IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Using A Multinomial Logit Model To Study The Mode Choice Behavior Of Commuters In Flanders

[1]Amit Pothina, [2]Prof. Dr. Muhammad Adnan

[1]Master's Graduate, [2]Associate Professor

[1]Transportation Sciences,

[1]Hasselt University, Hasselt, Belgium

Abstract: This paper demonstrates the possibility and the viability of combining new data resources with traditional surveys to estimate a mode choice model for work/school trips. These modelling tools are used to evaluate the effects of behavior change. Its purpose is to examine the travel mode choice preference of commutes to work/school. The proposed model is tested using a travel survey conducted in the Flanders region in Belgium referred to as the Onderzoek Verplaatsingsgedrag 5 (OVG), which translates to research on travel behavior. The objective is to analyze the correlation between mode choice, mode characteristics and socio-economic attributes of individuals. These are computed using a multinomial logit model (MNL) through a utility function. After testing out the significance and correlation of over 50 mode-specific and individual- specific characteristics using the Biogeme package, travel time, education level, and income level were identified as the attributes with the highest utility in the model. The OVG 5 data contained more than 100 individual attributes and mode characteristics, all of which were tested in the model's performance. The final utility function uses the 15 most significant parameters and yields the highest final log-likelihood score compared to the other models. Findings of this research contribute knowledge towards making transport policies in Flanders, for example, inducing a modal shift to more sustainable modes of transport amongst commuters. There are also behavioral implications and transport planning benefits associated with the findings of this research. The dataset was collected until early 2020, thus making it the last pre-COVID travel survey in Flanders. Future research like the OVG 6 and higher can use this model as a base model and build upon it because the study area will remain the same.

Keywords **- Biogeme, Multinomial Logit Model (MNL), utility function, mode choice model, travel behavior, travel survey.**

I. INTRODUCTION

For the last three decades, the Flemish Government has been actively researching the relocation and travel behavior of the Flemish people. This research is called travel behavior research (Onderzoek Verplaatsingsgedrag Vlaanderen in Dutch) or OVG. This research aims to analyze several mobility characteristics of families and individuals collected. The focus is on mapping out the travel behavior of the Flemish as accurately as possible (Onderzoek Verplaatsingsgedrag Vlaanderen 5, 2020). Increasing traffic congestion in intercity highways has raised severe issues regarding the detrimental effects of such congestion on local financial development, countrywide productiveness, sustainability, and environmental quality. In 2014, Brussels and Antwerp, the two largest cities in Belgium, were also ranked as the two most congested cities in both Europe and North America (INRIX, n.d.). To overcome the current trends and prevent further exacerbation, attention has been focused on identifying and proposing solutions to improve the transport systems within cities. For example, the EU has announced a budget of 2 billion euros towards cycling initiatives, and the region of Flanders has carried out successful initiatives in this domain (Times, 2021). There has also been a lot of development in transport infrastructure and facilities across Belgium under the Flemish Ministry of Mobility and Public Works. An integral part of these developments is the work/school trips made by individuals as these contribute to most (close to 40% of all trips) people make (Passenger Mobility Statistics, 2021).

Flanders Region

Roads and public transport are well connected in the region of Flanders. Still, developing a mode choice model for this region also poses challenges as it is relatively small, 13,625 km2 and its residents have a complex network of transport alternatives to choose from. This paper analyses the OVG data, the fifth consecutive study into travel behavior in Flanders, conducted between January 2015 and January 2020. The OVG 5 is a continuous study that is spread over five years. However, the travel survey was not targeted for mode-choice modelling of its commuters and is missing relevant attributes such as origin-destination travel time and distance using different modes of transport. As the utility function compares all the available travel modes, we also need to have this data available for modes that were not chosen. Belgium also has a high share of cars per household. The Flemish government has implemented several strategies and plans to promote carpooling, thereby increasing the vehicle occupancy ratio and decreasing congestion on roads. These strategies have existed since the 2001 Mobility Plan of Flanders. The government has always worked with employers to promote carpooling, and thus, this research can help them further improve these strategies to curb traffic congestion (Cools, 2013).

II. REVIEW OF LITERATURE

Transport modelling is a tool used to assess transport infrastructure and its complex relationship with the behavior of its users. These tools can be used to bring about the needed behavioral change. Conventional modelling methods prolonged with activity-based models help evaluate the impact of traffic and travel behavior where the main issue is identifying the most relevant modelling parameters (Binder et al., 2019). Mode choice is the selected travel mode for a trip from a set of available transport alternatives known to the individual. The individual considers all the elements such as alternatives, socio-economic factors, and time constraints before executing their travel activities. When there are more than two discrete mode choices available to the individual, a multinomial logit model is implemented (binary logistic regression is limited to precisely 2 available discrete choices). It is a statistical regression model that generalizes situations with more than 2 available discrete outcomes. The model predicts the possibilities of each of these outcomes using a dependent variable, the mode choice in this context, and a set of independent variables – the attributes and constraints of the individuals. These results can easily be interpreted, and they also explore the individual effect of the parameters on the mode choice (Paulssen et al., 2014). Thus, this research aims to study the correlation between the mode choice, individual attributes, socio-economic characteristics, and the available transport infrastructure using a multinomial logistic regression model. It identifies those variables that have the most significant effect on these discrete choices made by the individual.

Along with this is the formulation of the utility function and how its efficiency is maximized to suit the available dataset. The inferred results will be crucial to transport planners and policymakers interested in examining the influencing factors of behavior change. The paper has been presented in sections where the research goal is defined in the objectives section, the technique of data cleaning and preparation is explained in methodology, and the inferences and other results of the research are highlighted in the results and discussion section.

Four-Step Model

Developed by the Chicago Area Transportation Study (CATS), the four-step model is one of the most widely used forecasting models, which breaks down fractionates the estimation into four steps (McNally, 2008). The sequential four steps of this model are as follows:

1. Trip Generation, this step estimates the number of trips originating or ending in respective zones.
2. Trip Distribution, this step distributes the various destinations obtained from the first step.
3. Mode Choice, each of the trip volumes is distributed based on the transportation modes.
4. Trip Assignment, the final step assigns routes and network links for their corresponding trips.

This paper focuses on the third step of the model - mode choice within the context of commuting between home and work. Although the four-step model has its shortcomings, it remains a popular and practical approach to forecasting travel demand.

Multinomial Logit Model (MNL)

The multinomial logit model is a type of logit model that addresses more than two discrete alternatives. By discrete, it is implied that if a commuter walks a little to catch a bus to get to work, their corresponding mode choice would be public transport. Discrete in their choice is distinct and absolute. It is primarily based on the assumptions that the error elements comply with a Gumbel distribution and are independently distributed throughout the transport alternatives and individuals (Koppelman & Bhat, 2006). Discrete choice modelling is based on the principle of utility maximization, where an individual chooses the alternative that offers the highest utility amongst the available choices. This utility is the amount of value or functionality that the choice provides to them (Schiffer et al., 2012).

Biogeme

"Biogeme is an open-source Python package designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models" (Bierlaire, 2020). Biogeme (**BI**erlaire's **O**ptimization package for **GE**V **M**odels **E**stimation) was developed to estimate many random utility models using optimization algorithms and other python libraries. According to Biogeme's developer (Bierlaire, 2003), the objective of developing Biogeme is to provide researchers with an appropriate and efficient tool enabling them to explore new models tailored according to their specification without worrying about the estimation part. Biogeme package was used to conduct statistical tests and estimate the coefficients of all the parameters in the utility function using a standard technique called the maximum likelihood estimation. A given dataset is edited only to select a particular optimization algorithm for most users. The data file contains in its first line a list of labels corresponding to the available data, and every line that follows includes the same number of numerical data with each row corresponding to an observation.

III. RESEARCH OBJECTIVES

Developing a mode choice model for work/school trips in the Flanders region in Belgium can statistically reflect the unique transport behavior displayed by the individuals of this region. It can serve as a reference for future trends and research. This data was collected just until the coronavirus pandemic broke out (Onderzoek Verplaatsingsgedrag Vlaanderen 5, 2020). Thus, making it the latest available model to the Flemish Ministry of Mobility and Public works regarding work/school trips of Flemish people before the pandemic.

This model addresses the concerns of Mladenovic & Trifunovic (2014), such as discrete choice models not taking walking and bicycling into consideration which are the most sustainable modes of transport. The methods used in the scope of this research are through APIs that are openly available to the public. This methodology of enriching the available dataset to derive trip characteristics of both chosen and unchosen mode choice attempts to circumvent this roadblock caused by privately owned data, thus heralding all transport researchers to have an open-access approach to studying travel behavior. The findings of this research aim to identify the most influential attributes in the decision-making behavior of commuters, and the models themselves seeks to serve as a foundation for future research.

IV. RESEARCH METHODOLOGY

Sources of Data

The data was collected by the Flemish Ministry of Mobility and Public works from 2015 to 2020. A face-to-face survey was conducted with these respondents. The respondents had to complete two questionnaires. First, a family questionnaire - containing several questions about the family characteristics and second - a personal questionnaire containing various questions about the travel and personal characteristics. In addition, each respondent also received a travel booklet in which their travel movements were made a note of daily.

Data Enrichment and Preparation

The original OVG dataset shared had data stored following the questionnaire that the respondents filled. The data had to be enriched first because some crucial data required for the model was still missing. For example, mode choice, the respondents' home address and work/school address were filled available, but the distance and travel time was unavailable neither for the chosen mode nor for the alternate modes. To make this enrichment, additional data was processed from a Google Maps Application Programming Interface (API), a trip- building web platform with comprehensive multimodal travel information aggregated from multiple sources. The Google Maps API gives access to 2 crucial functions named "GoogleMaps_Distance" and "GoogleMaps_Duration". These two functions can be used to calculate the same for four different modes of transport: walking, bicycling, car, and transit (public transport). Using this API, the travel distance and trip time for all modes and origin- destination pairs were calculated.

In the raw dataset, 20374 respondents participated, but many of them did not complete their survey correctly. This led to multiple data inconsistencies such as missing work locations and, for some, invalid address. Identifying these errors and filtering them out was thus a crucial step to minimize the discrepancies in the model. After processing all the data, the final number of valid home to work/school trips that were estimated totaled up to a sample size of 9384.

Finalizing the Biogeme Model

Based on the literature on Biogeme, a specific discrete choice model can be finalized based on multiple conditions. Some of these include the significance of the variables. Only those variables whose p-values are less than 0.05 are retained. This ensures that only significant trip characteristics and individual attributes remain in the utility maximization. One way to cross- check this is by checking the final log-likelihood value. This value can be compared with multiple models, and a higher value for this is a more favorable one.

The other important element of a model is the coefficient. For example, the coefficient of time and distance are usually negative because when distance increases, time also increases and vice versa. The correlation of all the variables can be used to make inferences about the travel behavior of commuters. Not all categories of income and diploma were found to be significant for the model. This model considers walking as the base model. For character attributes, it factors respondents with diploma levels 2, 7, and 8 using bikes and respondents with diploma level 5, 6 and 7 using cars. Income level 2 was found to be a good match for bike and income levels 3 and 6 corresponded to cars. Travel time was found to be the most significant trip characteristic. Thus, this model is a good fit and the data used here is also an appropriate match. The final log-likelihood test is also conducted for the MNL model.

V. DISCUSSION

After multiple iterations and attempts at formulating a significant utility function, the useful utility functions were identified. In the models created by Biogeme, the p-values are used as parameters to select the relevant attributes. A p-value of less than 0.05 proves that the corresponding attribute is significant to our model. For bicycle as a mode of transport, it is interesting to note that people who completed only lower education, people who attended a college and people who attended university were all found to be significant to the model. This is further elaborated in the descriptive analysis section. Similarly, for car users, those who just finished their schooling whose inputs were significant to the model. Income also played an important role, especially because income is directly proportional to their car usage.

Role of Income

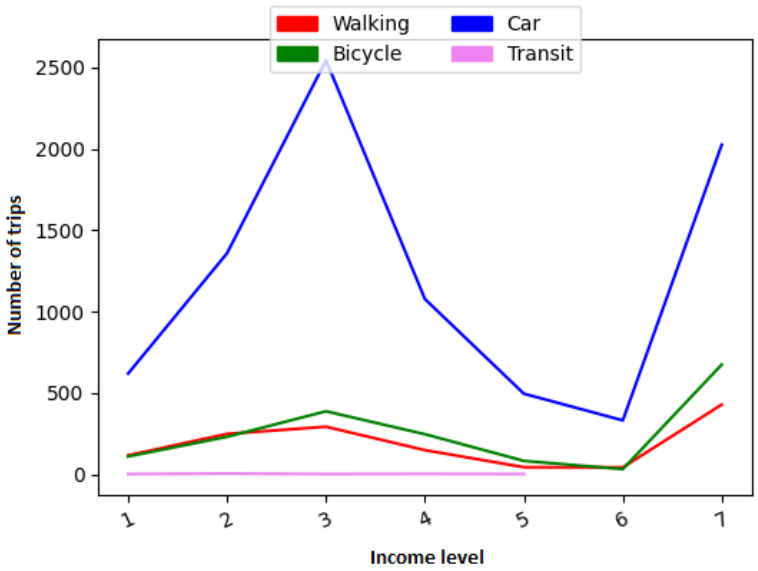

Figure 1. Line graph of Income level vs Number of trips

Figure 1 shows the correlation between income level and usage of transport modes. The income level is indicated as a categorical value where 1 is the lowest and 7 is the highest. As suggested by the Biogeme model, income level 3 and income 7 has a lot of users for bicycling. However, it is interesting to note that those with a high income (level 6) have decreased usage of the car. However, they have reduced usage of other transports too. This goes on to show a particular data bias while collecting the sample. Regardless, it still can be said that the commuters with high income (level 6) still use the car more often than other modes of transport.

Role of Education

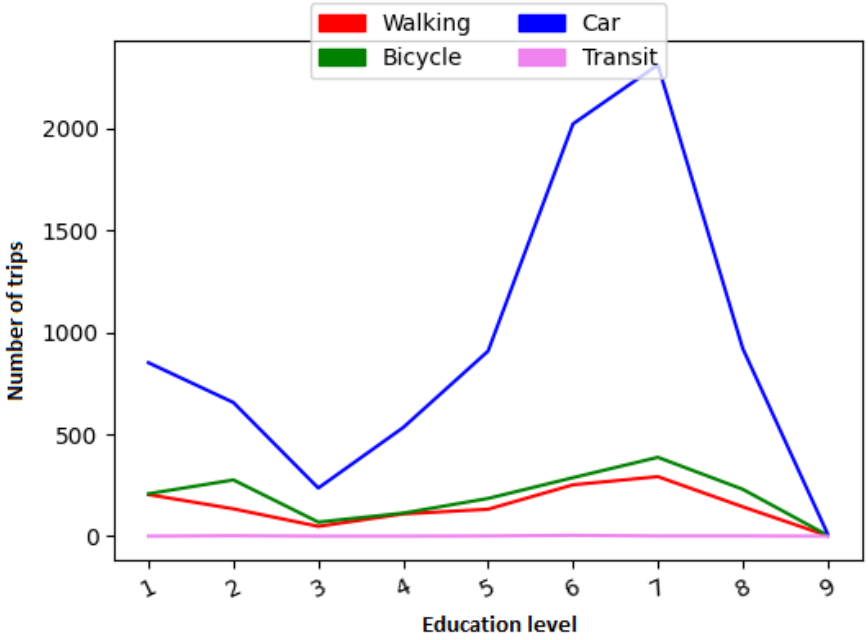

Figure 2. Education level vs Number of trips

Figure 2 shows the correlation between education (diploma) and the modal. This graph is in adherence with the model generated. It is interesting to note that car usage is directly proportional to the education level until level (8) which means the respondent has graduated from college or higher. This could also be associated with their income level as a higher level of education is likely to yield a higher paying job, making the car an affordable option.

Role of Age

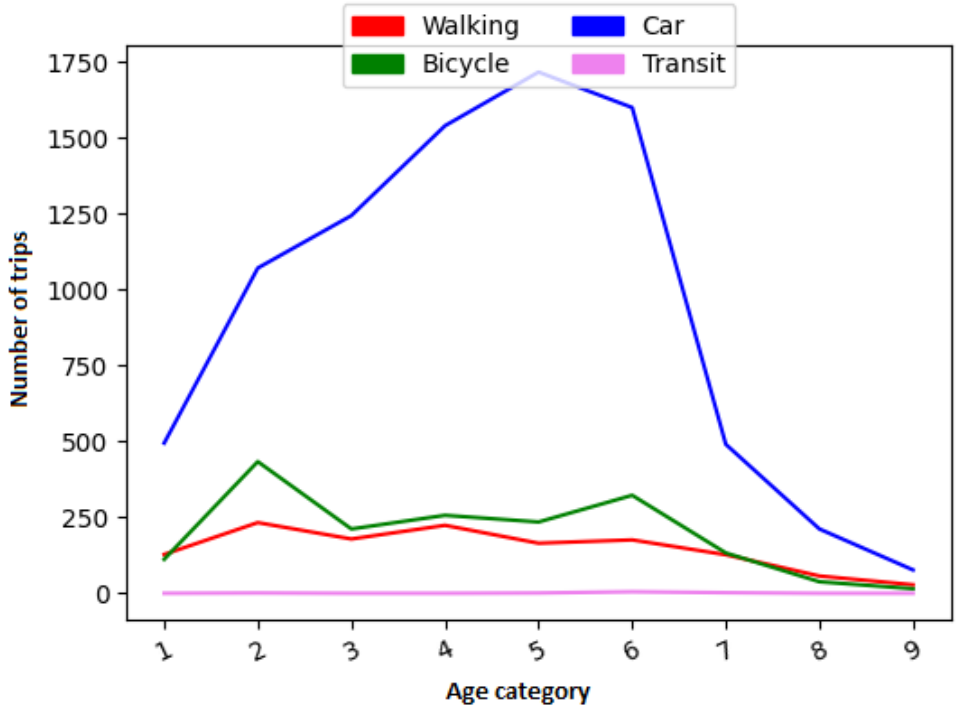

Figure 3. Age category vs Number of trips

Age in the OVG dataset was in years which has been segmented categorically in the figure 3. Here 1 represents the age 0 to 10, and it increases in increments of 10, with each category finishing at 80-90 at 9. It is evident that travelers who belong to the age group of 10 to 20 are more likely to walk or bike to get to school or their place of work. It is interesting to note that many minors are users of cars. This is because the mode choice car is a combination of those use car as driver and those who use the car as a passenger. In addition, there is a sudden plummet in the usage of bikes after the age of 20, which can be attributed to the fact that these commuters now possess a driving license. There is also a sudden spike between commuters having a driving license before and after the age of 20.

VI. DISCUSSION

This study documents the development of a multinomial logit mode choice model for work/school trips in the region of Flanders, Belgium. The model is based primarily on the OVG 5 data collected by the Flemish Ministry of Mobility and Public Works. The data was further enriched using web-based data sources. To achieve this, the available home and work/school address were converted into coordinates using the Google Maps API. The lack of mode-specific data was addressed by utilizing this API embedded into the apps script of Google Docs. The origin and destination of each trip position were defined since the exact location of both home and school/work was available. The API queried all possible mode choice data for all OD pairs. The sample size was filtered here because of missing data and invalid addresses. Only work/school trips were studied, thus making a uniform travel objective for all the respondents in the study. This was followed by the analysis of the relationship between mode choice and possible explanatory variables. The parameters were chosen based on theoretical consistency and statistical relevance (only those less than the 5% interval). After testing multiple trip characteristics and individual attributes like age, gender, travel time, parking time, parking fees, diploma, income level and others until I arrived at an optimal utility function consisting of coefficients like travel time, income level, and highest diploma attained. Other trip characteristics and individual attributes variables gender, age, travel group size, type of dwelling, etc., were gradually filtered out. To achieve a theoretically consistent value-of-time, a generalized time variable was used. It was calculated by converting travel cost into travel time using a specific value of time.

The study aimed to estimate the influence of the individual attributes and trip characteristics on the transport mode choice of individuals. An MNL was used to analyze these results and was processed using the Biogeme package. Although the modelling techniques and variables used were justified, the estimation confirmed that most individual attributes do not significantly influence the mode choice of individuals. Rather, the trip characteristics have a strong influence. The coefficients of the p-value support this finding. Thus, the trip characteristics, namely, trip distance by chosen mode, trip distance of unchosen modes, travel time and chosen and unchosen modes and referred to as availability of transport modes by the individual, were identified as highly influential factors of mode choice of individuals. The model closely matches the overall modal share of the original dataset.

Scope for Future Research

The work of Koppelman & Bhat (2006) has suggested the that the Nested Logit (NL) model is superior to the MNL model. However, in the scope of this study, the MNL model was applied because of the lack of data to apply an NL model. Data for the number of trips was available, but nothing regarding the location was available in the dataset. Using the Google Maps API allowed compilation of all modes from one single database, but in future research, analysts must explore the other platforms too. Meanwhile, the value of the trip should not be calculated using such media. These platforms are primarily built for those unfamiliar with routes. When these platforms calculate the cost of a trip, they assume that the commuter would buy a single ticket for their journey while the commuter is most likely to be in possession of a season pass. Thus, a better way to calculate the cost is to calculate a time cost based on the individual attributes of every individual.

More than 70% of all the work/school trips in this sample were made by car. This makes it challenging to predict the behavior of individuals who use the other three transport modes. The utility function predicts a car as the desired choice, and it also turns out to be accurate because of the sheer volume of car users. Schiffer et al. (2012) pointed out that it is almost impossible to reuse these models in other regions even if the dataset and attributes of people are the same. However, this model can be used as a base and modified according to the region. Despite all that, this model was tailored for the region of Flanders and can still help urban planners and policymakers make strategies in the future. Future research should further explore the use of databases like Google Maps API, TomTom traffic index, Rom2Rio and Bing Maps API in modelling, as it would significantly reduce the time and effort spent on data collection and can potentially increase the accuracy of the model.

REFERENCES

[1] Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Swiss Transport Research Conference*, *CONF*.

[2] Bierlaire, M. (2020). A short introduction to PandasBiogeme. *Transport and Mobility Laboratory, ENAC, EPFL*, 22.

[3] Binder, R. B., Lancaster, Z., Tobey, M., Jittrapirom, P., & Yamagata, Y. (2019). *TRANSPORT MODELING WITH A PURPOSE: HOW URBAN SYSTEMS DESIGN CAN BRIDGE THE GAPS BETWEEN MODELING, PLANNING, AND DESIGN*. 85–96. https://doi.org/10.2495/UT190081

[4] Cools, M. (2013). *Unravelling the determinants of carpool behaviour in Flanders, Belgium: Integration of qualitative and quantitative research*. 12.

[5] INRIX. (n.d.). *INRIX traffic congestion ranking*. Inrix. Retrieved January 1, 2022, from https://inrix.com/scorecard/

[6] Koppelman, F. S., & Bhat, C. (2006). *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. https://trid.trb.org/view/793000

[7] McNally, M. G. (2008). *The Four Step Model*. 19.

[8] Mladenovic, M., & Trifunovic, A. (2014). The Shortcomings of the Conventional Four Step Travel Demand Forecasting Process. *Journal of Road and Traffic Engineering*.

[9] *Onderzoek Verplaatsingsgedrag Vlaanderen 5*. (2020). www.vlaanderen.be. https://www.vlaanderen.be/mobiliteit-en-openbare-werken/onderzoek-verplaatsingsgedrag-vlaanderen-ovg/onderzoek-verplaatsingsgedrag-vlaanderen-5

[10] *Passenger mobility statistics*. (2021, November). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger_mobility_statistics

[11] Paulssen, M., Temme, D., Vij, A., & Walker, J. L. (2014). Values, attitudes and travel behavior: A hierarchical latent variable mixed logit model of travel mode choice. *Transportation*, *41*(4), 873–888. https://doi.org/10.1007/s11116-013-9504-3

[12] Schiffer, R. G., National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2012). *Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models* (p. 22661). Transportation Research Board. https://doi.org/10.17226/22661

[13] Times, T. B. (2021, December 31). *€2 billion for bikes: EU triples spending on cycling initiatives*. https://www.brusselstimes.com/belgium-all-news/199862/two-billion-for-bikes-eu-triples-spending-on-cycling-initiatives-in-recent-years