

Analysis of Individual Choices in Selecting Car as Mode of Transportation for Work or Study Commutes During the Business Week

SABD Project: Urban Mobility Analysis, Mode of Transport Choice.

Doga Yilmaz

31 / January / 2024

Table of Content

Introduction	2
Literature Review	2
Data Preparation	3
Data Preprocess / Analyse	3
Immobil Person	3
Selecting and Creating	3
Selecting Variables and Records	3
Creating New Variables and Tidying	4
Factorisation	4
One-hot Encoding	5
Relevel	5
Model Selection	6
Model Application	6
Logistic Regresion	6
Full Model	6
Reduced Model	9
Cross Validation	10
Decision Tree	10
Conclusion	11
References	13

Introduction

For the present report, we utilized a database extracted from a mobility survey conducted in Grenoble in 2010, provided by Professor Joly I. This database was employed to apply concepts developed during the semester in the subject of Smart Analytics for Big Data. The dataset encompasses information regarding the trips made by members of Grenoble households in 2010, as reported by Cerema (2013). It includes data related to individual and household characteristics, such as household structure, motorization, sex, age, professional status, and more, along with details about mobility practices, such as the number of trips, travel times, locations of trips, modes used, reasons for trips, etc.

For the specific case of this analysis, the focus was on analyzing the choice of the ‘car’ as the mode of travel for individuals. The aim was to identify the determinants of this choice among individual or household characteristics and modes of transport. Factors considered include the type of transport individuals usually take, the status of their driving license, age, and a specific emphasis on the influence of the household. Throughout the document, we will delve into data management, variables handling, model choice and fit, and more.

The *scientific question* guiding the present analysis was defined as follows: When applying logistic regression to the structured data, how significant are the effects of Age, Driving License status, and the different household characteristics in influencing the choice of car over an alternative transportation modes?

Literature Review

Travel mode choice is the complicated process by which people choose their modes of transportation. This process is impacted by a variety of variables, including environmental concerns and personal preferences. Understanding travel behaviour is essential to informing transport management and planning. (Dong Wang D., Liu Y., 2015) In addition, travel mode choice research reveals the impact of the transportation sector on climate change. The widespread public transportation network in cities reduces people’s use of private vehicles, which creates a positive environmental impact. (Cheea W., Fernandez J., 2013)

Article	Features	Method
Determinants of transport mode choice: a comparison of Germany and the USA (Buehler R., 2011, Virginia Tech University)	Trip purpose, car ownership, income, gender, household status, age	Multinomial logit
How Household Roles Influence Individuals’ Travel Mode Choice under Intra-household Interactions? (Ji Y. et al, 2018, Korean Society of Civil Engineers Journal)	Family structure, number of cars, household role, age, education level, employment status	K-means
The determinants of travel mode choice: the case of Łódź, Poland (Wójcik S., 2019, Bulletin of Geography)	Age, gender, driving licence, car ownership	Multinomial logit model
Transportation mode choice binary logit model: a case study for Johor Bahru city (Puan O. et al., 2019)	Age, gender, income, car ownership, comfort of car	Binary logit model

Figure 1: State of art

When the studies conducted in this field are examined, the features that give statistically significant results from the data obtained are almost similar to each other. Buehler (2011) conducted a comparative study examining mode choice in Germany and the USA, categorizing key attributes into four groups: socio-economic and demographic characteristics (e.g., household income, vehicle ownership), spatial development patterns (e.g., population density), transportation policies (e.g., fuel prices, parking costs), and culture and attitude (e.g., eco-socialization). (Puan O. et al., 2019) Buehler’s research revealed that vehicle owners who have a job and a driver’s license prefer the car at a higher rate.

In the article written by Ji Y. et al in 2018, it was examined how household roles affect travel mode choice. The article found that women are less inclined to travel long distances because they are usually busy with

household chores. However, when the distance exceeds the acceptable limit for cycling or walking, it has been observed that women's driving rates increase compared to their husbands.

In a study conducted by Wójcik S. in 2019 to determine the factors affecting travel mode choice in Poland, it was observed that owning a car increases the tendency to use private vehicles instead of public transportation. In addition, as age increased, the probability of regular use of private transport also increased up to a certain age, and then a decline was observed but according to the article the effect of age was linear for the public transport model. A one-year increase in age reduced the probability of choosing public transport by 0.004. Gender differences were clearly visible in the results of this study. Women were more likely (by 0.09) than men to use public transportation on a daily basis. The status of being a student had a significant impact on the regular usage of public transportation. Car ownership is the strongest positive determinant of choosing private transportation.

In the research conducted by Puan O. et al. in 2019, binary logit model was used as the artificial intelligence model since only the bus and car preference of the participants was examined. According to this research the mode choice of transport in Johor Bahru city is statistically correlated with age of users, income, vehicle ownership, car-comfortability, reliability of bus service, affective motives and instrumental motives. ROC curve was used to measure the accuracy of the model.

Data Preparation

For answering our research question, We will reduce the data to person level, selecting interested variables and creating new variables to describe family structure for each person.

Data Preprocess / Analyse

Immobil Person

In the data, we conclude that records (rows) which have a missing value in `id_depl`, have also a missing value in `mode_depl_ag`. It means that missing values in `mode_depl_ag` are attributed to persons being immobile (having no trip). So, we point person who has at least one trip in the day, and since the analysis focuses on transport choice (`mode_depl_ag`), records of immobile persons (who have no transport choice) are not relevant and are thus removed from the data.

Factorisation of “mot_o_red” We define categorical variable “mot_o_red” as factor and we associate labels to their modalities to make them usable in process.

Data Reduction For the beginning, we reduce the data to trips. It means we have more than one record based on their trips for each person.

Selecting and Creating

Selecting Variables and Records

We are selecting the variables that we are interested in for our case and make a new subset with them.

We filter the data frame ‘selected data’ by the level [2:5] which are “Travail, Ecole Primaire, Lycée, Université” values in the ‘mot_o_red’ column with the aim of focusing daily trip to go work or study. If we take just work, we eliminate children. Without children, it makes no sense to count child per household. If a household has child, we want to consider them and see them in analysis. So we take the persons going out from home to do their tasks on weekdays.

When we check our record counts, we see that there are 6309 records in our data, equal to the sum of “Travail,Ecole Primaire,Lycée, Université”, but 3965 unique ID for persons. It means we have more than one trip for each person.

Therefore, we choose to take the trip that has the longest duration for each person, expecting to have only one record for each person.

After this step, our data still does not have the same counts for records(rows) and unique IDs.

REMARK: Count of unique IDs and the count of rows are not the same, because there can be more than one record with the maximum value in the duree variable, i.e. multiple rows for an id_pers can have the same maximum duree value. This causes more than one row for the same id_pers to remain in the result data frame after the filter operation.

When we examine these records,we conclude that we have different mode of transport choices for same person. We don't want to select just one record and drop the others, because we focus mode of transport choices for a person and we want to analyse her/his choice based on familial structure. That person who has same family structure variables, choose different modes for going same destination because of different reasons that we are not interested. So we can/have to take that person with her/his all choices.

So we assign a unique ID for each combination of id_pers, duree, and mode_depl_ag.

From here we will use “combo_id” as unique identifier for each person.(with new created persons).

As a result, we have only one record for a person who has different choice even if (s)he has same familial structure.

```
## [1] 4674
```

```
## [1] 4674
```

,And as we can see in the outputs, the count of row and count of unique identifier are same.

Creating New Variables and Tidying

-In permis variable, we take ‘in progress’ and ‘yes’ values as “yes”.

-We create new variables that describe the family structure for the household: “child_count_hh” is to count the children in each household. “elderly_count_hh” is to count the elderly in each household. “size_hh” is for counting the number of people living in a household. “permis_hh” explains whether one of the people living in the same household has a permit. “car_count_hh” is for counting the cars in each household.

So we have our first clean data to go further with has variables below: “id_pers” , “id_men”, “NO_PERS”, “sexe”, “age”, “permis”, “VP_DISPO”, “lien”, “id_depl”, “mot_o_red”, “jourdepl”, “duree”, “mode_depl_ag”, “combo_id”, “child_count_hh”, “elderly_count_hh”, “size_hh”, “permis_hh”, “car_count_hh”

Factorisation

We define categorical variables which are “sexe,lien,mode_depl_ag, permis_hh” as factor and we associate labels to their modalities to make them usable in process.

We remove the variables “id_pers, NO_PERS, VP_DISPO, id_depl” that we will no longer use.

One-hot Encoding

We apply one-hot encoding for our “lien” variable. This means we create new columns for each value in the “lien” variable with “true/false” values for each household.

From the “sexe” variable, we create new variables that count men and women per household.

As a result of these preprocessing steps, we have a refined dataset with 4674 records, each representing an individual. The image below provides an overview of this dataset, showing statistics on gender, age groups, driving license status, reasons for trips, and the trip days.

N=4674		
Gender	Male	55%
	Female	45%
Age(Years)	<18	31%
	18-29	16%
	30-39	15%
	40-49	21%
	50-59	15%
	60>	2%
Driving Licence	Yes	65%
	No	35%
Trip Reason	Primary School	15%
	College and High School	17%
	University	6%
	Work	62%
Trip Day	Monday	19%
	Tuesday	25%
	Wednesday	16%
	Thursday	18%
	Friday	22%

Figure 2: Descriptives of the Data

Relevel

In our modeling efforts, we aim to focus exclusively on intraurban transportation and recognize the necessity of distinguishing between ‘by walk’ and ‘bus intercity’ modes, as they can not be grouped together. Therefore, we will exclude the ‘intercity’ category, which also has the fewest occurrences, and introduce a new category named ‘Others.’ This category combines ‘Other,’ ‘by Walk,’ and ‘BUS in City,’ allowing us to proceed with binomial regression modeling effectively.

Var1	Freq
Car	2574
Others	1957

Factorisation Check

We re-execute the factorization process to ensure that our categorical variables are correctly classified as factors. This step is crucial because previously, columns derived from one-hot encoding did not have the ‘factor’ data type.

For the purpose of predicting and analyzing “car” choice, we set the “Others” category as the reference level.

REMARK: Upon reviewing the factor variables, it is observed that the “Has_Grandparent_HH” column contains only ‘False’ values. Consequently, this column is not necessary for our analysis and can be omitted.

Model Selection

Once the explanation of the data and its handling has been done, it was implied that it was needed to determine the choice of car vs other alternatives. That's why a binomial model would be a nice fit for this specific case, choosing then to apply a Logistic Regression Model. The reference as it can be seen in blue in the image bellow, was seated to Else, this because it is wanted to analyse the effects of the chosen variables on car choice.

The team was also interested on the analysis of the effects of family structure and personal attributes, to determine if there is a significant insight it should be consider and to compare the behaviors of variables within a same model but different perspectives.

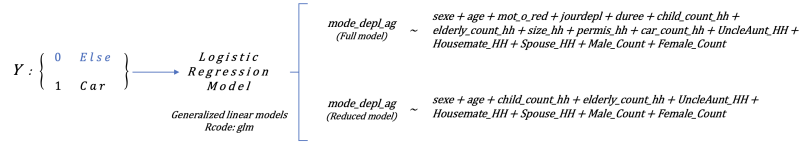


Figure 3: Model selection

During the development of this phase it was thought that it could also be beneficial to consider other algorithms that are less sensitive to heteroscedasticity, such as decision trees, being a viable option due to their inherent robustness and notable interpretability. That is why as an extra model, the team decided to implement decision tree to understand how it behaves with the selected data.

Model Application

Logistic Regresion

Full Model

We take all variables that we select and create in the Data Preparation:

```
model_full_binary <- glm(mode_depl_ag ~ sexe + age + mot_o_red + jourdepl + duree + child_count_hh + elderly_count_hh + size_hh + permis_hh + car_count_hh + UncleAunt_HH + Housemate_HH + Spouse_HH + Male_Count + Female_Count, family = binomial, data = selected_data)
```

% latex table generated in R 4.3.1 by xtable 1.8-4 package % Fri Feb 2 13:50:17 2024

	Estimate	Pr(> z)
mot_o_redECOLE PRIMAIRE	-1.37	0.00
mot_o_redC.E.S.-LYCEE	-2.54	0.00
mot_o_redUNIVERSITE	-1.53	0.00
jourdepl4	0.35	0.00
jourdepl5	0.34	0.00
duree	-0.01	0.00
child_count_hh	0.18	0.00
permis_hhNo	-0.78	0.00
car_count_hh	1.08	0.00
Male_Count	-0.21	0.00
Female_Count	-0.23	0.00

[1] "Null deviance: 6197.0 on 4530 degrees of freedom, Residual deviance: 4609.5 on 4510 degrees of freedom, AIC: 4651.5"

The output from the full logistic regression model provides insights into the variables that influence the selection of a “Car” compared to the “Others” reference category. Here are observations regarding the coefficients and their levels of significance:

1. **Gender (sexefemale):**

- The coefficient for the “sexefemale” variable is approximately -0.1497, indicating that being female is associated with lower log-odds of choosing a car. However, this effect is marginally significant ($p = 0.082$), suggesting that gender might have a minor influence on the choice of car.

2. **Trip Reason (mot_o_red):** Different destinations have a significant impact on the choice of car:

- “ECOLE PRIMAIRE” is associated with a significant decrease in the log-odds of choosing a car ($p < 0.001$). “C.E.S.-LYCEE” is associated with an even larger decrease in log-odds ($p < 0.001$). “UNIVERSITE” is also associated with a significant decrease in log-odds ($p < 0.001$).

3. **Day of trip (jourdepl):**

- The “jourdepl” variable represents the day of trip. Only “jourdepl4” and “jourdepl5” have statistically significant coefficients. It means they are associated with higher log-odds of choosing a car compared to “Others” ($p < 0.01$).

4. **Duration (duree):**

- The coefficient for “duree” is approximately -0.0114, and it is statistically significant ($p < 0.001$). This suggests that as the journey duration increases, the log-odds of choosing a car decrease.

5. **Household Characteristics:**

- “child_count_hh” is associated with higher log-odds of choosing a car ($p < 0.001$). “size_hh” is associated with lower log-odds of choosing a car ($p < 0.05$). “permis_hhNo” (lack of a driving license in the household) is associated with lower log-odds of choosing a car ($p < 0.001$). “car_count_hh” (the number of cars in the household) is strongly associated with higher log-odds of choosing a car ($p < 0.001$).

6. **Counts of Males and Females (Male_Count, Female_Count):**

- “Male_Count” is associated with lower log-odds of choosing a car ($p < 0.001$). “Female_Count” is also associated with lower log-odds of choosing a car ($p < 0.001$).

Overall Model Fit: - The model’s null deviance and residual deviance are provided, with the residual deviance being substantially lower. This indicates that full model provides a significantly better fit to the data compared to a null model. - The AIC value (Akaike Information Criterion) is 4651.5, which is a measure of model goodness-of-fit. Lower AIC values suggest a better trade-off between fit and complexity.

In summary, several variables have statistically significant effects on the choice of a car versus Others. These variables include trip reason, day of trip, duration, household characteristics (such as the presence of children, household size, and the number of cars in the household), and the counts of males and females. Other variables, such as gender and the presence of specific household members, have less significant effects.

Based on the **Diagnostic Plots** :

1. **Residuals vs Fitted:** This plot shows if there are any non-linear patterns in the residuals, and whether the model is correctly specifying the functional form. The residuals should be randomly dispersed around the horizontal line at 0. In our plot, the residuals don’t display a clear pattern, which is good. However, there is a slight ‘fanning out’ effect indicating potential heteroscedasticity (non-constant variance of residuals).
2. **Normal Q-Q:** This plot is used to check the normality of residuals. If the residuals are normally distributed, the points should fall approximately along the reference line. Our Q-Q plot shows that the residuals deviate slightly from normality at the tails, but overall, they seem reasonably well-behaved.

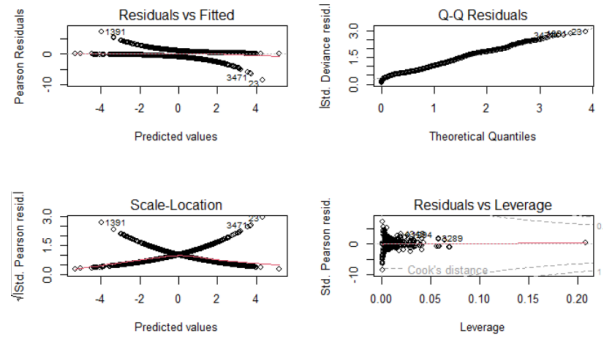


Figure 4: Descriptives of the Data

3. **Scale-Location (or Spread-Location):** This plot shows if residuals are spread equally along the ranges of predictors. This is another way to check for homoscedasticity. Our plot shows that the residuals' spread increases with the fitted values, indicating possible heteroscedasticity.
4. **Residuals vs Leverage:** This plot helps to identify influential cases (outliers), which are points with high leverage that can have a disproportionate impact on the model. The Cook's distance lines help to identify these points. In our plot, there are a few points outside the Cook's distance lines, it means we may have some influential outliers in dataset.

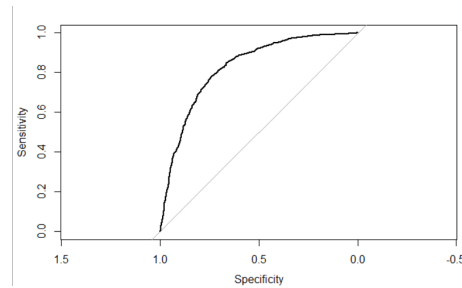


Figure 5: Descriptives of the Data

Based on the **ROC Curve and AUC:**

Our ROC curve shows a good separation between the positive and negative classes, as it bows towards the top left corner of the plot.

The AUC (Area Under the ROC Curve) value is 0.8245, which is a good score. An AUC of approximately 0.82 suggests that the model has a good predictive ability to distinguish between the classes in our outcome variable.

To sum up for our full model, it seems to perform well in terms of AUC, indicating good predictive power. However, the diagnostic plots suggest that there might be some issues with heteroscedasticity and influential outliers. These might affect the reliability of the coefficient estimates and standard errors.

Let's evaluate:

In the descriptive statistics results, we can see a high difference between duree ranges. Looking at the standard deviations (25.86 in full data, 57.43 in influential data), there is a much higher standard deviation for duree in the influential subset which indicates greater variability.

Looking at the density plot for age, the shapes of the two distributions (lines) are similar but the red line is higher than the blue line in the 20-30 age range. This suggests that the influential points are skewing the results of the analysis. For example, it is possible that the influential points are all from a single population,

such as a group of people who were all born in the same year. This would artificially increase the density of people in the 20-30 age range. We can conclude same result from boxplots and descriptive statistics results for age variable.

Looking at the proportions of the categories within the influential subset versus the full dataset. The frequency proportions are similar. The influential points are not causing a significant shift in the distribution of our outcome.

In conclusion, the influential points in our dataset do not appear to be clear outliers, but they do exhibit some characteristics that set them apart from the rest of the data. Their influence on our model may be due to the variability in numeric variables like “duree” and the concentration of points within certain ranges of the “age” variable.

As a result, since influential points don’t change the outcome proportion and they represent a meaningful subgroup within data that we want our model to capture, and the higher variability in some variables among the influential points is not problematic for our analysis as they are valid data points, it may be appropriate to keep them.

Reduced Model

By eliminating variables such as permis and car count that are statistically significant in the full model, we reduce our model only to family structure (variables containing age,relation, gender information) to see if there is any change.

```
model_reduced_binary <- glm(mode_depl_ag ~ sexe + age + child_count_hh + elderly_count_hh + UncleAunt_HH + Housemate_HH + Spouse_HH + Male_Count + Female_Count, family = binomial, data = selected_data)
```

% latex table generated in R 4.3.1 by xtable 1.8-4 package % Fri Feb 2 13:50:18 2024

	Estimate	Pr(> z)
(Intercept)	-0.84	0.00
sexefemale	-0.21	0.01
age	0.04	0.00
Spouse_HHTRUE	0.62	0.00
Female_Count	-0.16	0.00

[1] “Null deviance: 6197.0 on 4530 degrees of freedom, Residual deviance: 5475.8 on 4521 degrees of freedom, AIC: 5495.8”

1. Intercept (Intercept):

- The intercept represents the log-odds of choosing a car when all other predictor variables are zero. The log-odds is approximately -0.844, and it is statistically significant ($p < 0.001$). This means that when all other predictors are zero, there is a significant difference in the odds of choosing a car compared to “Others.”

2. Gender (sexefemale):

- The coefficient for the “sexefemale” variable is approximately -0.2083, indicating that being female is associated with lower log-odds of choosing a car. This effect is statistically significant ($p = 0.007$), suggesting that gender has a significant influence on the choice of a car.

3. Age (age):

- The coefficient for age is approximately 0.0412, and it is highly statistically significant ($p < 0.001$). This suggests that as age increases, the log-odds of choosing a car also increase significantly.

4. Living with Spouse [Spouse_HH]:

- The coefficient for Spouse_HH is approximately 0.6219, and it is highly statistically significant ($p < 0.001$). This suggests that as Spouse_HH increases, the log-odds of choosing a car also increase significantly.

5. Counts of Males and Females (Male_Count, Female_Count):

- “Male_Count” is associated with lower log-odds of choosing a car ($p = 0.011$).
- “Female_Count” is highly statistically significant and associated with lower log-odds of choosing a car ($p < 0.001$).

The residual deviance is lower than the null deviance, indicating that our model provides a better fit to the data compared to a null model. but the AIC value (5495.8), which is a measure of model goodness-of-fit, is greater than the Full model’s.

REMARK: Some variables that were not initially statistically significant in the full model have become significant in the reduced model.

- We have similiar results for the reduced model with the full model in Diagnostics Plots.
- The Area Under the Curve (AUC) for the full model is 0.8245, while the AUC for the reduced model is 0.7241. This suggests that the full model has a better ability to distinguish between the two classes than the reduced model.

Comparing Models: The choice between the full model and the reduced model should be based on a trade-off between complexity and predictive power. While the full model has a higher AUC, indicating better predictive performance, it’s also more complex and could be overfitting. The reduced model, while simpler and potentially more generalizable, has a lower AUC, indicating less predictive accuracy.

Let’s move on choosing full binary model.

Cross Validation

```
## [1] 0.1677044 0.1676156
```

Lower values for these metrics indicates better model performance. The two values are close to each other shows that the model’s performance is relatively consistent across the different cross-validation folds, indicating that our model is likely to generalize well to unseen data.

It could also be beneficial to consider other algorithms that are less sensitive to heteroscedasticity, such as decision trees or ensemble methods like random forests or gradient boosting machines.

In terms of model selection, if we are seeking a model with greater interpretability, a decision tree could be a good choice, as it provides a clear visualization of the decision paths. We might also try ensemble methods or support vector machines, which often perform better in terms of predictive accuracy but are less interpretable than decision trees.

Decision Tree

We apply a decision tree model to understand and evaluate the model’s metric scores.

	Others	Car
Others	407	184
Car	180	588

Accuracy: 0.732156

Our decision tree model has an overall accuracy of approximately 73.21%. This means that the model correctly predicts the transport choice for about 73 out of 100 cases. Looking at the specific metrics:

- Sensitivity (Recall) for 'Car': The model has a sensitivity of about 69.34%, indicating that it correctly identifies 69.34% of the actual 'Car' transport choices. This is a measure of the model's ability to identify positive results.
- Specificity for 'Others': The specificity is approximately 76.16%, which means that the model correctly identifies 76.16% of the instances where transport choices other than 'Car' are made.
- Precision for 'Car': Precision is about 68.87%, which shows that when the model predicts 'Car' as the transport choice, it is correct approximately 68.87% of the time.
- F1 Score for 'Car': The F1 score, which balances precision and recall, is approximately 69.10%. This suggests that the model has a fairly balanced performance between precision and recall, which is generally desirable.

From these scores, we can infer that our model has moderate performance in predicting the transport choice based on household structure. While the accuracy is relatively high, there is still room for improvement, especially in terms of precision and recall. This could potentially be improved by feature engineering, hyperparameter tuning, or using a more complex model.

Conclusion

The researchers initially considered employing the K-means algorithm for clustering. However, given the categorical nature of the dataset, this approach was deemed unsuitable. Consequently, the researchers opted for a Binary Logistic Regression model, a choice in line with the literature's recommendation for analyzing categorical outcomes. This model choice supports the research question which is: "When applying logistic regression to the structured data, how significant are the effects of Age, Driving License status, and the different household characteristics in influencing the choice of car over an alternative transportation modes?" and is justified based on the results and conclusions obtained from the logistic regression analysis.

In the full model we applied; gender, household size, number of children in the household, number of men and women in the household, trip reason, trip duration, specific trip days, presence of a driver's license in the household and the number of cars owned by the household appear to be significant.

In the evaluations we make for model analysis, metric scores showing the accuracy of the model give good results. In addition, as a result of the examinations, it was found that there is a certain age group that directs the results of the full model we applied.

This shows that although the age variable does not appear to have a statistically significant effect in the full model, it has an effect that directs the whole model. Since the distribution of this particular age group on the binary output matches the entire data and represents an important subset of the data, it was concluded that it should be kept in the dataset and evaluated in the model.

As seen in the results of the articles and in the full model we applied, the variables of driving license, number of cars, and trip duration, trip day, and trip destination related to the trip were statistically significant.

In the reduced model, these household variables were dropped and the effect of the family structure variables and the accuracy scores of the new model were examined. As a result of the applied reduced model, it was seen that gender, age, living with a spouse and the number of women in the household had a statistical effect on the choice of car as a transport mode.

In the Diagnostic Plot results examined for model analysis, similar results were encountered with the full model.

The AIC score and AUC score of the reduced model gave worse results than the full model. In this case, driver's license ownership, car count, trip destination, trip duration and certain trip days, which were significant in the full model, contributed more positively to the model. As a result of the analysis, the variables age, which are not significant in the full model but whose impact we see on the model, and living with Spouse appear as significant in the reduced model.

In both models, gender being female and the number of women in the household appear as significant variables that have a negative impact on the choice of car for transportation.

Given these significant findings, it is evident that age, driving license status, and various household characteristics play crucial roles in shaping the preferences of car over an alternative transportation mode. The research question is justified as it successfully captures and evaluates the impact of these variables on the transportation choices.

The results we obtained are consistent with the results obtained in the articles. As in the literature review section, the results obtained at the end of the study showed that age and gender were important factors in travel mode choice. Household roles are also examined, as in the article written by Ji Y. et al. (2018), and found that the number of women, men and children in the household was a statistically significant measure.

References

- Determinants of transport mode choice: a comparison of Germany and the USA (Ralph Buehler, 2011, Journal of Transport Geography 19 (2011), 644–657)
- How Household Roles Influence Individuals' Travel Mode Choice under Intra-household Interactions? (Ji Y. et al., 2018, Korean Society of Civil Engineering Journal, 22(11), 4635-4644)
- The determinants of travel mode choice: the case of Łódź, Poland (Szymon Wojcik, 2019, Bulletin of Geography Socio-economic Series, No.44, 93-101)
- Transportation mode choice binary logit model: a case study for Johor Bahru city (Puan O. et al., 2019, IOP Conference Series: Materials Science and Engineering, 527)