Cluster Feature Analysis Report for 200 Turbines (Nearest Neighbors)

This report summarizes the clustering results for 200 wind turbines using a nearest-neighbor clustering approach. Each cluster is interpreted based on power output, variation, ramp behavior, and operational status.

Summary

- **Cluster 0**: Faulty or shut-down turbines (2)
- **Cluster 1**: High-performance turbines with minimal downtime （29）
- **Cluster 2**: High ramping behavior turbines （3）
- **Cluster 3**: Limited operation or ramp-down group （2）
- **Cluster 4**: Mildly stable with low ramping （6）
- **Cluster 5**: Dominant, stable baseline group （126）
- **Cluster 6**: Downtime-prone mid-risk group （29）
- **Cluster 7**: Ramp-sensitive edge group （3）

Conlusion: Cluster 1,2,3,4,5,6,7 come into prediction. (Cluster 0 are outliers)

Evaluation Metrics for Wind Turbine Clustering

1. `mean_power_scaled_mean`: Overall Average Power (Standardized)

- **> 0** → Higher-than-average power output, good performance
- **< 0** → Lower-than-average power, possibly due to environmental or performance issues

2. `std_power_scaled_mean`: Power Output Fluctuation (Standardized)

- **> 0** → Large variations in power, possibly caused by wind changes or control system behavior
- **< 0** → Smaller fluctuations, more stable output and easier to forecast

3. `cv_scaled_mean`: Coefficient of Variation (Relative Instability)

- **> 0** → High variability per unit of power, potentially less reliable
- **< 0** → Relatively stable output and more consistent operation

4. `zero_ratio_mean`: Proportion of Downtime

- **Close to 0** → High availability, turbines run continuously
- **Close to 1** → Long shutdown periods or faults, affecting generation

5. `ramp_mean_scaled_mean`: Average Power Change Trend

- **> 0** → Ramp-up trend, possibly due to increasing wind speed or gradual engagement
- **< 0** → Ramp-down trend, possibly due to decommissioning, maintenance, or performance decline

6. `ramp_std_scaled_mean`: Volatility of Power Changes

- **> 0** → Sharp or frequent power swings, harder to predict
- **< 0** → Smoother changes, more stable and easier to model

7. Cluster Standard Deviation (After Scaling)

- **std > 1** → The group's variability is greater than average
- **std < 1** → The group's variability is lower than average

📊 Cluster Mean Statistics

Cluster	Count	mean_power_scaled_mean	std_power_scaled_mean	cv_scaled_mean	zero_ratio_mean	ramp_mean_scaled_mean	ramp_st
0	2	-0.7757	-0.9051	9.3432	0.9987	0.2329	
1	29	2.1545	2.1329	-0.181	0.0699	-1.1817	
2	3	1.0346	1.271	-0.1343	0.133	3.9492	
3	2	-0.185	-0.0515	-0.0885	0.2577	-1.1753	
4	6	0.5352	0.7771	-0.1263	0.09	0.5775	
5	126	-0.356	-0.3503	-0.1089	0.1737	0.0964	
6	29	-0.7082	-0.7916	0.0514	0.4035	0.1785	
7	3	-0.4917	-0.4425	0.0436	0.4251	1.1719	

📉 Cluster Standard Deviation Statistics

Cluster	mean_power_scaled_std	std_power_scaled_std	cv_scaled_std	zero_ratio_std	ramp_mean_scaled_std	ramp_std_scaled_std
0	0	0.0018	4.665	0.0011	0	0.0085
1	0.9598	0.8792	0.013	0.0225	1.8051	0.9089
2	0.6542	0.5339	0.0473	0.0498	1.2317	0.4925
3	0.1283	0.2094	0.0091	0.067	0.1443	0.3406
4	0.0637	0.1205	0.012	0.0183	0.4147	0.1467
5	0.0941	0.0956	0.0392	0.0903	0.2209	0.0935
6	0.0527	0.085	0.1556	0.2156	0.1093	0.0691
7	0.1932	0.1833	0.1909	0.3022	0.1412	0.1732

Cluster Details

Cluster 0 — Non-operational Group (Very High CV & Downtime)

- **Count**: 2 turbines
- **Highlights**:
  - `cv_scaled_mean`: 9.34 (extremely unstable)
  - `zero_ratio_mean`: 0.9987 (almost completely offline)
  - `ramp_mean_scaled_mean`: 0.23 (slightly positive), `ramp_std_scaled_mean`: -0.86 (very low variation)
  - Std devs are all nearly 0 → the two turbines behave almost identically.
- **Interpretation**: Severely underperforming or faulty turbines; likely not suitable for forecasting.

Cluster 1 — High-performance turbines with minimal downtime

- **Count**: 29 turbines
- **Highlights**:
  - `mean_power_scaled_mean`: 2.15, `std_power_scaled_mean`: 2.13
  - `zero_ratio_mean`: 0.07 (very low downtime)
  - High ramp fluctuation: `ramp_std_scaled_mean`: 2.14
  - Moderate variation across turbines (std devs between 0.01–1.8)
- **Interpretation**: Strong performers with high output and consistent availability. Ideal for power forecasting models.

Cluster 2 — High ramping behavior turbines

- **Count**: 3 turbines
- **Highlights**:
  - `ramp_mean_scaled_mean`: 3.95 — very aggressive ramping behavior
  - `ramp_std_scaled_mean`: 1.25 — high variability
  - Output: `mean_power_scaled_mean`: 1.03, decent performance
- **Interpretation**: Likely located in turbulent wind zones. Useful for studying ramp forecasting.

Cluster 3 — Limited Operation + Ramp-Down Group

- **Count**: 2 turbines
- **Highlights**:
  - `mean_power_scaled_mean`: -0.185 (below average)
  - `ramp_mean_scaled_mean`: -1.17 (significant ramp-down trend)
  - `zero_ratio_mean`: 0.26 (moderate downtime)
  - Very small standard deviations — similar behavior
- **Interpretation**: Possibly operating in declining performance or controlled shutdown states.

Cluster 4 — Mildly stable with low ramping

- **Count**: 6 turbines
- **Highlights**:
  - `mean_power_scaled_mean`: 0.53, `std_power_scaled_mean`: 0.78 — above average
  - `ramp_mean_scaled_mean`: 0.58 — healthy ramp-up tendency
  - All std devs are small → internally consistent group
- **Interpretation**: Stable, well-performing turbines with predictable behavior.

Cluster 5 — Dominant, stable baseline group

- **Count**: 126 turbines
- **Highlights**:

    - All metrics close to zero → balanced power, low ramp activity
    - `zero_ratio_mean`: 0.17 — moderate availability
    - Low standard deviations — homogeneous performance
- **Interpretation**: Mainstream turbines with reliable output, ideal for baseline modeling.

Cluster 6 — Downtime-prone mid-risk group

- **Count**: 29 turbines
- **Highlights**:
    - `zero_ratio_mean`: 0.40 — relatively high downtime
    - `ramp_mean_scaled_mean`: 0.18 (slightly ramp-up), `ramp_std_scaled_mean`: -0.76 (low ramp variability)
- **Interpretation**: Mid-risk turbines with reduced reliability. Worth modeling separately.

Cluster 7 — Ramp-Sensitive Edge Group

- **Count**: 3 turbines
- **Highlights**:
    - `ramp_mean_scaled_mean`: 1.17 — high ramping behavior
    - `zero_ratio_mean`: 0.43 — frequent shutdowns
- **Interpretation**: On the edge of operational stability, but may contain interesting dynamic patterns.