Warehouse Dataset Documentation (W_data)

Dataset Title

Warehouse Storage Capacity and Spatial Distribution Dataset (W_data)

Dataset Summary

This dataset contains 4,132 warehouse records across 11 well-defined attributes, including location coordinates, storage capacities, utilization status, and ranking. It is designed to support research in logistics optimization, spatial clustering, capacity forecasting, and IIoT-based smart warehouse management.

File Contents

- warehouse_data.csv: Main dataset file (4,132 rows x 11 columns)
- README.txt: This documentation file
- dataset_image.png: Visual summary of dataset structure and capacity status

Data Dictionary

1. Warehouse ID: Categorical - Unique identifier for each warehouse (4,132 unique IDs)
2. Warehouse Name: Categorical - Human-readable name of the warehouse (4,132 unique names)
3. X_Coordinate: Numerical - X-axis location (normalized coordinate) [4 - 150]
4. X_Squared: Numerical - Square of X_Coordinate [16 - 22,500]
5. Y_Coordinate: Numerical - Y-axis location [3 - 149]
6. Y_Squared: Numerical - Square of Y_Coordinate [9 - 22,201]
7. Total Capacity: Numerical - Maximum storage capacity [623 - 5,962]
8. Used Capacity: Numerical - Currently occupied capacity [0 - 5,962]
9. Available Capacity: Numerical - Remaining capacity [0 - 5,910]
10. Capacity Status: Categorical - Utilization level {Full, Near Max, Optimized, Balanced, Under}
11. Rank: Ordinal - Ranking based on utilization [1 (highest) - 5 (lowest)]

Data Source and Synthesis

- Structure based on TI Supply Chain API - Storage Locations specification.
  https://developers.google.com/ti-supply-chain/docs/specification/logistics/wms/v1/pilot/storage_locations

- Synthesized using Gretel.ai to ensure realistic distributions and dependencies.
- No missing values, no duplicates, and balanced categorical distributions.

Instructions for Use

You can use this dataset for:

Warehouse Dataset Documentation (W_data)

- Clustering warehouses based on spatial and capacity features
- Detecting outlier facilities (e.g., underutilized or overfull warehouses)
- Benchmarking smart storage systems or routing algorithms
- Evaluating privacy-preserving clustering techniques in IIoT logistics

Preprocessing Example (Python):

```
import pandas as pd
df = pd.read_csv('warehouse_data.csv')
df['Rank'] = df['Rank'].astype('category')
df = pd.get_dummies(df, columns=['Capacity Status'], drop_first=True)
```

License