UNDERSTANDING THE USER CHARACTERISTICS FOR SUBSTITUTING TRIPS BY TELEWORKING AND ONLINE SHOPPING

Juan Ramón López Soler
European Commission - Joint Research Centre. Seville, Spain
Universidad Politécnica de Madrid. Madrid, Spain
Panayotis Christidis
European Commission - Joint Research Centre. Seville, Spain
José Manuel Vassallo Magro
Universidad Politécnica de Madrid - Centro de Investigación del Transporte (TRANSyT), Madrid, Spain

ABSTRACT

This paper explores the user profile, spatial aspects, and mobility patterns affecting the adoption of teleworking and shopping online as a replacement for the working and shopping trip. The study is based on an EU-wide survey on mobility patterns and preferences with 26.500 respondents. The questionnaire combined socio-economic, mobility behaviour and the use of ICT in mobility-related issues, as well as teleworking and online shopping.

The methodology used combines classical statistical inferences and a classification Machine Learning (ML) algorithm to find the most important factors affecting the typology of the users adopting these tools.

The results suggest that while the new consumption patterns are high on average, the new working arrangements still have a margin to improve. Furthermore, there are significant differences among countries and between different socio-economic profiles. Online shopping was already prevalent, but teleworking was still limited (given that the survey was conducted in 2018, i.e., before the COVID-19 pandemic).
The findings of this work can be useful for the analysis of policies to encourage the uptake of new technologies in transport and mobility. Also, they can be a good reference point for future studies on the ex-post analysis of the impacts of the pandemic on mobility.

1. INTRODUCTION

Work and shopping constitute two of the main trip purposes for urban mobility and are responsible for the largest share of passenger transport activity (Eurostats, 2020).

Teleworking and e-commerce are two technology-enabled options that can modify individual daily mobility patterns and potentially reduce total transport demand and its associated impacts (energy consumption, CO2 and pollutant emissions, congestion, etc.).

The objectives of this paper are, on the one hand, exploring the main characteristics of individuals adopting teleworking and online shopping as a substitution of the trip to the workplace and to the traditional shop (socio-demographic, mobility-related characteristics, and the use of ICT). On the other hand, our target is to find the main election drivers by analysing the factors affecting the uptake of teleworking and online shopping.

2. METHODOLOGY

We used the second wave of the EU survey on issues related to transport and mobility, carried out in 2018, which applied a CAWI (Computer Aid Web Interview) methodology (Scarcella & Fiorello, 2017). The sample includes 26.500 questionnaires along with the 27 members of the European Union and the United Kingdom, with 1.000 respondents in each country, except Cyprus, Luxembourg, and Malta with 500 each. The sample was stratified by socio-economic characteristics based on age, gender employment status, level of education, and region of residence.

The questionnaire contains information on four categories: socio-demographics and car availability questions, information on the most frequent trip, details on medium and long-distance trips, and use of Information and Communication Technologies (ICT) related to transport, where trip substitution by teleworking and online shopping is included.

The analysis is based on a classification model, made by each of the studied variables, in which the outcome or **dependent variable** takes the value explained as follow.

On the one hand, it is defined a discrete dichotomous dummy variable **Telework (Y)** which takes the value 1 if the individual has ever replaced commuting by teleworking, once per month, 3-4 times per month, or more than 4 times per month. Likewise, the variable Y takes the value 0 if the respondent has substituted the trip to work only once or never.

On the other hand, and similarly to the previous one, the second analysis was made for the shopping trip substitution by e-commerce.

The variable **Online Shopping (Y)** takes the value 1 if the respondent has substituted the shopping trip by online commerce, rarely, sometimes, or often. Otherwise, the variable takes the value 0 if the individual has replaced the trip to the shop by buying online only once, or never.

The independent variables considered in the analysis, as explained before, include four categories (socio-economic and car availability, daily mobility, long trips and use of ICT).

3. SURVEY DESCRIPTION

In this section, we explore the impact of individual respondent characteristics on their choices, concerning telework and online shopping, by descriptive analysis, including frequency distributions and odds ratios. Later on, we use the machine learning algorithm XGBoost to obtain the relationship between the variables, highlighting the most important factors affecting individuals' choice and the overall impact on the outcome.

3.1. Trip substitution by teleworking

Males tend to telework at a higher proportion than female respondents. The odds ratio between the two genders is 1,4:1. The difference is probably due to a higher share of male respondents employed in jobs that are more suitable for teleworking. The relevance of the job type can be also induced from the correlation with education and income level. As a general trend, the ratio of teleworkers increases as the higher the education and the income level of the respondent is. Respondents with a university degree (or higher) are 2,4 times more likely to telework than respondents with just primary education. Similarly, the group of individuals with higher-middle or high-income levels are 3,8 and 6,1 times (respectively) more likely to telework than respondents with low-income level. Furthermore, looking at the income distribution by grouping teleworkers by age, we find the same pattern, so that, the higher the income, the more likelihood for teleworking. A teleworker profile that might be intuited based on the above is a male independent professional or manager with high education and high-income level.

Concerning age, the data shows an inverse association with teleworking, being the younger workers more likely to telework than advanced aged ones. This association may be related to younger's education in ICT and the lack of this type of skills in older people, but also because the group of students, normally young, might combine studies and telework.

The working day duration seems not to be an important factor for teleworking, neither seems to be for students who declare some type of telework. By contrast, people over 65 presents more probability to telework when part- or full-time work is declared, probably because they are linked to liberal and managerial professions.

Regarding the household place of residence, teleworkers present a higher likelihood of teleworking when they live in metropolitan areas or big cities rather than in small cities or rural areas. This might occur because teleworking is often linked with big companies usually placed in big cities.

Other factors affecting the probability of teleworking are car-related and mobility questions. Thus, people holding a driving license (motorcycle or car) presents a higher proportion of teleworkers. Taking this into account, owning a car presents also more teleworking probabilities despite this factor can be confounding and related to income level.

Nevertheless, individuals with a car subscription also tend to telework more than people without a car subscription, probably because this profile is associated with young professionals with low car availability and living in metropolitan areas where this kind of service is offered.

The most used mode of transport for the most frequent trip between teleworkers and not teleworkers is the car, followed by walking, private bicycle and bus services. However, using the car, the bus or going by foot are more used between non-teleworkers compared with teleworkers. Furthermore, the odds of teleworking are 5,5 times greater for bike-sharing relative to private car drivers, as do car-sharing users by 2,5 times.

In reference to the destination of the most frequent trip, the principal travel is made within the same urban area (49%) and to another urban area (34%), being this last movement more prevalent in the group of teleworkers. On the contrary, travelling outside an urban area is more prevalent in non-teleworkers. Once more, it could be due to the fact that teleworkers tend to live in metropolitan areas. Furthermore, the proportion of individuals commuting every day (or every working day) is higher in non-teleworkers compared with teleworkers.

On the contrary, people travelling two to four days per week is more prevalent in the group of teleworkers, travelling longer distances compared with non-teleworkers. This can be explained because of the size of the metropolitan city where teleworkers tend to live.

Another factor associated with teleworking is the number of long and medium distance trips, finding more trips for work business and study reason in the group of teleworkers, as well as for leisure and personal reasons. This behaviour may be associated with qualified jobs and the high-income level of the teleworker profile.

3.2. Trip substitution by online shopping

E-commerce is widely extended across Europe and is frequently used by male and female in the same proportion, being more prevalent in youngers compared with old-aged people.

Online shopping is also more used among people with high education level, for instance, graduates are 2,4 times more likely to shop online than individuals with primary education.

This could be due to higher ICT education but also because of the link with higher income level since above average income implies shopping online more frequently. For instance, people with high income are 2,3 times more likely to shop online than low-income individuals.

Similarly to the teleworking case, online shoppers tend to live in metropolitan areas and large cities, despite they have within easy reach a large product offer in the city. What is more, the bigger difference between online shoppers and traditional ones is presented in people living in the centre of a metropolitan area. At this point, this phenomenon can be explained because of gentrification of the city centre, the higher income level of the people living in central areas or the lack of car availability because of the vehicles ban in historical areas. This last assumption is reinforced with the fact that online and traditional shoppers holding a car driving license is similar. By contrast, the proportion of motorcycle license is bigger in the group of online shoppers, probably because they live in the city centre and this type of vehicles is more accessible within the downtown. Additionally, the proportion of people without a driving license is smaller in the group of online shoppers, hence not holding a driving license seems not to be a determinant for buying online.

The number of cars in the household, particularly two, presents a higher probability to buy online than households without a car. This could be explained because of the income level, nonetheless, households with more than two cars present the same proportion between online shoppers and traditional ones. This is not the case of car subscribers, where the relation is clearer, being 2,2 times more likely to buy online than non-subscribers.

The car remains to be the predominant mean of transport for the most frequent trip in both types of purchases. Concerning other means of transport, and in line with general mobility trends, walking, going by private bike and commuting by bus services are the most used means to reach the destination in the most frequent trip. This destination is usually located in the same urban area for almost half of the interviewed and more than one third are travelling to a different urban area, presenting a similar share between both types of shoppers. Given the fact that online shopping could avoid trips, we observe that people buying online tend to commute more frequently than traditional shoppers. For instance, the odds ratio of teleworkers travelling one day per week and daily commuters is 1:1,3. In addition, online shoppers spend more time and travel longer distances in the most frequent trip than traditional shoppers.

This could be due to the fact online shoppers living in metropolitan areas are more likely to commute more time and more distance to reach the frequent destination.

Lastly, online shoppers present a similar share of long-distance travels for work, business, or study purposes, but they are more likely to trip long distances for personal and leisure reasons. Furthermore, online shoppers do more medium distances trips for both purposes.

3.3. Methods

The initial data analysis suggests a strong correlation between teleworking or online shopping and certain respondent characteristics, but also suggest the existence of several confounding factors that can limit the possibility of interpreting the importance of each specific characteristic. Education level and income -for example- are to a certain degree correlated and a simple statistical analysis would not be sufficient to quantify their individual impact on the respondent's choice.

In order to solve that, we constructed a classification model that allows the generalization of the relationship between the variables taking into account the various collinearities. The model applied a tree-based approach using the well-known machine learning XGBoost algorithm. XGBoost has been tested and compared with Multinomial Logit Models in travel mode choices by Wang and Ross (2018) getting better performance. Other machine learning classifier in transport has been conducted in user choice modelling resulting in higher precision than conventional methods (Hagenauer & Helbich, 2017). Christidis and Focas (2019) analysed the uptake of electric and hybrid vehicles and car use (Focas & Christidis, 2017) within the EU using gradient boost decision trees and Random Forest respectively.

The model is set up in three parts randomly split. The first one is selected to perform the training model with 40% of the observations. The second one is the test set, with 40% of observations, which is used to evaluate the performance of the model trained before, using the previous model to predict the outcome. The third one is the validation set (20% of observations left), used to ensure the generalization of the model with unseen data.

Feature engineering is used to adapt the variables to the algorithm. We used One Hot Encoding (OHE) for the categorical variables.
XGBoost hyperparameters were selected based on the best AUC evaluation score, while the final variable election was carried out based on the outcome of the most important feature, from the predictor feature importance, and Shape Values.

The performance of the model has been evaluated with the AUC (Area Under the Curve) measure. The range of this evaluation goes from 0 to 1, being 0 when all predictions are wrong and 1 when all predictions are correct.

4. RESULTS

In this section, we describe the main results obtained from the analysis of the binary classification model regarding teleworking and online shopping.

We applied the XGBoost classification algorithm (tree-based Machine Learning model, non-linear model) to obtain the main factors affecting the uptake of teleworking and online shopping as a substitution of the trip for work and shop, respectively. For the teleworking model, the dependent variable takes the value 1 if the respondent has used this work system once per month or more, while takes the value 0 if has substituted the trip by teleworking once in his life or never.

Similarly, for the online shopping model, the outcome variable takes the value 1 if the shopping trip has been substituted rarely, sometimes, or often. If the respondent has used this service once or never, the variable takes the value 0.

In both cases, data pre-processing, feature selection, model training and evaluation has been performed. From all analysed variables, the most important features were selected in a second analysis to obtain a better performance. This selection was also based on how well these variables might explain the reasons why people perform telework and online shopping, thinking about causes and not effects.

The selected variables for the classification model were gender, age, education level, employment status, household members, income group, urbanization type, urban situation, driving license, number of vehicles per household, urban frequent destination, number of passengers in the last trip, country, population in 2018, vehicles per household member, car sharing subscription and urban-centre (combination of Urbanization type and Urban situation).

The number of observations after the data cleaning is 23.931.

4.1. Determinants of Teleworking

Once applied to the XGboost algorithm, the main factors affecting the trip substitution by teleworking are displayed in the next figure. The next graph represents the feature importance score for the most important variables ordered by how much they are helping in the prediction outcome. Thus, the more used to make the decision, the higher relative importance the variable will have.

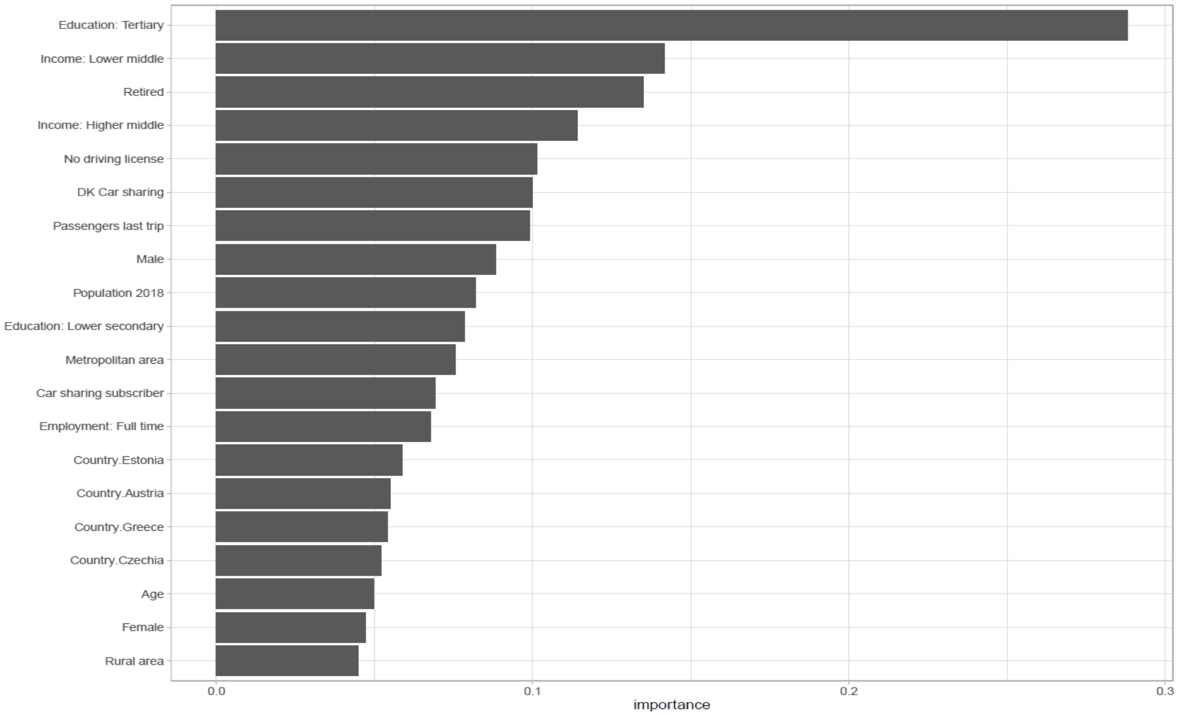

Fig. 1 – Predictor feature importance. Teleworking.

Prediction about using or not teleworking as a substitution of the working trip is influenced in the first positions by people with a high education level (tertiary or higher), with lower-middle and higher-middle income group, with employment status, with people without driving license, with gender (males), location of the household, country and age.
Figure 1 represents the main factors affecting the uptake of the trip substitution by teleworking, however, explainability is not the main feature of this kind of graphs. To resolve this, we use SHapley Additive exPlanation (SHAP), one of the most advanced methods to interpret results from tree-based models. SHAP values show the importance of each feature, the direction, and its contribution to the model outcome.

For the teleworking model, the contribution of each variable (impact on model output) is given through the shape value and the feature value as seen in figure 2. For instance, the first (and most important) variable found by the model to perform telework is having a tertiary education level, i.e., university degree or higher. The contribution to the model is given by the number on his side (0,288), which is the score. Following, we observe two dots, one yellow and one blue, meaning telework (value 1) and no telework (value 0). If the dot is on the right side, the contribution is positive, on the contrary, if the dot is on the left side will be negative. So that, having a tertiary level education will be positive for the model (feature value 1) and not having it will be negative (feature value 0)

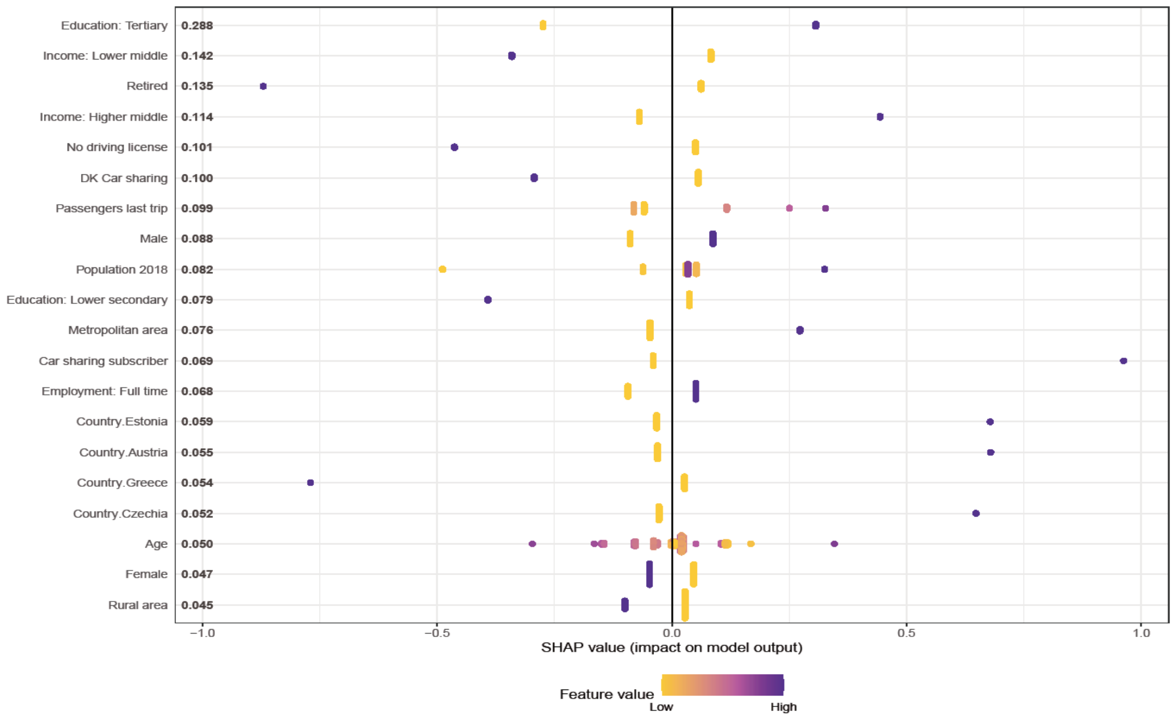

Fig. 2 – SHAP values. Teleworking.

The most important variable describing the likelihood for teleworking is to have a university degree (or higher) giving a positive value for the model, i.e., the model finds that people with high education level present more probabilities to telework. On the contrary, for individuals with a lower education level (lower secondary), the model finds a negative shape value, reducing in this case the teleworking likelihood.

Income level is normally linked with education level such as the model captures, so those individuals with lower-middle income present a negative shape value when replacing the working trip by teleworking and with higher-middle income present a positive shape value. In other words, people with higher-middle incomes tend to telework more frequently, while having a lower-middle income reduces the likelihood of telework. Gender appears to be also determinant. While male workers are more likely to telework, female professionals tend to work traditionally, probably because of the type of job developed typically by both genders.

Age presents a high variability as regards teleworking adoption, as seen in figure 2. While young people present a higher score (i.e., more prone to telework), the likelihood decreases until around 65 years old. According to the previous analysis, young professionals tend to telework more than seniors. However, when retirement age reaches, usually just liberal and managerial workers extend their professional career, precisely those with more probabilities to perform telework.

Regarding the employment status, we can identify respondents with full-time contract present a positive SHAP value (hence, full-time workers present more teleworking likelihood), in contrast to retired people, being negative, so that, retired people declaring having worked in the past are more likely to perform traditional work. This may be confounding with the previous paragraph, but it could be explained as far as retired declare to have teleworked in the past, while other people over 65 might remain working as an extension of his professional career.

Mobility patterns may also influence the teleworking choice. On the one hand, individuals without driving license are more prone to perform traditional work. On the other hand, car subscribers tend to telework more frequently, even though this profile normally is linked with high education level and living in metropolitan areas (where those services are present), which may be confounding for the model.

As for spatial factor, in accordance with the previous analysis, our model finds a positive relationship with metropolitan areas, meaning people living there present more chances to telework. By contrast, living in rural areas reduces the probability of teleworking. In addition, countries with a high population tend to telework more than countries with a low population, although work culture in every country is a determinant factor regarding remote work. For instance, individuals living in Estonia, Austria and Czechia got a positive shape value, resulting in more telework possibilities, while on the contrary, living in Greece gives a negative outcome, meaning fewer telework probabilities.

4.2. Determinants of Online Shopping

The following chart shows the main factors affecting the uptake of online shopping as a replacement of the traditional shopping trip.

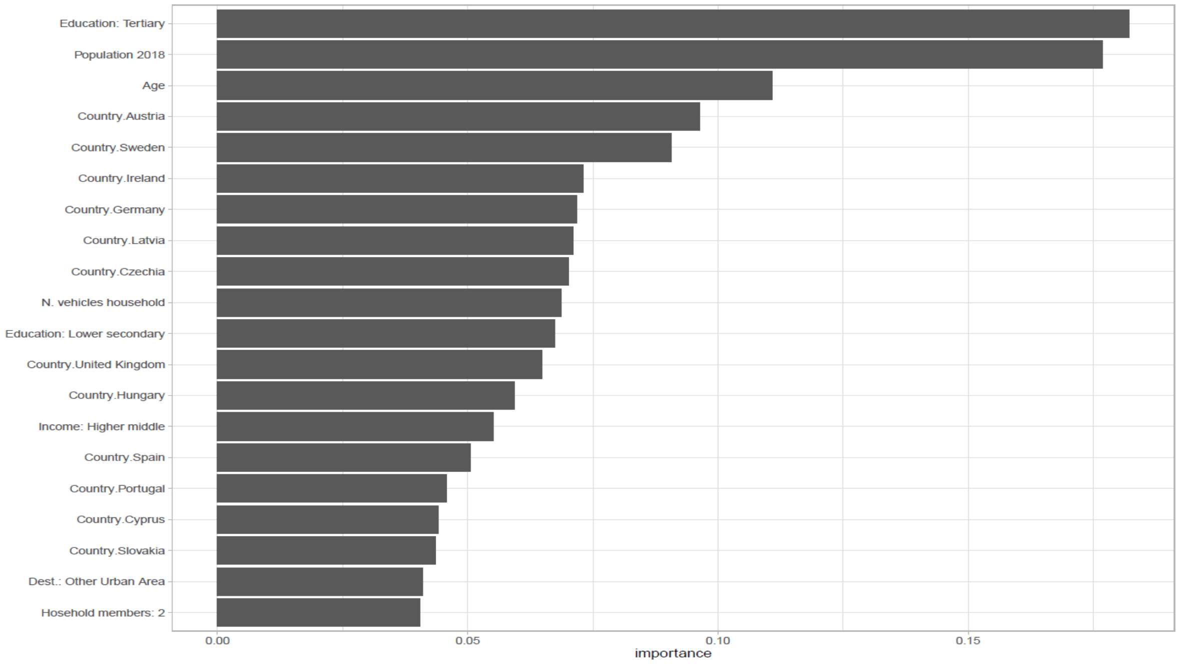

Fig. 3 – Predictor feature importance. Online shopping.

Predictions about using or not using e-commerce as a substitution of a traditional shopping trip are influenced in the first positions by people with high education level (tertiary or higher), with the level of population, age, and with people living in different countries like Austria, Sweden, Ireland, Germany, Latvia, or Czechia. It is also important the number of vehicles per household and the income level.

Next, we will explain with SHAP values the contribution of each feature affecting the prediction outcome found by the model.

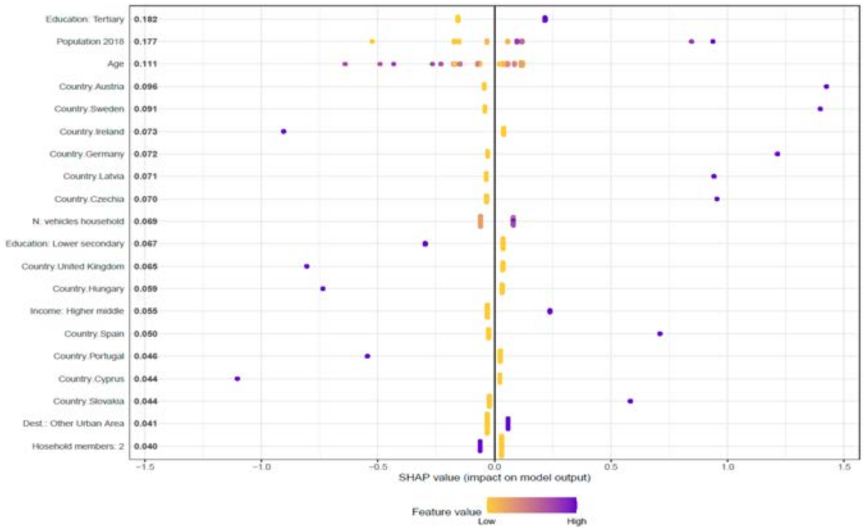

Fig. 4 – SHAP values. Online shopping.

In the same way as the teleworking case, having a high education level support the willingness to shop online as a substitution for the shopping trip, whereas people with lower secondary education tend to buy online less. Similarly, when income is higher-middle the predisposition to buy online increases. Furthermore, people between 19 and 46 years tend to shop online more frequently than younger and older individuals. In the first case probably because the acquisition power is lower, and in the second because of the lack of digital education.

Regarding individuals’ location, the model suggests high-populated countries tend to buy online more frequently than smaller countries. The country of residence appears to be a clear determinant factor affecting the online shopping choice. Countries like Austria, Sweden, Germany, Latvia, Czechia, Spain or Slovakia present a higher share of e-commerce than countries like Ireland, the United Kingdom, Hungary, Portugal or Cyprus. This result is difficult to explain given that there is not a clear pattern between both country groups, but differences in economic structure, culture and education factors, mobility schemes and accessibility or even supply chain may affect the uptake of online shopping.

4.3. Model performance

In this section, the analysis of the XGBoost classification model performance is presented. The most frequently used metric for classification problems is the Area Under the Curve (AUC) or Receiver Operating Characteristic curve (ROC). This aggregated measure of the model performance summarizes the True Positive rate (TP), versus False Positives rate (FP), by using different probability threshold.

In table 1 we present the results of the AUC-ROC test for teleworking and Online Shopping models for the test and validation datasets. The AUC range goes from 0 to 1, being 0 when all predictions are wrong and 1 when all predictions are correct.

Model	AUC Test	AUC Validation
Teleworking	0,712	0,710
Online Shopping	0,706	0,706

Table 1 – Summary of model performance.

Both models achieve a fair level of prediction, given the extensive number of variables included in the model.

5. CONCLUSIONS

This research explores the main determinant factors of teleworkers and online shoppers' profiles, replacing the trip with online procedures. The XGBoost algorithm used returns the most important factors and the impact on the model output, allowing us to determine the profile of individuals using this kind of services.

Firstly, our results confirm that the most important factors to substitute the working trip by teleworking are in general terms, having a high education level and a high-middle economic status. Gender remains determinant as far as men continue to be more likely to telework than women. As seen in the data analysis, living in low-populated countries perform fewer probabilities to telework than high-populated countries, being more likely in metropolitan areas with over one million inhabitants. However, living in certain countries favours the adoption of teleworking.

Secondly, our analysis also shows the profile of online shoppers replacing the shopping trip by e-commerce, being more likely to perform this activity by people with higher-middle education level, living in countries with high population, even though there is a high disparity between countries. The age continues to be determinant, but the range spreads from young people to the mid-age population.

The high dispersion between countries on the teleworking and online shopping adoption shows a high diversity among different regions and can be useful to bridge the gap for those who have less adoption for both ways of reducing trips through ICT.

Teleworking and online shopping can be two excellent tools to avoid trips and their negative effects (energy, congestion, pollution, transport externalities) but they must be monitored to address negative rebound effects. These rebound effects do not have to be necessarily negative. For instance, as far as more people can telework full time outside the default place of work, it can help rural or low-populated cities to attract new citizens looking for a bigger and cheaper house to live in, which eventually could change real estate market prices in big cities.

It has been mentioned high differences among European countries in the uptake of teleworking and online shopping to avoid trips. Likewise, the intensity of use shows a high variation.

On the one hand, Eastern Europe presents the lowest rate of teleworking, whereas most intensive teleworkers (declaring teleworking “often”) are placed in Finland, Austria, and the Netherlands. Besides, people declaring teleworking “sometimes” is more prevalent in Sweden, Iceland, the UK, the Netherlands, and Switzerland.

On the other hand, countries like Austria, Germany, Czech Republic, France, Sweden, and Spain present the most active (“Often” & “Sometimes”) share of online shoppers. On the contrary, countries like Cyprus, Ireland, Hungary, and Portugal present the lowest rate of e-commerce interaction (“Never” & “Once”).

As far as clean transport become more spread in the transport sector (higher share of rail mode, generalization of the electric vehicles, rising successful deliveries and optimization routes), the benefits of online shopping will be higher than now. However, congestion and most of the associated externalities will not improve with this change of technology, even though other technologies as automated cars could enhance this situation.

The scope of teleworking and online shopping remains uncertain and future work should be compared in upcoming years in order to see the evolution of these tools and the contribution to the GHG emissions mitigation. This upcoming work should answer questions about the real adoption of teleworking and online shopping habits in Europe in the post-pandemic era when both tools perform in a stable situation.

ACKNOWLEDGEMENTS

This research was carried out with the collaboration of the Joint Research Centre of the European Commission within the framework of the Doctoral Collaboration Agreement No. 35364.

REFERENCES

CHRISTIDIS, P., & FOCAS, C. (2019). Factors Affecting the Uptake of Hybrid and Electric. Energies.

EUROSTATS. (2020). Passenger mobility statistics Statistics Explained. https://ec.europa.eu/eurostat/statisticsexplained/

FOCAS, C., & CHRISTIDIS, P. (2017). Peak Car in Europe? Transportation Research Procedia, 25, 531-550.

HAGENAUER, J., & HELBICH, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications, 78, 273-282.

SCARCELLA, G., & FIORELLO, D. (2017). EU Survey on issues related to transport and mobility (N.º 933497-2017 A10-IT). EC Joint Research Centre.

WANG, F., & ROSS, C. L. (2018). Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. Transportation Research Record, 2672(47), 35-45.