

1 Questions

- Tracking what types of plants get planted where over the course of each planting season?
- Is there any correlation between geographical effects and plantation of various trees?

2 Data Sources

2.1 Description of Data Sources

- **Dataset 1: City of Neuss: tree plantings 2023**

Data on tree plantings in 2023. Numerous different tree species were selected for autumn tree planting - including oaks, walnut and cherry trees, various types of linden trees and sweetgum trees, which are particularly resistant to heat and drought. Because of these properties, the sweetgum tree is just as considered a “climate tree” as are hackberry trees, whitebeams or locust trees, all of which are also among the selected species

- **Dataset 2: World Monthly Atmospheric CO₂ Concentrations**

One of the major challenges of the program in past years has been accurately tracking what types of plants get planted where over the course of each planting season. In this project, I took data from many different sources and compiled them into a usable database format for analysis. Two different ways of cleaning data are reflected in the associated Jupyter Notebook.

2.2 Data Structure and Quality

- **Annual Tree Planting** The image shows a dataset of tree planting events, including details such as date, location, common and Latin names of plants, and quantities. The data is summarized with percentages for location and species distribution. For example, Lymburn and Twin Brooks had 14% and 13% of the planting events, respectively, while Saskatoon Berry and Trembling Aspen were common species planted, each making up 9% and 8% of the total. The highest quantity planted for a single species is 310.

| Date | Location | Common Name | Latin Name | Quantity |
|------------------------|-------------------------------|-----------------------------|-------------------------|--------------------------|
| Date of planting event | Location of planting event | Non-botanical name of plant | Botanical name of plant | Count of species planted |
| | Lymburn | 14% | Saskatoon Berry | 9% |
| | Twin Brooks | 13% | Trembling Aspen | 8% |
| | Other (917) | 72% | Other (1049) | 83% |
| 5/5/2023 | Sir Wilfrid Laurier Park 2023 | Saskatoon Berry | Amelanchier alnifolia | 58 |
| 5/5/2023 | Sir Wilfrid Laurier Park 2023 | Trembling Aspen | Populus tremuloides | 36 |
| 5/5/2023 | Sir Wilfrid Laurier Park 2023 | Pin Cherry | Prunus pensylvanica | 15 |
| 5/5/2023 | Sir Wilfrid Laurier Park 2023 | Red Osier Dogwood | Cornus sericea | 24 |
| 5/10/2023 | Lymburn | Saskatoon Berry | Amelanchier alnifolia | 24 |

- **Locations and Names** This dataset show that the types of tress that has been planted at different loactions.

| | M | N | O | P | Q | R | S |
|----|-----------|--------------------|-----------------|-------------------|----------|---|---|
| 1 | Date | Location | Common Name | Latin Name | Quantity | Species | |
| 2 | 5/5/2023 | Sir Wilfrid Laurie | Saskatoon Berry | Amelanchier alni | 58 | Saskatoon Berry - Amelanchier alnifolia | |
| 3 | 5/5/2023 | Sir Wilfrid Laurie | Trembling Asper | Populus tremulo | 36 | Trembling Aspen - Populus tremuloides | |
| 4 | 5/5/2023 | Sir Wilfrid Laurie | Pin Cherry | Prunus pensylva | 15 | Pin Cherry - Prunus pensylvanica | |
| 5 | 5/5/2023 | Sir Wilfrid Laurie | Red Osier Dogw | Cornus sericea | 24 | Red Osier Dogwood - Cornus sericea | |
| 6 | 5/10/2023 | Lymburn | Saskatoon Berry | Amelanchier alni | 24 | Saskatoon Berry - Amelanchier alnifolia | |
| 7 | 5/10/2023 | Lymburn | Lodgepole Pine | Pinus contorta le | 43 | Lodgepole Pine - Pinus contorta latifolia | |
| 8 | 5/18/2023 | Lymburn | Lodgepole Pine | Pinus contorta le | 46 | Lodgepole Pine - Pinus contorta latifolia | |
| 9 | 5/18/2023 | Lymburn | Prickly Rose | Rosa acicularis | 30 | Prickly Rose - Rosa acicularis | |
| 10 | 5/18/2023 | Lymburn | Saskatoon Berry | Amelanchier alni | 23 | Saskatoon Berry - Amelanchier alnifolia | |
| 11 | 5/18/2023 | Lymburn | Western Larch | Larix occidentali | 30 | Western Larch - Larix occidentalis | |
| 12 | 5/18/2023 | Lymburn | White Spruce | Picea glauca | 24 | White Spruce - Picea glauca | |
| 13 | 5/18/2023 | Lymburn | Lodgepole Pine | Pinus contorta le | 40 | Lodgepole Pine - Pinus contorta latifolia | |
| 14 | 5/18/2023 | Lymburn | Prickly Rose | Rosa acicularis | 24 | Prickly Rose - Rosa acicularis | |
| 15 | 5/18/2023 | Lymburn | Snowberry | Symphoricarpos | 21 | Snowberry - Symphoricarpos albus | |
| 16 | 5/23/2023 | Allard | White Spruce | Picea glauca | 72 | White Spruce - Picea glauca | |
| 17 | 5/23/2023 | Allard | Balsam Poplar | Populus balsami | 30 | Balsam Poplar - Populus balsamifera | |
| 18 | 5/23/2023 | Allard | Sandbar Willow | Salix exigua | 6 | Sandbar Willow - Salix exigua | |
| 19 | 5/23/2023 | Allard | Sandbar Willow | Salix exigua | 30 | Sandbar Willow - Salix exigua | |
| 20 | 5/23/2023 | Allard | Pin Cherry | Prunus pensylva | 30 | Pin Cherry - Prunus pensylvanica | |
| 21 | 5/21/2023 | Lymburn | Jack Pine | Pinus banksiana | 15 | Jack Pine - Pinus banksiana | |
| 22 | 5/21/2023 | Lymburn | Lodgepole Pine | Pinus contorta le | 28 | Lodgepole Pine - Pinus contorta latifolia | |
| 23 | 5/21/2023 | Lymburn | Buffaloberry | Shepherdia can | 34 | Buffaloberry - Shepherdia canadensis | |
| 24 | 5/21/2023 | Lymburn | Jack Pine | Pinus banksiana | 4 | Jack Pine - Pinus banksiana | |

2.3 Licenses and Permissions

The data sources are publicly available on IMF under open-data licenses. Detailed license information can be found at: [License](#)

3 Data Pipeline

The data pipeline has three main modules: extractor, transform, and loader. Each of the modules has their respective functions. First extract csv from extractor module is used to extract the data source from URL, then delete columns from transform module deletes the list of useless columns specified for every dataset, then a flag of "date column" is present in configs which only applies standardize date column to standardize the date format across necessary datasets, then renaming of the date columns is done using rename _year columns function, it is only triggered for those datasets where flag of "rename year columns" is equals to true, once all the transformations have been applied, dataset is then loaded to sqlite database using load df to sqlite from loader module.



Figure5:ETLPipelineDiagram

4 Result and Limitations

Output datasets of the pipeline for all data sources are stored in sqlite database as tables as it was faster and easier to handle as a collective database, The pipeline is coded in a way that data quality dimensions were of the utmost priority and that the output datasets of the pipeline

- reflect the real word and are correct indicators

- contain all necessary information which is required to answer selected questions
- are consistent in their formats
- time period of datasets are appropriate and intersecting
- presentation of the datasets aligns with the requirements of the questions need to be answered

The two datasets aim to tackle environmental challenges from different angles. The first dataset addresses the broad issue of climate change by analyzing human-generated greenhouse gas emissions and exploring solutions to reduce these emissions globally. The second dataset focuses on enhancing local environmental management by improving the tracking and analysis of tree planting activities, which helps in carbon sequestration and urban greening. Together, they contribute to understanding and mitigating climate change impacts at both global and local scales.

References

- [1] Open Government. <https://www.govdata.de/ckan/dataset/stadt-neuss-baumplantungen-2023.rdf>
- [2] Kaggle. <https://www.kaggle.com/datasets/chadmottershead/clean-planting-data-2023/data>