



1. Question

How do total vehicle kilometres of goods vehicles influence the air emission in Dutch?

2. Data Source

2.1 Data Source 1

Data URL: [Emissions to air on Dutch territory; road traffic](#)

Data Type: CSV

Description: Data on air emissions in the Netherlands from road traffic, including foreign vehicles, and average emissions per vehicle kilometre, covering various sources of emissions (2017-2022).

License: [CC BY 4.0](#)

2.2 Data Source 2

Data URL: [Vehicle kilometres goods vehicles; kilometres, vehicle weight 2001-2020](#)

Data Type: CSV

Description: Data on total and average annual kilometres of goods vehicles in the Netherlands, segmented by vehicle origin, territory, type, construction year, and weight (2015-2020).

License: [CC BY 4.0](#)

3. Data Quality and Structure

3.1 Emission Dataset

Dimensions	Description
Accuracy	Reflects real world data and is correct
Completeness	Missing values in many columns
Consistency	Data formatting is inconsistent
Timeliness	Updated on 8 February 2024
Relevancy	Contains updated information until Feb 2024

3.2 Vehicle Dataset

Dimensions	Description
Accuracy	Reflects real world data and is correct
Completeness	Missing values in many columns
Consistency	Data formatting is inconsistent
Timeliness	Updated on 10 November 2021
Relevancy	Contains updated information until Nov 2021

4. Data Pipeline

4.1 Technologies Used

- Python for data cleaning, preprocessing, and analysis
- Packages: Pandas and matplotlib

4.2 Steps

1. **Download CSV from URL:** Using Selenium WebDriver with Chrome.
2. **Read CSV:** Load into Pandas DataFrame.
3. **Load to Database:** Write DataFrame to SQLite database table.

4.3 Functions

- **download_csv_from_url(url):** Downloads CSV files.
- **read_csv():** Reads CSV into DataFrame.
- **load_to_database(df, table_name):** Writes DataFrame to database table.

Execution: Call `run_pipeline()` to execute the process.

4.4 Problems and Solutions

Problems	Solutions
Defining table header	Manually rename header and drop unnecessary values

5. Results and Limitations

5.1 Results

Dimensions	Description
Accuracy	Reflects real world data and is correct
Completeness	No missing values
Consistency	Data formatting is consistent
Timeliness	Updated on 10 November 2021
Relevancy	Contains updated info until Feb 2024

Dimensions	Description
Accuracy	Reflects real world data and is correct
Completeness	No missing values
Consistency	Data formatting is consistent
Timeliness	Updated on 10 November 2021
Relevancy	Contains updated info until Nov 2021

5.2 Limitations

- **Correctness:** Aggregation method not specified.
- **Completeness:** Dropped unit of measurement column.
- **Representativeness:** Vehicle data limited to the Netherlands until 2021.