

COMP9321 Report

This report analyzes the ATM dataset to identify the features of a location that can increase revenue for a new ATM machine.

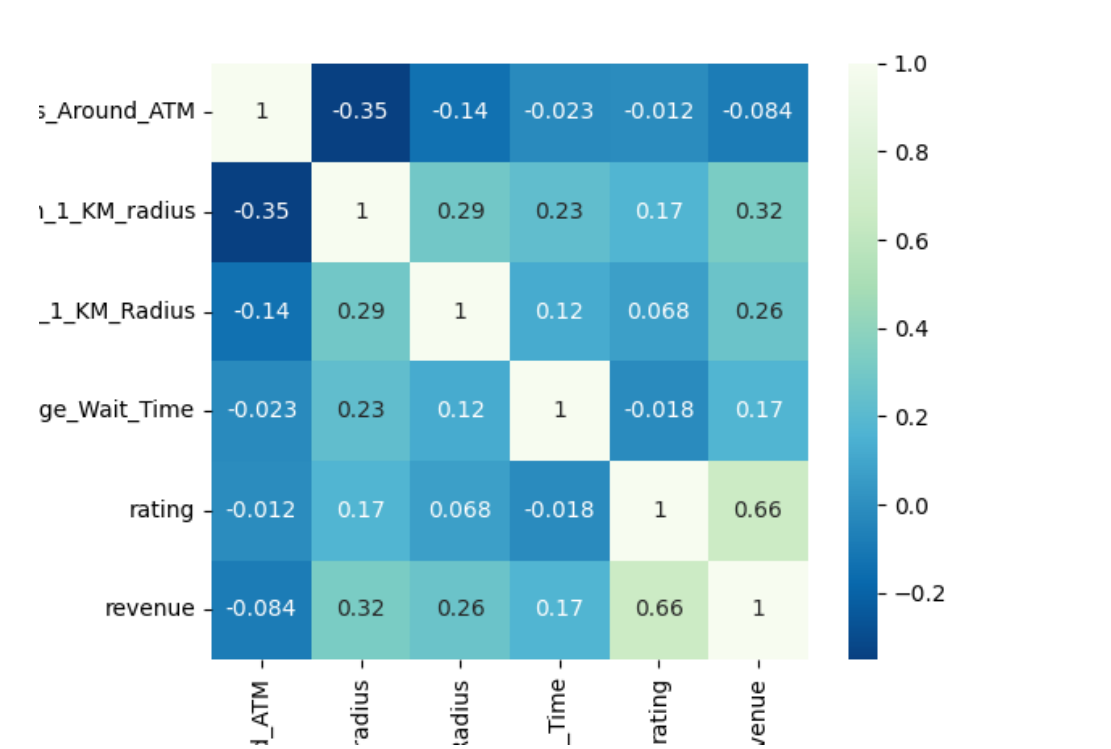
Exploratory Data Analysis

The ATM dataset contains information on various ATM machines, their locations, and other attributes that describe their surroundings. The dataset consists of more than 150,000 observations and 13 variables. The following table summarizes the summary statistics for each variable in the dataset:

Variable	Mean	Standard Deviation	Min	Max
Number_of_Shops_Around_ATM	62	42.39	21	200
No_of_Other_ATMs_in_1_KM_radius	70	22.60	21	213
Estimated_Number_of_Houses_in_1_KM_Radius	10517	9980.76	1300	215245
Average_Wait_Time	3	0.81	0	8
Rating	3	0.66	2	5
Revenue	186133	79474.90	35400	764900

Correlation analysis

To identify the relationship between different variables, we perform correlation analysis. The following table shows the correlation matrix for each variable:



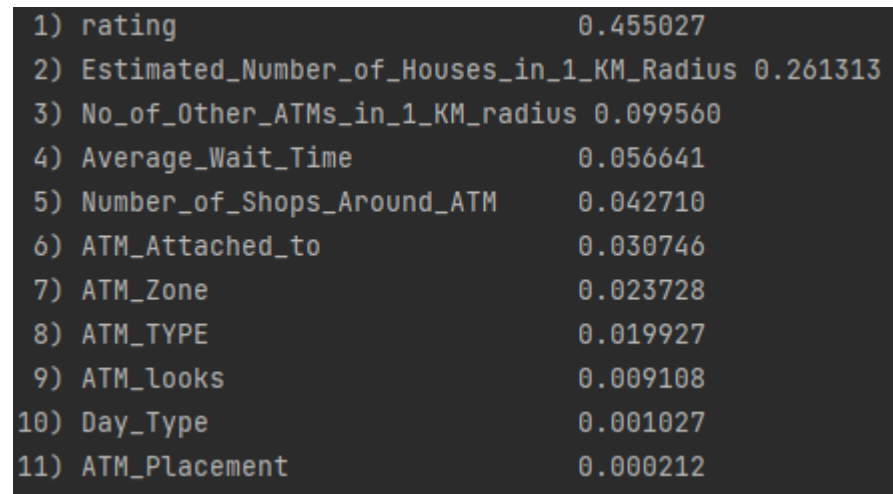
From the correlation matrix we could see that, the 'rating' variable is of the strongest

relationship with the 'revenue' variable. Besides, the 'No_of_Other_ATMs_in_1_KM_radius' variable also has some correlations with 'Number_of_Shops_Around_ATM' variable. Which we could summarize that the rating of ATM and the number of shops around ATM can influence the revenue of the ATM.

Regression analysis

To further investigate the relationship between the variables and revenue, we perform regression analysis. We build a Random Forest model with revenue as the dependent variable and the other variables as independent variables.

After training, our model has achieved 0.99 score in training set and 0.86 score in testing set. By interpreting feature importance of trained Random Forest regressor, we acquired the ranking feature importance as follows.



1) rating	0.455027
2) Estimated_Number_of_Houses_in_1_KM_Radius	0.261313
3) No_of_Other_ATMs_in_1_KM_radius	0.099560
4) Average_Wait_Time	0.056641
5) Number_of_Shops_Around_ATM	0.042710
6) ATM_Attached_to	0.030746
7) ATM_Zone	0.023728
8) ATM_TYPE	0.019927
9) ATM_looks	0.009108
10) Day_Type	0.001027
11) ATM_Placement	0.000212

This means the rating of the variable is the features that most related to the revenue of the ATM and that is also in accordance with the correlation analysis. The rating of the ATM is positively correlated with ATM revenue, indicating that ATMs with higher ratings tend to generate higher revenue than those with lower ratings. This finding suggests that customers are more likely to use an ATM with a higher rating, which may be seen as more reliable or trustworthy.

The number of Houses near the ATM location is also another important features that need to be considered, indicating that ATMs located near these establishments tend to generate higher revenue than those located in areas with few or no such establishments.

Summary and conclusion

Based on these findings, we recommend that the company target locations with a high number of shops and houses in the surrounding area, a low density of other ATMs, short average wait times, high ratings, and a mix of nearby commercial and retail establishments. It's important to note that above analysis can only identify correlation, not causation. Therefore, it's important to consider other factors that may impact ATM revenue, such as marketing efforts, accessibility, and customer behavior.