Topic 1

Exploring the Predictive Power of Demographic and Health Variables on Autism Spectrum Disorders (ASD) Traits

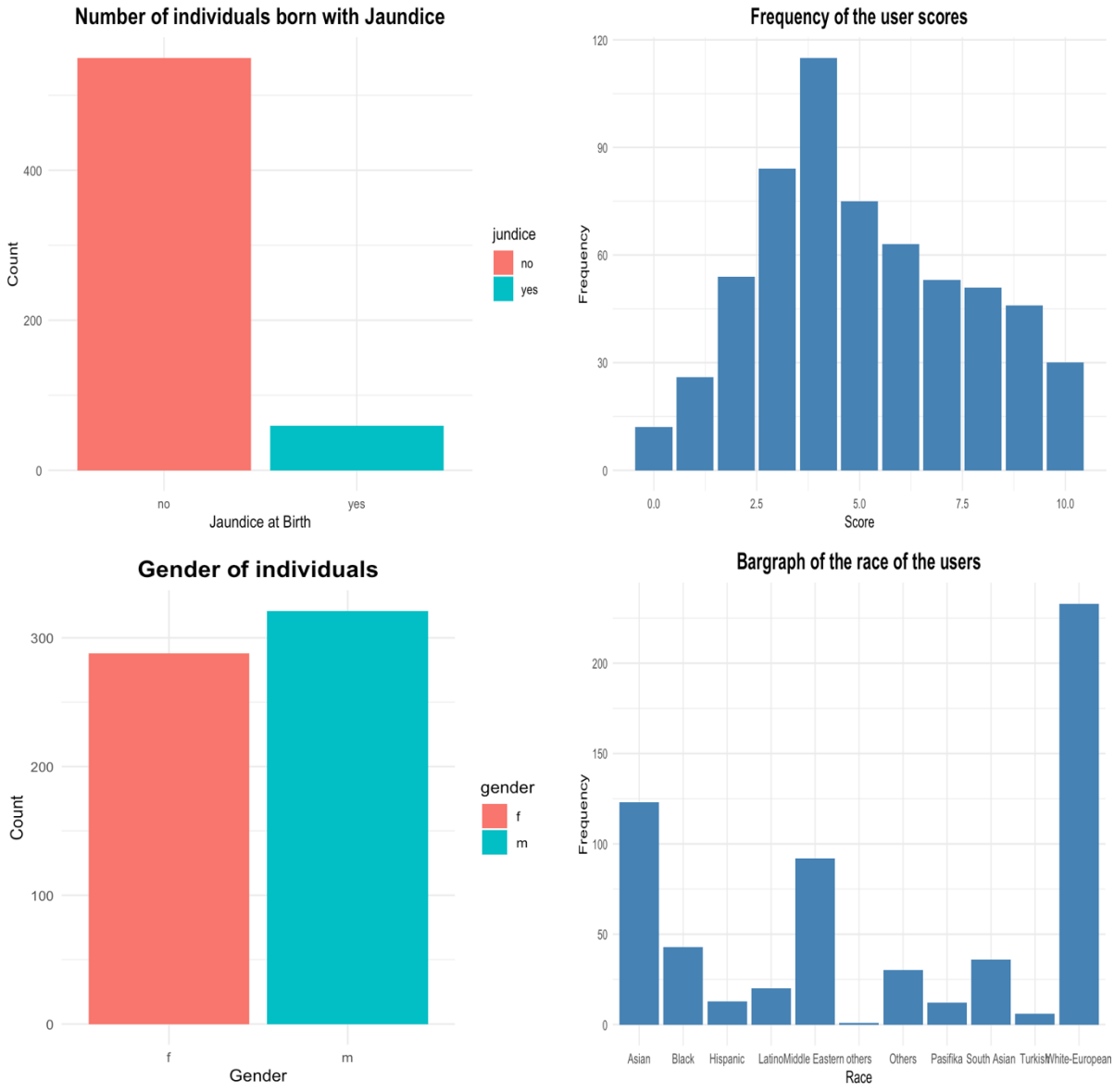
Autism Spectrum Disorders (ADS) is a neurological and developmental disorder, which characterised by difficulties in social interaction, communication and learning ability. Autism typically develops in early childhood, with most children showing symptoms in the first two years of their lives. According to the autism research charity Autistica, 70% of the autistic people also experience mental health problems such as anxiety, depression and attention deficit hyperactivity disorders. Earlier detection and intervention of ADS traits can significantly help improve the wellbeing and mental health of the autistic people as timely supports and health cares can be provided. This study aims to explore the predictiveness of various demographic and health factor, with the goal of providing timely alerting signals for the parents to seek further clinic diagnosis for their children.

The dataset that this study will be based on is a set of ADS screening dataset provided by Dr. Fadi Fayes Thabtah from the Department of Digital Technology, Manukau Institute of Technology, New Zealand. The dataset is collected through a mobile ASD screening app, which gathers personal information such as age and sex and presents users with 10 questions assessing traits commonly associated with autism. Numerous studies have evaluated the accuracy of this screening method, demonstrating that the results from these questions are reliable indicators of ASD traits.

The dataset consists of 609 observations after removing the observations with missing values. The variables include the age, sex, country of residence, race, whether the user was born with jundice, whether any immediate family member has diagnosed with ADS. The dataset also contains the total score obtained for the 10 questions which each has only one answer Yes=1 or No=0. The user is classified having autism traits if the total score is higher than 3.

Below shows an overview of the dataset:

- There are few users who were born with jaundice than those who were not born with jaundice,
- From the frequency of the users with different values of scores for the 10 questions, we can see that it shows a normal distribution, with most users scoring between 3 and 7. Based on the classification criteria, a significant proportion of users is classified as having autism traits,
- The dataset has a fair gender distribution, with slightly more males than females,
- In terms of race distribution, white-European group has the highest representation and Asian and Middle-east groups also have notable frequencies.



Possible multivariate statistical methods that will be applied are:

1. Clustering analysis: The use of clustering analysis aims to cluster the observations according to their demographical and health characteristics in order t**o** analyze whether the clusters have any relationship with the ADS status classified by the app. For instance, which cluster has higher number of autism cases.
2. Latent class analysis: Since the variables are mostly categorical, latent class analysis can also be used to identify the probability that a user would be diagnosed with ADS.
3. Generalized linear model: Since the response is whether the person has ADS or not, binary logistic regression can be fit to predict the likelihood of diagnosing ADS based on these variables.

Topic 2

Validation of self-reported barriers among foreign-born individuals in European countries

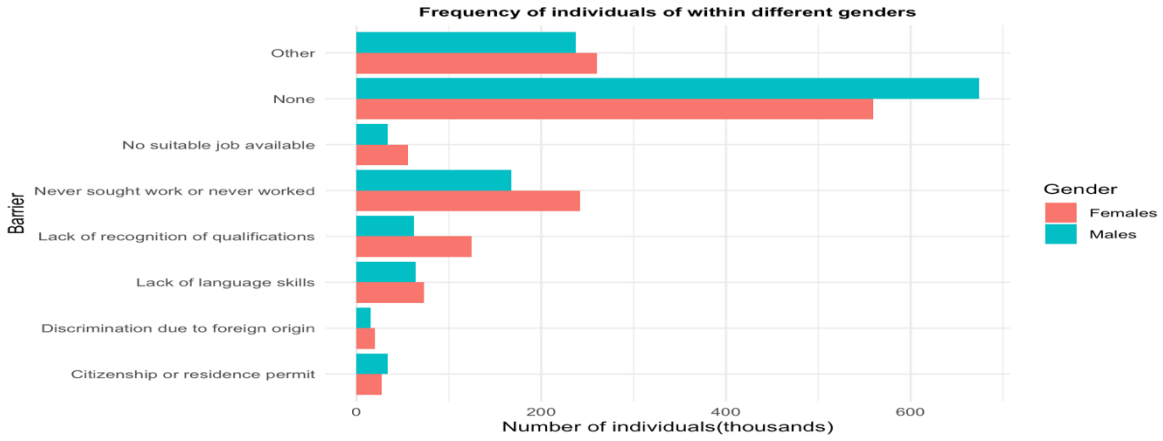
The participation of diverse labour force is crucial for the economic growth and innovation development of a country as a wider range of perspectives, knowledge, skills and experience can be brought into the country. Foreign-born individuals, who are also part of the labour market, often facing unique challenges in their host countries, including legal barriers, discrimination and language proficiency. This study aims to predict the likelihood of experiencing specific barriers by analyzing the factors such as educational attainment, city of residence, age and sex. In the process of analyzing, The finding would also provide some meaningful insight onto whether the data collected for migration studies are reliable so that to improve the efficiency of their use in policymaking, which aiming to support the foreign-born individuals in the workplace.

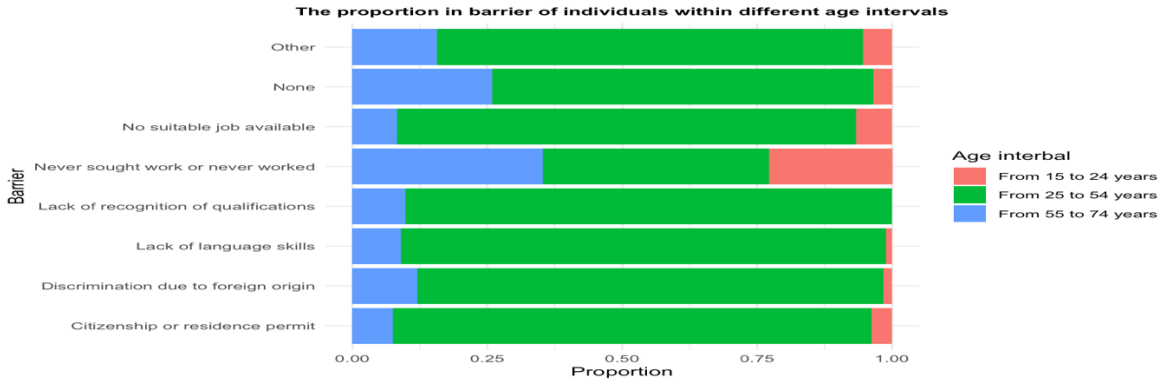
This study will utilize the dataset from Eurostat, which is collected in year 2021 from 27 countries which consist of both EU members and non-EU members such as Switzerland, Spain, Italy and Austria. The dataset comprises 2105 groups of observations with each categorized by different combination of gender, age, country of birth, educational attainment and the perception of types of discrimination experienced by the labour force. In this study, the predictors will be the gender (Male or female), the age (from 15 to 24 years, from 25 to 54 years and from 55 to 74 years), educational attainment (Less than primary, primary and lower secondary education (levels 0-2), Upper secondary and post-secondary non-tertiary education (levels 3 and 4) and Tertiary education (levels 5-8)), perception of discrimination (“Lack of language skills”, ”None”, “No suitable job available", "Never sought work or never worked", "Other", "Discrimination due to foreign origin", "Lack of recognition of qualifications" and "Citizenship or residence permit") and country of birth ("EU27 countries (from 2020) except reporting country", "Foreign country" and "Non-EU27 countries (from 2020) nor reporting country").

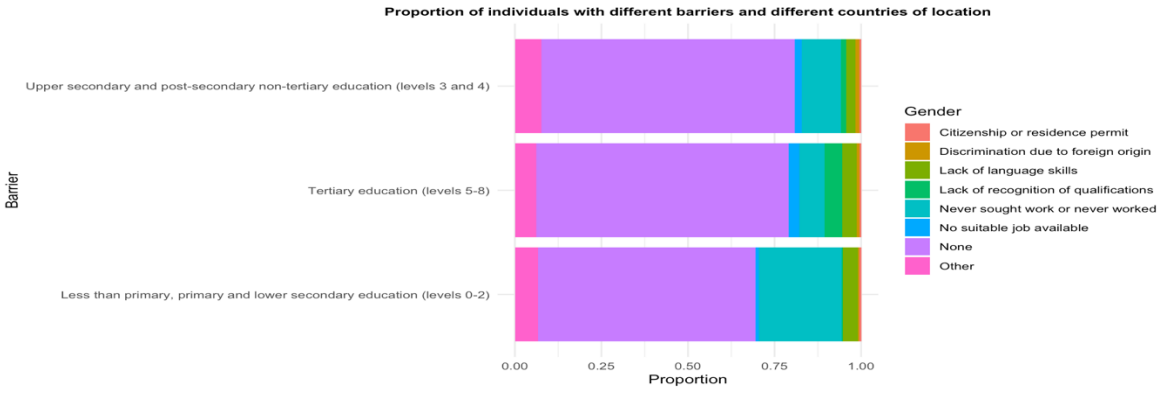
Below are the overview of the dataset:

- In the first plot below, females have a significantly higher proportion in barrier “Never sought work or never worked”. This could indicate there’s gender-related labor market participation differences, potentially due to cultural norms. While in other barriers, the proportion of males and females are similar.
- In the second plot below, individuals with lower educational attainment (levels 0-2) face more challenges such as "No suitable job available" and "Lack of language skills". Even among individuals with tertiary education, some still face issues like qualification recognition and foreign origin discrimination, this means that as shown by the dataset, tertiary education does not eliminate all the barriers.

- In the third plot below, the majority of individuals belong to the 25-54 age group, which is expected as this represents the core working population. Younger individuals (15-24 years) are more likely to have never sought work or worked, which is likely due to ongoing education or lack of work experience. Meanwhile, older individuals (55-74 years) appear to face difficulties in finding suitable jobs, possibly due to skill mismatches or age-related employment challenges.







The potential methods to apply since the variables are mostly categorical are:

1. Logistic regression: generalized multivariate regression can be used to predict the likelihood of individuals experiencing different types of barriers by assessing the effect of all the factors on the response.

2. Latent class analysis which clusters the observations based on their characteristics. After clustering the observations, each cluster can be analysed to check if they have reported the same types of discrimination.

Topic 3

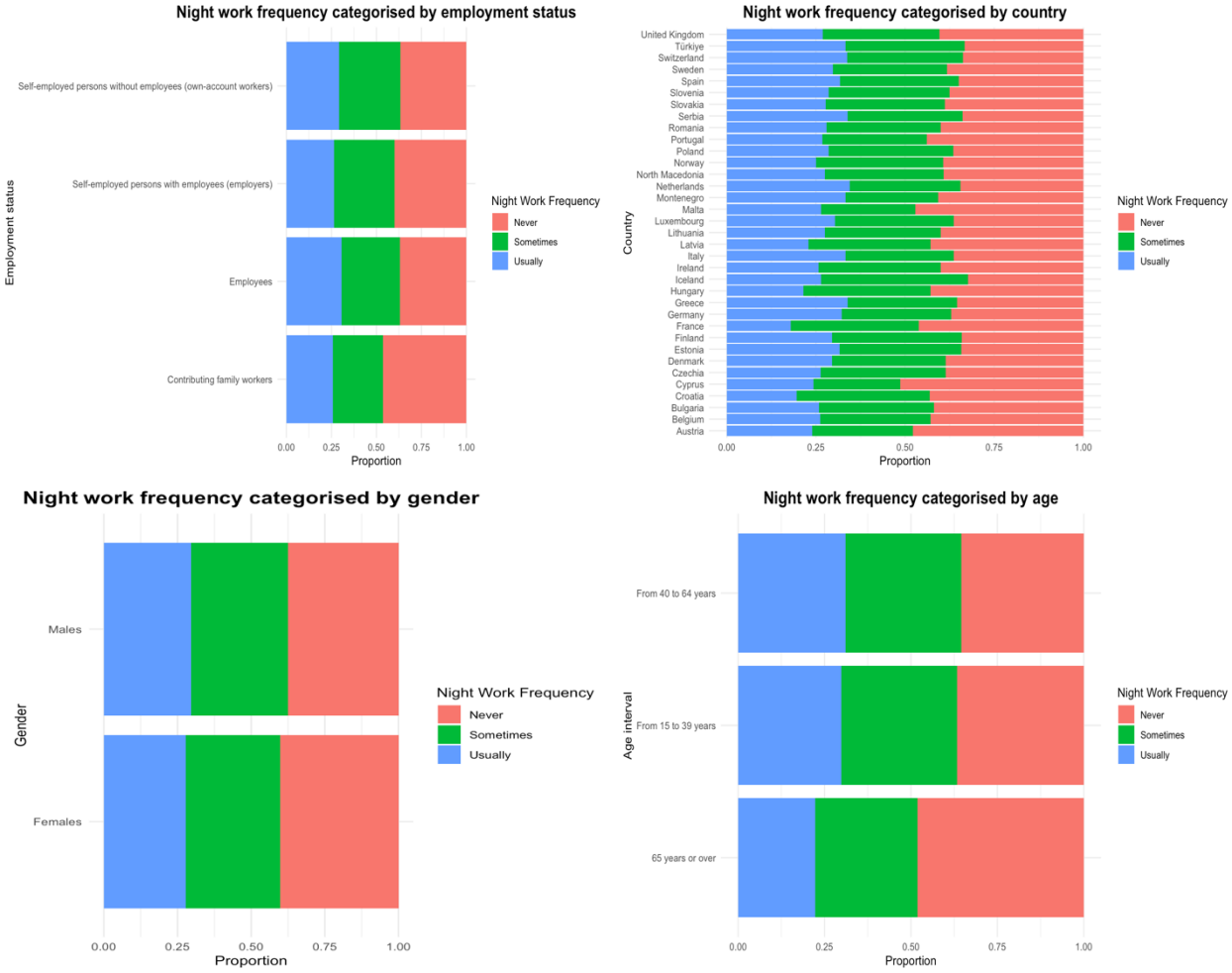
Understanding the Dynamics of Night Work: Examining Socioeconomic, Demographic, and Occupational Influences Across Countries

Working pattern has undergone more diverse transformation over recent decades. Due to the rise of 24/7 economies, there is an increasing need for the employees to work at night shifts and overtime in order to maintain the continuous operation of some businesses. In additional, the technology advancement has also accelerated the new pattern of remote working, providing the employees flexibility to work at night. This study investigates the influence of demographic, socioeconomic and occupational factors on the frequency of working at night using a multivariate analysis approach with the goal of providing more useful insights for the labour organizations and policymakers who seeking to address employment inequality and wellbeing of the citizens who frequently work at night.

This study will utilize the dataset from Eurostat, which reports the percentage of employed persons working at night, categorized by sex, age and professional status. The data is collected from year 1923 to 2023 by Eurostat from 36 countries which consist of both EU members and non-EU members. In this study we will only take the data from the year of 2019 which comprises 1658 groups of observations.

Below shows an overview of the dataset:

- The proportion of males and females working at night follows similar pattern and the plot shows that there is a significant proportion of employed who have never worked at night and a small proportion who have worked at night.
- In the plot of “Night working pattern categorized by employment status”, it is shown that “Employees” and “self-employed individuals without employees” have a similar distribution, with most never working at night. “Self-employed persons with employees” has a greater proportion of working at night.
- In the plot of “Night working pattern categorized by age”, it shows that workers such as those between 15 and 64 have high proportion of working at night compared to those who are beyond 64. This suggests that night work is more common at early career stages or for older workers.
- In the plot of “Night working pattern categorized by country”, it shows that there is clearly a country-level differences in the night work frequency. Countries such as Cyprus, Malta and Austria have less labour force working at night while countries such as Serbia, Switzerland and Finland have more labour force working at night.



When applying the generalized linear model in order to estimate the effect of each variable on night work pattern, the predictors of interest are age (from 15 to 39 years old, from 40 to 64 years and 65 years or over), sex (male and female), geographical location (36 European countries), employment status(Contributing family workers, Employees, Self-employed persons without employees (own-account workers), Self-employed persons with employees (employers)). The response of interest would be working at night frequency (Never, Sometimes, Usually).

When performing cluster analysis on the dataset in order to understand the segmentation of countries in terms of night work pattern, the night work frequency would become one of the

predictors and the response of interest would be the clusters of countries with similar night work pattern.