task-2

May 15, 2025

1 Task 2: Data Transformation & Feature Engineering

[12]:
```
from sklearn.preprocessing import MinMaxScaler

# let copy original data
df_transformed = df.copy()

# Normalize columns used in scoring
scaler = MinMaxScaler()
df_transformed[['Normalized_Irradiance', 'Normalized_GridAccess',
                'Normalized_Infrastructure', 'Normalized_Cost']] = scaler.
 ↪fit_transform(
    df_transformed[['Solar_Irradiance_kWh_m2_day',
                    'Grid_Access_Percent',
                    'Infrastructure_Index',
                    'Electricity_Cost_USD_per_kWh']]
)

# Invert normalized grid access: we want higher score if access is lower
df_transformed['Inverse_GridAccess'] = 1 -␣
 ↪df_transformed['Normalized_GridAccess']

# Apply weights:
# - Irradiance: 35%
# - Inverse grid access: 25% (priority if underserved)
# - Infrastructure index: 20%
# - Electricity cost (normalized): 20%

df_transformed['Solar_Access_Score'] = (
    0.35 * df_transformed['Normalized_Irradiance'] +
    0.25 * df_transformed['Inverse_GridAccess'] +
    0.20 * df_transformed['Normalized_Infrastructure'] +
    0.20 * df_transformed['Normalized_Cost']
)

# Sort by score for inspection
```

```python
df_transformed_sorted = df_transformed.sort_values(by='Solar_Access_Score',␣
↪ascending=False)
df_transformed_sorted
```

[12]:

	Region	Solar_Irradiance_kWh_m2_day	Rural_Pop_Density_per_km2 \
31	Region_32	7.35	111
6	Region_7	7.08	376
2	Region_3	6.15	64
47	Region_48	6.56	304
30	Region_31	4.90	456
12	Region_13	5.74	188
9	Region_10	6.04	178
0	Region_1	6.00	90
33	Region_34	4.44	342
44	Region_45	4.02	306
7	Region_8	6.27	58
41	Region_42	5.67	150
4	Region_5	5.27	114
48	Region_49	5.84	447
1	Region_2	5.36	206
34	Region_35	6.32	148
45	Region_46	4.78	54
17	Region_18	5.81	338
20	Region_21	6.97	280
42	Region_43	5.38	480
40	Region_41	6.24	276
38	Region_39	4.17	84
32	Region_33	5.49	265
22	Region_23	5.57	77
3	Region_4	7.02	350
24	Region_25	4.96	250
25	Region_26	5.61	377
27	Region_28	5.88	467
28	Region_29	4.90	82
39	Region_40	5.70	498
21	Region_22	5.27	90
5	Region_6	5.27	394
26	Region_27	4.35	317
18	Region_19	4.59	428
11	Region_12	5.03	112
36	Region_37	5.71	409
43	Region_44	5.20	180
19	Region_20	4.09	310
8	Region_9	5.03	393
10	Region_11	5.04	185
37	Region_38	3.54	263
35	Region_36	4.28	221

16	Region_17	4.49	468
23	Region_24	4.08	184
15	Region_16	4.94	212
46	Region_47	5.04	267
14	Region_15	3.78	441
29	Region_30	5.21	97
49	Region_50	3.74	408
13	Region_14	3.59	130

	Grid_Access_Percent	Infrastructure_Index	Electricity_Cost_USD_per_kWh \
31	46.4	0.48	0.39
6	55.7	0.68	0.38
2	28.3	0.49	0.36
47	73.4	0.82	0.37
30	20.0	0.86	0.28
12	35.2	0.46	0.39
9	30.4	0.59	0.27
0	23.0	0.39	0.31
33	32.3	0.79	0.37
44	39.0	0.90	0.40
7	62.2	0.57	0.34
41	61.9	0.89	0.27
4	35.1	0.44	0.37
48	31.1	0.33	0.32
1	73.3	0.88	0.35
34	60.1	0.67	0.23
45	38.5	0.88	0.24
17	66.8	0.48	0.40
20	54.2	0.31	0.25
42	50.3	0.56	0.33
40	36.4	0.32	0.25
38	38.3	0.69	0.36
32	42.9	0.56	0.25
22	51.2	0.69	0.21
3	53.0	0.22	0.22
24	44.3	0.43	0.32
25	29.2	0.35	0.19
27	88.0	0.77	0.25
28	40.4	0.44	0.29
39	32.6	0.36	0.17
21	36.4	0.56	0.15
5	87.2	0.64	0.36
26	46.7	0.70	0.26
18	42.2	0.77	0.15
11	60.5	0.26	0.38
36	71.9	0.35	0.29
43	24.9	0.38	0.11

19	27.9	0.76	0.11
8	72.2	0.51	0.30
10	65.3	0.61	0.21
37	40.2	0.58	0.30
35	56.4	0.71	0.21
16	86.0	0.89	0.23
23	86.2	0.80	0.32
15	72.1	0.53	0.25
46	72.2	0.59	0.18
14	64.9	0.76	0.19
29	68.6	0.27	0.18
49	94.8	0.40	0.39
13	90.7	0.37	0.19

	Terrain_Ruggedness_Score	Normalized_Irradiance	Normalized_GridAccess	\
31	0.19	1.000000	0.352941	
6	0.19	0.929134	0.477273	
2	0.57	0.685039	0.110963	
47	0.30	0.792651	0.713904	
30	0.63	0.356955	0.000000	
12	0.93	0.577428	0.203209	
9	0.37	0.656168	0.139037	
0	0.33	0.645669	0.040107	
33	0.40	0.236220	0.164439	
44	0.86	0.125984	0.254011	
7	0.27	0.716535	0.564171	
41	0.16	0.559055	0.560160	
4	0.08	0.454068	0.201872	
48	0.54	0.603675	0.148396	
1	0.55	0.477690	0.712567	
34	0.05	0.729659	0.536096	
45	0.25	0.325459	0.247326	
17	0.25	0.595801	0.625668	
20	0.85	0.900262	0.457219	
42	0.98	0.482940	0.405080	
40	0.33	0.708661	0.219251	
38	0.44	0.165354	0.244652	
32	0.07	0.511811	0.306150	
22	0.71	0.532808	0.417112	
3	0.98	0.913386	0.441176	
24	0.30	0.372703	0.324866	
25	0.42	0.543307	0.122995	
27	0.61	0.614173	0.909091	
28	0.08	0.356955	0.272727	
39	0.67	0.566929	0.168449	
21	0.14	0.454068	0.219251	
5	0.31	0.454068	0.898396	

26	0.26	0.212598	0.356952
18	0.30	0.275591	0.296791
11	0.84	0.391076	0.541444
36	0.03	0.569554	0.693850
43	0.84	0.435696	0.065508
19	0.32	0.144357	0.105615
8	0.49	0.391076	0.697861
10	0.39	0.393701	0.605615
37	0.58	0.000000	0.270053
35	0.89	0.194226	0.486631
16	0.36	0.249344	0.882353
23	0.55	0.141732	0.885027
15	0.67	0.367454	0.696524
46	0.04	0.393701	0.697861
14	0.21	0.062992	0.600267
29	0.01	0.438320	0.649733
49	0.33	0.052493	1.000000
13	0.07	0.013123	0.945187

	Normalized_Infrastructure	Normalized_Cost	Inverse_GridAccess	\
31	0.382353	0.965517	0.647059	
6	0.676471	0.931034	0.522727	
2	0.397059	0.862069	0.889037	
47	0.882353	0.896552	0.286096	
30	0.941176	0.586207	1.000000	
12	0.352941	0.965517	0.796791	
9	0.544118	0.551724	0.860963	
0	0.250000	0.689655	0.959893	
33	0.838235	0.896552	0.835561	
44	1.000000	1.000000	0.745989	
7	0.514706	0.793103	0.435829	
41	0.985294	0.551724	0.439840	
4	0.323529	0.896552	0.798128	
48	0.161765	0.724138	0.851604	
1	0.970588	0.827586	0.287433	
34	0.661765	0.413793	0.463904	
45	0.970588	0.448276	0.752674	
17	0.382353	1.000000	0.374332	
20	0.132353	0.482759	0.542781	
42	0.500000	0.758621	0.594920	
40	0.147059	0.482759	0.780749	
38	0.691176	0.862069	0.755348	
32	0.500000	0.482759	0.693850	
22	0.691176	0.344828	0.582888	
3	0.000000	0.379310	0.558824	
24	0.308824	0.724138	0.675134	
25	0.191176	0.275862	0.877005	

27	0.808824	0.482759	0.090909
28	0.323529	0.620690	0.727273
39	0.205882	0.206897	0.831551
21	0.500000	0.137931	0.780749
5	0.617647	0.862069	0.101604
26	0.705882	0.517241	0.643048
18	0.808824	0.137931	0.703209
11	0.058824	0.931034	0.458556
36	0.191176	0.620690	0.306150
43	0.235294	0.000000	0.934492
19	0.794118	0.000000	0.894385
8	0.426471	0.655172	0.302139
10	0.573529	0.344828	0.394385
37	0.529412	0.655172	0.729947
35	0.720588	0.344828	0.513369
16	0.985294	0.413793	0.117647
23	0.852941	0.724138	0.114973
15	0.455882	0.482759	0.303476
46	0.544118	0.241379	0.302139
14	0.794118	0.275862	0.399733
29	0.073529	0.241379	0.350267
49	0.264706	0.965517	0.000000
13	0.220588	0.275862	0.054813

	Solar_Access_Score
31	0.781339
6	0.777380
2	0.713849
47	0.704733
30	0.680411
12	0.664989
9	0.664068
0	0.653889
33	0.638525
44	0.630592
7	0.621306
41	0.613033
4	0.602472
48	0.601368
1	0.598685
34	0.586468
45	0.585852
17	0.578584
20	0.573809
42	0.569483
40	0.569182
38	0.557360

32	0.549148
22	0.539406
3	0.535253
24	0.505822
25	0.502817
27	0.496004
28	0.495596
39	0.488869
21	0.481697
5	0.480268
26	0.479796
18	0.461610
11	0.449487
36	0.438254
43	0.433175
19	0.432945
8	0.428740
10	0.420063
37	0.419403
35	0.409404
16	0.396500
23	0.393765
15	0.392206
46	0.370429
14	0.335976
29	0.303961
49	0.264417
13	0.117586

1.0.1 Task 2: Data Transformation & Feature Engineering

1.0.2 Objective

To assist in prioritizing regions for solar energy investment, I've created a composite **"Solar Access Score"** based on four weighted indicators:

- **Solar Irradiance (35%)** – Primary driver for solar yield
- **Inverse Grid Access (25%)** – Regions with poor grid access are higher priority
- **Infrastructure Index (20%)** – Indicates readiness for deployment logistics
- **Electricity Cost (20%)** – Higher cost regions offer greater economic return

---

1.0.3 Calculation Breakdown

Each component was **normalized** using Min-Max scaling to ensure comparability:

- **Inverse Grid Access** = `1 - (normalized grid access)` to prioritize underserved areas

- **Solar Access Score** is computed as a weighted sum of the scaled inputs

---

1.0.4 Business Justification for Weighting

The chosen weights reflect Prime Frontier's operational focus on:

- **Maximizing ROI and energy yield** (heavier weight to solar irradiance)
- **Targeting under-electrified regions** (emphasized via inverse grid access)
- **Feasibility of implementation** (logistics and access depend on infrastructure)
- **Financial leverage** (higher electricity cost = greater savings from solar)