Assignment 2 Open Ended Project Report (COMP20008)

Group W{05}G{1}

Emerson Leishman	Jason Jiang	Benedict Yong	Zhengrong Yan
1460385	1460397	1600075	1343546
emersonl@student.unimelb.edu.au	jasjiang@student.unimelb.edu.au	yongbw@student.unimelb.edu.au	zhengrongy1@student.nimelb.edu.au

On The Predictive Relationship of House Prices from Criminal Offences and Gambling

This report explores the potential predictive relationship between house factors and two key factors: criminal offence rates and electronic gambling rates across various local government areas (LGA) across Victoria. The objective of this report is to identify the relationship between house prices, EGM data, and Offences data. Through our exploration, we have found a moderate correlation between EGM data and House prices using linear regression as well as a predictive relationship between both EGM and Offences, and House Prices using the K-nn model. The Linear regression model suggests that the average individual spending on electronic gambling is significantly higher in areas with lower average House Prices. As a result, it is advised that targeted government interventions such as anti-gambling ads be displayed in areas of lower house prices.

Introduction

The purpose of this report is to explain and elaborate on the key findings of our research project, which aimed at investigating whether a region's house prices could be predicted by their rates of crime and gambling. Specifically, the research question we have decided on for this project is "Can we find a predictive relationship of house prices based on criminal offences and gambling?" With this research question, our project essentially contained two sub-explorations:

1. The effect of crime rates on house prices. In particular, we aimed to investigate if there was a notable correlation between house prices in each LGA (dependent variable) and their corresponding offences rate (independent variable), and, if possible, determine a specific division of offences.

2. The effect of EGM rates on house prices. Similar to the above, we sought to determine whether or not a rise in EGM rates (independent variable) would correlate with a fluctuation in house prices (dependent variable).

For this analysis, we primarily used datasets containing detailed information on each region's crime rates, EGM rates, and house prices across Victoria. By analysing these data sets, we intend to uncover and highlight trends, patterns, and key insights that can inform local government authorities in Victoria in their planning efforts. The significance of this project lies in its potential to provide valuable insights into the relationship between housing dynamics and social factors, such as crime and gambling. If it is the case that we can find a predictive model of this relationship, then this research could identify critical areas for support, investment opportunities, and the unique challenges faced by each region, suburb, or LGA, allowing for more effective decision-making across Victoria.

Methodology

Throughout this project, our team decided to use a shared GitHub repository as this provided us with a convenient way to delegate and maintain collaborative work across each aspect of the project. Through this delegation process, each team member was responsible for different aspects of the project, such as preprocessing different datasets, employing different machine learning models, etc. The code language used was Python 3.12.6, with Anaconda as the Python interpreter. The data preprocessing and preparation stages of this project used a Python file accompanied by relevant comments detailing the coding, while meaningful graphs and data modelling techniques were written using Jupyter Notebook with .ipynb files. The revised datasets that underwent data cleaning were exported using .csv files.

Additionally, multiple libraries and packages were used for various purposes throughout the data preprocessing and modelling stages, including:

- 'Pandas' for data manipulation
- 'Sklearn' for machine learning
- 'Matplotlib' and 'Seaborn' for graph visualisation
- 'NumPy', 'Random' and 'Scipy' for mathematical and statistical calculations

Data Exploration and Preprocessing

House Prices by LGA

We noticed that except for the houses by LGA which recorded the data by suburbs, all data that we received records data by LGA. Therefore, we decided to explore our data according to LGAs. The corresponding LGA to suburbs can be found on communities.csv. Given that an LGA contains many suburbs and different suburbs have different house prices, we introduce the concept of weighted house price of LGA:

$$Weighted\ house\ price\ of\ LGA\ =\ \sum_{suburbs\ in\ LGA} \frac{Population\ of\ suburb}{Total\ population\ of\ LGA} \times\ House\ price\ of\ suburb$$

We noticed that there is missing data in Houses by Suburb. Therefore only the suburbs which have house prices given will be involved in the calculation. Otherwise, the excess total population of LGA will lower the weighted house price of LGA and produce a huge error. From the population information from the communities.csv, we only had access to the population data for 2007 and 2012. Since there is not enough data on the population in each community, it was hard for us to find the relationship between population over the years, making it difficult to predict the population over time. To get close to the realistic population situation, we decided to use the 2012 population data as the populations for future years in order to minimise the error in population calculation.

To improve the performance of our data in the supervised model, we decided to classify the house price into three levels: Low, Medium and High:

- The Low-range house price is under \$600,000
- The Medium-range house price is between \$600,000 and \$1,000,000
- The High-range house price is over \$1,000,000

We used the price of \$600,000 and \$1,000,000 as a bound because we found that most weighted house prices remained in that range from the data, which can be shown through the histogram, which is positively skewed:

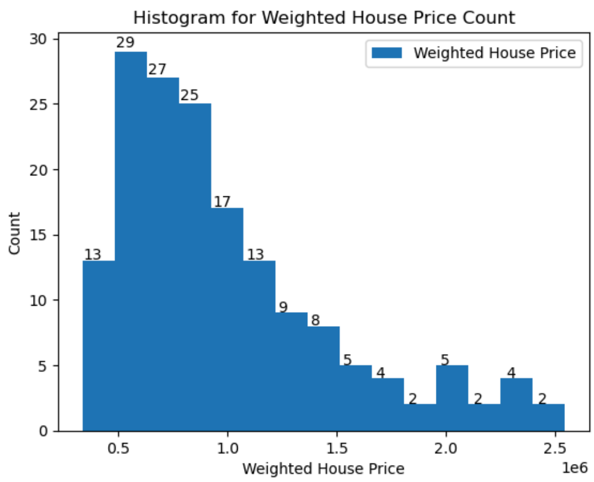


Figure 1: house price distribution histogram (from File Zhengrong.ipynb of the code).

For simplicity in calculation and analysis, we round our boundary value to the whole number. Since the size of the data is small, it is easier and more accurate for the model to predict the house price level based on the EGM rate and offence rate. In this case, even though we cannot accurately predict the weighted house price of LGA, we can predict the level of weighted house price interval that LGA locates in a specific year.

EGM by LGA

From the EGM csv we took the total amount of electronic gambling from years 2016-2020 and the LGA names, we were required to clean the data by removing the prefixes like "City of " or "Suburb of ". Additionally, we were required to split some of the areas into their composite LGA's. This calculation is as follows:

$$EGM\ amount\ for\ LGA = EGM\ Total\ for\ Greater\ Area \times \frac{Population\ \ of\ the\ LGA}{\Sigma\ Population\ \ of\ each\ LGA\ in\ Greater\ Area}$$

This calculation used the most recent population data found in the communities csv which is the 2012 estimated resident population:

$$EGM\ rate = \frac{EGM\ total\ for\ LGA}{population\ of\ the\ LGA}$$

In the EGM data set, we were required to create a new measurement expressing the amount of money as a rate by dividing by the population of the respective LGA, generating a weighted EGM rate in dollars per person (using the 2012 estimate of the resident population). The new measurement of the EGM rate was made necessary by our observations of the EGM dataset where we noticed that the highly populated areas spend much more on gambling, so to normalise the data we decided to express it as a rate.

Offences by LGA

From the six tables of data provided in the LGA Offences.csv, we decided to analyse only the Table 2 data as it provided the most relevant information to our research, allowing us to reduce redundant and otherwise noisy data. In particular, we only required the LGA name, Year, Offence Division, Offence Count, and LGA Rate per 100,000 population as all other meaningful analyses could be derived from these columns. Additionally, Since each entry in the dataset was separated by its individual offence subgroup, we first had to combine the entries into their respective LGAs. This was simply achieved by iterating through each LGA and only using the entries that belonged to the specific LGA. Moreover, to maintain compatibility with the Houses and EGM datasets, the offences dataset only targeted years from 2016 - 2020. To determine which division of crime was most influential to house prices, we needed a way to quantify the distribution of each crime division within each LGA. This was calculated using the following formula:

$$Percentage\ of\ each\ crime\ division = 100 \times \frac{\Sigma Count\ of\ crime\ in\ specific\ division}{\Sigma\ Count\ of\ total\ crimes\ in\ \ LGA}$$

However, this spread of specific crime divisions did not end up amounting to any meaningful correlation in our exploration and hence was not included in our further interpretation and analysis stages.

Data Modelling

Supervised Learning Model - K Nearest Neighbour

K Nearest Neighbor (KNN) is a supervised machine learning algorithm which uses proximity to make classifications or predictions about the grouping of an individual data point. We decided to use the k-nn model because of the following reasons:

- It is non-parametric, which means it does not need any pre-knowledge about the data itself.
- It uses simple comparisons to find similar records in the training data, KNN is sometimes very effective at making good predictions. Therefore, when there comes new LGA data with offence rate and EGM rate, we can predict the weighted house price level by comparing its data to the old data, finding LGA data which has similar situations to the new one and then predicting the weighted house price level based on these similar data.
- It can be effective at capturing the complex interactions among variables without having to define a separable statistical model, such as coefficients and weights. In this case, the weighted house price level is affected by the fusion effect of the EGM rate and offence rate. It

is not rational to weigh the effect of the EGM rate and offence rate separately. Therefore, the K-nn model is more suitable for managing this fusion effect.

We used cross-validation with 5 equal-sized partitions to find the k, the number of neighbours, to make our model more accurate. We use 1 partition for testing and the rest of 4 partitions for training, which means we have 132 rows of data for training and 33 rows. For each number of k, we repeat the testing 100 times and calculate the average number of accuracy.

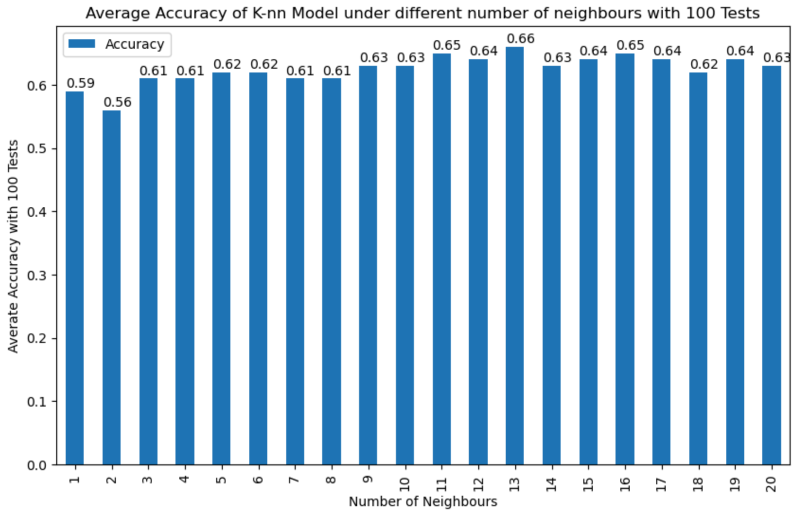


Figure 2: Accuracy of K-nn Model with varying numbers of neighbourhoods histogram (from File knn.ipynb of the code).

From the result, we can observe that when the number of neighbours is equal to 13, we have the highest accuracy of 0.66 despite the differences in accuracy in different numbers of neighbours not being predominant. Hence, we decided to let the number of neighbours be equal to 13 in order to get better accuracy in our model and further analysis.

Still separating the data into 5 equal-size partitions randomly with k = 13 and using 4 partitions for training and 1 partition for testing, we got the highest accuracy of about 72%. Considering that our training data size is small, we agreed that our model has high accuracy in the circumstances and can predict the weighted house price level successfully based on the offence rate and EGM rate. Below is the confusion matrix of the result:

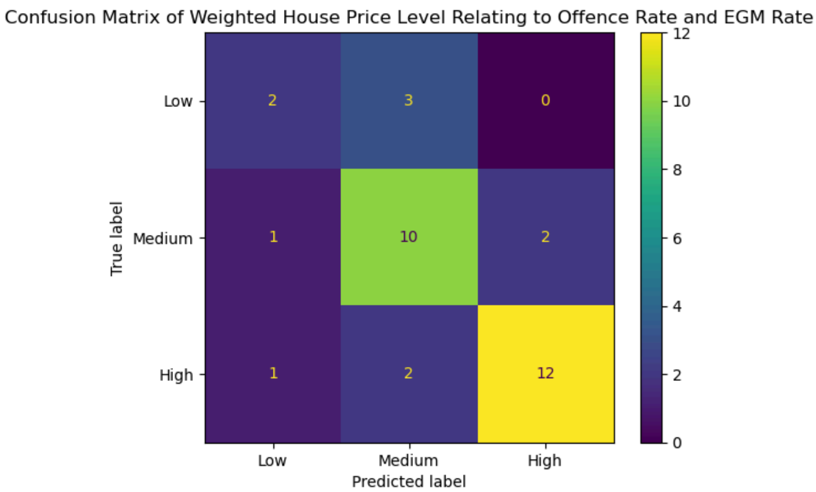


Figure 3: Confusion Matrix for k-nn prediction of house price (from File knn.ipynb of the code).

Then we further evaluate the confusion matrix to identify which group the model is good at predicting by using classification metrics:

Table 1: Performance of Prediction Analysis in terms of actual vs predicted values (Round to two decimal places).

	#TP	#FN	#FP	#TN
Low	2	3	2	26
Medium	10	3	5	15
High	12	3	2	16

Table 2: Prediction Analysis of K-nn using precision recall and F1 (Round to two decimal places).

	Low	Medium	High
Precision ($\frac{\#TP}{\#TP+\#FP}$)	0.50	0.67	0.86
Recall ($\frac{\#TP}{\#TP+\#FN}$)	0.40	0.77	0.80
F1 ($2 \times \frac{Precision \times Recall}{Precision + Recall}$)	0.44	0.72	0.83

Supervised Learning Model - Linear Regression

Linear regression is a commonly used supervised learning technique to create models based on data with an 'output' dependent variable and one or more 'input' independent variables and forms a predictive linear model based on the dataset. From our research question, these are the variables that we have used:

- Weighted House Price per LGA
    - Dependent variable
    - For the weighting method, see the above Data Exploration and Preprocessing section
- Electronic Gaming Machine rate per LGA
    - Independent variable
    - Financial losses per person, for calculations, see the above Data Exploration and Preprocessing section
- Offences rate per LGA
    - Independent variable
    - LGA Rate per 100,000 Population
- Year
    - Data used from 2016 to 2020, the years the different datasets overlap

The basic linear regression model is based on the following formula: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

where Y_i is the dependent variable, X_i are the independent variables and ε_i is the error component. β_0 is the intercept and β_1 (and β_2 for 2 independent variables) is the coefficient of the predicted linear regression line.

We essentially created 3 linear regression models based on the above variables:

- House Price vs EGM rate
- House Price vs Offences rate
- House Price vs EGM rate + Offences rate

Each of these linear regression models are created using the *LinearRegression* function from the 'Sklearn' package (calculating the intercept and coefficients). The error (MSE and RMSE) for each model is also determined using the *mean_squared_error* function from the 'Sklearn' package. The Pearson Correlation, which can be a measure of how good is the fit of the regression line, is determined using the *pearsonr* function from the 'Scipy' package.

Individual Model - House Price vs EGM rate - 2016 to 2020

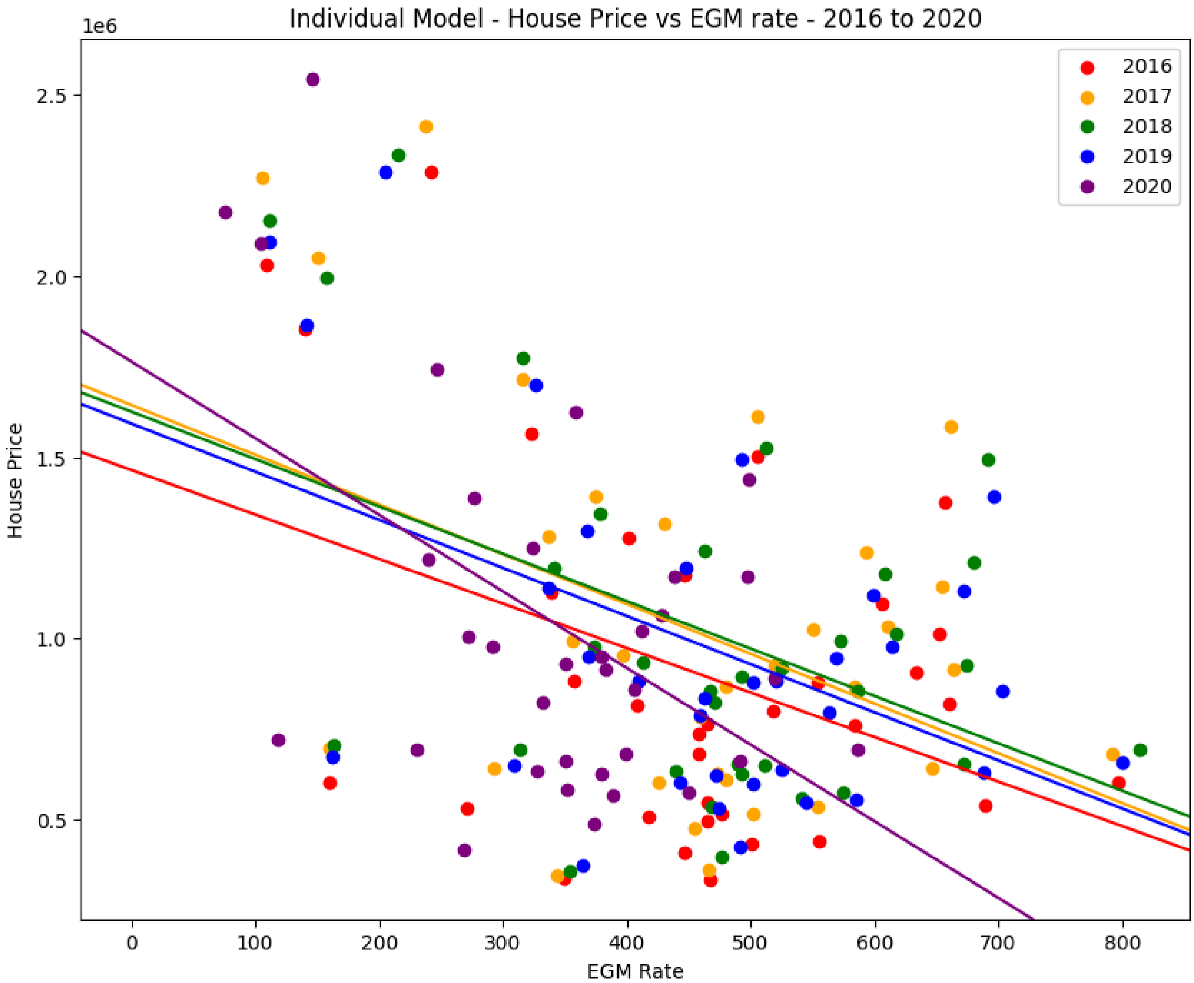


Figure 4: Individual model for House Prices vs EGM rate - stacked plot (from File Ben.ipynb of the code).

Table 3: Individual model for House Prices vs EGM Rate parameters (From File Ben.ipynb of the code)

Year	Intercept	Coefficient	RMSE	Pearson Correlation
2016	1465002.4399	-1229.5951	443205.6321	0.3990
2017	1644587.9942	-1374.5316	471807.6413	0.4096
2018	1626196.5955	-1309.0945	434346.5769	0.4338
2019	1593236.3632	-1329.1603	408780.7135	0.4650
2020	1764659.1271	-2116.5355	432004.3856	0.5015
Average			438028.9899	0.44178

Individual Model - House Price vs Offences Rate - 2016 to 2020

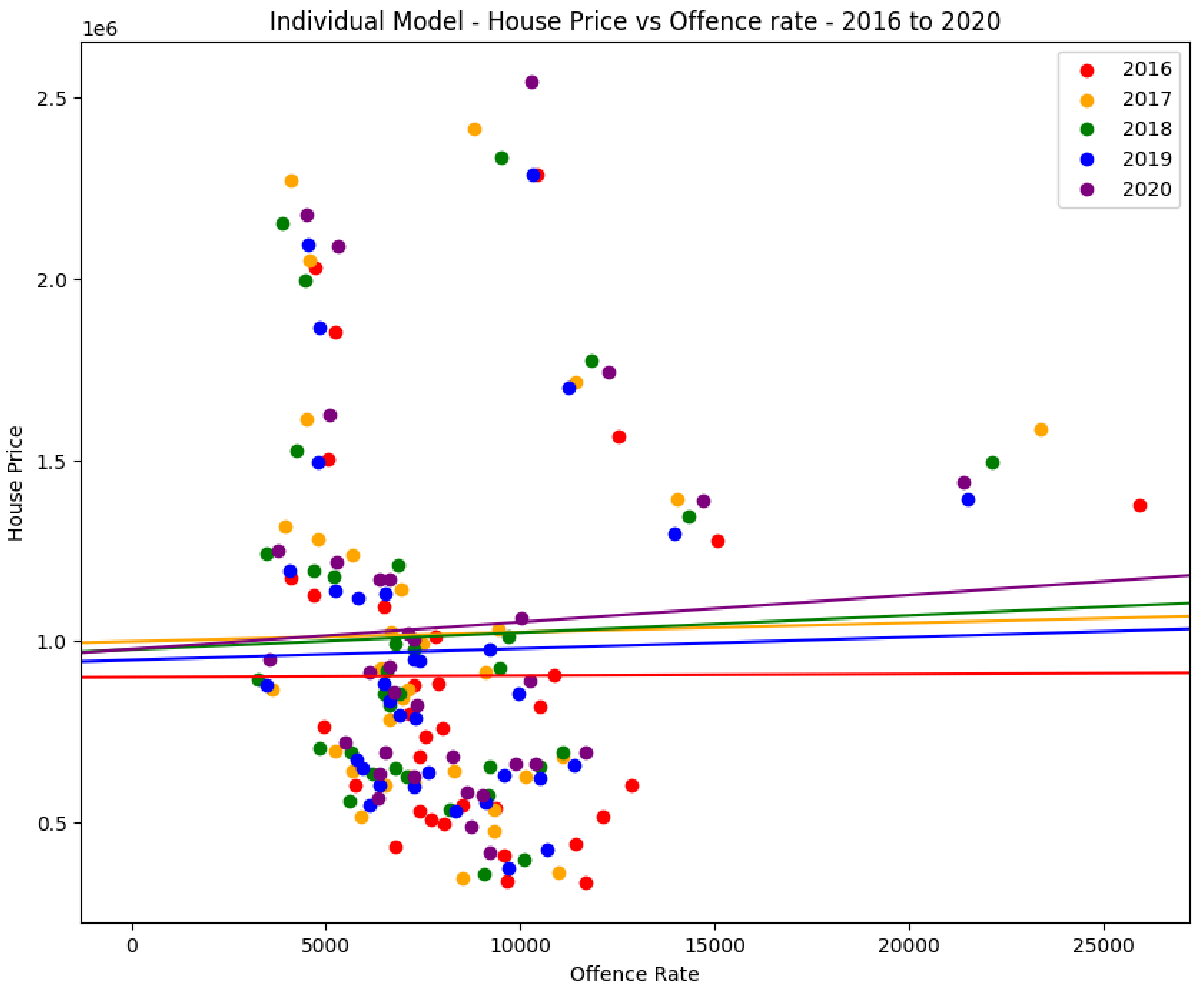
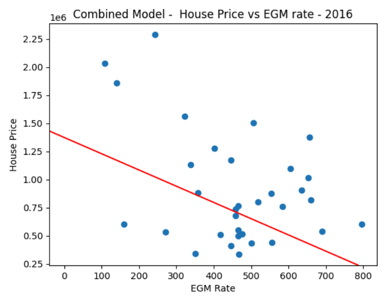


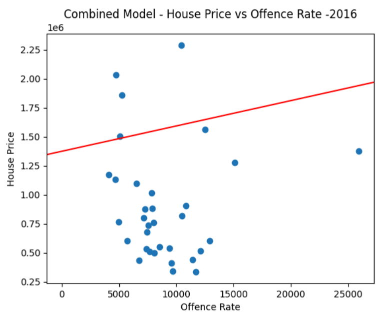
Figure 5: Individual model for House Prices vs Offences rate - stacked plot (from File Ben.ipynb of the code).

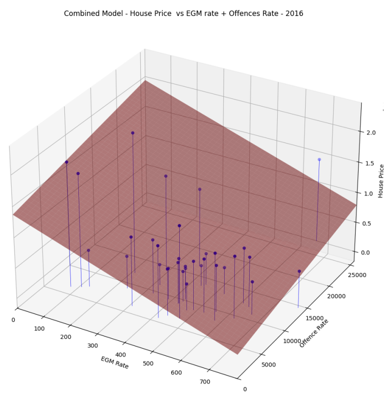
Table 4: Individual model for House Prices vs Criminal Offences parameters (From File Ben.ipynb of the code)

Year	Intercept	Coefficient	RMSE	Pearson Correlation
2016	900374.7505	0.4475	490366.4674	0.0037
2017	999089.0215	2.5567	524759.8566	0.0177
2018	976502.0641	4.7228	488827.2206	0.0347
2019	947774.8981	3.1560	468318.7182	0.0227
2020	977778.4352	7.4824	506083.4793	0.0503
Average			495671.1484	0.02582

Combined Model - House Price vs EGM rate + Offences rate - 2016 to 2020







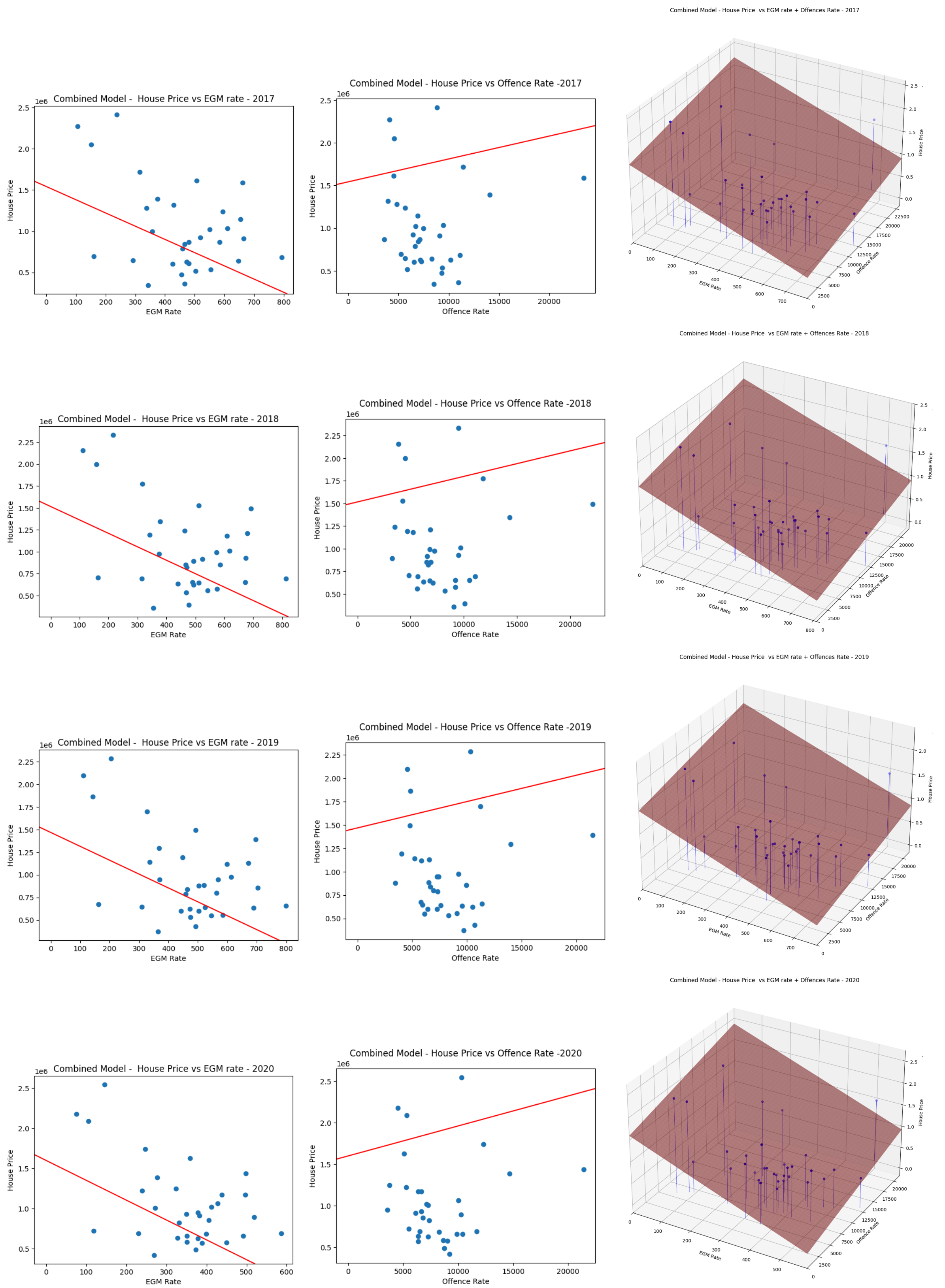


Figure 6: Combined models for House Prices vs Offences rate (from File Ben.ipynb of the code).

Table 5: Combined model for House Prices vs EGM vs Offences parameters (From File Ben.ipynb of the code)

Year	Intercept	Coefficient 1	Coefficient 2	RMSE	Pearson Correlation
2016	1373109.4644	-1445.1124	21.8800	441674.4751	0.4344
2017	1542513.8023	-1606.4953	26.8747	469376.5150	0.4474

Year	Intercept	Coefficient 1	Coefficient 2	RMSE	Pearson Correlation
2018	1516451.0188	-1533.0064	28.2049	429582.1832	0.4782
2019	1468316.4594	-1536.4651	28.2069	404053.2229	0.5060
2020	1599430.1128	-2475.3336	36.1605	421706.6628	0.5544
Average				433278.6118	0.48408

Discussion and Interpretation

Supervised Learning Model - K Nearest Neighbour

In cross-validation accuracy (**Figure 2)**, we would identify the accuracies of different numbers of neighbours are consistent since there is not too much difference. This implies that our model does not have a serious issue with overfitting, which occurs when a model is trained too well on the training data and performs poorly on the new, unseen data. Moreover, cross-validation provides a more realistic estimate of the model's generalisation performance, this means our model has the ability to perform well on new, unseen data.

For the evaluation metrics, we use precision, recall and F1 for assessing how good the model predicts for each group. Precision measures how often predictions for the positive class are correct. From **Table 2**, we can see High weighted house price level has the highest precision, while the Low weighted house price level has the lowest. Therefore, we can conclude that our model has decent precision in High and Medium, but the prediction in Low needs to be improved. Recall measures how well the model finds all positive instances in the dataset. **Table 2** demonstrates that our model is good at finding Medium and High instances, but cannot identify the Low weighted house price level.

F1 is used to examine the performance of the model on each class individually rather than considering overall performance. A higher number correlates to a better performance in the prediction. From **Table 2**, we can observe that the F1 scores for Medium and High are much higher than for Low. Again, this implies that our model has better performance in predicting the Medium and High weighted house price levels while having poor ability in predicting the Low group.

From the metrics, we can conclude that our model predicts better in High and Medium weighted house price levels while having a suboptimal prediction in the Low weighted house price level. This is mostly because our training data set has more data for Medium and High weighted house price levels, allowing the model to have better understanding and make more accurate predictions on these two classifications. However, when it comes to Low weighted house price level, we simply do not have enough data for training and testing, meaning that the model will be impacted more by noises and outliers in predicting Low weighted house prices.

Notice that we only evaluate the result with high accuracy here. Since we select training and testing data at random, it is possible that different partitions of the dataset will make the model have different levels of understanding for each group and therefore generate different results. Also, because our data size is small, accuracy may vary in a noticeable interval.

All in all, our model has decent performance in predicting weighted house price level based on offence rate and EGM rate with the highest accuracy about 72%. Considering the data size is small, we agree that the model has good performance in predicting the weighted house price level based on offence rate and EGM rate, particularly in Medium and High house price levels. This consolidates our idea that house price is related to the offence rate and EGM rate.

Supervised Learning Model - Linear Regression

As can be seen from the House Price vs EGM rate model, there appears to be a negative correlation between the 2 variables which is shown by the linear regression model predicted and plotted against the scatter plots for the years of the available dataset (2016 to 2020). This can be seen visually on the plot as well as confirmed by the average of the Pearson Correlation calculated for each of the years

(~0.44). This is also represented by a large negative coefficient β_1 , which is the downward slope of the line. This indicates a negative trend in House Prices as the EGM rate increases.

Similarly looking at the House Price vs Offence rate model, there appears to be a near-zero correlation between the 2 variables. The linear regression model produces a flat linear regression line (near zero coefficient) and can't be considered to be a good model for any predictive relationship.

The combined model of House Price vs EGM rate + Offence rate appears to have a similar negative correlation between House Price and the EGM rate component (large negative coefficient β_1 and slope), whereas the Offence rate component has a slight positive coefficient β_2 and slope. I believe this is caused mostly by the much stronger relationship of House Price vs EGM rate compared to House Price vs Offence rate (as seen from the individual models) and the need for the combined model to essentially form a 2D flat plane when viewed in a 3D plot.

From *Slide 31 of Week 6 of the Lecture Slides* for interpreting Pearson Correlation values, "*0.5 is large, 0.3-0.5 is moderate, 0.1-0.3 is small, less than 0.1 is trivial*". This further validates that the House Price vs EGM rate (Individual model) and the House Price vs EGM rate + Offence rate (Combined model) have "moderate correlation" and can be considered reasonable models for use as a predictive relationship, whereas the House Price vs Offence rate (Individual model) has "trivial correlation" and is not a reasonable model to use in this case. The combined model has a similar (but slightly higher) Pearson Correlation possibly suggesting that the additional data from the Offence rate does slightly improve the model, although this is not clear if this is possibly due to error or some other hidden factor in the data like a relationship between EGM and Offence rate.

Table 6: Summary of Linear Regression Model effectiveness (From File Ben.ipynb of the code)

Linear Regression Model	RMSE (average)	Pearson Correlation (average)
House Price vs EGM rate	438028.9899	0.44178
House Price vs Offence rate	495671.1484	0.02582
House Price vs EGM rate + Offence rate (Combined model)	433278.6118	0.48408

Attempts at removing outliers were undertaken, like the LGA of City of Melbourne which has a very large Offence Rate, likely caused by residents from outside of Melbourne travelling into the city and disproportionately affecting the data. However it did not appear to significantly alter the regression models or the Pearson Correlation enough to change the discussion above.

One thing noticed was a higher negative coefficient (steeper downwards slope) for the House Price vs EGM rate model for year 2020 compared to all the previous years, as can be visually seen on the stacked plot. We suspect that this is a result of the COVID-19 pandemic that occurred primarily in 2020, resulting in significantly reducing gambling rates due to enforced state-wide lockdowns.

The negative relationship between EGM data and House prices are most likely correlated as a result of sharing the same cause. It is likely that this cause is that of individuals' earnings, since an individual with a low income is unlikely to afford an expensive house, similarly the same individuals are more likely to suffer from excessive problem gambling. However it is also possible that a causal link connects the two, where an individual's gambling habits can worsen a person's financial situation making it far less likely that they can afford a house in an area of high average house price. More research is needed to deduce which of these situations is the cause of the correlation observed.

Limitations and Improvement Opportunities

Data constraints:

From the original criminal offences dataset we used a category which recorded the crimes per 100,000 people; however, such a calculation is likely flawed since the crimes are not necessarily committed near someone's place of residence. This could cause inflation in crime statistics in areas with large amounts of daily migration like the City of Melbourne, as the daily migration means that many individuals who do not count towards the population of the City of Melbourne can still count towards the criminal statistics.

Throughout our investigation, we performed many techniques aimed at modifying the data provided to improve the accuracy of the analysis of the research question. One such modification was the weighting of statistics like total dollar amount gambled, and the transition from house prices for suburbs to house prices for LGA's both of these modifications use the total population statistics provided in the communities csv. However, this data is a 2012 estimate and thus not entirely reliable when used in combination with our other data collected from 2016-2020. As a group, we concluded that this error was acceptable since its impact is systematic (shown by the similar gradients of the population estimates between different LGAs) and applied to almost all of our data. Thus, in comparing the data, this systematic error would have a negligible impact. Furthermore, if we attempted any imputation techniques to generate data for the missing years, this would introduce random error since many communities might continue to grow as predicted, and many others may not. This flaw could easily be improved by collecting newer data, perhaps from a more recent census. However, if this is an impossibility, we could use the data collected from 2007 and 2012 in tandem to create a prediction of the population of the LGA's for the years in question. Such a prediction would have to be carefully designed since it would have to factor in both the exponential growth from births and the linear growth seen in immigration. To look at a simplified case, we could assume that the populations grew at either a linear or exponential, giving us the populations seen in **Figure 7** respectively; however, these predictions are unlikely to be accurate.

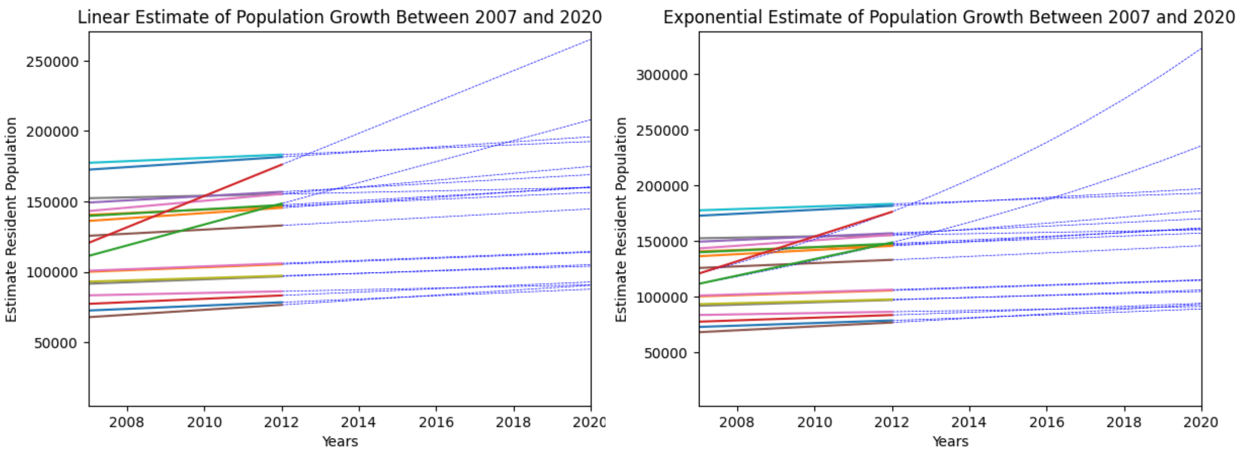


Figure 7: Estimates of populations assuming linear vs exponential growth (from File pop_est_err.ipynb of the code).

Larger Sample Size for Machine Learning:

By the end, our data frame had 165 rows and we split the majority of data into a training data set and the rest into the test dataset. This indicates that our data size is not large enough to make our model with high accuracy in order to find more detailed correlations among weighted house prices, offence rate, and EGM rate. Also, due to insufficient data, our model will be heavily impacted by outliers and noises. For example, in our k-nn model, since we do not have enough data in Low weighted house price level, we have bad performance in predicting the Low group, while we have much better performance in predicting Medium and High because of sufficient data. We expect to increase the model accuracy in predicting the Low group by having more data.

Similarly many rural LGAs might have only a few thousand people. This can make these area's criminal offences much more prone to distortion from outliers, ultimately compromising the predictive analysis to be less accurate than it otherwise would be.

Assumptions:

Our research question seeks to identify a relationship between gambling and house prices, however the EGM data set collects only electronic gambling data, and thus may not be a complete representation of gambling. For example, a large portion of gambling in communities such as the City of Melbourne happens in person through casinos like Crown, or horse races like the Melbourne Cup. It is likely that if this in person gambling data were to be included, our results would be significantly different. Furthermore, some remote towns do not have access to in person gambling, making the EGM data the complete gambling statistics of the area. Thus, through our analysis, we are comparing the complete sum of gambling done by some towns to what may only be a small portion in others. To improve this, it's recommended that in-person gambling data be collected as well as that of the EGM data.

Our research question states that we seek to find a relationship between criminal offences and house prices, which we analysed through the total crime count for each LGA. However, this analysis cannot find a relationship between different types of crime or crime of different severities and house prices. This could be easily remedied by producing many different dimensions in the models which take into account each different type of crime.

For our groups house price calculations we used the formula:

$$Weighted\ house\ price\ of\ LGA\ = \sum_{suburbs\ in\ LGA} \frac{Population\ of\ suburb}{Total\ population\ of\ LGA} \times\ House\ price\ of\ suburb$$

However this is only an approximation to the true value:

$$Weighted\ house\ price\ of\ LGA\ = \sum_{suburbs\ in\ LGA} \frac{Houses\ in\ suburb}{Houses\ in\ LGA} \times\ House\ price\ of\ suburb$$

This approximation assumes that the ratio of the population of the suburb to the population of the LGA is approximately equal to the ratio of the total number of houses in the suburb to the total number of houses in the LGA. This may hold for the vast majority of LGAs, however in newly built areas such as Bonnie Brook or Mambourin, it is likely that the ratio of the number of houses greatly exceeds the ratio of the populations, which could result in an underestimation of the weighted house price of the LGA. Additionally, To divide the data into the categories of High, Medium and Low house prices, we decided to choose approximate values which will split the data into hopefully significant categories. However, upon further analysis, we have identified that the house pricing data is positively skewed as seen in **Figure 1.** This could mean that our selected values to divide the data into thirds are much closer than they would be if the data followed a standard normal distribution. As a result, this is likely to increase the number of errors in the prediction since the difference in house prices between Low and Medium house price categories are far smaller than they would be. This could explain why the accuracy of the predictions for the low house price data is much lower than that of the medium or high house prices. To fix this, careful consideration must be placed in choosing appropriate values for the splits of house prices so that the majority of the data will fall within the selected bounds.

Conclusion

Our report has identified a moderate correlation between the EGM data and that of house prices, while little to no correlation has been identified between house prices and Criminal Offences. From this, we have managed to generate a predictive relationship by using the K-nn algorithm with the highest accuracy of 72%. From these findings, there likely exists a relationship between electronic gambling in an area and the area's house price, however, it is unlikely that this relationship persists for criminal offences. This data suggests that gambling as a problem is most prevalent in areas with lower house prices, and so government interventions such as anti-gambling messages and gambling support groups will be put to best use if established in these areas. However, it should be noted that while a definitive correlation has been proven between EGM rates and house prices, that is not to say that house prices are directly influenced by EGM rates. Indeed, further research should extend the scope of this exploration to other potential factors that may affect house prices so that we may better confirm and understand the relationship between house prices and EGM rates.

References

Hendricks, R. (2024). *What is a good accuracy score in Machine Learning?* [online] Deepchecks.com. Available at: https://www.deepchecks.com/question/what-is-a-good-accuracy-score-in-machine-learning/ [Accessed 2 Oct. 2024].