

# Feature Engineering of Geo- Location feature.

**A new approach to deal with location feature in  
datasets.**

Rohan Mahesh Patil

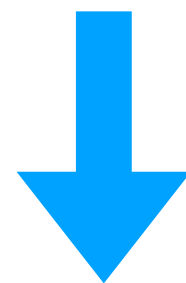
# Methodology

## Steps involved

Get the co-ordinates of locations (city, country, localities) in the dataset.



Using unsupervised algorithms like KMeans and Hierarchical to cluster the locations



Finally replace the location feature column with outcome of clustering model

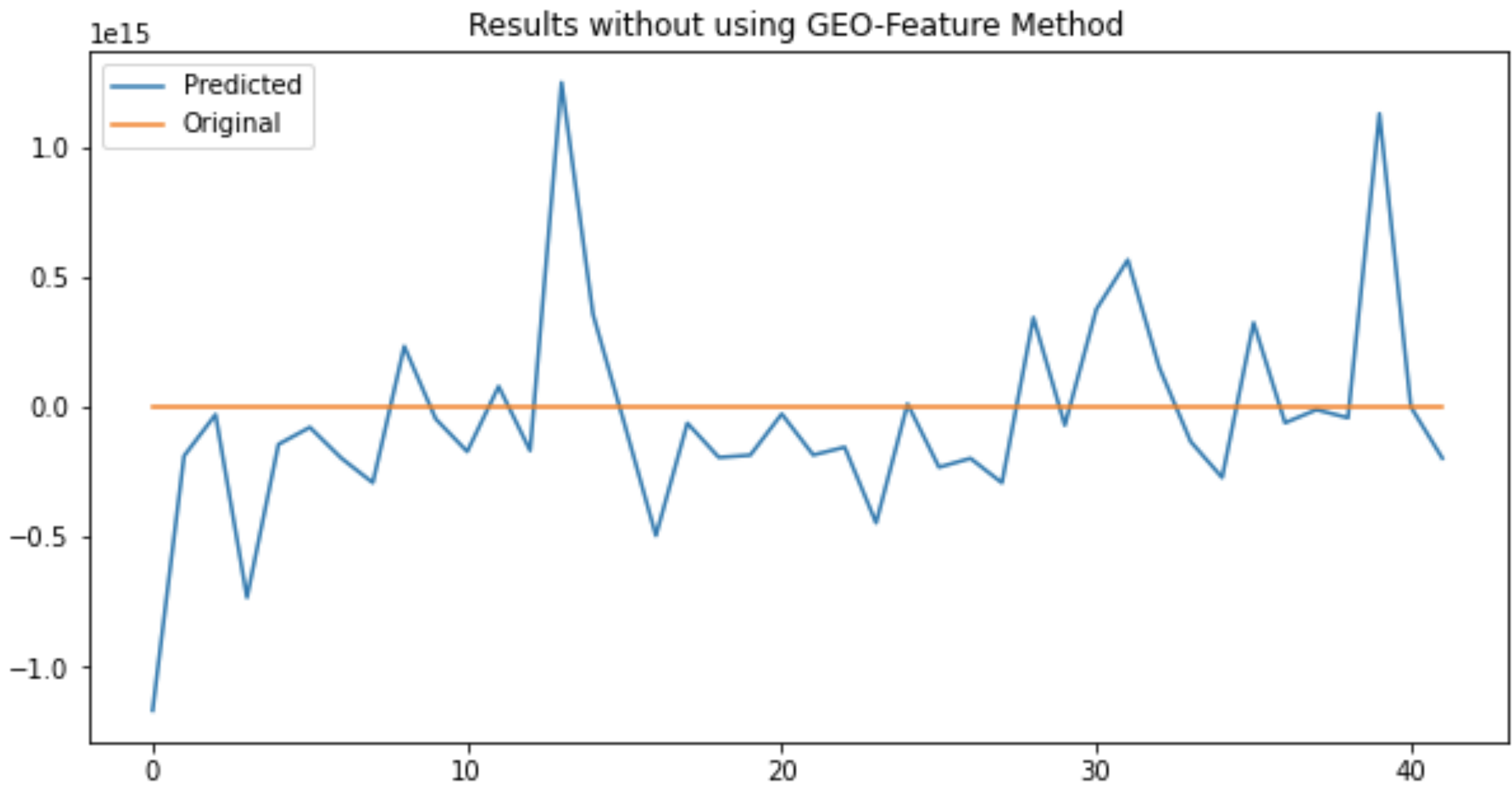
	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

# Dataset after feature engineering

	area_type		availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area		19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area		Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area		Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area		Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area		Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

	area_type	availability	location	size	total_sqft	bath	balcony	price
0	3.0	15751584000000000000	0	2	1056	2.0	1.0	39.07
1	1.0	16383168000000000000	0	4	2600	5.0	3.0	120.00
2	2.0	16383168000000000000	0	3	1440	2.0	3.0	62.00
3	3.0	16383168000000000000	0	3	1521	3.0	1.0	95.00
4	3.0	16383168000000000000	0	2	1200	2.0	1.0	51.00

# Results



## Results

	Normal-Method	GEO-Featuring
<b>MSE</b>	1.610517E+29	6178.271836
<b>MAE</b>	2.697909E+14	38.971816
<b>RMSE</b>	4.013125E+14	78.601984