Final Report

Matthew Hickson, Hasne Hossain, Gan Li
Group: 122

mhickson@student.unimelb.edu.au, hhossain@student.unimelb.edu.au,
gl1@student.unimelb.edu.au

Since the introduction of electronic gaming machines (EGMs) in 1956 to local pubs, clubs, hotels, and casinos, Australia now homes 20% of all gaming machines worldwide despite being only 0.3% of the global population. In 1999, it was estimated that 2.1% of adults are *problem gamblers* requiring either treatment or urgent intervention.[4] It's clear that gambling addiction is harmful and has the capacity to damage lives. Despite around 70% of people believing gaming machines do more harm than good and 92% not wanting further expansion of gaming machines, the number of EGMs and the industry revenue has steadily grown each year. In Victoria, currently $7,446,423 is spent at EGMs every day.[2]

With the emergence of online gambling, never before has the pastime been more accessible and heavily promoted. However, even with the development of more easy to access avenues for gambling, half of the entire gambling revenue is collected by gaming machines. It's estimated 3 out of 4 people categorised as being severely harmed by gambling addictions play on EGMs[1]. It has been shown previously that gaming machine operators often locate machines in suburbs with low incomes and the least ability to cope with the losses caused by gaming machines.[3] Vulnerable groups are often targeted by operators due to the inadequate harm prevention policies currently implemented by the government. A report released by the Victorian Responsible Gambling Foundation in February, 2020 states there is limited research that has examined the behaviours and attitudes of older adults. The report goes on to conclude that older adults are one of the more vulnerable groups for experiencing gambling harm, and over time that rates of gambling participation and problem gambling is increasing for this group.[5]

The aim of this report, therefore, is to identify any statistical relationship between the amount of gaming expense and the proportion of older adults, specifically retirement age and higher, in Victorian regions.

1. Data Wrangling

To demonstrate this relationship, datasets detailing median income, age distributions, and net gaming statistics in Victorian regions were examined. These datasets had a temporal range from 2012-2016 and include key research figures delineated by local government area (LGA). The main question guiding this research is whether there is evidence suggesting EGM operators are targeting older adults, defined in this report as people of 65 years of age or older. Demonstrating a statistical connection between these two data records can aid in constructing meaningful cases for a decrease in the number of gaming machines prevalent in Australia. Previous research has concluded that the number of factors influencing gambling behaviours is quite large and as such the ability to predict gambling frequency and problem gamblers is unreliable with only a handful of key statistics.[6] It is believed this research will also face challenges related to reliable

indicators for gambling expenditure due to the complexity of hidden factors not being captured by our source datasets.

This research sourced the following three datasets: Historical Gaming Expenditure Statistics, Population Statistics from the Australia Bureau of Statistics (ABS) and Personal Income Statistics, also from ABS. The data was preprocessed to remove undesired feature fields from each dataset before merging the datasets by LGA codes. Once merged, a summary table was created containing population distribution numbers, categorised by Age Groups, Number of EGMs and Venues, Net Expenditure on these EGMs, and Median Income. Using this data, the Proportional Percentage of Adults Over 65 compared to the Population Above 19 was determined. This proportion was chosen for examination as it represents the percentage of adults above retirement age compared to the population expected to be able to gamble. Links for the datasets are below.

- Historical Gaming Statistics from Vic. Commission for Gambling and Liquor Regulation:

  ```
  https://discover.data.vic.gov.au/dataset/historical-lga-population-density-\
  gaming-expenditure-statistics
  ```

- Population by Age and Sex, Regions of Australia from ABS

  ```
  https://search.abs.gov.au/s/search.html?collection=abs-search&form=simple&
  profile=_default&query=%22Population+by+Age+and+Sex%2C+Regions+of+Australia\
  %22
  ```

- Personal Income in Australia from ABS

  ```
  https://www.abs.gov.au/statistics/labour/earnings-and-work-hours/personal-\
  income-australia/latest-release
  ```

During the data wrangling, regions that had no EGMs were removed from the combined data, as they also subsequently recorded 0 Net Gaming Expenditure. This decision was made to ensure only regions with any gaming expenditure were influencing the analysis methods. It is important to note that after pruning our dataset, and restricting the temporal range, the sample size reduced to ($\mathbf{n} = 350$), which admittedly is rather low given the complexity of the issue. Although more years were available for examination, because they lacked other class features present in our target temporal range, it was decided to not include them. However, it is expected this sample size would need to be increased to produce better conclusions.

A. Permutation Feature Importance

After wrangling, the importance score of each feature influencing Net Expenditure was determined. The figure below illustrates the strength of connection between the Net Expenditure and the next four most influential class features (weaker class features were excluded in the figure).
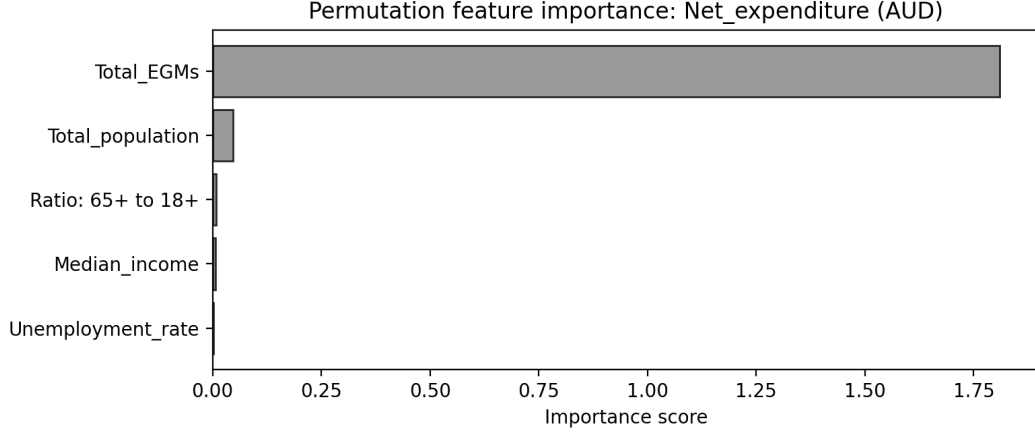

Figure 1: Importance score of most influential class features on Net Expenditure (AUD)

2. Data Analysis

As can be seen from Figure 1, the most influential feature was the Total Number of EGMs in a region. For this reason, the Total Number of EGMs were repeatedly examined with all other features versus Net Expenditure and the greatest influencing pairs were determined. These pairs were determined to be Median Income and Total EGMs, and Proportion over 65 and Total EGMs.

To ensure the robustness of a multiple linear regression model, both of these influential pairs were examined for collinearity. The results are shown below.
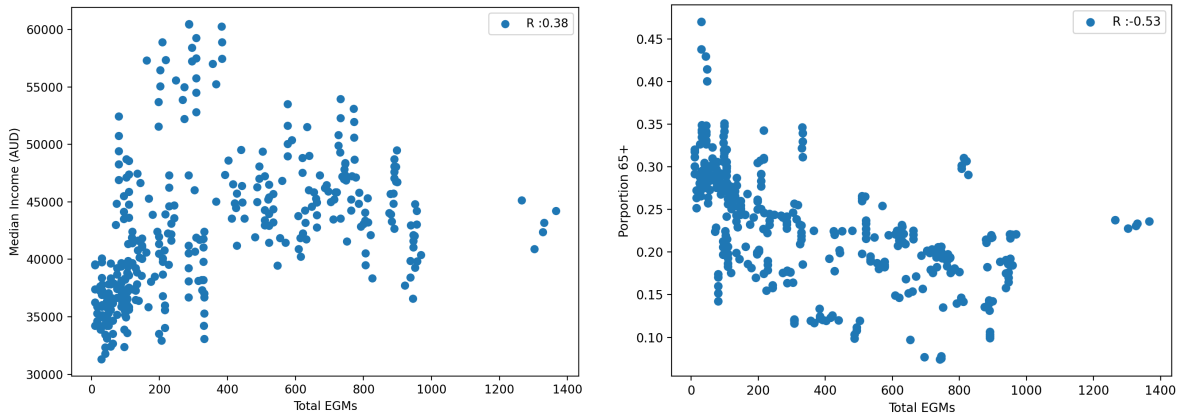

Figure 2: Total EGMs vs Median Income (left) and Total EGMs vs Proportion over 65 (right)

Figure 2 demonstrates a weak positive linear relationship between Median Income and Total EGMs (R = 0.38), and a moderate negative linear relationship between Proportion over 65 and Total EGMs (R = -0.53). This suggests that Median Income will be an appropriate class feature

to examine with Total EGMs, however, the result generated with Proportion over 65 will not be as reliable an indicator for gambling expenditure.

The multiple linear regression models for both pairs were trained and tested against the sample data.
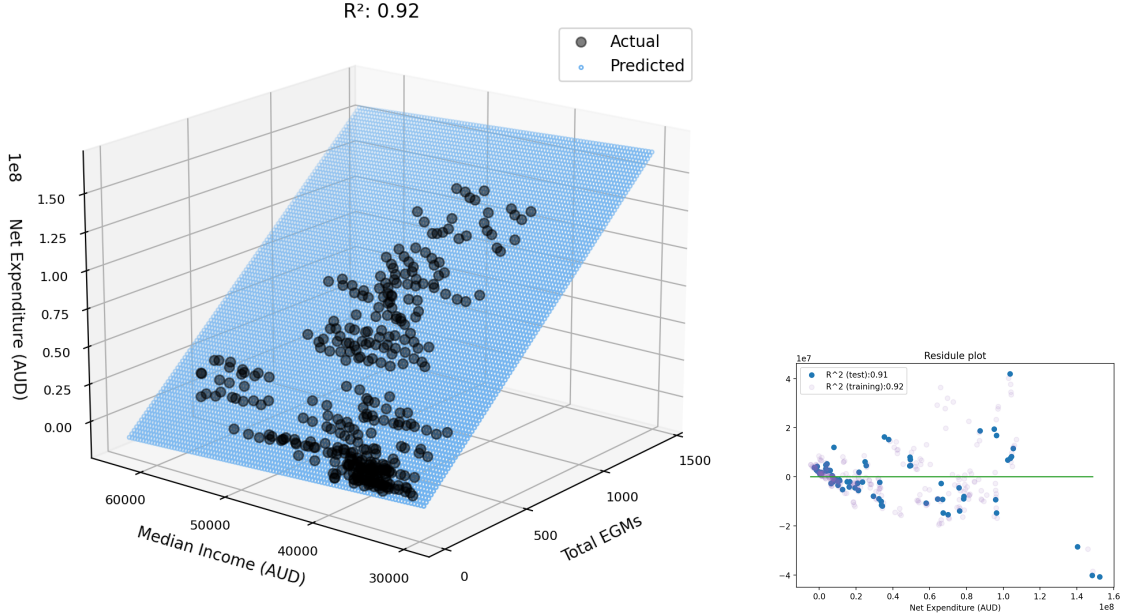

Figure 3: Multiple Linear Regression models of Median Income and Total EGMs vs Net Expenditure (AUD) (left), and corresponding residual plot (right)

As seen in Figure 3, the Coefficient of Determination was 0.92 for the training data and 0.91 for the test data. This indicates a strong correlation between the pair of features and Net Expenditure. Of which they can explain 92% of the variance in Expenditure. Observing the 3D plot, it can be readily determined that an increase in the Number of Gaming Machines directly links to an increase in Gaming Expenditure. This was to be expected as more EGMs in a region increase the capacity for that region to gamble. Conversely, it could also indicate that within high expenditure regions, machine operators choose to increase their quantity of machines to increase their business revenues. This relationship is shown in Figure 4.
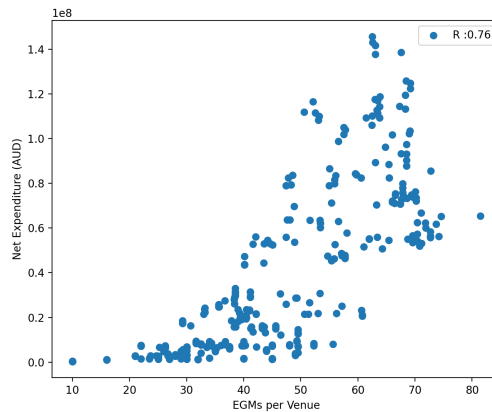

Figure 4: EGMs per Venue vs Net Expenditure (AUD)

As expected the regions with higher EGMs per Venue also had higher net expenditure, most likely due to a greater income-stream from more EGM users.

Interestingly though, in accordance with past research our results showed that the highest earning regions did not gamble as much as the moderate-low Median Income regions. It's worthy of note that many of the regions in the lowest Median Income range were also low Net Expenditure LGAs. The majority of the high expenditure regions had Median Income values between $35,000 to $50,000.
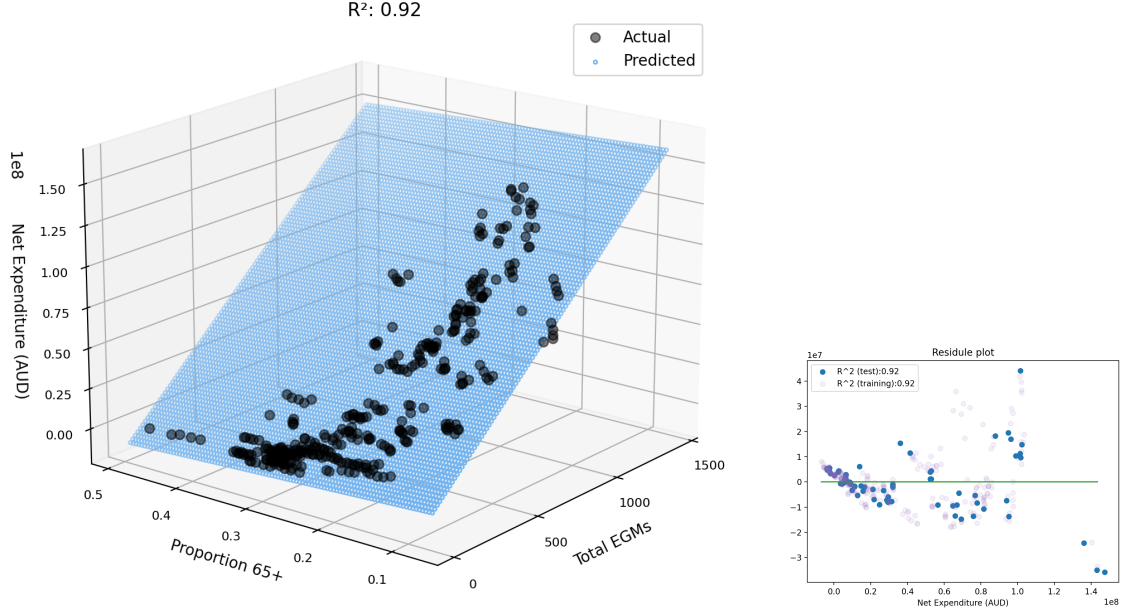

Figure 5: Multiple Linear Regression models of Median Income and Total EGMs vs Net Expenditure (AUD) (left), and corresponding residual plot (right)

The model in Figure 5 above shows a negative relationship between the Proportion over 65 and both Net Expenditure and Total EGMs. The residual plot also illustrates that the model is accurate for Net Expenditures below $40,000,000 and after which the error increases more and more. Similarly to Figure 3, 92% of the variance in Net Expenditure can be explained by a change in Total EGMs and Proportion over 65. It can be seen as the Net Expenditure increases, the percentage of people over 65 decreases. This negative relationship was the opposite of what was expected, and the highest expenditure regions were ones with Proportions over 65 between 10% and 25%.

3. Discussion and Conclusion

Since both regression models were dominated by the Total EGMs, it is suspected that the reason for the negative relationship is due to a large array of hidden factors not captured by our datasets influencing regional gambling habits. This does not indicate that older adults are not contributing, either minimally or significantly, to the Net Expenditure but that this conclusion cannot be inferred from the data selected. All of the sourced data did not describe the attributes of patrons playing EGMs, and exclusively using the Proportion over 65 was insufficient at showing a relationship between these two variables.

The amount of expenditure per person was also determined for each region. This statistic is useful for normalising the net expenditure, and thus was chosen for examining the entropy in combination with the Proportion over 65. The respective class features were discretised into variable bin sizes and the normalised mutual information was calculated. The result indicated that 67% of the change in Expenditure per Person could be explained by a change in Proportion over 65.

In conclusion, it was determined that the Proportion of adults over 65 years of age was not a defining characteristic of regions with higher gaming expenditure or total quantity of EGMs. The Net Expenditure was most influenced by the number of EGMs in a region and then next most influenced by the Median Income, then Proportion over 65. Thus, it is not possible to draw reliable conclusions between these target variables and more research will need to be performed. Increasing the sample size by examining more years, or including other states, would assist in producing statistical significance. As it currently is, a connection between the percentage of adults over 65 and Net Gaming Expenditure could not be shown.

References

[1]   Web Page. 2008. URL: http://www.problemgambling.gov.au/facts/.

[2]   Web Page. 2021. URL: https://responsiblegambling.vic.gov.au/resources/gambling-victoria/pokies-across-victoria/.

[3]   Web Page. 2021. URL: https://www.pokiesplayyou.org.au/the_damage_done.

[4]   Productivity Commission. *Australias Gambling Industries*. Report Report No. 10. 1999. URL: https://www.pc.gov.au/inquiries/completed/gambling/report/gambling1.pdf.

[5]   Victorian Responsible Gambling Foundation. *Factors that shape the gambling attitudes and behaviours of older adults in Victoria*. Report. Victorian Responsible Gambling Foundation, 2020. URL: https://responsiblegambling.vic.gov.au/documents/747/VRGF_RR_OlderAdults_0220_FA.pdf.

[6]   Piers D. L. Howe et al. "Predictors of gambling and problem gambling in Victoria, Australia". In: *PLOS ONE* 14.1 (2019), e0209277. DOI: 10.1371/journal.pone.0209277. URL: https://doi.org/10.1371/journal.pone.0209277.