

# Beijing Air Quality Analysis

Chenghui Zhu, Tuanzhang Li, Advisor: Rajendra Raj | Rochester Institute of Technology | Department of Computer Science

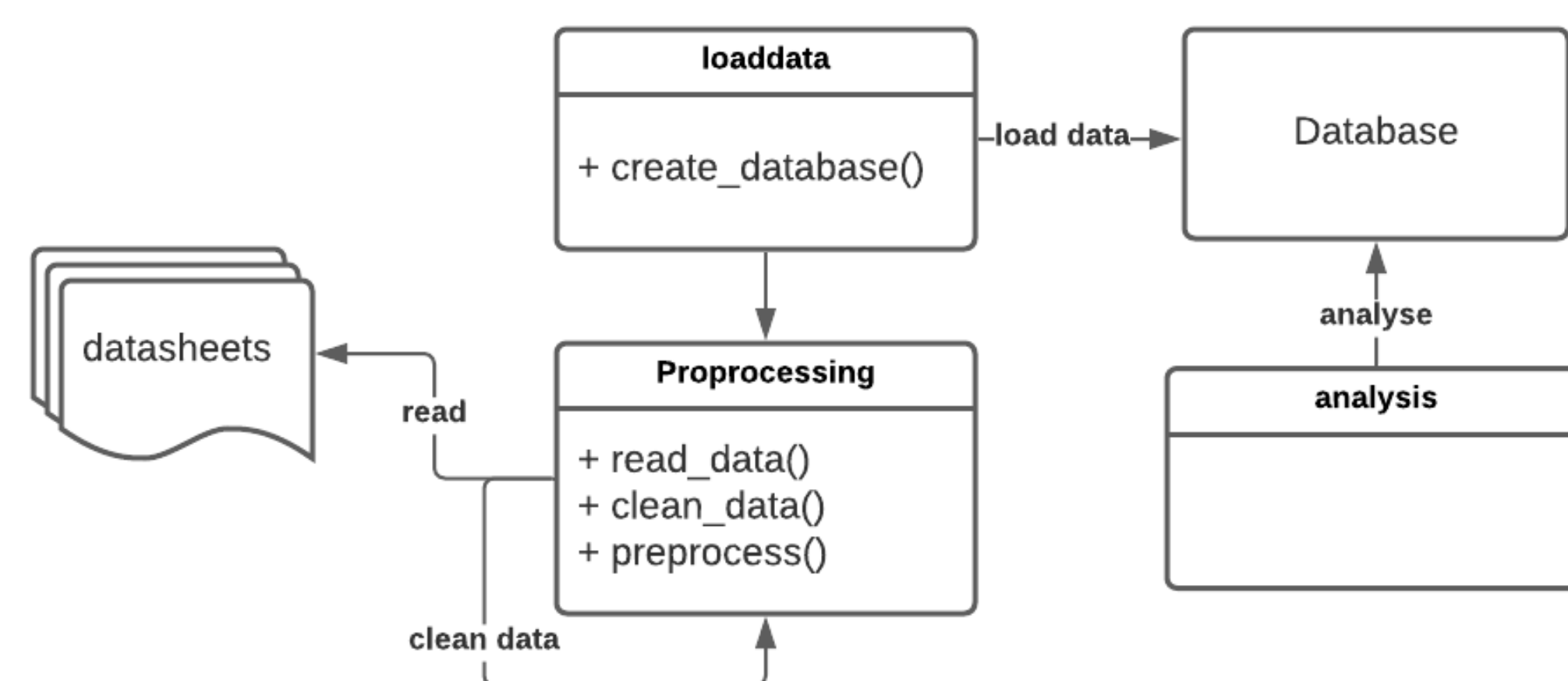
## Motivation & Goal

- Sustainable development has been a universal and urgent concept. Analysis for air quality is required to provide possible solutions.
- Finding critical factors that determine the air quality which is significant to the environment.
- Learning python and SQL to build database models and exploring data mining techniques to make reasonable analysis.

## Background

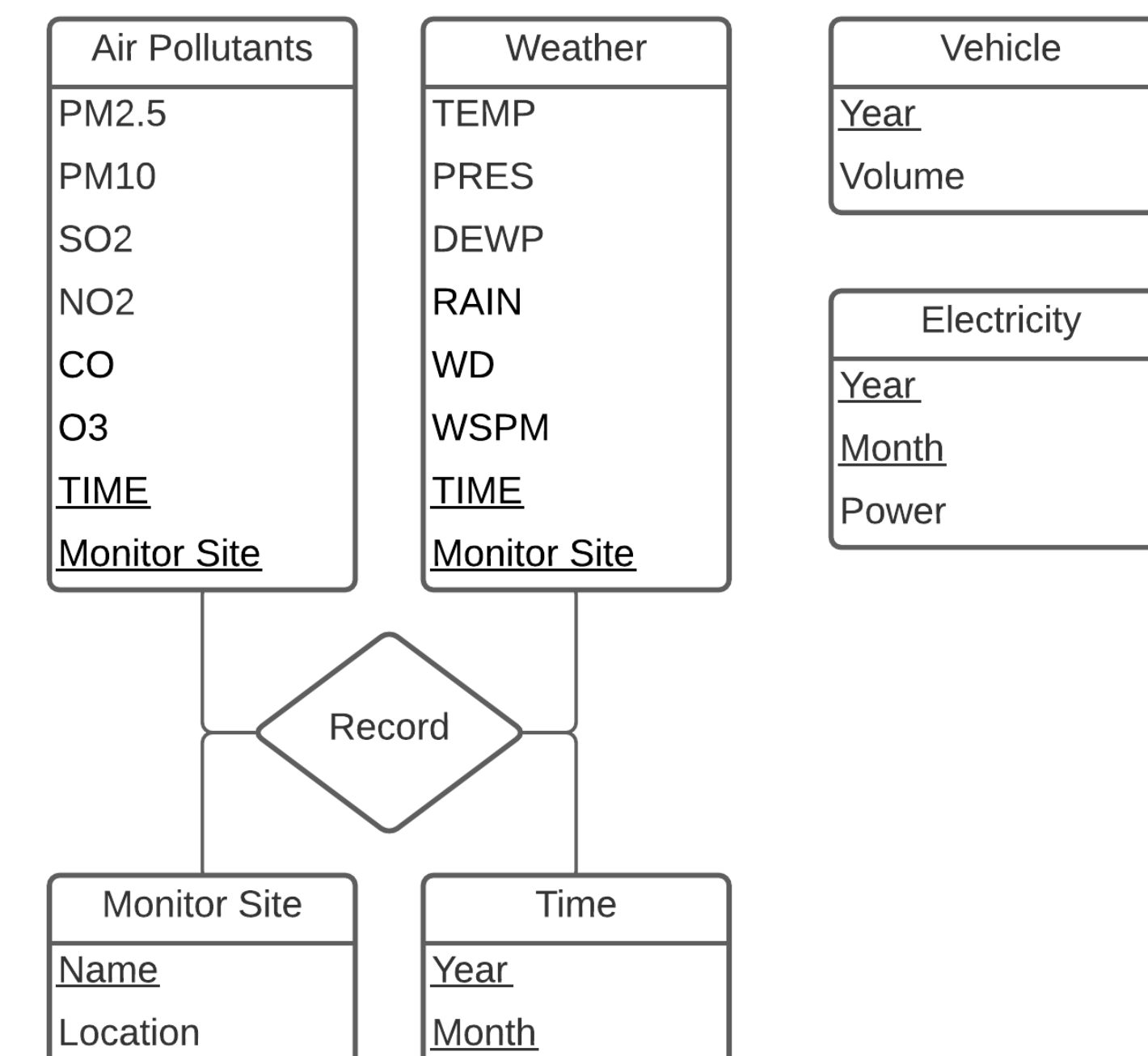
- Air pollution requires both ethical considerations and legal protections are essential to solve it.
- Data mining techniques like clustering [1], outlier detection, and association [2], are applied in a lot of fields to solve particular problems.
- SQLALchemy used to connect the SQL server and python.
- Pandas library used for data preprocessing and cleaning.

## Approaches & solutions



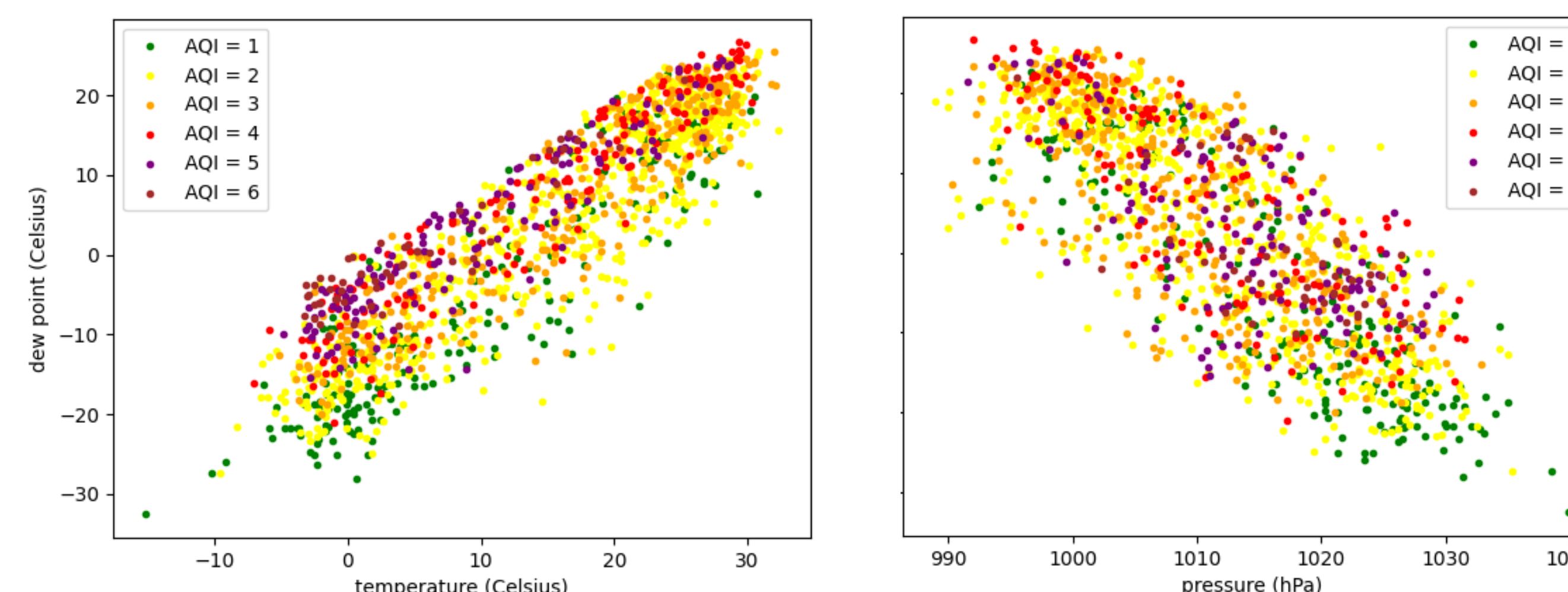
- Designing entity relationship (ER) model based on attribute types and topics.
- Creating SQL database and importing original csv data into database via SQLALchemy.
- Managing data cleaning and preprocessing to smoothen data and exclude outliers. Different strategies are applied.
- Analyzing air quality by clustering. Investigating the relationship between concentrations of air pollutants and the weather conditions.
- Analyzing air quality by association. Exploring the influence of different factors on air quality

## ER Model Design



ER model built for Beijing Multi-Site Air-Quality Data dataset. Air pollutants and weather conditions were recorded at a certain time in certain monitor site. Vehicle and electricity are two separated tables.

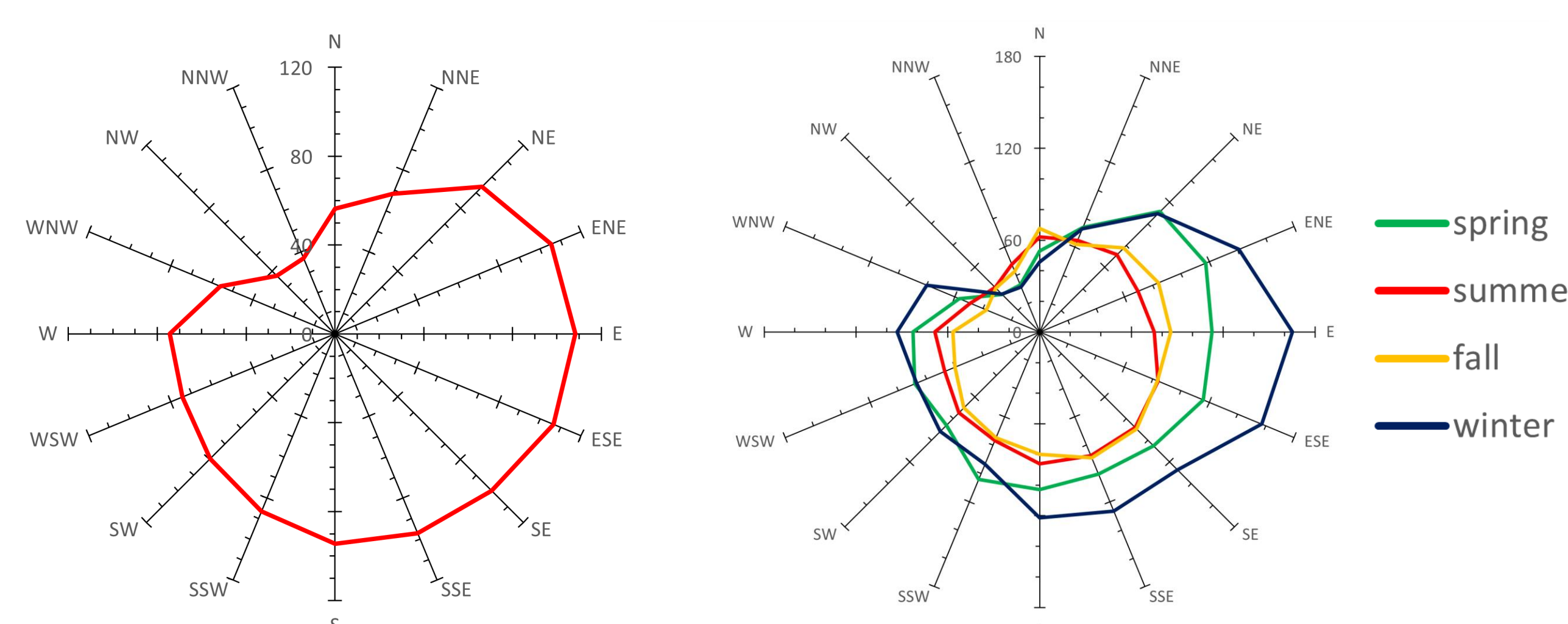
## Clustering Analysis



Finding potential relations between weather conditions and air quality via clustering (at Aotizhongxin). The air quality index (AQI)\* is well clustered by temperature and dew point (left figure), but badly clustered by pressure and dew point (right figure).

\*AQI was calculated via Technical Regulation on Ambient Air Quality Index (on trial) [3].

## Associating Analysis



Sample analysis of data collected in Aotizhongxin monitor site: the influence of wind direction on PM2.5. Data is averaged annually (left figure) or quarterly (right figure).

## Observation & Results

- AQI clustered better with temperature and dew point rather than pressure and dew point.
- A blurred separation between high AQI and low AQI can be observed in temperature and dew point clusters.
- The average PM 2.5 concentration is relatively low when the wind comes from the northwest direction.
- The PM 2.5 concentration does not fluctuate by seasons obviously.

## Lessons Learned

- Mistakes in data preprocessing are well-solved. The improvement also suggest the directions of analysis.
- The cluster relation between temperature and dew point has potential to be used to forecast the air quality.
- Pollutant concentration affected by wind direction and seasons can be used to guide the geographical design and distribution of the city.

## Status & Concluding Remarks

- Designing the ER model based on the air quality data as well as the supporting data.
- Preprocessing the original data using different strategies and successfully building the database in SQL via python.
- For the data mining component, multiple techniques were proposed and some primary visualizations were displayed as well.
- For future work, more efficient SQL querying can be used to locate the specific data. It also benefits further data mining component by finding either the hidden patterns or the relationship between attributes.

## Works Cited

- [1] T Parimalam and K Meenakshi Sundaram. 2017. Efcient Clustering Techniques for Web Services Clustering. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). 1–4. <https://doi.org/10.1109/ICCIC.2017.8524480>
- [2] Liye Song. 2017. Impact Analysis of Air Pollutants on the Air Quality Index in Jinan Winter. In 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Vol. 1. 471–474. <https://doi.org/10.1109/CSE-EUC.2017.89>
- [3] Zhang Shuyi, Guo Bin, Dong Anlan, He Jing, Xu Ziping, and Chen Song. Xi. 2017. Cautionary tales on air-quality improvement in Beijing. Proc. R. Soc. A. 473 (2017). <https://doi.org/doi.org/10.1098/rspa.2017.0457>