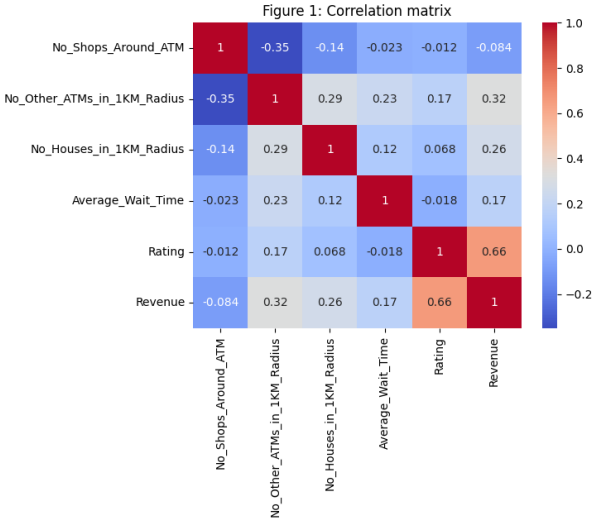
Report on Finding an Optimal ATM Location

Introduction

This report aims to help a bank identify the optimal location for their new ATM machines by analysing the provided dataset and using machine learning algorithms to predict monthly revenue and customer rating for new ATMs.

Figure 1: Correlation matrix



Correlations to the Revenue

To gain an understanding of the correlation between the different numerical variables, a Pearson correlation matrix was generated and represented as a heatmap as seen in Figure 1 to the right. Clearly, there is a high positive correlation between rating and revenue which indicates that ATMs that are higher rated by the public are more likely to be ones that generate superior profit than lower rated ones.

Also, the number of other ATMs within 1km, and number of houses within 1km also show a low correlation with revenue and thus should be considered as factors when trying to maximise revenue when deciding where to place a new ATM. On the other hand, the number of shops around the ATM and average wait time have a very low correlation to revenue so are much less likely to be significant factors affecting it.

Once categorical features were converted into indicator variables, it was found that the Residential Low-Density (RL) ATM zone has a 0.2450 Pearson score, but the Residential Medium-Density (RM) zone has a score of -0.2879. This indicates that ATMs in RL areas are more likely to have higher revenue than not, but the opposite is true for ones in RM zones. This is also the case with 'Town' settlement type (ATM_Type column) being 0.2232, and 'Urban' being -0.2676.

Thus, from this correlation analysis, a highly rated new ATM placed in a high ATM and house density area while within an RL zone in a town are optimal.

Trends in the Categorical Data

Categorical data were also analysed by visualising the median revenue for each category. The median value was used to estimate the typical revenue for an ATM falling under each category.

From Figure 2, the median revenue in commercial (C) zones are nearly half that of the other zones and floodplain zones have greater than 160% higher revenue than commercial. Also, ATMs attached to sheds seem to outperform other locations and those attached to shops lag behind others, as seen in Figure 3.

Figure 2: Median Revenue for an ATM by Zone

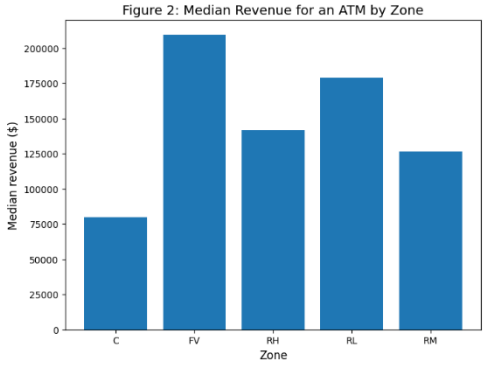
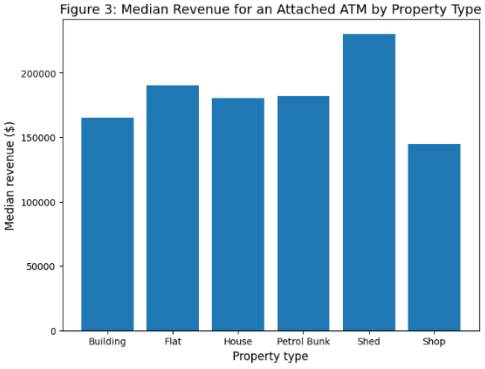


Figure 3: Median Revenue for an Attached ATM by Property Type

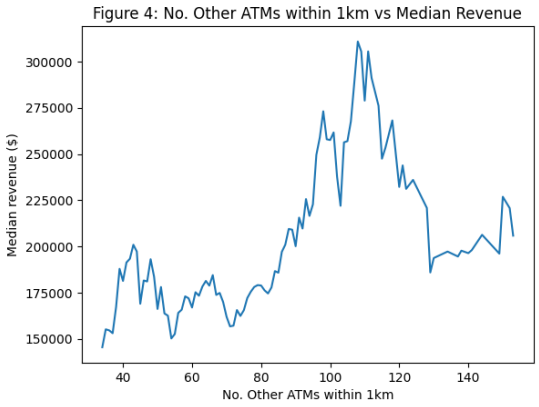


Additionally, ATMs that are placed a 'Little Inside' perform 17% worse in terms of revenue than ones that are 'Facing Road' on average.

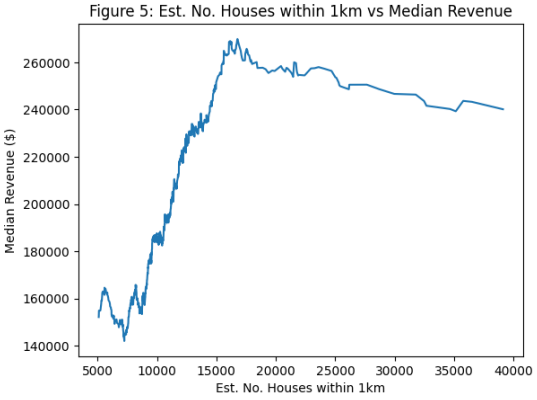
Thus, from these statistics, it will be most optimal to place a new ATM in a floodplain zone in a non-urban settlement, attached to a shed, facing the road.

Trends in the Numeric Data

The number of other ATMs in a 1km radius was plotted against the revenue for these ATMs, as seen in figure 4, to gain better insights into their relationship. The correlation between the number of other ATMs in a 1km radius and revenue as found above is seen to actually be non-linear with the median revenue trend starting low and then peaking at around 110 nearby ATMs.



In addition, the number of houses in a 1km radius was also plotted against revenue, with a similar growth then drop-off. According to this graph, the optimal value for the number of houses in a 1km radius is around 16,000.



Thus, an ATM at a location with around 100 – 120 ATMs and 15,000 – 30,000 homes within 1km could lead to higher revenue figures based on these trends.

Machine Learning Model

The Scikit-learn library was used to train and evaluate a Random Forest Regression model which predict revenues with high accuracy and has predictions with a high correlation to actual data. This can be seen with its 0.9959 accuracy score and 0.9980 Pearson correlation score against the provided test data and similar results when cross-validated.

When predicting the revenue, the model performed best without taking into account the transaction type (ATM_Location_TYPE column), ATM placement, settlement type, and the ATM looks. Thus, these are likely less important factors when deciding a new ATM's location as, their presence have a negative effect on the regression model used.

Recommendations

Based on the above analysis, it is recommended that a new ATM be placed in a residential low-density or floodplain zone. The new ATM is optimally placed in areas with around 100 – 120 ATMs and 15,000 – 30,000 homes within 1km.

It is also recommended that the business ensure that its new ATM is rated highly as an ATMs rating has a high correlation with its revenue. External factors affecting the rating must be analysed and considered when installing a new ATM. For example, the effect of the user interface, user experience or convenience of the location it's in should be studied.

Note that, on average, the best area for a new ATM is in a town, attached to a shed and facing the road, but the findings in the regression models have found these not to be driving factors in increasing ATM revenue and thus are not stressed as important.

Conclusion

In conclusion, this report provides insights into the features that must be considered when deciding the optimal location for a new ATM machine to increase its revenue. The analysis and recommendations provided in this report will allow the business to make more informed decisions when finding a good spot for a new ATM.