Data-Driven Reduction in Gambling Harm for Stronger Community Resilience

Datasets used, purpose, and cleaning notes

All gambling datasets used were only related to the expenditure from gaming machines at venues in Victoria. Hence, the analysis from this paper should only be extrapolated to gaming machines at venues in Victoria, and should not be used for other contexts such as online gambling unless deemed appropriate.

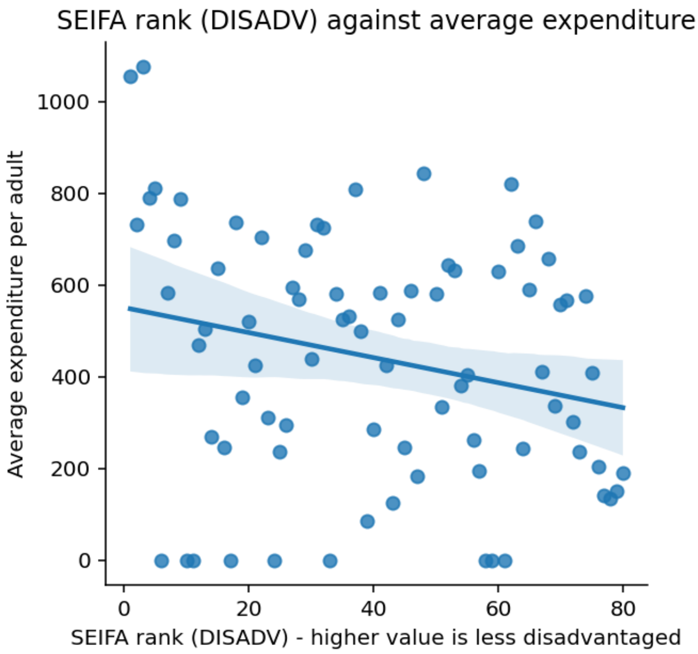
The yearly gaming expenditures across LGAs and quarterly rental report datasets were used to construct linear regressions. In the yearly gaming data, we looked at SEIFA scores (a linear combination of various socioeconomic variables) and unemployment rates across LGAs (Local Government Areas) for the 2023-24 financial year, and we also looked at the yearly median rent for each LGA in the 2023-24 financial year.

The yearly gaming data had detailed data on the gaming expenditures for each financial year. Some extraneous rows were removed from the detailed data sets for the yearly gaming data. A similar procedure was performed for the quarterly rental data. The yearly gaming data and quarterly rental data also had different names for LGAs, so to combine them in Python the LGAs needed to be renamed, with some manual modifications to the LGA names in the rental Excel data file. For more details see the "Code for analysis.pyscript.

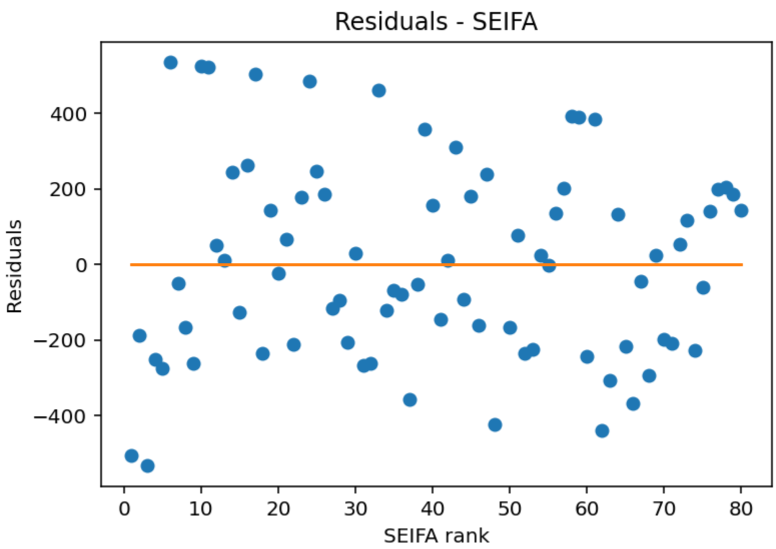
The monthly gaming expenditures across LGAs dataset was used to understand monthly trends and figure out potential times within the year when people were more likely to spend more on gambling. Time series graphs were constructed and analysed to understand this. As only two years of data are currently available, more data is likely needed to determine whether the observed trends are seasonal or not. Similar to the yearly gaming data, detailed data for each financial year was stored in separate sheets, and some extraneous rows were removed from these sheets.

Linear regression

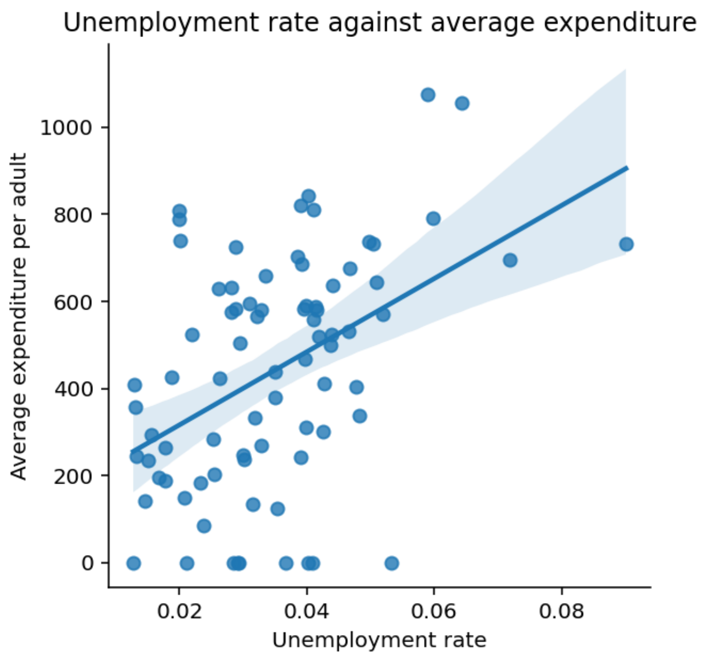
To understand if socio-economic factors in general would affect gambling, a simple linear regression was first created for SEIFA DIS rank (where each LGA is ranked by their SEIFA DIS rank, with higher ranks indicating less disadvantage (e.g. an LGA ranked 50 is less disadvantaged on average than an LGA ranked 20)). Note that SEIFA DIS is a linear combination of various variables that reflect overall socioeconomic disadvantage, so as per advice from the ABS here the SEIFA DIS indexes themselves were not used, and the ranking system was used instead. SEIFA DIS rank was considered in isolation for this initial foray, as using SEIFA with other socioeconomic variables like unemployment rate or rent would induce multicollinearity due to SEIFA being a linear combination of said variables (amongst others).



Figur 1: Linear regression for SEIFA DIS rankings for each LGA against average gaming expenditure per person.



Figur 2: Plot of residuals against SEIFA DIS rankings.

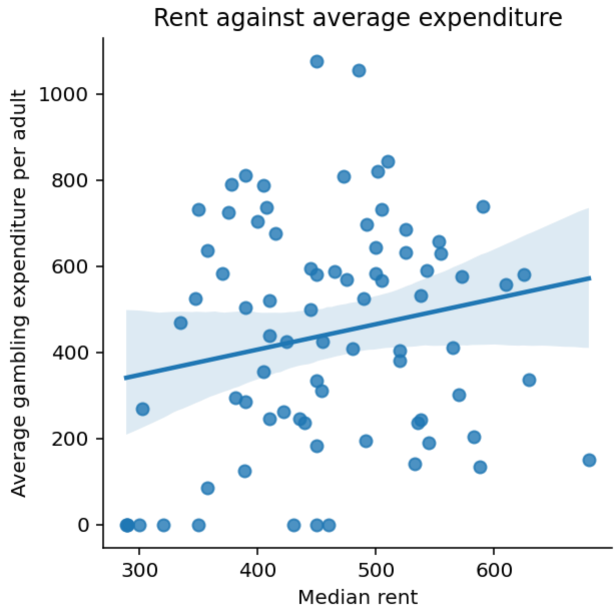


Figur 3: Linear regression model for unemployment rates against average gaming expenditure per person.

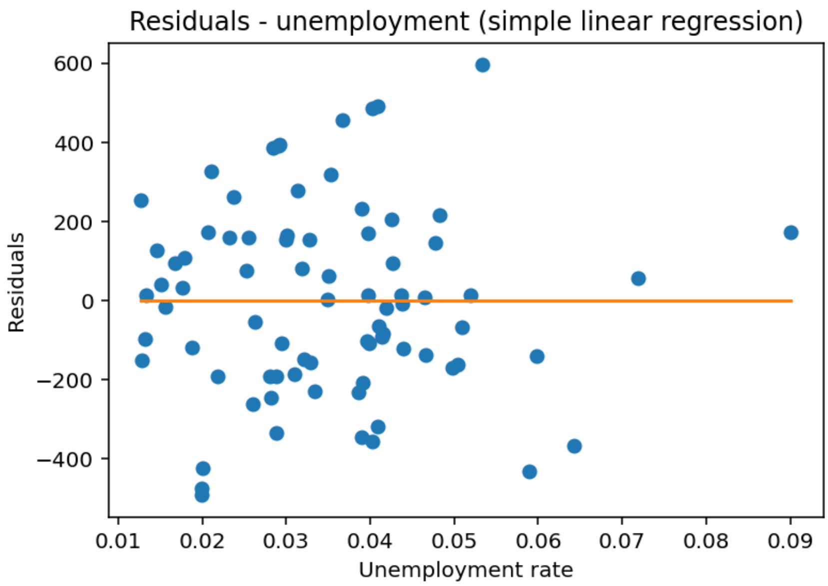
As Figure 1 shows, there exists a weak negative correlation between SEIFA DIS rankings and average gambling expenditure, suggesting that for areas less socio-economically disadvantaged on average there tends to be a higher spend on gambling. Figure 2 shows that the plot of residuals against the SEIFA DIS rankings is approximately symmetrical about 0, indicating that the data is approximately linear and that a linear regression is appropriate.

To determine possible risk factors for gambling, we looked at the variables that contributed the most to SEIFA DIS scores and found that low rent and unemployment were potential factors that contributed to the scores. To investigate this further, a simple linear regression was also performed for each of rent and unemployment. Figures 3 and 4 show a positive correlation between unemployment rate and average expenditure, as well as rent against average expenditure. This suggests that unemployed people are more likely to spend more on gambling, and similarly so for people living in LGAs with higher rents.

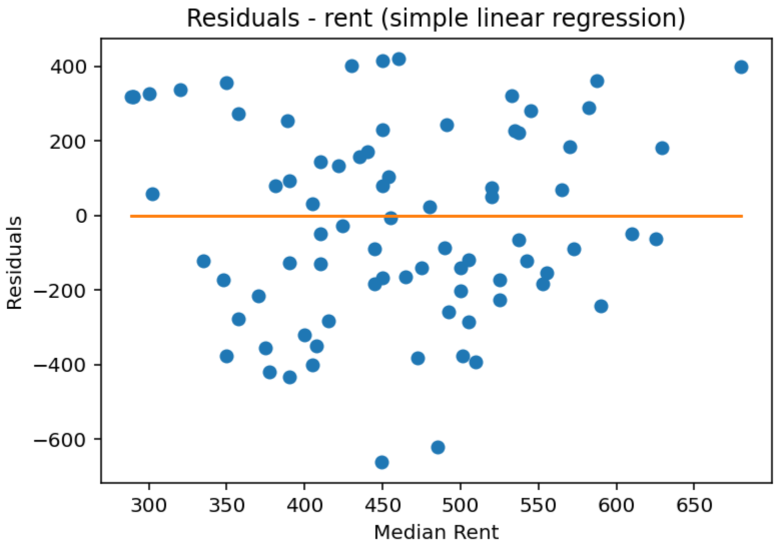
A multiple linear correlation was also attempted using both unemployment and rent, but Figures 7 and 8 show the plots of residuals not being symmetric about 0 for both unemployment rate and rent, suggesting that a multiple linear regression is not appropriate. This is possibly due to collinearity between unemployment rates and rental rates.



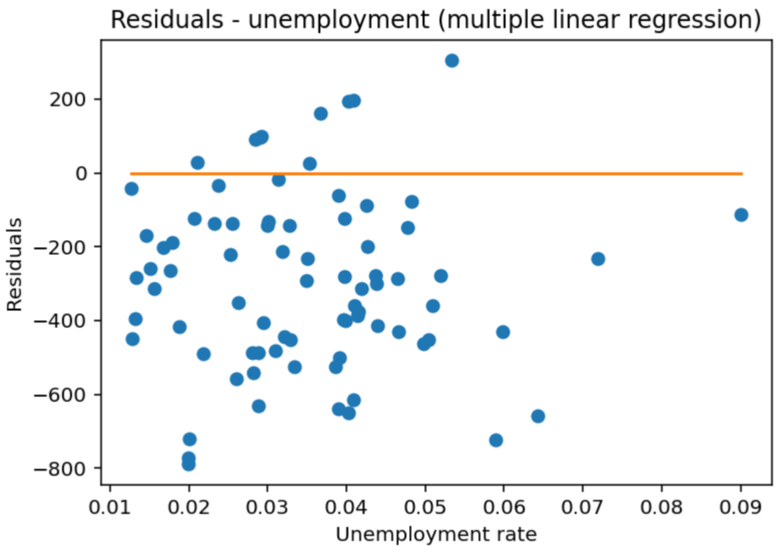
Figur 4: Linear regression model for median rent per LGA against average gaming expenditure per person.



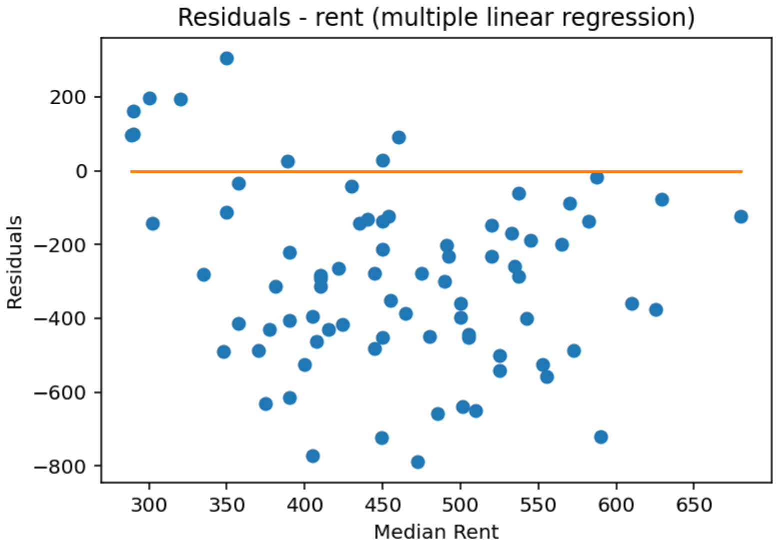
Figur 5: Plot of residuals against unemployment rates - simple linear regression.



Figur 6: Plot of residuals against rent - simple linear regression.



Figur 7: Plot of residuals against unemployment rates - multiple linear regression.



Figur 8: Plot of residuals against rent - multiple linear regression.

Time Series Analysis

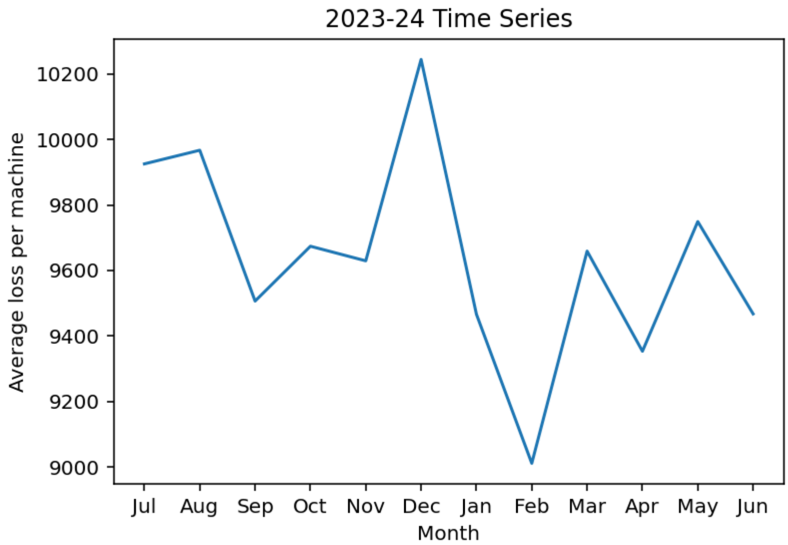
To understand potential seasonal trends in how much people spend on gambling, a time series analysis of the monthly loss per machine was conducted for the 2023-24 and 2024-25 financial years. As gaming machines are biased to win for the house on average, the average loss per machine is still a good measure of how much people tend to spend on gaming machines.

Figures 9 and 10 both show a significant dip in average loss per machine from December to February (corresponding to after the Christmas period), followed by a sharp spike in losses from March onwards. In 2023-24 there was also a very large spike from November to December (corresponding to the start of the Christmas period). Looking at both time series superimposed on the same plot in Figure 11, we also see that there has been a fairly consistent increase in average losses to gaming machines across all months except December and February.

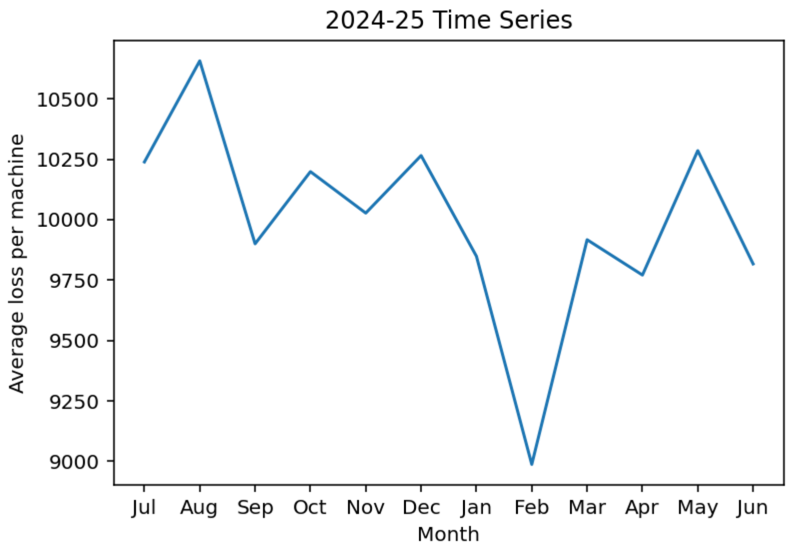
Discussion on findings

As noted, we have found that on average higher unemployment rates are correlated to higher spending on gaming. Additionally, we have found that Christmas may be a potential trigger for increased gambling spend.

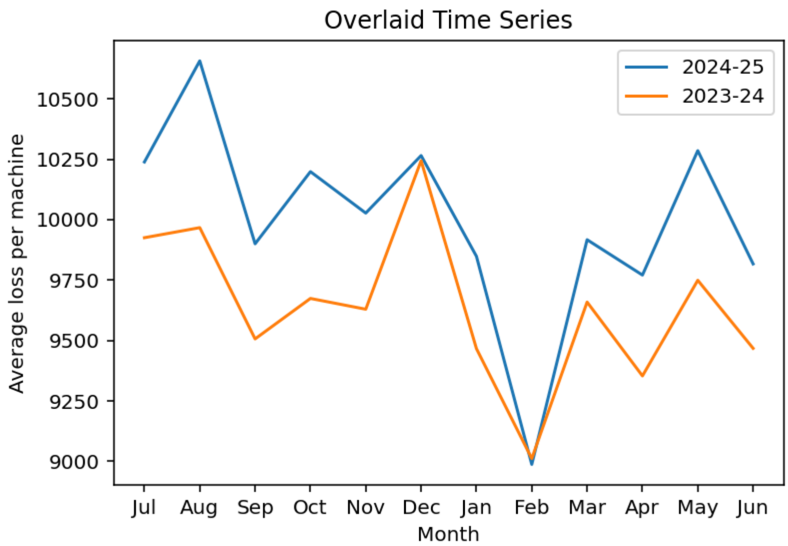
To resolve this, other community-led activities could be conducted for particularly high-risk times, such as community Christmas celebrations. Young people could also be encouraged to



Figur 9: Time series of average loss per machine for each month over the 2023-24 financial year.



Figur 10: Time series of average loss per machine for each month over the 2024-25 financial year.



Figur 11: Time series of average loss per machine for each month over the 2023-24 and 2024-25 financial years, superimposed on the same graph.

participate in other activities that are unrelated to gambling at gaming machines. Additionally, more activities could also be conducted in January and February, ideally with the focus on continuing them long-term, to take advantage of the trend to spend less in gambling in January and February.

For unemployed people, job support/finding groups could also encourage them to find additional activities to keep themselves engaged during Christmas, and such activities should be doable long-term to also take advantage of the trend to spend less in gambling in January and February.

More broadly, additional funding could be dedicated to constructing new areas such as playgrounds or non-gambling gaming venues. This way, people have more alternatives to gambling at gaming machines and may be less likely to return to gaming machines.

Limitations

The analysis conducted was fairly surface-level; it is highly likely other socioeconomic factors also influence gambling, such as school results and school completion status. Additionally, while data analysis has been conducted to understand potential risk factors, more data analysis should be conducted on the suggestions provided to understand their potential efficacy (or lack thereof, depending on what the data shows).