Week 4 - Imputation

David Ewing (82171165)

2024-08-05

Introduciton - Imputation Tecnique

Imputation is a common Data Engineering Technique for cleaning up data. This report is the results of a lab that compares accuracy of different imputation techniques.

A client representing the Green Party collected data about the Cannabis referendum in 2020. They wish to know from their data if it is worth pursuing a citizen initiated referendum in the next election cycle.This report is to address the following:

- What proportion of people in the sample supported legalisation?
- Who in the sample supported legalisation?

This report os organised as follows:

- Table on missingness
- Visualisation on demographics
- Inline reporting of proportions
- Tables of each logistic regression
- A clear conclusion
- Extra for experts for those game enough

Data (and Table on Missingness)

The data is a provided csv file containing : "Referendum Survey Lab Version". This CSV file containing survey responses with Age, Gender, and whether someone voted yes (1) or no (0) in the referendum. There was a considerable amount of missingness in the data. It is believed this is because some people, particularly younger people, were reluctant to give their voting preference. Table 1 shows the Missingness. The Age across Genders for the participants is shown in the graphic that follows.

Table 1: Missingness

Variable	Missing Count
age	12
gender	18
referendum	85

Visualisation of Demographics

Distribution of Age Across Genders

Count

Age
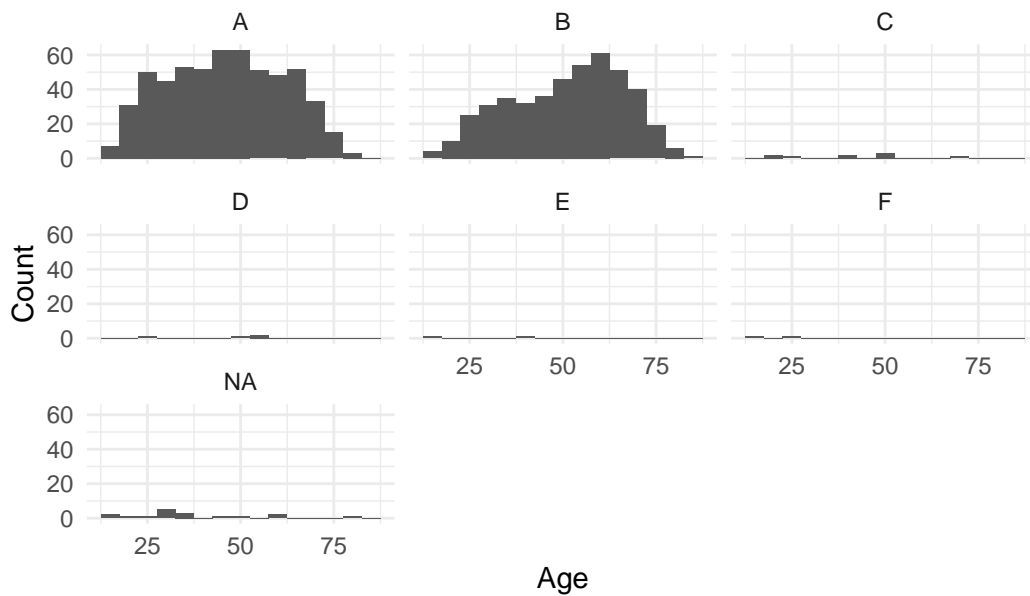
Table 2: Gender Legend

Label	Description
A	Female
B	Male
C	Genderqueer, neither exclusively male nor female
D	Other (Please State)
E	Transgender Female (Male to Female: MTF)
F	Transgender Male (Female to Male; FTM)
G	NA

Table 3: logistic regression results

term	estimate	std.error	statistic	p.value
(Intercept)	1.99	0.24	8.23	0.00
Age	-0.03	0.00	-6.36	0.00
Queer (Gender)	1.35	1.08	1.25	0.21
Male	-0.41	0.14	-3.00	0.00
Other (Gender)	0.25	1.23	0.21	0.84
Transgender Female (MTF)	13.39	608.77	0.02	0.98
Transgender Male (FTM)	13.16	621.63	0.02	0.98