Determinants for the energy, environmental and cost impacts of mobility patterns in the Lisbon Metro Area

Jérémy Daniel Diacre

Thesis to obtain the Master of Science Degree in

Energy Engineering and Management

Supervisors: Prof. Filipe Manuel Mercier Vilaça e Moura

Dr. Patrícia de Carvalho Baptista

Examination Committee

Chairperson: Prof. Francisco Manuel da Silva Lemos

Supervisor: Prof. Filipe Manuel Mercier Vilaça e Moura

Member of the Committee: Dr. Catarina Maria da Cunha de Eça Calhau Rolim

December 2021

Abstract

Following the need to decarbonize and reduce the energy consumption in order to limit the effects of climate change, urban areas have begun to change their transportation policies in order to attain sustainable urban mobility. This notion ties together mobility, environment and sociology. The goal of this work is to find mobility or socio-demographic or economic indicators that have an impact on the energy consumption, the CO_2 emissions as well as the cost.

To do so, data resulting from the Imob survey realized in the Lisbon Metropolitan Area in 2017 was analyzed. A new dataset of 50 city pairs was created in order to find relationships between selected independent variables and dependent variables being both total and per passenger.km energy consumption, CO_2 emissions and cost. An exploratory data analysis was realized to determine which IV can fit the best in a following Multiple Linear Regression. It resulted that the share of public cars or of public transportation is almost always the most impactful indicator on per pkm DVs. It has also been determined that the share of people with a high education and the unemployment rate have an impact on the per pkm DVs. Concerning the totals DVs, as shown in the well-known IPAT identity, the population has a big impact on the number, but for this work with a surprising inverse relationship.

Keywords : Energy efficiency; Decarbonization; Multimodal emission factors;Sustainable urban mobility; AML

Resumo

Na sequência da necessidade de descarbonizar e reduzir o consumo de energia a fim de limitar os efeitos das alterações climáticas, as áreas urbanas começaram a alterar as suas políticas de transporte a fim de alcançar uma mobilidade urbana sustentável. Esta noção liga a mobilidade, o ambiente e a sociologia. O objectivo deste trabalho é encontrar mobilidade ou indicadores sócio-demográficos ou económicos que tenham impacto no consumo de energia, nas emissões de CO2, assim como no custo.

Para tal, foram analisados os dados resultantes do inquérito Imob realizado na Área Metropolitana de Lisboa em 2017. Foi criado um novo conjunto de dados de 50 pares de cidades a fim de encontrar relações entre variáveis independentes seleccionadas e variáveis dependentes sendo tanto o consumo energético total como por passageiro.km, as emissões de CO2 e o custo. Foi realizada uma análise exploratória de dados para determinar qual o IV que melhor se pode encaixar numa Regressão Linear Múltipla seguinte. O resultado foi que a percentagem de carros públicos ou de transportes públicos é quase sempre o indicador com maior impacto por pkm DVs. Foi também determinado que a percentagem de pessoas com uma educação elevada e a taxa de desemprego têm um impacto sobre os DV por pkm. Em relação aos totais de DV, como mostra a conhecida identidade do IPAT, a população tem um grande impacto no número, mas para este trabalho com uma surpreendente relação inversa.

Palavras-chave : Eficiência energética; Descarbonização; Factores de emissão multimodais; Mobilidade urbana sustentável; AML

Acknowledgements

I’d like to address a special thanks to Filipe Moura and Patricia Baptista who accepted to help me with this thesis considering the situation I was in.

I thank as well my family and friends for the support during these difficult times.

Table of contents

Figure Index

Table index

Acronyms

€ - Euro

CO2 – carbon dioxid

DV – Dependent Variable

EU – European Union

EDA – Exploratory Data Analysis

EV – Electric Vehicle

g – grams

ICT – Information and Communication Technology

IPAT – I = PAT identity

IPCC – Intergovernmental Panel on Climate Change

IUTP – International Association of Public Transport

IV – Independent Variable

km – kilometer

MJ – Megajoule

NUTS – Nomenclature of Territorial Units for Statistic

OD – Origin-Destination

PC – Private Car

pkm – passenger.km

PT – Public Transport

SUMP – Sustainable Urban Mobility Plan

TAZ – Traffic Analysis Zone

WBCSD – World Business Council for Sustainable Development

1. Introduction

1.1 Motivations

The IPCC Assessment Report published in 2014, was already damning and warned everyone about the effect climate change would have on our lives and was followed by the historic Paris Agreement in 2015. The last update provided in 2021 by the Working Group I contribution to the sixth Assessment Report some information that previews an even worse situation than previously thought. The ideal limit of increase in temperature by 1,5°C above pre-industrialisation levels will possibly be exceeded as soon as 2025. Even with more CO2 emissions limitations, the impact on the planet will be severe and irreversible (IPCC, 2021).

“Every fraction of a degree matters” insists the IPCC. To try to slow the process and limit the repercussions on mankind , it suggests a radical transformation on our societies on all levels. The European Union has tackled this issue by presenting the European Green Deal in 2019. The goal of this plan is to ensure that Europe becomes the first climate-neutral continent by 2050, safeguard biodiversity, create a circular economy and get rid of pollution while increasing the competitiveness of European industry (European Commission, 2019). To attain these goals the EU should first make sure to reach its objective of emissions reduction by 2030. These objectives have even been updated in 2020 to be set at 55% in emissions reduction instead of 40%. To be able to implement well the measures needed to attain the goal fixed the EU presented in 2021 a series of legislative proposals under the name ‘Fit for 55’ (European Commission, 2021). Some of these directives impact the transportation field.

This latter is indeed responsible for around 25% of the global CO2 emissions (AIE, 2020) and will need to be the first to experiment these drastic changes. Transport represents a crucial part of everyday life and efficient and affordable operation is necessary for a good quality of life. With growing social concern and in order to achieve a more sustainable future, proper management of transport is necessary for the population especially in urban areas, where the impact of pollution is felt the most. Urban scenery also offers a large diversity of transports available for people and that’s why the changes to decarbonize should be made first in cities as it will be easier to begin with. Attaining sustainable urban mobility is a key to reaching the goals fixed to drastically reduce our emissions.

The concept of sustainability was first introduced in 1987 in the World Commission on Environment and Development, known as the Brundtland report. Sustainable Development is described as a way of development to “ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs” (WCED, 1987). The definition has evolved in different ways ever since and there is now a lack of consistent definitions in the literature (Moore et al. ,2017). Moore et al. tried in 2017 to develop a comprehensive definition of sustainability based on those available in the literature. Their revised definition is the following : “ [Sustainability is when] after a defined period of time, a program, clinical intervention, and/or implementation strategies continue to be delivered and/or individual behavior change (i.e., clinician, patient) is maintained; the program and individual behavior change may evolve or adapt while continuing to produce benefits for individuals/systems." (Moore et al., 2017).

The definition is broad and can be applied to a lot of different subjects and transportation is one of them. In 2001, the EU adopted in the so-called April Resolution during the meeting of the EU Council of Ministers a definition of a sustainable transport system. This latter should ensure basic access and development needs as well as equity within and between generations; be affordable, efficient, diverse and well-balanced; limit emissions and waste using preferably renewable resources. This definition has been accepted in its time as it has been used in a variety of papers all around the world. Having a clear and accepted definition of sustainable transportation is helpful to focus the attention on the required actions and policies (The Centre for Sustainable Transportation, 2005).

However transportation, while linked with it, should not be confused with mobility. The latter has even recently replaced the former in the public discourse (Forum for the Future, 2020). According to the Forum for the Future, "Transportation ("across-carry" in Latin) describes the act of moving something or someone, whereas mobility ("capable of movement") describes the ability of a person to move or be moved." Therefore mobility disposes of a more socio-economical approach than transportation (Forum for the Future, 2020). The concept of sustainable mobility appeared in 1992 in the EU Green Paper on the Impact of Transport on the Environment which is a response from the EU to the challenges presented in the 1987 Brundtland report mentioned earlier. The EU stated in this paper that sustainable mobility should "enable transport to fulfil its economic and social role while containing its harmful effects on the environment" (Green Paper EU ,1992).

WBCSD stated in 2015 that "Sustainable mobility is the ability to meet society's need to move freely, gain access, communicate, trade and establish relationships without sacrificing other essential human or ecological values, today or in the future." (WBCSD, 2015)

Following the birth of this concept, the understanding and interpretation of it has evolved (Holden, 2019). Holden et al. identified in 2019 four different generations of studies on mobility from 1992 to 2018. While the first generation (1992-1993) focused more on transport efficiency and environmental impacts, the subjects broadened more and more each generation and now comes from diverses fields such as sociology, psychology, innovation studies, political science on top of more traditional ones for this kind of research like planning or transport economy (Holden, 2019). The authors concluded that to achieve sustainable mobility, "we know fairly well what we need to do and we have sufficient knowledge about how to do it".

1.2 Objectives

In order to implement sustainable urban mobility, it will be necessary to understand how impact variables of mobility patterns such as energy consumption, CO_2 emissions or cost can be related to measurable indicators characterizing an urban area. This work aims to identify if mobility, socio - economic or demographic indicators have an influence on the aforementioned variables.

1.3 Structure

This work is divided into five parts : introduction, literature review, data and methods, results and discussion and conclusions and future work. The first part presents the motivation on why doing this kind of work as well as the objectives of the latter. In the second part is reviewed what has already been done in the literature in this area of work. Sustainable urban mobility, energy and environmental impacts of mobility, travel costs as well as IPAT identity and Origin-Destination analysis are topics tackled in this part. The third part presents the methodology of the work with different subsections : first a characterization of the case study of the AML is done, then the data analytical methods that will be used are presented and finally the methodology approach is detailed. The fourth part displays the results given by the analytical methods presented just before, first the results of the exploratory data analysis and then of the multiple linear regression. To end the work conclusions as well as future possible works will be presented.

2. Literature review

Sustainable Urban Mobility Challenges

Tools and strategies for sustainable mobility began to appear in the same document that introduced the concept of sustainable mobility. The Green Paper presented potential measures such as "promoting fast, safe, and convenient urban and regional transport services and reducing urban car traffic" and even "encouraging low transport demand" while acknowledging that the root of the problem is human behaviour which will require fundamental changes to be modified (Green Paper EU, 1992).

In what is one of the most cited papers on the subject, Banister laid in 2008 new foundations for sustainable mobility. He stated that the main goals of this concept from a transport planning standpoint are : reducing the need for travel; shifting the modal split towards more environmentally-friendly than car transport modes such as public transport, cycling and walking; reducing travel distances; increasing the efficiency of transport (Banister, 2008). However, Banister argued, like Holden et al. in 2019, that these measures are already known and applied with varying degrees of success. He emphasized the need to be more open-minded when studying transportation and include "why people travel and how they use time". This way of implementing people in sustainable mobility is key to make policies more understood and accepted.

In order to do so, we need stakeholders to be on the same wavelength when talking about sustainable mobility. In 2020, Foltynova et al. surveyed 36 Czech stakeholders using the Q statistical method in order "to reveal the main shared viewpoints on the preferred paths towards sustainable urban mobility" (Foltynova, 2020). They found out that stakeholders may want to implement things differently even if they share the same definition of sustainable urban mobility which was given by the authors : " a city that motivates its inhabitants to change their travel behaviour towards minimising their emissions and noise impacts on health and the environment". While stakeholders shared the same point of view on increasing quality, affordability and use of public transport, regulating transport based on environmental impacts and constructing dedicated infrastructure for bicycles, there is no consensus on car restriction and development of infrastructure for motor traffic and also on the legitimacy of new alternative transport modes to supplement cars. For example, a non-negligible part of stakeholders share the opinion that building more transport infrastructure would be beneficial, which is a belief environmentalists would consider non-sustainable. This illustrates perfectly the "wide gap between sustainable mobility research and sustainable mobility practice" (Foltynova, 2020). Researchers on this topic are faced with the challenge of defining and explaining clearly the concept of sustainable mobility in order to support in the best way possible stakeholders for them to make the correct step towards achieving sustainable urban mobility. Foltynova et al. also questioned if, regarding their findings, Sustainable Urban Mobility Plans were correctly and practically implemented.

Sustainable Urban Mobility Plans (abbreviated SUMP) were first introduced in 2013 by the European Commission. The objectives of such plans were clearly defined by the commission : “A Sustainable Urban Mobility Plan has as its central goal improving accessibility of urban areas and providing high-quality and sustainable mobility and transport to, through and within the urban area. It regards the needs of the 'functioning city' and its hinterland rather than a municipal administrative region.” (European Commission, 2013). Attributes that an urban transport system should possess to obtain sustainable urban mobility were stated in this report : accessibility, diversity of services, good integration of transport modes, sustainability, efficiency, adequate use of urban space, safety and security. To implement them, a SUMP has to present a long-term strategy to develop future transport and mobility infrastructure and services accordingly but it should also incorporate a delivery plan for short-term action. The European Commission also emphasized integrated planning at all mobility levels to enhance new forms of sustainable urban mobility, in order to reduce externalities associated with the transport sector. Finally, it was highlighted that a “high level of cooperation, coordination and consultation between the different levels of government and relevant authorities” as well as a “transparent and participatory approach” are required for a suitable development and implementation of a SUMP (European Commission, 2013).

SUMPs concepts have been accepted from the get go and were applied to several cities in Europe (Kiba-Janiak et al. ,2019). Consequently there were also a great deal of studies analyzing the performance of such plans as well as proposing methodologies to integrate successfully objectives within the SUMP. Given the tremendous number of different topics that SUMP tackle, the majority of researchers focus on specific issues. For example, Pirlone and Spandaro targeted in 2021 safe mobility in the wake of the COVID-19 pandemic proposing specific indicators and good practices regarding accidents, risk perception and health emergencies (Pirlone, 2021). Pisoni et al. simulated in 2019 the application of SUMPs concepts to 642 european cities to evaluate the impact on background air quality, finding that results on air quality improvement are reasonable - though not impressive - and should be considered as a positive outcome of SUMP (Pisoni, 2019). Maria Diez et al. studied in 2018 the case of the city of Burgos in Spain in which a SUMP-like plan was introduced as soon as 2005. Their goal was to assess the cost effectiveness of SUMP, mainly from a CO_2 savings standpoint (Maria Diez, 2018). One of their main conclusions was that this plan could be regarded as a success especially given the drastic modal split change towards more sustainable means of transportation. Considered on its own the cost of each tonne of CO_2 is high but this modal shift led to greater productive efficiency which added savings hence an acceptable cost per tonne of CO_2 saved at the end. Savings in health costs should also be considered to even lower the cost (Maria Diez, 2018). Lopez-Ruiz et al. tried in 2013 to quantify the effects on CO_2 emissions from measures proposed in SUMP for each NUTS3 in Europe. They found out that a reduction of around 7% of each country’s emissions of CO_2 could be achieved (Lopez-Ruiz, 2013). The most prolific measures to save CO_2 emissions would be : public transport coverage, improvement of efficiency using ICT, reallocation of road space to sustainable transport modes and congestion charging zones (fee to circulate).

However, while SUMP seem to be beneficial in regards to CO2 emission and by extension technological performance, the lack of holistic strategies that also include citizens well-being or socio-economical aspects is concerning (Mozos-Blanco, 2018). In the vein of what was said earlier, public participation continues to be mentioned only in theory and rarely in practice (Mozos-Blanco, 2018).

Urban Index to evaluate cities

In 2019 Kiba-Janiak and Witkowski evaluated the level of advancement of 15 capital cities of EU countries in the sustainable urban mobility planning field. They established 3 clusters of criteria each subdivided into groups. Those were strategy formulation, implementation of measures towards sustainable mobility, and results regarding the implementation of measures in terms of sustainable urban mobility (Kiba-Janiak et al., 2019). The first cluster focused on planning and cooperation, the second on policies to attain sustainable mobility and the last on tangible results and surveys. After running a computer simulation using the Promethee method they unearthed three groups of cities with different levels of advancement regarding the implementation and formulation of SUMP : high-level, medium-level and low-level. The city of Lisbon found itself to be in the low-level group alongside Tallinn, Riga, Bucharest and Athens. The authors especially mentioned that Lisbon is one of the three cities in this study that as of the date of writing the paper weren't part of any initiatives and projects dedicated to the SUMP.

In 2014 Van Audenhove et al. sorted for Arthur D. Little and the IUTP 84 cities throughout the world based on their Urban Mobility Index 2.0 which assess the maturity and performance of a public transportation service through multiple criterias such as 'Share of public transport in modal split' or 'Mean travel time to work' for examples. While Lisbon's network was finally ranked as average it was also the third to last ranked city in Europe only ahead of Athens and Rome (Arthur D. Little, 2014).

Energy and environmental impacts of mobility

Evaluating energy consumption and CO2 emissions of different modes of transport is necessary in the process of studying sustainable mobility to assess the environmental impacts of different policies. Numerous studies have compared energy consumptions between thermal cars and electric ones. In fact, it makes more sense to compare them on energy consumption rather than on pollutants emissions since EVs produce very few of them. In 2017 Braun and Rid compared the consumption between internal combustion and electric version of the same car model being the Renault Kangoo (Braun and Rid, 2017). They chose different speeds and driving scenarios for their study. Routes were separated into three categories : urban, extra-urban, and highway. In an urban setting, it was found that for the same speed and travel distance the thermal version consumed 4.3 times more energy than the electric one. This number decreased to 2.6 and 2.8 times in extra-urban and highway settings respectively. Regarding these results, EVs are at a clear advantage compared to regular vehicles in an urban context. Wu et al.

shared the same conclusion in their 2015 evaluation of the energy consumption of EVs. They collected data on a converted electric car doing daily commute to work but taking 4 different ways to do so. they found out that the energy consumption was at its lowest when taking the urban road. On top of that, driving in the city is even more beneficial for EVs due to a higher number of regenerative braking events (Wu et al., 2015).

Lin et al. tried in 2015 to estimate energy consumption of transport modes using EDA (Exploratory Data Analysis). Their results show that optimal energy efficiencies are observed for rail and water transport whereas the opposite can be said for aviation and road transport (Lin et al., 2015). While trying to estimate the energy and environmental impacts of alternative scenarios for the Portuguese transportation sector, Baptista also estimated the energy consumption and emission factors for different technologies. Concerning cars, diesel and gasoline ones consumed respectfully 1.96 MJ/km and 2.12 MJ/km and emitted 146 and 154 gCO2/km while diesel and natural gas buses consumed respectfully 10.72 and 13.72 MJ/ km and emitted 798 and 1022 gCO2/km (Baptista et al., 2012)

In 2011 Jung et al. studied the emissions of pollutants (including CO2) by analyzing 76 vehicles in South Korea. They obtained the emissions for six different routes with different average speeds for each. This way they statistically obtained a formula for emission (in g/km) in function of speeds for each of these pollutants. Here is the one for CO2 (Jung et al., 2011) :

$$FE(CO2) = 1113.8 * V\text{^}(-0.4928) \quad (1)$$

These results show that emissions are higher for lower speeds and decrease rapidly with speed increasing. The researchers also found that emissions of CO2 aren't really sensitive to the mileage of a car but rather to fuel consumption. They show a linear relationship to this latter as seen in the following formula :

$$FE(CO2) = 22.866*FC+6.8689 \quad (2)$$

Alimujang et al. compared 2020 emissions of private cars, taxis, and buses running with gasoline to electric vehicles ones (taking into account the emission of electricity production) in order to determine the reduction of emission when replacing firsts with the latter. As anticipated the emissions for mostly all pollutants were higher for thermal vehicles than for electric ones except for SO2 as we can see in the figure 1 (Alimujang et al., 2020) :

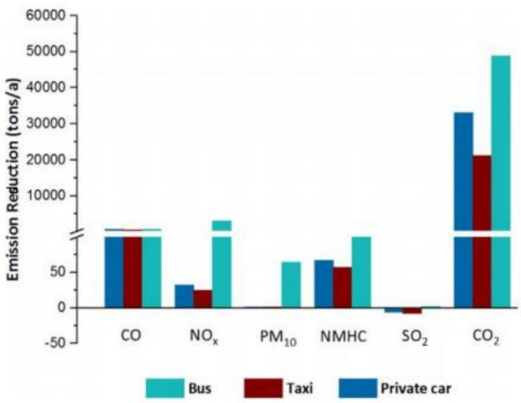

Figure 1- Emission reduction of pollutants depending on vehicle type (from Alimujang et al., 2020)

In 2020 Ghaffarpasand compared the emissions from different categories of vehicles (gasoline, diesel, cab, bus, light goods vehicles and heavy goods vehicle) each divided in EURO class (Ghaffarpasand, 2020). The results show that in most cases vehicles of the most recent EURO class emit fewer pollutants compared to older EURO class except for cabs when the opposite happens. Furthermore, we can see in figure 2 that a bus emits around 3 times more CO2 than private cars (on a vehicle basis - not considering the amount of passengers per vehicle).

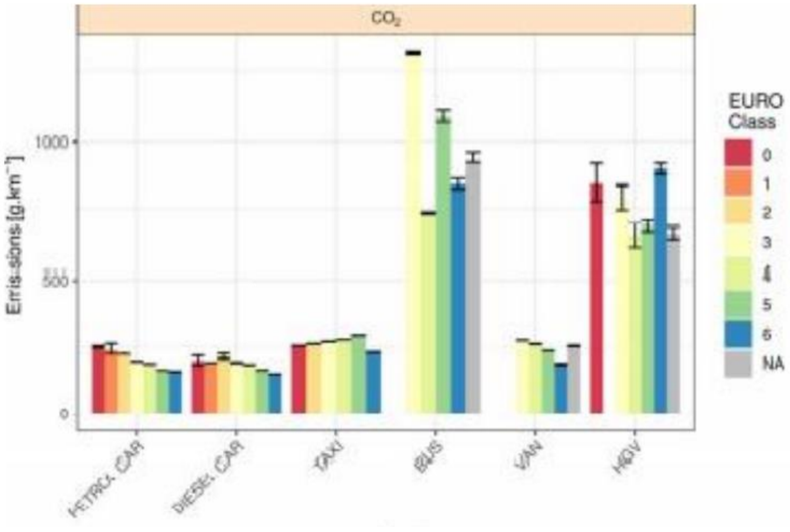

Figure 2- Emissions of different types of vehicles depending on EURO class (from Ghaffarpasand, 2020)

In 2013 Gomes et al. looked at the entire costs, both internal and external, connected with various forms of transportation in Portugal, including light passenger, light freight, heavy passenger, heavy cargo vehicles, motorcycles, mopeds, bicycles, pedestrians, and rail transit. The costs are based on a 30-year time horizon (the European Commission's suggested time horizon for public investments in the water and environment, energy, highways, and other sectors) and a mileage of 11111 kilometers each year. Based on APA, the authors choose to utilize vehicle fleet data from 2007. (Portuguese Environment Agency). Accidents, noise, air pollution, and urban air pollution and urban consequences are among the external expenses that have been calculated. They found out that the internal costs of light passenger vehicles, motorcycles, mopeds and bicycles were respectively 0.117€/km, 0.1741€/km, 0.0561€/km and 0.0188€/km (Gomes et al., 2013).

Wu et al. analyzed and compared in 2015 the total costs of hybrid, plug-in hybrid, electric, and internal combustion vehicles as a function of the range of distance driven. The case study assumes three distance ranges where short distance corresponds to a daily mileage of 14-21 km, medium distance corresponds to 42 km, and long distance 78 km. They also divided vehicles into three different classes to better represent the vehicle fleet : mini and small (A/B), medium small and medium (C/D), sports car (J) (Wu et al., 2015). Their findings are summarized in the table 1 below :

Table 1 - TCO (€/km)of different types of vehicle for an intermediate distance range (from Wu et al., 2015 in Belga, 2021)

Classe/Categoria	**ICEV**	**HEV**	**PHEV**	**BEV**
A/B (31%)	0.340	0.340	0.355	0.375
C/D (50%)	0.460	0.460	0.500	0.520
J (19%)	0.530	0.540	0.613	0.613
Média literatura	0.436	0.438	0.477	0.493

I = PAT identity

I = (PAT) is a mathematical notation introduced in 1971 by Ehrlich and Holdren. Its main goal is to evaluate the human impact on the environment as a function of three factors. It was first designed to show that population is a real contributor to the environmental crisis (Ehrlich and Holdren, 1971). The expression is the following :

$$I = P \times A \times T \tag{3}$$

where I is the environmental impact, P is the population, A is a measure of affluence and T a measure of technology.

This equation became noteworthy because it is a rather simple formulation to tackle such a complex issue as climate change. The three variables of population, affluence and technology supply researchers with measurable but also adaptable tools (DeHart and Soulé, 2000). I=(PAT) can be interpreted in different ways. First it can be used to determine which of the three variables is the most impactful on the environment. But its purpose can also be to show that impacts of an increase in population and affluence can be negated by new technologies that would reduce the toll on the environment (Chertow, 2000).

Although I=(PAT) is first designed to be used on a significant scale and has been used so many times since its creation for example for China (Ma et al., 2016) or the former Soviet Union (Brizga et al., 2013), it can also be scaled down to urban sceneries. In 2000 DeHart and Soulé were wondering if the IPAT formula would be useful on a more local scale. They tried to determine relationships between social driving forces and greenhouse gas emissions in northwestern North Carolina. One of their conclusions was that indeed IPAT identity can be used successfully in smaller scales (DeHart and Soulé, 2000). In recent years more studies appeared where the IPAT formula was used in urban situations. Dai et al. analyzed in 2013 the relationship between energy consumption and economic growth in the Chanzhutan urban agglomeration (Dai et al., 2013). In 2016 Hu also used IPAT to determine that an income gap between urban and rural areas leads to environmental deterioration (Hu, 2016). Yu et al. studied in 2020 the relationship between urban agglomeration and CO_2 emissions in the Yangtze River Delta in China. They found that there was a negative correlation between urban size and emission, meaning that bigger urban areas are more emissions efficient, with a 1% increase of urban population leading to a 0,22% decrease in CO_2 emissions (Yu et al., 2020).

Origin-Destination Analysis

The Origin-Destination (OD) matrix is a significant part of transportation and travel behaviour analysis (Egu and Bonnel, 2020; Abrahamsson, 98). The matrix presents data on the number of passengers between a point A - the Origin - and a point B - the destination (Abrahamsson, 1998). Usually this matrix is created from survey data but with the rise of new technologies other ways of collecting data have emerged (Egu and Bonnel, 2020). A great part of the papers written on the OD subject concern these means of collecting data. Egu and Bonnel compared in 2020 three OD sets of data for the public transportation system of the city of Lyon in order to better evaluate the data sources when estimating public transit. The three means they compared were : large-scale origin-destination survey, household travel survey and automatic fare collection via smart card. They found that the household survey tends to overestimate quite a bit the number of trips (Egu and Bonnel, 2020). Furthermore, transportation data could now be collected much easier and cheaper thanks to the mobile phone network. This has been studied by various authors such as Wideberg et al. in 2013, Iqbal et al. in 2014 or Tolouei et al. in 2017 (Wideberg et al., 2013 ; Iqbal et al., 2014 ; Tolouei et al., 2017).

Once the data has been collected analysis can be made. The bulk of the literature investigates passenger flows thanks to the OD data. Guo et al. used in 2012 OD pairs data in a new approach to comprehend spatio-temporal patterns in the movements of people (Guo et al., 2012). In 2020 Barroso et al. as well as Saberi et al. in 2016 utilized OD trips to study travel behavior in urban networks (Barroso, 2020 ; Saberi et al., 2016). Martin et al. analyzed in 2017 around 26 million trips to reveal patterns in the structure of commuting flows (Martin et al., 2017). An OD matrix was used in order to find significant commuting flows in Flanders by Van Nuffel in 2007 (Van Nuffel et al., 2007). Khadem et al. determined in 2019 the most suitable bike station locations for a bike sharing program thanks to an OD matrix (Khadem et al., 2019).

Furthermore, as said by Verreault in 2009, OD datasets notably those obtained via surveys present a colossal analytic potential going way further than just classic flow analysis (Verreault, 2009). It can be used for example to evaluate the impact of different variables on the aforementioned flows. In 1998 Mamoghli studied the impact of prices, transportation time and frequencies on railway flows in the Paris basin using OD surveys (Mamoghli, 1998). But we can also use OD data to evaluate the impact of flows on external indicators.

Finally, OD datasets can also be used as a means to link variables not directly linked with flows together. In 2013 Sider et al. tried to correlate land-use and socio-economics with traffic emissions and air pollution in the city of Montreal (Sider et al., 2013). Using the process in Figure 3 they managed to estimate the emissions generated per person in each TAZ as well as the emissions experienced per km² in each TAZ thanks to a combination of OD data, emissions functions, vehicle database and traffic estimations.

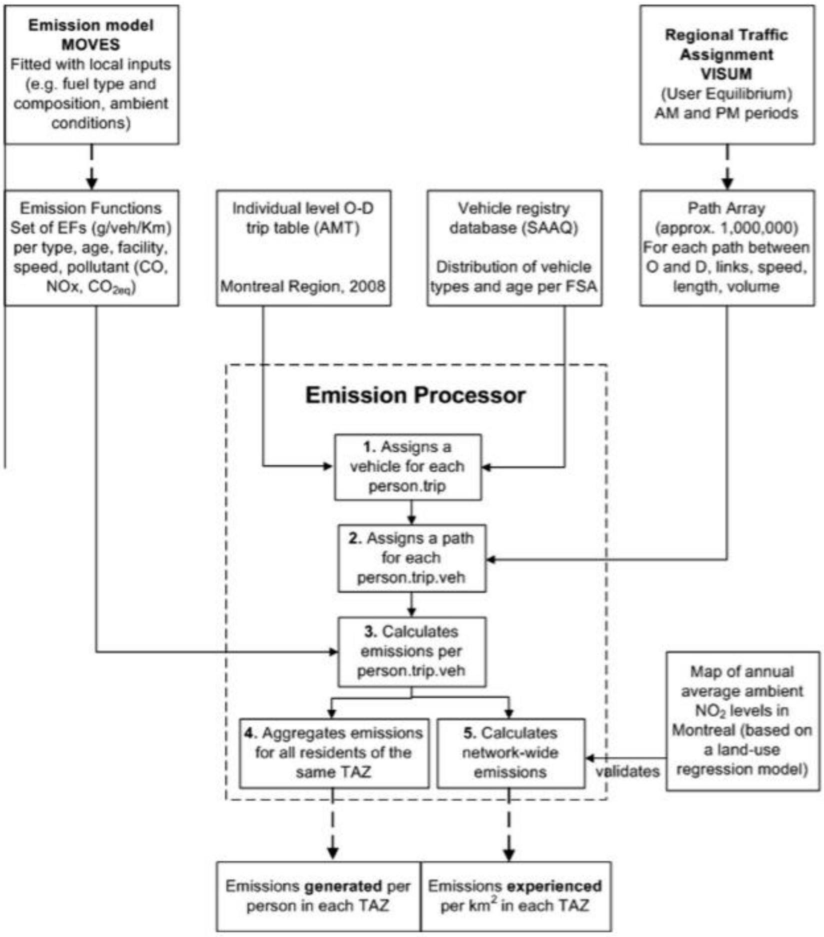

Figure 3 - Methodology to calculate emissions generated and experienced in each TAZ (from Sider et al., 2013)

They then applied a multivariable regression between the two factors calculated and a range of socio-economic, land-use and transportation related variables such as population age, residential density and highway length to find any correlation. Their findings showed major inequalities between the generation and exposure of traffic-related pollution, as people living in dense areas aren't the ones emitting most of the pollutants while being those who are the most exposed to them whereas for people living outside the city in areas of lesser densities it's the opposite (Sider et al., 2013).

However, this kind of research is very scarce in the literature. OD datasets are not often used to study environmental impacts of transportation and the potential links they could have with socio-economical markers. That's why this study focuses on understanding in more detail the more determinant variables influencing urban trips.

3. Data and Methods

3.1 Case study : characterization

Imob data

In 2018 the Portuguese National Institute of Statistics (INE) released the Mobility Survey in the Metropolitan Areas of Porto and Lisboa 2017 referred more commonly as Imob (Imob, 2018). This survey was done with the support of the European Commission and specially its statistics division Eurostat in an effort to develop furthermore transport and mobility statistics in the context of the European Statistical System (Imob, 2018). This way Imob followed good practices put in place by Eurostat as well as other similar international projects. It was conducted in the 4th quarter of 2017.

The goal of Imob was to identify and quantify the trips made by the inhabitants in metropolitan areas (Porto and Lisbon). This way the travel undertaken by the population living in each region can be characterized and used for further policies. The work was divided in two chapters, one for each metropolitan area. Each chapter is divided in 3 parts :

- characterization of the resident population from the perspective of mobility (mobile population), including socio-economic information and mobility expenditure
- analysis of the mobility in the metropolitan area according to two perspectives:
  - total trips, according to the residence of the respondent .
  - intra-metropolitan trips (origin and destination in the metropolitan area)
- opinions of the residents, showing reasons for using individual and public transport, as well as their evaluation of public transport

The survey can be qualified as a household one as it was conducted through a mostly online but also offline questionnaire which asked for each household various precise elements :

- detailed trip for each resident (origin and destination, distance, time, transport mode, reason)
- informations of each resident (age range, education)
- transport informations (vehicles owned, parking, passes)
- financial informations (income, expenditures on transportation)
- opinions on transportation system

It has to be noted that each survey answer especially concerning the trips are only for one given day. A binary variable is then accorded to each trip to tell if it corresponds to a work day or not. The same reasoning is applied to the hours of each trip that are also specified in the survey. This way another binary variable is applied to determine if the trip happens on the peak hours (8 to 10 am, 5 to 7 pm) or not.

A little less than 30000 households spread across AML's eighteen municipalities were surveyed collecting around 120000 trips made by more than 60000 inhabitants of the metropolis. In order to describe precisely the transportation in the AML each household was assigned a weighing coefficient (coded as "pesofin") based on their living area that makes it correspond to the real number of people with potentially similar mobility patterns. As multiple modes of transport are being used by the inhabitants of the metropolis and are detailed in the data, it is also necessary to take into account the different occupancies of these different categories of vehicles. The occupancy rate may vary from studies to studies and Imob surveyors have settled on a particular set of rates that can be found in Table A 1. The number of passengers corresponding to each trip in Imob can be calculated by multiplying the extrapolation factor "pesofin" with the corresponding occupancy rate. This allows for example to evaluate the modal share at the AML level or for more specific areas.

Of all the trips presented in the Imob data, some may present incomplete data as it is possible in the survey to answer that you don't know or don't have a response. Some trips also present time and distance values that are not possible under real conditions. Both of these kinds of trips are cut from the dataset that will be utilized. Furthermore, the ferry is a transport mode that can be used in the AML but since it's really different from other modes, usually less common in other areas and not significant in terms of passengers, it has been deemed that the trips made in ferry won't be used as well. Finally trips that concern municipalities outside of the AML won't be taken into account either as we will concentrate on the AML. This reduces the number of trips to around 90000.

Energy consumption, CO2 emissions and cost data

As stated in the objectives, our main goal is to determine what impacts the most energy consumptions, CO2 emissions and cost of transportation. To do that we need the data corresponding to these variables. Based on a previous work (Belga, 2021), energy consumption, CO2 emissions and cost were associated with each trip. The methodology used by Belga in 2021 is the following.

In order to obtain the energy consumption and pollutants emissions for the average national fleet, the software COPERT (Computer Program to Calculate Emissions from Road Transport) 5.0 was used. The program requires some input parameters to calculate precisely the values corresponding to the area studied. To do so data from the Portuguese national inventory on the national fleet of Portugal was used. Since light passenger cars represent the majority of cars, they were the only ones included in the calculation with buses and motorcycles. COPERT also distinguishes on-peak and off-peak transport behaviour therefore the data from Imob was used to have a good representation of trips' distribution for both traffic environments (Table A 2).

When all the required data is given it is possible to obtain values for the energy consumption and emission factors of pollutants for a certain speed and traffic condition. The process was here repeated 18 times from 10 to 50km/h in peak and off-peak conditions. Energy consumption and emission factors were then regressed against average circulation speeds for a range of travel distances (up to around 20km) and mathematical expressions were defined. They are presented in Table A 3. Now that these expressions were known, the speed of each trip in Imob was calculated based on the respondents data in order to obtain the energy consumption and emission factors for each trip.

The total cost associated with private vehicles was calculated by estimating the costs of acquisition and the cost of operations. The vehicles were divided in 5 categories depending on the engine. Based on several studies as well as an anonymous dataset, a table summarizing all the different costs related to the ownership of a vehicle could be done (Table A 4). From this a cost per kilometer could be calculated and coupled with the aforementioned data of occupancy rate we can obtain the final cost/passenger.km for each category of private vehicles. Concerning public transportation, the value was obtained by dividing by passenger*kilometers the total revenue of each public transport mode as stated in the corresponding operator's yearly accounting reports. Other similar calculations were realized to evaluate the cost/pkm of alternative transport modes. The summary of these values can be found in the table 2 below :

Table 2 - Cost of trasnportation per passenger.km for each transportation mode (from Belga, 2021)

Transporte	€/(pass.km)
Bicicleta	0,019
Comboio	0,062
Motociclo	0,122
Metro	0,126
Autocarro	0,229
BEV	0,227
HEV	0,267
DV	0,311
GV	0,315
PHEV	0,319
e-bicicleta	0,570
e-trotinete	0,760
e-ciclomotor	1,038
TVDE	1,090
Táxi	1,230

It has to be noted that for clarification purposes, only the cost/pkm of diesel vehicles will be used for all private vehicles in the following of the study, as it is the most common kind of vehicle in Portugal and presents similar numbers as gasoline vehicles.

Once the values per passenger.km of energy consumption, emissions and costs have been determined for each trip, it is possible to group trips from the same origin and destinations together in order to quantify energy and environmental impacts of each OD pair. As referred previously, each origin and destination correspond to a TAZ (traffic analysis zone). In our study, each TAZ corresponds to a municipality of the AML. Again, our goal is to evaluate the energy, emission and cost performance of the main mobility dynamics ("corridors") between the 18 municipalities and within each one. These values can be turned into an OD matrix. The figure 4 presents the OD matrix for energy consumption (MJ/pass.km). Each cell represents the energy performance of that OD pair or transport "corridor".

Municipios	Alcochete	Almada	Amadora	Barreiro	Cascais	Lisboa	Loures	Mafra	Moita	Montijo	Odivelas	Oeiras	Palmela	Seixal	Sesimbra	Setúbal	Sintra	Vila Franca De Xira	Total
Alcochete	2.28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Almada	-	1.32	-	-	-	1.07	-	-	-	-	-	-	-	1.25	-	-	-	-	3.64
Amadora	-	-	2.28	-	-	1.22	-	-	-	-	1.59	1.65	-	-	-	-	1.83	-	8.57
Barreiro	-	-	-	1.89	-	-	-	-	1.84	-	-	-	-	-	-	-	-	-	3.73
Cascais	-	-	-	-	2.12	1.56	-	-	-	-	-	2.08	-	-	-	-	1.80	-	7.56
Lisboa	-	1.06	1.24	-	1.31	1.41	1.13	-	-	-	0.88	1.52	-	1.37	-	-	1.51	1.51	12.94
Loures	-	-	-	-	-	1.12	1.82	-	-	-	1.82	-	-	-	-	-	-	1.92	6.69
Mafra	-	-	-	-	-	-	-	2.63	-	-	-	-	-	-	-	-	-	-	2.63
Moita	-	-	-	1.93	-	-	-	-	2.16	-	-	-	-	-	-	-	-	-	4.09
Montijo	-	-	-	-	-	-	-	-	-	2.30	-	-	-	-	-	-	-	-	2.30
Odivelas	-	-	-	-	-	0.83	1.06	-	-	-	2.04	-	-	-	-	-	-	-	3.94
Oeiras	-	-	2.24	-	1.80	1.74	-	-	-	-	-	2.46	-	-	-	-	1.50	-	9.74
Palmela	-	-	-	-	-	-	-	-	-	-	-	-	2.29	-	-	1.59	-	-	3.88
Seixal	-	1.18	-	-	-	1.48	-	-	-	-	-	-	-	1.73	-	-	-	-	4.40
Sesimbra	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.16	-	-	-	2.16
Setúbal	-	-	-	-	-	-	-	-	-	-	-	-	1.70	-	-	2.03	-	-	3.73
Sintra	-	-	1.76	-	1.90	1.53	-	-	-	-	-	1.32	-	-	-	-	2.25	-	8.75
Vila Franca De Xira	-	-	-	-	-	1.32	1.80	-	-	-	-	-	-	-	-	-	-	2.04	5.15
Total	-	3.55	7.51	3.82	7.13	13.30	5.81	2.63	4.00	2.30	6.34	9.03	3.99	4.35	2.16	3.62	8.88	5.46	93.88

Figure 4 - Origin-Destination matrix for MJ/pkm in AML (from Belga, 2021)

The same matrix but for CO2 emissions can be found in the annex (figure A 2) . The values presented correspond to the OD pairs that generate 80% of the total number of trips thus avoiding zero-value cells that would hinder our statistical analysis. It is expected that a significant number of OD pairs (over 50%, many times) in a metro area do not share trips, meaning that the corresponding zones do not attract each other in terms of socioeconomic activities.

We can see in the figure 4 that energy consumption is far from being similar for all the OD pairs. Except for some municipalities like Lisbon or Almada, it seems that trips that take place inside the same city consume more energy than a trip between two different cities. It can also be noted that as expected trips from and to Lisbon constitute a big part of the total number of trips but they seem to generate less energy consumption as trips between other cities as well. This behaviour can be observed as well for CO2 emissions. Some assumptions can be made to try to explain these values such as population or modal share. The goal of this work is to determine which determinants are the most impactful (and with statistical significance) on energy consumption, CO2 emissions and cost in order to be able to explain the differences detected in the OD matrixes.

3.2 Data Analatycal Methods

3.2.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach to apprehend and analyze datasets using graphs or other data visualization methods as well as preliminary calculations. It means looking at the data before processing it in order to see if there is anything noteworthy or that could be detrimental to further analysis.

In this work EDA will be used to preemptively filter the variables that won't have any impacts or tackle variables that are too similar to one another. Furthermore, the variables that will be the most useful later on can also be detected in EDA. To do so several actions will be taken. First of all, plotting the data is an elemental part of EDA and will be realized. Then a correlation table will be built using the corresponding function on the Excel software that calculates the Pearson correlation coefficient between variables. Finally p-values between the variables will be calculated.

The Pearson correlation coefficient (often noted r) is a normalized measurement of the covariance which therefore presents results between -1 and 1. This calculation can only highlight a linear relationship between variables though as it is not designed for other types of relationships. A coefficient below 0 means that there is a negative correlation between tested variables meaning that when one increases the other decreases, while the opposite happens for a positive coefficient. The higher the absolute value of r the higher the correlation.

The *p*-value is used in the context of null hypothesis testing in order to quantify the statistical significance of a result. The lower the *p*-value is, the lower the probability of getting that result if the null hypothesis were true. A result is said to be *statistically significant* if it allows us to reject the null hypothesis. All other things being equal, smaller p-values are taken as stronger evidence against the null hypothesis.

3.2.2 IPAT identity

The IPAT identity, as seen in the literature review, is an equation that quantifies the environmental impact of mankind as a function of three factors : Population, Affluence and Technology. The expression is the following :

$$I = P \times A \times T \tag{3}$$

In this work we will use total energy consumption, total CO2 emissions and total cost as the impact in the IPAT identity because this latter isn't really suited for variables per passenger.kilometer. In order to put into light an equation corresponding to IPAT with our dataset, we will need to use a Multiple Linear Regression. However, since the results of a MLR are presented as a sum and the IPAT identity is a product , we will need to utilize exponential and logarithm to make the two correspond.
Our IPAT identity will look like that :

$$I = \exp(P) \times \exp(A) \times \exp(T) \tag{4}$$

which once applied the logarithm function is transformed to :

$$\ln(I) = \alpha P + \beta A + \gamma T \tag{5}$$

This last expression and especially the coefficients will be determined through an ordinary least square MLR.

3.2.3 Multiple Linear Regression

The multiple linear regression is a mathematical regression method used to describe the variation of a dependent variable based on the variations of independent variables. The theoretical model is the following :

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon \qquad (6)$$

where: y is the dependant variable, and (X1, …, Xn) are the independent variables

β_0 is the intercept

(β_1,..., β_n) are the regression coefficients

ε is the model error

Multiple linear regression is based on various assumptions that need to be tested :

- the dependent variable is normally distributed. To verify it we will use the Shapiro test.
- dependent and independent variables show a linear relationship.
- independent variables shouldn't be highly correlated to one another. The Variance Inflation Method is used here.
- the variance of residuals is constant. The data will be plotted to test the assumption.
- the values of residuals should be independent. This is tested via the Durbin-Watson test.

The regression is usually done in a software. In this work the software Rstudio will be used. The first regression that is done uses variables selected beforehand thanks to the exploratory data analysis. If the data doesn't pass the tests or isn't satisfying enough meaning that the p-value of one of the variables is too high we eliminate this variable to test another one. It is possible though to keep a variable that won't be significant in order for this model to respond better to tests.

Once the regression is done and the model is viable we can interpret the regression coefficients. If the variables are not standardized we can't compare the coefficients directly however it is possible to evaluate the magnitudes of each variable thanks to the elasticity. This latter is calculated by multiplying the coefficients calculated by the regression by the mean of the IV on the mean of the DV.

3.3 Methodological approach

The flowchart in Figure 5 explains roughly how the work was conducted.

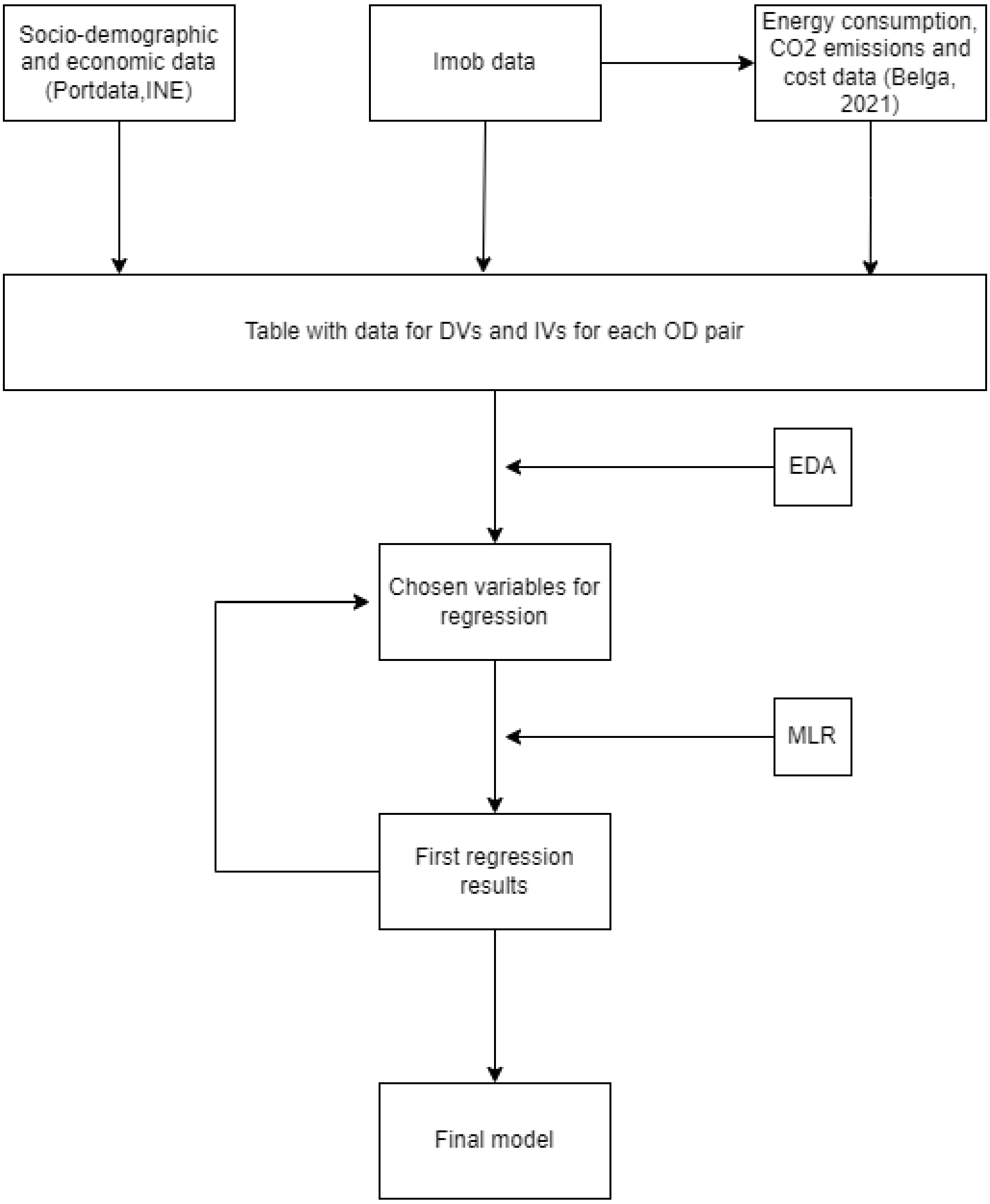

Figure 5 - Methodology flowchart

The first part of the work was to gather all the data that will be needed . The one retrieved with Imob and complemented with Belga's data presents itself under the form of a listing of 90000 trips with details on origin, destination, trip length and more, as well as the added values for energy consumption and CO_2 emissions. Ultimately the data that will be used for the correlation concerns Origin-Destination pairs rather than trips. That's why instead of studying trips separately, they will be gathered together for each OD pair they correspond to. This way we can obtain in particular the total number of trips, total number of weighed passengers, total quantity of energy consumed and total quantity of CO_2 emitted. As each trip also has its transport mode indicated, it is also possible to calculate the modal share for each OD pair by simply dividing the number of weighed passengers for one transport mode by the total number of ponderated passengers for an OD pair.

Once trips were gathered into OD pair, one important decision of this work was to split the data in two : one file that will tackle pairs where the origin and destination are the same municipality - these trips will be from now on called the 'Intra trips' - and the other pairs which will be called the 'Inter trips'. The breakdown of the number of trips and passengers is described in the following table :

Table 3 – Trips and passengers repartition in function of the nature of trips

	% trips	% passengers
Intra trips	72,5	57,6
Inter trips	27,5	42,4

As we can see, even if Intra trips represent a little less than three quarters of the total number of trips, the distribution of passengers is way more balanced and explains why Inter trips are as significant as Intra ones. The reasoning behind separating the two types of trips comes mainly from the fact that transport mode supply and mobility behaviour within a city (shorter trips) and between cities (longer trips) aren't the same. For instance, walking and biking are the modes used by 18% of the Intra trips whereas they only represent 0,3% of them for Inter trips.

Furthermore, filtering needs to be done as well on the Inter trips dataset. There are indeed several pairs that only account for a negligible amount of trips and therefore present skewed statistics due to the low statistical sample. From the original dataset containing all the Inter trips two children datasets are created and will be compared in order to determine which approach is the best : one dataset will contain the 50 Inter pairs with the biggest number of trips and the second one consists of the 50 city pairs regardless of the origin and destination with the biggest number of trips. For the second one, total number of trips, passengers, emissions and energy consumptions are added from the two OD pairs that form a city pair. The rest of the variables will be calculated thanks to an average of the values for each OD pair. For example, 60974 trips are made from Lisboa to Loures and 57541 are made from Loures to Lisboa. So the Lisboa-Loures pair present 118515 trips in total. The first ten city pairs are presented in the table 3 below :

Table 4 - 10 city pairs with the biggest number of trips

Origin	Destination	# of trips
LISBOA	LOURES	60974
LOURES	LISBOA	57541
LISBOA	AMADORA	56630
AMADORA	LISBOA	56541
OEIRAS	LISBOA	47676
LISBOA	OEIRAS	46407
LISBOA	SINTRA	44029
SINTRA	LISBOA	38460
LISBOA	ODIVELAS	37427
ODIVELAS	LISBOA	35873
OEIRAS	CASCAIS	27304

As we can see Lisboa is a common denominator to a lot of these pairs, as expected knowing the gravitational influence Lisboa has on the overall metropolitan area's socioeconomic activity and mobility dynamics. The 50 chosen city pairs represent 96% of the total number of Inter trips made.

In order to be able to examine what determinants can impact the dependent variables which are energy consumption, CO2 emission and cost we need to craft carefully a set of relevant and pertinent independent variables. The variables are divided into three sections : mobility, socio-demographic and -economic indicators and respondents. The set of variables used in this work contains 25 different of them but only the more relevant will be detailed here. The complete list is available in the annex in the table A 4.

Mobility variables concern variables that indicate how certain aspects of mobility are shaped in the area studied. In this study the variables belonging to the mobility section are mostly linked to the modal share. For simplification purposes, the modal split retrieved from Imob data is split in three : **% of public transportation** obtained by adding the share of bus, metro, train and buses; **% of active modes** obtained by adding the share of bike and walking; **% of private vehicle** obtained by adding the share of private cars and taxis.

Socio-demographic and economic variables combine factors that define people in a specific population on a social or economic scale. The data for the following variables are retrieved from Contemporary Portugal Database called Pordata that presents official and verified statistics about Portugal and Europe (Pordata, 2021) as well as from the government office for national statistics of

Portugal INE (INE, 2021). The main variables for this section are : **% of unemployment** describing the unemployed registered at public employment offices as a % of resident population in 2019; **purchase power** describing the purchasing power per capita in 2017; **income** describing the the median value of gross reported income less personal income tax paid per taxable person in 2019; **# of non-residential buildings** describing the number of buildings that don't have the purpose of housing people in 2011. The **population** of a city is also an important variable, especially for the IPAT identity.

The respondents section gathers information about the respondents of the Imob survey based on their answers. The variables can be about different topics. The ones that are the most relevant are : **% of higher education** obtained by dividing the number of people that stated having a higher education by the total number of people (for a specific OD pair); **average size of household** describing the average number of people living in one family unit. Each respondent also indicated which age range they belonged to, them being less than 14, 15-24, 25-44, 45-64, 65-84 and more than 85. The percentage of people belonging to each of these age ranges is a variable by itself although some have a minimal significance.

Once the data has been collected and attributed to each OD pair, it is now possible to realize an Exploratory Data Analysis to evaluate how the variables behave together. The first and easiest way to explore the dataset is to plot variables together in order to visually analyze if they are more or less correlated, if their relationship is linear or non-linear and if their correlation corresponds to the common sense, the assumptions made beforehand as well as fellow studies. To give an example before detailing the results in section 4, the following figure 5 represents gCO2/pkm by % of public transportation for the Intra trips.

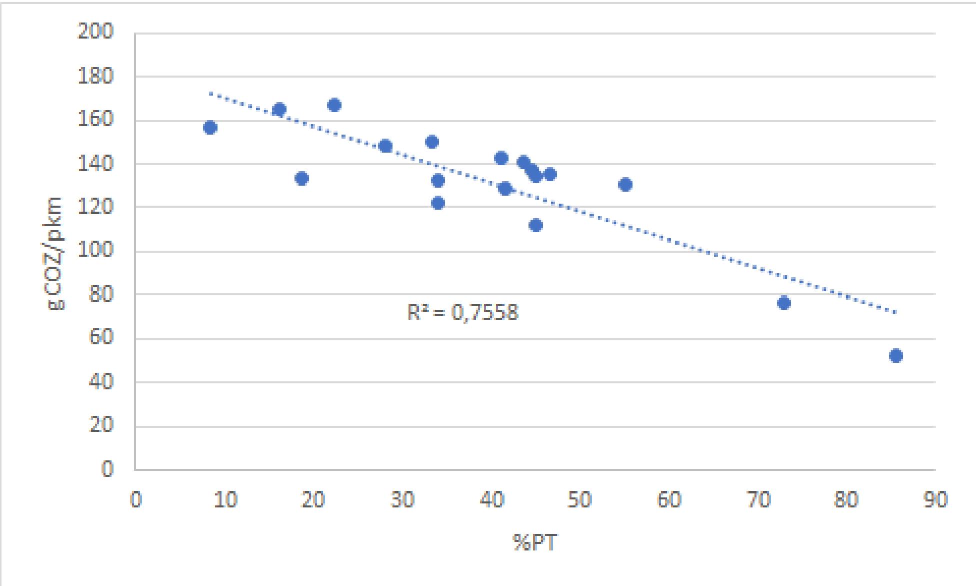

Figure 6 - gCO2/pkm by %PT plot for Intra trips

This plot gives us several interesting results that can be interpreted and used for further analysis. Firstly it is clear that the CO2 emissions per passenger.km are strongly linearly related with a r-squared of 0,7558. Secondly we can state that the correlation is negative meaning that the bigger the share of public transportation in the modal split, the less emissions per passenger.km there is. This follows previous results for example from Belga 2021 since public transportation emits less CO2 per passenger.km than cars, it makes sense that the more people use public transportation the less CO2 is emitted per passenger.km.

Then a correlation table with all the variables is created. This relates how much the variables are correlated and if the latter is negative or positive. For example, the good correlation assumed between gCO2/pkm and %PT is related here as well as the correlation coefficient is -0,8694. The absolute value is close to one so the correlation is really good. A correlation is deemed to be non-significant if the absolute value of the coefficient is lower than 0,3. This way we filter all the variables that won't make the cut for the MLR. Another interesting value to look at is the correlation coefficient between independent variables. If this number is above 0,5 this could indicate correlation and therefore be problematic for the aftercoming regression where the variables tested should not be correlated and avoid heteroscedasticity in the regression This would be useful to know beforehand what variable to select when testing the models.

Once the variables that won't have any impact are removed models for the regression can start to be tested thanks to the R programming language. In order to do the Multiple Linear Regression several functions dedicated to various necessary tests are available in R. The code can be found in the annex. Those for which their results will be presented alongside the models are the r-squared, the Durbin-Watson test (which detects the presence of autocorrelation among the residuals - the null hypothesis) as well as the Condition Indexes (which detects multicollinearity among the independent variables and therefore potential heteroscedasticity in the regression analysis). The best models will be found through empirical testing with different variables, the first ones tested being those that delivered the best results in EDA.

Ultimately we will settle with 3 different kinds of models. The first one will be the one used for the IPAT identity with the total values of each dependent variable, i.e., total energy consumption in each OD pair, total CO2 emissions and total costs. The second one will be the one that fits the best all pkm DVs with 3 fixed variables in order to be able to compare the behaviour of each. The final one will be the best model possible regardless of the variables. Each type of model will be realized for each DV and for Intra and Inter trips separately leading to a theoretical total of 18 models. However it is possible that a model corresponding to the second kind will also be the best possible, therefore reducing the final number of models.

4.Results and discussion

4.1 Exploratory Data Analysis results

As stated in the methodology, the exploratory data analysis realized on the dataset consisted of two separate things : calculating the Pearson coefficient as well as p-value to decipher the relationship between the dependant and independent variables and plotting them in order to see if this relationship is linear or not. We will perform the first task to decide which of the variables should be the basis for our regressions. Then we will plot the dependent variables with those that have been chosen to check for linearity. The following data concerns the Inter trips only, and it was determined that the Inter dataset which uses a city pair instead of the one who uses OD pair because of the better results it provides.

Energy consumption per pkm

Table 5 - Results of EDA for MJ/pkm

Independant variables	r	p-value
X_pc	0,49474662	8,1113E-08
ltrip	-0,3899301	3,9267E-05
OD_purchasepower	-0,3809313	6,1003E-05
OD_pop	-0,3731228	8,8492E-05
HH_size	0,31248044	0,00117322
D_NR_building	-0,3070153	0,00144446
OD_Density	-0,3014012	0,00178127

As we can see in Table 4, the variable MJ/passenger.km isn't strongly correlated to any of the independent variables. There are however several variables that still present a big enough correlation to work with. The two variables that are correlated the most are two mobility ones : % of private cars in the modal mix and length of trips. The positive r for %PC is justified by the fact that private cars show a bigger MJ/pkm than public transportation therefore by increasing the share of PC the MJ/pkm increases as well. The other variable positively correlated is the average household size. The explanation is less straightforward but we can imagine that the bigger a family is the higher the usage of the car is. This is partially demonstrated by the fact that the average household size is correlated with a 0,55 r coefficient to another variable present in the analysis but that won't be used being the number of cars a household possesses. Variables belonging to the socio-demographic and economic type we defined in the methodology can also be found : purchase power, non-residential buildings, population and density. Finally, there are no issues with the p-values presented.

CO2 emissions per pkm

Table 6 - EDA results for gCO2/pkm

Independant variables	r	p-value
X_pc	0,94211245	1,0942E-50
X_clean	-0,8976124	1,9761E-38
Metro_._Train	-0,7666158	1,5507E-21
OD_purchasepower	-0,5969887	1,7997E-11
OD_pop	-0,5704524	2,1155E-10
OD_income	-0,4770489	2,6756E-07
X.high_ed	-0,4439399	2,0999E-06
ltrip	-0,4428613	2,2377E-06
OD_Density	-0,4320899	4,1715E-06
D_NR_building	-0,4072332	1,6226E-05

We can observe a better overall correlation between gCO2/pkm and the independent variables than with MJ/pkm. It is especially strongly correlated to three mobility variables % private cars, % clean public transportation and the presence of the metro and train in the cities. It has to be noted that these variables are logically strongly correlated with each other (r around 0,8) as there are interdependent. If the share of private cars increases the share of public transportation is very likely to decrease. The emission per passenger.km also shows good correlation with several socio-economic and demographic variables that are purchase power, population, density, gross income less income tax (simplified as income) and to a lesser extent number on non-residential buildings. Two of them concern the financial situation of the inhabitants and present a negative correlation coefficient, meaning that the more comfortable financially should inhabitants be the lesser the emissions per passenger.km. We will verify this take with the regression models and discuss it later on.

Cost per pkm

Table 7 - EDA results for cost/pkm

Independant variables	r	p-value
X_clean	-0,9622048	5,3113E-60
X_pc	0,86329555	2,34E-32
Metro_._Train	-0,7832472	5,4648E-23
OD_purchasepower	-0,5625888	4,213E-10
OD_pop	-0,5117754	2,4095E-08
OD_income	-0,5006593	5,3616E-08
X.high_ed	-0,4706231	4,0598E-07
ltrip	-0,3911203	3,7009E-05
OD_Density	-0,36908	0,00010689
D_NR_building	-0,3499177	0,00025315

The exact same variables as for gCO2/pkm are present here with similar correlation coefficients. This can be explained by the fact that cost/pkm and gCO2/pkm are heavily correlated with a r of around 0,9 whereas the correlation between cost per pkm and energy per pkm is way less evident with a correlation coefficient of around 0,3.

Total energy consumption

Table 8 - EDA results for ln(totalMJ)

Independant variables	r	p-value
OD_pop	0,73615918	3,6479E-19
X_pc	-0,6269517	8,3706E-13
OD_purchasepower	0,58911782	3,8263E-11
X_clean	0,57568933	1,3236E-10
OD_income	0,48650876	1,4256E-07
Metro_._Train	0,47735015	2,6232E-07
D_NR_building	0,44748479	1,7015E-06
OD_Density	0,39063093	3,7922E-05

The results of the EDA on the total consumption through its logarithm are a little bit different than those per passenger.km. The correlations are much better, especially for population, modal shares and purchase power. However the signs of correlations are changed for the most to become positive. This is explained by the fact that usually totals increase with other variables. However, the correlation has become negative for the share of private cars. This is surprising as we would expect the total energy consumption to increase as the share of cars increases.

Total CO2 emissions

Table 9 - EDA results for ln(totalCO2)

Independant variables	r	p-value
OD_pop	0,63203954	4,8095E-13
OD_purchasepower	0,42593638	5,898E-06
X_pc	-0,3842782	5,1861E-05
D_NR_building	0,36696543	0,00011788
OD_income	0,35420139	0,00020975
ltrip	-0,3432667	0,00033721
OD_Density	0,32921533	0,00060539

As for the energy consumption, the logarithm of the emission has become positively correlated to most variables except the two mobility ones share of cars and length of trips. This could be attributed to the fact that as population is more important in the big urban municipalities such as Lisbon it leads to more trips therefore more total CO2 emitted (as well as total energy consumed). As the share of public cars and length of trips are lower in those areas it would makes sense that the correlation coefficients would be negative.

Total cost

Chosen Indicators

Based on the results we collected previously, we can select several variables that will constitute the basis for our models. These variables differ between DVs per passenger.km or totals.

Table 10 - Chosen variables for DVs per pkm

Type of IV	Independent variable	r MJ/pkm	r gCO2/pkm	r cost/pkm
Mobility	% private cars	0,49474662	0,942112449	0,863295553
Socio-economic/demographic	Purchase power	-0,3809313	-0,596988719	-0,56258876
Respondents	% high education	-0,2947293	-0,443939901	-0,4706231

The three variables chosen for the base model are presented in the table 9. The variable from the mobility type is the share of private cars. The choice of this indicator makes sense as it is the one presenting one of the best correlation coefficients for each DV. Concerning the socio-economic or demographic determinant, the choice of the purchase power indicator is logical as it presents a valid correlation coefficient for each variable and has the best one of the independent variables of its type. Finally for the variables from the respondents the choice of the percentage of respondents with a high education has been made as it presents a good correlation coefficient for emissions and cost and it's the only variable of its category with a coefficient above 0,3 on absolute value. It does not however pass this threshold for energy consumption but only just barely therefore it has still been chosen.

The graphs of the DVs in function of the chosen IVs are available in the annex (figure A 2-10).

Following the same process, the variables chosen for total DVs are : share of public transportation, population and share of people with high education.

4.2 Regression analysis

Preliminary comments

Once the basis variables have been determined we can launch the regression process in order to determine if we can find an expression relating the relationship between the dependent variables and the indicators we presented. The first regressions are done with the variables chosen in the part before that are supposedly the best ones. However the better results can't be found only thanks to an exploratory data analysis and it can happen that the better models include variables that haven't been deemed relevant in the EDA. This is the case for the model concerning the per passenger.km variables for Inter trips. Purchase power was supposed to be the best determinant from a socio-economical standpoint however the results with it weren't concluant at all. After some trials and errors, the socio-economic variable that gives the better results is unemployment despite its Pearson coefficient being lower than 0,3 in absolute. It was also determined that the share of private cars gives better results than the share of public transportation for total DVs, while the inverse is true for pkm DVs.

The residuals plots for each model are available in the annex, from Figure A 12 to Figure A 17.

4.2.1 Total DVs results

Total energy consumption

Table 11 - Regression results for ln(totalMJ)

IVs	p-value
%PC	0,00175
%high ed	0,03892
Population	1,23E-06
Tests	
r-squared	0,6338
DW test	stat : 1,047975 / p-value = 0
Highest Condition Index	9,576689

The first regression concerns the total energy consumption. As we can see in table 10, the three variables chosen beforehand seem to make for a good model. In fact, they both have a p-value lower than 0,05 meaning that they are significant. The population variable is the most important one followed by the share of private cars and then by the share of people with high education.

Concerning the tests needed to be validated to approve the model, the r-squared has a very good value of 0,6338 and the higher condition index being lower than 10 shows that there are no multicollinearity problems with the model. Finally, the Durbin-Watson test is lower than 1,5 however the p-value being 0 means that there is not necessarily autocorrelation in the model.

As the variables used are not standardized we cannot compare directly the impact they have through their estimates alone. In order to evaluate the magnitude we should calculate the elasticity of each variable.

Table 12 - Estimates and elasticities for the chosen IVs in the total energy consumption regression

IV	estimate	elasticity
%PC	-2,11E-02	-5,79E-02
%high ed	8,22E-06	1,64E-05
Population	-1,51E-02	-2,26E+02

The elasticity results showcase three different levels of magnitude for the Independent variables. By far the most important one is the population which as said previously is expected when working with totals as the IPAT identity can show. Then the share of public cars is way less significant than population and finally the importance of the impact of the share of people with a high education is really low. One thing that is interesting and already discussed before is the signal meaning that population has an inverse relationship with total energy consumed. This will be more thoroughly explored later in the work.

Total CO2 emissions

Table 13 - Regression results for ln(totalCO2)

IVs	p-value
%PC	0,27208
%high ed	0,00286
Population	1,77E-06
Tests	
r-squared	0,4964
DW test	stat = 1,055038 / p-value = 0
Highest Condition Index	9,576689

As for the energy consumption the regression concerning total CO2 emissions results in a model that seems acceptable. The p-values of population and % high ed are below 0,05 meaning that both IVs are significant. However the p-value of %PC is higher than 0,05 being around 0,27. This means that the variable is not really significant but the value is still low enough to consider keeping it in the model.

The Condition Index and Durbin-Watson test shows similar behaviour as the former model. The r-squared of this one is lower meaning that overall the quality of the model will be worse but an adjusted r-squared around 0,5 is still good enough to validate the model.

Table 14 - Estimates and elasticities for chosen IVs in the total CO2 emissions regression

IV	estimate	elasticity
%PC	-6,73E-03	-1,46E-02
%high ed	7,70E-06	1,21E-05
Population	-2,14E-02	-2,52E+02

The behavior of the elasticity is the same as for total energy consumed with similar order of magnitude and signs.

Total cost

Table 15 - Regression results for ln(totalCost)

IVs	p-value
%PC	0,000353
%high ed	0,008602
Population	3,56E-09
Tests	
r-squared	0,7336
DW test	stat = 1,018467 / p-value = 0
Higher Condition Index	9,576689

The three variables show p-values below 0,05, the tests have the same results as before and the adjusted r-squared is quite good. This means that this model is viable as well.

Table 16 - Estimates and elasticities for chosen IVs in the total cost regression

IV	estimate	elasticity
%PC	-1,90E-02	-7,38E-02
%high ed	8,34E-06	2,35E-05
Population	-1,52E-02	-3,20E+02

The behavior of the elasticity is the same as for total energy consumed and total emissions with similar order of magnitude and signs.

4.2.2 Per pkm results

Energy consumption per pkm

Table 17 - Regression results for MJ/pkm

IVs	**p-value**
%PT	8,10E-06
%high ed	0,192
Unemployment	0,585
Tests	
adjusted r-squared	0,4188
DW test	stat : 1,938208 / p-value = 0,718
Highest Condition Index	20,321857

The model presented in the table 16 is not the best but presents decent enough results to be considered acceptable. The p-value for %PT is very good, meaning that this variable has a big significance in the model, contrary to unemployment which shows little significance with a p-value over 0,5. However we can still keep the variable in the model. The p-value of %high education is higher than the usual threshold of 0,05 but below 0,2 it can still be considered somewhat significant for the model.

The adjusted r-squared of this model is pretty bad at 0,4188. It passes very well the Durbin-Watson test with a value around 2 and the highest condition index is a little high but still acceptable. It means that there are some multicollinearity issues but we can disregard them.

Table 18 - Estimates and elasticities for chosen IVs in the MJ/pkm regression

IVs	**estimate**	**elasticity**
%PT	-0,010611	-0,4541544
%high ed	-0,00367	-0,0677755
Unemployment	0,03438	0,13221653

As the r-squared of this model is low the validity of the conclusions has to be taken with more precautions than models with better r-squared. We can see that the most impactful variables on energy consumption per passenger.km is the share of public transportation, followed by unemployment which shows a similar level of magnitude but a slightly lower value as well as a different sign. However as seen earlier the p-value of unemployment is pretty high so we should not consider it accurate. Finally the share of people with high education is less significant but has still its importance.

CO2 emissions per pkm

Table 19 - Regression results for gCO2/pkm

IVs	p-value
%PT	2E-16
%high ed	0,00065
Unemployment	0,07894
Tests	
adjusted r-squared	0,927
DW test	stat : 1,988652 / p-value = 0,896
Highest Condition Index	20,321857

This model looks very good. First of all the adjusted r-squared is excellent with a value above 0,9. Furthermore all the p-values are good. The one corresponding to unemployment is a little bit above the 0,05 threshold but it doesn't change much of its significance in the model. Finally the Durbin-Watson test is passed without any issue with a DW-value around 2. The only hindering thing about the model is the relatively high condition index.

Table 20 - Estimates and elasticities for chosen IVs in the gCO2/pkm regression

IVs	estimate	elasticity
%PT	-2,72E-04	-2,45E-04
%high ed	-8,80E-05	-3,42E-05
Unemployment	-1,42E-03	-1,15E-04

Surprisingly the impactfulness on gCO2/pkm of %PT and unemployment is very similar. We can spot a significant difference between these results and the one for energy consumption as the values for unemployment are negative meaning an inverse relationship between unemployment and gCO2/pkm whereas this relationship is direct for MJ/pkm. As for share of higher education the same conclusions can be applied though with it being slightly less significant.

Cost per pkm

Table 21 - Regression results for cost/pkm

IVs	p-value
%PT	2,02E-14
%high ed	0,00974
Unemployment	0,05949
Tests	
adjusted r-squared	0,7965
DW test	stat : 2,456588 / p-value = 0,104
Highest Condition Index	20,321857

This model looks good as well. The adjusted r-squared is great and p-values are low. We have the same issue than the other models with this high condition index but it's still manageable. Contrary to the other two models, the Durbin-Watson test is not passed as comfortably but it is still acceptable with a value just below the 2,5 threshold.

Table 22 - Estimates and elasticities for chosen IVs in the cost/pkm regression

Ivs	estimate	elasticity
%PT	-1,45624	-3209,0036
%high ed	-0,33728	-320,69161
Unemployment	-3,73212	-738,96777

Like for CO2 emissions all of the signs are negative here. However the share of public transportation seems more important regarding the other variables than for the two other DVs. Furthermore the importance of %high ed and unemployment are closer for cost/pkm than for the other.

4.3 Discussion

Intra trips

First of all, a discussion can be made on the non-inclusion of Intra trips in the regression. The thing is that despite relatively good correlation coefficients, finding a good model with one variable of each kind was really difficult. Furthermore, tests were also really hard to pass. Several things can explain it. Firstly, the low amount of inputs from the dataset : only 17. A small sample is likely to give skewed results in the end. To top it of 2 municipalities from the 17 were outliers that derailed the data analysis, Lisboa primarily and Almada. In order to analyze the Intra trips they should be treated at a bigger scale, i.e., at the parish level to better grasp the local mobility dynamics.

IPAT identity

Now that we have the results of the regression on the logarithm of total DVs we can come back to the IPAT identity. The three regressions of each total DV present similar results : big impact of the population with an inverse relationship with the DV, small impact of the share of private cars with a direct relationship with the DV and a even lower impact for the share of people with high education with an inverse relationship with the DV. The two lasts relationships can be explained in different ways but the first one is surprising. As the population increases we could expect the total CO2 emissions or energy consumption to increase as well. However a couple of things are to be noted. First of all the so-called 'Population' variable here is in fact the average population between the two cities in the OD pair. This could provide results different than expected because the variable is not as straightforward as the population of one city. Furthermore as stated in the work it could be explain by the fact that areas with larger populations tend to bet hose with a more diversified public transportation service whereas inhabitants of smaller cities may be required to use the car.

All in all the population have a big impact on the dependent variable and that correlates with the creation of the IPAT identity that was done to show that population also have a big impact on the human impact on the environment.

Interpretations of the results for DV per passenger.km

For cost/pkm and MJ/pkm we can see that the share of public transportation has a big impact on them with an inverse relationship. This means that the more the public transportation in the modal share the less energy is consumed and the less the cost per passenger and per km. Concerning the cost it is logical when we look at the Figure 2 as the cost/pkm of pubic transportation is lower than the cost/pkm of a car regardless the mean of public transportation. The same can be said for the energy consumption as for the kind of trips studied the rail is the mean of public transportation the most used and it is also one of the most energy efficient. Concerning the emissions, it makes also sense that emissions decrease when public transportation increases in the same vein as energy however its impact is lower.

For the 3 DV per pkm the share of higher educated people has an inverse relationship with the DV, meaning that when people are more educated the energy consumption, CO2 emissions and cost per pkm are lower. We can conjecture on the explications of this result. People with high education are more likely to have jobs with higher responsabilities. This kind of work can be found more likely in big cities where economic point of interest lie. In the AML this place is Lisboa. This would mean that people with high education may be more likely to travel to Lisboa, where public transportation to attain it is more available than to other cities. As said earlier public transportation offers better results on pkm variables therefore explaining the relationship found in the regression.

Concerning the unemployment, except for MJ/pkm for which the model has less viability, the relationships with gCO2/pkm and cost/pk mis an inverse one, meaning the more the share of unemploymed people the lower the CO2 emissions and cost per passenger.km. We can conjecture that as unemployed people tend to have lesser financial stability, they are less likely to possess a private car that can be a bif financial burden and more likely to use public transportation. As said earlier public transportation offers better results therefore explaining the relationship.

5 Conclusion and future works

As the climate change situation becomes worse and worse each year and transportation plays a big part in it, it has become important to determine what should be done in order to limit its impact on the environment. One of the big challenge for urban areas around the world is reaching sustainable urban mobility. This will align the needed changes for the enivronment with social and economic improvements for people.

In order to have a better grasp on the relationship between transportation and socio-economic and socio-demographic variables, it can be interesting to determine and quantify the impact these indicators can have on energy consumption and CO2 emissions as well as the cost per passenger.km of mobility patterns. This was the goal of this work, with a focus on the AML.

The data used was retrieved from the 2017 Imob survey relaized in the Lisbon Metropolitan Area. This recorded with details on transport mode, length and duration, social and economic situation of respondents around 120000 trips (cut down to 90000) from a little less than 30000 households. The energy consumption, CO2 emissions and cost were calculated for each trip by Belga in 2021. Socio-economic and socio-demographic indicators were added to the dataset in order to create one that could answer the goal of this work after analysis. The trips were gathered in so-called city pairs that presented the sum for total variables and average for the others.

It was first determined the more interesting indicators thanks to an exploratory data analysis. The three indicators for the per pkm variables were : share of public transportation in the modal split, share of people with a high education and purchase power of inhabitants. For the total variables the share of public cars was rather used as well as the population instead of unemployment.

These variables were then used in a multiple linear regression. The first thing noteworthy was that even if it presented good results in the EDA, purchase power didn't fit well in the model of regression. The unemployment rate was used instead beacause of the better results it gave even if this variable didn't show up in the EDA.

Except for MJ/pkm and total CO2 emissions for which some reserves one can have, all the models found show a clear relationship between DVs and IVs. For most of them the most impactful factor is the mobility one, it being the share of cars or of public transportation, with an inverse relationship with the DV. Depending on the variables the impactfulness of the other variables can be more or less important but was always significant.

Future works

As mentioned in the work one of the future work is to tackle the Intra trips by studying them at a parish level to have a better understanding of the mobility dynamics.

We could also integrate other variables such as the opinion of inhabitants on transportation, a variable present in the Imob data but that required a little bit more work like a factorial analysis.

The process shown in this work can also be done with data from other urban areas to check if the results are similar or not and then conclude about what this means.

6 Bibliography

A. Alimujiang and P. Jiang, "Synergy and co-benefits of reducing CO2 and air pollutant emissions by promoting electric vehicles—A case of Shanghai," Energy Sustain. Dev., vol. 55, pp. 181–189, Apr. 2020, doi: 10.1016/j.esd.2020.02.005.

Abrahamsson, Torgil. (1998). Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey.

Banister, D. (2008). The sustainable mobility paradigm. Transport Policy, 15(2). https://doi.org/10.1016/j.tranpol.2007.10.005

Baptista, P. C., Silva, C. M., Farias, T. L., & Heywood, J. B. (2012). Energy and environmental impacts of alternative pathways for the Portuguese road transportation sector. Energy Policy, 51. https://doi.org/10.1016/j.enpol.2012.09.025

Barroso, J. M. F., Albuquerque-Oliveira, J. L., & Oliveira-Neto, F. M. (2020). Correlation analysis of day-to-day origin-destination flows and traffic volumes in urban networks. Journal of Transport Geography, 89. https://doi.org/10.1016/j.jtrangeo.2020.102899

Belga, C. (2021), Energy, environmental and economic performance of multimodal urban trips in different distance ranges.

Braun and W. Rid, "The influence of driving patterns on energy consumption in electric car driving and the role of regenerative braking," in Transportation Research Procedia, Jan. 2017, vol. 22, pp. 174–182, doi: 10.1016/j.trpro.2017.03.024.

Brizga, J., Feng, K., & Hubacek, K. (2013). Drivers of CO2 emissions in the former Soviet Union: A country level IPAT analysis from 1990 to 2010. Energy, 59. https://doi.org/10.1016/j.energy.2013.07.045

Brůhová Foltýnová, H., Vejchodská, E., Rybová, K., & Květoň, V. (2020). Sustainable urban mobility: One definition, different stakeholders' opinions. Transportation Research Part D: Transport and Environment, 87. https://doi.org/10.1016/j.trd.2020.102465

Chertow, M. R. (2000). The IPAT Equation and Its Variants. Journal of Industrial Ecology, 4(4). https://doi.org/10.1162/10881980052541927

DeHart, J. L., & Soulé, P. T. (2000). Does I = PAT Work in Local Places? The Professional Geographer, 52(1). https://doi.org/10.1111/0033-0124.00200

Diez, J. M., Lopez-Lambas, M. E., Gonzalo, H., Rojo, M., & Garcia-Martinez, A. (2018). Methodology for assessing the cost effectiveness of Sustainable Urban Mobility Plans (SUMPs). The case of the city of Burgos. Journal of Transport Geography, 68. https://doi.org/10.1016/j.jtrangeo.2018.02.006

Egu, O., & Bonnel, P. (2020). How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? An empirical investigation in Lyon. Transportation Research Part A: Policy and Practice, 138. https://doi.org/10.1016/j.tra.2020.05.021

European Commission. (2013). A Concept for Sustainable Urban Mobility Plans.

G. Wu, A. Inderbitzin, and C. Bening, "Total cost of ownership of electric vehicles compared to conventional vehicles: A probabilistic analysis and projection across market segments," Energy Policy, vol. 80, pp. 196–214, May 2015, doi: 10.1016/j.enpol.2015.02.004

G. Wu, A. Inderbitzin, and C. Bening, "Total cost of ownership of electric vehicles compared to conventional vehicles: A probabilistic analysis and projection across market segments," Energy Policy, vol. 80, pp. 196–214, May 2015, doi: 10.1016/j.enpol.2015.02.004.

Guo, D., Zhu, X., Jin, H., Gao, P., & Andris, C. (2012). Discovering Spatial Patterns in Origin-Destination Mobility Data. Transactions in GIS, 16(3). https://doi.org/10.1111/j.1467-9671.2012.01344.x

HU, L. (2016). Would Urban–Rural Income Gap Affect Carbon Dioxide Emissions? Empirical Research Based on the Extended IPAT Model. Chinese Journal of Urban and Environmental Studies, 04(02). https://doi.org/10.1142/S2345748116500147

Holden, E., Gilpin, G., & Banister, D. (2019). Sustainable Mobility at Thirty. Sustainability, 11(7). https://doi.org/10.3390/su11071965

Holdren, J. and P. Ehrlich. 1974. Human population and the global environment. American Scientist 62: 282–292

INE website - https://ine.pt/

IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

Instituto Nacional de Estatística (2018) “Inquérito à mobilidade (Imob) - Sessão de apresentação dos resultados definitivos”

Iqbal, Md. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies, 40. https://doi.org/10.1016/j.trc.2014.01.002

Khadem, N. K., Kabir, M. M., Banerjee, S., & Jeihani, M. (2019). Bike Station Suitability on University Campus Using Origin–Destination Matrix—A Morgan State University Case Study. Urban Science, 3(3). https://doi.org/10.3390/urbansci3030074

Kiba-Janiak, M., & Witkowski, J. (2019). Sustainable Urban Mobility Plans: How Do They Work? Sustainability, 11(17). https://doi.org/10.3390/su11174605

Lin, W., Chen, B., Xie, L., & Pan, H. (2015). Estimating Energy Consumption of Transport Modes in China Using DEA. Sustainability, 7(4). https://doi.org/10.3390/su7044225

López Ruiz, H. G., Christidis, P., Demirel, H., & Kompil, M. (2013). Quantifying the Effects of Sustainable Urban Mobility Plans.

Ma, M., Yan, R., Du, Y., Ma, X., Cai, W., & Xu, P. (2017). A methodology to assess China’s building energy savings at the national level: An IPAT–LMDI model approach. Journal of Cleaner Production, 143. https://doi.org/10.1016/j.jclepro.2016.12.046

Mamoghli, M. (1998). Impact de la tarification, des fréquences et des temps de transfort sur les flux ferroviares dans le Bassin parisienThe impact of pricing, train frequencies and travel times on railway passenger flows in the Paris Basin. Recherche - Transports - Sécurité, 59. https://doi.org/10.1016/S0761-8980(98)90051-6

Martin, D., Gale, C., Cockings, S., & Harfoot, A. (2018). Origin-destination geodemographics for analysis of travel to work flows. Computers, Environment and Urban Systems, 67. https://doi.org/10.1016/j.compenvurbsys.2017.09.002

McKay, J. (2019, November 13). Transport or Mobility: What’s the difference and why does it matter? Forum for the Future.

Ministère de la transition écologique, 2021, 'Répartition sectorielle des émissions de CO2 dans le monde', , retrieved from https://www.statistiques.developpement-durable.gouv.fr/

Moore, J. E., Mascarenhas, A., Bain, J., & Straus, S. E. (2017). Developing a comprehensive definition of sustainability. Implementation Science, 12(1). https://doi.org/10.1186/s13012-017-0637-1

Mozos-Blanco, M. Á., Pozo-Menéndez, E., Arce-Ruiz, R., & Baucells-Aletà, N. (2018). The way to sustainable mobility. A comparative analysis of sustainable mobility plans in Spain. Transport Policy, 72. https://doi.org/10.1016/j.tranpol.2018.07.001

O. Ghaffarpasand, D. C. S. Beddows, K. Ropkins, and F. D. Pope, “Real-world assessment of vehicle air pollutant emissions subset by vehicle type, fuel and EURO class: New findings from the recent UK EDAR field campaigns, and implications for emissions restricted zones,” Sci. Total Environ., vol. 734, p. 139416, Sep. 2020, doi: 10.1016/j.scitotenv.2020.139416.

P. Gomes, M. Lopes, H. Martins, and J. Carvalho, “Custos e Benefícios, à escala local, de uma Ocupação Dispersa - Anexo 8: Custos internos e externos de Mobilidade em Portugal,” Aveiro, Mar. 2011.

Pisoni, E., Christidis, P., Thunis, P., & Trombetti, M. (2019). Evaluating the impact of “Sustainable Urban Mobility Plans” on urban background air quality. Journal of Environmental Management, 231. https://doi.org/10.1016/j.jenvman.2018.10.039

Pordata website – https://www.pordata.pt/

S. Jung, M. Lee, J. Kim, Y. Lyu, and J. Park, “Speed-dependent emission of air pollutants from gasolinepowered passenger cars,” Environ. Technol., vol. 32, no. 11, pp. 1173–1181, 2011, doi: 10.1080/09593330.2010.505611

Sider, T., Alam, A., Zukari, M., Dugum, H., Goldstein, N., Eluru, N., & Hatzopoulou, M. (2013). Land-use and socio-economics as determinants of traffic emissions and individual exposure to air pollution. Journal of Transport Geography, 33. https://doi.org/10.1016/j.jtrangeo.2013.08.006

Spadaro, I., & Pirlone, F. (2021). Sustainable Urban Mobility Plan and Health Security. Sustainability, 13(8). https://doi.org/10.3390/su13084403

The Centre for Sustainable Transportation. (2005). Defining Sustainable Transportation.

Tolouei, R., Psarras, S., & Prince, R. (2017). Origin-Destination Trip Matrix Development: Conventional Methods versus Mobile Phone Data. Transportation Research Procedia, 26. https://doi.org/10.1016/j.trpro.2017.07.007

Verreault, H. (2009). Méthodes et outils pour l’utilisation cohérente des enquêtes origine-destination.

Wideberg, Johan & Caceres, Noelia & Benitez, Francisco. (2013). Deriving origin-destination data from a mobile phone network.

World Commission on Environment and Development. (1987). Our Common Future.

Yu, X., Wu, Z., Zheng, H., Li, M., & Tan, T. (2020). How urban agglomeration improve the emission efficiency？A spatial econometric analysis of the Yangtze River Delta urban agglomeration in China. Journal of Environmental Management, 260. https://doi.org/10.1016/j.jenvman.2019.110061

Zhang, H. (2017). Research on the Relationship between Economic Growth and Environmental Technologies in China. MATEC Web of Conferences, 100. https://doi.org/10.1051/matecconf/201710005087

van Audenhove, F.-J., Korniichuk, O., Dauby, L., & Pourbaix, J. (2014). The Future of Urban Mobility 2.0: Imperatives to Shape Extended Mobility Ecosystems of Tomorrow.

van Nuffel, N. (2007). Determination of the Number of Significant Flows in Origin–Destination Specific Analysis: The Case of Commuting in Flanders. Regional Studies, 41(4). https://doi.org/10.1080/00343400701281808

Annex

Table A 1 - Occupancy rates by transport mode in AML (from Imob)

Transport mode	Occupancy rate (pass/vehicle)
Private cars (Peak Hour)	1.23
Private cars (Off-Peak Hour)	1.26
Bus	14.55
Motorbike	1.00
Bicycle	1.00
By foot	1.00
Train	22.20
Subway	31.00

Table A 2 - Repartition of trips in AML in function of the day and the peak/off-peak hours

-	Workday	Weekend	All week
Peak Hour	54.1%	35.5%	48.8%
Off-Peak Hour	45.9%	65.5%	51.2%

Table A 3 - Equation for the pollutant emisions of each transpotr mode in function of peak hours and velocity (from Belga, 2021)

Transport mode	Traffic mode	Variable	Speed Range (km/h)	Equation
Private cars (PC)	Urban - Peak Hour (P)	CO_2	10-50	$FE(CO_2) = 0.12 * V^2 - 10.92 * V + 425.22$
		CE	10-50	$CE = 0.0016 * V^2 - 0.15 * V + 5.79$
		CO	10-50	$FE(CO) = 0.00010 * V^2 - 0.040 * V + 2.26$
		NOx	10-50	$FE(NOx) = 0.00020 * V^2 - 0.019 * V + 0.97$
		COVNM	10-50	$FE(COVNM) = -0.10 \ln V + 0.56$
		COV	10-50	$FE(COV) = -0.0040 * V + 0.35$
		P	10-50	$FE(PM\ TSP) = 9.00 * 10^{-7} * V^2 - 0.00040 * V + 0.079$
	Urban - Off-Peak Hour (OP)	CO_2	10-50	$FE(CO2) = 0.14 * V^2 - 12.09 * V + 426.41$
		CE	10-50	$CE = 0.0018 * V^2 - 0.16 * V + 5.77$
		CO	10-50	$FE(CO) = -0.015 * V + 1.08$
		NOx	10-50	$FE(NOx) = 0.00040 * V^2 - 0.036 * V + 1.21$
		COVNM	10-50	$FE(COVNM) = -4.00 * 10^{-5} * V^2 + 0.0019 * V + 0.13$
		COV	10-50	$FE(COV) = -5.00 * 10^{-5} * V^2 + 0.0022 * V + 0.13$
		PM	10-50	$FE(PM\ TSP) = -5.00 * 10^{-5} * V^2 + 0.0032 * V + 0.012$
Motorbike (M)	Urban - Peak Hour (P)	CO_2	10-50	$FE(CO_2) = 0.15 * V^2 - 12.23 * V + 347.46$
		CE	10-50	$CE = 0.0021 * V^2 - 0.17 * V + 4.82$
		CO	10-50	$FE(CO) = 0.0020 * V^2 - 0.18 * V + 8.24$
		NOx	10-50	$FE(NOx) = -0.013 lnV + 0.77$
		COVNM	10-50	$FE(COVNM) = -0.52 lnV + 3.66$
		COV	10-50	$FE(COV) = -0.52 lnV + 3.75$
		PM	10-50	$FE(PM\ TSP) = -2.00 * 10^{-6} * V^2 + 9.00 * 10^{-5} * V + 0.048$
	Urban - Off-Peak Hour (OP)	CO_2	10-50	$FE(CO_2) = 0.026 * V^2 - 3.57 * V + 206.45$
		CE	10-50	$CE = 0.00030 * V^2 - 0.048 * V + 2.85$
		CO	10-50	$FE(CO) = 0.0017 * V^2 - 0.16 * V + 7.94$
		NOx	10-50	$FE(NOx) = 3.00 * 10^{-5} * V^2 - 0.0021 * V + 0.14$
		COVNM	10-50	$FE(COVNM) = 0.00050 * V^2 - 0.045 * V + 2.79$
		COV	10-50	$FE(COVNM) = 0.00050 * V^2 - 0.045 * V + 2.88$
		PM	10-50	$FE(PM\ TSP) = -3.00 * 10^{-6} * V^2 + 0.00010 * V + 0.048$
Taxi (T)	Urban - Peak Hour (P)	CO_2	10-50	$FE(CO_2) = 0.11 * V^2 - 10.16 * V + 412.54$
		CE	10-50	$CE = 0.0015 * V^2 - 0.14 * V + 5.56$
		CO	10-50	$FE(CO) = 8.00 * 10^{-5} * V^2 - 0.010 * V + 0.42$
		NOx	10-50	$FE(NOx) = 0.00030 * V^2 - 0.030 * V + 1.32$
		COVNM	10-50	$FE(COVNM) = -0.0011 * V + 0.068$
		COV	10-50	$FE(COV) = -0.0011 * V + 0.071$
		PM	10-50	$FE(PM\ TSP) = -0.014 lnV + 0.12$
	Urban - Off-Peak Hour (OP)	CO_2	10-50	$FE(CO_2) = 0.21 * V^2 - 17.35 * V + 514.00$
		CE	10-50	$CE = 0.0029 * V^2 - 0.23 * V + 6.92$
		CO	10-50	$FE(CO) = 0.00020 * V^2 - 0.016 * V + 0.49$
		NOx	10-50	$FE(NOx) = 0.00070 * V^2 - 0.055 * V + 1.71$
		COVNM	10-50	$FE(COVNM) = -0.0011 * V + 0.064$
		COV	10-50	$FE(COV) = -0.0011 * V + 0.066$
		PM	10-50	$FE(PM\ TSP) = 6.00 * 10^{-7} * V^2 - 0.00050 * V + 0.085$
Bus (B)	Urban - Peak Hour (P)	CO_2	10-50	$FE(CO_2) = 0.45 * V^2 - 41.32 * V + 1601.40$
		CE	10-50	$CE = 0.0064 * V^2 - 0.59 * V + 22.88$
		CO	10-50	$FE(CO) = -1.30 lnV + 5.90$
		NOx	10-50	$FE(NOx) = -4.26 lnV + 21.76$
		COVNM	10-50	$FE(COVNM) = -0.31 lnV + 1.29$
		COV	10-50	$FE(COV) = -0.34 lnV + 1.63$
		PM	10-50	$FE(PM\ TSP) = 1.00 * 10^{-4} * V^2 - 0.0095 * V + 0.47$
	Urban - Off-Peak Hour (OP)	CO_2	10-50	$FE(CO_2) = 0.45 * V^2 - 41.32 * V + 1601.40$
		CE	10-50	$CE = 0.0064 * V^2 - 0.59 * V + 22.88$
		CO	10-50	$FE(CO) = -1.30 lnV + 5.90$
		NOx	10-50	$FE(NOx) = -4.26 lnV + 20.76$
		COVNM	10-50	$EF(COVNM) = -0.31 lnV + 1.28$
		COV	10-50	$FE(COV) = 0.00040 * V^2 - 0.038 * V + 1.19$
		PM	10-50	$FE(PM\ TSP) = 1.00 * 10^{-4} * V^2 - 0.0095 * V + 0.47$

Municípios	Alcochete	Almada	Amadora	Barreiro	Cascais	Lisboa	Loures	Mafra	Moita	Montijo	Odivelas	Oeiras	Palmela	Seixal	Sesimbra	Setúbal	Sintra	Vila Franca De Xira	Total
Alcochete	167	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Almada	-	77	-	-	-	34	-	-	-	-	-	-	-	65	-	-	-	-	177
Amadora	-	-	136	-	-	24	-	-	-	-	67	117	-	-	-	-	58	-	402
Barreiro	-	-	-	137	-	-	-	-	76	-	-	-	-	-	-	-	-	-	213
Cascais	-	-	-	-	129	28	-	-	-	-	-	62	-	-	-	-	124	-	343
Lisboa	-	30	25	-	18	52	46	-	-	-	29	52	-	11	-	-	13	32	308
Loures	-	-	-	-	-	49	131	-	-	-	130	-	-	-	-	-	-	80	390
Mafra	-	-	-	-	-	-	-	151	-	-	-	-	-	-	-	-	-	-	151
Moita	-	-	-	87	-	-	-	-	133	-	-	-	-	-	-	-	-	-	220
Montijo	-	-	-	-	-	-	-	-	-	165	-	-	-	-	-	-	-	-	165
Odivelas	-	-	-	-	-	22	72	-	-	-	141	-	-	-	-	-	-	-	234
Oeiras	-	-	91	-	55	58	-	-	-	-	-	135	-	-	-	-	74	-	412
Palmela	-	-	-	-	-	-	-	-	-	-	-	-	143	-	-	40	-	-	183
Seixal	-	63	-	-	-	10	-	-	-	-	-	-	-	122	-	-	-	-	196
Sesimbra	-	-	-	-	-	-	-	-	-	-	-	-	-	-	157	-	-	-	157
Setúbal	-	-	-	-	-	-	-	-	-	-	-	-	87	-	-	148	-	-	235
Sintra	-	-	54	-	130	14	-	-	-	-	-	88	-	-	-	-	112	-	398
Vila Franca De Xira	-	-	-	-	-	24	88	-	-	-	-	-	-	-	-	-	-	132	244
Total	-	170	305	224	331	317	337	151	209	165	367	455	230	198	157	188	381	245	4429

Figure A 1 - OD matrix of gCO2/pkm for AML (from Belga, 2021)

Table A 4 - Cost details for different types of car (from Belga, 2021)

	Petrol	Diesel	Plug-in hybrid	Hybrid	Electric
		Fixed costs			
Acquisiton price (€) *	36 190	38 970	73 090	47 170	53 860
Depreciation (%) (t=12y) *	31	28	31	31	31
Incentives (€) [21]	0	0	0	0	-3 000
ISV (€) [18], [22]	1 423.00	4 054.00	355.75	853.80	0.00
IUC (€/year) [17], [18], [23]	122.64	132.88	122.64	122.64	0.00
Periodic inspection (€/inspection) [24]	25.61	25.61	25.61	25.61	25.61
Insurance (€/year) [18], [25]-[28]	266.94	266.94	266.94	266.94	266.94
		Variable costs			
Parking (€/year) [2]	125.40	125.40	125.40	125.40	125.40
Toll (€/year) [2]	184.50	184.50	184.50	184.50	184.50
Fuel (€/km) *, [30]-[32]	0.130	0.080	0.058	0.078	0.033
Maintenance (€/km) [33]	0.082	0.082	0.041	0.041	0.042

Table A 4 - List of the independant variables with details and sources

Variable	Details	Source
avg_km	Average distance of the trip in km	Imob
Population	Average population of the two cities	Portdata
Density	Average population density of the two cities in people/km²	Portdata
Unemployment	Average unemployment share of the two cities	Portdata
Crimes	Average number of crimes by city	Portdata
Income	Average median value of gross reported income less personal income paid tax (€)	INE
Purchase_power	Average purchase power of the two cities	Portdata
#non-residential_buildings	Average number non-residential_buildings of the two cities	Portdata
<= 14	Average share of people aged less than 14 years old	Imob
15-24	Average share of people aged between 15 and 24 years old	Imob
25-44	Average share of people aged between 25 and 44 years old	Imob
45-64	Average share of people aged between 45 and 64 years old	Imob
65-84	Average share of people aged between 65 and 84 years old	Imob
>= 85	Average share of people aged more than 85 years old	Imob
>=65	Average share of people aged more than 65 years old	Imob
Car_ownership	Average car ownership per household	Imob
Avg_size_of_household	Average size of household	Imob
%pay_Parking	Average share of people paying for parking	Imob
%high_education	Average share of people with a higher education	Imob
%daily_driving	Average share of people driving daily	Imob
%pass_intermodais	Average share of people possessing an intermodal pass	Imob
%PT	Average share of public transport in the modal split	Imob
%active	Average share of active modes of transportation in the modal split	Imob
%PC	Average share of public cars in the modal split	Imob
ageing	Average ageing index of the two cities	INE

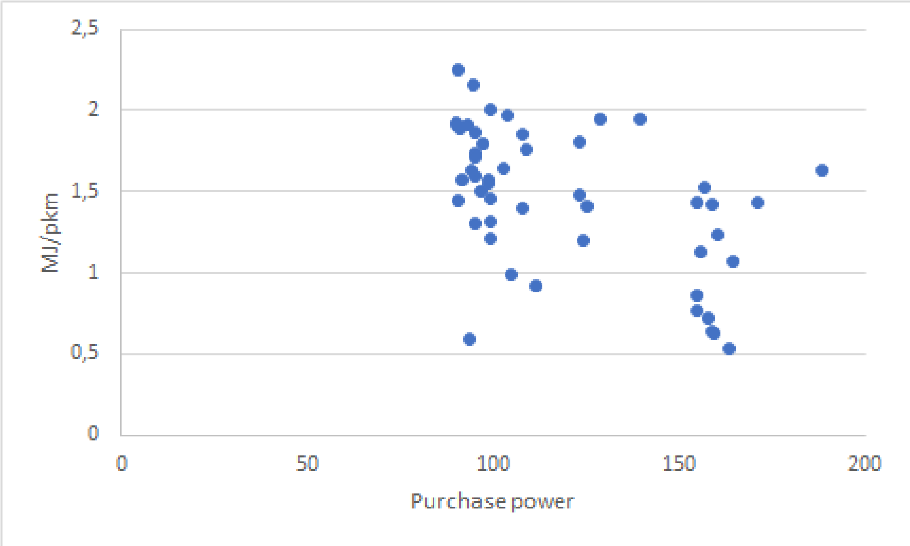

Figure A 2 - MJ/pkm in function of purchase power for Inter trips

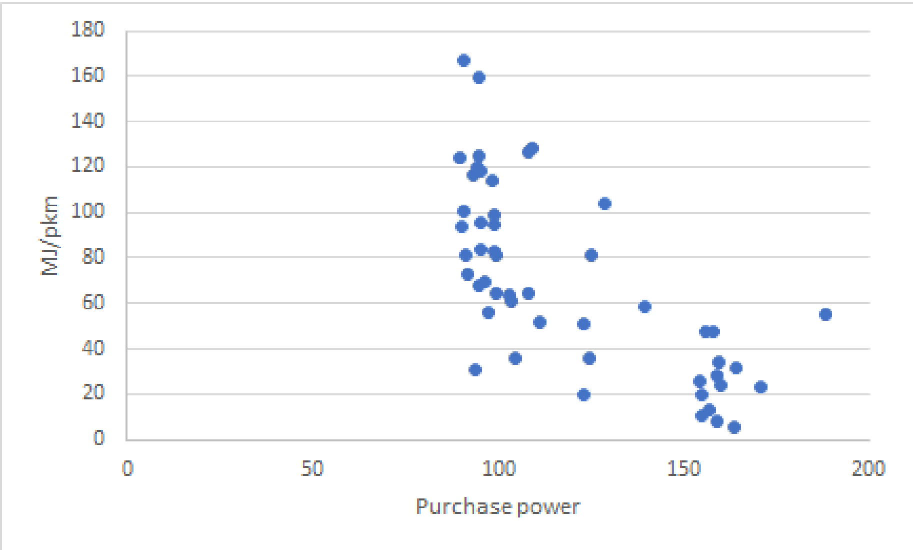

Figure A 3 - gCO2/pkm in function of purchase power for Inter trips

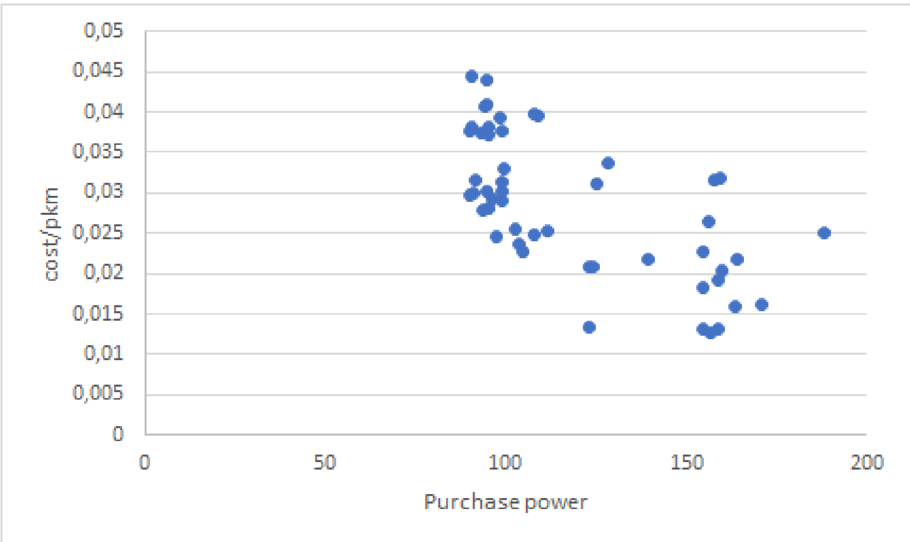

Figure A 4 - cost/pkm in function of purchase power for Inter trips

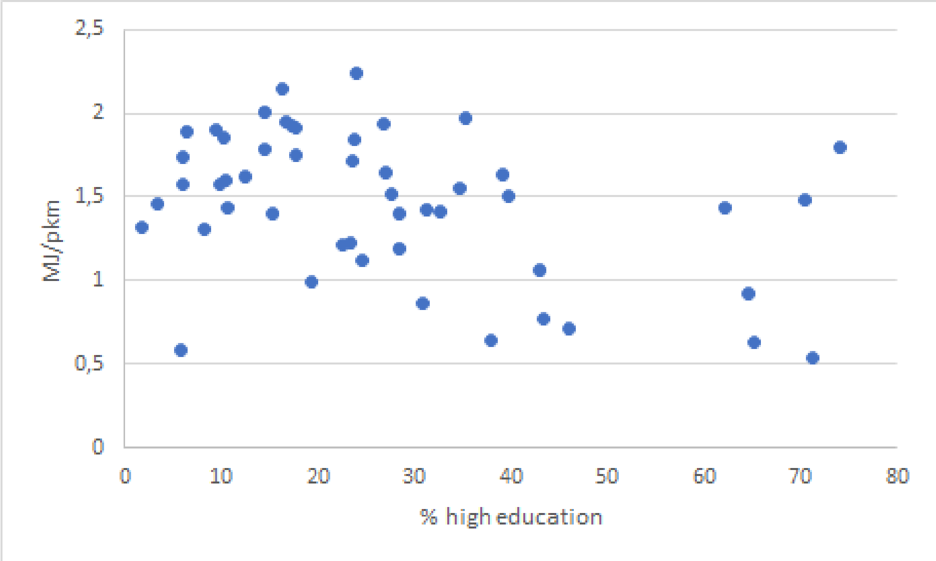

Figure A 5 - MJ/pkm in function of % high education for Inter trips

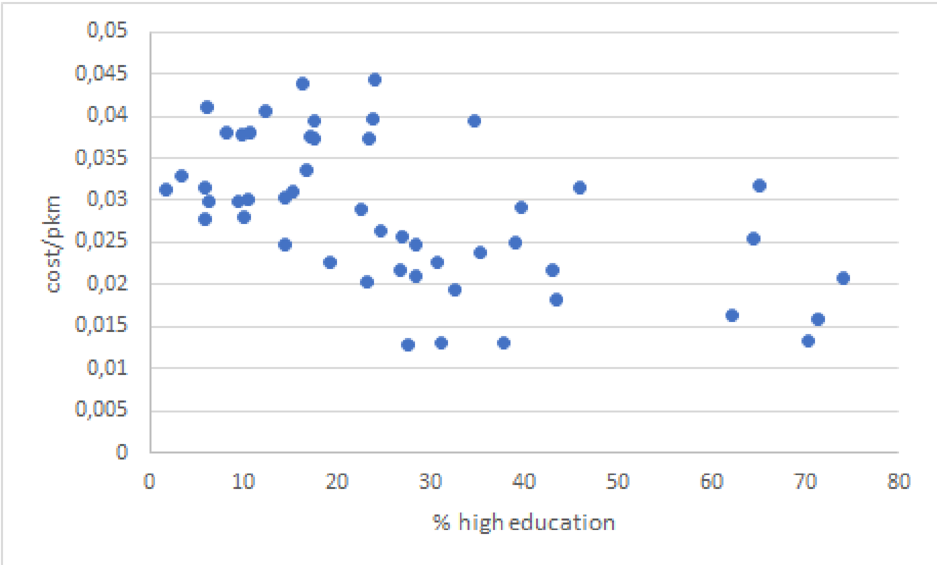

Figure A 6 - gCO2/pkm in function of % high education for Inter trips

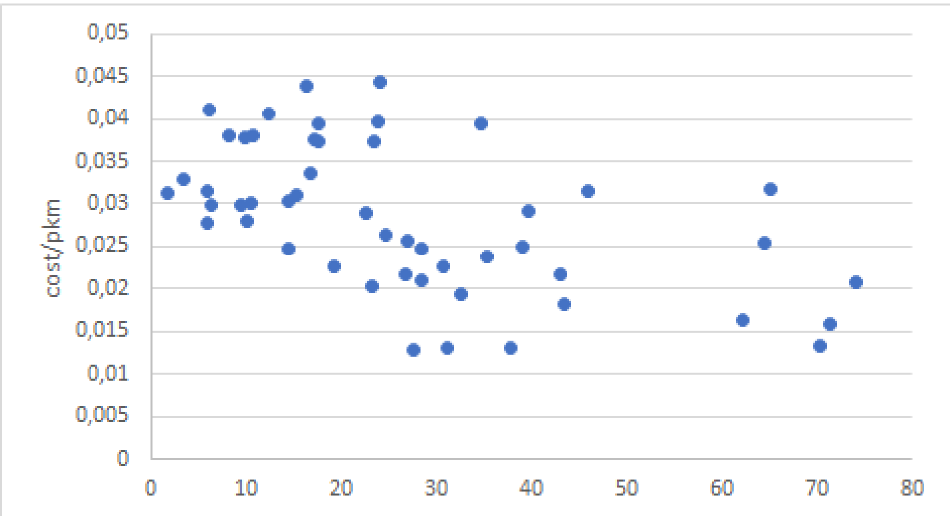

Figure A 7 - cost/pkm in function of % high education for Inter trips

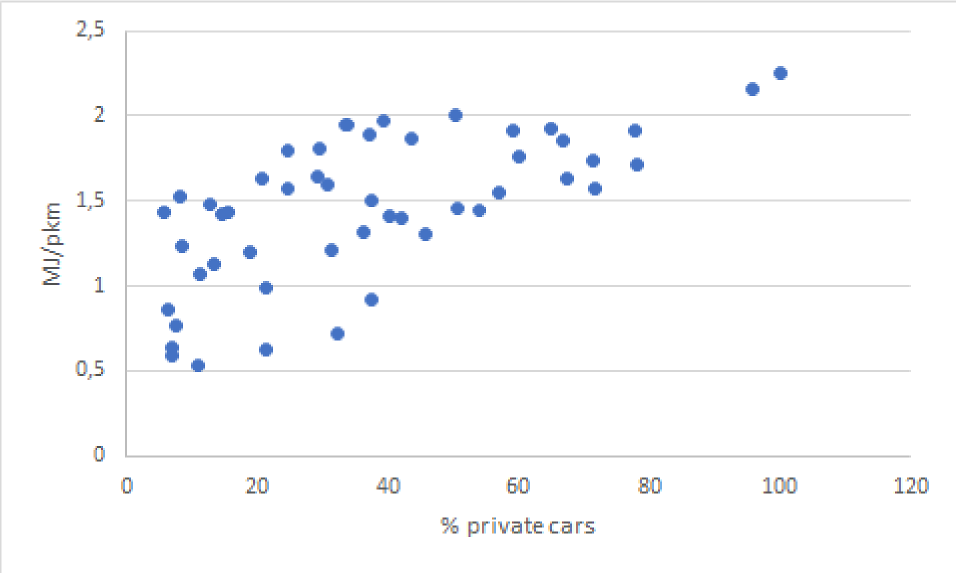

Figure A 9 - MJ/pkm in function of % private cars for Inter trips

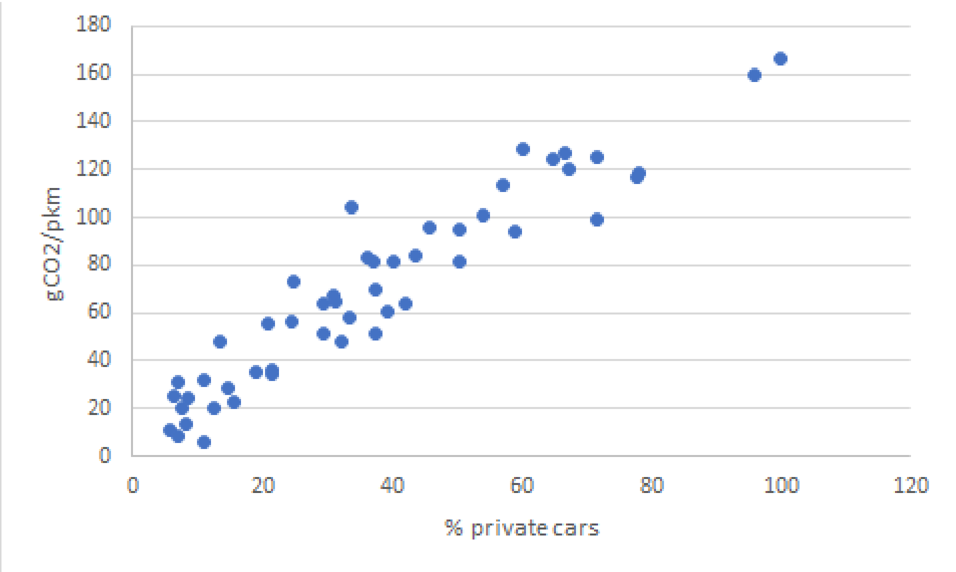

Figure A 8 – gCO2/pkm in function of % private cars for inter trips

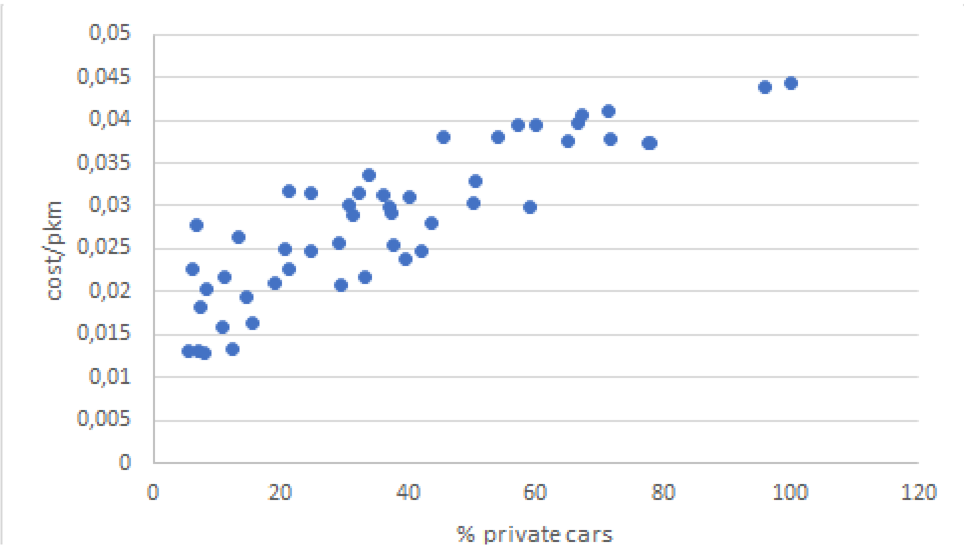

Figure A 10 - cost/pkm in function of % private cars for Inter trips

```r
#' ##### Import Libraries
library(readxl) #Library used to import excel files
library(tidyverse) # Pack of most used libraries
library(skimr) # Library used for providing a summary of the data
library(DataExplorer) # Library used in data science to perform exploratory data analysis
library(corrplot) # Library used for correlation plots
library(car) # Library used for testing autocorrelation (Durbin Watson)
library(olsrr) # Library used for testing multicollinearity (VIF, TOL, etc.)


#' ##### Transform the dataset into a dataframe

dfOD <- data.frame(DataOD)


##### Show summary statistics
skim(df)
summary(df)


shapiro.test(dfROD$gCO2.pass.km)

model <- lm(MJ.pass.km ~ X.PT + Unemployment + X.high_education , data = dfOD)
summary(model)

model <- lm(cost.pass.km ~  X.PT + X.high_education + Unemployment, data = dfOD)
summary(model)

model <- lm(gCO2.pass.km ~  X.PT + X.high_education + Unemployment, data = dfOD)
summary(model)

model <- lm(lnTotalMJ ~  X.PC + Population + X.high_education  , data = dfOD)
summary(model)

model <- lm(lnTotalCost ~ X.PC + Population + X.high_education  , data = dfOD)
summary(model)

model <- lm(lnTotalCO2 ~ X.PC + Population + X.high_education , data = dfOD)
summary(model)


#Tests

par(mfrow=c(2,2))
plot(model)

durbinWatsonTest(model)

ols_coll_diag(model)
```

Figure A 11 - R code

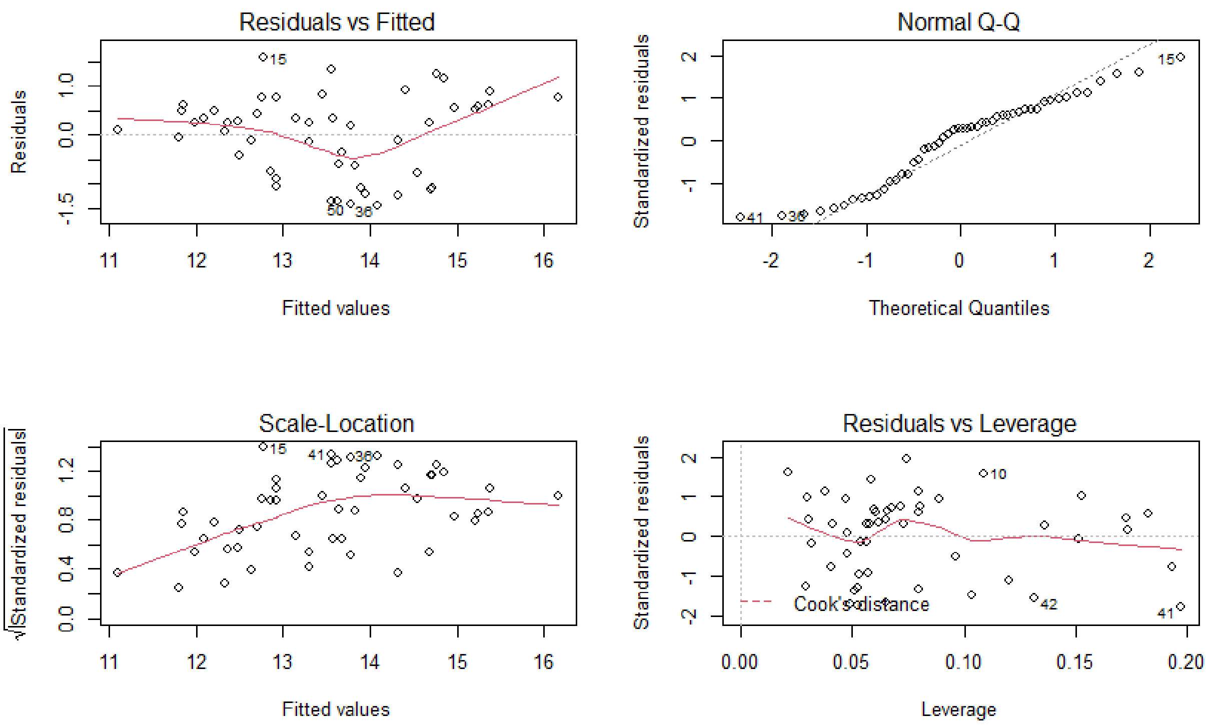

Figure A 12 - Residuals plots of the regression for total energy consumption

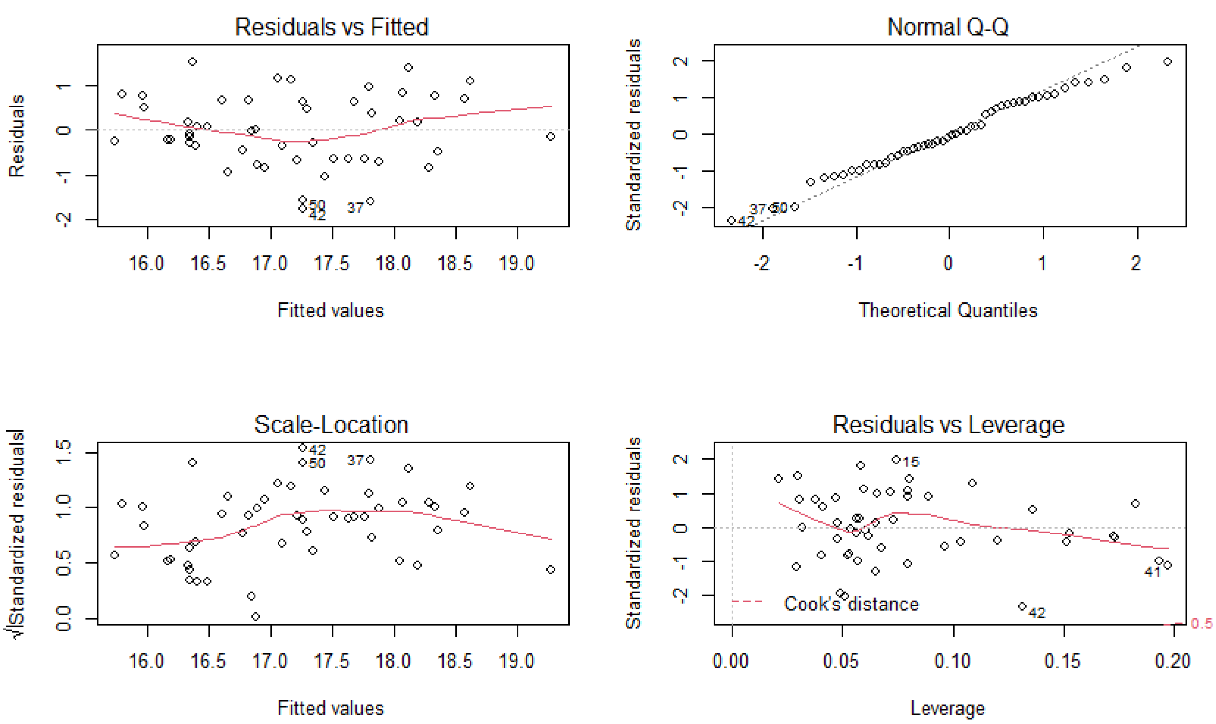

Figure A 13 - Residuals plots of the regression for total CO2 emissions

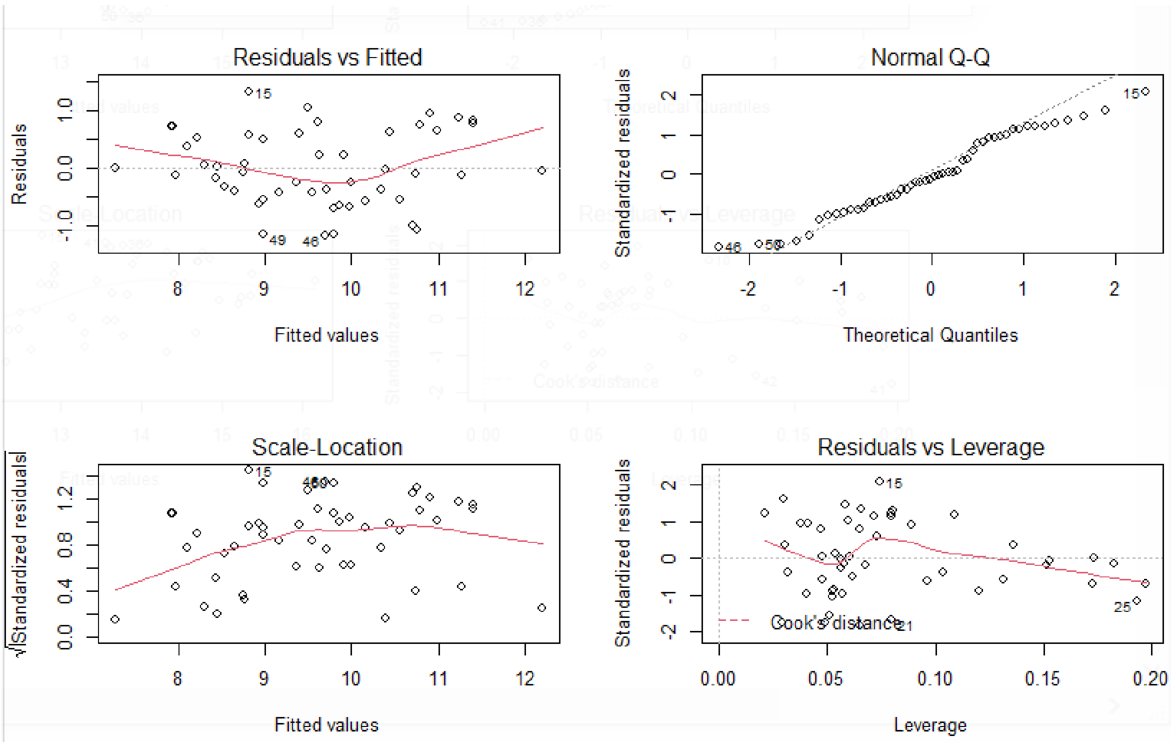

Figure A 14 - Residuals plots of the regression for total cost

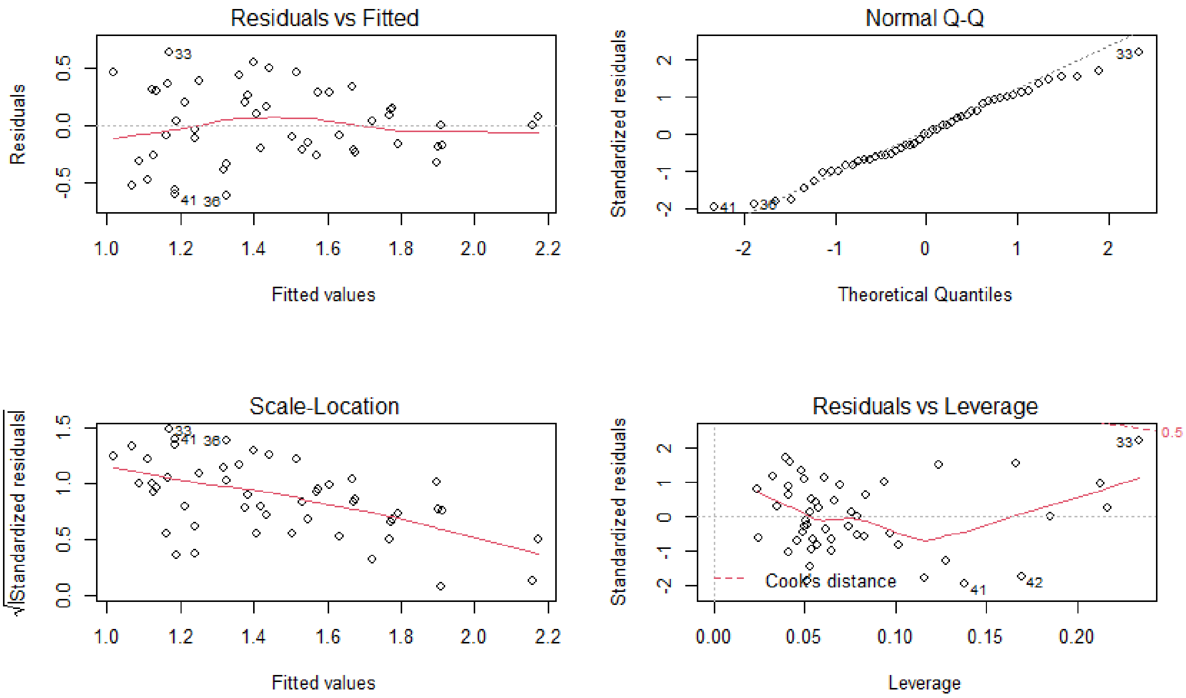

Figure A 15 - Residuals plots of the regression for MJ/pkm

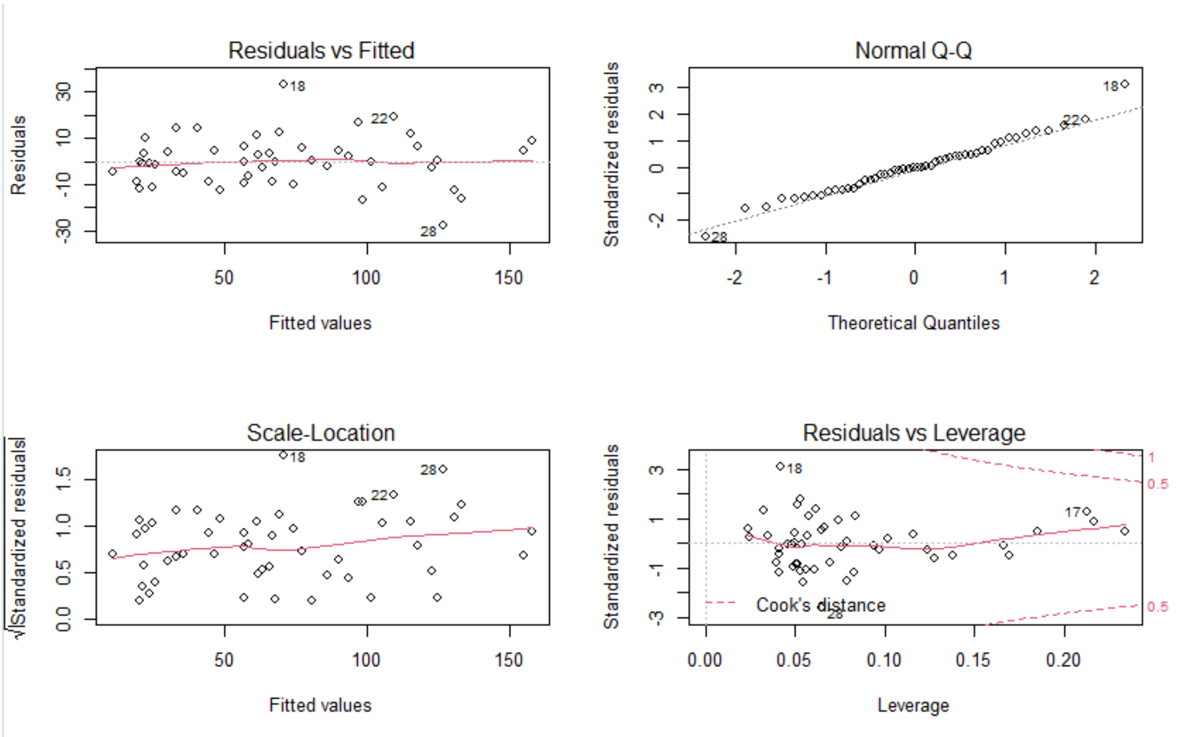

Figure A 16 - Residuals plots of the regression for gCO2/pkm

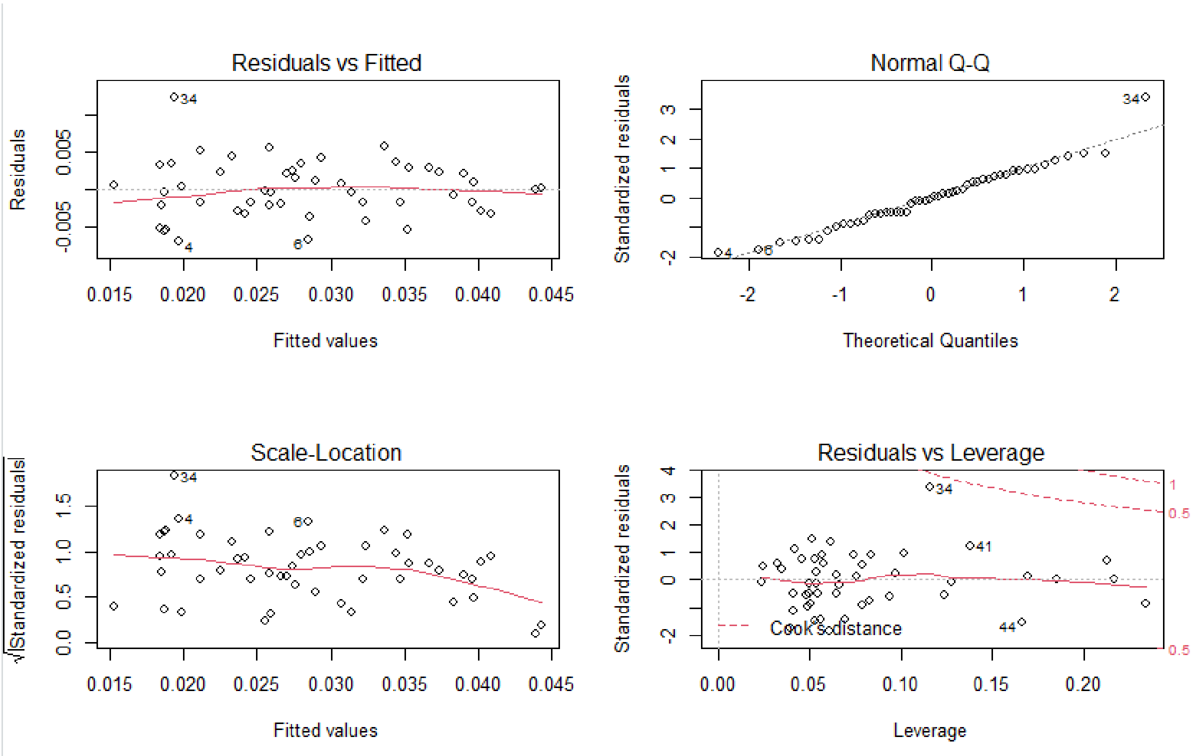

Figure A 17 - Residuals plots of the regression for cost/pkm