The Data Professional, Assignment 3- Data Analytics Report

Transport for Wales (Tfw) is the public body who are responsible for transport and travelling in Wales. The transport system consists of personal travel by car, bicycles, and walking. Along with public travel by buses, taxis, and more. The purpose and aim for TfW, is to provide high-quality, safe, integrated, affordable, and accessible transport network (Transport for Wales 2023). By achieving these aims, economic growth will occur, mainly by creating job mobility for workers from deprived areas. £800 million was invested into a new fleet of trains for TfW named 'Happy Valley' (https://tfw.wales/about-us/our-story). With the new fleet being used on the recent travel line linking the North of Wales to Liverpool. Providing 215 trains a week, into one of the major city hubs of the UK. Sustainability has also been at the forefront for TfW. Minimising environmental effects has become one of their top priorities. Improving access to walking and cycling routes has been one of their main methods to minimising congestion and reducing the need to use fossil fuels to transport around Wales. However, economic, and environmental benefits usually act a trade-off, making it difficult to achieve both. More doesn't always mean better, as mentioned Tan (2019), congestion is responsible for 'poor air quality and urban pollution, which in turn contributes to global warming'. The impact of the recent pandemic has presented extra pressures for TfW. With O'Neil (2023) reporting that passenger numbers on trains have fallen by 33% compared to pre-pandemic levels. Mainly due to the fear of using public transport and contracting Covid-19. As a result, congestion has increased.

From April 2013 to March 2014, the national survey for Wales was conducted. The national survey was conducted amongst 22 regions in Wales, with approximately 200 interviews occurring in each region, to give a sample total of 4600. The survey involved members of the public participating in a 25-minute face-to-face interview. The purpose of the survey was to find satisfaction levels amongst various public services. These services included health, education, and the Welsh government. But most importantly for TfW, transport satisfaction. Figure 1 shows the questions asked in the national survey. As a result, TfW extracted the relevant information into their findings, shown by the 42 data tables in their report (National Survey for Wales - Transport 2014). The 42 statistical tables show the average score as a mean. For example, table 1 shows the mean score of transport satisfaction by household type, along with providing a confidence interval (CI) to show the variation around the mean.

My ULM diagram in Figure 2, shows the one-to-one relationship between the survey users' information (e.g., Gender, employment status, etc) and the 5 classes/questions that TfW extracted from the national survey. These 5 travel questions were:

1. Overall satisfaction with state of transport system in Wales.
2. Have use of a car.
3. Ease of getting to and from GP surgery
4. Ease of getting to and from Hospital
5. Feeling of safety travelling by public transport after dark

I've also referred to 'Dummy Variables' for questions with only two options. For Example, you can only be employed, or unemployed so you'd represent those answers with a 0 (False) or 1 (True).

From there, TfW can infer insights from creating comparisons. For example, is the ease of getting to hospital, dependant on the users' employment status. As we can see, 98% of employed individuals found it easy to get to hospital, while only 90% of non-employed individuals found it easy. Which makes sense because being employed means you have greater monetary power to travel to the

hospital, compared to someone who is unemployed. Similar patterns occur for ease of transport to hospital for those with access to a car. 71% of car users found it very easy to get to hospital, compared to the 48% of those without access to a car (National Survey for Wales - Transport 2014). From a policy perspective, does travel to the hospital need to be more affordable, so that those unemployed can get they receive treatment? Or does public transport need more availability to make it easier for all members of the public (e.g., extended bus and train timetable).

Overall, from the 42 data tables, the relationships show that:

Tables	Observations
1-14 -Satisfaction with the transport system in Wales	- Smaller variation in satisfaction with travel, by household type, gender, area free from heavy traffic and lifelong illness. - Higher variation in satisfaction with transport by area, use of car, deprivation scores, satisfaction with life, age, employment status, and safety when dark.
15-22 - Have use of a car.	Higher variation for those satisfied with life, employment status, and deprivation score for owning a car.
23-30 - Ease of getting to and from GP surgery.	- Negligible variation when observing ease of access to GP services by deprivation scores (WIMD), and ACORN classification. - General health, limiting illness, employment status, and access to a car, have impact on accessing GP based on high variation in dataset. - Little variation based on household type and age, but for categories 'Single pensioner (no children)' and '75 and over'.
31-38 - Ease of getting to and from Hospital.	- Little variation when observing deprivation scores and ACORN classification impacting access to Hospital. Little difference when looking at GP Access. - Access to Hospital by age varies for different ends of the scale (Young vs. Old). Coefficient of Variation (CV) larger for 16–24-year-olds could be because of smaller sample size in younger participants.
39-42 - Feeling of safety travelling by public transport after dark	Distribution of age has variation in the users feeling safe using public transport in the dark. However, CV scores are larger for 16-24 and 75+ year olds,

	potentially due to sample size of these age groups.
	• Higher variation in safety based on Gender. Males feeling safer than women when travelling in the dark.
	• Small variation for deprived areas by safety on transport after dark. Outlier group being the '20% most deprived'.

My data analysis strategy would be to group table findings together and inferring conclusive insight, with the use of different visualisations. Visualisations are one the best methods for deciphering and making sense of big data. However, different visualisations have pros and cons which need to be taken into consideration when deciding which representation to use. This is outlined below:

Type of Visualisation	Pros	Cons
Heat Maps	+Good at showing disparity between regions. +Easy to understand and gain insight from.	-Reliance on big sample set, to avoid a skewed distribution
Bar Chart	+Easy to generate and gain basic insight. +Able to show relationships between two categories.	-Good for showing correlations but not causations.
Line Graph	+Great for showing time series data. +Can use multiple lines to show difference between control groups	-Easy to clutter graph with too many lines -Doesn't explain causation of fluctuations by itself.
Histogram	+Good at showing the distribution of data around the mean. +Can be adjusted to use log values to reduce skewness in results	-Reliance on the size of the 'groups', especially when using the Age as the 'groups.

My choice of data representation and visualisations will be dependent on the variables we'll be analysing. Therefore, I would use the following visuals with the associated variables:

• Area/deprivation with Heat Maps– I will use heat maps to show correlations between areas of Wales and deprivation. Investigating to see if similar trends in the heat maps are present with deprivation scores.

• Gender/Urban or Rural with Line Graphs – For comparing the dummy variables, Line graphs would be most appropriate for showing differences in control groups whilst both using the same graph.

• Satisfaction scores and Bar Charts – I'll be using bar charts for grouped data such as satisfaction scores (Very Safe, Fairly Safe, Fairly Unsafe, Very Unsafe).

• Histograms and age – With having age on the x-axis we can look to use a range of bars to show the distribution amongst questions like 'Ease of getting to and from a GP'.

From there, I'll be looking outline correlations in the visualisations and grouping them together to prove causations for fluctuation and deviations in the dataset.
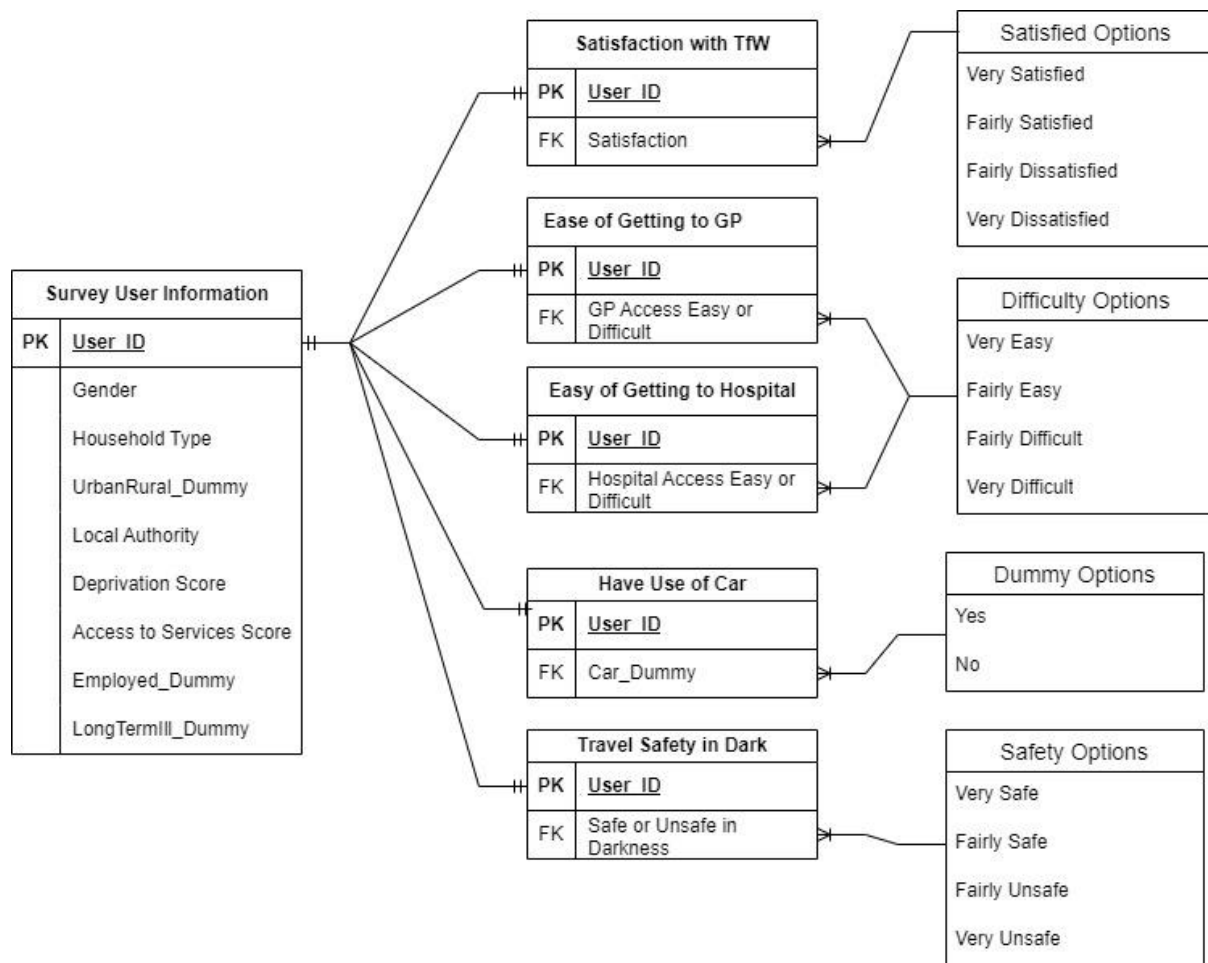
To conclude, TfW have invested heavily into the recent improvements of Welsh public transport, over the last 10 years. Economic and environmental benefits act as a trade-off, which is one of the main problems TfW faces when implementing new strategies and policies. The introduction of big data does mean that TfW are able to visualise insights, as a way of deciphering and making sense of data. My data strategy involves using the dataset from TfW, which is based on data from The National Survey for Wales 2013/14. My visualisations will help to highlight correlations between survey users' information and transport questions. Grouping visuals together will help to show causations behind some of the correlations.

Figures

Figure 1 – Questions Covered in National Survey of Wales

Questionnaire module	Areas covered
Household grid	• Confirmation of number and names of adults aged 16 or over in household (based on information provided in the electronic contact sheet) • Names of all under 16s in household • Date of birth and gender of all household members • Marital status of respondent • Relationship of respondent to all other household members • Housing tenure
Smoke alarms	• Number of smoke alarms in household and number currently working
Welsh language	• Ability of respondent to understand, speak, read and write Welsh • How regularly respondent speaks Welsh • Whether other members of the household can speak Welsh
Local area	• Proximity to natural green or blue space • Perception of belonging to local area • Feeling of safety in various situations in local area • Perception of other aspects of local area • Perceptions of people in local area
Discrimination	• Experience of discrimination, harassment or abuse
Education[21]	• Perceptions of primary school (if attended by child in household in last 12 months) • Perceptions of secondary school (if attended by child in household in last 12 months) • Interest in / awareness of how to participate in decision-making at schools attended by respondent's child / children • Mode of travel and distance to school • Aspirations for child when they reach 16 and leave school (if child in household aged 11-15) • Involvement in activities with children (asked for children aged 3-7 and 8-17)

Questionnaire module	Areas covered
Health demographic questions	• General health condition • Experience of physical or mental illness and impact on day-to-day activities
Health	• Perceptions of the NHS and local health services • Whether seen a GP or family doctor about own health in last 12 months and, if yes, experience of this • Whether had an appointment at an NHS hospital in the last 12 months and, if yes, experience of this
Satisfaction with local authority services	• Perception of whether council provides high quality services • Sources of information about council services • Perception of ability to make decisions with regard to local area
Welsh Government	• Overall satisfaction with the way the Welsh Government is doing its job • How much respondent has seen or heard about work of Welsh Government in last 12 months
Satisfaction with public services	• Satisfaction with education / health / transport services[22]
Wellbeing (EU-SILC questions)	• Satisfaction with various aspects of life • Estimation of positive and negative feelings over the past four weeks • Whether respondent has a support network • Trust in people and national institutions
Demographic questions	• Country of birth, national identity and ethnicity • Sexual identity • Religion • Qualifications • Access to a car • Economic status / employment questions
Financial inclusion	• How well respondent has been keeping up with bills and credit commitments • Whether used any debt advice / support services in last 12 months

Questionnaire module	Areas covered
Internet use	• Internet access and use • Barriers to internet use
Public service communications	• Whether visited any of a range of Government or other public service websites in the last 12 months • Barriers to accessing these websites • Preferred method for contacting/ using public services
Active travel	• Mode of transport used to get to specified destinations • Modes of transport used in past 7 days • Purpose of most recent 'active travel' • Whether respondent can ride / owns a bicycle and if so how frequently used as a means of transport in last 3 months • Barriers to using a bicycle • Whether respondent walked more than 5 minutes as a means of transport in last 3 months • Barriers to walking
Recontact and data linkage	• Whether happy for personal details to be passed to Welsh Government for the purpose of future research • Whether willing to have survey answers linked to information held by the NHS and other public service organisation[23]

Figure 2 – UML Entity Relationship Diagram

References

National Survey for Wales (2014) Available at:
http://doc.ukdataservice.ac.uk/doc/7767/mrdoc/pdf/7767_nsw_2013-14_technical_report_en.pdf
[Accessed: 2nd July 2023].

National Survey for Wales - Transport (2014) Table of Contents Available at:
https://www.swansea.gov.uk/media/3570/National-Survey-for-Wales-2013-14-LA-tables/excel/National_Survey_for_Wales_2013-14_LA_tables.xlsx?m=1632908655937 [Accessed: 9th July 2023].

Tan, M. (2019) Challenges of Public Transport. Available at:
https://medium.com/@SupaDynaMike/challenges-of-public-transport-35f4ca5d9588 [Accessed: 1st July 2023].

O'Neil, R. (June 16, 2023) Passenger numbers on Welsh trains well down on pre-pandemic levels.
Wales Online. Available at: https://www.walesonline.co.uk/news/wales-news/passenger-numbers-welsh-trains-well-27125006 [Accessed: 1st July 2023].

Transport for Wales (2023) Homepage. Available at: https://tfw.wales/about-us [Accessed: 2nd July 2023]