardly. The results of the analyses reveal that while certain comparisons yield similar results across both datasets, others show significant differences. These variations suggest that relying on a single dataset may lead to incomplete or potentially inaccurate insights, emphasising the value of using both datasets to gain a more comprehensive view. The analysis also identified that not all attributes are available in both datasets.

Therefore, the choice of dataset depends on the specific goals. Each dataset offers unique advantages; however, using them together would provide the most comprehensive understanding of travel behaviour. Moving forward, developing an effective method for integrating both datasets is essential to leverage their combined strengths fully.

References

1. Axhausen, Kay & Zürich, ETH. (2007). Concepts of Travel Behaviour Research.
2. Mwale, M., Luke, R., & Pisa, N. (2022). Factors that affect travel behaviour in developing cities: A methodological review. *Transportation Research Interdisciplinary Perspectives, 16*, 100683. Available online 22 September 2022. https://doi.org/10.1016/j.trip.2022.100683
3. Subbarao, Saladi Sv & Rao, K V. (2020). Activity based approach to travel demand modelling: An overview. European Transport\Trasporti Europei. 1-16.
4. DARIDO, G. Concerns over the growth of motorcycles in cities. 2012 [cited 2022 13th June]; Available from: https://blogs.worldbank.org/transport/concerns-over-the-growth- of-motorcycles-in-cities.
5. Dissanayake, Dilum & Morikawa, Takayuki. (2002). Household Travel Behaviour in Developing Countries: Nested Logit Model of Vehicle Ownership, Mode Choice, and Trip Chaining. Transportation Research Record. 1805. 45-52. 10.3141/1805-06.
6. Firdausi, M., Ahyudanari, E., & Herijanto, W. (2023). Study on the analysis of travel behaviour: A review. *E3S Web of Conferences*, 434, 02022. https://doi.org/10.1051/e3sconf/202343402022
7. Cao, Jason & Mokhtarian, Patricia & Handy, Susan & Ltd, Francis & Sgm,. (2009). Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. Transport Reviews. 29. 359-3951464. 10.1080/01441640802539195.
8. Cao, X., Mokhtarian, P. L. and Handy, S. L. (2008) Examining the Impacts of Residential Self-Selection on Travel Behaviour: Methodologies and Empirical Findings. Research Report No. UCD-ITS-RR-08-25 (Davis, CA: University of California – Davis). Available at: http://pubs.its.ucdavis.edu/ publication_detail.php?id=1194 (accessed 1 August 2008).
9. Firdausi, M., Ahyudanari, E., & Herijanto, W. (2023). Study on the analysis of travel behaviour: A review. E3S Web of Conferences, 434, 02022. https://doi.org/10.1051/e3sconf/202343402022

10. Shariff, Norhazlin & Shah, Muhammad. (2008). Factors influencing travel behaviour and their potential solution : a review of current literatures.

11. Servizi, V., Pereira, F. C., Anderson, M. K., & Nielsen, O. A. (2021). Transport behaviour-mining from smartphones: a review. *European Transport Research Review*. https://doi.org/10.1186/s12544-021-00516-z

12. Graells-Garrido, E., Opitz, D., Rowe, F., & Arriagada, J. (2023). A data fusion approach with mobile phone data for updating travel survey-based mode split estimates. *Transportation Research Part C, 155*, 104285. https://doi.org/10.1016/j.trc.2023.104285

13. Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11, 141–155.

14. Goulias, K. G. (2016). Travel behaviour models. In D. R. Montello (Ed.), *Handbook of Behavioural and Cognitive Geography* (Chapter 5). GEOTRANS, University of California, Santa Barbara, CA, USA.

15. City of Seattle & Puget Sound Regional Council. (2019). *Introductory guide to household travel survey data*. RSG, Inc. https://www.rsginc.com

16. Lawson, C. T., Krans, E., Rentz, E. (G.), & Lynch, J. (2023). Emerging trends in household travel survey programs. *Social Sciences & Humanities Open, 7*(1), 100466. https://doi.org/10.1016/j.ssaho.2023.100466

17. Miller, E. J., Srikukenthiran, S., & Chung, B. (2018). Workshop synthesis: Household travel surveys in an era of evolving data needs for passenger travel demand. *Transportation Research Procedia, 32*, 374-382. https://doi.org/10.1016/j.trpro.2018.10.051

18. GSM Association. (2019). *Mobile big data: Solutions for a better future*. PricewaterhouseCoopers Australia. Retrieved from www.gsma.com

19. Xie, Jiyang & Song, Zeyu & Li, Yupeng & Ma, Zhanyu. (2018). Mobile big data analysis with machine learning.

20. Ubioworo, E. (2020). *SWOT analysis as a tool supporting the strategic management at DHL* (Master’s thesis, University of Economics and Human Sciences in Warsaw).

21. Namugenyia, C., Nimmagadda, S. L., & Reiners, T. (2019). Design of a SWOT analysis model and its evaluation in diverse digital business ecosystem contexts. *Procedia Computer Science, 159*, 1145–1154. Available online at https://www.sciencedirect.com/science/article/pii/S1877050919314219

22. Gürel, Emet. (2017). SWOT ANALYSIS: A THEORETICAL REVIEW. Journal of International Social Research. 10. 994-1006. 10.17719/jisr.2017.1832.

23. Sammut-Bonnici, T., & Galea, D. (2015). SWOT Analysis. In *Wiley Encyclopedia of Management* (Vol. 12). https://doi.org/10.1002/9781118785317.weom120103

24. Nayak, P., Thejas, R., Tejaswini, G., Vaishali, V., Shetty, V. M., & Varsha, A. M. (2023). A review paper on big data analytics in mobile networks. *International Journal of Advanced Research in Science, Communication and Technology*, 3(2).

25. J. Kolodziej and H. Gonz´alez-V´elez (Eds.): cHiPSet, LNCS 11400, pp. 163–185, 2019. https://doi.org/10.1007/978-3-030-16272-6_6

26. Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behaviour (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies, 68*, 285-299. https://doi.org/10.1016/j.trc.2016.04.005

27. Arunkumar, A., & Yogeswari, K. (2021). The challenges of using big data analytics in smart transportation. *OSR Journal of Engineering (IOSRJEN), 11*(5), 19-29. https://www.iosrjen.org

28. GSMA. (2017). *Mobile privacy and big data analytics*. https://www.gsma.com

29. Davis, K., & Patterson, D. (2012). *Ethics of big data*. O'Reilly Media, Inc.

30. Peng, Z., Bai, G., Wu, H., Liu, L., & Yu, Y. (2021). Travel mode recognition of urban residents using mobile phone data and MapAPI. *Urban Analytics and City Science, 48*(9), 2574-2589. https://doi.org/10.1177/2399808320983001

31. Essadeq, Imane; Janik, Thibault (2021) : Use of mobile telecommunication data in transport modeling: A French case study, International Transport Forum Discussion Paper, No.

2021/04, Organisation for Economic Co-operation and Development (OECD), International Transport Forum, Paris, https://doi.org/10.1787/90483afc-en.

32. Deutsch, K., McKenzie, G., Janowicz, K., Li, W., Hu, Y., & Goulias, K. (2012, July 15-20). *Examining the use of smartphones for travel behaviour data collection*. 13th International Conference on Travel Behaviour Research Toronto.

33. CNR, University of Pisa, & WINDTRE. (2020). *Mobile phone data analytics against the COVID-19 epidemics in Italy: Flow diversity and local job markets during the national lockdown*.

34. Sun, Y., Liu, C., & Zhang, C. (2021). Mobile technology and studies on transport behaviour: A literature analysis, integrated research model, and future research agenda. *Mobile Information Systems, 2021*, Article ID 9309904. https://doi.org/10.1155/2021/9309904

35. Breyer, N. (2019). *Analysis of travel patterns from cellular network data* (Licentiate thesis, Linköping University, Thesis No. 1845). Linköping Studies in Science and Technology.

36. Davenport, T. H. 2006. "Competing on Analytics," Harvard Business Review (84:1), p. 98-107.

37. Amoako, B. T. (2013). *The importance of business intelligence as a decision-making tool: Case study Electricity Company of Ghana (E.C.G)* (Master's thesis). University of Borås.

38. Siddiqui, Ahmad. (2021). Data Visualization: A Study of Tools and Challenges. 11. 18-23.

39. Zheng, Jack. (2023). Data Visualization for Analytics and Business Intelligence: A Comprehensive Overview.

40. Bharadiya, J.P. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. American Journal of Artificial Intelligence, 7(1), 24.

41. Srivastava, Deepmala. (2023). An Introduction to Data Visualization Tools and Techniques in Various Domains. International Journal of Computer Trends and Technology. 71. 125-130. 10.14445/22312803/IJCTT-V71I4P116.

42. Parthe, Rupesh. (2023). Comparative Analysis of Data Visualization Tools: Power BI and Tableau. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 07. 1-11. 10.55041/IJSREM26272.

43. Devadharshini, L., & Boomika, D. M. (2022). Comparative analysis of data visualization using tools in data science. *International Journal of Research Publication and Reviews, 3*(5), 2598-2600. https://www.ijrpr.com.

44. Parthe, Rupesh. (2023). Comparative Analysis of Data Visualization Tools: Power BI and Tableau. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 07. 1-11. 10.55041/IJSREM26272.

45. Taherdoost, Hamed. (2021). Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects.

46. Kabir, S. M. S. (2016). Methods Of Data Collection Basic Guidelines for Research: An Introductory Approach for All Disciplines (first ed., pp. 201-275).

47. He, S., Miller, E., & Scott, D. M. (2017). Big data and travel behaviour. *Travel Behaviour and Society, 12*, 100-106. https://doi.org/10.1016/j.tbs.2017.12.003

48. Methodological elements for the analysis of passenger mobility using Big Data, Ferrovie dello Stato Italiane Group ISTAT Ministry of Infrastructure and Transport, which year ?

49. Zorrilla Salgador, Juan Pablo. (2018). Data Analytics with Tableau: The Trend Lines Models. SSRN Electronic Journal. 10.2139/ssrn.3282727.

50. Andry, J. F., Tannady, H., Limawal, I. I., Rembulan, G. D., & Marta, R. F. (2021). Big data analysis on YouTube with Tableau. *Journal of Theoretical and Applied Information Technology, 99*(22), 5460. https://www.jatit.org