📌 Real Estate Data Scraping & Analysis

📝 Project Overview:

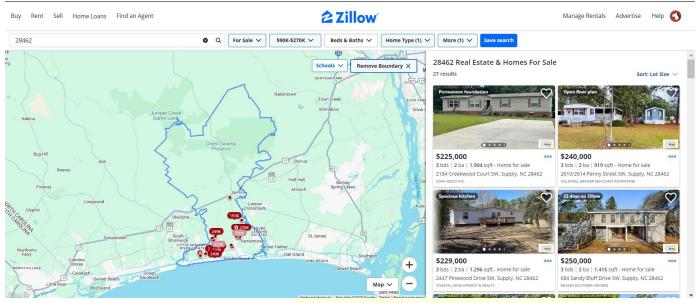|	SALE		SOLD			
SITUS ZIP	Min	Max	Min	Max	Avg. Price	Avg. Price/acre
28462	0.12	0.52	0.1	3	$124,618	$198,413

In this project, I conducted an in-depth **Real Estate Data Scraping & Analysis** for properties listed on **Zillow** and **TheLandPortal**, along with similar platforms such as **Redfin, Realtor.com, DataTree, ParcelFact, and Google Earth**. The primary objective was to extract, clean, and analyze property data to identify key market trends, optimize pricing strategies, and improve investment decision-making.

🔍 Scope of Work:

1️⃣ Data Collection from Zillow & Similar Platforms

✓ Manually searched properties using **ZIP Code (e.g., 28462)**
✓ Filtered properties categorized as **Manufactured Property**
✓ Extracted **Min Acre & Max Acre and Avg. Price, Avg. Price/acre** for properties priced between **$90K – $270K**
✓ Documented property details, including location, acreage, and pricing

2 Identifying Comparable Properties on TheLandPortal & Other Platforms

✓ Used **APN Code, State Name, and County Name** to locate specific parcels
✓ Instead of extracting direct property information, I collected **neighboring property addresses**
✓ Leveraged these addresses on **Zillow & Realtor.com** to obtain **Zestimate values**

3 Data Compilation & Cleaning

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Lot Acres	Purch Yr	Purch Price	20% PP	Adj Home Value	5% Adj HV	Adj Home Addre	TFR Offer %	TFR Offer Price	Average PPA	Avg Offer %	Avg Offer Price	Parcel County	Parcel Zip	APN	OWNER 1 FU
2	0.413		$15,000.00	$ 3,000.00	$ -	$ -		7%	$ 4,040.35	$139,756.00	20%	$ 11,543.85	Garland	71913	030912	Johnson Will
3	0.68		$0.00	$ -	$ -	$ -		7%	$ 4,910.13	$103,154.00	20%	$ 14,028.94	Garland	71901	100-01968-00(	Mauldin Che
4	1.181	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 11,553.63	$139,756.00	20%	$ 33,010.37	Garland	71913	025262	Stout Charles
5	2.31		$0.00	$ -	$ -	$ -		7%	$ 16,680.00	$103,154.00	20%	$ 47,657.15	Garland	71901	083488	Wages Vicki
6	0.413		$0.00	$ -	$ -	$ -		7%	$ 4,040.35	$139,756.00	20%	$ 11,543.85	Garland	71913	015726	Oseguera De
7	2		$0.00	$ -	$ -	$ -		7%	$ 9,408.28	$ 67,202.00	20%	$ 26,880.80	Garland	71964	012893	Childers Card
8	10.46	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 75,529.36	$103,154.00	20%	$ 215,798.17	Garland	71901	100-01792-00(	Heinemann F
9	2.72	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 26,609.54	$139,756.00	20%	$ 76,027.26	Garland	71913	088521	Scholl Lyndo
10	1.904	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 6,588.83	$ 49,436.00	20%	$ 18,825.23	Garland	71956	081296	Castle David
11	2.7	2019-03-28	$55,000.00	$ 11,000.00	$ -	$ -		7%	$ 8,167.64	$ 43,215.00	20%	$ 23,336.10	Garland	71909	100-00860-00(	Womble Mic
12	0.287	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 706.18	$ 35,151.00	20%	$ 2,017.67	Garland	72087	056798	Webb Ozie
13	0.82	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 3,857.39	$ 67,202.00	20%	$ 11,021.13	Garland	71964	012958	Mccaslin Ran
14	0.407	2018-05-29	$15,250.00	$ 3,050.00	$ -	$ -		7%	$ 3,981.65	$139,756.00	20%	$ 11,376.14	Garland	71913	089060	Jenkins Walla
15	3.01	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 16,968.72	$ 80,535.00	20%	$ 48,482.07	Garland	71968	090373	Crank Michae
16	0.273	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 944.72	$ 49,436.00	20%	$ 2,699.21	Garland	71956	056754	Lewis Elbert
17	3		$0.00	$ -	$ -	$ -		7%	$ 3,292.17	$ 15,677.00	20%	$ 9,406.20	Garland	71949	090704	Barfield Jean
18	0.275		$0.00	$ -	$ -	$ -		7%	$ 1,985.71	$103,154.00	20%	$ 5,673.47	Garland	71901	042383	Reynolds Bri
19	0.97		$0.00	$ -	$ -	$ -		7%	$ 4,563.02	$ 67,202.00	20%	$ 13,037.19	Garland	71964	082561	Foreman Ste
20	1.85	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 18,098.40	$139,756.00	20%	$ 51,709.72	Garland	71913	023666	Holder Roger
21	5.07	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 15,337.00	$ 43,215.00	20%	$ 43,820.01	Garland	71909	100-03518-00(	Welch Stewa
22	0.956		$208.00	$ 41.60	$ -	$ -		7%	$ 5,389.40	$ 80,535.00	20%	$ 15,398.29	Garland	71968	024174	Bright Seth
23	0.304	1905-06-15	$0.00	$ 4,200.00	$ -	$ -		7%	$ 2,974.01	$139,756.00	20%	$ 8,497.16	Garland	71913	026277	Dalrymple Fa
24	0.548		$32,750.00	$ 6,550.00	$ -	$ -		7%	$ 5,361.04	$139,756.00	20%	$ 15,317.26	Garland	71913	108610	Brown Prope
25	0.813		$0.00	$ 23,200.00	$ -	$ -		7%	$ 5,870.49	$103,154.00	20%	$ 16,772.84	Garland	71901	017411	Tak Integrati
26	4.47	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 15,468.52	$ 49,436.00	20%	$ 44,195.78	Garland	71956	101003	Estell Ricky E
27	0.264	1905-06-22	$0.00	$ 3,600.00	$ -	$ -		7%	$ 2,582.69	$139,756.00	20%	$ 7,379.12	Garland	71913	028082	Reed Stephe
28	1.22	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 5,739.05	$ 67,202.00	20%	$ 16,397.29	Garland	71964	013363	Morehead El
29	0.38	2018-01-12	$28,500.00	$ 5,700.00	$ -	$ -		7%	$ 3,717.51	$139,756.00	20%	$ 10,621.46	Garland	71913	081172	Eubanks Cha
30	3.66	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 4,016.45	$ 15,677.00	20%	$ 11,475.56	Garland	71949	015195	Head Andrea
31	1.44	1900-01-00	$0.00	$ -	$ -	$ -		7%	$ 6,773.96	$ 67,202.00	20%	$ 19,354.18	Garland	71964	012874	Castorina An
32	1.08		$35,000.00	$ 7,000.00	$ -	$ -		7%	$ 10,565.55	$139,756.00	20%	$ 30,187.30	Garland	71913	087179	Smith Barbar

✅ Organized data in **Google Sheets / Excel** for analysis
✅ Applied **filters, sorting, and basic statistical calculations** for insights
✅ Created a structured dataset with **property acreage, pricing, and trends**

4 Data Analysis & Insights
✅ Identified high-value areas based on **Zestimate valuation trends**
✅ Compared property price fluctuations across different counties
✅ Suggested potential investment opportunities based on **ROI and property appreciation trends**

🛠 Tools & Platforms Used:

🔶 **Real Estate Data Platforms:** Zillow, TheLandPortal, **Redfin, Realtor.com, DataTree, ParcelFact, Google Earth**
🔶 **Manual Data Collection & Entry:** Excel / Google Sheets
🔶 **Data Visualization & Reporting:** Power BI / Google Data Studio
🔶 **Web Scraping (Planned for Future Use):** Selenium + BeautifulSoup
🔶 **Database Management (Planned):** SQL for structured data storage

🚀 Future Automation Implementation Plan

To improve efficiency and scalability, I am currently working on automating the process using **Python (Selenium + BeautifulSoup)**. The next phase will include:

✅ **Automated Data Scraping:** Eliminating manual searching and extracting property data programmatically
✅ **Database Integration:** Storing property records in a structured **SQL database**
✅ **Automated Data Cleaning:** Using **Pandas & NumPy** for faster processing
✅ **Real-Time Analysis & Dashboards:** Implementing **Power BI & Python Visualizations**

🎯 Key Takeaways & Business Impact

✔ Improved accuracy in **property valuation and trend forecasting**
✔ Enhanced efficiency in **lead generation for real estate investments**
✔ Reduced manual effort through **semi-automated processes**
✔ Prepared for **full automation implementation** to handle large-scale datasets

🔊 Conclusion

With my current experience in **manual data collection and analysis**, I have gained in-depth knowledge of **real estate data extraction** using platforms like **Zillow, TheLandPortal, Redfin, Realtor.com, DataTree, ParcelFact, and Google Earth**. Moving forward, I am actively working on automating the entire process using **Python and web scraping tools** to increase efficiency and deliver more scalable solutions for real estate market analysis.