

Investigating the Impact of Weather Conditions on Traffic Flow in Dublin

This study seeks to investigate the relationship between weather and traffic flow in Dublin. I want to see if there is a relation between daily traffic data and weather variables such as temperature, rainfall, humidity, and wind speed. Understanding this correlation can help city planners and traffic managers improve traffic flow and reduce congestion.

1 Question

The primary question for this project is: "How do daily weather conditions affect traffic flow in Dublin? Is there a significant correlation between weather parameters and traffic patterns?"

2 Data Sources

2.1 Traffic Flow Data

- **Metadata URL:** https://data.gov.ie/dataset/traffic-flow-data-jan-to-june-2021-sdcc/resource/62f76e7d-a289-4675-b208-cf9078730b0b?inner_span=True
- **Data URL:** <https://data-sdublincoco.opendata.arcgis.com/datasets/dublincoco::traffic-flow-data-jan-to-june-2021-sdcc.csv>
- **Data Type:** CSV (Structured and Tabular Dataset)
- **Description:** This file contains South Dublin County Council (SDCC) Traffic Congestion Saturation Flow Data for January to June 2021.
- **License:** Creative Commons Attribution 4.0
Canonical URL: <https://creativecommons.org/licenses/by/4.0/>.

2.2 Weather Data Provided by Visual Crossing

- **Metadata URL:** <https://www.visualcrossing.com/resources/documentation/weather-data/weather-data-documentation/>
- **Data URL:** <https://weather.visualcrossing.com/VisualCrossingWebServices/rest/services/retrievebulkdataset?&key=JFUWSDF8WDRYFZZFBZ3W9M8FE&taskId=82943c3c20942672294f3f2292ede040&zip=false>
- **Data Type:** CSV (Structured and Tabular Dataset)

- **Description:** This file contains temperature, humidity, and other weather-related data of the city Dublin for January to June 2021.
- **License:** Visual Crossing Corporation. Visual Crossing Weather. 2021, <https://www.visualcrossing.com/>. Accessed May 20, 2024.

2.3 Compliance with Data Source Licenses

The traffic flow data is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which allows for sharing and adaptation of the data, provided appropriate credit is given. The weather data from Visual Crossing is provided under their own usage terms, which typically involve acknowledging the source and not using the data for commercial purposes without permission. The data will be used strictly for educational and research purposes. Also, all reports, visualizations, and publications that use these datasets will include proper attribution. No modifications or transformations will be applied to the data.

3 Data Pipeline

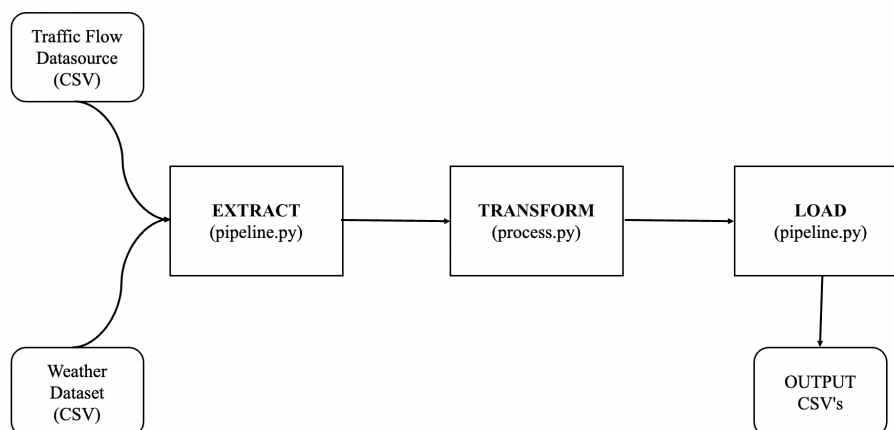


Figure 1: Data Pipeline

3.1 Running the Pipeline

The pipeline can be run using the bash script `pipeline.sh` present in the project directory. It executes `pipeline.py`, which in turn executes `process.py` for the data transformation steps.

3.1.1 Extract

- **get_url_data():** Fetches the data from the given URL and loads it into a DataFrame. Technically, this method sends a GET request to the provided URL, reads the response into a DataFrame. Also, handles any errors.

3.1.2 Data Preparation

The functions in `process.py` handles dataset-specific cleaning:

- **process_traffic_flow_data():** Process traffic flow data by converting dates, dropping unnecessary columns, aggregating data by date, and rounding flow values. Initially, the dataset contained hourly values of traffic flow for every day of the month. Then had to aggregate the data by date, and round the flow values. This method ensures the data is in the correct format and aggregated appropriately for analysis. The columns considered for further analysis are *date* and *flow*.
- **process_weather_data():** Process weather data by removing unnecessary columns. This method simplifies the dataset by keeping only relevant columns for analysis. The columns considered for further analysis are *datetime*, *temp*, *humidity*, *precip*, *windspeed*, *visibility*, and *icon*.

3.1.3 Saving Dataframes

`save_df_to_csv()`: The final step saves the cleaned dataframes as separate CSV files for data analysis, interpretation, and visualization.

4 Results and Limitations

4.1 Preferred Data Format for Pipeline Output

The pipeline outputs are saved in CSV format because it is simple, compatible, and human-readable, making it ideal for managing structured datasets. The columns selected from the datasets are chosen to facilitate the correlation analysis between traffic flow and weather conditions. The *date* column from the traffic flow dataset and the *datetime* column from the weather dataset are used to align the data temporally. This alignment allows for the examination of potential relationships between traffic flow (using the *flow* column) and various weather parameters (such as *temp*, *humidity*, *precip*, *windspeed*, *visibility*, and *icon*).

4.2 Assessment of Data Quality and Potential Challenges

- **Traffic Flow Data:** The dataset is of good quality, however it may not capture all variations in traffic patterns because of unexpected events or anomalies. Unexpected events, such as road construction, accidents, or public events, might produce data anomalies that differ from normal traffic circumstances. Also, the dataset may not cover the entire city of Dublin, resulting in gaps in traffic statistics for specific regions.
- **Weather Data:** The data measured hourly and averaged over longer periods such as days, weeks, or months, may not fully capture short-term variations or anomalies, potentially affecting the accuracy of weather analysis. It may not cover all geographical areas within Dublin or may have gaps in temporal coverage, potentially limiting the scope and comprehensiveness of the analysis.