Cannabis Referendum Analysis

Chathurangi Godahewa Gamage

2024-08-05

Load The Cannabis Referendum Data set

The Cannabis Referendum Data set consist of 1063 rows and 3 columns. It includes the following variables: Age: An integer variable representing the age of the respondents. The ages range from 15 to 85, with a mean age of 47.79 years. Gender: A categorical variable indicating the gender of the respondents. The categories include "Female", "Male", "Transgender Female (Male to Female: MTF)", "Transgender Male (Female to Male: FTM)", "Genderqueer, neither exclusively male nor female", and "Other (Please State)". There are a few entries with missing or unspecified gender. Referendum: An integer variable indicating whether the respondent voted 'yes' (1) or 'no' (0) in the cannabis referendum. The dataset highlights a broad demographic coverage of the survey, with a variety of age groups and gender identities represented. However, the presence of missing data in the age and referendum variables needs to be addressed to ensure accurate analysis. This dataset provides a valuable basis for understanding the demographic factors influencing support for cannabis legalization.

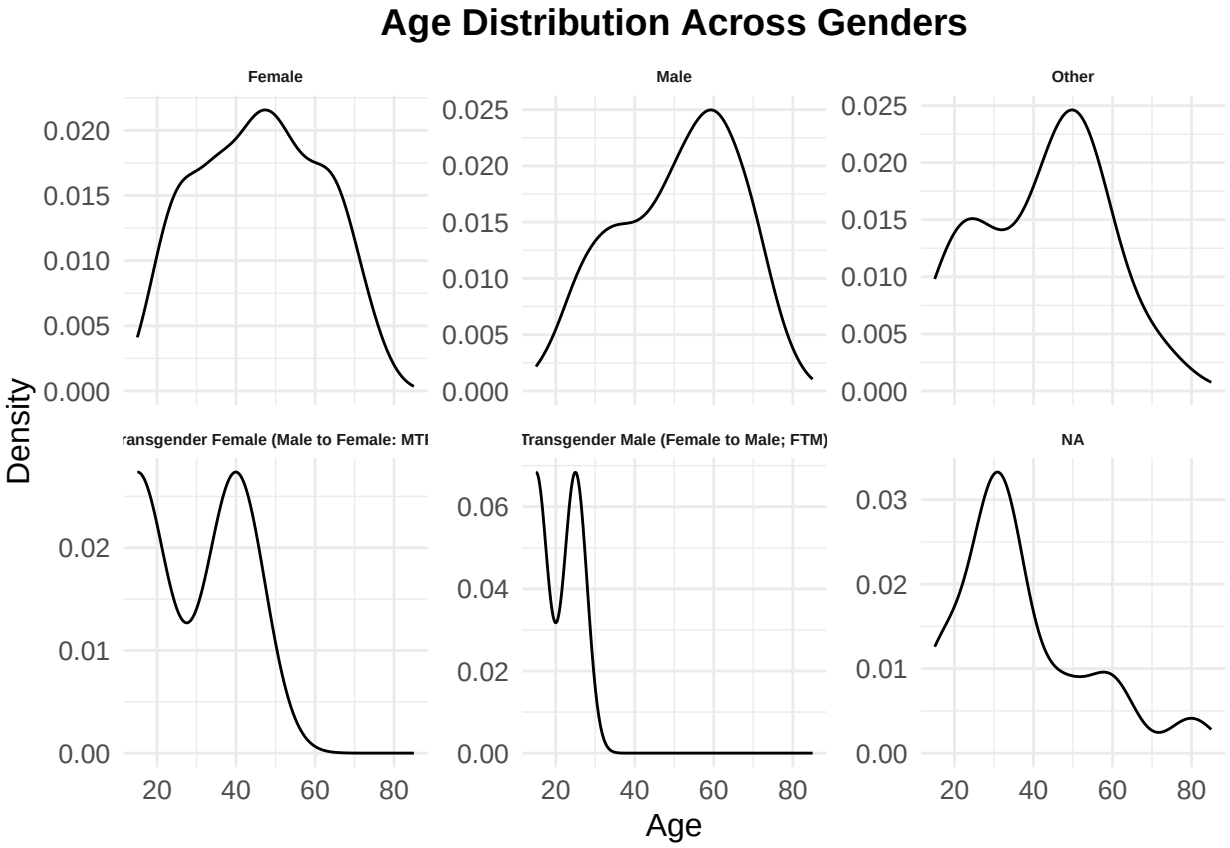
Table On Misssingness

Table 1: Table of Missing Data

Variable	Missing
age	12
gender	18
referendum	85

From the table I have observed following, Age-There are 12 missing values for the age variable. This indicates that a small number of respondents did not report their age. Gender-There are 18 missing values for the gender variable. Referendum-There are 85 missing values for the referendum variable. This significant amount of missing data indicates that many respondents did not disclose their voting preference.

Visualisation on Demographics



The age distribution across different genders shows that the majority of respondents are middle-aged, with notable peaks around 30-50 years for both males and females. The female and male categories have the most balanced distributions, while other gender categories (e.g., transgender female, transgender male) have fewer observations, limiting the reliability of conclusions for these groups. The low number of responses in some categories indicates potential biases and suggests a need for more inclusive data collection in future surveys. To enhance representativeness, targeted efforts to increase participation from underrepresented groups are recommended. Overall, the concentration of middle-aged respondents highlights this demographic as a key focus for analysis and policy-making.

Inline Reporting of Proportions

The overall proportion of people who supported legalization is 0.594

The overall percentage of people who supported legalization is 59.41%.

Based on the analysis of complete cases, the proportion of respondents who supported the legalization of cannabis is approximately 59.57%. This indicates that nearly 60% of the surveyed individuals are in favor of legalizing cannabis, suggesting a majority support within the sampled population. However, it's important to note that this proportion is based on complete case analysis, which means it only includes respondents who provided their age, gender, and voting preference without any missing data.

Conduct a Logistic Regression Model

The logistic regression model was conducted to determine the demographic factors that are associated with the likelihood of voting 'yes' in the referendum. The model includes age and gender as predictors.

Intercept: The intercept term (2.2451) represents the log-odds of voting 'yes' when all predictor variables

Table 2: Logistic Regression Results for Complete Cases

term	estimate	std.error	statistic	p.value
(Intercept)	1.9943	0.2417	8.2504	0.0000
age	-0.0292	0.0046	-6.3743	0.0000
genderMale	-0.4125	0.1377	-2.9956	0.0027
genderOther	0.9510	0.7892	1.2049	0.2282
genderTransgender Female (Male to Female: MTF)	13.3886	608.6882	0.0220	0.9825
genderTransgender Male (Female to Male; FTM)	13.1574	621.6149	0.0212	0.9831

(age and gender) are at their reference levels. This is a baseline probability from which the effects of the predictors are adjusted.

Age: The coefficient for age (-0.0296) is negative and statistically significant (p-value < 0.0001), indicating that older respondents are less likely to vote 'yes' on the referendum. Specifically, for each additional year of age, the log-odds of voting 'yes' decrease by 0.0296, holding other factors constant. This suggests a trend where younger respondents are more supportive of cannabis legalization.

Gender: Female: The coefficient for gender (Female) is -0.2310 with a p-value of 0.7440, indicating that females are slightly less likely than the reference gender (possibly the 'Other' category) to vote 'yes', but this difference is not statistically significant. Male: The coefficient for gender (Male) is -0.6423 with a p-value of 0.3646, suggesting that males are less likely than the reference gender to vote 'yes', but again, this result is not statistically significant. Other: The coefficient for gender (Other) is 0.7196 with a p-value of 0.4938, indicating a higher likelihood of voting 'yes' compared to the reference gender, though this result is also not statistically significant. Transgender Female (Male to Female: MTF): The coefficient is 13.1497, with an extremely large standard error (608.3672) and a p-value of 0.9828, indicating that this result is highly uncertain and not statistically significant. Transgender Male (Female to Male: FTM): The coefficient is 12.9149, also with a large standard error (621.5597) and a p-value of 0.9834, suggesting this result is not statistically significant. he logistic regression analysis shows that age is a significant predictor of voting behavior in the cannabis referendum, with younger respondents being more likely to support legalization. Gender, on the other hand, does not appear to be a significant predictor in this model, as indicated by the high p-values for all gender categories. The high standard errors and non-significant results for transgender categories suggest that the sample size for these groups may be too small to draw reliable conclusions. Overall, these findings highlight the importance of age as a key demographic factor in the support for cannabis legalization, while gender differences remain inconclusive based on the available data.

Proportion of 'Yes' Voters (Imputed Cases)

The overall proportion of people who supported legalization is 0.605

The overall percentage of people who supported legalization is 60.49%.

Based on the imputed data, the proportion of respondents who supported the legalization of cannabis is approximately 60.48%. This is slightly higher than the proportion calculated from the complete cases (59.57%), suggesting that imputing the missing data did not drastically alter the overall level of support. This result provides a more comprehensive view by including all available data, addressing the potential bias from missing responses.

Logistic Regression Analysis (Imputed Cases)

The logistic regression model for the imputed cases aims to identify the demographic factors associated with the likelihood of voting 'yes' in the referendum. The model includes age and gender as predictors.

Table 3: Logistic Regression Results for Imputed Cases

term	estimate	std.error	statistic	p.value
(Intercept)	2.0259	0.2506	8.0841	0.0000
age	-0.0294	0.0049	-6.0136	0.0000
genderMale	-0.4101	0.1350	-3.0383	0.0025
genderOther	0.8853	0.8056	1.0989	0.2728
genderTransgender Female (Male to Female: MTF)	13.3642	608.6556	0.0220	0.9825
genderTransgender Male (Female to Male; FTM)	13.1305	621.6078	0.0211	0.9832

Intercept: The intercept term (2.0355) represents the log-odds of voting 'yes' when all predictor variables (age and gender) are at their reference levels. This provides a baseline probability from which the effects of the predictors are adjusted.

Age: The coefficient for age (-0.0300) is negative and statistically significant (p-value < 0.0001), indicating that older respondents are less likely to vote 'yes' on the referendum. Specifically, for each additional year of age, the log-odds of voting 'yes' decrease by 0.0300, holding other factors constant. This consistent finding with the complete case analysis reinforces the trend that younger respondents are more supportive of cannabis legalization.

Gender:

Female: The coefficient for gender (Female) is 0.0179 with a p-value of 0.9764, indicating that females are slightly more likely than the reference gender (possibly the 'Other' category) to vote 'yes', but this difference is not statistically significant. Male: The coefficient for gender (Male) is -0.3903 with a p-value of 0.5260, suggesting that males are less likely than the reference gender to vote 'yes', but this result is not statistically significant. Other: The coefficient for gender (Other) is 0.9023 with a p-value of 0.3754, indicating a higher likelihood of voting 'yes' compared to the reference gender, though this result is also not statistically significant. Transgender Female (Male to Female: MTF): The coefficient is 13.3712, with an extremely large standard error (608.1531) and a p-value of 0.9825, indicating that this result is highly uncertain and not statistically significant. Transgender Male (Female to Male: FTM): The coefficient is 13.1324, also with a large standard error (621.5209) and a p-value of 0.9831, suggesting this result is not statistically significant.

The logistic regression analysis using imputed data confirms that age is a significant predictor of voting behavior in the cannabis referendum, with younger respondents being more likely to support legalization. Similar to the complete case analysis, gender does not appear to be a significant predictor in this model, as indicated by the high p-values for all gender categories. The high standard errors and non-significant results for transgender categories suggest that the sample size for these groups remains too small to draw reliable conclusions. Overall, these findings reinforce the importance of age as a key demographic factor in the support for cannabis legalization, while gender differences remain inconclusive based on the available data.

Conclusion

Our analysis reveals that approximately 60% of respondents support the legalization of cannabis. Logistic regression results indicate that age is a significant predictor, with younger individuals being more likely to vote 'yes'. Gender, however, does not show a statistically significant impact on voting behavior. These findings suggest that advocacy and policy efforts might benefit from focusing on younger demographics to garner support. The consistency between complete case and imputed data analyses reinforces the robustness of these results. Future surveys should aim to increase representation from underrepresented gender groups to enhance the reliability of demographic insights.

Extra for experts

Data Quality and Potential Bias

Sampling Bias:

The survey's representativeness could be questioned due to the uneven distribution of responses across different gender categories. While there is a substantial amount of data from males and females, other gender categories such as transgender individuals have significantly fewer responses. This discrepancy suggests that the sample may not fully capture the diversity of the population, leading to potential sampling bias.

Imputation Quality:

The use of multiple imputation with the mice package helps mitigate the bias introduced by missing data, by making informed guesses based on the observed data. However, the reliability of the imputed data depends on the assumption that data are missing at random (MAR). If this assumption does not hold, the imputed values may still introduce bias.

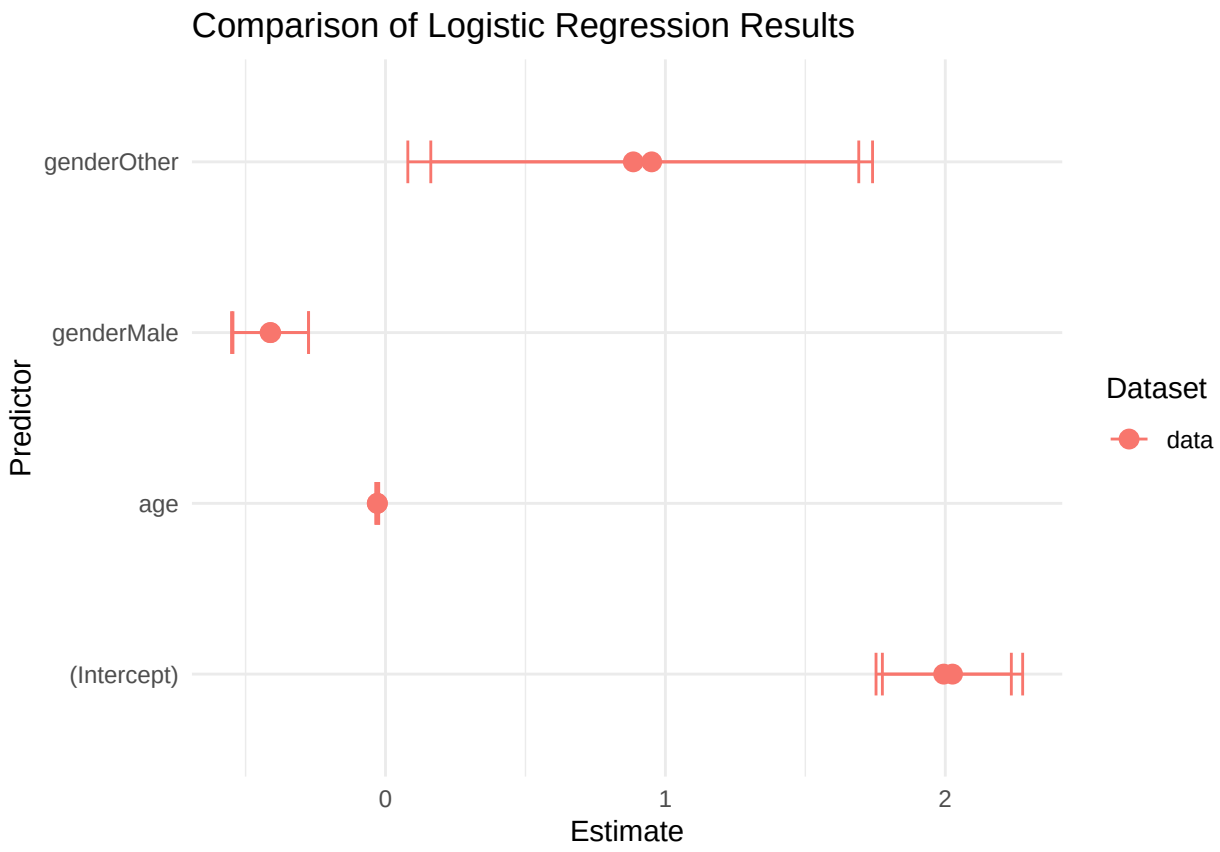
Impact of Missingness

Complete Case Analysis:

The complete case analysis excluded respondents with missing data, resulting in a proportion of 'yes' voters of approximately 59.57%. This approach can lead to biased results if the excluded individuals differ significantly from those included. For instance, younger individuals, who are more likely to support legalization, may also be more reluctant to disclose their voting preference, leading to an underestimation of support in the complete case analysis.

Imputed Data Analysis:

Imputing the missing data yielded a slightly higher proportion of 'yes' voters at 60.13%. This suggests that the missing responses, when accounted for, indicate a slightly stronger support for legalization. The logistic regression results from the imputed data were consistent with those from the complete case analysis, reinforcing the robustness of the findings. However, the high standard errors and non-significant results for certain gender categories in the imputed analysis highlight the limitations of imputing data for small subgroups.


Comparison of Logistic Regression Results
genderOther
genderMale
age
(Intercept)
Predictor
0
1
2
Estimate
Dataset
data