Assignment: Transport Planning Methods

Liliana Cedro
Lucio Tassone

Assignment **Fall 2021**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Assignment: Transport Planning Methods

Liliana Cedro
ETH Zürich
CH-8093 Zurich
lcedro@student.ethz.ch

Lucio Tassone
ETH Zurich
CH-8093 Zurich
ltassone@student.ethz.ch

Fall 2021

Abstract

This paper consists in a descriptive statistical analysis on mobility data from the MOBIS 2019 survey for the canton of Zurich. First the data has been cleaned and then a two step analysis follows. The first step focuses on travel behaviour of the respondents. The second step focuses on socio-demographic characteristics of the sample.

A linear regression has been created, observing how the number of trips per participant per week vary respect to various variables, as income and general accessibility. A gravity and an MNL model are then generated on the basis of the data, which allow to forecast the probable trip distribution and modal split between each pair of zones.

In the last part of our analysis, it has been modelled a scenario where autonomous vehicles are the only vehicles present in the canton. In particular, scope has been set on the trip flow between Zurich city borders. It has been assumed that the vehicles have been introduced substituting public transport and cars in year 2020. A Cost Benefit Analysis has been calculated considering 40 year time frame of project effectiveness, as recommended by the Swiss norms. In particular, construction Costs, travel time savings, induced traffic benefit, and decrease incident rates have been calculated.

Suggested Citation

Axhausen, K. W. (2016) Style guide for student dissertations, *Working Paper*, **1140**, IVT, ETH Zurich, Zurich

1 Introduction

1.1 Descriptive analysis

The data used in this exercise was provided by MOBIS (Mobility in Switzerland) and was collected entirely in Switzerland. The aim of this study was to understand the rhythms of travel behaviour and the influence of mobility pricing measures on the people travelling. The participants were asked to track their mobility with a mobile-phone based app called "Catch-my-Day". In parallel, many web-based surveys were conducted in order to gather data on the personal and household level. In this paper, a subset of the MOBIS dataset focusing on the Canton of Zurich area has been analysed. A descriptive statistical analysis has been carried, assessing people behaviour. Social variables as income, household size, or education level, have been paired with quantitative variables as number of trips, length or duration of the trips. This analysis is found in section 2.

1.2 Trip production: Regression and models

Using a Linear Regression, it has been investigated the relationship that exist between number of trips per week per participant and various regression variables, primarily income and general accessibility. A further step has been to add various control variables. This analysis is found in section 3.

Through the Furness method, an gravity model is then generated. This allows the comparison between the real trip distribution observed through the data and the modelled trip distribution. The second implemented model is an MNL model, created with the help of a special *R* package. This allows to predict the modal split between each pair of zones. Indicators such as the goodness of fit and the value of travel time are also investigated. The two model combined offer an overview on the forecasted average number of trips with each mode between each pair of zones. Those models are found in section 4.

1.2.1 Cost Benefit Analysis

Cost benefit analysis are an essential tool to establish whether a potential policy should be implemented or not. We performed one on the project of introducing autonomous vehicles in the canton of Zürich, as both public and private transport.

The scope of the projected is limited on the flows entering or exiting the city of Zürich (i.e. zone 112). This policy causes increased travel time for these trips of 30%. This paper will calculate whether the benefits of this policy out-weight the costs. The benefits are the net benefit of induced traffic (our policy will decrease the trips to and from Zürich due to the increased travel times) and the decreased accident costs. The costs will be represented by the initial fleet investment and the costs due to increased travel time.

A profitability calculation with a discount rate is then performed in order to compare costs on the same reference date. A sensibility analysis on the construction costs, the discount rate and the value of travel time will then be performed. This allows us to assess the robustness of our calculations.

The project will be recommended if the total net present value is positive.

2 Part 1: Descriptive analysis

The data sets used for this first part are the following:

- *participants*: contains socio-demographic characteristics of participants such as income, household size, education, age, sex, employment, travel behavior,...
- *cb_participants*: is a Codebook for the data set *participants*
- *trips*: contains informations about the participants' trips for a randomly sampled week in fall 2019 such as time, length, mode, location,...
- *cb_trips*: is a Codebook for the data set *trips*

It is important to note that the data was not gathered on the same calendar week. The participants had to choose a week between September and October 2019 and log all their trips that started and finished in a district of the Canton of Zurich.

In this report the data will be initially cleaned and prepared for the analysis. Then the correlation between various socio-demographic parameters and travel behavior will be highlighted through graphs. In the end a short overview of the information about the participants will help better understand some behaviors observed in the previous section.

2.1 Data import and cleaning

Before directly processing the data in R, some cleaning of the data set *trips* is necessary. This process is described in the following points:

- **Cleaning mode variables**: in the variables *mode* and *mode_ agg* an error occurred, possibly during the data export. The additional string "Mode::" is present before the actual mode. This is corrected by simply removing the string for each element of the two variables.

- **Rounding of variables *length* and *duration***: this two variables are represented with a precision that is not necessary. Working with the rounded numbers is enough for the scope of this assignment and thus all these measurements will be floored.

- **Determining of the main leg of the trip**: every trip is divided into legs, which are defined as the smallest unit of movements using one mode of transport. To be able to analyze data based on trips, it is necessary to condense this information into one for every trip. Here the decision made was to consider the longest (i.e. with maximal length) leg as the trip itself.

After these cleaning steps the most important information on the data set *participant* was moved in the data set *trips*. This was the case for the variables *household size* and *income*, which were needed to conduct the following statistical analysis.

For some sensitive questions like income or employment some people chose the option "Prefer not to say" or directly didn't responded. These participants are not to already be completely disregarded, because some analysis on the other information is still possible. This cleaning step will be then undertaken wherever necessary.

Moreover, variables as length or duration of the trip presented various outliers. Trip behaviour could vary a lot, therefore even those have been left in the data set.

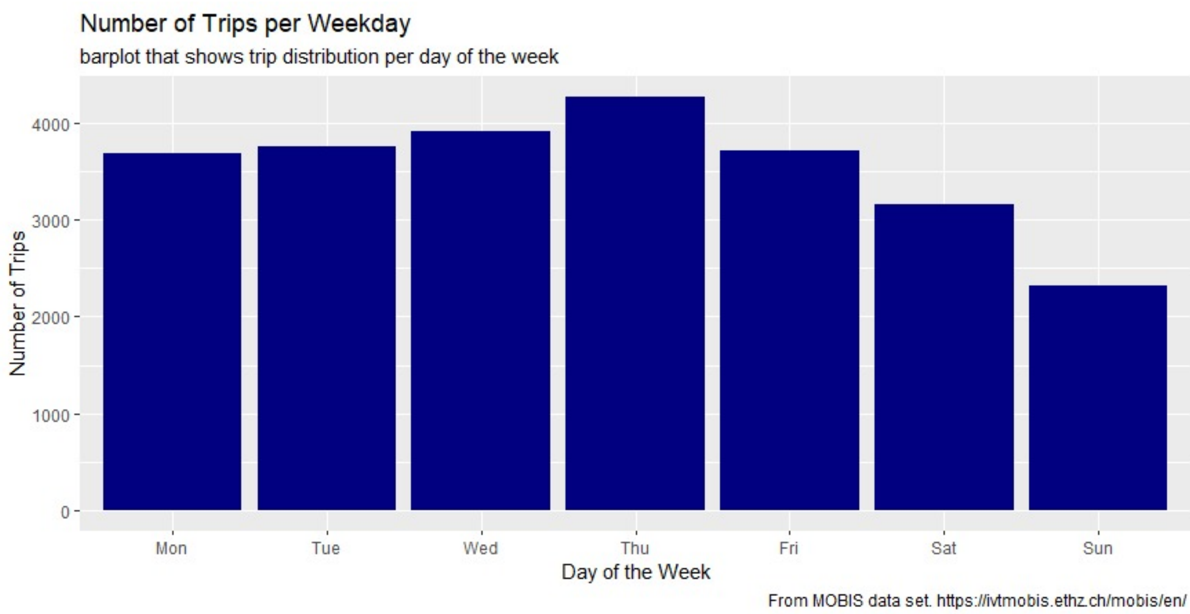
2.2 Travel behaviour

After the cleaning step the descriptive analysis has been carried. Results are shown in the following subsections.

2.2.1 Number of trips per weekday

A first analysis on the given data can be made on the total number of trips made each day of the week. As already stated before, these numbers don't refer themselves necessarily to the same day of the same week but they are the sum of the total trips made on that weekday over the observed period of time.

Figure 1: Trip distribution per day of the week



As expected, the days registering the highest number of trips are the working days from Monday until Friday. Commuting to and from work is one of the main reason why people travel, while on the weekends less people have to make work-related work trips.

Sundays registers the lowest number of trips, which is also not a surprised. In Switzerland Sundays are (mostly) holidays and this means that also shopping- and commissions-related trips don't occur as much as on the other weekdays.

2.2.2 Number of trips per income

Another interesting relation which can be analyzed is the number of trips made per income class. Income classes are defined in the following table:

Table 1: Income classes definition

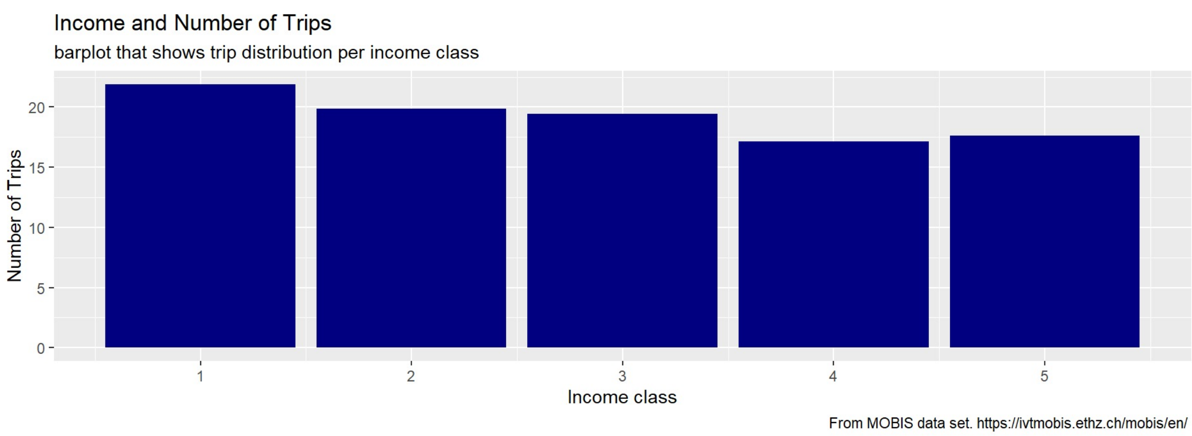
Income class	Household income
1	4'000 CHF or less.
2	4'001-8'000 CHF.
3	8'001-12'000 CHF.
4	12'001-16'000 CHF.
5	More than 16'000 CHF.

Source: MOBIS data set. https://ivtmobis.ethz.ch/mobis/en/

To avoid bias due to the number of participants for each class, the following number of trips will always be normalized over the number of people in each income class. This value assumes then the meaning of average number of trips per person.

Since income is a sensitive topic, not all the data was used for this analysis since it was missing or simply not provided. The total amount of deleted data was around 12% of all the responses.

Figure 2: Trip distribution per person per income class



From Figure 2 it can be seen that people with a lower income class tend to undertake a slightly higher average number of trips per week. This could be explained by the fact that usually people with higher incomes can afford to live in places with higher accessibility and easier mobility. This allows them for example to sum up multiple activities in the same trip. Another aspect is home delivery, which is more expensive than picking things up yourself and for this reason preferred by people who can afford it. This results in less trips made by the person themselves.

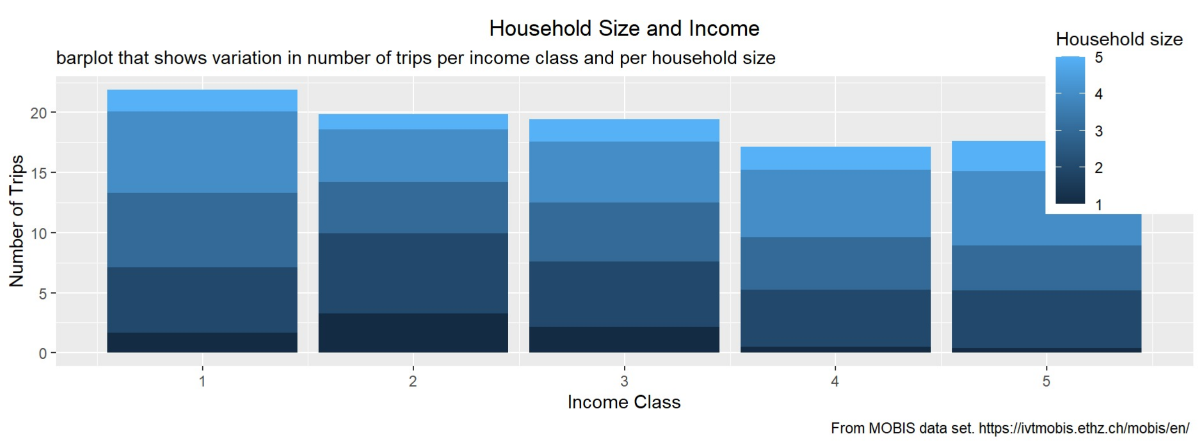
2.2.3 Number of trips per income and household size

This analysis is similar to the one conducted in the last chapter, but it takes household size into consideration as well.

With household size is meant the number of people living in the household of the participant in the study, with 5 meaning five or more people.

The first data visualization allows to interpret the number of trips per income presented in Figure 2 also by household size. It is important to note that the normalization is always done based on the participants per income class.

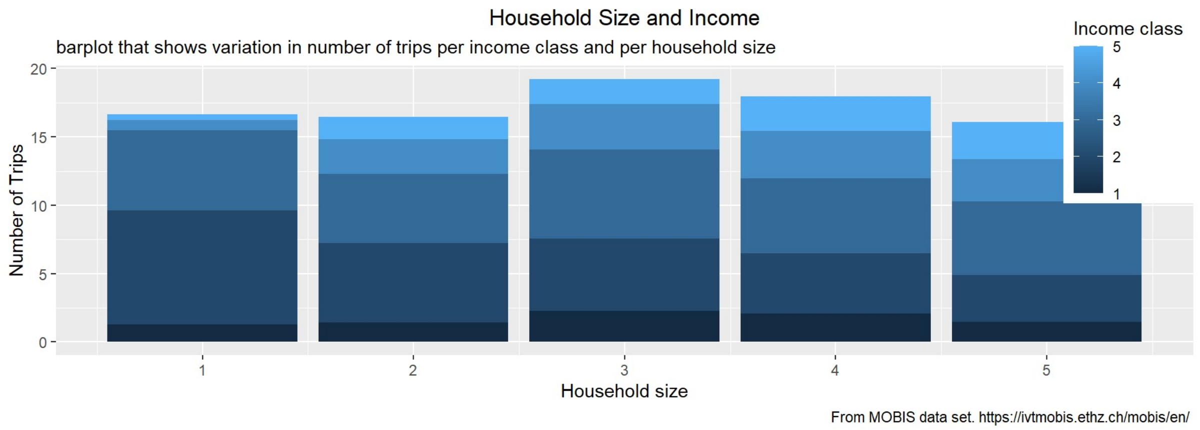
Figure 3: Trip distribution per income class and household size



A second visualization consists in analysing the number of trips by household size directly. This time it will obviously be necessary to normalize the number of trips by the amount of participants for each household size, in order to have an idea of the average number of trips based on how many people live together.

From Figure 4 it can be seen how 3- and 4-people households travel slightly more than the other categories. Families with kids are likely to move more to pick up the children and do activities with them. On the contrary, families with more than two kids are probably more limited in their movements and prefer more to stay at home to manage the children easily (for example they are maybe less likely to go to restaurants in the evening).

Figure 4: Trip distribution per household size and income



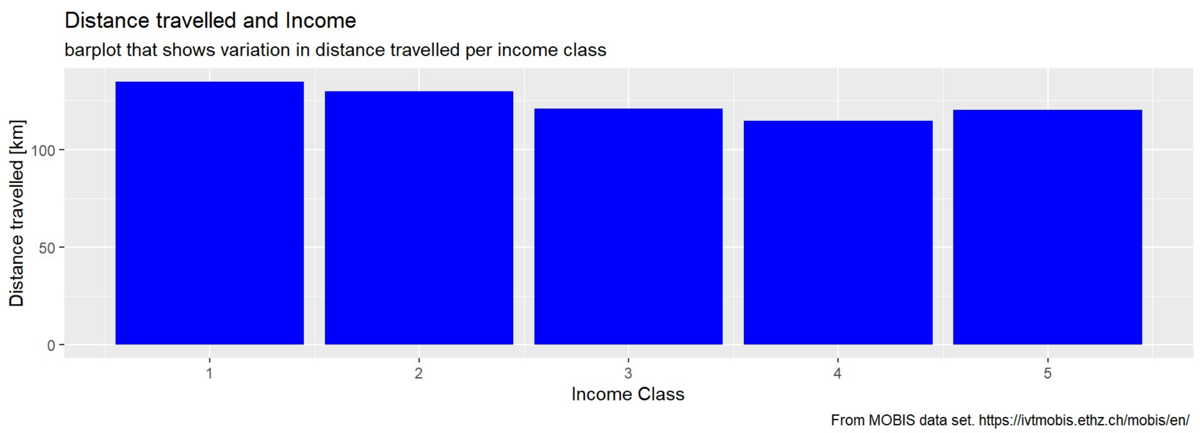
Income plays a big role in the number of trips for people that live alone. Participants from lower-middle income classes (i.e. 2 and 3 especially) seem to travel significantly more than people with higher income.

2.2.4 Distance by income

This section focuses on the distance travelled based on the income class of the participant.

Here as well the total distance travelled will be divided by the number of participants for each income class, which allows to purely observe correlation between the two variables. For this reason the distance travelled presented in Figure 5 is to consider equal to the average distance travelled by one participant of the respective income class.

Figure 5: Average distance travelled per person per income class



The shape of this graph follows closely the one in Figure 2, which compared income class to number of trips. This is to be expected, since a higher number of trips will automatically produce longer distances travelled.
In this case as well we can attribute the longest distance travelled of participants from lower income classes to the lower accessibility that they can afford. For example, it is not unusual for people working in the center of Zurich to not live in the proximity, since housing can be very expensive. They will instead commute from other neighbourhoods or entirely other cities and this will have an impact on their total distance travelled.

2.2.5 Mode Choice

To analyse the participants' mode choice two different methods were chosen: counting the total number of trips made with each mode and summing the total distance travelled with each mode.

Figure 6: Number of Trips per chosen travel mode

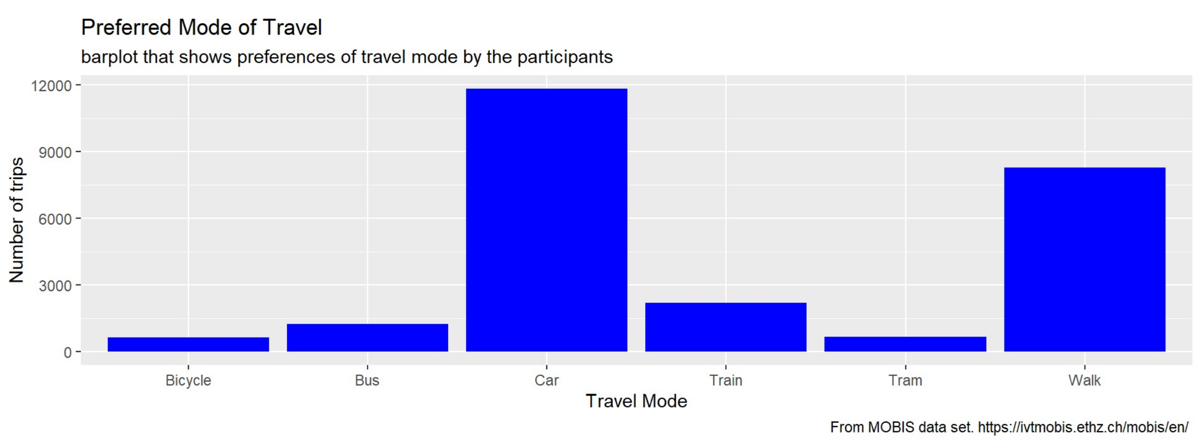
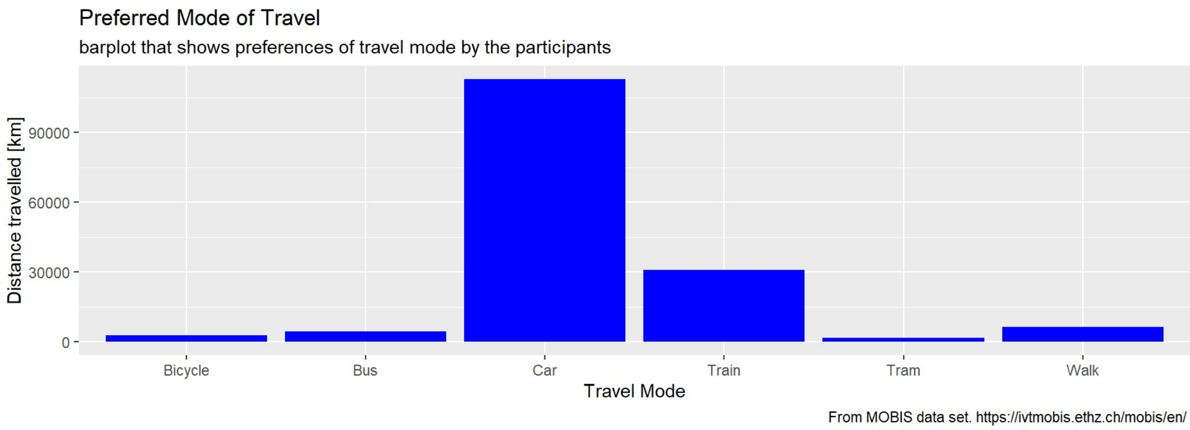


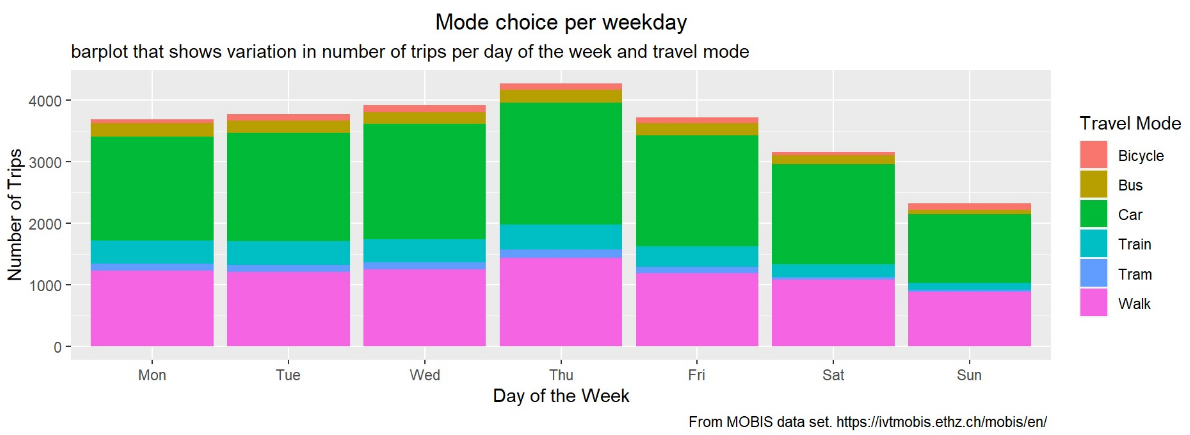
Figure 7: Distance travelled per chosen travel mode



By looking only at Figure 6 it seems that travelling by car and walking are the preferred mode of choice. This is true in terms of the number of trips made, but by looking at Figure 7 one can see clearly how small the total distance travelled by foot is. People make a lot of trips by foot, but they are usually very short and don't compare at all to the distance travelled by car.

The graph in Figure 8 shows how the day of the weekday influences the mode choice.

Figure 8: Number of trips per day of the week per mode choice

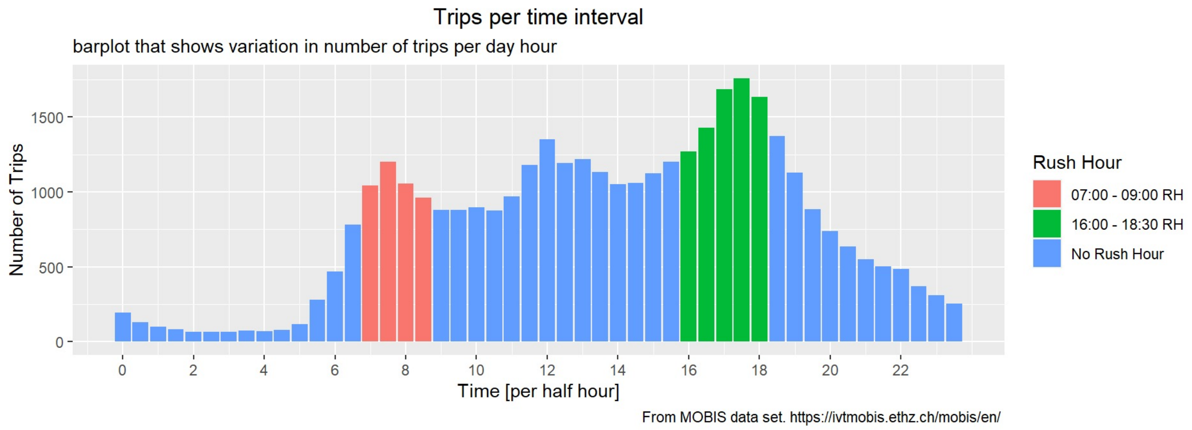


Public transport seems to be slightly less used on weekends, especially on Sundays. This may be due to the more restricted offer of buses and trams during the weekends, which often happens in more rural areas. People also probably prefer to use private transport for leisure to have more flexibility and comfort.

2.2.6 Peak Hour

The number of trips is furthermore strongly influenced by the time of the day.
To analyse this behavior the number of trips for each half-hour-interval of the day was plotted. The result is shown in Figure 9.

Figure 9: Number of trips per half-hour-interval



As expected, rush hours do play an important part in the number of trips generated. Two peaks are present when people usually go to and come back from work. Another important peak can be found around midday, when people go out of their workplaces to eat.
The higher peak on the latest rush hour can be explained by people making additional trips to go shopping or doing other activities before dinner.

2.3 Socio-demographic characterstics

In this section the focus has been on the participants data, to better understand the social characteristics of the dataset. Results were shown in the following subsections.

2.3.1 Basic Informations

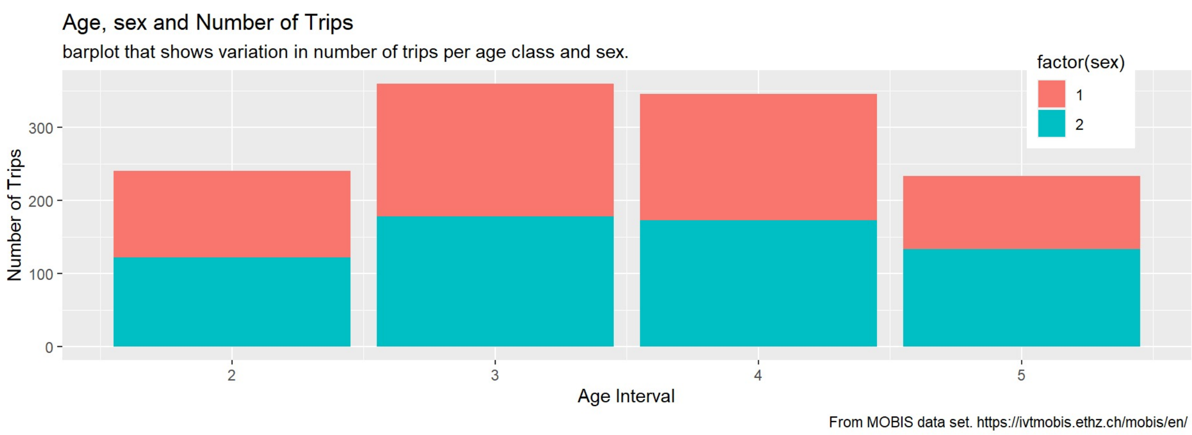
The number of participants is 1’254 and the reported variables are 43. These range from physical characteristics (age, sex), personal information (postcode, household size, education), employment information (income, type of employment, work percentage), travel behavior (means of transport owned, public transport passes, means of transport used) to behavior during the study (number of tracked days).

2.3.2 Age distribution

To analyse age distribution, age intervals that comprised the ages of all participants were chosen. They are defined as follows: interval 2 for people between 10 and 25 years old, 3 for people between 25 and 40, 4 for people between 40 and 55 and 5 for people between 55 and 70.

Being sex sometimes a sensible question, it has to be noted that about 70 participants didn't respond. This data was deleted to be able to build the graph presented in Figure 10.

Figure 10: Age distribution divided by sex



Young people travel less because they don't have the same financial means of people who work, while older people tend to travel less because undertake less activities or are already retired.

With the sex factor 1 = females and 2 = males it can be noticed that between the older participants male are more likely to travel than females. A possible explanation can be found in the discrepancy of number of driver licenses between sexes, which was still present for that generation. Figure 6 highlighted how the number of trips by car dominates over the other and for this reason it does make sense that older women travel less than men.

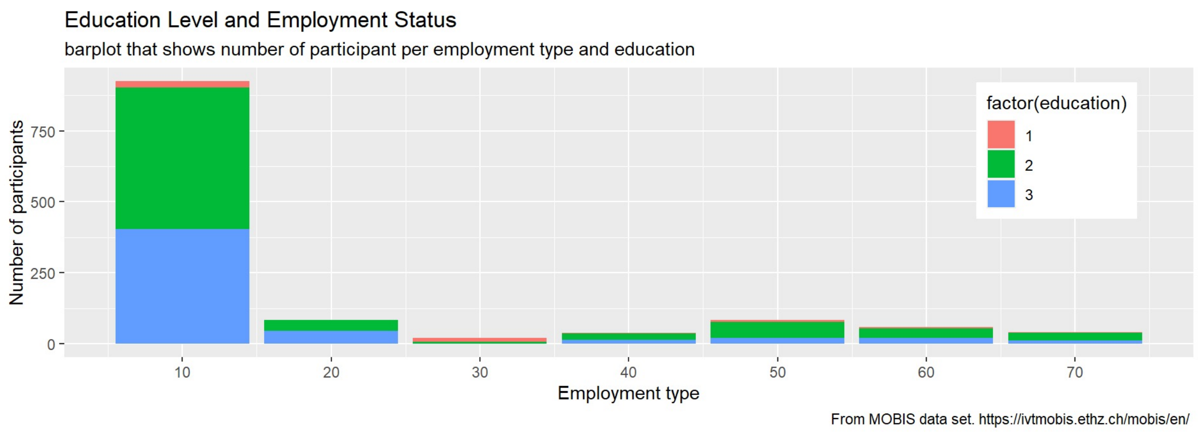
2.3.3 Education

The classification of the education level is in the codebook defined in 3 categories: completed mandatory education for class 1, secondary education (apprenticeship or diploma) for class 2 and higher education (university) for class 3.

The analysis shows that mainly people with education level 2 or 3 participated. Usually people with higher education are more likely to undertake this kind of survey because they have the financial means to travel more and they are more interested in it. This explains why the number of people with low education is so small.

Employment type is defined by MOBIS as following: category 10 for employed, 20 for self-employed, 30 for apprentice, 40 for unemployed, 50 for student, 60 for other, 70 for retired.

Figure 11: Number of participant per employment type and education



The low numbers of unemployed or retired people confirms our interpretation of the data based on working hours. Almost all the participants are regularly commuting from and to work or school.

2.3.4 Income

The aim of this subsection is to assess the influence of income on the number of trips. A linear regression has been carried, showing a weak and negative, not significant relation between number of trips per participant (n variable) and income. A next step has been to add to the model household size as correlation variable. Results of the second model confirmed the results of the first.

Table 2: Linear regression between income and number of trips

	Dependent variable:	
	n	
	(1)	(2)
income	−0.013	−0.015
	(0.015)	(0.015)
household_size		0.719**
		(0.346)
Constant	19.202***	17.108***
	(0.443)	(1.102)
Observations	1,246	1,246
R^2	01	04
Adjusted R^2	−002	02
Residual Std. Error	14.443 (df = 1244)	14.424 (df = 1243)
F Statistic	0.696 (df = 1; 1244)	2.502* (df = 2; 1243)
Note:		*p<0.1; **p<0.05; ***p<0.01

Source: MOBIS data set. https://ivtmobis.ethz.ch/mobis/en/

3 Part 2: Trip production - Regression

In this section, the focus will be on trip production. It will be assumed that the variable "trips per week per person" is a continuous variable. We will investigate on the variables we want to use to explain the trip production based on the participants data. All steps of the modelling process will be described in this report.

3.1 Data set

Assuming to upload the data as if it would have been just downloaded, as done in the precedent part, a first cleaning step and data set creation/modification is needed. From Section 2.1 of the report, the third point named: "Determining of the main leg of the trip" will be repeated. Please note that this step includes the creation of the *trip id* variable, necessary to calculate the main leg per trip id. The first two points of the 2.1 cleaning step are not necessary for our analysis, as our focus will be on the participants data and not on the trips data. After sub-setting the trips data frame, including only the main mode per trip, number of trips per week per participant is joint to the participants data. Also note that at this step the income variable has been cleaned by not significant values. No other significant cleaning steps have been included, as regressions in R automatically take off not available data.

3.2 Trips per week per participant

A first analysis of the dependent variable "trips per week per participant" or "trips w", will underline the presence of outliers in the distribution (fig box-plot simple). Those have been taken off from the data set, as we decided to focus on the statistical behaviour of the participants, to have a more accurate prediction model. In figure 12, it is shown the trips per week per participant distribution including and excluding outliers. In figure 13, it is showed the plot difference between trips w and income and general accessibility, including and excluding outliers.

Figure 12: Box-plot of trips per week per participant

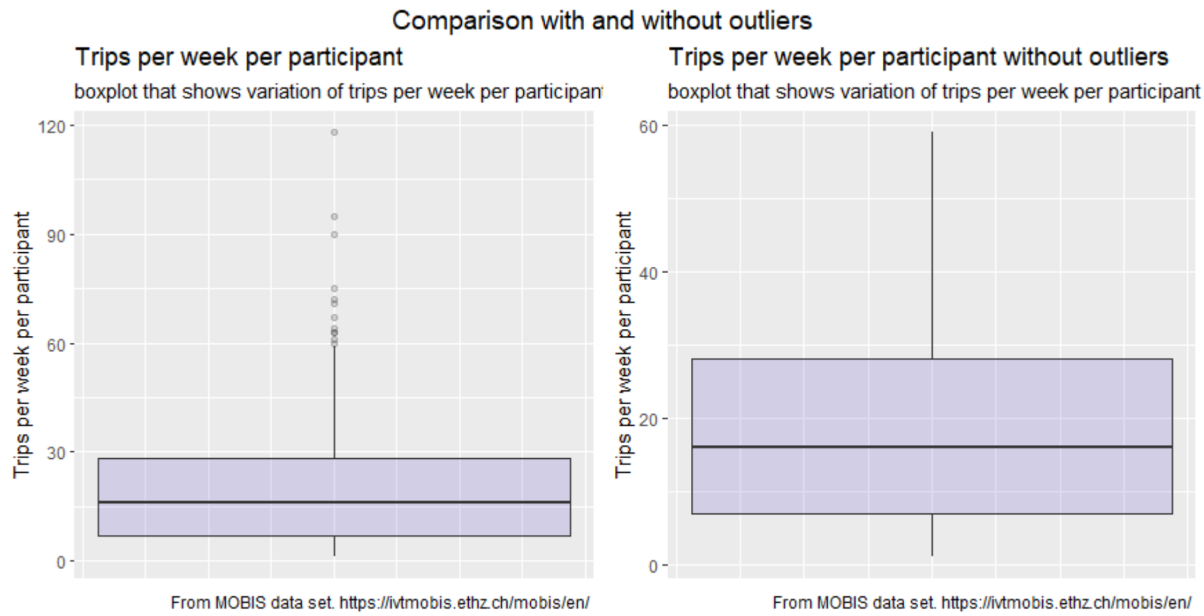
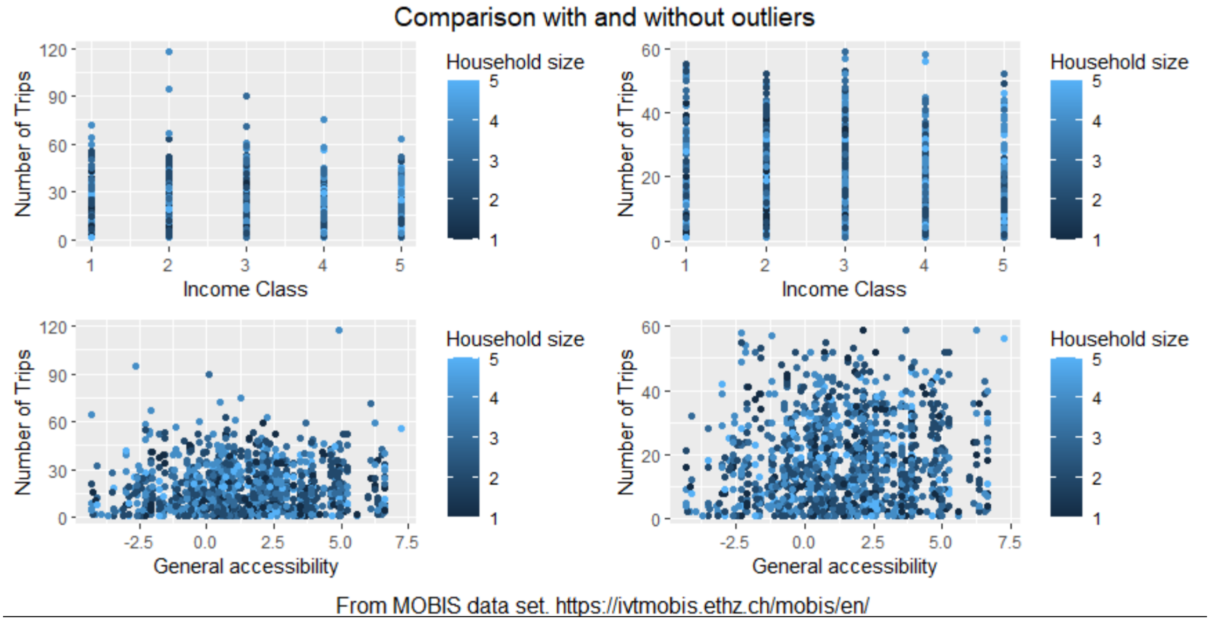


Figure 13: Scatter-plot of trips per week per participant per income and accessibility



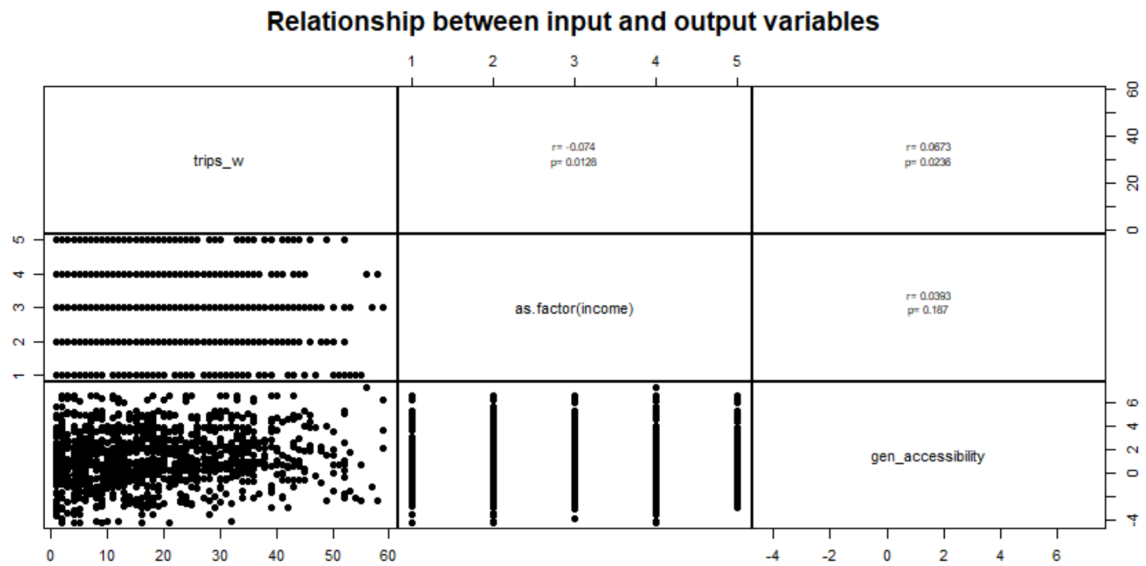
3.3 Regression and Hypothesis Formulation

After numerous meeting and confrontations, a perfect agreement between all the group components has been reached. It has been jointly decided to investigate the relationship that exists between the numbers of trips per week per participants and income and general accessibility. To be noted, other control variables will be included in the analysis, to strengthen the model significance and to compare the different effects of the independent variables on the dependent one, trips per week per participants. Predictors that have been included in the analysis are: income (income classes 1 to 5), general accessibility (continuous variable 0 to 1 values), age (numerical discrete variable), household size (numerical discrete variable), education (discrete variable), sex (binary variable), has GA pass (binary variable), has half fare (binary variable), has regional pass (binary variable), has car (discrete variable), has motorbike (discrete variable), has bike or e-bike (discrete variable), first employment type (discrete variable, comparatively better accessibility by individual transport (continuous variable 0 to 1 values), comparatively better accessibility by public transport (continuous variable 0 to 1 values). Moreover, the it has been checked for location fixed effects using the variable postcode home (ZIP number).

In Figure 14, it will be found a correlation matrix of the variables taken into account, created using the pairs() function in R. From one side, the use of this matrix allows to detect the variables that influence the output variable number of trips the most. From the other side, it is also possible to detect multicollinearity between the input variables.

Both graphical and quantitative information is included in the output matrix.

Figure 14: Relationship between output and input variables



The approach that has been used to strengthen our model has been progressive: initially the model has been ran using an input variable only (income in our case), and next it has been ran adding one category of variables per run. This allowed to check every time the effect of the added variables in the regression. Moreover, after choosing the best regression, thanks to multicollinearity tests, it has been possible to determine the variables with multicollinearity issues. Dummy variables have been created to run the regressions, and also product of variables has been included to eliminate multicollinearity between the predictors. This is better explained in the following sections.

3.4 Construct dummies and transform necessary variables

In this section, it will be advocated the necessity to construct dummy variables for discrete variables. Income, education, and employment variables have been treated as discrete variables. Income is represented by 5 different income classes, so it does not reflect a monetary value but a category. This applies also to education, employment, household size and sex as to all the binary variables, as for example the "has" variables family (for example "has regional pass"). All those has been factorized inside the regression, allowing to treat them as dummies. The age variable have been squared, as it changes the significance of the model enough to justify the transformation. Location fixed effects have been added factorizing the variable "postcode home" inside the regression. In the code, regressions have been ran both with modification and without, to confirm our assumptions.

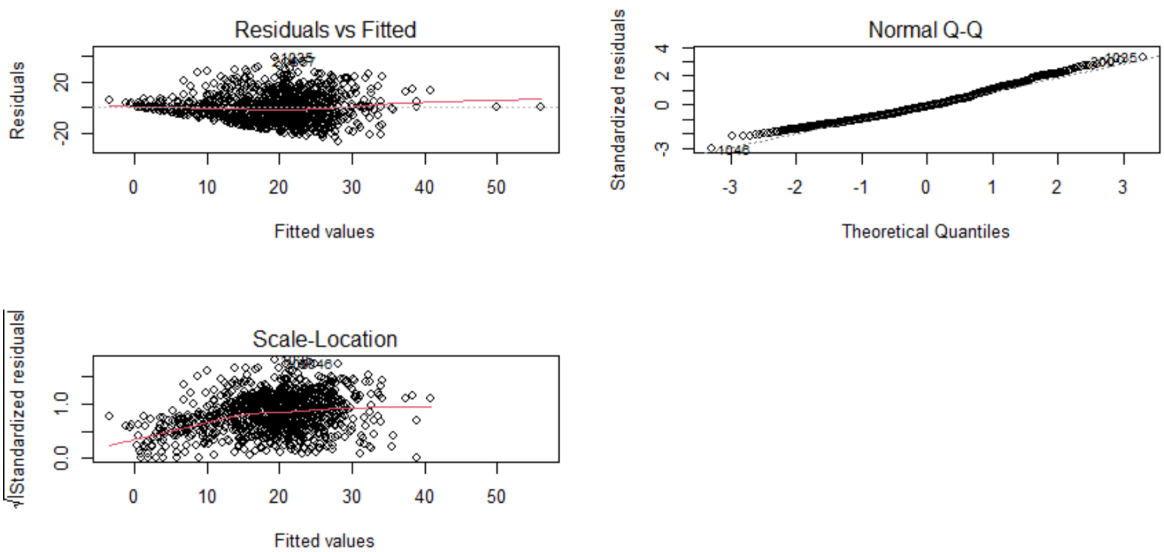
3.5 Linearity, multicollinearity and goodness of fit

Every time after running the various regressions, the plot() function has been used to check the linear relationship assumptions, to examine whether the residuals are normally distributed and to check the homogeneity of variance of the residuals (homoscedasticity). Those assumptions have been graphically checked using respectively the first, the second and the third plot of the plot() function (Figure 15). For the final model we decided to analyse, linearity assumptions could be said to be almost fulfilled. In fact, a not perfectly straight line is present in the first regression plot. A horizontal line, without distinct patterns is an indication for a linear relationship. Looking at the second plot, it can be said that residuals are normally distributed, as residuals points follow the straight dashed line. In the third, horizontal line with equally spread points is a good indication of homoscedasticity. The model seems not to respect completely homoscedasticity assumption, as the line showed in the plot seems more curved than linear.

To check for multicollinearity, the vif() function has been used. If the output included values major than 5, it is present a strong collinearity between the input variables. This does not happen in our model. A step further for the multicollinearity analysis has been to create a Pearson correlation matrix. If output included values bigger than 0.3, multicollinearity between two variables has been detected (weak collinearity). To overcome this problem, control variables created by the product of the variables with collinearity issues have been included in the model . In our regression, general accessibility and comparatively better accessibility by individual transport show a Pearson coefficient of -0.38, that indicates the presence of multicollinearity between the two input variables. Same for general accessibility and comparatively better accessibility by public transport, that shoes a Pearson coefficient of -0.57. Finally, the anova() function has been used to check significance of the discrete variables. This allowed us to remove not significant variables from the model such as age squared (not categorical but lowering adjusted R squared consistently), sex, has pt pass ga, has pt pass halffare, own car, own motorbike own bike or ebike. This final cleaning step takes us to the final regression model.

Considering the final regression, regarding the goodness of fit, it is easy to notice a huge difference in Table 3 between R squared (0.3329) and Adjusted R squared (0.1121). In fact, Adjusted R-squared value can be calculated based on the value of R-squared, the number of independent variables (predictors) and the total sample size. This means that the sample studied is too small to be able to effectively predict the number per trips per week per individual using all the input variables chosen. To solve this, a larger dataset is suggested to be used. To also consider the possibility to use less input variables to reduce the difference between the two Rs. This will not increase the Rs anyway. Trade-off between model fit and significance have to always be considered when running a linear model.

Figure 15: Regression Plots



3.6 Regression recapitulae and conclusions

To conclude, the model has apparently a great significance level, as the p-value is small enough (<0.05). This unfortunately cannot be said for both the input variables we decided to analyse. In fact, looking at Table 3, it is evident that income class has not a great significance level. Instead, general accessibility has been found having a 99% confidence level. But what does that exactly means? Looking as reference the third regression in Table 3, the constant of 28.9654 indicates the number per trips per participants if any of the control variables has value 0. We can notice how a greater level of general accessibility (0 to 1, so a 1% increase), generally reduces the number of trips per week by 1.4580. The results meet our expectations, as with such a small dataset and such an elaboration complexity of social variables, it would be ambitious to have a better model fit. If we would apply our model to a new dataset, we would look first at the new dataset. In fact, from different datasets, different model types could be created. If the new dataset presents the same characteristics of the one we used, preferably with a bigger sample, we would have applied the same steps. Moreover, it would have been possible to estimate the different behaviour between Urban and Rural areas of the Canton. First, we would clean the data and look at the output variable we are interested in. Then we would have qualitatively assessed how different variables are linked between each other, to then run the regression. Then, visual checks would allow us to understand the characteristics of the new dataset. From the other side, in this model the assumption that the output variable has been treated as a continuous variable may have limited the precision of our estimates. If we would have decided to treat it as a discrete variable, it may have been possible for us to use a logit model to have a better regression fit and significance.

Table 3

	Dependent variable:		
	trips_w		
	(1)	(2)	(3)
as.factor(income)2	−1.0843	−1.0213	−0.5772
	(1.7107)	(1.8880)	(1.7813)
as.factor(income)3	−0.4140	−0.4416	−0.3248
	(1.7079)	(1.8565)	(1.7853)
as.factor(income)4	-3.4857^{**}	-3.5159^{*}	−2.7855
	(1.7629)	(1.9088)	(1.8303)
as.factor(income)5	−1.8411	−1.6030	−1.0247
	(1.8624)	(248)	(1.9495)
gen_accessibility	−0.1001	-1.6621^{***}	-1.4580^{***}
		(0.2008)	(0.1825)
I(ageˆ2)	−004	−003	
household_size	0.3685		
as.factor(household_size)2		−0.0682	0.2848
as.factor(household_size)3		3.0506	2.7678
as.factor(household_size)4		0.9983	1.3233
as.factor(household_size)5		0.5786	0.7952
as.factor(education)2	−1.4360	−0.9969	−0.1326
as.factor(education)3	−3.3970	−3.0495	−2.0754
as.factor(sex)2	0.3670	0.4475	
as.factor(has_pt_pass_ga)1	−0.0299	0.1440	
as.factor(has_pt_pass_halffare)1	0.6254	0.7456	
as.factor(has_pt_pass_regional)1	3.5058	3.6907	3.9992
as.factor(own_car)1	−15.6642	−15.8725	
as.factor(own_car)9	−14.9617	−15.2078	
as.factor(own_motorbike)1	−0.4539	−0.4830	
as.factor(own_motorbike)9	2.5496	2.8693	
as.factor(own_bike_ebike)1	0.8594	0.7288	
as.factor(own_bike_ebike)9	0.7884	1.1264	
miv_accessibility	−1.7529	5.2635	3.1621
oev_accessibility	−0.0134	−3.4097	−2.5544
as.factor(employment_1)20	−1.4410	−1.5599	−1.4204
as.factor(employment_1)30	3.6439	3.9650	5.1344
as.factor(employment_1)40	−5.5363	−5.3795	−6.3790
as.factor(employment_1)50	−0.4098	−0.4251	0.1956
as.factor(employment_1)60	−5.2136	−4.3805	−4.8443
as.factor(employment_1)70	−3.3358	−3.0346	−307
gen_accessibility:miv_accessibility		−7.3360	−6.6349
gen_accessibility:oev_accessibility		1.4227	1.1586
Constant	42.1928^{***}	46.8869^{***}	28.9654^{***}
	(1.5474)	(1.7323)	(2.0743)
Observations	1,065	1,065	1,131
R^2	0.3363	0.3455	0.3329
Adjusted R^2	0.0970	0.1038	0.1121
Residual Std. Error	12.7069	12.6595	12.5312
F Statistic	1.4054^{***}	1.4292^{***}	1.5075^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01		

4 Gravity model

The goal of this subsection is to perform the trip distribution for the given data set using a loop or a function in R. This is possible by using the *trips* data, where start and end district are specified for every trip.

A first step consists in calculating the number of trips registered for each pair of zones. The results are reported in the following table:

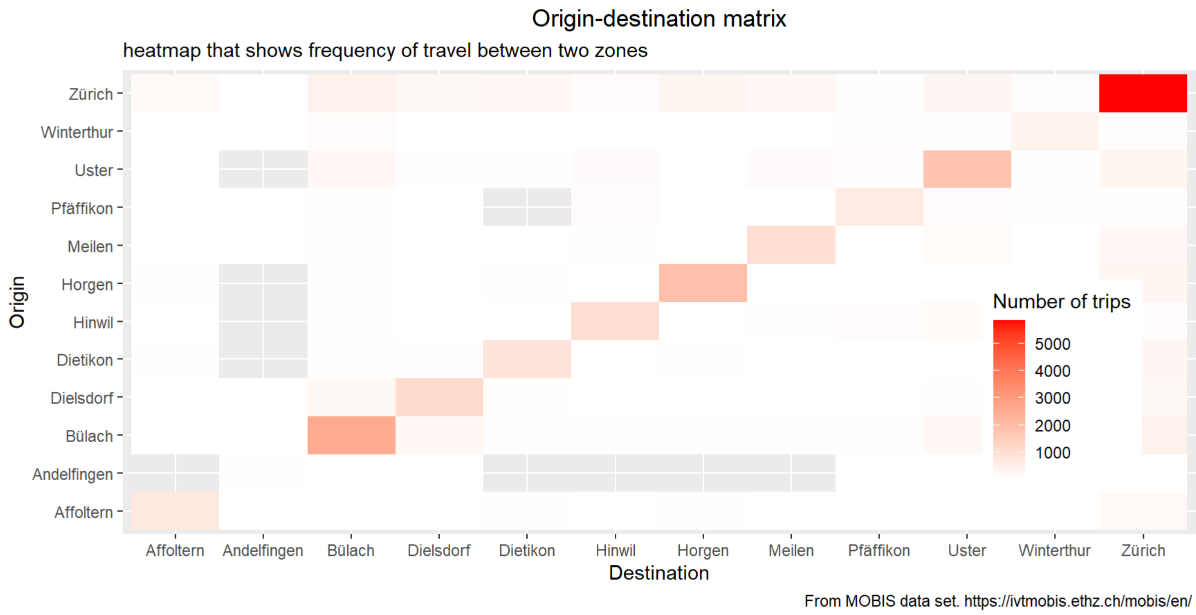
Table 4: Number of trips between each pair of zones

Start district	End district											
	Affoltern	Andelfingen	Bülach	Dielsdorf	Dietikon	Hinwil	Horgen	Meilen	Pfäffikon	Uster	Winterthur	Zürich
Affoltern	716	1	11	1	45	1	35	1	4	3	5	146
Andelfingen	0	32	12	2	0	0	0	0	2	1	2	2
Bülach	5	10	2529	216	26	39	24	29	44	204	93	395
Dielsdorf	1	3	196	1107	41	4	5	11	4	30	8	195
Dietikon	41	0	28	45	839	2	17	8	2	13	10	265
Hinwil	1	0	27	2	3	983	4	50	90	134	10	90
Horgen	31	0	23	5	19	5	1949	15	1	8	4	281
Meilen	2	1	31	3	10	52	15	994	8	106	7	204
Pfäffikon	1	2	43	3	0	88	1	10	626	100	50	49
Uster	3	0	205	31	17	123	8	109	98	1801	18	288
Winterthur	2	6	93	7	7	7	4	6	46	18	378	69
Zürich	161	3	412	200	249	79	282	208	42	278	72	5843

Source: MOBIS data set. https://ivtmobis.ethz.ch/mobis/en/

For a better visualization of this data, a heatmap can be used. On the y axis, the trip origin is represented, while on the x axis the trip destination can be found. The white cells represents the less traveled combinations of zones, while red indicates a higher number of trips.

Figure 16: Heatmap of the origin-destination matrix



The first very noticeable thing is that the majority of trips happen inside the same zone, that is on the diagonal of the matrix. The pair Zürich-Zürich is the one that counts the most number of trips. This doesn't come as a surprise, since it is for sure both the most populated and the most attractive zone of them all.

The next step is to use the Furness method to predict an origin-destination matrix based on the total number of trips attracted and produced for each zone and compare it to the actual data. The impedance function is based on the mean travel time between each pair of zones as the costs and is defined as following:

$f(c) = 1/c^2$

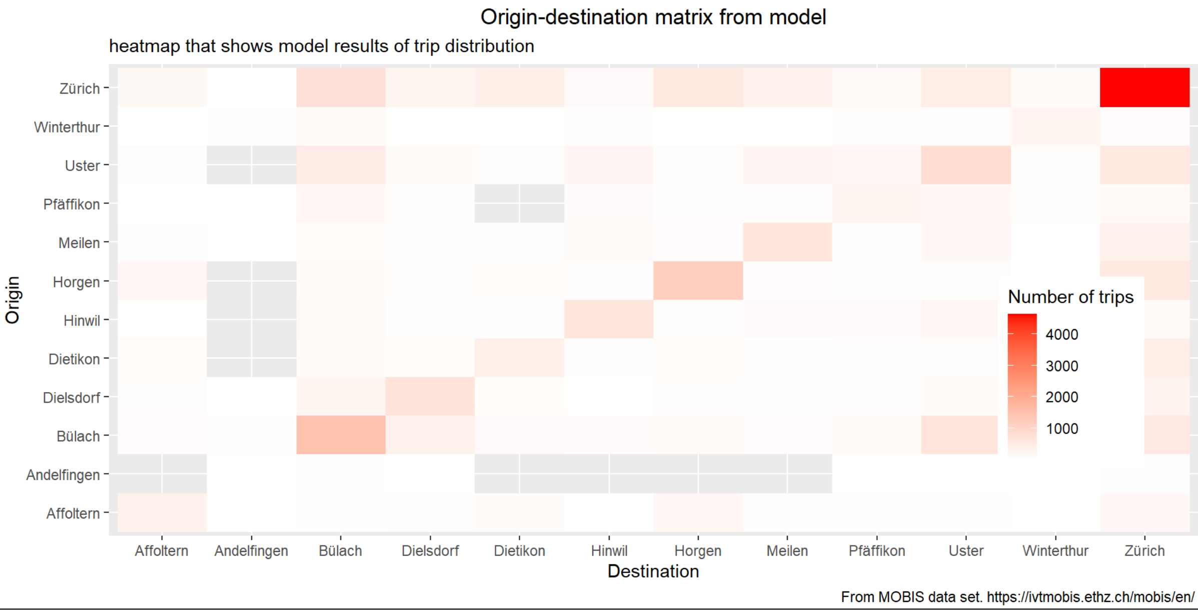
With the help of a *while* loop in R it's possible to compute the modelled origin-destination matrix by correcting attraction and production with the iterative parameters α and β. The loop is set to stop when the error margin:

$$\epsilon = \sum_{\forall j} \frac{|\beta_j^{k+1} - \beta_j^k|}{\beta_j^k}$$

is smaller than 00001.

After the simulation has reached its end, the resulting origin-destination matrix is described with the heatmap below:

Figure 17: Heatmap of the origin-destination matrix obtained with the Furness method



The model result is quite similar to the heatmap of the real trip distribution and confirms the goodness of the Furness method in order to construct the origin-destination matrix.

In the next page, the whole matrix obtained through the model is shown.

Table 5: Number of trips between each pair of zones

Start district	End district										
	Affoltern	Andelfingen	Bülach	Dielsdorf	Dietikon	Hinwil	Horgen	Meilen	Pfäffikon	Uster	Winterthur	Zürich
Affoltern	328	2	41	60	105	5	176	21	21	32	4	167
Andelfingen	0	6	18	1	0	0	0	0	6	3	1	14
Bülach	74	14	1442	326	115	89	109	70	125	630	90	524
Dielsdorf	51	3	243	668	82	13	58	14	39	102	23	297
Dietikon	81	0	101	83	370	19	85	19	34	48	15	410
Hinwil	11	0	118	23	15	617	46	93	93	201	39	133
Horgen	178	0	108	50	85	60	1149	70	29	49	13	545
Meilen	19	1	83	3	15	99	66	586	23	174	9	337
Pfäffikon	8	3	175	30	0	91	38	33	223	165	50	150
Uster	46	0	453	103	57	237	48	210	180	775	48	539
Winterthur	5	15	100	12	11	31	11	8	58	62	250	75
Zürich	158	10	722	247	397	115	552	305	129	449	109	4629

Source: MOBIS data set. https://ivtmobis.ethz.ch/mobis/en/

5 Choice modelling

The goal of this chapter is to find the utility function for the four modes used in the provided Mobis dataset: walk, bike, car and public transport (PT). The general utility formulas will look like this:

$Walk : U_w = \epsilon_w + \beta_{tt,w} * tt_w$
$Bike : U_{bike} = \epsilon_{bike} + \beta_{tt,bike} * tt_{bike}$
$Car : U_{car} = \epsilon_{car} + \beta_{tt,car} * tt_{car} + \beta_{cost} * cost_{car}$
$Public\ transport : U_{PT} = \epsilon_{PT} + \beta_{tt,PT} * tt_{PT} + \beta_{cost,pt} * cost_{pt} + \beta_{dist,PT} * dist, PT$
$+ \beta_{accegr,t,PT} * accegr, t, PT + \beta_{freq,PT} * freq, PT + \beta_{trans,t,PT} * trans, t, PT$
$+ \beta_{trans,nr,pt} * trans, nr, PT$

The ϵ represent the unobserved part of the utility (assumed identically and independently distributed across all alternatives and respondents), while tt indicates the travel time for each mode.

Walking and biking are assumed to be cost-less and thus a cost parameter will not be part of their utilities. Car and public transport are assumed to have different β parameter to describe the cost part of the utility.

The utility function for public transport takes into consideration a couple things more such as the distance routed in km the access and egress time in minutes, the frequency in minutes (i.e. every how many minutes there is a connection), the transfer time in minutes and number of transfers between other PT modes.

An MNL model can then be computed with the help of the R package *Apollo*, which allows to determine the different β and ϵ parameters through the maximization of the log-likelihood. The goal is to find the model which estimates at best the actual choice the people made by knowing if they were available and how attractive they were.

The estimation was successful and the parameters found are summarized in the table below:

Table 6: Results for betas parameter

Parameter	Value
ϵ_{car}	0.0000
ϵ_{w}	0.029590
ϵ_{bike}	-2.271458
ϵ_{PT}	1.188124
$\beta_{tt,w}$	0.0000
$\beta_{tt,bike}$	0.080711
$\beta_{tt,car}$	-0.293538
$\beta_{tt,pt}$	-0.045664
$\beta_{dist,pt}$	0.674602
$\beta_{accegr,t,pt}$	-0.372216
$\beta_{freq,pt}$	-0.02394
$\beta_{trans,t,pt}$	0.199922
$\beta_{trans,nr,pt}$	-3.545287
$\beta_{cost,pt}$	-0.180038
β_{cost}	4.180844

ϵ_{car} and $\beta_{tt,w}$ are estimated with value 0 because they chosen to be the fixed constant at the start of the modelling process. This means that the other ϵ and β are relative to those.

The negative coefficients of β indicate the decreased utility associated with an increase of the respective variable relative to $\beta_{tt,w}$. The positive β related to the car costs can be explained by the fact that it is related to the cost of the car in CHF and that people who are willing to pay a lot for a car they are then more prone to use it instead of the alternatives. Other than that, the signs meet our expectations: longer travel time generally impact negatively the utility, as well as egress/regress times and number of PT changes.

An important value that defines how good the model is is the goodness of fit:

$\rho = 1 - LL(\hat{\beta})/LL(0)$

ρ compares the log-likelihood of the model to the one obtained with no data. For the model presented here, the goodness of fit equals to about 0.37, which is not very high but is to be expected with such a relatively small sample of observation. This will be better discussed in the conclusions section.

The model allows then a prediction in percents of the chosen mode. These probabilities are summarized in the table below.

Table 7: Predicted choice probability for each mode

Car	PT	Bike	Walk	Chosen
Min. :0.06712	Min. :0000	Min. :0000	Min. :0000	Min. :000594
1st Qu.:0.42887	1st Qu.:0.01791	1st Qu.:0.02840	1st Qu.:0.09932	1st Qu.:0.3127092
Median :0.61544	Median :0.08865	Median :0.04147	Median :0.17833	Median :0.5676281
Mean :0.59549	Mean :0.17413	Mean :0.04248	Mean :0.18790	Mean :0.5297440
3rd Qu.:0.77619	3rd Qu.:0.25817	3rd Qu.:0.05422	3rd Qu.:0.26809	3rd Qu.:0.7551533
Max. :100000	Max. :0.91902	Max. :0.16312	Max. :0.54694	Max. :100000

With a mean of almost 60%, our model predicts that cars will be the most chosen mean of transport. The other alternatives would then be chosen with a percentage of under 20%, with biking being the less attractive and chosen only on 4% of the times. This looks likely to happen also in reality and confirms this sanity check on our model.

Next, a brief discussion about value of travel time (VOT/VTT) is conducted. This important parameter can be defined as the price people are willing to pay to acquire an additional unit of time for a determined mean of transport. The formula used to calculate it is the following:

$$VOT = \beta_{tt}/\beta_{cost}$$

The parameters used are those found through the *Apollo* simulation. β_{tt} is specific for every mean of transport, while for β_{cost} the two cost-specific variables for cars and public transport are used. In the table below, the different VOT for the different combinations of β are calculated.

Table 8: Results for VOT

VOT	Value [CHF/min]
$\beta_{tt,car}/\beta_{cost}$	-0.07021
$\beta_{tt,bike}/\beta_{cost}$	-0.01930
$\beta_{tt,pt}/\beta_{cost}$	-0.01092
$\beta_{tt,pt}/\beta_{cost,pt}$	0.25360
$\beta_{tt,car}/\beta_{cost,pt}$	1.63000
$\beta_{tt,bike}/\beta_{cost,pt}$	-0.44830

People are ready to spend about 0.25 CHF to gain a minute while using public transport, while they are not more prone to buy more expensive cars to save on travel time.

The last step is to put together the findings from the MNL and Gravity models. This allows the generation of the table below, which indicates on average how many trips are undertaken with which mode of transport for each pair of zones according to the model. The probabilities come from the MNL model and are further tied to the specific origin and destination. The results from the Gravity model allow to know the trip distribution for each pair, which are then multiplied with the respective probabilities.

Origin	Destination	Average car trips	Average PT trips	Average bike trips	Average walk trips
Zürich	Zürich	1963	1264	237	1163
Meilen	Zürich	164	108	13	50
Uster	Uster	505	94	33	141
Hinwil	Hinwil	450	58	21	86
Zürich	Horgen	319	172	12	46
Horgen	Horgen	811	104	44	188
Horgen	Zürich	239	190	22	93
Zürich	Dietikon	300	45	12	39
Dietikon	Zürich	262	89	12	44
Bülach	Bülach	914	209	59	258
Dietikon	Dietikon	235	75	11	47
Pfäffikon	Pfäffikon	149	31	7	33
Winterthur	Winterthur	138	37	11	62
Hinwil	Zürich	62	64	1	5
Bülach	Zürich	321	86	23	92
Zürich	Uster	257	100	17	74
Uster	Zürich	230	198	26	83
Affoltern	Affoltern	247	27	11	41
Dielsdorf	Dielsdorf	487	42	27	110
Bülach	Uster	402	110	25	91
Uster	Bülach	282	92	17	60
Meilen	Meilen	388	71	23	101
Zürich	Meilen	148	88	11	57
Zürich	Bülach	486	75	35	124
Meilen	Uster	136	6	6	24
Winterthur	Pfäffikon	53	1	0	2
Pfäffikon	Winterthur	39	6	1	2
Dielsdorf	Bülach	199	6	8	27
Bülach	Dielsdorf	270	8	10	36
Dielsdorf	Zürich	192	63	10	30
Zürich	Dielsdorf	201	18	6	21
Affoltern	Bülach	21	4	3	11
Affoltern	Zürich	95	36	5	28
Pfäffikon	Bülach	105	29	6	33
Bülach	Pfäffikon	17	98	1	8
Zürich	Hinwil	42	67	1	3
Meilen	Hinwil	75	11	3	9
Zürich	Winterthur	44	40	4	19
Bülach	Horgen	30	69	1	8
Uster	Pfäffikon	155	13	2	7
Uster	Hinwil	165	53	4	12
Uster	Meilen	149	27	6	26
Hinwil	Meilen	64	13	3	11
Zürich	Pfäffikon	20	64	9	33
Hinwil	Uster	155	32	3	10
Zürich	Affoltern	64	81	2	9
Horgen	Affoltern	138	27	3	8
Andelfingen	Andelfingen	4	0	0	0

6 Part 3: Cost benefit analysis (CBA)

This last section focuses on the cost benefit analysis of a potential policy that influences the mobility of the city of Zurich. In this last section the introduction of shared autonomous vehicles in the district of Zurich will be taken into consideration. It is assumed that this policy will cause an increase by 30% of the travel time for all the cars that enter or leave the zone 112.

The structure of this analysis will follow the Swiss standard. Cost and benefits of this potential policy will be calculated and compared thanks to some assumptions that will be further explained. The result will show whether this project could actually be advantageous if implemented.

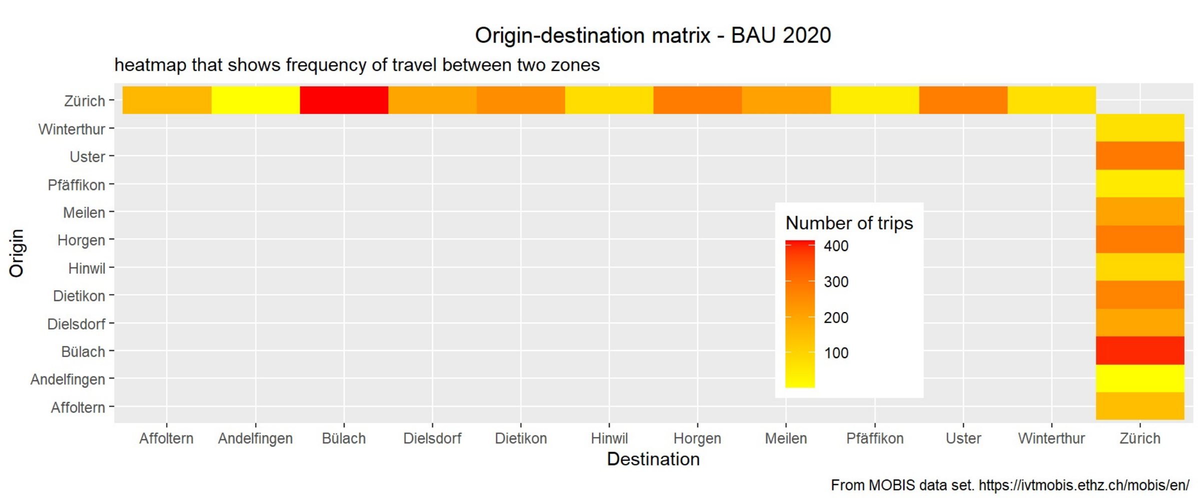
6.1 Project definition

The study area is the border of the city of Zürich (number 112). Only the trips that go in and out will be considered. All the trips that happen in the city and all the trips between the other zones will not be in the scope of our research.

The study period defined in the norms is 40 years. This means that our models and calculations will be made for the year 2060. To forecast this future number of trips, we incorporated the information from the *Kommunaler Richtplan* of the city of Zürich, where the population for the year 2040 has been forecast. The growth rate is assumed to be constant until 2060 and equal for all the zones of the canton of Zürich. For our model we assume that the number of trips growth reflects the population growth rate.

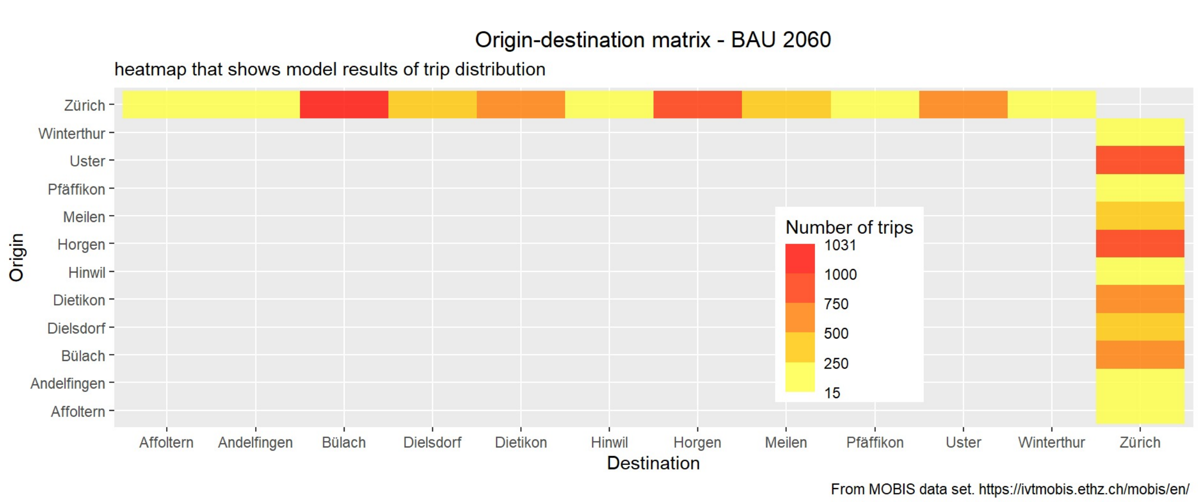
The reference case is the MOBIS dataset filtered. In the figure below is shown the origin-destination matrix representing the scope of our research.

Figure 18: Heatmap of the origin-destination matrix from the MOBIS dataset (2019)



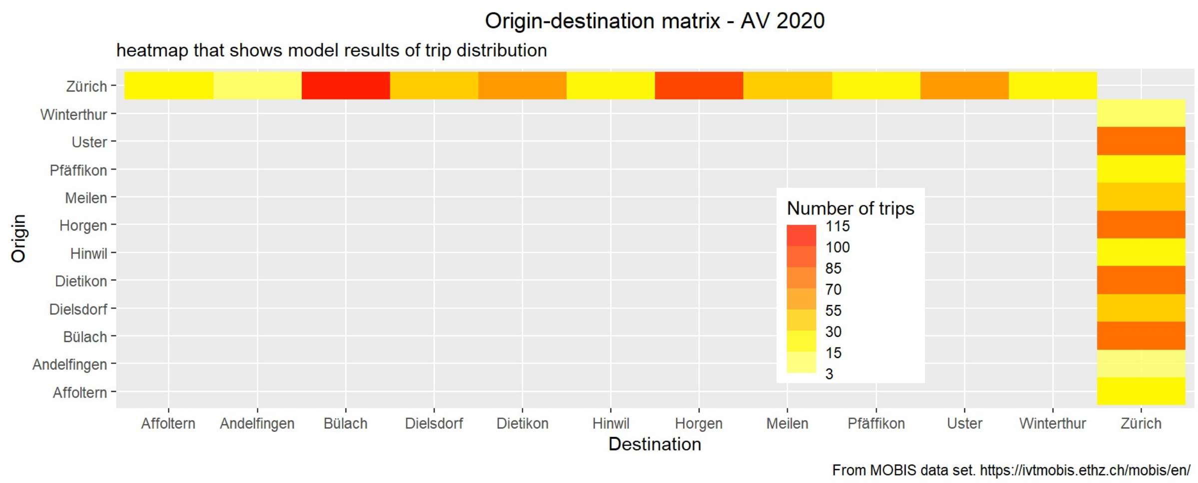
The origin-destination matrix for 2060 is calculated through a Furness method loop similar to that used in the previous assignment. The only difference is the increase in the production and attraction for each zone. The mean travel times are assumed to be the same as for 2020.

Figure 19: Heatmap of the origin-destination matrix for the forecasted trips in 2060



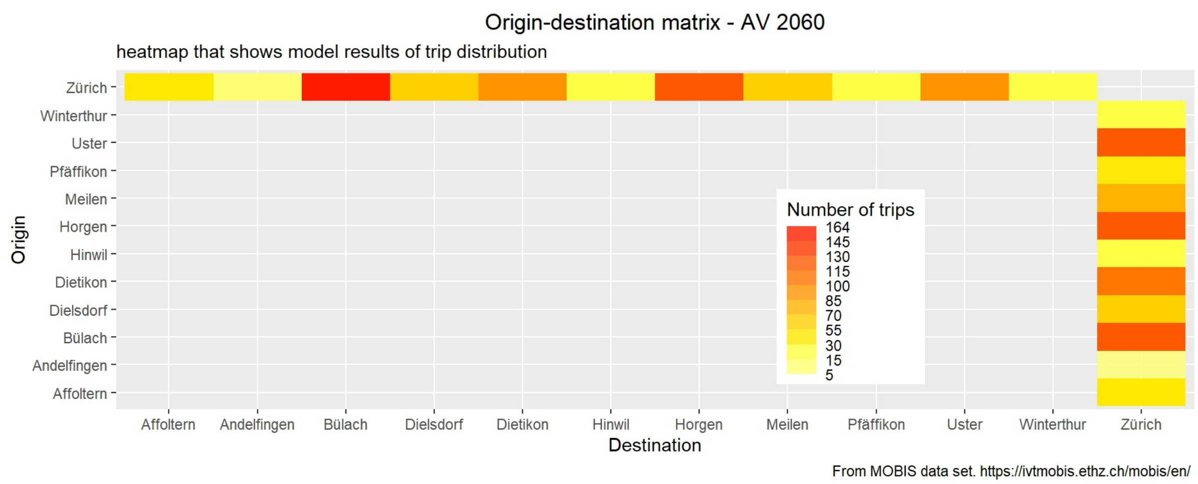
It is important to note that these number of trips refer to all the trips and not just car trips.

Figure 20: Heatmap of the origin-destination matrix for the forecasted trips in 2020 with AV



It can be noted right away that the number of trips decreases with the introduction of this policy. This can be explained by the fact that the increased travel time for only the zone 112 pushes the inhabitants of canton Zürich to travel elsewhere. In our model we did not constraint the people to make the same trip and thus it seems reasonable that people will avoid to travel from and to zone 112.

Figure 21: Heatmap of the origin-destination matrix for the forecasted trips in 2060 with AV



The trips in 2060 will obviously increase because of the bigger population, but will not be as many as the BAU (Business as usual) 2060 because of the increased travel times for zone 112.

6.2 Determination of the indicators set

The indicators considered for our CBA are the following:

6.2.1 Construction costs

The construction costs refer to the costs of the shared autonomous vehicle fleet bought as initial investment. It is assumed that a certain number of shared autonomous vehicles will be bought on the first year and for each year new vehicles are bought in order to accommodate the population growth.

In the AV scenario private ownership is still considered (Bösch *et al.*, 2017). This creates two kinds of AV users. PAV (Private AV) and SAV (Shared AV). For sake of simplicity, the share of PAV will reflect the share of CAR ownership, while PT users will use the SAV option.

6.2.2 Travel time savings

Travel time savings are the change in travel time for the project variant in comparison to the reference case multiplied for the minimal number of trips in our scope area between the reference case and our scenario.

6.2.3 Net benefit of induced traffic

The net benefit of induced traffic reflects the decrease of the operating costs that the introduction of the shared AV causes. This will influence the operating cost rate itself and the net benefit as a whole because of the fewer total trips (and thus also fewer km travelled).

6.2.4 Accidents and accident cost rates

A further indicator that we consider important for a CBA of a policy involving autonomous cars is the decreased number of accidents. This is arguably one of the main benefit of introducing autonomous vehicles and thus should not be forgotten in our analysis.

6.3 Capturing the impacts (Performance indicators)

6.3.1 Construction Costs

In order to calculate how many shared autonomous vehicles are necessary to successfully substitute public transport, the Apollo model is run again. The utility functions remain the same, but the model is reran with the following modifications to the mode choice dataset: both the travel times for autonomous private cars and autonomous shared cars increase by 30% for every trip that enters or exits zone 112. It has been estimated that the acquisition cost of an autonomous vehicle, compared to a not autonomous one, would increase by an average of 20%, due to the necessary technology. In our model, this has been reflected with an increase of the car trip cost of 20%. The costs for this new type of public transport decrease and is the 45% of the non autonomous ones (Bösch *et al.*, 2018). Additionally, access and egress time is not considered anymore, since it is assumed that anyone will be able to call up a free shared autonomous directly where they need it. The transfer time and the number of transfers is also not taken into consideration anymore, since every vehicle operates from the origin to the destination without stopping. The distance remains the same and so does the frequency, since it is assumed that there is not one shared autonomous vehicle per passenger and sometimes people will have to wait for a free one.

The new predictions of the mode choice is shown in the table below:

Car	PT	Bike	Walk	Chosen
Min. :0.01448	Min. :0.0000	Min. :0.00000	Min. :0.00000	Min. :0.0001842
1st Qu.:0.25510	1st Qu.:0.4379	1st Qu.:0.01056	1st Qu.:0.03330	1st Qu.:0.1764391
Median :0.35299	Median :0.5363	Median :0.01655	Median :0.06933	Median :0.3735094
Mean :0.38854	Mean :0.5194	Mean :0.01627	Mean :0.07576	Mean :0.3776316
3rd Qu.:0.49450	3rd Qu.:0.6169	3rd Qu.:0.02146	3rd Qu.:0.11497	3rd Qu.:0.5507186
Max. :1.00000	Max. :0.9840	Max. :0.04959	Max. :0.21283	Max. :1.0000000

The choice of public transport increases from 17% to 52%, while the choice of car decreases from 60% to 39%. This is expected since the car costs increase and the public transport costs decrease.

This information allows us to calculate the number of trips with public transport. Summing all these trips allow us to estimate the capacity needed to provide this public service. We assume that no more buses are present and that the new fleet will be composed by midsize autonomous vehicles (i.e. VW Golf). We assume that one autonomous vehicle is provided every two trips to keep the fleet dimension realistic. We counted 1790 PT trips, therefore the provided fleet will be of 895.

An autonomous vehicle will cost 20% more than a non-autonomous one (IHS, 2014). The cost of a non-autonomous VW Golf is 35'000 CHF (Bösch *et al.*, 2018), which leads to an autonomous vehicle costing 42'000 CHF. With this data the total construction costs (i.e. the total sum of the autonomous cars purchase) are 37.6 Mio. CHF.

The fleet has to be continuously expanded to accommodate the population growth, so the cost of the infrastructure in 2060 will have increased. With the recalculation of the number of trips for the AV 2060 case and the predictions of the mode choice it's possible to estimate the PT trips in 2060 and thus the fleet size. Again, one person every two AV is considered and the 2060 fleet size should amount to 1279 vehicles. Assuming that the AV price stays the same, this translated into 57.6 Mio. CHF.

6.3.2 Travel time savings

The travel time savings are calculated according to the swiss norms formula:

$$RG = \sum_k B_k \sum_i \sum_j F_{i,j,k} \Delta t_{i,j,k} [h]$$

With:
$F_{i,j,k}$ Minimum $F^0_{i,j,k}, F^P_{i,j,k}$
$F^0_{i,j,k}$ Number of vehicle trips in the reference case 0 for the relation from i to j done by vehicle type k
$F^P_{i,j,k}$ Number of vehicle trips in the reference case P for the relation from i to j done by vehicle type k
B_k Occupancy rate of the vehicle type k
$\Delta t_{i,j,k}$ Change in travel time for the project P variant in comparison to the reference case 0 for the relation from i to j for the vehicle type k

For sake of simplicity we did not calculated each sum for each pair of zone but we took into consideration the mean for the whole set. $\Delta t_{i,j,k}$ is calculated as the total mean difference in travel time between the scenario with and without the policy. The occupancy rate is always considered to be one person per vehicle, either for public and for private transport.

Table 9: Travel time savings costs 2020

Vehicle type	$F^0_{i,j,k}$	$F^P_{i,j,k}$	$\Delta t_{i,j,k}$[h]	RG[h]
Private vehicles	2149	1790	0.15	268.5
Shared vehicles	1684	1342	0.11	147.62

Source: SN 641 822a (2009)

The results are calculated for the 2060 scenario, where the number of trips increases due to population growth, while the travel times are assumed to be the same as their respective 2020 scenarios, since increased travel times due to congestion are not taken into consideration in our model. The choice probability for the reference case is considered the same as 2020.

Table 10: Travel time savings costs 2060

Vehicle type	$F^0_{i,j,k}$	$F^P_{i,j,k}$	$\Delta t_{i,j,k}$[h]	RG[h]
Private vehicles	4872	1919	0.15	287.85
Shared vehicles	1380	2559	0.11	151.8

Source: SN 641 822a (2009)

The total RG are then negative (i.e. a cost) because of the increased travel times ($\Delta t_{i,j,k}$ is positive) and amount to 416.12 h for 2020 and 439.65 h for 2060.

6.3.3 Net benefit of induced traffic

The net benefit of induced traffic are calculated according to the swiss norms formula:

$$NNM = 0,5 \sum_k \sum_i \sum_j |\Delta F_{i,j,k}| \Delta BK_{i,j,k} + 0,5 \sum_k B_k \sum_i \sum_j |\Delta F_{i,j,k}| \Delta ZK_{i,j,k} [MCHF]$$

With:
$|\Delta F_{i,j,k}|$ Absolute difference in the number of vehicle trips in the project variant P compared to the reference case 0 for the relation from i to j by vehicle type k
$\Delta BK_{i,j,k}$ Changed vehicle operating costs (including fuel taxes and tolls) in the project variant P compared to the reference case 0 on a path from i to j by vehicle type k
B_k Occupancy rate of the vehicle type k
$\Delta ZK_{i,j,k}$ Changed hourly costs (travel time and reliability) in the project variant P compared to the reference case 0 on a path from i to j by vehicle type k

The second part of the formula will be neglected in this analysis, since the changed hourly costs for travel time and reliability are not known and thus considered equal to zero. The occupancy is always considered equal to one person per vehicle, $|\Delta F_{i,j,k}|$ is the trip difference with or without policy obtained with Tables 9 and 10.

The difference in operating costs is estimated calculating the differences between the operating costs for regional non-autonomous buses, 0.89 CHF/km*p and regional autonomous midsize cars, 0.31 CHF/km*p (Bösch *et al.*, 2018). This amounts to 0.58 CHF/km*p. With an occupancy of one, all these costs can be simply expressed with CHF/km.

As Furness did not model the increase in travelled km but only the trips, we assumed that they increased and decreased by the same rate. The total travelled length for the BAU 2020 scenario was 25'635 km, for the 2020 AV scenario is 23'072 km (due to a trip decrease of 10%), for the BAU 2060 is 54'346.2 km (trips increase of 112%)and for the AV 2060 39'129.26 km (trips increase of 28%). The length difference will then be multiplied by the changed vehicle operating costs to obtain $\Delta BK_{i,j,k}$.

Table 11: Net benefit of induced traffic

Scenario	$\lvert\Delta F_{i,j,k}\rvert$	Changed vehicle operating costs [CHF/km]	$\Delta BK_{i,j,k}$ [CHF]	NNM [Mio. CHF]
2020	391	0.58	1'486.83	0.3
2060	3199	0.58	8'825.86	14.1

Source: SN 641 822a (2009)

6.3.4 Accidents and accident cost rates

In Bosch *et al.*, 2018 it is assumed that safer driving due to the use of autonomous vehicles decreases the accident rates by 40%, as reported by Tesla.

The actual accident rate on main roads out of town (our analysis does not take into consideration the inner city of Zürich) for accidents, injured and dead is given in the norms and presented in the table below with the decrease that the introduction of shared autonomous vehicles should cause.

Table 12: Actual accident rates and their decrease 2020

Accident, Injury, Fatality	Unit	Rate	Decrease in rate with AV
Accident Rate	Accidents per 10^6 vehkm	0.85	0.34
Injury Rate	Injuries per 10^8 vehkm	49.84	19.94
Fatality Rate	Fatalities per 10^8 vehkm	1.59	0.64

Source: SN 641 824 (2013)

For this section, bike and walk trips will be neglected.

For the unit it is necessary to calculate the total amount of km travelled by PT and car. This was already calculated for the net benefit of induced traffic. The number of vehicles is the number of trips by car plus the number of vehicles in our shared autonomous fleet. This amounts to 2'685 vehicles for 2020 and 3'198 vehicles for 2060.

6.4 Evaluation of monetary impacts (Reference Values)

6.4.1 Construction costs

These costs are already calculated as a monetary impact because of their nature. In the previous section they have already been estimated with 37.6 Mio. CHF for 2020 and 57.6 Mio. CHF for 2060.

6.4.2 Travel time savings

According to the norms, the monetization of the travel time savings depends on the knowledge of the trip purpose and length distribution of the trips. To simplify things we use Table 2 of the Norm SN 641 822a and consider the travel time costs for all-purpose trips. The results are summarized in the table below:

Table 13: Travel time savings costs

Type of vehicle	In-vehicle travel time costs [CHF/h]	Total increase in travel time costs [CHF] 2020	Total increase in travel time costs [CHF] 2060
Motorized individual traffic	23.29	6'253	6'704
Public transport	14.43	2'130	2'190

Source: SN 641 822a (2009)

The total travel time savings costs amount to 8'383 CHF for 2020 and 8'894 CHF for 2060.

6.4.3 Net benefit of induced traffic

The formula in the norms the net benefits of induced traffic are already expressed in CHF, so the results presented here have already been calculated and amount to 0.3 Mio. for 2020 and CHF 14.1 Mio. CHF for 2060.

6.4.4 Accidents and accident cost rates

In the norms the costs of each accident, fatality and injured is presented. In the previous section we calculated the decrease in accident rates if AV were to be introduced. By multiplying the two data the monetization of this benefit can be calculated.

For the accidents and injured costs the average costs were taken into consideration.

Table 14: Accidents and accident cost rates 2020

Accident, Injury, Fatality	Economic Accident Costs [CHF]	Cost decrease [Mio. CHF]
Accidents	103'500	2.18
Injured	89'900	1.11
Fatalities	3'191'400	1.27

Source: SN 641 824 (2013)

Table 15: Accidents and accident cost rates 2060

Accident, Injury, Fatality	Economic Accident Costs [CHF]	Cost decrease [Mio. CHF]
Accidents	103'500	4.40
Injured	89'900	2.24
Fatalities	3'191'400	2.56

Source: SN 641 824 (2013)

The total benefit in accident costs for the introduction of autonomous vehicles is the sum of the three row and amounts to 4.56 Mio. CHF for 2020 and 9.2 Mio for 2060.

6.5 Profitability calculation

According to the norms, all costs and benefits should be discounted to the same reference date in order to make them comparable. The reason for this is that revenues and expenses occurring at different times have different values. To do this, a discount rate is estimated and used. The norms perform this calculation based on the annual mortality rate, the future consumption growth rate and the marginal utility elasticity and reached a result of a discount rate of 2%. It should be noted that the used norm dates back to 2006 and could for this reason be dated. The sensitivity analysis will consider a fluctuation of this value and show its impact.

6.5.1 Discount Rate

Individual inter temporal preferences are influenced by subjective factors. Social discount rate is defined as a normative concept: it gives a price to time and can be interpreted as the minimum required rate of return from a socially desirable project. In other words, the social discount rate is the rate at which the society is willing to trade present versus future consumption, and, as a result, it is not affected by individual characteristics. It developed countries usually the discount rate is low (about 5%), while for developing countries is usually higher (generally more than 10%). From the Swiss Norms, in this analysis it has been used a discount rate of the 2%. This allow us to calculate the Net Present Value (NPV) of the project, that amount -33.82 MCHF.

6.5.2 Decision Criteria

From the cost-benefit analysis perspective (i.e. neglecting the non-monetary indicators), essentially a recommendation should be made for the implementation of a project (or project variant) where the discounted benefits exceed the discounted costs and, consequently the net present value is positive. To classify further various favorable projects or variants, the cost-benefit ratio as a profitability assessment should be used. In fact, the corresponding ranking shows the most profitable projects, given a limited budget. For our analysis, a cost-benefit ratio (CBR) of 0.34 has been determined. This make the project unattractive from a rational point of view.

6.6 Sensitivity analysis

The sensitivity analysis is a useful toll to examine the critical components of the CBA, and to assess the robustness of the results. According to the Swiss Norms, the indicators that should undergo this analysis are the following: Discount Rate, Construction Costs and the Value of Travel Time. For our analysis, the following changes have been applied: 0% 3% and 5% as new discount rates, 10% for the construction costs, and 10% for the value of travel time.

Construction Costs are the components contributing the most in our analysis. In fact, the highest and lowest NPV and CBR have been determined adding and subtracting 10% from the original calculated value. Those values are -38.95 MCHF for the NPV and -0.94 for the CBR when adding a 10%, and -28.69 MCHF for the NPV and -0.85 for the CBR when dimishing the original value of a 10%.

6.7 Presentation and interpretation of results

Table 16: Results CBA

Project: Shared autonomous Fleet	Initial investment costs: 37.6 Mio. CHF	
Description: Introduction of autonomous vehicles in canton Zürich		
Indicator	Present Values in MCHF	
	Costs	Benefits
Calculation of Net Present Values		
Construction Costs (Initial Investment Cost)	37.6	
Construction Costs (Vehicle acquisition through years)	13.7	
Travel Time Savings	0.009	
Net Benefit of Induced Traffic		9.7
Accident Savings		7.7
Total	51.3	17.5
Net Present Value	-33.8	

The net present value is negative, this means that the costs of this project largely surpass the benefits and thus it will not recommended.

It is to be noted that in the CBA we only included a handful of possible indicators. This is only a rough estimate of what the real results are and thus should be evaluated carefully.

7 Summary and Conclusion

This is the concluding section of the report. Here it can be found the conclusions of the descriptive analysis (Assignment 1), of the Regression, the gravity model and the choice modelling (Assignment 2), and, as soon as it will be added, of the CBA of the selected policy intervention (Assignment 3).

7.1 Descriptive analysis

By analysing the distribution of number of trips per day of the week, it could be generally said that people travel more during the working week compared than in the weekend. Moreover, it has been observed that people travel less on Sundays. It has also been noticed a slightly higher average number of trips per week for the people of a lower income class. This has been confirmed by additionally looking at household size. While income and household size seems not to be related, participants from lower-middle income classes seem to travel more than the ones from high income classes. Car and walking seem to be the preferred modes for number of trips, while car and train are the more prominent by distance. Public transport was overall less used during the weekends, probably due to less flexibility and comfort in comparison to an own vehicle. Finally, rush hours also do play an important role in the number of trips generated, creating noticeable number of trips peaks.

Then, a brief analysis on the social characteristics of the participants was performed. It's observed a discrepancy in travel behaviour between old males and females, probably due to the difference in the distribution of driver licences between sexes in the past. Generally, people from 25 to 55 have been observed to travel the most. It has also been underlined how people with higher level of education are more likely to answer the surveys, creating sampling biases to the data set. Then, analysing the employment status of the respondents, it has noticed how most of the participants were regularly commuting from and to work or school during the analysis. Finally, it has been created a linear regression model to assess how income influence travel behaviour. Results shows a weak and not significant correlation. This has been confirmed by adding household size as control variable. To sum up, it could be learned from this analysis than when having social data, analysis has to be carried with major attention, regarding the meaning of the variables themselves and of the result shown.

7.2 Trip production: Regression and models

Regarding the linear regression, the model has apparently a great significance level, as the p-value is small enough (<0.05). General accessibility has been found having a 99% confidence level and a negative coefficient, that underlines how the more accessibility one person in canton Zurich has, the less trips they are expected to perform in a week. A bigger dataset is highly desirable to have more significance and a better Adjusted R-squared, that in the final model presented showed how our model effectively explains only approximately 11% of the cases.

The model obtained through the Furness method seems to well reproduce the real trip distribution in the reality. The number of trips that stay inside of one zone are maybe slightly underestimated in favor of more trips between different zones, but the numbers are relatively good compared to reality. Of course, in this case some information from the data was used (total number of trips attracted and produced), but it's possible to model production and attraction on other parameters, such as the population and the number of workplaces in each zone. The highest number of trips happens for both the model and the real data within each zones. Zürich is obviously the zone with most of attracted and produced trips.

The MNL model, on the contrary, doesn't fit the data as well. The goodness of fit is lower than 0.4, which doesn't indicate a very good prediction. This is probably due to the relatively low number of data: some of them had to be deleted due to routing issues, other because of missing values. In the end the usable observations left are 2'613, which probably don't allow the creation of a very accurate model. The two models can then be combined into a single table that predicts the number of average trips for each mode and each pair of zones. Biking is the less used mean of transport, especially for very far zones. Car trips are dominant for almost all zones combination, except for a couple.

7.3 Cost benefit analysis

In this last section, it has been modeled a scenario where autonomous vehicles are introduced in the canton of Zurich. Following the Swiss Norms at best, it has been calculated a CBA that forecasted the impact of this potential policy in 40 years. In particular, the flow between the city of Zurich and the other cities of the canton has been in the scope of our research, with an increased travel times for cars and public transportation of 30%. The communal masterplan from Zurich City allowed us to forecast population growth in the future, while academic literature supported our assumptions on the benefits and costs of the autonomous vehicles. Results shows, with a NPV of -33.82 Mio. CHF and a CBR of 0.34, that even with a diminish of accident rate of 40% still the introduction of autonomous vehicles in the canton is not viable from an economical point of view. Performing a sensitivity analysis, it has been determined that the major bottleneck for this policy option are the construction costs of the fleet. A more accurate analysis, considering road capacity and different types of automated vehicles could strengthen the results of this analysis.

8 References

Ortuzar and Willumsen (2011) Modelling Transport, Wiley, Chichester.

Axhausen, K. W. (2016) Style guide for student dissertations, Working Paper, 1140, IVT, ETH Zurich, Zurich

Bösch et al. (2018) Cost-based analysis of autonomous mobility services, (doi.org/10.1016/j.tranpol.2017.09.005)

Meyer et al. (2017) Autonomous vehicles: The next jump in accessibilities?, (https://doi.org/10.1016/j.retrec.2017.03.005)

SN 641 820 (2006)

SN 641 821 (2006)

SN 641 822a (2009)

SN 641 824 (2013)

Data Source: https://ivtmobis.ethz.ch/mobis/en/