Urban Tree Management for Climate Resilience in Würzburg

1 Introduction

Urban areas worldwide are increasingly facing significant environmental challenges due to climate change. Rising temperatures, deteriorating air quality, and more frequent extreme weather events pose serious threats to urban living conditions. Urban trees, often regarded as the "green lungs" of cities, play a crucial role in mitigating these adverse effects. They provide shade, reduce urban heat island effects, improve air quality, and contribute to the overall aesthetic and ecological health of urban environments.

This project, titled "Urban Tree Management for Climate Resilience in Würzburg," aims to leverage urban tree data to enhance climate resilience and improve living conditions in the city of Würzburg, Germany. By analyzing detailed data on tree species, sizes, locations, and soil moisture levels, we seek to understand how strategic urban forestry can contribute to sustainable city planning and climate adaptation strategies.

The main question guiding this project is: How can urban tree data be leveraged to enhance climate resilience and improve living conditions in Würzburg?

2 Methods

In this project, two different data source were utilized.

Data Source Name	Data Type	License	Data Quality
Baumkataster der Stadt Würzburg	CSV	Open-data license	Requires cleaning for missing values and name standardization.
Würzburger Klimabäume Bodenfeuchte	CSV	Open-data license	High-quality, requires alignment with tree data.

Table 1: Overview of Datasets

2.1 Data Source 1: Baumkataster der Stadt Würzburg

- **Data URL:** https://www.govdata.de/web/guest/suchen/-/details/baumkataster-der-stadt-wurzburg
- **Data Type:** CSV
- **Description:** This dataset contains comprehensive information on over 40,000 public trees in Würzburg. It includes attributes such as species (both common and Latin names), trunk circumference, height, crown width, and geographical coordinates. These attributes are essential for understanding the distribution, health, and ecological impact of urban trees.
- **Data Structure & Quality:** The dataset is structured with well-defined columns representing various attributes of the trees. However, the dataset requires cleaning to address missing values and standardize species names for consistency.

- **License:** This dataset is available under an open-data license, which permits usage with appropriate attribution. To comply with the license requirements, all usage of the dataset in our reports and publications will include proper citation of the source.

2.2 Data Source 2: Würzburger Klimabäume – Bodenfeuchte

- **Data URL:** https://www.govdata.de/web/guest/suchen/-/details/wurzburger-klimabaeume-bodenfeuchte
- **Data Type:** CSV
- **Description:** This dataset includes sensor data from selected trees planted in various soil types across Würzburg. It tracks soil moisture levels, which is critical for understanding tree health and the effectiveness of watering schedules under different urban conditions.
- **Data Structure & Quality:** The dataset is presented in a CSV format, with columns representing various soil moisture metrics and sensor readings. The data is generally high quality but requires alignment with the tree data for effective integration and analysis.
- **License:** Similar to the Baumkataster dataset, this data is also under an open-data license. Proper attribution will be ensured in all project outputs, adhering to the license obligations.

By combining these datasets, we aim to create a comprehensive view of Würzburg's urban forestry and its impact on climate resilience.

3 Data Pipeline

The project's backbone is an automated data pipeline designed to extract, transform, and load the Baumkataster and Klimabaeume datasets.

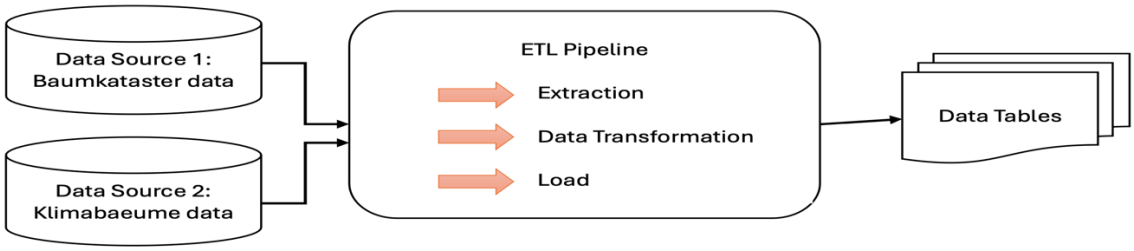

Figure 1: Pipeline Structure

Technological Framework: The technological framework utilized is built upon the Python programming language. Pandas library is employed for data manipulation and analysis, while Requests library facilitates data downloading. This technological infrastructure proves highly effective in maintaining data integrity and streamlining operations.

Data Extraction and Cleaning: The process of data extraction and cleaning involves steps such as downloading data from the internet and removing extraneous whitespace. Particularly crucial measures have been taken to address missing values and ensure the consistency of datasets. This process is pivotal in maintaining data integrity and accuracy.

Data Transformation: In the data transformation stage, column names have been renamed for clarity. Additionally, operations such as standardizing data formats and integrating different datasets have been conducted. These steps have contributed to making the data more meaningful and usable.

Data Loading: The cleaned datasets have been loaded into a SQLite database. This operation ensures that the data is stored more permanently and made ready for analysis later on. This stage enhances data integrity and accessibility, thereby contributing to the data analysis process.

Transformation and Cleaning Steps

1. **Download Data:** Use the requests library to download CSV files.
2. **Load Data:** Read CSV files into Pandas dataframes.
3. **Strip Whitespace:** Remove leading and trailing whitespaces from column names.
4. **Rename Columns:** For readability and consistency.
5. **Handle Missing Values:**
    - Drop rows with essential missing data.
    - Impute missing numerical values with medians.
    - Impute missing categorical values with modes.
6. **Drop Irrelevant Columns:** Remove unnecessary columns.
7. **Convert Data Types:** Ensure appropriate data types and handle invalid data.

This set of features makes the pipeline robust and adaptable, capable of addressing various data quality issues and changes in input data structure. As a result, consistently prepared datasets have been successfully loaded into the database. We can now provide reliable insights into urban tree management and climate resilience in Würzburg.

4 Limitations

Key limitations of the data pipeline include:

- **Data Completeness:** Potential information loss from removed rows due to missing data.
- **Imputation Bias:** Risk of bias from median and mode imputation methods.
- **Geospatial Accuracy:** Critical reliance on precise geographical data for accurate dataset merging.

5 Conclusion

The processed data is primed for analysis to explore the effective management of urban trees in enhancing Würzburg's climate resilience. Despite facing certain limitations like data completeness and potential imputation bias, the insights derived will be instrumental in informing sustainable urban planning efforts. Future enhancements will focus on refining data integration techniques and updating methodologies to reflect new data and urban dynamics.