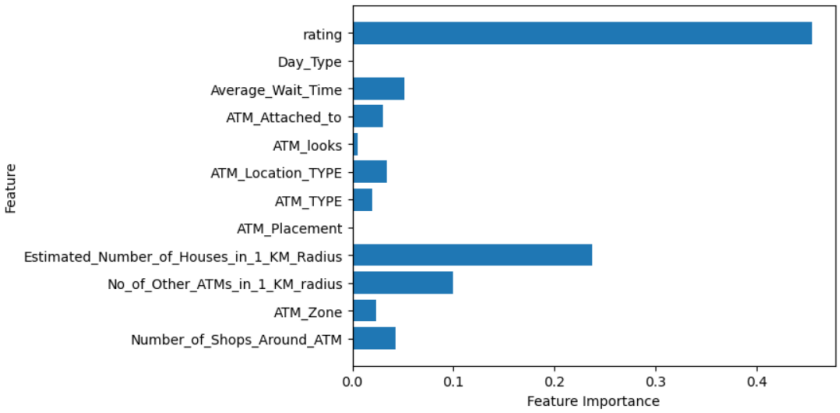
Data Analysis Report

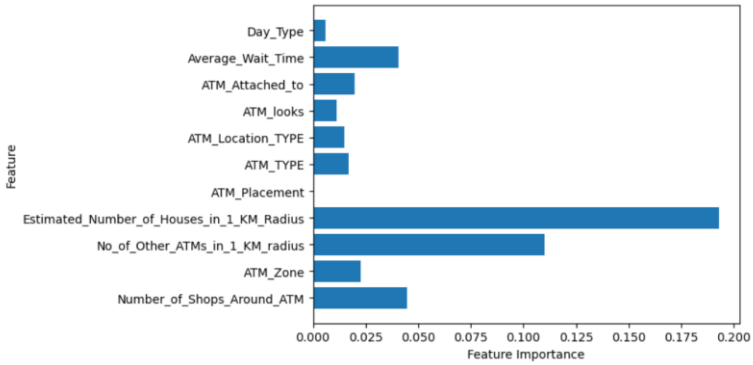
I found that there are two areas where ATM machines can stabilize revenue at around 75k.

After I used the random forest to predict the revenue of the dataset, I observed that among the features related to revenue, rating is the most influential, followed by 'Estimated_Number_of_Houses_in_1_KM_Radiu', 'No_of_Other_ATMs_in_1_KM_radius', 'Average_Wait_Time' and 'Number_of_Shops_Around_ATM'.

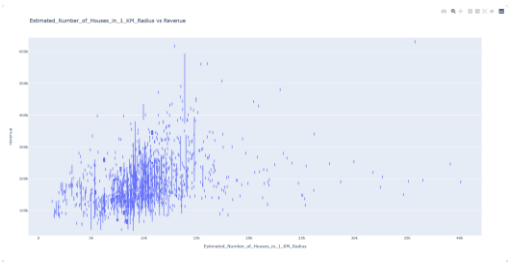


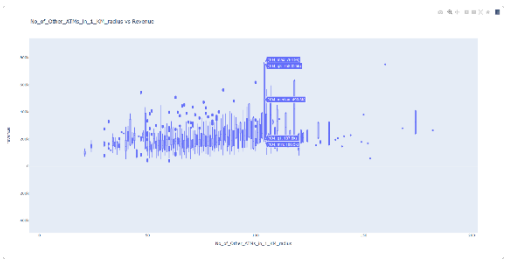
The most influential 'rating' among them may be the rating given by most users to ATM machines. Although higher ratings are more likely to result in higher revenue, which is in line with the actual situation, this is the behavior of most users but not an objective condition that we can control.

Thus I went on to plot the impact of various features on 'rating'. It can be seen that the features impact on 'rating' is the same as that on 'revenue'.



Among them, 'Average_Wait_Time' is also an uncontrollable objective factor with relatively little impact. Without it, The impact of 'Estimated_Number_of_Houses_in_1KM_Radiu', 'No_of_Other_ATMs_in_1_KM_radius', and 'Number_of_Shops_Around_ATM' on 'revenue' is shown in the figure below.





We should first pay more attention to the plot of 'Number_of_Shops_Around_ATM', because it's abnormal with too centralized data. I speculate that the ATMs in the dataset are centrally placed in batches, so there are areas where revenue can be stabilized. So I screened the feature values corresponding to the maximum revenue values in the three box plots and found that there are indeed two regions that can stabilize revenue at around 750k.

	Number_of_Shops_Around_ATM	No_of_Other_ATMs_in_1_KM_radius	Estimated_Number_of_Houses_in_1_KM_Radius	ATM_Attached_to	revenue
1182	69	160	15623	Petrol Bunk	749600
2642	70	160	15623	Petrol Bunk	747800
4102	68	160	15623	Petrol Bunk	752500
5562	69	160	15623	Petrol Bunk	752500
7022	70	160	15623	Petrol Bunk	746700
...	...	...	...	...	...
144262	66	160	15623	Petrol Bunk	748400
145722	62	160	15623	Petrol Bunk	746700
147182	68	160	15623	Petrol Bunk	748800
148642	63	160	15623	Petrol Bunk	752700
150102	63	160	15623	Petrol Bunk	752600

103 rows × 5 columns

	Number_of_Shops_Around_ATM	No_of_Other_ATMs_in_1_KM_radius	Estimated_Number_of_Houses_in_1_KM_Radius	ATM_Attached_to	revenue
691	70	104	21535	Building	761700
2151	67	104	21535	Building	764700
3611	63	104	21535	Building	757900
5071	67	104	21535	Building	764600
6531	61	104	21535	Building	762100
...	...	...	...	...	...
143771	68	104	21535	Building	756200
145231	70	104	21535	Building	758800
146691	65	104	21535	Building	761500
148151	65	104	21535	Building	763400
149611	65	104	21535	Building	756100

103 rows × 5 columns