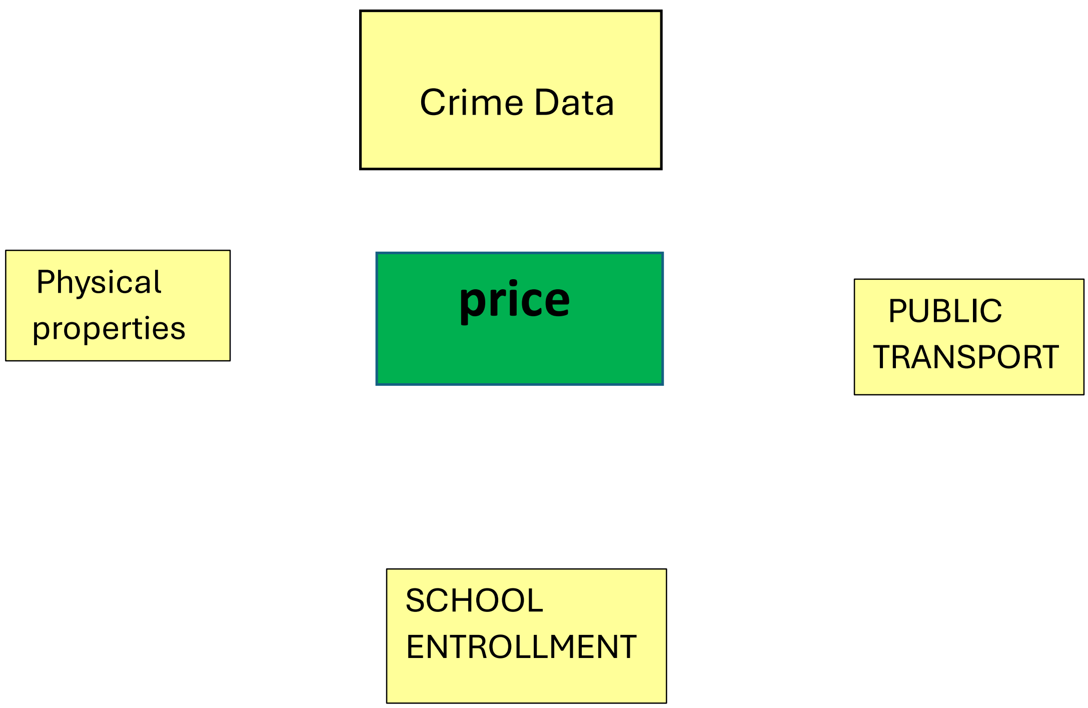
Comparative Analysis of property Price Predictions in Milwaukee District 1: LLM vs. Traditional Models

INTRODUCTION

The debate around the effectiveness of large language models (LLMs) for structured data analysis is growing. While LLMs excel at handling unstructured data like PDFs, they are often thought to fall short compared to traditional machine learning models in tasks like price prediction. To explore this, I've chosen Milwaukee District 1 as a case study, using datasets such as house prices, crime statistics, nearby public transport, and school proximity. The goal is to compare the predictions made by LLMs and machine learning models to see if they align or show noticeable differences. This study aims to provide a better understanding of whether LLMs can compete with traditional models in structured data tasks or if their strengths are better suited for other applications.

DATA MODEL



PHYSICAL PROPERTY DATA

This dataset contains property information for Milwaukee from 2018 to 2023, sourced from the City of Milwaukee Open Data Portal. Each property is uniquely identified by attributes like **_id,**

Property ID, and **tax key**. Key details include the **Address**, **District**, and **Neighbourhood (nbhd)**.

Property features include **Style**, **Extwall**, **Stories**, **Year Built**, **Rooms**, **Finished Sqft**, **Bdrms** (bedrooms), and bathrooms (**Fbath** and **Hbath**). It also specifies **Units** for multi-unit properties and **Lotsize** for land area.

Sales information includes **Sale date** and **Sale price**, making the dataset valuable for analysing real estate trends and property characteristics across neighbourhoods.

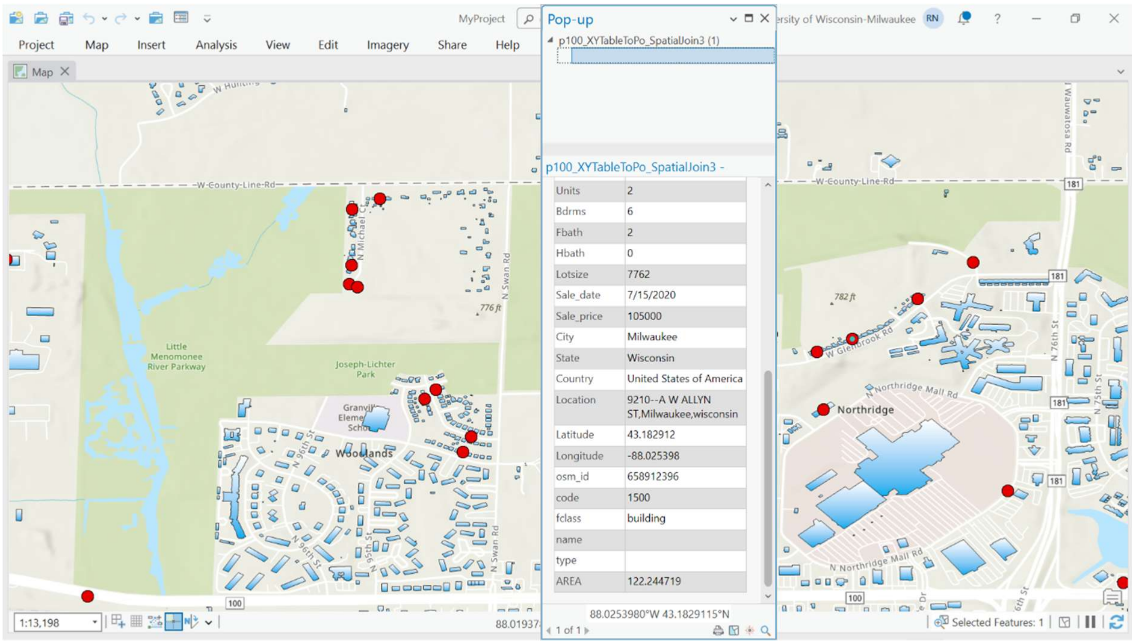
Data Cleaning:

The dataset underwent a systematic cleaning process to ensure its reliability for analysis. Duplicate rows were removed to eliminate redundancy, and columns Condo Project and Extwall were dropped due to approximately 90% missing values, making them unsuitable for meaningful analysis. Null values in the H Bath (Half Bathroom) column were replaced with 0, based on research indicating that null entries likely represent property without half bathrooms.

In case of Finished Sqft I have used ArcGIS pro, where I have geolocated the houses in open street map and spatially joined with the GIS OSM building data (contains the shapes of the building in the Wisconsin state) used field calculator to calculate the size of the house(Finished Sqft)

_price	Location	latitude	longitude	osm_id	code	fclass	name	type	AREA
0000.0	3102 N HACKETT AV,Milwaukee,wisconsin	43.074923	-87.876505	660573431.0	1500.0	building			310.808371
7600.0	3218 N DOWNER AV,Milwaukee,wisconsin	43.076113	-87.877580	499328674.0	1500.0	building		house	147.531235
9000.0	3234 N HACKETT AV,Milwaukee,wisconsin	43.076458	-87.876441	660573311.0	1500.0	building			134.644470
3000.0	3235 N SHEPARD AV,Milwaukee,wisconsin	43.076450	-87.874725	660573308.0	1500.0	building			120.952958
4000.0	3218 N SHEPARD AV,Milwaukee,wisconsin	43.076120	-87.874173	660573341.0	1500.0	building			109.889064

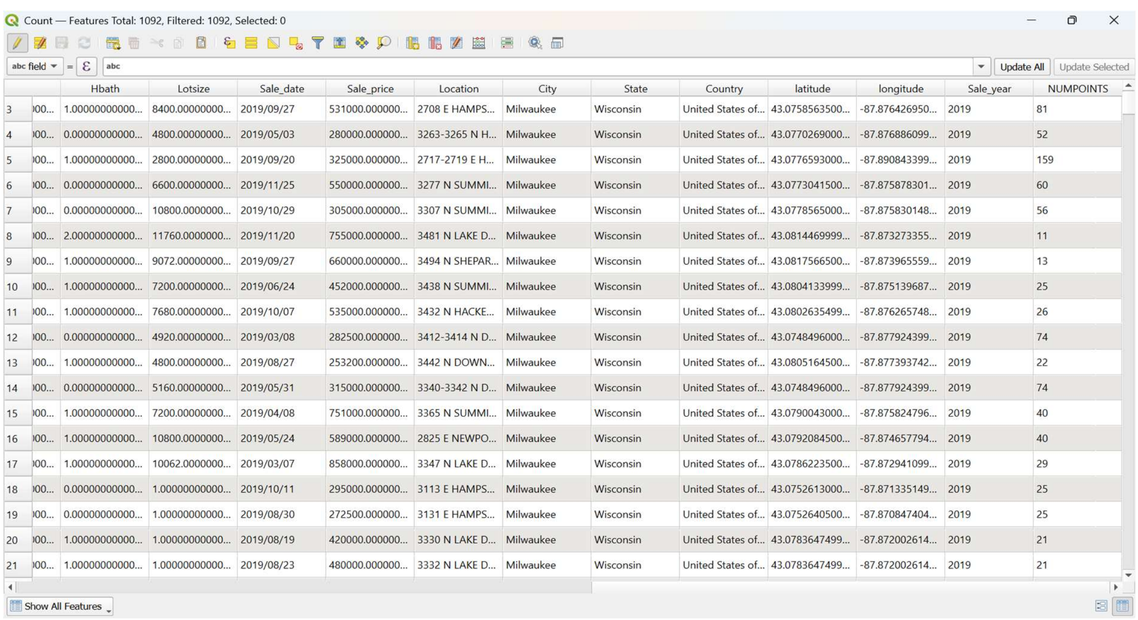
We got the area of the property(finished Sqft) using ArcGIS pro.



CRIME DATA

The dataset provides crime statistics for specific locations within Milwaukee's police District 1&5. Data includes the number of crimes reported annually at various addresses. Geographic details such as longitude and latitude are provided for spatial analysis. The data is collected between 2016 to 2022.

Using QGIS software, I'm going to calculate the number of crimes that occurred in 300m area near the property.



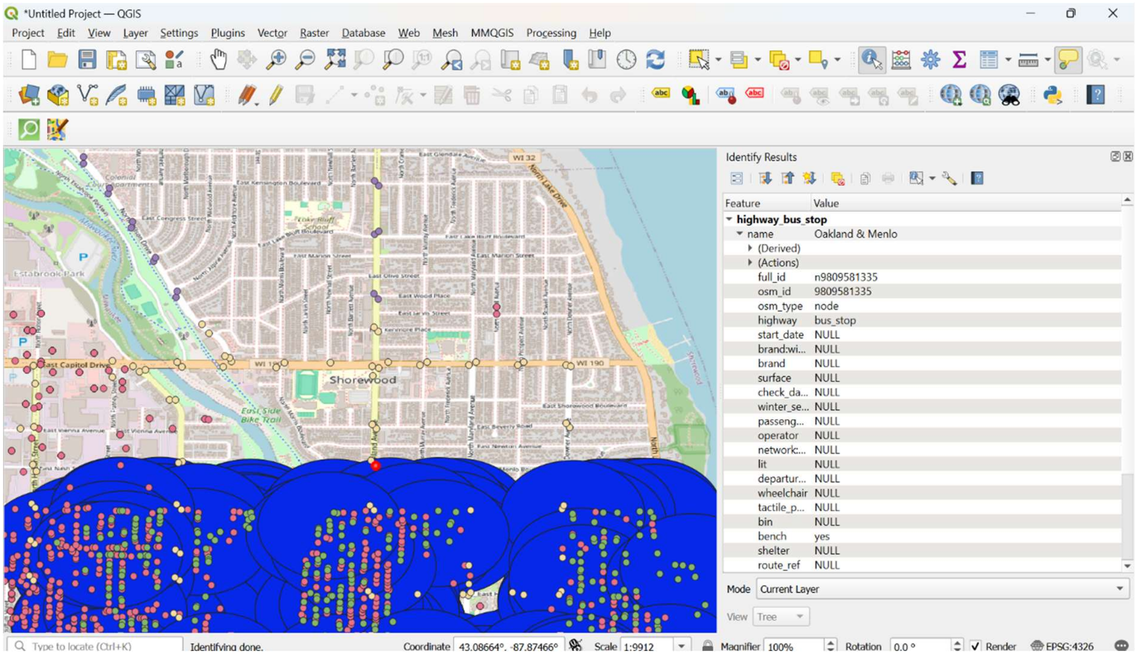
Count — Features Total: 1092, Filtered: 1092, Selected: 0

		Hbath	Lotsize	Sale_date	Sale_price	Location	City	State	Country	latitude	longitude	Sale_year	NUMPOINTS
3	00...	1.00000000000...	8400.00000000...	2019/09/27	531000.000000...	2708 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.0758563500...	-87.876426950...	2019	81
4	00...	0.00000000000...	4800.00000000...	2019/05/03	280000.000000...	3263-3265 N H...	Milwaukee	Wisconsin	United States of...	43.0770269000...	-87.876886099...	2019	52
5	00...	1.00000000000...	2800.00000000...	2019/09/20	325000.000000...	2717-2719 E H...	Milwaukee	Wisconsin	United States of...	43.0776593000...	-87.890843399...	2019	159
6	00...	0.00000000000...	6600.00000000...	2019/11/25	550000.000000...	3277 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0773041500...	-87.875878301...	2019	60
7	00...	0.00000000000...	10800.0000000...	2019/10/29	305000.000000...	3307 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0778565000...	-87.875830148...	2019	56
8	00...	2.00000000000...	11760.0000000...	2019/11/20	755000.000000...	3481 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.0814469999...	-87.873273355...	2019	11
9	00...	1.00000000000...	9072.00000000...	2019/09/27	660000.000000...	3494 N SHEPAR...	Milwaukee	Wisconsin	United States of...	43.0817566500...	-87.873965559...	2019	13
10	00...	1.00000000000...	7200.00000000...	2019/06/24	452000.000000...	3438 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0804133999...	-87.875139687...	2019	25
11	00...	1.00000000000...	7680.00000000...	2019/10/07	535000.000000...	3432 N HACKE...	Milwaukee	Wisconsin	United States of...	43.0802635499...	-87.876265748...	2019	26
12	00...	0.00000000000...	4920.00000000...	2019/03/08	282500.000000...	3412-3414 N D...	Milwaukee	Wisconsin	United States of...	43.0748496000...	-87.877924399...	2019	74
13	00...	1.00000000000...	4800.00000000...	2019/08/27	253200.000000...	3442 N DOWN...	Milwaukee	Wisconsin	United States of...	43.0805164500...	-87.877393742...	2019	22
14	00...	0.00000000000...	5160.00000000...	2019/05/31	315000.000000...	3340-3342 N D...	Milwaukee	Wisconsin	United States of...	43.0748496000...	-87.877924399...	2019	74
15	00...	1.00000000000...	7200.00000000...	2019/04/08	751000.000000...	3365 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0790043000...	-87.875824796...	2019	40
16	00...	1.00000000000...	10800.0000000...	2019/05/24	589000.000000...	2825 E NEWPO...	Milwaukee	Wisconsin	United States of...	43.0792084500...	-87.874657794...	2019	40
17	00...	1.00000000000...	10062.0000000...	2019/03/07	858000.000000...	3347 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.0786223500...	-87.872941099...	2019	29
18	00...	0.00000000000...	1.00000000000...	2019/10/11	295000.000000...	3113 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.0752613000...	-87.871335149...	2019	25
19	00...	0.00000000000...	1.00000000000...	2019/08/30	272500.000000...	3131 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.0752640500...	-87.870847404...	2019	25
20	00...	1.00000000000...	1.00000000000...	2019/08/19	420000.000000...	3330 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.0783647499...	-87.872002614...	2019	21
21	00...	1.00000000000...	1.00000000000...	2019/08/23	480000.000000...	3332 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.0783647499...	-87.872002614...	2019	21

Show All Features

PUBLIC TRANSPORT DATA

This part of the data we go it from open street map using QGIS quick OSM, which gives us view of all the public transport available in the Milwaukee area.



Using Distance to nearest hub we calculated the distance between the property location and the nearest bus stop.

Hub distance — Features Total: 1092, Filtered: 1092, Selected: 0

		Lotsize	Sale_date	Sale_price	Location	City	State	Country	latitude	longitude	Sale_year	HubName	HubDist
1	1	4800	11/1/2019	572000	3118 N MARIET...	Milwaukee	Wisconsin	United States of...	43.07514305	-87.873036463...	2019	n698134843	408.235329110...
2	1	5400	6/28/2019	395000	3233 N HACKE...	Milwaukee	Wisconsin	United States of...	43.07637795	-87.877039050...	2019	n3930647897	112.526979760...
3	1	8400	9/27/2019	531000	2708 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.07585635	-87.876426950...	2019	n3930647897	131.610148398...
4	0	4800	5/3/2019	280000	3263-3265 N H...	Milwaukee	Wisconsin	United States of...	43.0770269	-87.8768861	2019	n698134837	118.242811112...
5	1	2800	9/20/2019	325000	2717-2719 E H...	Milwaukee	Wisconsin	United States of...	43.0776593	-87.8908434	2019	n9809581333	241.108813238...
6	0	6600	11/25/2019	550000	3277 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.07730415	-87.875878301...	2019	n698134837	176.188452957...
7	0	10800	10/29/2019	305000	3307 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0778565	-87.875830148	2019	n698134837	173.547798791...
8	2	11760	11/20/2019	755000	3481 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.081447	-87.873273355...	2019	n9812322053	377.780857136...
9	1	9072	9/27/2019	660000	3494 N SHEPAR...	Milwaukee	Wisconsin	United States of...	43.08175665	-87.873965559...	2019	n9812322053	317.066021658...
10	1	7200	6/24/2019	452000	3438 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0804134	-87.875139687...	2019	n698134831	250.484886102...
11	1	7680	10/7/2019	535000	3432 N HACKE...	Milwaukee	Wisconsin	United States of...	43.08026355	-87.876265748...	2019	n698134831	162.720328542...
12	0	4920	3/8/2019	282500	3412-3414 N D...	Milwaukee	Wisconsin	United States of...	43.0748496	-87.8779244	2019	n698134843	13.8278862119...
13	1	4800	8/27/2019	253200	3442 N DOWN...	Milwaukee	Wisconsin	United States of...	43.0805164500...	-87.877393742...	2019	n698134831	128.113301469...
14	0	5160	5/31/2019	315000	3340-3342 N D...	Milwaukee	Wisconsin	United States of...	43.0748496	-87.8779244	2019	n698134843	13.8278862119...
15	1	7200	4/8/2019	751000	3365 N SUMMI...	Milwaukee	Wisconsin	United States of...	43.0790043	-87.875824796...	2019	n698134831	175.740716339...
16	1	10800	5/24/2019	589000	2825 E NEWPO...	Milwaukee	Wisconsin	United States of...	43.07920845	-87.874657794...	2019	n698134831	265.523351443...
17	1	10062	3/7/2019	858000	3347 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.07862235	-87.872941099...	2019	n698134831	413.963318124...
18	0	1	10/11/2019	295000	3113 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.0752613	-87.871335149...	2019	n3930647897	546.438159149...
19	0	1	8/30/2019	272500	3131 E HAMPS...	Milwaukee	Wisconsin	United States of...	43.07526405	-87.870847404...	2019	n3930647897	586.011640469...
20	1	1	8/19/2019	420000	3330 N LAKE D...	Milwaukee	Wisconsin	United States of...	43.07836475	-87.872002614...	2019	n698134837	489.662746934...

Show All Features

Regression Analysis

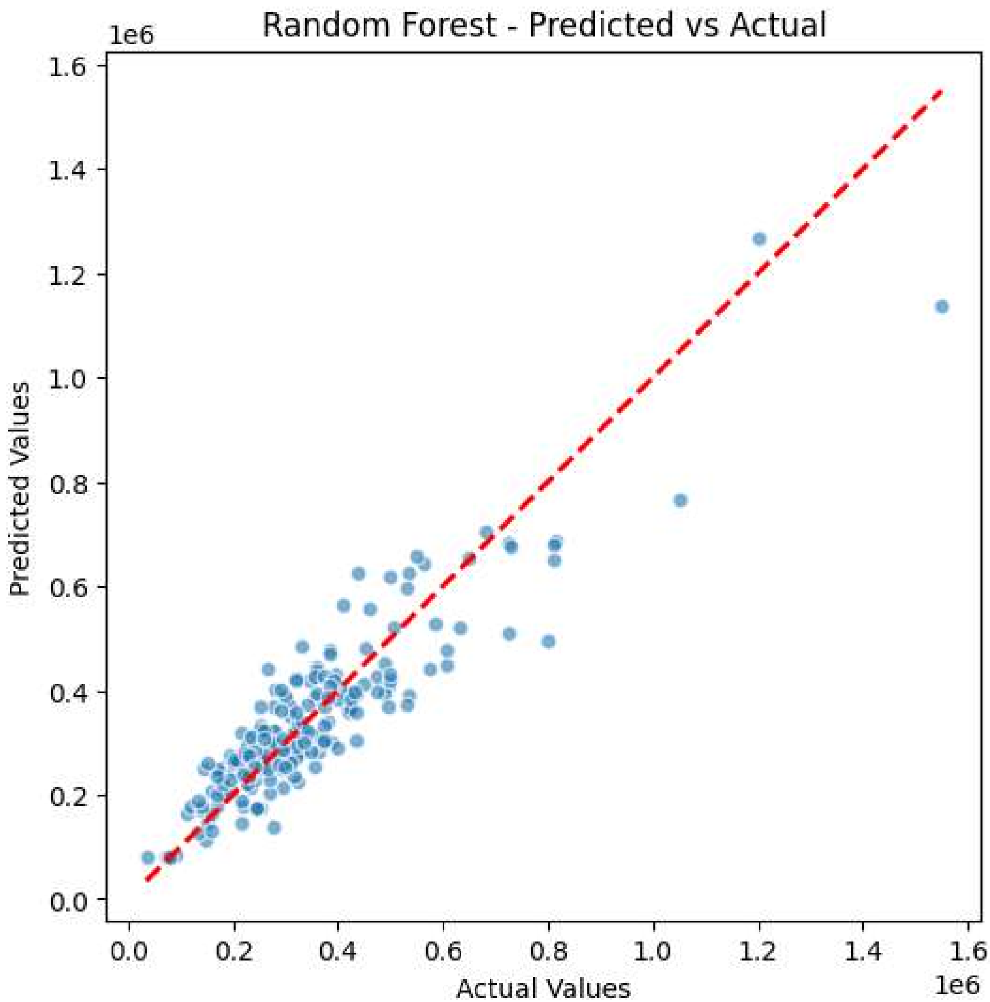
Combined all the data like property data, crime data, public transport data performed regression analysis to get predicted price of the house

Random forest

Mean Squared Error for Random Forest: 6001637364.975826

R^2 Score for Random Forest: 0.8195807770359443

Root Mean Squared Error for Random Forest: 77470.2353486539



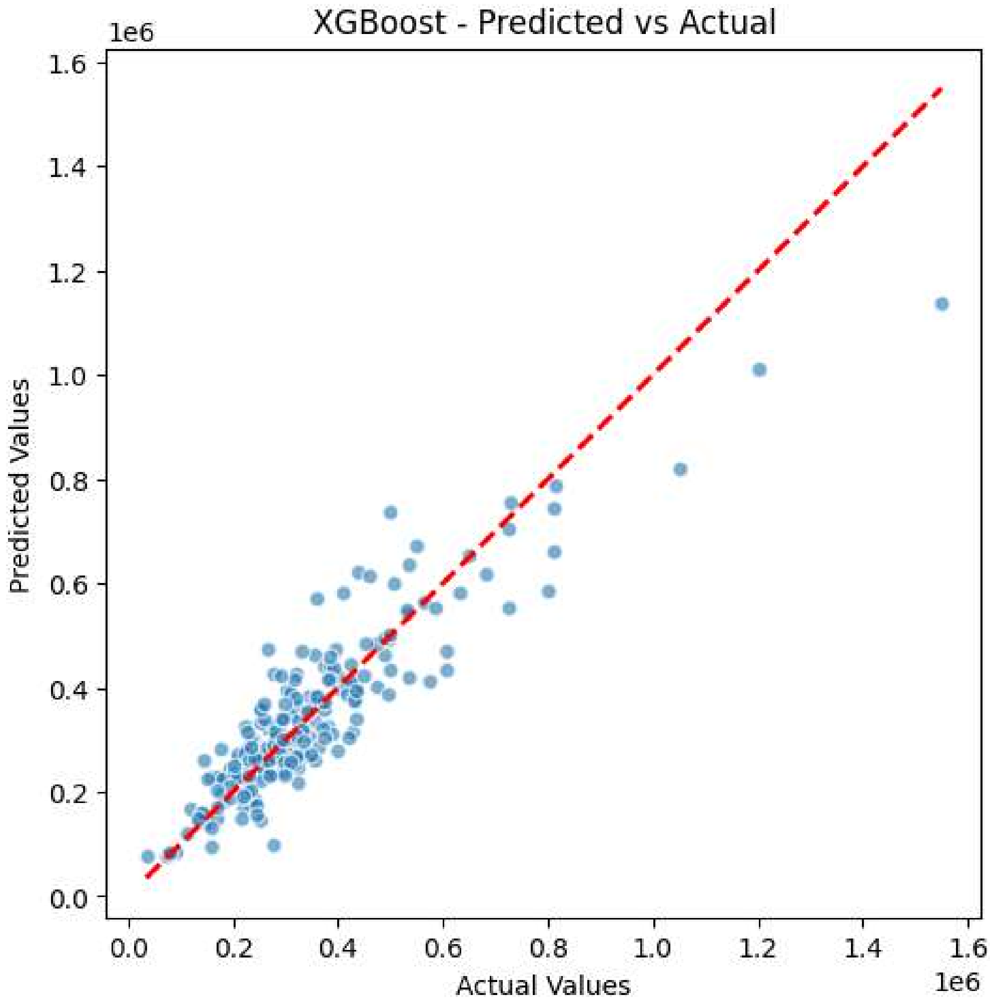
The R square value of 0.81 indicates the model is a good fit.

XGBOOST

Mean Squared Error for XG Boost: 5988610598.128285

R^2 Score for XG Boost: 0.8199723833609255

Root Mean Squared Error for XG Boost: 77386.1137293267



Test Case

Machine learning output

test_case_1= {

"PropertyID": [482], "PropType": ["Residential"], "Address": ["2128-2130 N CAMBRIDGE AV"], "District": [3], "nbhd": [3190], "Style": ["Dplx Bungalow"],"Stories": [2.0], "Year_Built": [1924], "Rooms": [12], "FinishedSqft": [2788.0], "Units": [2], "Bdrms": [6],"Fbath": [2], "Hbath": [0],"Lotsize": [6000], "Sale_date": ["2022/08/12"], "Location": ["2128-2130 N CAMBRIDGE AV,Milwaukee,wisconsin"], "City": ["Milwaukee"], "State": ["Wisconsin"], "Country": ["United States of America"], "latitude": [43.072868], "longitude": [-87.891558], "Sale_year": [2024], "HubDist": [197.156468], "NUMPOINTS": [166]

}

Predicted price for the test case_1:$ 380658.12

LLM OUTPUT

PROMPT-you are a property manager, give me the property price of the property based on the following property location- "PropertyID": [482], "PropType": ["Residential"], "Address": ["2128-2130 N CAMBRIDGE AV"], "District": [3], "nbhd": [3190], "Style": ["Dplx Bungalow"],"Stories": [2.0], "Year_Built": [1924], "Rooms": [12], "FinishedSqft": [2788.0], "Units": [2], "Bdrms": [6],"Fbath": [2], "Hbath": [0],"Lotsize": [6000], "Sale_date": ["2022/08/12"], "Location": ["2128-2130 N CAMBRIDGE AV,Milwaukee,wisconsin"], "City": ["Milwaukee"], "State": ["Wisconsin"], "Country": ["United States of America"], "latitude": [43.072868], "longitude": [-87.891558]

Thus, the estimated price for this property is approximately **$418,200**.

Conclusion

Based on the output the price estimated by large language model and the traditional machine learning is approximately similar.