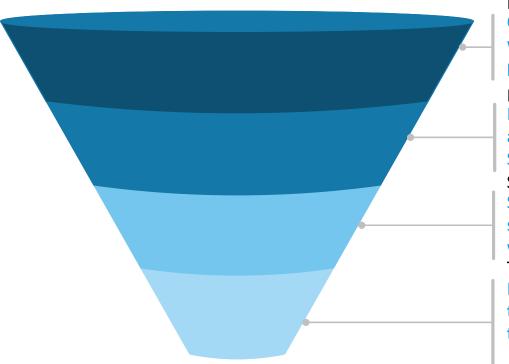
# **Data Preprocessing Pipeline**



#### Raw Dataset

COVID-19 time series of cases for cities, combined with spatial data on city connectivity and population.

#### **Preprocessing Dataset**

Preprocess spatial data into nodes, edges, and normalized population weights.
Standardize time series data using z-score.

## Snapshots

Segment the preprocessed data into snapshots, each defined by a specific window size and forecast horizon.

### Training and Tests

Partition the snapshots into training and test sets, with 80% of the data used for training and 20% for testing.