Journal of Transport Geography 109 (2023) 103594

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Estimation of trip purposes in public transport during the COVID-19 pandemic: The case of Santiago, Chile

Raúl Pezoa[a,*], Franco Basso[b,c], Paulina Quilodrán[a], Mauricio Varas[d]

[a] Escuela de Ingeniería Industrial, Universidad Diego Portales, Santiago, Chile
[b] School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[c] Instituto Sistemas Complejos de Ingeniería, Chile
[d] Centro de Investigación en Sustentabilidad y Gestión Estratégica de Recursos, Facultad de Ingeniería, Universidad del Desarrollo, Santiago, Chile

ARTICLE INFO

Keywords:
COVID-19
Public transportation
Trip purpose
XGBoost

ABSTRACT

The COVID-19 pandemic strongly affected the mobility of people. Several studies have quantified these changes, for example, measuring the effectiveness of quarantine measures and calculating the decrease in the use of public transport. Regarding the latter, however, a low level of understanding persists as to how the pandemic affected the distribution of trip purposes, hindering the design of policies aimed at increasing the demand for public transport in a post-pandemic era. To address this gap, in this article, we study how the purposes of trips made by public transport evolved during the COVID-19 pandemic in the city of Santiago, Chile. For this, we develop an XGBoost model using the latest available origin-destination survey as input. The calibrated model is applied to the information from smart payment cards during one week in 2018, 2020, and 2021. The results show that during the week of maximum restriction, that is, during 2020, the distribution of trips by purpose varied considerably, with the proportion of trips to work increasing, recreational trips decreasing, and trips for health purposes remaining unchanged. In sociodemographic terms, in the higher-income communes, the decrease in the proportion of trips for work purposes was much greater than that in the communes with lower income. Finally, with the gradual return to in-person activities in 2021, the distribution of trip purposes returned to values similar to those before the pandemic, although with a lower total amount, which suggests that unless relevant measures are taken, the low use of public transportation could be permanent.

1. Introduction

Public transport plays a fundamental role in the functioning of cities worldwide. For example, in some of the world's main capitals, such as Hong Kong, London, and Paris, public transport represents more than 40% of motorized trips (Celi Ortega, 2018). In the case of Santiago, Chile, where the data of this study come from, this percentage reaches almost 50% (Muñoz et al., 2015). However, during the recent COVID-19 pandemic, the use of public transport has been strongly impacted worldwide, with reductions of up to 80% in some cities in Europe (Das et al., 2021). In the same way, in Santiago, Chile, during 2020, a decrease in ticket validations in public transport of 84% was observed compared to the previous year.[1]

In recent decades, the widespread use of technologies, such as smart cards and GPS in buses, has generated massive information on the operation of public transport systems, allowing support for planning and policy decision-making (Gschwender et al., 2016). This type of technology has been beneficial in the context of COVID-19, allowing us to analyze the changes in mobility in public transport in a large-scale manner. In this regard, a large number of recent contributions have analyzed the loss of public transport users using this type of information (Wielechowski et al., 2020; Jenelius and Cebecauer, 2020; Przybylowski et al., 2021; Kopsidas et al., 2021; Awad-Núñez et al., 2021; Das et al., 2021; Gramsch et al., 2022).

However, to the best of our knowledge, studies have yet to analyze the trip purposes in public transport during the pandemic using passive data from the complete system. This gap is mainly explained by the fact that the purpose of the trip is a piece of information that is not available directly from the data collected passively and is usually obtained through surveys. Furthermore, studying trip purposes is relevant

* Corresponding author.
E-mail addresses: raul.pezoa@udp.cl (R. Pezoa), francobasso@gmail.com (F. Basso), paulina.quilodran@mail.udp.cl (P. Quilodrán), mavaras@udd.cl (M. Varas).

[1] https://www.mtt.gob.cl/archivos/25502. Accesed 4 August 2022. In Spanish.

https://doi.org/10.1016/j.jtrangeo.2023.103594
Received 11 September 2022; Received in revised form 15 March 2023; Accepted 17 April 2023
Available online 21 April 2023
0966-6923/

because, due to the pandemic, how passengers value different attributes (e.g., overcrowding) of public transport (Aghabayk et al., 2021; Cho and Park, 2021) has changed. Therefore, studying the disaggregated changes in mobility by trip purpose allows us to understand what type of trip has been most affected and, in this way, develop focused transport policies.

This article studies the purposes of public transport trips during the COVID-19 pandemic in Santiago, Chile. For this, a methodology is developed that estimates the purposes using as input the last origin-destination survey conducted, which provides information such as the purpose stated by the respondent, which is considered the dependent variable to be modeled in the calibration phase. This information is complemented with independent variables computed using land use data and points of interest (POIs) in the city. Then, the calibrated model is applied to the information of smart cards, the only payment method available in the system. These cards provide information on the place and time of boarding of each passenger but do not include the purpose of the trip.

The contribution of this article is twofold. First, a supervised machine learning model is built, which uses both active and passive sources of information and identifies the purposes of travel disaggregating into multiple categories: Work, Study, Return home, Recreational, Health, and Other. Second, a case study is carried out in Santiago, Chile, analyzing how activity patterns are affected as the COVID-19 pandemic progressed in three temporal sections. These three periods correspond to the period before the pandemic, complete quarantine, and relaxed measures.

The rest of this article is organized as follows. Section 2 reviews the literature. Section 3 describes the data used in this article. Section 4 presents the methodology used, and Section 5 applies the methodology to the case study and presents the results obtained. Finally, Section 6 presents the conclusions of this article.

2. Literature review

This literature review is structured into two subsections. The first presents the methodologies developed to determine the trip purposes in public transport. The second section presents the contributions related to the mobility data analysis during the COVID-19 pandemic.

2.1. Determination of trip purposes

The purpose of a trip is the reason for a trip, explaining why a person moves from one point to another. Therefore, understanding the purpose of the trips enriches the underlying information in the data that are generated (Faroqi and Mesbah, 2021) and allows, among other things, understanding the user's choice of mode of transport (Perchoux et al., 2019). Moreover, the literature shows that the perceptions associated with different attributes of public transport (e.g., time values) differ depending on the purpose of the trip (Wardman, 2004). Therefore, having information on the objectives of the trips allows the design of transport policies that make use of these differences.

Traditionally, the information associated with trip behavior, including purpose, is obtained through active methods, such as surveys (Bohte and Maat, 2009). In these, people describe their trip behavior on one or more previous days. Two articles that illustrate the use of active information to study trip purposes in public transport are described next. Muñoz et al. (2015) analyze the results of the origin-destination survey for Santiago, Chile, obtaining trip distributions according to three purposes: work, study, and other. On the other hand, Zailani et al. (2016) examine the intention to use public transport in the central areas of Kuala Lumpur, Malaysia, according to three types of purposes: work/study, shopping, and leisure, using surveys.

However, the use of active information has several drawbacks, including high costs of collecting data and a low frequency of updating (Basso et al., 2022b). Due to these drawbacks, passive data information has gained relevance in recent decades. This type of information corresponds to data that are collected using sensors and technologies, often already available (Yang et al., 2018), and therefore do not require an active role. The availability of passive information in public transportation systems in the world, mainly linked to smart cards and GPS, has enabled the development of new approaches in detecting trip purposes (Faroqi et al., 2018).

The contributions using passive data to detect trip purposes in public transport differ, mainly from the methodological approach used, which can be supervised or unsupervised (Faroqi and Mesbah, 2021). On the one hand, unsupervised learning methods group trips with similar characteristics in clusters and then assign a purpose through an expert criterion or decision rule. This method has the advantage that it does not require the trip purposes variable for training the models but at the expense of needing some prior expert knowledge to assign the purpose to the clusters after calibration accurately (Schwenker and Trentin, 2014). On the other hand, supervised learning methodologies fit a model using historical information to later apply it in another time frame (Gareth et al., 2013). This approach has the advantage that it is more like a 'blank page,' in which the model finds the patterns without using prior expert knowledge. However, as stated in Wang et al. (2015), using supervised learning methodologies might require collecting expensive labeled data to cope with the lack of purpose information in passive data. Relevant examples of articles that use these methodologies to estimate trip purposes in public transport are described below.

Regarding articles that use unsupervised methodologies, Devillaine et al. (2012) identifies the purpose of travel for public transport, performing two case studies, one in Santiago. Chile and another in Gatineau, Canada. In both cases, transaction data from the network are used. However, in the case of Gatineau, information on land use is available. Through a set of rules, this article identifies the purpose of the trip within four options: return home, study, work, and others. On the other hand, Goulet-Langlois et al. (2016) infer the patterns of activities using public transport transactions and sociodemographic information for a subset of users in the city of London, United Kingdom. The authors group the users into trip clusters with similar characteristics to later relate the patterns with the sociodemographic characteristics of the people. Similarly, Yang et al. (2019) analyze data from smart cards in conjunction with points of interest, land use, and information from social networks in the city of Shenzhen, China. The authors group the users according to their spatial and temporal system use patterns to infer the purpose of the trip using the change in land use in the areas of origin and destination of the trips. Finally, Faroqi and Mesbah (2021) estimated the trip purposes in public transport in southern Queensland, Australia, using information from smart cards in conjunction with their temporal attributes and a mobility survey. Using a clustering algorithm, the authors label trips according to five purposes: return home, work, education, shopping, and recreation.

On the other hand, and confined to the use of supervised methodologies, Wang et al. (2017) study the activity patterns using the transactions of the metro network for users in Shanghai, China. Using a logit model, the authors classify trips into three purposes: home, work, and secondary activity. Alsger et al. (2018) study the purposes of travel with information on transactions, land use, and origin-destination surveys, among others, for the southeast of Queensland in Australia. Specifically, the authors use a probabilistic choice model, classifying trips into five purposes: work, education, shopping, home, and recreation. Similarly, Medina (2018) uses transactional information and sociodemographic data for a subset of public transport users in the city-state of Singapore. Using two discrete logit models, the authors establish the primary trip purposes of the users based on three categories, namely, work/study, home, and others. In addition, Kim et al. (2021) estimated the purpose of travel in public transportation for Seoul, Korea, using information from smart cards in conjunction with surveys. Subsequently, using a random forest model, the authors estimate the purpose of the trips within four categories: trips to work, business, leisure, and return home. Finally, Sari Aslam et al. (2021) estimate the trip purposes for the city of London

Table 1
ODS variables of interest.

Variable	Description
Boarding Time	Trip start time (CLT).
Alighting Time	Trip end time (CLT).
Trip Time	Duration of the trip (minutes).
Duration Activity	Duration of the activity performed at the alighting point (Hours).
Stages	Number of stages in a trip.
x.Boarding	X coordinate of boarding the mode of transport (UTM).
y.Boarding	Y coordinate of boarding the mode of transport (UTM).
x.Alighting	X coordinate of alighting the mode of transport (UTM).
y.Alighting	Y coordinate of alighting the mode of transport (UTM).
Purpose	Purpose of the trip

using data from smart cards and information on points of interest. Through a neural network model, the authors classify trip purposes into two categories, namely, primary (work and home) and secondary (leisure, food, shopping, recreational, and travel).

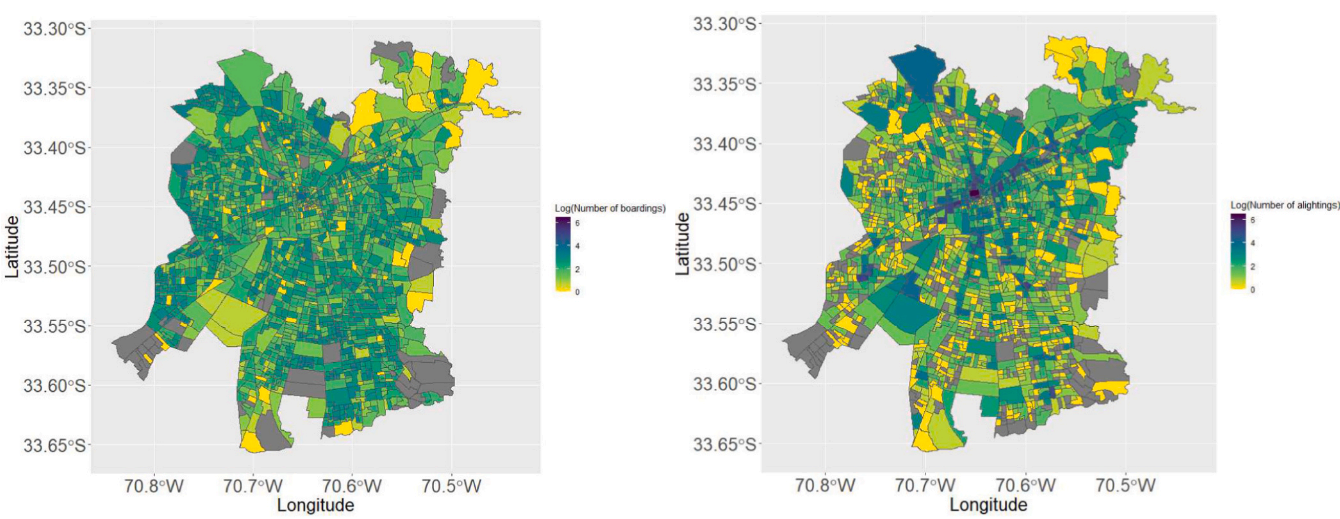
In this article, we estimate the trip purposes by public transport in Santiago, Chile, using a supervised learning methodology. In particular, an XGBoost model is fitted using information obtained actively (origin-destination survey). This methodology, which has not been used in this context, has shown promising results in different areas of transport, such as lane change prediction (Basso et al., 2022a) and accident detection (Parsa et al., 2020).

Table 2
Distribution of purposes in the ODS.

Purpose	Percentage (%)
Return home	46.23
To work	23.47
To study	7.88
Bureaucratic errands	5.65
Shopping	3.90
Visiting someone	3.20
Health	2.81
Recreation	1.69
For work	1.33
Looking for someone or dropping them off	1.21
Other activity	1.19
For study	1.10
Looking for something or dropping it off	0.17
Eat or drink something	0.07

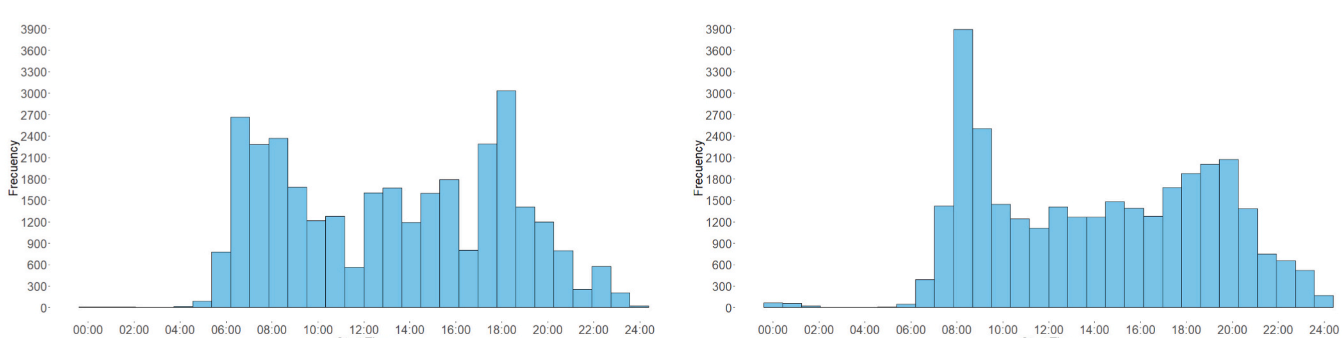
2.2. Mobility in public transport during the COVID-19 pandemic

The COVID-19 pandemic has produced profound changes in mobility in cities (Warren and Skillman, 2020). In particular, public transport has been strongly affected, with a sharp decrease in the number of trips (Gkiotsalitis and Cats, 2021; Vickerman, 2021; Gutiérrez et al., 2021). Chile was not the exception: bus transactions reduced up to 86% in June 2020 (Basnak et al., 2022). In this context, several studies in various parts of the world have described the impact of the pandemic on mobility in public transport using passive data. Some examples, grouped



(a) Boarding (b) Alighting

Fig. 1. Distribution of the first trip of the day.



(a) Boarding (b) Alighting

Fig. 2. Distribution of trip boarding and alighting times.

Table 3
Categorization of purposes.

Category	Purpose	Percentage (%)
Work	To work	24.81
	For work	
Study	To study	8.98
	For Study	
Return home	Return home	46.23
Recreational	Eat or drink something	4.97
	Visiting someone	
	Recreation	
Health	For health	2.81
Other	Look for or leave someone	12.14
	Look for or drop something off	
	Shopping	
	bureaucratic errands	
	Other	

Table 4
Trip information by category.

	Work	Study	Return home	Recreational	Health	Other
Number of trips	7767	2820	14,516	1561	897	3032
Average Activity Time (Hours)	10.09	7.53	15.30	9.38	3.77	4.74
Average trip time (Minutes)	64.82	50.53	61.95	55.95	52.88	45.37
Average distance (km)	10.01	6.97	8.30	7.25	6.46	5.94

Table 5
Variables of the land use base.

Variable	Description
Latitude	Latitude associated with the property.
Longitude	Longitude associated with the property.
Dimension	Dimension of each property (m^2).
Destination	Use or destination given to the property.

Table 6
Transactions base variables.

Variable	Description
Card number	Card identifier
Mode	Type of associated transport (Bus, Metro, Train)
Timestamp	Date and time of the transaction
Place	Place of boarding (License plate or Station)

Table 7
GPS base variables.

Variables	Description
License	Bus license
Timestamp	Date and time of registration
Latitude	Geographical position
Longitude	Geographical position
Service	Service that operates the bus
Sense	Direction of service (Outbound or Return)
Velocity	Instantaneous velocity

Table 8
Information on the application periods.

	Period 1	Period 2	Period 3
	(11–15 June 2018)	(06–10 July 2020)	(04–08 October 2021)
Transactions	27,276,935	3,490,541	15,782,024
Number of Buses	6582	6494	6744
Number of GPS pins	62,951,696	44,304,751	59,932,152

Table 9
Distribution of points of interest.

Type	Percentage(%)
Companies	83.04
Stores	8.48
Food	3.34
Education	2.33
Bureaucratic errands	0.86
Parks	0.80
Recreation	0.48
Health	0.47
Churches	0.15

by geographic location, are described below. The reader should note that the field of COVID-19 is still moving fast, so the overview may not be complete for this reason.

In the European context, Jenelius and Cebecauer (2020) analyzes the impact of COVID-19 on the number of travelers in public transport for the three most populated regions of Sweden. Using passive data, they find a decrease of between 40% and 60% of passengers. In addition, the authors show that for their data, there is no strong correlation between the number of passengers and the rate of transmission of the virus at an aggregate level. Almlöf et al. (2021) studied the factors that affect trip behavior during COVID-19 in Stockholm, Sweden. For this, data from smart cards is used to generate clusters of groups with similar characteristics according to sociodemographic factors. One of the results found indicates that the higher the income and the older the user, the more likely they are to abandon public transport.

On the other hand, in North America, Wilbur et al. (2020) analyze the impact of COVID-19 on transportation accessibility and its demand, differentiating by socioeconomic groups in the cities of Nashville and Chattanooga, USA. For the above, the authors use data from the boarding of passengers and socioeconomic information, which is obtained through surveys. Through a correlation analysis, it is found that the higher the average income of the neighborhood is, the greater the decrease in demand for public transportation. Similarly, Parker et al. (2021) combined mobile phone data and surveys collected from a sample of people from 26 states in the USA to study the sociodemographic factors that impact the behavioral changes of public transport users. The article establishes that public transport users decreased their trips more than users of other modes. Deschaintres et al. (2022) studies which COVID-19 measures have more impact on subway daily ridership in Montreal, Canada. The authors use automatic fare collection data, concluding that restrictive measures had an immediate but decreasing impact on subway use, while release measures led to a gradual recovery.

In Asia, Zhang et al. (2021b) analyze the changes in mobility in public transport of different groups of people in Hong Kong, China. For this, the authors use various sources of information, including smart cards, demographic information, and data on the total number of COVID-19 infections. The authors characterize the change in total trips for each group, showing that children reduced the use of the metro system by 86% and students by 73%, which is explained by the closure of schools. In addition, the authors highlight that mobility to work, shopping, and leisure areas decreased by 42%, 80%, and 99%, respectively. Mützel and Scheiner (2021) analyzed the spatiotemporal patterns and changes in the use of subways during COVID-19 in the city of Taipei, Taiwan. For the above, the authors use information from smart cards and information from metro stations. The authors find that trip patterns did not change significantly, despite the decrease in total trips.

Finally, in Chile, the country in which this research is focused,

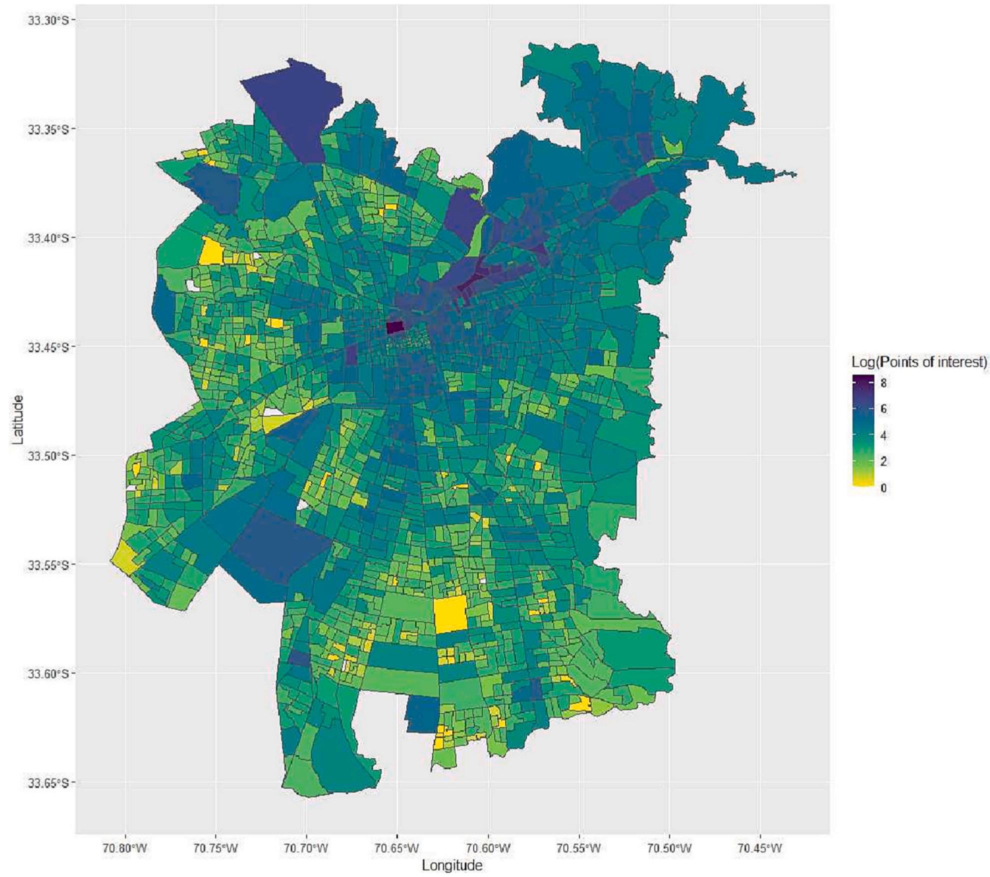


Fig. 3. Points of interest in the Metropolitan Region.

Gramsch et al. (2022) analyze the effect of dynamic quarantine on demand for public transport in Santiago using smart card data. One of the findings indicates that the suspension of classes and self-confinement at the beginning of the pandemic decreased the demand for public transport by up to 72.3%, while the partial quarantine by commune reduced the demand by 12.1% on average.

From the previous review, it is observed that passive information on public transport has been a key source for studying mobility during the pandemic due to its massiveness and high accessibility. However, to the best of our knowledge, no article has studied trip purposes and their variation in this era of the pandemic using this type of information for a complete network. Therefore, there is a poor understanding regarding whether the decrease in trips is uniform for different purposes, which makes it difficult to design policies aimed at increasing the demand for public transport in the post-pandemic era. It should be noted that a study similar to ours is that of Astroza et al. (2020), which quantifies, in Chile, the difference in the number of trips according to the purpose (work, study, others) and mode. However, unlike this work, in their article, the authors use active data collected from a sample of people during only the beginning of the pandemic.

3. Data description

For the estimation of trip purposes in public transportation in Santiago, Chile, multiple data sources are used. Subsection 3.1 describes the data from the origin-destination survey (ODS). Subsection 3.2 describes the land use data, which comes from the Internal Revenue Service (SII, for its acronym in Spanish). Subsection 3.3 describes the passive data associated with public transportation, namely, transactions and GPS records. Subsection 3.4 describes the data of points of interest in the city, coming from the OpenStreetMaps platform. Finally, Subsection 3.5 describes the social priority data in the region, which are generated by the Ministry of Social Development and Family.

The transportation data used in this study has become one of the key elements for public transport planning in Santiago, Chile, due to its reliability and availability. Indeed, as Gschwender et al. (2016) describes, many planning and operational decisions are supported by tools based on this passive data. For instance, origin-destination matrices built using the GPS and transaction data are used both by the operators and by the transport authority to evaluate, for example, the impact of new direct bus lines (Munizaga et al., 2014). Additionally, the origin-destination survey conducted in 2012 was used to evaluate the impact of transport projects and policies implemented since the previous mobility survey, conducted in 2001 (Contreras et al., 2016). Finally, the origin-destination survey and transaction data show similar mobility patterns, supporting the reliability of both data sources (Muñoz et al., 2015).

3.1. Origin-destination survey

The ODS is a mobility survey carried out usually every ten years in Santiago, Chile (SECTRA, 2014). This survey aims to capture the travel behavior patterns of Santiago's travelers and their socioeconomic characteristics by gathering disaggregated data (Muñoz et al., 2015). The 2012 ODS survey, which was conducted between July 2012 and November 2013 covered the 45 communes of the Metropolitan Region. It surveys almost 18,000 households in 866 zones through face-to-face interviews. People were asked about the trips made within public areas and their activities on a particular day (Kickhofer et al., 2016). The blocks surveyed in the city are chosen according to the *Probability Proportional to Size* method. Then, in each block, the number of households surveyed increases with the total number of households. This survey generated information on 113,592 trips, of which 31,398 correspond to trips on public transportation. In this paper, we consider only trips made on public transportation. Table 1 shows the subset of ODS variables used in this research.

Fig. 1 shows the distribution of boardings and alightings of the day's first trip in public transport, which is obtained from the coordinate variables detailed in Table 1. From Fig. 1a, it can be deduced that the boardings are concentrated mainly in the peripheral areas of the city, which have, in general, lower incomes. From Fig. 1b, on the other hand, it follows that the alightings of the first trip of the day occur mainly in the central and northeast areas. This is explained by the fact that these are the main centers of economic activity in Santiago.

On the other hand, Fig. 2 shows the distribution of the trips' boarding and alighting start times recorded in the ODS. In consideration of this, it is possible to observe that there are two activity peaks in the city of Santiago. The first peak begins at approximately 7:00 am and coincides with the start of the work activities. The second peak, on the other hand, begins at approximately 18:00, which corresponds to people returning to their homes.

Table 2 shows the distribution of trip purposes according to the ODS. This survey considers 14 trip purposes, the most common being *Return home* (46.09%), and the least common, *Eating or drinking something* (0.08%). Other non-daily activities, such as *Bureaucratic errands* (paying bills and visiting state offices), *Shopping*, *Visiting someone* and *Health* (medical check-up/treatment) range between 2.81% and 5.65%. Note that several purposes present in the ODS represent similar trip objectives. For example, *To work* and *For work* are strongly related. To simplify the analysis, the purposes contributing to the same objective are grouped into the same category. The distribution of these categories, as well as the purposes they include, are shown in Table 3.

Table 4 presents descriptive statistics of the trips grouped by category. In particular, the categories *Work* and *Going Home* dominate the rest in all the metrics presented. On the other hand, the *Recreational*, *Health*, and *Other* categories include trips that, on average, have considerably shorter activity times, trip times, and distances than the previous categories. One of the reasons that explain the latter is that, for the categories *Recreational*, *Health* and *Other*, people have greater flexibility in choosing a destination, so they may seek to minimize the distance or time of the trip. In contrast, for the purposes of *Work* or *Returning Home*, the destination corresponds to a strategic decision in the medium and long term, where the distance and trip time are only one of the factors that affect the decision of the workplace/location.

3.2. Land use

The land use base used in this research includes the variables described in Table 5 for each property in the Metropolitan Region. This information is provided by the Internal Revenue Service (SII), and in particular, it details the use or destination that is given to each property within the following options: Housing, Industry, Commerce, Education, Warehouse, Office, Sports, or Health.

Appendix A shows the distribution of land uses for each commune of Santiago. From this, it is observed that, for all communes, except Pudahuel and Quilicura, the predominant land use is Housing. For Pudahuel, the highest percentage corresponds to warehouses (46.1%), which is explained by the fact that this commune has access to the routes connecting the two main ports of Chile, Valparaíso and San Antonio, generating incentives to locate the warehouses there. In the case of Quilicura, the highest percentage corresponds to industry (40.55%). This is explained by the changes in the regulatory plans of the city of Santiago during the 1990s, which promoted the use of the northern part of the city for industrial purposes (Neira and Undurraga, 2003).

3.3. Transactions and GPS

Public transportation within the Metropolitan Region consists of an integrated network of buses, metros, and trains. In this network, the only possible payment method is smart cards, which generate a considerable volume of trip records daily (Devillaine et al., 2012). On the other hand, each bus in the public transport system emits a GPS signal every 30 s. Both sources of information, that is, transactions and GPS, are provided by the Metropolitan Public Transport Directory (DTPM), the entity responsible for planning the system. Table 6 shows the fields available in the system transactions database, while Table 7 shows the fields available in the GPS database of the public transportation system buses. It is important to note that ticket validation occurs only upon boarding, so there is no information on the point of alighting.

In this study, we consider three periods of interest. Period 1, which runs from June 11 to 15, 2018, corresponds to a situation before the pandemic. Period 2, which runs from July 6 to 10, 2020, corresponds to a complete quarantine period in Santiago. Finally, Period 3, which runs from October 4 to 8, 2021, corresponds to a period without quarantine in the city. Table 8 presents a descriptive statistic of the GPS information, transactions, and supply of the system for each study period. The public transport transactions decreased by 87.20% in the complete quarantine period, that is, period 2, compared to period 1, prior to the pandemic. However, the supply of buses was not diminished in the same way. In particular, the number of buses operating decreased by only 1.34%, while the number of GPS records was reduced by 29.62%. This occurs because, during the pandemic of COVID-19, the Chilean Government imposed curfews restricting public transport working hours. On the other hand, when the confinement measures were relaxed in period 3, the transactions in the system increased considerably. However, the number of recorded transactions is still 42.14% lower than pre-pandemic levels. Finally, both the number of buses operating and the GPS records during period 3 present levels of magnitude similar to the prepandemic (period 1).

3.4. Points of interest

The POI base corresponds to the location of different points of interest in the city of Santiago, Chile. This database is obtained through the use of the OpenStreetMap platform and includes the geographical location of 113,453 points of interest. In particular, these places are grouped into nine categories: *Companies*, *Stores* (grocery stores, marketplaces, malls, etc.), *Food* (restaurants, fast food joints, pubs, etc.), *Education* (universities, schools, etc.), *Bureaucratic errands* (banks, courthouses, post offices, etc.), *Parks*, *Recreation* (cinemas, theaters, etc.), *Health* (clinics, dentists, etc.) and *Churches*. Table 9 shows the distribution of the points of interest according to these categories, while Fig. 3 illustrates the spatial distribution of these points in the Metropolitan Region. In this regard, note that there is a greater concentration of points in the central and northeast of the city. This is explained by the fact that the main centers of economic activity in Santiago are located in these areas.

3.5. Social priority index

The social priority index (SPI) is an indicator created by the Ministry of Social Development (MDS), which quantifies the relative standard of living reached by the population of that commune.[2] This index ranges between 0 and 100, where a higher value indicates that the commune contains inhabitants of a lower socioeconomic level. Fig. 4 shows the distribution of the social priority index in the Metropolitan Region. From this figure, clear spatial segregation can be noted in Santiago. The communes with the lowest index are clustered in the northeast area of the city.

[2] https://www.desarrollosocialyfamilia.gob.cl/storage/docs/boletin_interno/INDICE_DE_PRIORIDAD_SOCIAL_2020(1).pdf

4. Research methodology

4.1. Overview of the proposed approach

In this subsection, we present a summary of the proposed method, as shown in Fig. 5. Our approach starts by fitting a multiclass classification model that aims to predict the purpose of a trip. This model is calibrated using the trips reported by the ODS (Subsection 3.1). As mentioned before, these trips include the reported purpose (that is, our dependent variable). Then, the independent variables considered in our model are built using the Land Use (Subsection 3.1), POI (Subsection 3.4), and SPI (Subsection 3.5) datasets. These independent variables aim to characterize the trip (duration, stages, etc.), as well as the boarding and alighting zones (land use, amenities, etc.), and are described in Subsection 4.2. The model proposed is a concatenation of three XGBoost models, as described in Subsections 4.3 and 4.4. Once this model is calibrated, we apply it to the three study periods (2018, 2020, and 2021). To apply the model, we build the same independent variables for each trip recorded in the transaction data (Subsection 3.3) during these periods. Note that this application requires knowing the boarding and alighting point of every trip. Although the boarding point is recorded in the transaction data, the alighting point is not. This issue is tackled in Subsection 4.5. Finally, the output of our procedure is an estimated purpose for every trip recorded in the transaction data for the three study periods.

Note that our method is based on information obtained actively (in particular, the ODS data). Consequently, it might suffer from some drawbacks of this type of information, e.g., misreport of travel times or distances (Stopher et al., 2002) and low survey application frequencies. However, our method should not require frequent updates of this active data. Indeed, as long as trips of a given purpose keep the same characteristics over time, our method will be able to classify them correctly. In other words, the underlying assumption of our method is that, for example, work-related trips during 2012 have similar characteristics to those made during 2018, 2020, and 2021 (in terms of land use and amenities near the origin and destination of the trip, trip distance, stages, etc.). We think this is a reasonable assumption. We discuss this in detail in Subsection 4.3. In particular, our method does not require that the distribution of trip purposes remain stable over time. In fact, one of our results shows (Subsection 5.2) that recreational trips during the complete quarantine period represent a much lower percentage of the total trips compared to the ODS data. This is explained because trips with characteristics similar to recreational-related trips in the ODS data are less frequent in the quarantine data. Finally, note that the application of our method requires only passive data (GPS and transactions).

4.2. Construction of input variables

The trips recorded in the ODS (described in Subsection 3.1), together with the purposes stated by the users, are used as the training dataset to fit a supervised model that allows classifying trips according to purpose. d To achieve this, each ODS trip is associated with a set of explanatory variables, which are shown in Table 10.

The explanatory variables considered are divided into five groups. The first corresponds to variables that describe the characteristics of the trip, using information from the ODS. The second corresponds to variables that characterize the socioeconomic information of the point of boarding and alighting of the trip using the SPI. The third includes variables that describe the land use of the place of boarding and alighting. The fourth corresponds to variables that quantify the points of interest in the boarding and alighting zone. Finally, the *Ratio* variable is added, which is constructed through the land use and POI database. To construct this variable, first, the city of Santiago is partitioned into 1000 hexagons, which are then grouped into 14 clusters with similar characteristics (according to land use and POI) using a K-means model. Thus, the *Ratio* variable is constructed following Eq. (1). In this equation, the numerator corresponds to the distance between the origin and the closest point in the destination cluster. The denominator represents the distance of the trip. This variable, therefore, compares the distance of the trip made to the minimum distance that allows accessing a similar land use type and points of interest. Intuitively, when this variable takes values close to 1, it suggests that the trip has a more interchangeable destination, and thus, people choose one with a short distance to the origin (e.g., shopping). In contrast, when this variable takes values close to 0, the distance of the trip is much greater than the minimum possible, suggesting that people have a fixed destination (e.g., work). Note that the underlying reasoning behind the intuition for the variable Ratio is that, for example, working destinations are less interchangeable than convenience store destinations. In other words, the distance is a more relevant factor when choosing where to shop than when choosing where to work, resulting in distances closer to the minimum possible for some types of trips (e.g., shopping) than others (e.g., work).

$$\text{Ratio} = \frac{\text{Distance}_{\text{origin}-\text{cluster}}}{\text{Distance}_{\text{origin}-\text{destination}}} \tag{1}$$

4.3. XGBoost

To estimate the trip purposes, we build a supervised multiclass classification model. As Fig. 6 illustrates, this model aims to separate the different classes of observations (different trip purposes) using the variables described in Subsection 4.2 as predictors. In our case, the model is built using the ODS data (training data), described in Subsection 3.1. Once the model is built, we use it to predict the purpose of trips made during the three periods of interest (Periods 1, 2, and 3). In terms of the illustrative Fig. 6, the yellow, blue, and green points represent the training observations, while the red point represents a new observation. The model would then declare this new observation as class blue.

In this paper, we use the extreme gradient boosting (XGBoost) model (Chen and Guestrin, 2016; Mitchell and Frank, 2017). XGBoost is a supervised machine-learning algorithm that assembles sequentially constructed decision trees. In this boosting model, each tree learns from the errors made by the previous trees. XGBoost combines individual learners that use the boosting concept in order to create dependencies, while effectively building and running processes in parallel. The best splits within each tree are found by enumerating all the possible splits on all the features. This procedure is called the *exact greedy algorithm* and is performed efficiently by sorting the data according to feature values and visiting the data in the sorted order. However, the sorting algorithm might be pretty time-consuming; thus, the implementation proposed by Chen and Guestrin (2016) proposes to store the data in in-memory units called *blocks*. Within each block, a compressed column format is used. This sorted input data layout is computed only once before training, being reusable in future iterations (Rahman et al., 2021).

The objective function L used in the fit of this model is presented in Eq. (2), which establishes a trade-off between model performance and its complexity. In effect, the function l corresponds to a convex loss function, which measures the difference between the predicted $\widehat{y}_j$ and the real y_j, where N corresponds to the number of observations. On the other hand, the second component corresponds to a regularization term, which seeks to control the complexity of the K trees generated. In particular, as shown in Eq. (3), $\phi(f_k)$ quantifies the complexity of the k-th tree, increasing with the depth of the tree. In this equation, the term T corresponds to the number of leaves on a tree, λ and γ are parameters of the model, and w is the weight vector of the leaves. Model. A grid search is used to select the hyperparameters of the model. Overall, the regularized objective function aims to select a model employing simple but predictive functions (Brownlee, 2016; Boehmke and Greenwell, 2019).

$$L(\phi) = \sum_{j=1}^{N} l\left(\widehat{y}_j, y_j\right) + \sum_{k=1}^{K} \phi(f_k) \tag{2}$$

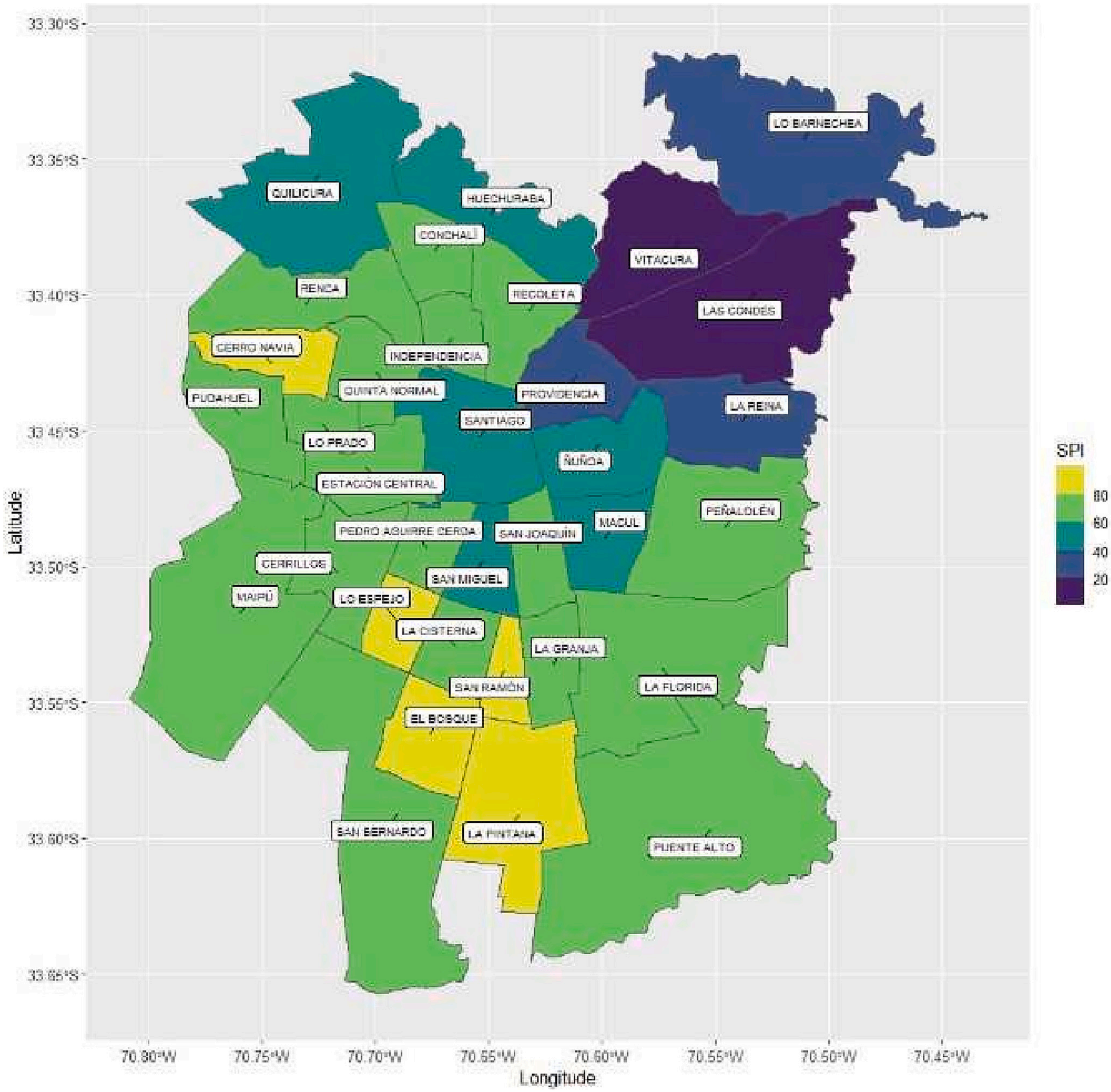


Fig. 4. Classification according to SPI in each commune of Santiago.

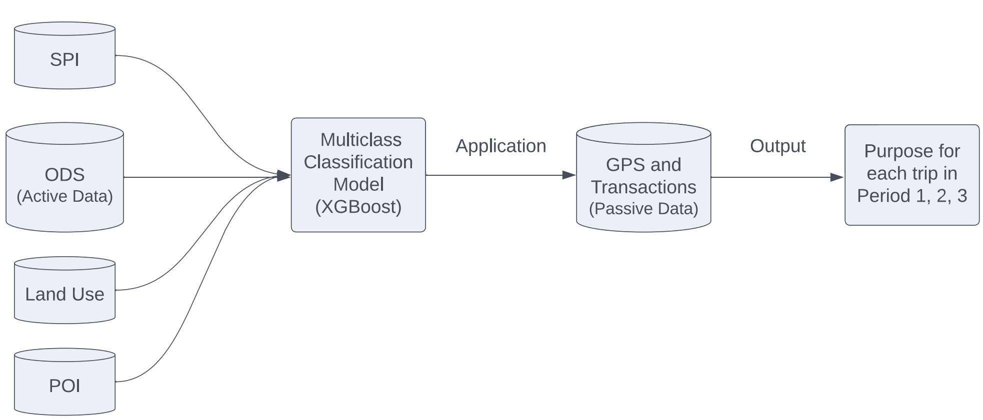


Fig. 5. Overview of the proposed approach.

Table 10
Variables used for the purpose estimation model.

Name of variables	Definition	Source
Purpose	Purpose associated with the trip	
Stages	Number of stages in a trip	
TripTime	Duration of trip from origin to destination in minutes	
ActivityDuration	Duration of activity performed	
Boarding.Time	Time of trip boarding	
Alighting.Time	Time of trip alighting	ODS
Distance	Distance between the point of boarding and alighting in meters	
FirstTrip	Indicator, takes value 1 if the trip corresponds to the first of the day	
LastTrip	Indicator, takes value 1 if the trip corresponds to the last of the day	
SPI Boarding	Social priority index of the commune associated with the trip boarding	SPI
SPI Alighting	Social priority index of the commune associated with the trip alighting	
Percentage.Y. Boarding	Percentage of land use of type Y within a 500 m radius from the point of boarding	
Percentage.Y. Alighting	Percentage of land use of type Y within a 500 m radius from the point of alighting	
Meters.Y.Boarding	Square meters of land use of type Y within a 500 m radius from the point of boarding	Land Use
Meters.Y.Alighting	Square meters of land use of type Y within a 500 m radius from the point of alighting	
X.Boarding	Number of points of interest of type X within a 500 m radius from the point of boarding	POI
X.Alighting	Number of points of interest of type X within a 500 m radius from the point of alighting	
Ratio	Value between 0 and 1 that responds to the division between the origin distance cluster and the origin destination distance	Land Use and POI

$$\phi(f_k) = \gamma T_k + \frac{1}{2}\lambda \|w_k\|^2 \tag{3}$$

The choice of the XGboost model over others is explained by the fact that this algorithm has shown remarkable prediction accuracy in the literature (Torlay et al., 2017; Li et al., 2019; Ogunleye and Wang, 2019). Even though these promising results, XGBoost has received some critics for being complex and challenging to interpret (Carmona et al., 2022). Nevertheless, our research aims to analyze the results of the predicted trip purposes rather than understanding the model itself. Additionally, note that the data we use to train the model is over ten years old. In this context, it is possible that the phenomenon we study may have experienced some type of *dataset shift*, that is, a change in the distribution of a single feature, a combination of features, or the class boundaries (Moreno-Torres et al., 2012). The underlying assumption in this paper is that *population drift* (also called *covariate shift*) is the only type of dataset shift that may be present. In other words, we assume that the conditional probability $p(y|x)$ remains unchanged between the training dataset (ODS survey) and the application datasets (the three study periods), but the input distribution $p(x)$ may differ, where y is the trip purpose, and x is the set of independent variables (Bickel et al., 2007). Note that this assumption allows, for example, a different distribution of trip purposes $p(y)$ between the training dataset and the application datasets, as we obtain in Subsection 5.2. Previous research shows that complex classifiers (such as decision tree based models) are more robust to this type of dataset shift, remaining more accurate than simpler models (Alaiz-Rodrguez and Japkowicz, 2008). Thus, we think XGBoost is the appropriate method to accomplish this task.

4.4. *Concatenated model*

For the construction of the model, two groups of purposes are generated. The purposes included in each of these groups are presented in Table 11. The reasoning behind this separation is to generate groups of purposes with similar characteristics to improve the predictive capacity in each of them. In particular, as described in Subsection 3.1, trips from *Group 1* tend to have longer distances and trip times than trips from *Group 2*. Additionally, trips from *Group 1* present a lower average value for the variable *Ratio*, compared to trips from *Group 2* (0.20 versus 0.34), supporting the intuition that some trips (those from *Group 2*) have a more interchangeable destination. Then, the resulting model (from now on, the concatenated model) consists of the concatenation of three submodels, with a fitting scheme shown in Fig. 7. The first of these models corresponds to a binary model that determines if the trip is of *Group 1* or *Group 2*. The second and third models determine the specific purpose within these categories, taking into consideration the first binary decision. Computational experiments showed that using this modeling scheme leads to better performance results than using a single model (that is, a model that directly classifies a trip into one of the six purposes considered).

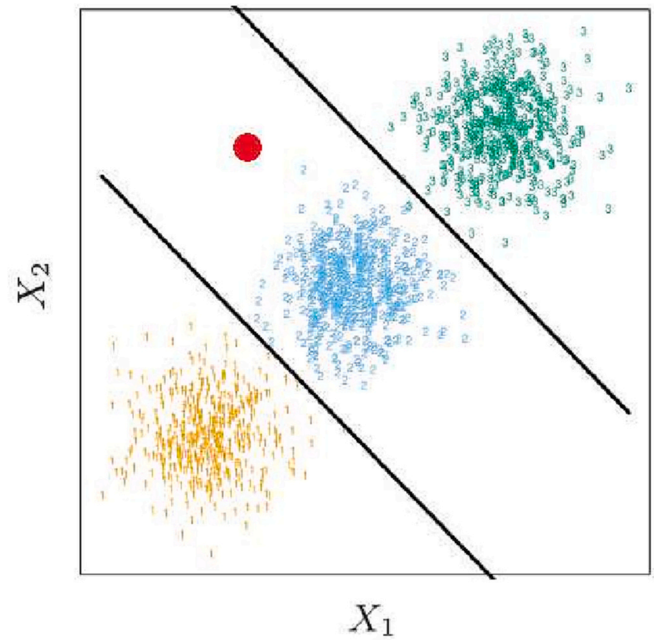


Fig. 6. Example of multiclass classification. Based on Hastie et al. (2009).

4.5. *Estimation of alighting points*

In the city of Santiago, Chile, transactions in the system only record the time and place of boarding of passengers. However, to apply the fitted model to the three study periods, it is necessary to know the alighting location. Therefore, to obtain information on the alighting point, it is necessary to make estimates. In this work, we use the methodology proposed by Munizaga and Palma (2012)), where the alighting point of a user is estimated minimizing the generalized trip time, as shown in Problem (4).

$$\begin{aligned} \min_{i \in \text{Stops}} \quad & t_i + f_w \cdot \frac{d_{i\text{-post}}}{s_w} \\ \text{s.t.} \quad & d_{i\text{-post}} < d \end{aligned} \tag{4}$$

In this Problem, the generalized trip time if getting off at bus stop i corresponds to the weighted sum between the bus trip time if the user gets off at the bus stop i, t_i, and the walking time between the bus stop i and the stop where the user gets their next validation. The latter, in particular, is defined as the ratio between $d_{i-\text{post}}$ and s_w, which corresponds to the distance between the whereabouts i and the place of the next validation, and the walking speed, respectively. The penalty factor, f_w, reflects that walking leads to greater disutility than in-vehicle travel

Table 11
Types of purposes.

Group 1	Group 2
Work	Recreational
Study	Health
Return Home	Other

time. In this way, Problem (4) performs a search on all bus stops i that are within a maximum walking distance d of the location of the next validation. The stop that reaches the optimum of this Problem is considered the point of alighting. Then, note that in this paper, we aim to characterize the trip destination as the place where a person goes to conduct an activity. However, we do not have information on the actual place where people conduct activities, only their behavior given by smart card data transactions. Thus, we proxy the destination place as the last stop in a trip, under the assumption that the activity was made in the proximity of this point. However, to find this point, it is necessary to differentiate trips from trip segments. In other words, we are not interested in intermediate destinations caused by a transfer from one bus to another, only in final destinations. For this purpose, we follow (Munizaga and Palma, 2012): (i) if a person (card) stays for longer than 30 min in a particular point, then it is a final destination, in any other case, it is a transfer point (ii) if two transactions in a row are made in the Metro or the same bus route, a final destination is assumed between the two, regardless of the time interval.

5. Results and discussion

This section describes the results of the application of the methodology described in Section 4 to the data presented in Section 3. In particular, Subsection 5.1 shows the parameters used to fit the trip purpose prediction model, as well as its performance metrics. Then, Subsection 5.2 analyzes the results in the case of Santiago, Chile.

Table 12
Parameters of each XGBoost model.

Parameters	First stage model	Second stage model 1	Second stage model 2
objective	binary:logistic	multisoftmax	multisoftmax
min_child_weight	16.71622	24.06922	17.73402
num_class	2	3	3
gamma	0	0	0
subsample	0.75	1	1
colsample_bytree	1	0.4	1
Max_depth	10	6	8
eta	0.02	0.02	0.016
F1 test	0.888	0.860	0.724

5.1. Fit and validation of the model

For fitting the concatenated model, the trips in public transport reported in the ODS are considered, which is described in Subsection 3.1 and calibrated using the procedure described in Section 4. The hyperparameters of the resulting XGBoost models, as well as the F_1 of each of them, are presented in Table 12. Note that, due to the black-box nature of the XGBoost model, the hyperparameter values provide limited interpretability about the effect of the independent variables over the prediction (Sagi and Rokach, 2021). Indeed, the choice of the XGBoost model in this paper is motivated by its predictive performance rather than its interpretability. In the context of prediction performance, Fig. 8 compares, for the test base, the percentage of trips of each purpose in the ODS with the predictions of the concatenated model. Based on the above, it is established that the concatenated model is very capable of replicating the ODS observations at an aggregate level, presenting differences not greater than 2.7 percentage points for each purpose considered.

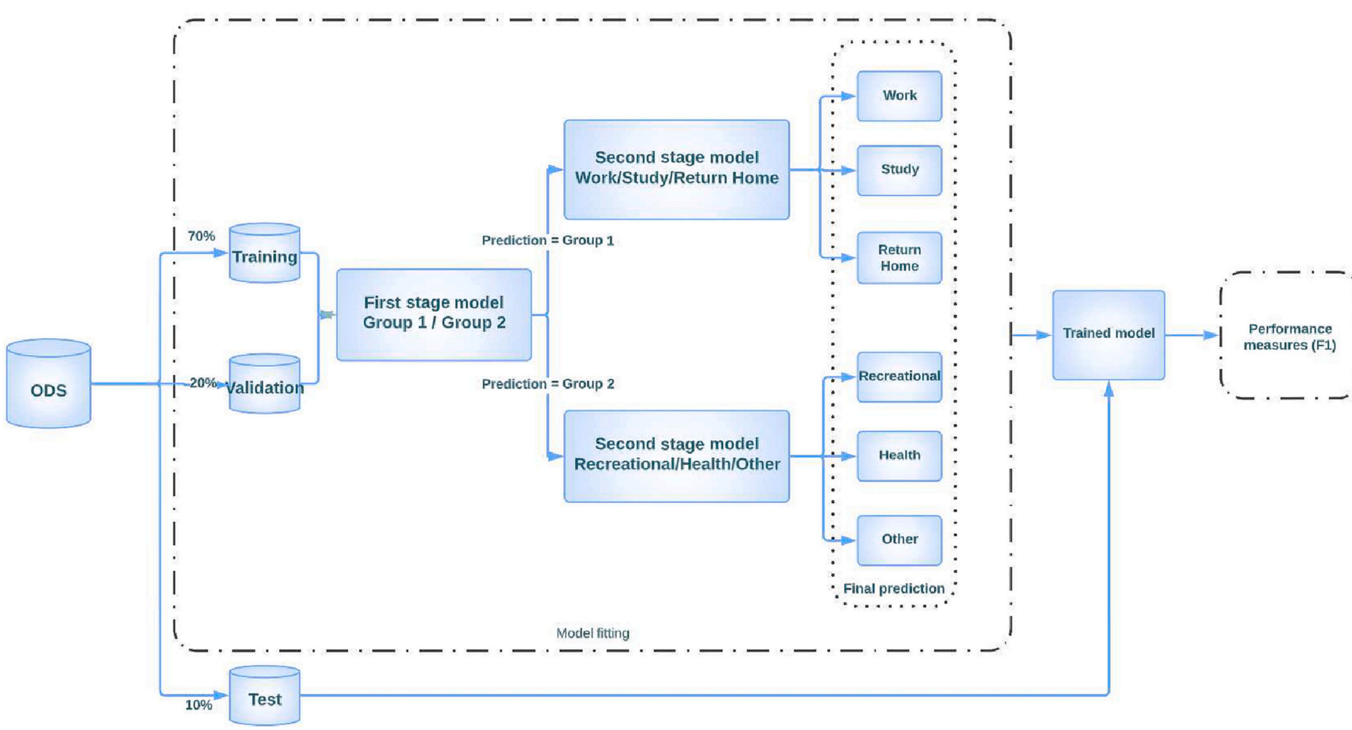


Fig. 7. Description of the concatenated model.

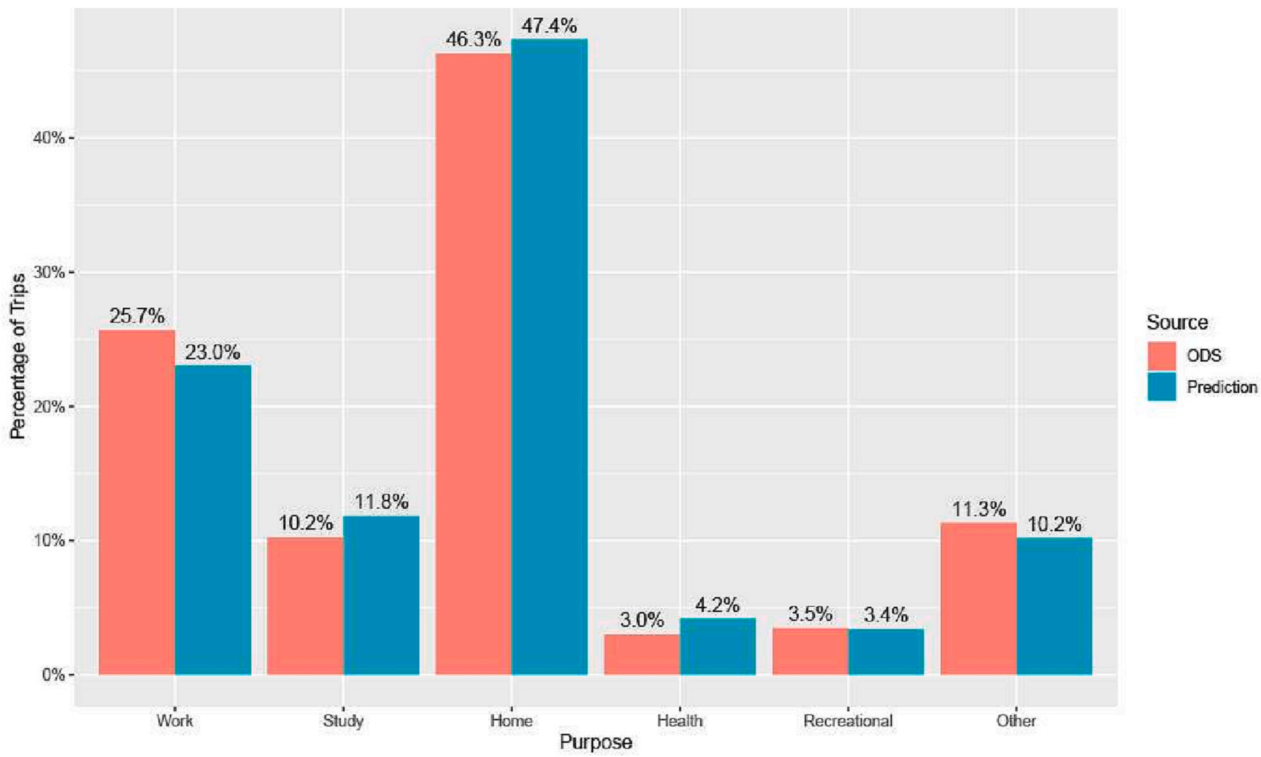


Fig. 8. Comparison of the actual vs. predicted trip distribution in the test base.

Table 13
Proportion of trips according to purpose for each Period.

$$\Delta T_{ijkl} = \frac{T_{ijk}}{\sum_{i' \in P} T_{i'jk}} - \frac{T_{ijl}}{\sum_{i' \in P} T_{i'jl}} \tag{5}$$

Purposes	Period 1		Period 2		Period 3	
	N	Percentage	N	Percentage	N	Percentage
Work	3,875,581	23.81%	304,918	22.86%	1,847,961	23.26%
Study	1,834,722	11.05%	81,506	6.16%	731,059	9.09%
Return Home	6,716,247	41.56%	629,067	47.36%	3,358,104	42.60%
Recreational	620,052	4.14%	20,259	1.62%	275,503	3.64%
Health	673,717	4.87%	46,913	4.27%	375,782	5.46%
Other	2,356,830	14.55%	235,490	17.70%	1,256,596	15.83%
Total	16,077,149	100%	1,318,153	100%	7,845,005	100%

5.2. Application to the case of Santiago, Chile

This study considers three periods of interest in Santiago, Chile. As described above, Period 1 corresponds to a situation before the pandemic, which ran from June 11 to 15, 2018. Period 2 corresponds to a complete quarantine period from July 6 to 10, 2020. Finally, Period 3 corresponds to a period without quarantine in the city, spanning from October 4 to 8, 2021.

During 2020, the government reacted quickly and implemented several measures in response to the COVID-19 crisis, including a state of exception, closure of the borders, policies regarding the use of masks, school and university closures, quarantines, and limited contact with the elderly population. A detailed description of these and other implemented measures can be found in Cámara de Diputados (2021). These measures were effective during the first months after their implementation, particularly in some regions (Barra-Sandoval et al., 2022), but they gradually lost acceptance by the population (Li et al., 2022). In 2021, these measures were subsequently lifted.

Based on the application of the concatenated model described in Subsection 4.4, Table 13 shows the distribution by purpose for each of the periods mentioned in the previous paragraph. First, it is important to note that the purposes of Work and Returning Home are those that represent the highest percentage in all periods, which is consistent with the distribution shown in the ODS. Then, it can be seen that trips with Study purpose show a decrease of 4.89 percentage points for Period 2 compared to Period 1. In nominal terms, trips for this purpose decrease by 95.55%. This is explained by the fact that several educational establishments (both schools and universities) opted for remote classes during the pandemic and the periods of reduced mobility. However, for Period 3, the trips for this purpose resumed, showing an increase of 2.93 percentage points with respect to Period 2. This caused the trips of this purpose (Study) to reach a percentage close to the one they presented before the pandemic. However, note that the total number of trips for this purpose is still much lower than that observed before the pandemic.

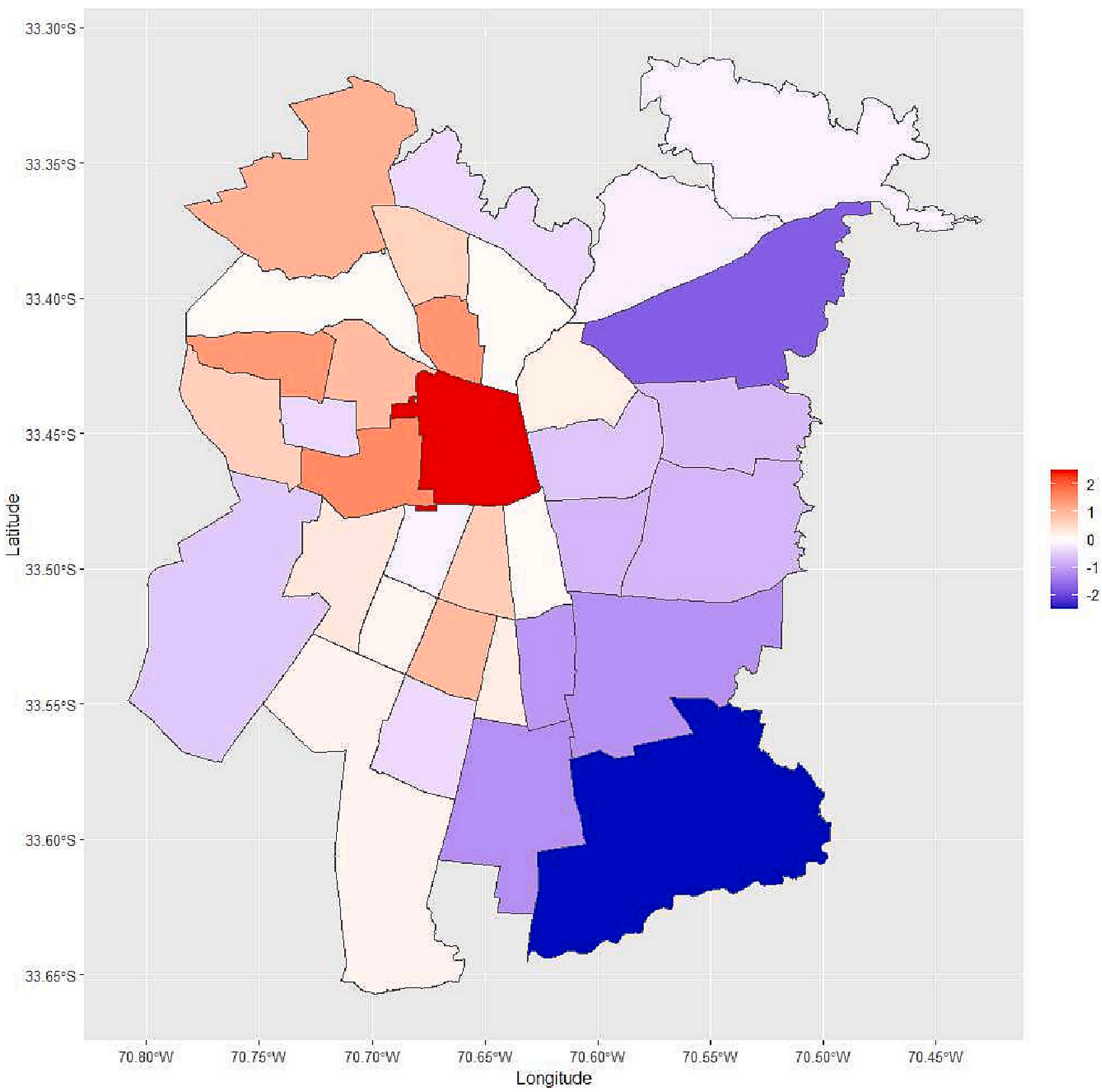


Fig. 9. $\Delta T_{\mathrm{Work},\, j,2,1}$ for each commune j.

Table 14
Average distance of trips with work purpose for each commune - Period 2.

Commune	Distance (Km)
CERRILLOS	6.77
CERRO NAVIA	7.35
CONCHALÍ	6.87
EL BOSQUE	10.28
ESTACIÓN CENTRAL	7.05
HUECHURABA	7.03
INDEPENDENCIA	6.37
LA CISTERNA	9.64
LA FLORIDA	8.73
LA GRANJA	6.71
LA PINTANA	10.06
LA REINA	6.66
LAS CONDES	8.03
LO BARNECHEA	10.34
LO ESPEJO	7.38
LO PRADO	7.37
MACUL	6.69
MAIPÚ	9.75
PEDRO AGUIRRE CERDA	6.71
PEÑALOLÉN	7.18
PROVIDENCIA	7.01
PUDAHUEL	8.08
PUENTE ALTO	11.95
QUILICURA	7.60
QUINTA NORMAL	6.79
RECOLETA	7.04
RENCA	7.11
SAN BERNARDO	14.34
SAN JOAQUÍN	7.26
SAN MIGUEL	9.23
SAN RAMÓN	10.04
SANTIAGO	6.80
VITACURA	11.89
ÑUÑOA	5.88

The latter is explained by the fact that, during relaxed measures, many educational centers continued to offer activities online.

Fig. 9 illustrates for each commune of the Metropolitan Region the value of $\Delta T_{\text{Work},\, j,2,1}$, that is, the difference in percentage points in the participation of the Work purpose in Period 2 with respect to Period 1. Looking at this figure, it is possible to see that there is spatial segregation, where the communes located in the eastern sector significantly decreased their participation in terms of trips to work, while the communes of the western sector, overall, increased their participation. The above is explained taking into consideration the territorial organization of the city of Santiago, which generates a dynamic of delimited social spaces, where the northeast sector corresponds to the high-income sector, while the rest of the population is settled in the rest of communes (Canales Cerón, 2021). Due to the above, of the total number of people who travel by public transport for work during the complete quarantine period, the percentage of people who come from higher-income communities is lower. This phenomenon has been previously reported in Chile, showing that citizens with higher incomes have greater opportunities for teleworking.[3]

To strengthen this idea, the R-Pearson correlation is analyzed to identify whether the variations in the participation of trips by commune are correlated with the income of each commune (presented in Appendix B and Appendix C). It was found that the correlation between the values plotted in Fig. 9 and the income by commune is -0.14, which indicates an inversely proportional relationship; that is, the higher the income of the commune is, the greater the reduction in work trips. However, because this correlation is relatively low, other factors in addition to income explain the decrease in work trips. One of these other factors could be the trip distance. Table 14 shows the average trip distance for work purposes in each commune during period 2. From this table, it can be seen that there is high variability in the trip distances, and therefore, it is possible that the inhabitants of communes with longer trip distances may be more inclined to telework. In contrast, those passengers with shorter trip distances may be inclined to continue in-person activities. This could be the case of the commune of Santiago Centro, which according to Table 14 shows one of the lowest average trip distances and which in turn, presents one of the largest increases according to Fig. 9. Another factor associated with the differences in participation between communes is the change of public transport mode to private vehicles or bicycles, as has been widely described in other parts of the world (e.g., Das et al., 2021; Šinko et al., 2021; Zhang et al., 2021a). One of the hypotheses is that in more peripheral communes or with longer trip distances, there are greater incentives to seek a different mode of transportation. This could be what explains why communes in the southern sector of Santiago have shown stronger decreases in public transportation, such as Puente Alto.

Fig. 10 shows for each commune of the Metropolitan Region the changes in the participation of the Work purpose of Period 3 with respect to Period 1. From the above, in general, the changes are of smaller magnitude compared to Period 2 (Fig. 9). This is because, in this period, the restrictions were being relaxed so that people began to resume the activities they carried out in the periods prior to the pandemic. In addition, Fig. 10 shows that there is again spatial segregation, where the communes of the eastern sector continue to participate less in work trips compared to period 1. As discussed above, this effect is explained by the differences in income and, therefore, in the possibility of continuing teleworking. Calculating the R-Pearson correlation between income and the changes presented in Fig. 10, a value of -0.24 is obtained, which supports this argument. However, again, additional factors explain the changes in participation. One possible explanation has to do with the profound changes observed in the labor market during the pandemic period: due to the shocks experienced in different industries, unemployment levels increased drastically[4].[5] However, unemployment does not affect the population uniformly, and therefore, it is possible that the communes with lower income (located in the southern and western sectors) have been more impacted, decreasing their participation in work trips. A second possible explanation is related to the modal shift experienced as a result of the pandemic, similar to that mentioned above for Period 2. In fact, there is ample evidence of an increase in the use of private transport as a precautionary measure against infection (e.g., Das et al., 2021; Abdullah et al., 2020, 2021) This effect again does not influence all communes in the same way, and it is possible that people who make longer trips (and therefore, with greater exposure to potential contagion) have tended to change their mode of transportation.

Finally, there is one factor that has a slight impact in the short term but could have a significant impact in the medium and long term: the distribution of work trips is the result of an urban balance, where people choose the location of their home and place of work. Because the pandemic brought changes in transportation costs to work (reducing them to zero in the case of people who telework), this balance is expected to change in the future. In effect, people who telework will have greater incentives to move to the city's periphery, where land costs are

[3] https://www.uchile.cl/noticias/162383/solo-1-de-cada-4-trabajadores-de-menores-ingresos-realizo-teletrabajo-. Accessed 23 August 2022, in Spanish.

[4] https://www.latercera.com/pulso/noticia/desocupacion-llega-a-su-mayor-nivel-en-16-anos-y-covid-destruye-15-millones-de-empleos/EFI52PEMMZG73KHSBMCS5FXBJI/. Accessed 23 August 2022, in Spanish.

[5] https://www.uchile.cl/noticias/166307/tasa-de-desempleo-en-el-gran-santiago-se-estanca. Accessed 23 August 2022, in Spanish.

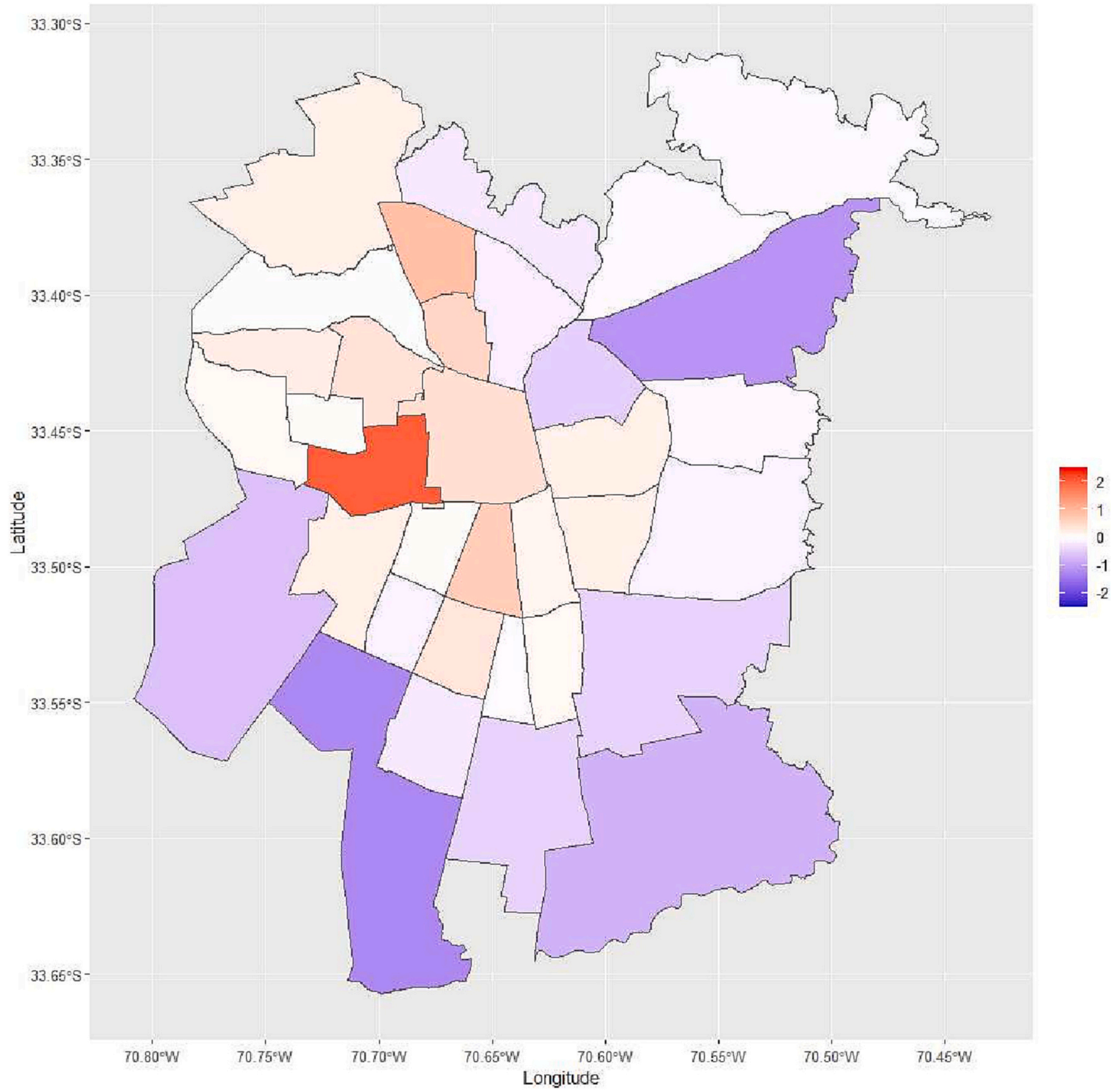


Fig. 10. $\Delta T_{\text{Work},\, j,3,1}$ for each commune j.

Table 15
Information by purpose Period 1.

	Work	Study	Return home	Recreational	Health	Other
Average Activity Time (Hours)	9.61	6.26	19.78	5.18	1.99	2.04
Average trip time (Minutes)	35.48	27.63	33.08	26.48	30.78	21.60
Average distance OD (Kilometers)	9.88	7.15	8.19	6.14	7.95	5.86

Table 16
Information by purpose Period 3.

	Work	Study	Return home	Recreational	Health	Other
Average Activity Time (Hours)	9.59	6.44	16.72	4.45	1.83	1.82
Average trip time (Minutes)	34.41	26.66	30.48	25.82	29.17	20.46
Average distance OD (Kilometers)	9.78	7.00	7.89	6.09	7.81	5.71

lower. Therefore, peripheral communes are expected to reduce their participation in work trips, a product of this new equilibrium.

For a more in-depth comparison between period 1 (prepandemic) and period 3 (relaxed measures), Tables 15 and 16 show the average time of activity, the average trip time and the average distance of the trips for each of these periods. First, it is important to note that the activity times obtained are consistent with the characteristics of the purposes considered. Thus, for example, the average activity times for work purposes are close to nine hours, which corresponds to the full working day according to Chilean regulations. Additionally, one of the main observations from what is presented in Tables 15 and 16 is that the characteristics of the trips during Period 3, for all purposes, strongly aligned with what was observed in Period 1. This reinforces the fact that the functioning of the city quickly recovered the conditions prior to the pandemic once the restriction measures were relaxed.

Overall, according to our findings, there was a significant shift in the

Commune	Warehouse (%)	Commercial (%)	Sport (%)	Education (%)	Housing (%)	Industry (%)	Office (%)	Health (%)
ALHUE	0.41	2.98	0.00	1.27	95.13	0.00	0.22	0.00
BUIN	2.73	4.18	1.19	3.33	66.6	21.3	0.29	0.38
CALERA DE TANGO	0.46	3.79	0.51	3.15	85.13	6.54	0.22	0.19
CERRILLOS	12.12	12.61	0.39	3.52	46.58	24.08	0.53	0.16
CERRO NAVIA	1.05	6.08	0.41	4.31	85.08	2.11	0.68	0.29
COLINA	2.29	3.31	0.49	3.45	83.96	5.74	0.53	0.22
CONCHALI	3.35	8.86	0.46	3.09	71.58	11.07	1.17	0.42
CURACAVI	0.96	3.59	0.37	1.30	88.88	4.31	0.42	0.16
EL BOSQUE	0.75	7.21	0.82	7.40	78.47	4.55	0.51	0.30
EL MONTE	0.55	2.73	0.30	4.47	82.88	8.75	0.27	0.05
ESTACION CENTRAL	3.68	8.67	0.95	6.40	64.31	11.43	1.68	2.88
HUECHURABA	4.85	8.66	2.27	4.75	61.57	6.12	11.36	0.43
INDEPENDENCIA	2.51	18.34	1.54	5.81	54.62	11.74	2.05	3.38
ISLA DE MAIPO	0.45	10.89	0.39	1.39	68.83	17.4	0.20	0.45
LA CISTERNA	1.62	8.26	3.59	4.66	75.8	4.39	1.32	0.35
LA FLORIDA	0.25	6.52	1.65	5.60	81.47	1.32	0.67	2.52
LA GRANJA	2.33	6.79	3.88	8.35	68.47	7.61	1.30	1.26
LA PINTANA	6.19	7.23	1.50	11.81	62.47	9.51	0.79	0.50
LA REINA	0.42	5.22	4.94	3.72	80.66	2.36	0.59	2.09
LAMPA	7.11	3.17	0.76	3.38	63.73	21.30	0.33	0.23
LAS CONDES	0.45	4.51	3.10	6.21	79.34	0.16	4.60	1.63
LO BARNECHEA	0.39	4.85	5.45	3.51	82.13	2.53	0.61	0.52
LO ESPEJO	1.98	13.73	1.99	4.00	70.05	5.85	1.66	0.75
LO PRADO	0.67	8.80	1.19	3.67	82.69	1.19	0.72	1.06
MACUL	3.44	5.69	2.43	9.24	61.95	13.80	1.38	2.06
MAIPU	5.84	6.94	0.42	4.02	64.67	17.25	0.48	0.39
MARIA PINTO	1.40	1.08	5.32	3.22	88.63	0.14	0.13	0.07
MELIPILLA	0.34	4.95	0.89	4.32	86.39	2.49	0.33	0.29
NUNOA	1.39	7.91	5.06	8.85	69.07	3.13	3.44	1.16
PAC	4.35	8.17	5.36	4.51	72.63	2.35	1.80	0.82
PADRE HURTADO	0.35	1.60	0.34	2.96	85.56	8.95	0.20	0.03
PAINE	3.36	4.80	0.43	3.36	67.56	19.37	0.85	0.28
PENAFLOR	1.85	4.59	1.16	1.76	82.30	8.04	0.10	0.21
PENALOLEN	0.20	5.98	2.28	11.21	76.58	2.62	0.29	0.82
PIRQUE	0.47	2.02	1.50	1.32	92.21	2.29	0.10	0.09
PROVIDENCIA	0.81	7.48	0.15	5.46	67.19	0.51	15.04	3.36
PUDAHUEL	46.08	3.28	1.18	1.24	32.27	13.37	2.25	0.34
PUENTE ALTO	0.40	4.74	3.01	5.24	75.89	8.79	0.71	1.21
QUILICURA	22.24	4.10	0.19	1.45	30.72	40.55	0.57	0.19
QUINTA NORMAL	4.21	8.96	0.86	3.90	66.47	13.34	1.30	0.97
RECOLETA	2.08	11.62	2.71	4.86	67.1	7.11	2.79	1.71
RENCA	25.08	3.68	1.18	2.94	45.64	19.91	1.21	0.35
SAN BERNARDO	15.75	7.75	1.52	3.66	45.96	23.79	1.17	0.39
SAN JOAQUIN	6.68	8.02	2.76	4.19	58.65	17.5	1.71	0.50
SAN JOSE DE MAIPO	0.12	10.22	0.45	1.57	70.76	15.60	0.69	0.59
SAN MIGUEL	2.55	6.27	0.74	3.41	75.87	7.12	1.88	2.17
SAN PEDRO	0.00	4.07	0.70	4.73	86.34	0.00	3.36	0.80
SAN RAMON	0.98	7.00	2.34	6.02	74.09	3.01	1.83	4.73
SANTIAGO	2.73	21.31	0.96	8.50	50.81	7.54	7.04	1.11
TALAGANTE	0.74	4.14	0.44	4.71	78.33	10.76	0.26	0.61
TILTIL	3.45	4.57	0.03	2.70	64.77	23.66	0.51	0.32
VITACURA	0.41	6.43	5.54	4.83	75.54	1.03	4.90	1.32

purpose of trips during the peak period of restrictions in 2020. As expected, while the proportion of trips for work increased, recreational trips decreased, and trips for health purposes remained steady. In other words, people continued using public transportation mainly for mandatory trips. However, the decrease in the proportion of trips for work purposes was more pronounced in higher-income communes than in lower-income ones, presumably due to differences in the possibility of working from home. As in-person activities gradually resumed in 2021, the distribution of trip purposes and trip conditions quickly returned to the pre-pandemic values, albeit with fewer trips. This has profound implications: our results show that the decrease in public transport trips in Period 3, compared to Period 1, is relatively proportional for all trip purposes. Therefore, this suggests that people stopped using the public transport system for all purposes and not only for those perceived as higher risk (e.g., peak hours or long trip distances). This suggests that the decline in public transportation usage could be permanent without appropriate measures.

6. Conclusions

The COVID-19 pandemic drastically changed the mobility landscape during 2020, significantly decreasing travel in all modes (Li et al., 2021; Astroza et al., 2020) and causing passengers to flee from public transport to other modes (Das et al., 2021). This generates challenges in the planning of public transport systems, which must implement measures to attract passengers again. However, there is a gap in the understanding of how the pandemic affected the distribution of trip purposes because the literature has focused mostly on the aggregate study of changes in mobility (e.g. Abdullah et al., 2020, 2021; Gramsch et al., 2022; Jenelius and Cebecauer, 2020). The above hinders the design of targeted measures aimed at increasing the demand for public transport. In fact, trip with different purposes are valued differently in terms of public transport attributes, such as trip time (Wardman, 2004) or overcrowding (Whelan and Crockett, 2009), so their response to transport policies is not uniform.

To address this gap, this paper builds a supervised learning model

(XGBoost) to estimate the trip purpose in public transport in Santiago, Chile, using the last origin-destination survey conducted in the region as the main source of data for calibration. This information is complemented with land use and POI data. For the application of the proposed model, we also used data from the smart payment cards and GPS of the buses. With this, the purpose is estimated for all trips made in the public transport system for three weeks in different periods: before the pandemic, during complete quarantine, and during relaxed measures.

First, our results show that the number of trips for all purposes decreased drastically in response to mobility restriction measures. However, these decreases were not uniform for different purposes. In effect, work-related trips maintained their percentage share of total trips, while study trips or trips for recreational purposes strongly decreased. This is explained by the fact that in this period, the student campuses and recreational facilities were mostly closed, in addition to the fact that the confinement measures allowed only trips that could not be postponed. In relation to work trips, our results also show that there is spatial segregation in the city of Santiago. In particular, the communes of the western sector had to continue making work trips, while the communes of the eastern sector decreased their participation, which is explained mainly by teleworking. This fact is related to income, which is in line with what other authors have mentioned (Wilbur et al., 2020; Astroza et al., 2020). Then, when analyzing the period of relaxed restriction measures, our results show that the number of trips for all purposes increased strongly with respect to the period of confinement. Despite this, the total number of trips in this period represents less than 50% of the trips in the pre-pandemic period. For this period, the distribution of trip purposes quickly resumed patterns close to those observed before the pandemic, showing that society quickly adapts to changes in restriction measures.

The main limitation of this research is the age of one of the datasets we use: the 2012 origin-destination survey, which provides the last official information available for trip purposes. Even though it would be desirable to have more recent information, we think using this data does not invalidate our results and conclusions. As discussed in Subsection 4.3, the main underlying assumption of our work is that population drift is the only type of dataset shift present in the studied phenomenon. This type of shift can be controlled by employing more complex classification methods, such as XGBoost. Consequently, as long as the conditional probability $p(y|x)$ does not dramatically change over time, our results will hold.

From what was found in this article, several recommendations for public transport can be drawn. First, although the distribution of trip patterns resumed patterns close to those observed prior to the pandemic, the number of trips decreased sharply. Particularly for the Work purpose, the greatest decreases are observed in peripheral communes and in higher-income communes. For these communes, one of the reasons that explain this is that there was a modal shift toward private transportation (Anwari et al., 2021; Das et al., 2021). Moreover, our results show that the decrease in public transport trips in the relaxed-measures period, compared to the pre-pandemic period, is relatively proportional for all trip purposes. Therefore, this suggests that people stopped using the public transport system for all types of trips, and not only those associated with a greater risk of contagion (peak hours, long trip distances, etc.). In this sense, it is necessary to promote measures to regain confidence and increase public transport's service standards. Examples include an increase in the frequency of routes (Mouwen, 2015), dedicated infrastructure (Basso and Silva, 2014), better standards in buses (Ngoc et al., 2017) or improvements in the accessibility of public transport (Basso et al., 2020). However, to attract users to public transport again, it is also necessary to directly discourage using other modes. Examples include implementing congestion charges (Eliasson and Mattsson, 2006) or public transportation subsidies (Van Goeverden et al., 2006).

Finally, based on what has been done in this article, several lines of research follow. First, it is possible to redo this study with data from more recent periods, already without restriction measures, to visualize how the effect of the pandemic is dissipated or maintained in the medium and long term. This might be of interest, especially considering that public transport usage has not recovered as fast as expected. In fact, during August 2022, the average daily transactions were still 27% less than in the pre-pandemic period. Additionally, analyzing the next ODS –planned for the following years in the case of Santiago, Chile– would allow us to validate our assumption regarding the type of dataset shifts present. In case other types of shifts are present, the use of methods proposed in recent streams of literature might help to develop models that combine old and new ODS (see, e.g., Zhuang et al., 2020, for a survey on transfer learning). Concerning the trip purposes, one would expect the behavior patterns to return to the pre-pandemic period during the following years. On the other hand, the use of new data sources could serve as input for this model. For example, the use of social networks (Steiger et al., 2014) or mobile phone data (Huang et al., 2019) has shown great potential for the study of mobility. Therefore, these types of sources could enrich the model, allowing a more microscopic analysis of the problem. Finally, the problem studied could be addressed through an unsupervised learning methodology, generating groups of trips with similar characteristics and analyzing their variation in the face of different measures of mobility restriction.

CRediT authorship contribution statement

Raúl Pezoa: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Franco Basso:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision. **Paulina Quilodrán:** Software, Validation, Investigation, Writing – original draft. **Mauricio Varas:** Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Data availability

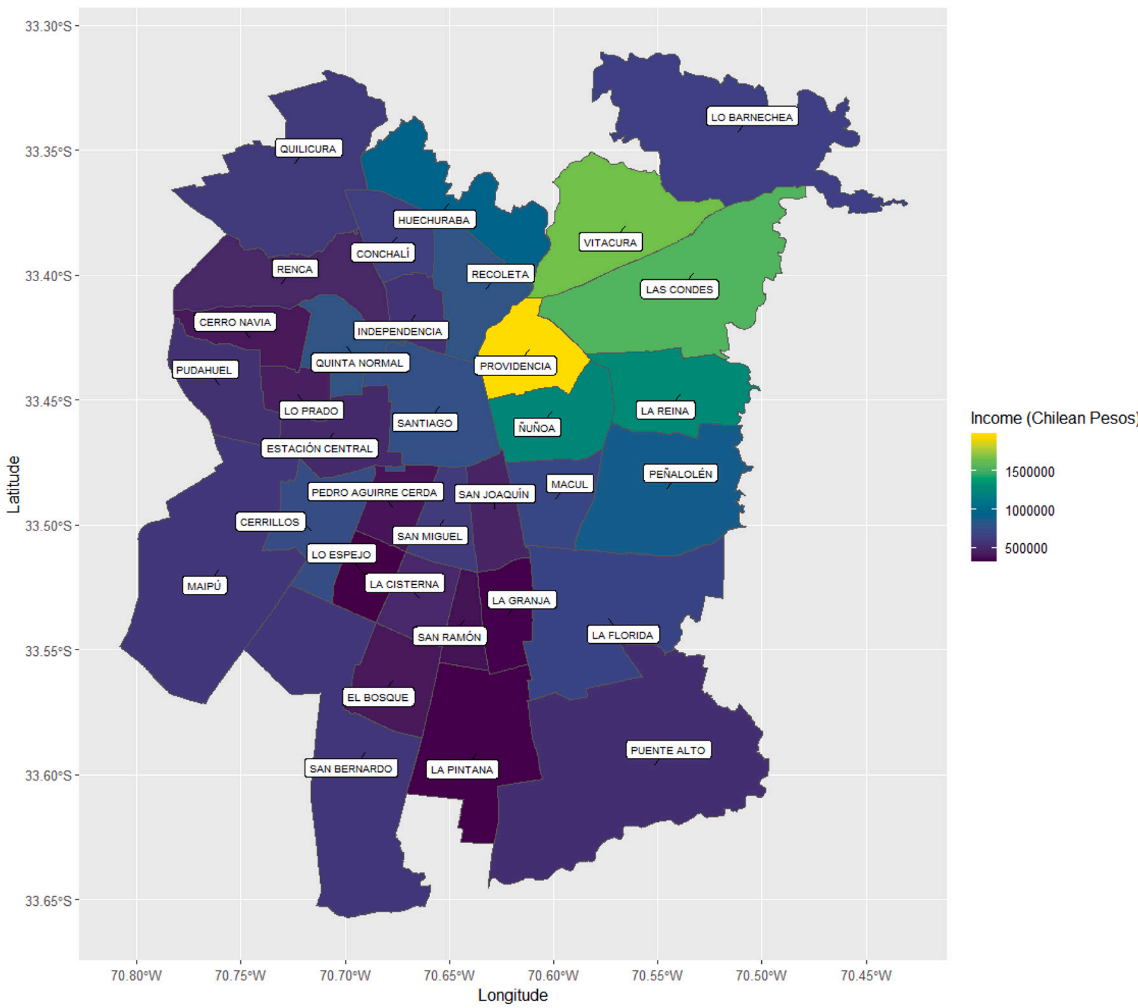
Data will be made available on request.

Acknowledgments

Raúl Pezoa thanks doctoral scholarship to ANID-PFCHA/Doctorado Nacional/2018-21181528. Franco Basso gratefully acknowledges the financial support from both the Complex Engineering Systems Institute, ISCI (grant ANID PIA AFB180003) and a grant from the National Agency for Research and Development (ANID) of the Chilean Government under the Ministry of Science and Technology, Knowledge and Innovation (FONDECYT Project 11200167). Mauricio Varas thanks a grant from ANID by FONDECYT Project 11190892.

Appendix A. Distribution of land use by commune

Appendix B. Average income by commune for the year 2020



Appendix C. Average income by commune for the year 2020 (Chilean Pesos)

Commune	Income
CERRO NAVIA	419,410
MAIPÚ	596,667
LA REINA	1,287,177
LO PRADO	458,188
PUDAHUEL	549,414
PUENTE ALTO	538,506
LAS CONDES	1,556,078
SAN BERNARDO	586,004
SANTIAGO	776,917
RENCA	486,575
MACUL	704,740
LA PINTANA	328,847
INDEPENDENCIA	558,087
PROVIDENCIA	1,989,530
CERRILLOS	744,007
ÑUÑOA	1,263,088
LA GRANJA	335,312
LA FLORIDA	687,740

(*continued on next page*)

(*continued*)

Commune	Income
QUILICURA	605,365
PEÑALOLÉN	870,402
QUINTA NORMAL	806,526
SAN JOAQUÍN	483,413
ESTACIÓN CENTRAL	505,281
VITACURA	1,664,926
RECOLETA	820,334
EL BOSQUE	418,780
LO BARNECHEA	659,923
HUECHURABA	931,341
LA CISTERNA	493,864
CONCHALÍ	640,567
PEDRO AGUIRRE CERDA	404,481
SAN MIGUEL	617,137
SAN RAMÓN	388,183
LO ESPEJO	320,136

References

Abdullah, M., Dias, C., Muley, D., Shahin, M., 2020. Exploring the impacts of COVID-19 on travel behavior and mode preferences. Transp. Res. Interdiscipl. Perspect. 8, 100255.

Abdullah, M., Ali, N., Hussain, S.A., Aslam, A.B., Javid, M.A., 2021. Measuring changes in travel behavior pattern due to COVID-19 in a developing country: a case study of Pakistan. Transp. Policy 108, 21–33.

Aghabayk, K., Esmailpour, J., Shiwakoti, N., 2021. Effects of COVID-19 on rail passengers' crowding perceptions. Transp. Res. A Policy Pract. 154, 186–202.

Alaiz-Rodrguez, R., Japkowicz, N., 2008. Assessing the impact of changing environments on classifier performance. In: Advances in Artificial Intelligence: 21st Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2008 Windsor, Canada, May 28–30, 2008 Proceedings, 21. Springer, pp. 13–24.

Almlöf, E., Rubensson, I., Cebecauer, M., Jenelius, E., 2021. Who continued travelling by public transport during COVID-19? Socioeconomic factors explaining travel behaviour in Stockholm 2020 based on smart card data. Eur. Transp. Res. Rev. 13 (1), 1–13.

Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., Hickman, M., 2018. Public transport trip purpose inference using smart card fare data. Transp. Res. Part C: Emerg. Technol. 87, 123–137.

Anwari, N., Ahmed, M.T., Islam, M.R., Hadiuzzaman, M., Amin, S., 2021. Exploring the travel behavior changes caused by the COVID-19 crisis: a case study for a developing country. Transp. Res. Interdiscipl. Perspect. 9, 100334.

Astroza, S., Tirachini, A., Hurtubia, R., Carrasco, J.A., Guevara, A., Munizaga, M., Figueroa, M., Torres, V., 2020. Mobility changes, teleworking, and remote communication during the COVID-19 pandemic in Chile. Findings 13489.

Awad-Núñez, S., Julio, R., Gomez, J., Moya-Gómez, B., González, J.S., 2021. Post-COVID-19 travel behaviour patterns: impact on the willingness to pay of users of public transport and shared mobility services in Spain. Eur. Transp. Res. Rev. 13 (1), 1–18.

Barra-Sandoval, C., Ferreira, G., Lagos, B., Bacigalupo, C.M., 2022. Assessing the effectiveness of quarantine measures during the COVID-19 pandemic in Chile using bayesian structural time series models. Infect. Disease Modell. 7 (4), 625–636.

Basnak, P., Giesen, R., Muñoz, J.C., 2022. Estimation of crowding factors for public transport during the COVID-19 pandemic in santiago, Chile. Transp. Res. A Policy Pract. 159, 140–156.

Basso, L.J., Silva, H.E., 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. Am. Econ. J. Econ. Pol. 6 (4), 1–33.

Basso, F., Frez, J., Martnez, L., Pezoa, R., Varas, M., 2020. Accessibility to opportunities based on public transport gps-monitored data: the case of santiago, Chile. Travel Behav. Soc. 21, 140–153.

Basso, F., Cifuentes, Á., Cuevas-Pavincich, F., Pezoa, R., Varas, M., 2022a. Assessing influential factors for lane change behavior using full real-world vehicle-by-vehicle data. Transp. Lett. 14 (10), 1126–1137.

Basso, F., Pezoa, R., Tapia, N., Varas, M., 2022b. Estimation of the origin-destination matrix for trucks that use highways: a case study in Chile. Sustainability 14 (5), 2645.

Bickel, S., Brückner, M., Scheffer, T., 2007. Discriminative learning for differing training and test distributions. In: Proceedings of the 24th International Conference on Machine Learning, pp. 81–88.

Boehmke, B., Greenwell, B.M., 2019. Hands-on Machine Learning with R. CRC Press.

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day gps-based travel surveys: a large-scale application in the Netherlands. Transp. Res. Part C: Emerg. Technol. 17 (3), 285–297.

Brownlee, J., 2016. XGBoost with Python: Gradient Boosted Trees with XGBoost and Scikit-Learn. Machine Learning Mastery.

Cámara de Diputados, 2021. Anexo 3. medidas adoptadas para enfrentar la pandemia de COVID-19. https://www.camara.cl/estado_excepcion/Anexo%203/Anexo%203.%20Medidas%20adoptadas%20para%20enfrentar%20la%20pandemia%20de%20COVID-19.pdf. Accessed: 28 December 2022.

Canales Cerón, A.I., 2021. La desigualdad social frente al COVID-19 en el área metropolitana de santiago (chile). Notas de poblacion.

Carmona, P., Dwekat, A., Mardawi, Z., 2022. No more black boxes! Explaining the predictions of a machine learning xgboost classifier algorithm in business failure. Res. Int. Bus. Financ. 61, 101649.

Celi Ortega, S.F., 2018. Análisis del comportamiento del transporte público a nivel mundial.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Cho, S.-H., Park, H.-C., 2021. Exploring the behaviour change of crowding impedance on public transit due to COVID-19 pandemic: before and after comparison. Transp. Lett. 13 (5–6), 367–374.

Contreras, R., Navarrete, C., Arias, G., 2016. Encuesta origen-destino santiago 2012 evolución de escenarios de accesibilidad. Estudios de Transporte 20 (1).

Das, S., Boruah, A., Banerjee, A., Raoniar, R., Nama, S., Maurya, A.K., 2021. Impact of COVID-19: a radical modal shift from public to private transport mode. Transp. Policy 109, 1–11.

Deschaintres, E., Morency, C., Trépanier, M., 2022. Assessing the Impacts of the COVID-19 Pandemic on Subway Ridership and on the Interactions with Other Transportation Modes.

Devillaine, F., Munizaga, M., Trépanier, M., 2012. Detection of activities of public transport users by analyzing smart card data. Transp. Res. Rec. 2276 (1), 48–55.

Eliasson, J., Mattsson, L.-G., 2006. Equity effects of congestion pricing: quantitative methodology and a case study for Stockholm. Transp. Res. A Policy Pract. 40 (7), 602–620.

Faroqi, H., Mesbah, M., 2021. Inferring trip purpose by clustering sequences of smart card records. Transp. Res. Part C: Emerg. Technol. 127, 103131.

Faroqi, H., Mesbah, M., Kim, J., 2018. Applications of transit smart cards beyond a fare collection tool: a literature review. Adv. Transp. Stud. 45.

Gareth, J., Daniela, W., Trevor, H., Robert, T., 2013. An Introduction to Statistical Learning: With Applications in R. Spinger.

Gkiotsalitis, K., Cats, O., 2021. Public transport planning adaption under the COVID-19 pandemic crisis: literature review of research needs and directions. Transp. Rev. 41 (3), 374–392.

Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. Transp. Res. Part C: Emerg. Technol. 64, 1–16.

Gramsch, B., Guevara, C.A., Munizaga, M., Schwartz, D., Tirachini, A., 2022. The Effect of Dynamic Lockdowns on Public Transport Demand in Times of COVID-19: Evidence from Smartcard Data. Transport Policy.

Gschwender, A., Munizaga, M., Simonetti, C., 2016. Using smart card and gps data for policy and planning: the case of transantiago. Res. Transp. Econ. 59, 242–249.

Gutiérrez, A., Miravet, D., Domènech, A., 2021. COVID-19 and urban public transport services: emerging challenges and research agenda. Cities Health 5 (sup1), S177–S180.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer.

Huang, H., Cheng, Y., Weibel, R., 2019. Transport mode detection based on mobile phone network data: a systematic review. Transp. Res. Part C: Emerg. Technol. 101, 297–312.

Jenelius, E., Cebecauer, M., 2020. Impacts of COVID-19 on public transport ridership in Sweden: analysis of ticket validations, sales and passenger counts. Transp. Res. Interdiscipl. Perspect. 8, 100242.

Kickhofer, B., Hosse, D., Turnera, K., Tirachinic, A., 2016. Creating an Open Matsim Scenario from Open Data: The Case of santiago de Chile. TU Berlin, Transport System Planning and Transport Telematics, Berlin, Germany.

Kim, E.-J., Kim, Y., Kim, D.-K., 2021. Interpretable machine-learning models for estimating trip purpose in smart card data. In: Proceedings of the Institution of Civil Engineers-Municipal Engineer, Vol. 174. Thomas Telford Ltd, pp. 108–117.
Kopsidas, A., Milioti, C., Kepaptsoglou, K., Vlachogianni, E.I., 2021. How did the COVID-19 pandemic impact traveler behavior toward public transport? The case of Athens, Greece. Transp. Lett. 13 (5–6), 344–352.
Li, Y., Yang, L., Yang, B., Wang, N., Wu, T., 2019. Application of interpretable machine learning models for the intelligent decision. Neurocomputing 333, 273–283.
Li, A., Zhao, P., Haitao, H., Mansourian, A., Axhausen, K.W., 2021. How did micro-mobility change in response to COVID-19 pandemic? A case study based on spatial-temporal-semantic analytics. Comput. Environ. Urban. Syst. 90, 101703.
Li, Y., Undurraga, E.A., Zubizarreta, J.R., 2022. Effectiveness of localized lockdowns in the COVID-19 pandemic. Am. J. Epidemiol. 191 (5), 812–824.
Medina, S.A.O., 2018. Inferring weekly primary activity patterns using public transport smart card data and a household travel survey. Travel Behav. Soc. 12, 93–101.
Mitchell, R., Frank, E., 2017. Accelerating the xgboost algorithm using gpu computing. PeerJ. Comp. Sci. 3, e127.
Moreno-Torres, J.G., Raeder, T., Alaiz-Rodrguez, R., Chawla, N.V., Herrera, F., 2012. A unifying view on dataset shift in classification. Pattern Recogn. 45 (1), 521–530.
Mouwen, A., 2015. Drivers of customer satisfaction with public transport services. Transp. Res. A Policy Pract. 78, 1–20.
Munizaga, M.A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, Chile. Transp. Res. Part C: Emerg. Technol. 24, 9–18.
Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behavior estimated from smartcard data. Transp. Res. Part C: Emerg. Technol. 44, 70–79.
Muñoz, V., Thomas, A., Navarrete, C., Contreras, R., 2015. Encuesta origen-destino de santiago 2012: Resultados y validaciones. Estudios de Transporte 19 (1).
Mützel, C.M., Scheiner, J., 2021. Investigating spatio-temporal mobility patterns and changes in metro usage under the impact of COVID-19 using Taipei metro smart card data. Publ. Transp. 1–24.
Neira, R., Undurraga, F., 2003. Localización industrial y planificación urbana: El caso de las formaciones industriales del norponiente de santiago. Universidad de Chile, Facultad de Arquitectura y Urbanismo, Seminario para optar al Ttulo de Arquitecto, p. 138.
Ngoc, A., Hung, K., Tuan, V., 2017. Towards the development of quality standards for public transport service in developing countries: analysis of public transport users' behavior. Transp. Res. Proc. 25, 4560–4579.
Ogunleye, A., Wang, Q.-G., 2019. Xgboost model for chronic kidney disease diagnosis. IEEE/ACM Trans. Comp. Biol. Bioinform. 17 (6), 2131–2140.
Parker, M.E., Li, M., Bouzaghrane, M.A., Obeid, H., Hayes, D., Frick, K.T., Rodrguez, D. A., Sengupta, R., Walker, J., Chatman, D.G., 2021. Public transit use in the United States in the era of COVID-19: transit riders' travel behavior in the COVID-19 impact and recovery period. Transp. Policy 111, 53–62.
Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. Accid. Anal. Prev. 136, 105405.
Perchoux, C., Brondeel, R., Wasfi, R., Klein, O., Caruso, G., Vallée, J., Klein, S., Thierry, B., Dijst, M., Chaix, B., et al., 2019. Walking, trip purpose, and exposure to multiple environments: a case study of older adults in Luxembourg. J. Transp. Health 13, 170–184.
Przybylowski, A., Stelmak, S., Suchanek, M., 2021. Mobility behaviour in view of the impact of the COVID-19 pandemic—public transport users in gdansk case study. Sustainability 13 (1), 364.
Rahman, M., Cao, Y., Sun, X., Li, B., Hao, Y., 2021. Deep pre-trained networks as a feature extractor with xgboost to detect tuberculosis from chest x-ray. Comput. Electr. Eng. 93, 107252.
Sagi, O., Rokach, L., 2021. Approximating xgboost with an interpretable decision tree. Inf. Sci. 572, 522–542.
Sari Aslam, N., Ibrahim, M.R., Cheng, T., Chen, H., Zhang, Y., 2021. Activitynet: neural networks to predict public transport trip purposes from individual smart card data and pois. Geo-spatial Inform. Sci. 24 (4), 711–721.
Schwenker, F., Trentin, E., 2014. Pattern classification and clustering: a review of partially supervised learning approaches. Pattern Recogn. Lett. 37, 4–14.
SECTRA, 2014. Actualización y recolección de información del sistema de transporte urbano, etapa, p. ix.
Šinko, S., Prah, K., Kramberger, T., 2021. Spatial modelling of modal shift due to COVID-19. Sustainability 13 (13), 7116.
Steiger, E., Ellersiek, T., Zipf, A., 2014. Explorative public transport flow analysis from uncertain social media data. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, pp. 1–7.
Stopher, P.R., Bullock, P., Horst, F., 2002. Exploring the use of passive gps devices to measure travel. In: Applications of Advanced Technologies in Transportation, 2002, pp. 959–967.
Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciu, M., 2017. Machine learning–xgboost analysis of language networks to classify patients with epilepsy. Brain Inform. 4 (3), 159–169.
Van Goeverden, C., Rietveld, P., Koelemeijer, J., Peeters, P., 2006. Subsidies in Public Transport.
Vickerman, R., 2021. Will COVID-19 put the public back in public transport? A UK perspective. Transp. Policy 103, 95–102.
Wang, X., Feng, H., Fan, Y., 2015. Fault detection and classification for complex processes using semi-supervised learning algorithm. Chemom. Intell. Lab. Syst. 149, 24–32.
Wang, Y., de Almeida Correia, G.H., de Romph, E., Timmermans, H., 2017. Using metro smart card data to model location choice of after-work activities: an application to shanghai. J. Transp. Geogr. 63, 40–47.
Wardman, M., 2004. Public transport values of time. Transp. Policy 11 (4), 363–377.
Warren, M.S., Skillman, S.W., 2020. Mobility changes in response to COVID-19. arXiv preprint arXiv:2003.14228.
Whelan, G., Crockett, J., 2009. An investigation of the willingness to pay to reduce rail overcrowding. In: Proceedings of the first International Conference on Choice Modelling, Harrogate, England, vol. 30. Citeseer.
Wielechowski, M., Czech, K., Grzeda, Ł., 2020. Decline in mobility: public transport in Poland in the time of the COVID-19 pandemic. Economies 8 (4), 78.
Wilbur, M., Ayman, A., Ouyang, A., Poon, V., Kabir, R., Vadali, A., Pugliese, P., Freudberg, D., Laszka, A., Dubey, A., 2020. Impact of COVID-19 on public transit accessibility and ridership. arXiv preprint arXiv:2008.02413.
Yang, X., Stewart, K., Tang, L., Xie, Z., Li, Q., 2018. A review of gps trajectories classification based on transportation mode. Sensors 18 (11), 3741.
Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. Who, where, why and when? Using smart card and social media data to understand urban mobility. ISPRS Int. J. Geo Inf. 8 (6), 271.
Zailani, S., Iranmanesh, M., Masron, T.A., Chan, T.-H., 2016. Is the intention to use public transport for different travel purposes determined by different factors? Transp. Res. Part D: Transp. Environ. 49, 18–24.
Zhang, J., Hayashi, Y., Frank, L.D., 2021a. COVID-19 and transport: findings from a world-wide expert survey. Transp. Policy 103, 68–85.
Zhang, N., Jia, W., Wang, P., Dung, C.-H., Zhao, P., Leung, K., Su, B., Cheng, R., Li, Y., 2021b. Changes in local travel behaviour before and during the COVID-19 pandemic in Hong Kong. Cities 112, 103139.
Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2020. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76.