How should we get there? Transport mode modelling in Brisbane, Australia

Craig Whiffin (02389703)
IFN703/4 Assessment 3, due 11:59pm Sunday 14 June 2020

Executive Summary

Policy makers are keen to understand how travellers can be persuaded to make greener and more sustainable travel mode choices. This document explores the literature on the research that has been done on travel mode choices as well as the modelling approaches that are best applied to this type of analytical work. Given various models have been used for decades, study into the best method is included as part of the literature review. Ultimately, a multinomial logit model was created to assess the relationship between the selected features. The approach consisted of standard pre-processing activities such as cleansing and mapping to reduce the number of classes and balance the remaining. Exploratory data analysis was performed to gain insight into the relationship between the variables. The model showed that travellers have a strong preference for driving themselves, even where trips did not involve carrying heavy or many items. It was also observed that despite the evidence in the literature, travel time had little effect on transport mode choice. The report concludes with reflection and future work.

Introduction

Clear insight into the factors that influence commuters' transport decisions assists policy developers in both private and public sectors to incentivise desirable behaviours. One general method involves discrete choice modelling (DCM) which differs from other consumption models where a product or service is assumed to be measured by how much is consumed. Instead, discrete choices are accepted to be made against available alternatives with defined probabilities. This type of analysis has been applied to transport mode choices for groups and individuals for decades with more sophisticated models continually developed to account for limitations in their predecessors. Several studies comparing the performance of models have been completed by Greene and Hensher [1] and Shen [2] concluding that there is no magic bullet when it comes to model performance. There are numerous studies [3]–[5] that investigate group and individual factors that influence transport mode decisions however very few have applied this to the travel choices made by commuters in Brisbane, Queensland. This project sets out to apply proven modelling techniques to identify factors that influence Brisbane travellers' transport mode choices among the available alternatives using updated data collected in 2018-19.

Literature Review

Transport mode choices: contributing factors

The benefits of accurately predicting consumer choices are obvious in the transport sector, particularly in large modern cities where demand often exceeds supply. However other adjacent policy initiatives can be informed by traveller behaviour including environmental, health and well-being; and urban planning. A common goal for strategists across these areas is the minimisation of road vehicles and the development of incentives that encourage passengers to adopt alternative modes of transport. While the choice of alternative modes may differ between regions, the most common include private vehicle, public vehicle, public transit, bike and walk. What leads an individual to choose a mode of transport varies and several studies have concluded that lifestyle-related factors such as household structure, residential location, income and vehicle ownership play a role in the utility, or ranked preference, attributed by commuters to their travel options. Ardeshiri et al [5, p. 343] conclude that there is consensus that factors in commuters' built environment, such as density and distance to travel, have influence on commuter behaviour with only the extent of that influence in question when compared to other factors. More subjective factors have also been considered. For example, Krueger et al [6] explored the relationship between normative beliefs, i.e. behavioural perceptions; and transport mode choices. Their proposed methodology used stated normative beliefs to categorise individuals into "latent", or hidden, classes which differed from preceding studies that focused mostly on socio-demographic variables for class categorisation. The latent normative beliefs were found to significantly contribute to the mode choices commuters made. However, Chorus and Kroesen [7] criticise the use of latent variables (LV) generally for travel choice prediction. They claim that individuals' perceptions are endogenous to transport choice which disallows any claim of causality and that the cross-sectional nature of the data prevents comparison of changes in choice at an individual level. This remains a hotly debated topic in the data modelling world. More recently, Bahamonde-Birke et al considered the criticism levelled at the use of LV and concluded that while they do affect individuals' travel choices, greater transparency in the feature selection process is required [8]. This indicates that the mathematical techniques are not in question but how features are selected requires a greater degree of reflection and communication in research papers.

Transport mode choice within Brisbane, Australia

Several theses have contributed to transport choice analysis for commuters in Brisbane, Australia however they lack the rigor of peer review to provide high levels of confidence in their findings. For example, Khan et al developed passenger mode choice models using the 2003/04 South East Queensland Travel Surveys, finding that white collar workers were willing to spend considerably more money travelling by car to work than their blue collar counterparts. They also identified that tertiary education students were likely to utilise public transport as a higher percentage of their modal split than other groups [3, p. 15]. Pourfarzad modelled passenger behaviour before and after the 2011 Brisbane floods [9] however this was limited only to the period directly preceding and following the natural disaster and was intended to identify changes in modal preferences as a result of the event. A gap exists to apply discrete choice modelling to updated travel survey results to identify factors that influence Brisbane residents' everyday transport choices. A review of the candidate models follows.

Model selection

Modern choice modelling theory can be traced back to Thurstone's proposal of the law of comparative judgement in 1927 in which he applied binary logistic regression to predict subjects' responses to presented stimuli [10]. By the 1980s, it had been applied in numerous fields due to advances in modelling frameworks, namely the development of multinomial logistic regression. Many modelling approaches have appeared throughout the literature over the last five decades. The most popular model during this period was the multinomial logit model (MNL), an extension of Thurston's proposed probit model. An important property of the MNL is that it assumes independence from irrelevant alternatives (IIA) which stipulates that the ratio of the probabilities of two alternative choices is independent of the option to choose a third alternative. This is based on economic utility theory whereby a person's preferences are represented by a utility function provided that their alternative choices can be numerically assigned. Alternative models were developed to overcome this limitation as it applies only to homogenous populations. Of the alternatives, the mixed logit model (MLM) gained popularity. This led to Greene and Hensher to compare the MNL with a modified latent class model (LCM), also based on the MNL. They claim that unlike MLMs, LCM do not demand that analysts make potentially incorrect assumptions about the lack of uniformity among individuals. However, the MLM provides greater capacity to specify this unobserved heterogeneity which may counter-balance the weakness of faulty assumptions [1, p. 697]. Shen later compared the LCM and MLM in the same transport choice context and reached the same overall conclusion: that although the statistical behaviours better support the LCM than MLM, this cannot be generalised to all situations [2, p. 2923]. Ultimately, neither model is best in all situations and each must be selected based on the context in which they're used.

More recently, four models were compared using specialised software by Yu and Sun [11]. Those included the MNL and MLM already mentioned as well as the nested logit (NL) and heteroscedastic extreme value (HEV). The technical differences, if not their impact, are minor between the models and Yu and Sun concluded that while the MLM is most likely superior in terms of performance, it is also more difficult to estimate and unless the IIA assumption fails, the MNL should be considered first preference for discrete choice modelling, simultaneously stipulating that no single model is universally superior [11, p. 301]. This is a common theme across these analyses. No single model is best in all situations and it is up to the analyst to decide based on the available data and intended application.

Approach

Data Collection

Survey response tables were programmatically extracted from the Microsoft Access Database. The project's original intent was to combine Australia Bureau of Statistics (ABS) datasets to compliment the survey responses. As such, the most appropropriate overlapping year 2017/2018 was chosen. When the decision to exclude the ABS data was made, the original dataset was kept due to the large amount of data cleansing and exploratory analysis already completed.

Data Cleansing
The following data cleansing tasks were performed over the various data sets extracted from the survey response database.

- Unique ID checks
- Removal of missing values
- Data type conversions

After initial inspection, two variables were mapped to reduce the number of unique categories: transport mode and trip purpose. Two factors motivated this merger. First, since transport mode is the response variable, the number of unique values was lowered from 7 to 4 to reduce model complexity. The second reason was an attempt to balance the response variable classes more evenly. Though in the end the classes remained unbalanced, the number of observations in the new, merged categories are still higher than before the mapping which is desirable. Figure 1 shows the class balances for transport mode before applied mappings and Figure 2 shows the balance after. It can be seen that the number of observations in the active class are still quite few compared to the dominant class Car driver, however grouping the public transport options appears to have made a reasonable difference with a single group containing just above 4,000 observations. Table 1 shows the original and new values for the transport mode mapping.

Figure 1: Class balance for transport mode before applied mappings. It can be seen that a few categories have just tens of observations such as Ferry and Light rail.

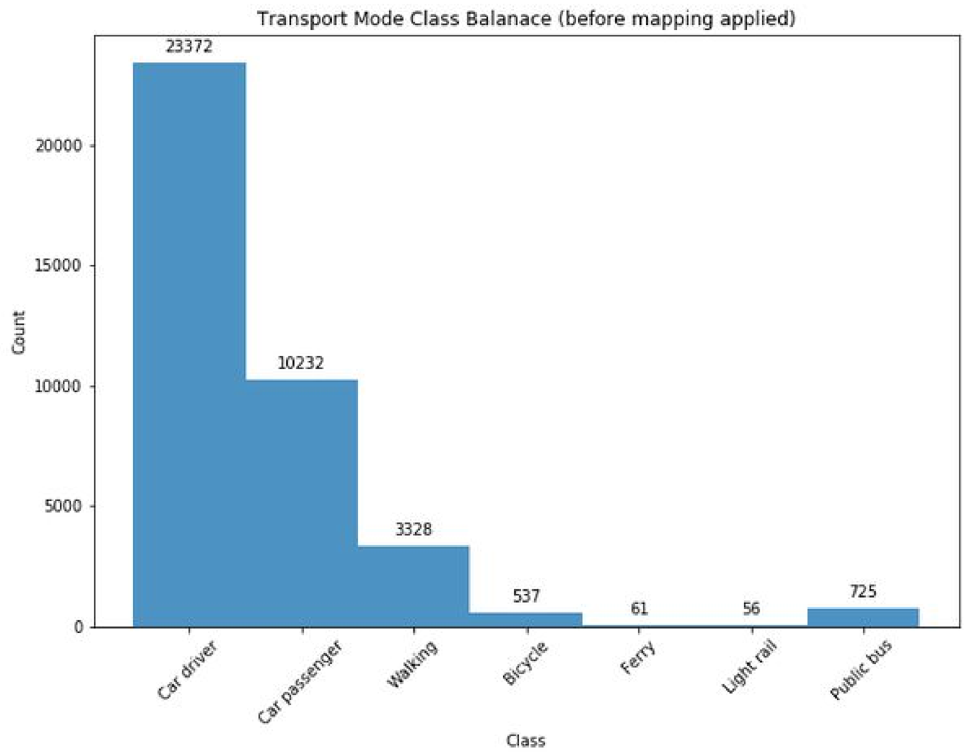


Figure 2: Class balance for transport mode after applied mappings. The active category now looks more reasonable with just under 4,000 observations and there has been an increase in the combined Public transport category, however small compared to the overall totals.

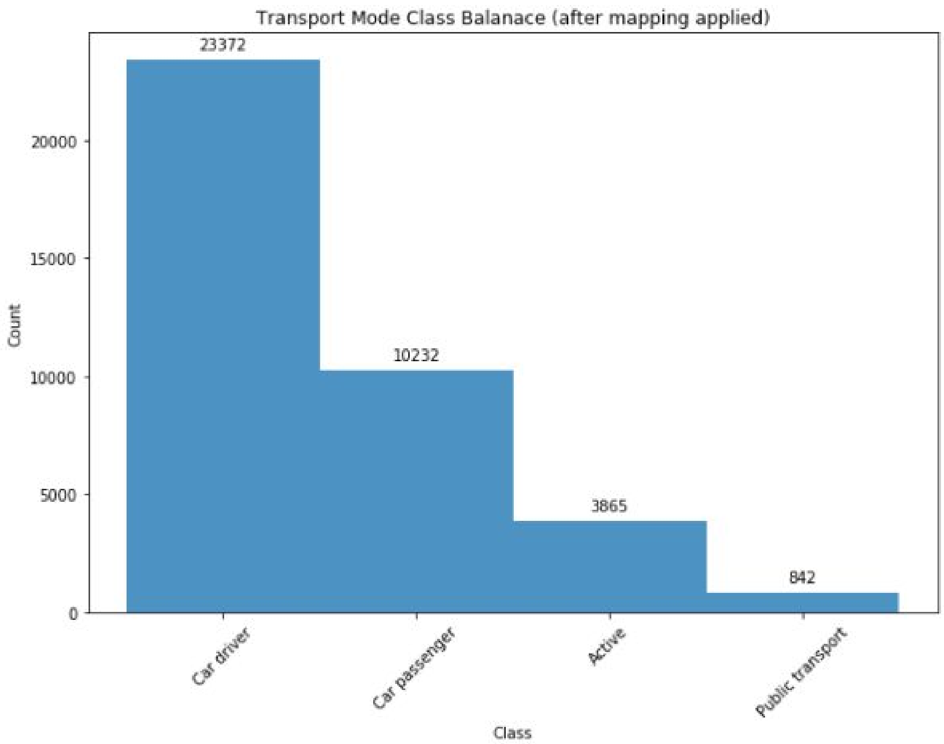


Table 1: Original and new mode mappings. *Car driver* and *Car passenger* remain unchanged while three categories were rolled up to Public Transport (Ferry, Light rail and Public bus) and two into Active (Bicycle and Walking).

Transport Mode Value	New Transport Mode Value	New Transport Mode Code
Car driver	Car driver (unchanged)	0
Car passenger	Car passenger (unchanged)	1
Ferry	Public Transport	2
Light rail	Public Transport	2
Public bus / “Bus”	Public Transport	2
Bicycle	Active	3
Walking	Active	3

Trip purpose was also merged however given it is not the response variable, its balance doesn’t deserve as much scrutiny. Table 2 below shows the mapping and code values.

Table 2: Original and new Trip Purpose values. *Direct Work Commute* and *Work Related* were rolled up into a single Trip Purpose category called Work. *Pickup/Deliver Something* and *Pickup/Dropoff Someone* were merged to Pickup/Dropoff. The remaining values were unchanged.

Trip Purpose Value	New Trip Purpose Value	New Trip Purpose Code
Accompany Someone	Accompany Someone (unchanged)	0
Direct Work Commute	Work	1
Work Related	Work	1
Education	Education (unchanged)	2
Other Purpose	Other Purpose (unchanged)	3
Personal Business	Personal Business (unchanged)	4
Pickup/Deliver Something	Pickup/Dropoff	5
Pickup/Dropoff Someone	Pickup/Dropoff	5
Recreation	Recreation (unchanged)	6
Shopping	Shopping (unchanged)	7
Social	Social (unchanged)	8

Exploratory Data Analysis

Variables selected for exploration were based on those found in the literature. In particular, Koppelman and Bhat provide a list of candidate variables for transport mode modelling [12]. Of those available in the data set, the following were chosen as explanatory variables.

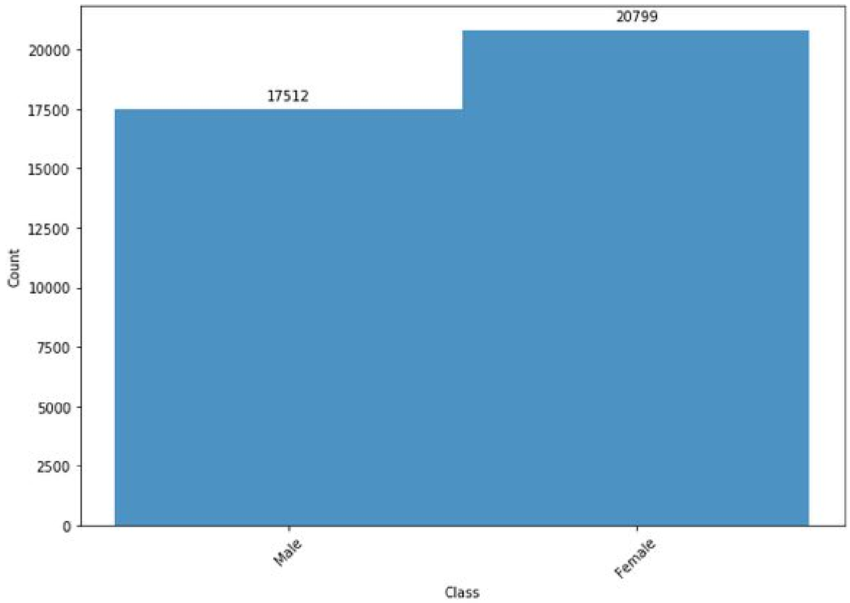
- Sex
- Age group
- Time travelled
- Trip purpose

Two key variables not available include income and travel cost. It is noted here that these two missing variables and / or interactions with them may change the model's predicted outcome for transport mode. Though consideration was given to approximate income by region, accurate data could not be sourced and applied sensibly so this was abandoned. Similarly, the complexities involved in mapping cost to each trip are great given travel costs change over time.

Transport Mode Type by Sex

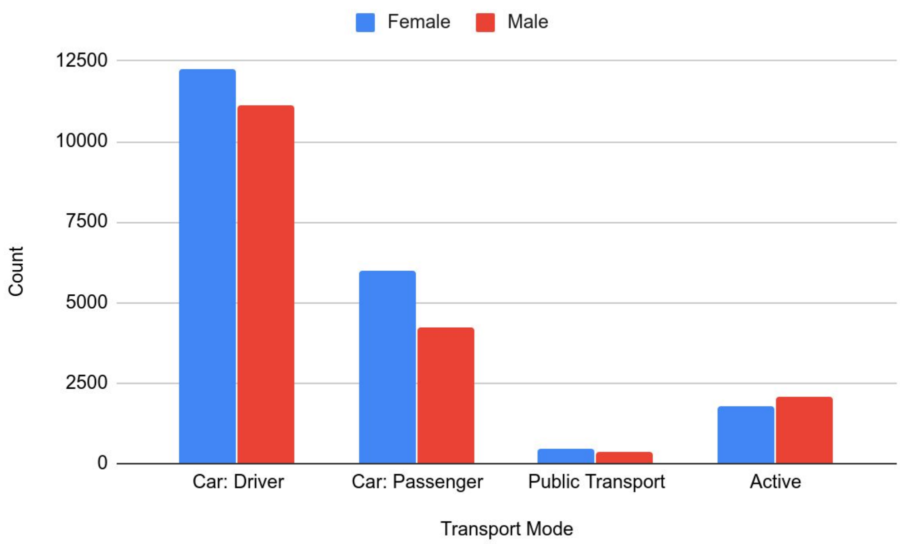
Figure 3 below shows the class balance for the variable, Sex. It can be seen that about 17,512 (~45%) out of a total 38,318 are male respondents and 20,799 (~55%) female.

Figure 3: Class balance of the Sex variable. The balance between male and female respondents is close to even at 45:55 ratio for male to female.



In Figure 4, it can be seen that there are lower counts for males and females but that most of the transport modes are proportional to the number of respondents for each sex. One exception is for the Active transport mode. It shows that more males choose Active whereas if there were no relationship, it would be expected that it would be lower in number than females. This hints at a relationship between sex and transport mode.

Figure 4: Transport Mode Type by Sex. As the survey is made up of around 45% to 55% respectively for females and males respectively, no relationship would show relatively even bars for each group however this is not the case except for Active.



Transport Mode Type by Age Group

A bar plot showing the total counts for each transport mode type was produced for the respondents' age group. Some reassuring signs are seen as the young do not appear to be driving cars. Public transport appears to be a fairly democratic choice with people of all ages choosing to travel by it. As expected, we see low numbers for the very elderly and Active transport however this is likely due to the small number of observations given that counts are very low across all transport mode types for these age groups. If the general shape of the bars in Figure 5 are compared to those in Figure 6, we see two shapes that resemble each other, indicating that there is no obvious indication of a relationship between age group and transport mode.

Figure 5: Transport Mode Type by Age Group. The Car:Driver is by far the most popular transport mode choice across most ages except those in the non-driving age ranges: the very young and the elderly.

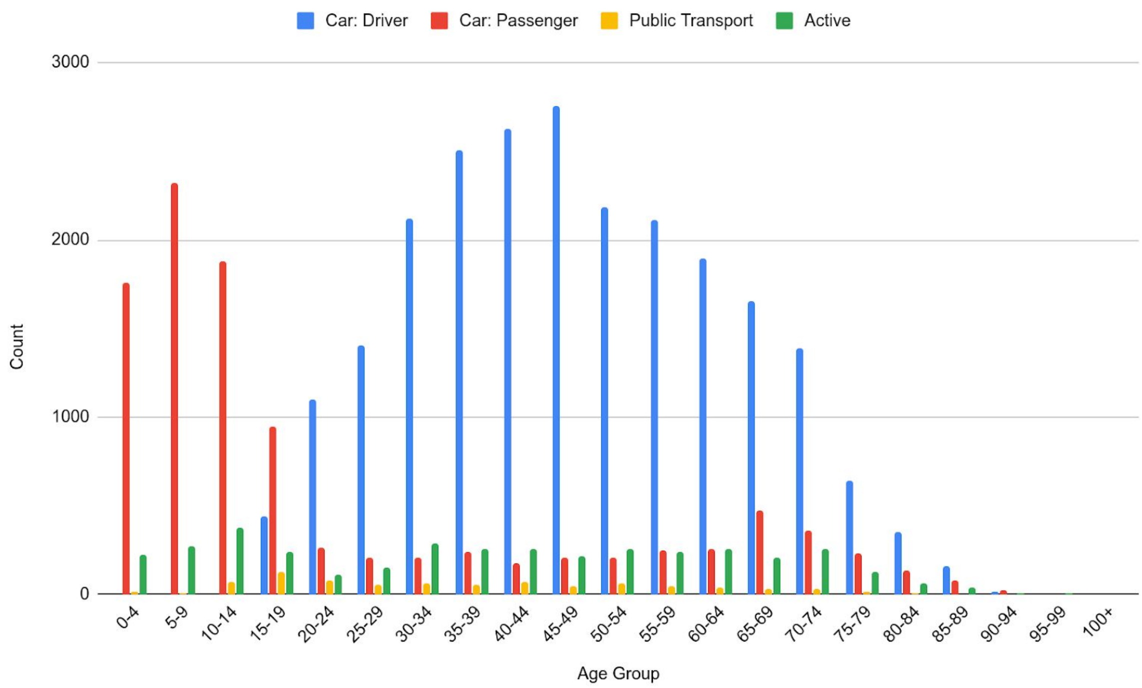
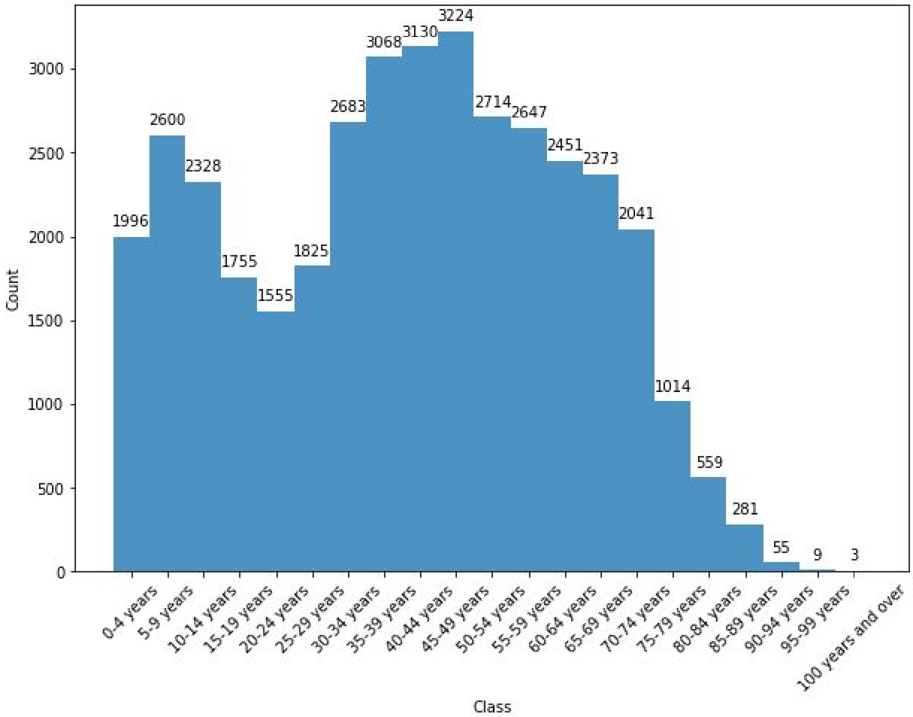


Figure 6: Histogram showing the counts of respondents by age group. Its similarity in shape compared to that in Figure 5 indicates there may be no obvious relationship between Age Group and Transport Mode.



Transport Mode Type by Purpose

An inspection of the class balance for the Purpose variable can be performed using Figures 7 and 8 below. Unlike in previous comparisons, it is not valid to compare figure shapes since Figure 8 is grouped by Transport mode and Figure 7 is not resulting in the height of the bars to be incomparable. Instead, Figure 7 is included to gain an impression of the class numbers, indicating here that Work travel is by far the most frequent travel purpose followed by Shopping and Pickup/Dropoffs. Interestingly, in Figure 8 it can be seen that for these three highest categories, travellers tend to prefer methods that are convenient, namely driving their cars. For Shopping and Pickup/Dropoffs, it makes intuitive sense for car use to be preferred given the challenges of moving objects between locations. However this rationale does not apply for the Work purpose necessarily and it is not fully clear from this exploratory analysis whether the high numbers simply represent the proportion of respondents.

Figure 7: Histogram showing class counts for the Purpose variable.

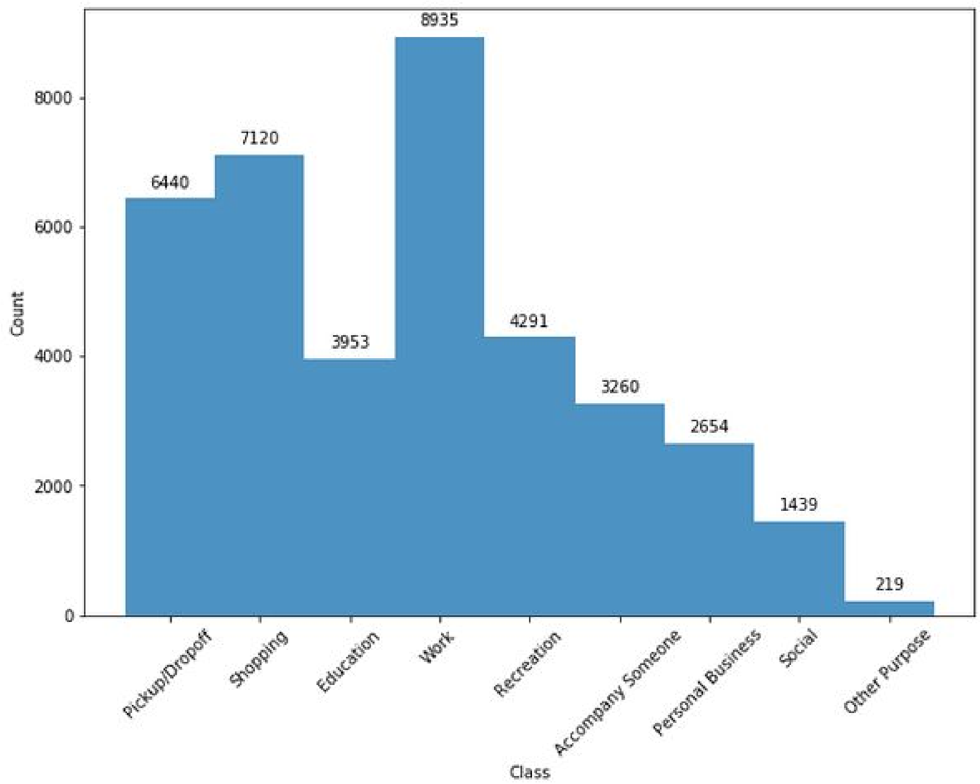
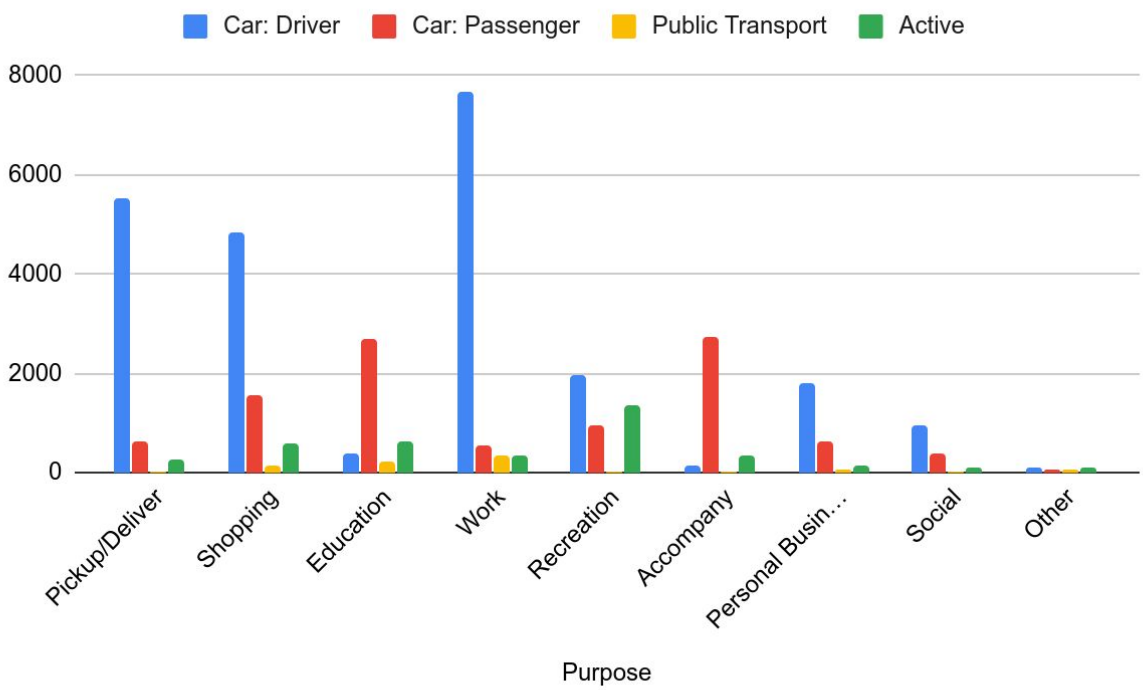


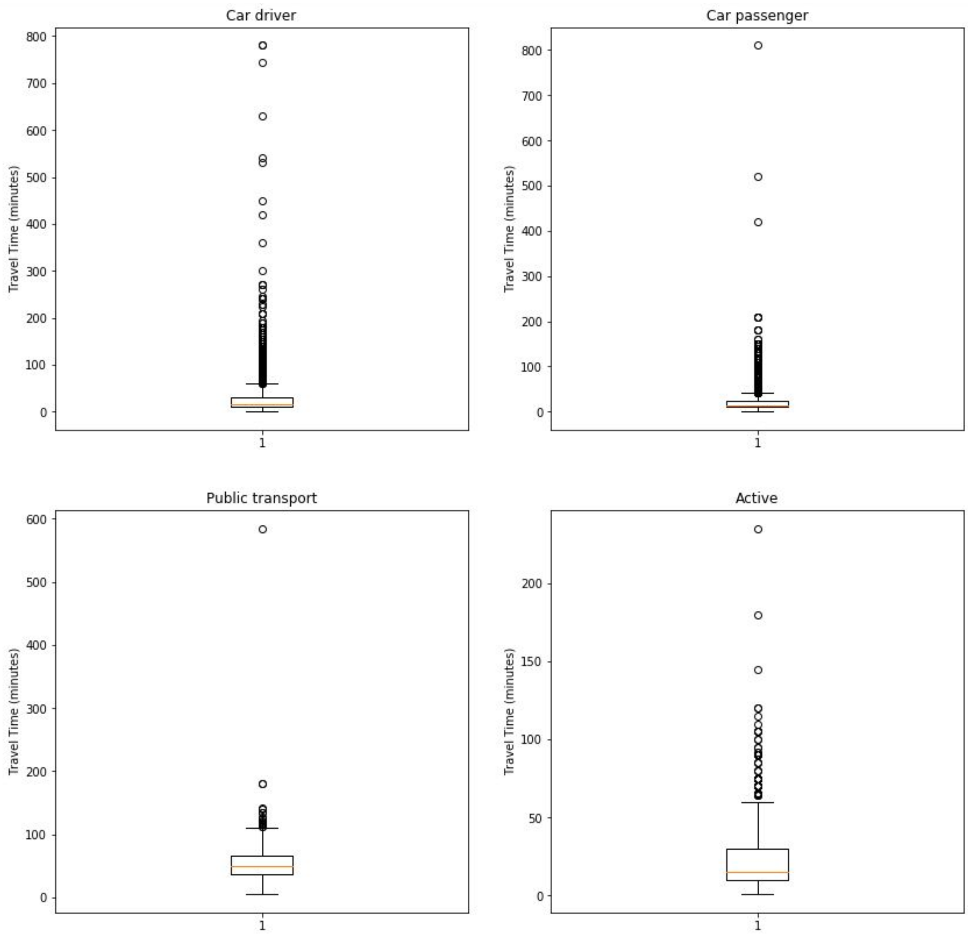
Figure 8: Bar plot of Trip Purpose by Transport Type



Transport Mode Type by Travel Time

An inspection of spread of data for Travel time grouped by Transport mode is shown below in Figure 9. As expected, the active modes have an overall lower travel time than the other three types. Recall that Active includes walking and cycling. It appears that the majority of trips for Car driver and Car passenger are short in duration. Although car travel is typically faster than the other forms, car travel may also attract long distance travel for comfort purposes. It can be observed that public transport appears to have a larger range indicating that the time it takes to complete a trip on public transport is typically longer than by car.

Figure 9: Box plots showing spread of data for Travel time for each Transport mode. The range of data for Car driver and Car passenger are longer however the majority of trips are well below 100 minutes, lower than that of Public transport.



Findings

A multinomial logit model was created based largely on the assessment by Shen [2, p. 2924]. The output of the model can be seen in Table 3. One of the drawbacks of the model includes its interpretability. Each group's figures are made with respect to what is called a reference group. In this case, the reference group is Car driver and is the category against which all other figures are measured. For example, for the Car Passenger group, the first three coefficients are less than one. Values less than one indicate a reduced chance to choose Car passenger over Car driver and values above one indicate an increased chance, for a single unit increase in the variable. For example, a single unit increase in Age group (age bracketed by 5 year increments) would result in a reduced chance of 73% of choosing that outcome over Car driver. Similarly, a single unit increase in Sex, which amounts to being male, results in a reduced chance to choose being a car passenger over driving yourself.

Table 3: Multinomial logit model output showing regression coefficients. All relevant figures have been exponentiated and represent a single unit increase in the variable listed.

Model Results				
Pseudo R-squared	0.152			
AIC	50434.7384			
BIC	50559.6933			
Group: Car passenger	**Coef**	**P>(|t|)**	**L. Conf**	**U. Conf**
(Intercept)	6.51050483	0.0000	5.983466008	7.083725713
Travel Time	0.9894767608	0.0000	0.9876765632	0.9913377355
Age Group	0.7326376776	0.0000	0.7262942814	0.7390423888
Sex	0.7189302038	0.0000	0.6760420505	1.307870183
Purpose	1.016940897	0.0110	1.003807229	1.030145444
Group: Public transport				
(Intercept)	0.08868814956	0.0000	0.07157567481	0.1098872982
Travel Time	1.025670967	0.0000	1.023368871	1.027984408
Age Group	0.8941989405	0.0000	0.875552643	0.9132916573
Sex	0.5974259117	0.0000	0.5067690003	0.7043358337
Purpose	0.8752523799	0.0000	0.8450157608	0.9065582434
Group: Active				
(Intercept)	0.4239893254	0.0000	0.3759496704	2.091329525
Travel Time	0.9915211499	0.0000	0.9891591897	0.9939185672
Age Group	0.8799220104	0.0000	0.8710115862	0.8888738096
Sex	1.030668891	0.4486	0.9532291052	1.11438201
Purpose	1.10266282	0.0000	1.084154044	1.121536926

Overall, the most interesting findings based on the interpretation of these regression coefficients are:

- As travellers age, their preference to drive themselves increases
- Males generally prefer to drive themselves rather than be passengers or catch public transport however have an increased chance of choosing an Active mode of transport over driving. It is noted that there is a lack of confidence in the significance of this variable.
- Overall, purpose and travel time had little effect on the mode of transport choice. This runs counter to what the literature proposes.

Reflection and Future Work

Some limitations of this project exist and are documented here for future reproducibility and extensions to this work.

1. Interactions between variables were not explored. These should be further analysed, with a view to include variables such as income and travel cost.
2. Adding latent classes as a model input could reveal even more interactions that are currently unknown. This would involve latent class analysis and modelling.

References

[1] W. H. Greene and D. A. Hensher, "A latent class model for discrete choice analysis: contrasts with mixed logit," *Transp. Res. Part B Methodol.*, vol. 37, no. 8, pp. 681–698, Sep. 2003, doi: 10.1016/S0191-2615(02)00046-2.

[2] J. Shen, "Latent class model or mixed logit model? A comparison by transport mode choice data," *Appl. Econ.*, vol. 41, no. 22, pp. 2915–2924, Oct. 2009, doi: 10.1080/00036840801964633.

[3] O. Khan, J. Kruger, and T. Trivedi, "Developing passenger mode choice models for Brisbane to reflect observed travel behaviour from the South East Queensland Travel Survey," p. 18.

[4] C. G. Prato, K. Halldórsdóttir, and O. A. Nielsen, "Latent lifestyle and mode choice decisions when travelling short distances," *Transportation*, vol. 44, no. 6, pp. 1343–1363, Nov. 2017, doi: 10.1007/s11116-016-9703-9.

[5] A. Ardeshiri and A. Vij, "Lifestyles, residential location, and transport mode use: A hierarchical latent class choice model," *Transp. Res. Part Policy Pract.*, vol. 126, pp. 342–359, Aug. 2019, doi: 10.1016/j.tra.2019.06.016.

[6] R. Krueger, A. Vij, and T. H. Rashidi, "Normative beliefs and modality styles: a latent class and latent variable model of travel behaviour," *Transportation*, vol. 45, no. 3, pp. 789–825, May 2018, doi: 10.1007/s11116-016-9751-1.

[7] C. G. Chorus and M. Kroesen, "On the (im-)possibility of deriving transport policy implications from hybrid choice models," *Transp. Policy*, vol. 36, pp. 217–222, Nov. 2014, doi: 10.1016/j.tranpol.2014.09.001.

[8] F. J. Bahamonde-Birke, U. Kunert, H. Link, and J. de D. Ortúzar, "About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models," *Transportation*, vol. 44, no. 3, pp. 475–493, May 2017, doi: 10.1007/s11116-015-9663-5.

[9] A. Pourfarzad, "On transport mode choice in natural disasters: A case study of the 2011 Brisbane floods," Queensland University of Technology, 2016.

[10] L. L. Thurstone, “A law of comparative judgment,” *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, Jan. 1927, doi: 10.1037/h0070288.

[11] L. Yu and B. Sun, “Four types of typical discrete Choice Models: Which are you using?,” in *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics*, Jul. 2012, pp. 298–301, doi: 10.1109/SOLI.2012.6273550.

[12] F. S. Koppelman and C. Bhat, “A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.” US Department of Transportation Federal Transit Authority, Jun. 30, 2006, Accessed: Jun. 14, 2020. [Online].