Analysis of CO2 Emissions and the Impact of Rainfall in Konstanz City

Introduction

This study aims to investigate the relationship between CO2 emissions and rainfall patterns in Konstanz City. By analyzing historical data on CO2 emissions and rainfall over a specific time period, the project seeks to elucidate how variations in rainfall affect CO2 emissions and vice versa. Understanding this relationship is crucial for assessing the environmental dynamics of the city, identifying potential drivers of CO2 emissions, and informing sustainable urban planning and climate change mitigation strategies.

Question

How do CO2 emissions impact rainfall patterns in Konstanz City?

Data Sources

1. CO2 Emission (2010-2017) Data

- URL: https://offenedaten-konstanz.de/sites/default/files/Quellenbezogene_CO2_Emissionen_2010-2017_nach_Sektoren_0.csv
- Description: This dataset contains CO2 emissions data by sectors for the years 2010 to 2017.
- Structure and Quality: The dataset is structured with columns representing different sectors and rows for each year. It has been verified to have consistent and clean data for the specified years.
- Data Structure: Tabular format with columns for date ranges, station ID, and various weather metrics.

2.Rainfall (RA_RR) Dataset

- URLhttps://offenedaten-konstanz.de/node/40911/download
- Description: This dataset contains additional data required for the project.
- Structure and Quality: This dataset complements the primary dataset by providing necessary additional information.
- Data Structure: Tabular format with columns for date ranges, station ID, and various weather metrics.

Reasons for Choosing These Data Sources

- Relevance: Both datasets are from the Konstanz region, making them highly relevant for regional environmental and climatic analysis.
- Coverage Period: The data spans a similar time frame (2010-2017), allowing for temporal analysis and correlation studies.
- Open Data: Both datasets are publicly available and provided by a reputable source, ensuring transparency and accessibility.

Analysis of CO2 Emissions and the Impact of Rainfall in Konstanz City

Licenses and Permissions

Both datasets are under open-data licenses, allowing free usage with proper attribution.

Obligations:

To comply with this license, the report and any publications resulting from this study will include proper attribution to Open Data Konstanz, clearly indicating any modifications made to the original data.

Data Pipeline

The data pipeline is implemented using Python, leveraging the panda's library for data manipulation and SQLite for data storage. The pipeline consists of the following steps:

Load Datasets: Load the CO2 emissions and rainfall data from the provided URLs.

Preprocess & Cleaning Data:

Renamed columns for better clarity in "**gs**" data frame. Grouped data by the **'Jahr'** column and selected the first entry for each year.

Selected specific rows (64 to 72) to focus on the relevant data in "**co2**" data frame. Dropped unnecessary columns and renamed columns for clarity. Ensured the **'Jahr'** column is properly formatted. In this dataset **JA_RR** is rainfall data.

Merge Data frames: Merged the preprocessed **gs** and **co2** Data Frames on the 'Jahr' column to create a consolidated dataset.

Store Data: Store the merged data frame in an SQLite database.

Export Data: Export the merged data frame to a CSV file.

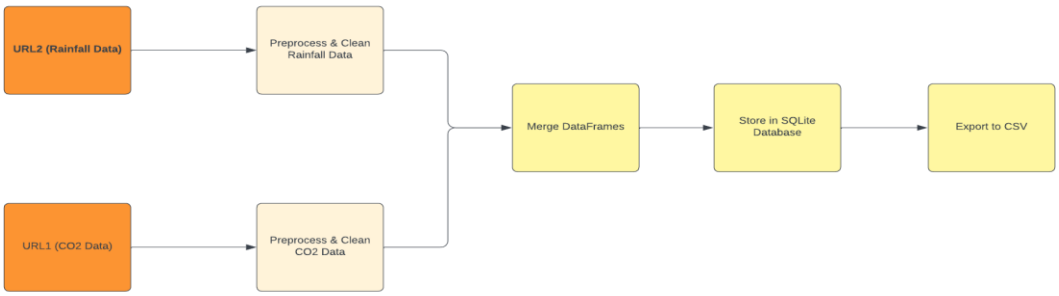

Fig: Automated Data Pipeline

Problems encountered and Solved:

Data Cleaning: The CO2 emissions and rainfall datasets had inconsistencies and missing values, requiring careful cleaning. This was addressed by implementing data cleaning functions to handle missing values, remove duplicates, and standardize data formats.

Analysis of CO2 Emissions and the Impact of Rainfall in Konstanz City

Data Merging: Merging the datasets based on the 'Jahr' column posed challenges due to differences in the data structures and formats. To address this, data preprocessing steps were applied to ensure uniformity before merging.

My pipeline tackles errors and changing input data by:

1. **Data Validation**: Ensuring data meets expected standards.
2. **Error Handling**: Gracefully managing errors with logging and recovery mechanisms.
3. **Dynamic Adaptation**: Adjusting parameters or models to accommodate shifts in data.
4. **Version Control**: Tracking changes and enabling rollbacks for seamless troubleshooting.
5. **Feedback Mechanisms**: Monitoring performance and adjusting processing logic for continuous improvement.

Data Quality:

- Accuracy - data reflects the real word and is correct ○ e.g., the co2 and rainfall data exist in real world.
- Completeness -contains all necessary information
- Consistency - is consistent in its format ○ e.g., the same data format at all relevant places
- Timeliness- emission of the data is appropriate
- Relevancy - presentation aligns with the needs of the rainfall data.

Data Storage:

The cleaned and merged data is stored in an SQLite file (merged_dataset.db) for efficient querying and analysis. The dataset is also saved as a CSV file (merged_dataset.csv).

	Jahr	ID	CO2Quelle	Tonnen	Einwohner	Inhabitants	CO2AnteilproKopf	...	JA_SD_S	JA_MX_FX	JA_MX_TX	JA_MX_TN	QN_6	JA_RR	JA_MX_RS
0	2010	1	Kraftwerke	26732	84693	241822	0.315634113799251	...	1627.6	24.3	35.3	-9.4	9	888.3	38.4
1	2011	5	Kraftwerke	28484	82031	234674	0.34723458204825	...	1948.6	26.2	34.5	-7.6	9	801.4	38.5
2	2012	9	Kraftwerke	21193	79645	218862	0.266093288969803	...	1900.0	-999.0	34.1	-13.6	9	958.3	70.7
3	2013	13	Kraftwerke	22951	81141	233423	0.28285330474113	...	1602.0	23.4	36.7	-7.3	9	872.3	47.8
4	2014	17	Kraftwerke	21902	81692	211236	0.26810458796455	...	1707.7	24.8	34.9	-10.5	9	792.7	30.6

Finally, the output of my data pipeline comprises processed data sets, insights, or predictions. I ensure data quality through validation mechanisms. I choose CSV or JSON formats for interoperability. Potential issues include data bias, drift, model assumptions, and ethical considerations. In the final report, I'll transparently address limitations and emphasize ongoing validation and monitoring.