

Exploring the Impact of Weather and Climate on Bicycle Traffic in Köln

Presented by:

Sujit Debnath

MSc Student in Artificial Intelligence

Data Engineering, SS'23

Outline

A roadmap to guide our presentation journey



01 Project Introduction

02 Data Source Details

03 Project Structure: ETL Pipeline

04 Data Exploration

05 Modifying and Combining Two Data Sources

06 Exploratory Data Analysis (EDA)

07 Questions and Analysis

08 Conclusion

This project aims to analyze the weather and climate conditions of Köln and its bicycle traffic generated from several automatic counting stations throughout the city to determine if Köln is a suitable city for an enthusiastic cyclist to live in.

Objectives:

- Explore the variation in temperature throughout the years in Köln.
- Analyse the trends and patterns in bicycle traffic across different seasons and years in Köln city.
- Investigate the relationship between temperature and bicycle traffic to understand the impact of temperature on bicycle traffic.
- Examine the potential connection between monthly precipitation and average sea-level air pressure with total bicycle traffic in Köln.
- Assess the influence of wind speed and sunshine duration on the total bicycle traffic in Köln.

Overall, the main goal is to focus on the following question,

“How do the weather conditions in Köln impact bicycle traffic throughout the year?”

Datasource1: Bicycle Traffic Data in Köln

- Source: [mobilithek](#)
- Metadata URL: <https://mobilithek.info/offers/-2236641647976866002>
- Sample Data URL: https://offenedaten-koeln.de/sites/default/files/Fahrrad_Zaehlstellen_Koeln_2016.csv
- Data Type: CSV

This data source contains Köln's bicycle traffic generated from several automatic counting stations throughout the city from 2009.

Datasource2: Weather and Climate Data of Köln

- Source: [meteostat](#)
- Metadata URL: <https://dev.meteostat.net/bulk/monthly.html>
- Sample Data URL: <https://bulk.meteostat.net/v2/monthly/{station}.csv.gz>,
Station-id of Köln = '10513', 'D2968'
- Data Type: CSV

This data source will provide weather and climate data in Köln, including average air temperature, daily minimum and maximum air temperature, monthly precipitation total, maximum snow depth, average wind direction and speed, peak wind gust, average sea-level air pressure, and monthly sunshine total.

Project Structure: ETL Pipeline

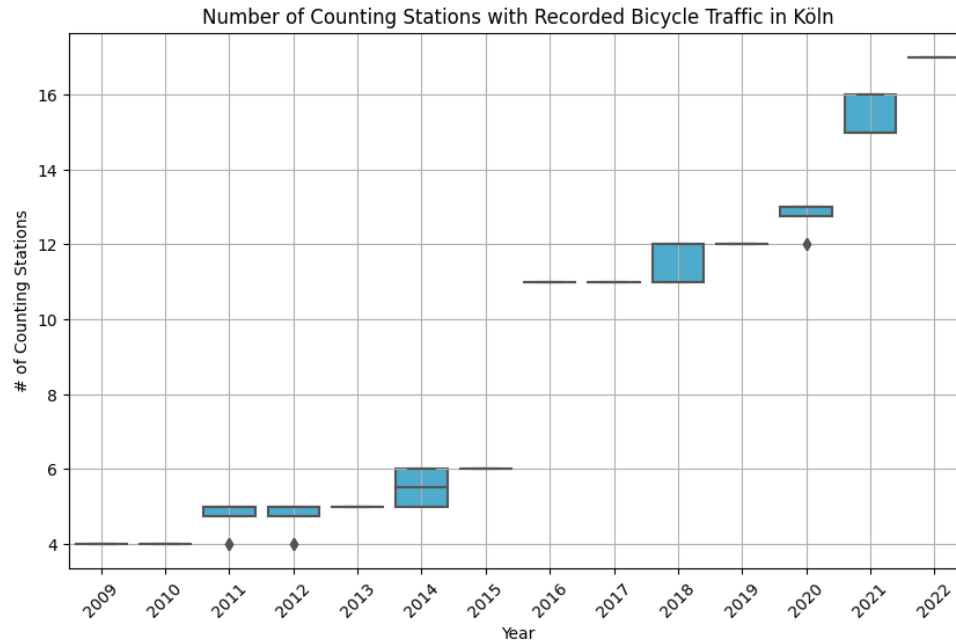
Explanation of the ETL process and overall project structure

```
project/
├── config/                                # Configuration files and settings
│   ├── __init__.py
│   ├── config_var.py                     # Configuration variables
│   └── source_info.json                  # Source information
├── data/                                 # Data directory
│   ├── processed/                       # Processed data
│   └── raw/                             # Raw data
├── etl/                                 # ETL (Extract, Transform, Load) pipeline modules
│   ├── __init__.py
│   ├── extract/                         # Extraction module
│   │   ├── __init__.py
│   │   └── data_extractor.py           # Data extraction logic
│   ├── transform/                      # Transformation module
│   │   ├── __init__.py
│   │   └── data_transformer.py        # Data transformation logic
│   └── load/                           # Loading module
│       ├── __init__.py
│       └── data_loader.py             # Data loading logic
├── pipelines/                          # Data pipeline modules
│   ├── __init__.py
│   └── data_pipeline.py               # ETL data pipeline implementation
├── utils/                              # Utility modules
│   ├── __init__.py
│   └── service_factory.py             # Service factory utility
├── tests/                              # Test modules
│   ├── __init__.py
│   ├── test_component.py               # Test cases for component testing
│   ├── test_pipeline.py               # Test cases for system testing
│   └── transformed_data.pkl           # Original transformed data for testing purposes
├── main.py                             # Entry point of the project
├── tests.sh                            # Bash script for running all the test cases
├── exploration.ipynb                   # Notebook for data exploration
├── report.ipynb                        # Notebook for final project report
└── project-plan.md                     # Project plan and documentation
```

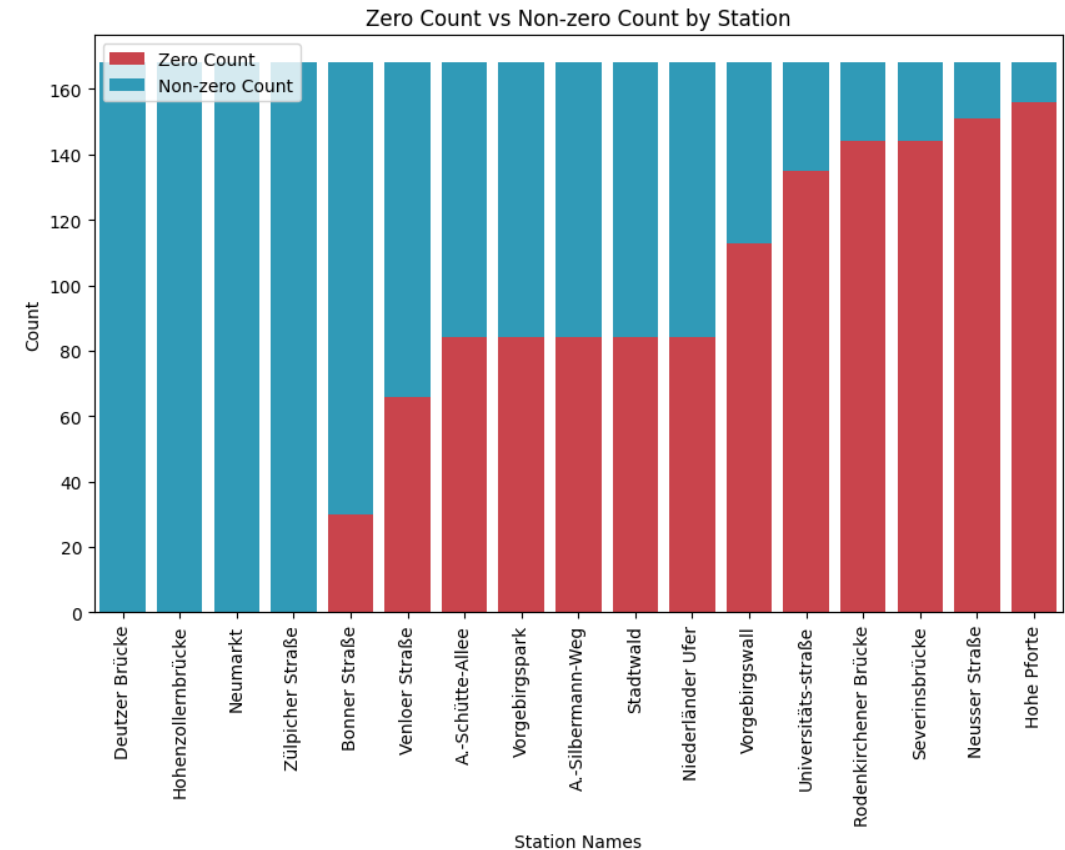
- The project follows a structured ETL pipeline approach, encompassing various directories and modules with specific functionalities.
- The main.py serves as the entry point for running the pipeline using the command `pdm run main.py`, resulting in the generation of the final dataset stored in an SQLite database.

Data Exploration

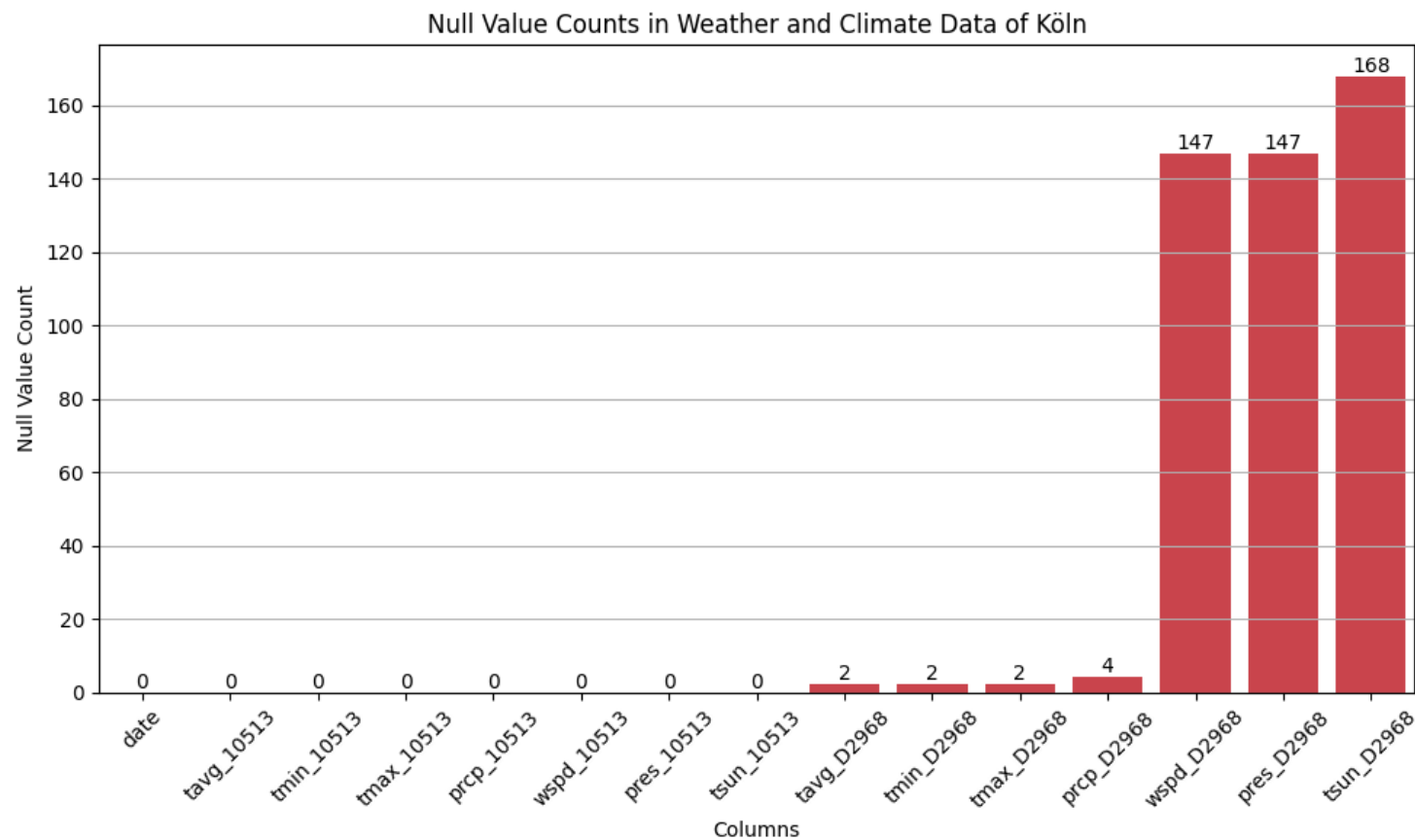
Overview of the initial exploration and analysis of the data



Number of stations that started counting Bicycle Traffic from the beginning varies a lot



Only 4 stations (Deutzer Brücke, Hohenzollernbrücke, Neumarkt, Zülpicher Straße) has started counting Bicycle Traffic from 2009



wspd, pres, and tsun of station Köln-Stammheim (D2968) has high number of null count

Modifying and Combining Two Data Sources

Description of how two data sources were modified and combined for analysis

Data columns (total 18 columns):

#	Column	Non-Null Count	Dtype
0	Date	168 non-null	object
1	Deutzer Brücke	168 non-null	int64
2	Hohenzollernbrücke	168 non-null	int64
3	Neumarkt	168 non-null	int64
4	Zülpicher Straße	168 non-null	int64
5	Bonner Straße	168 non-null	int64
6	Venloer Straße	168 non-null	int64
7	A.-Schütte-Allee	168 non-null	int64
8	Vorgebirgspark	168 non-null	int64
9	A.-Silbermann-Weg	168 non-null	int64
10	Stadtwald	168 non-null	int64
11	Niederländer Ufer	168 non-null	int64
12	Vorgebirgswall	168 non-null	int64
13	Universitäts-straße	168 non-null	int64
14	Rodenkirchener Brücke	168 non-null	int64
15	Severinsbrücke	168 non-null	int64
16	Neusser Straße	168 non-null	int64
17	Hohe Pforte	168 non-null	int64

dtypes: int64(17), object(1)

- To analyze the overall bicycle traffic in Köln, we will focus on first 4 columns and calculate the sum of their traffic values.

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	date	168 non-null	object
1	tavg_10513	168 non-null	float64
2	tmin_10513	168 non-null	float64
3	tmax_10513	168 non-null	float64
4	prcp_10513	168 non-null	float64
5	wspd_10513	168 non-null	float64
6	pres_10513	168 non-null	float64
7	tsun_10513	168 non-null	float64
8	tavg_D2968	166 non-null	float64
9	tmin_D2968	166 non-null	float64
10	tmax_D2968	166 non-null	float64
11	prcp_D2968	164 non-null	float64
12	wspd_D2968	21 non-null	float64
13	pres_D2968	21 non-null	float64
14	tsun_D2968	0 non-null	float64

dtypes: float64(14), object(1)

- Drop the columns with a high null count for D2968
- Fill in the missing values for the remaining columns
- Calculate the aggregated values for tavg, tmin, tmax, and prcp to represent the combined values for Köln
- Consider the column wspd_10513, pres_10513, and tsun_10513 as the combined 'wspd', 'pres', and 'tsun'

Modifying and Combining Two Data Sources

Description of how two data sources were modified and combined for analysis

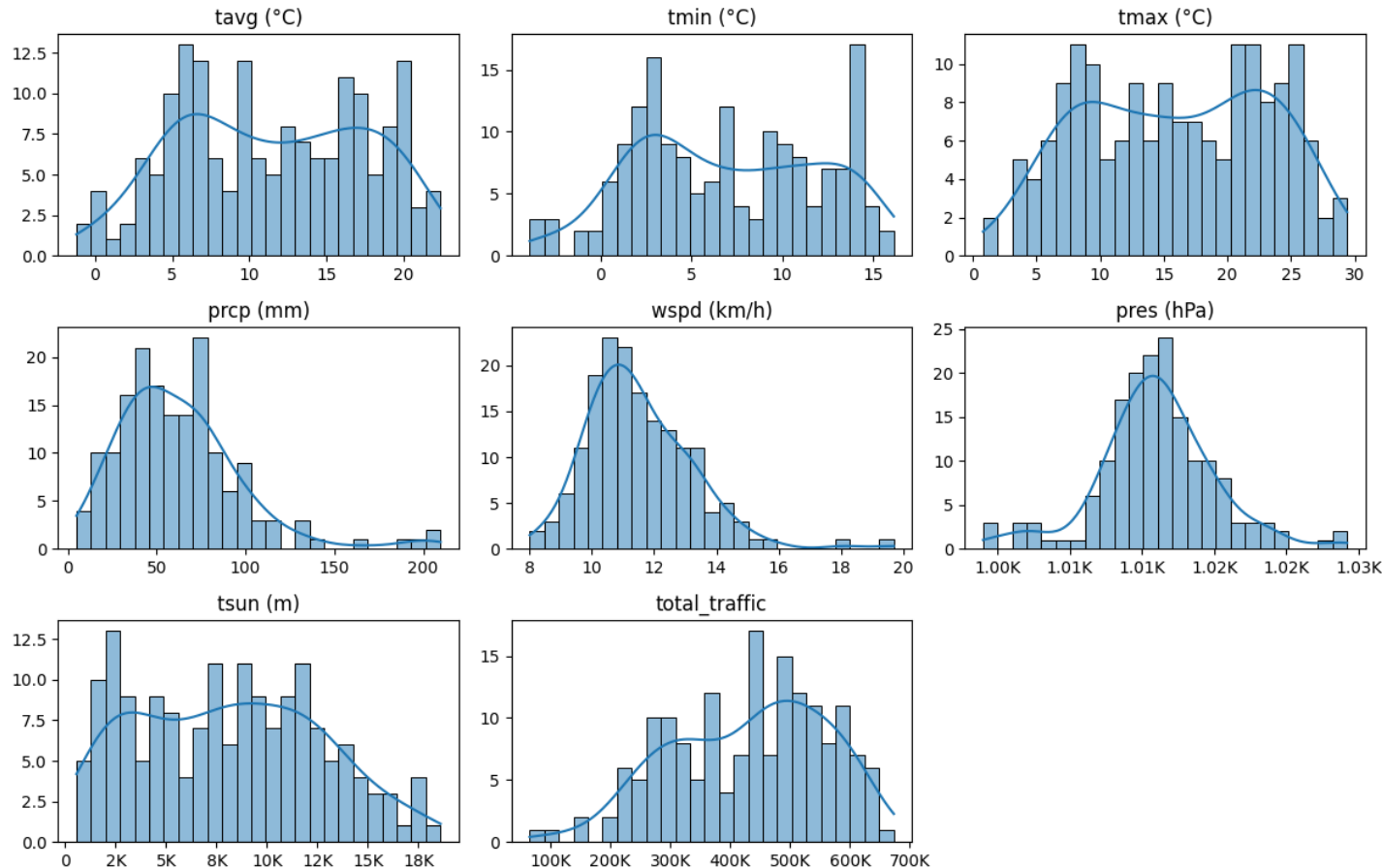
date	tavg	tmin	tmax	prcp	wspd	pres	tsun	total_traffic
January-2009	-0.05	-3.50	3.15	31.450	11.1	1015.2	5340.0	209000
February-2009	3.45	0.60	6.30	70.500	12.3	1014.5	1800.0	230657
March-2009	6.50	2.50	10.55	102.000	12.6	1014.9	5460.0	305112
April-2009	13.65	7.80	19.05	68.600	10.0	1014.0	11220.0	375060
May-2009	15.00	9.30	20.25	57.175	10.8	1019.2	11940.0	393464

- date: the date in the month-year format.
- tavg: the average air temperature in °C.
- tmin: the average daily minimum air temperature in °C.
- tmax: the average daily maximum air temperature in °C.
- prcp: the monthly precipitation total in mm.
- wspd: the average wind speed in km/h.
- pres: the average sea-level air pressure in hPa.
- tsun: the monthly sunshine total in minutes.
- total_traffic: the total bicycle traffic of the 4 stations

Exploratory Data Analysis (EDA)

Highlights of the EDA process and insights gained from analyzing the data

Histograms for each of the columns of combined dataset

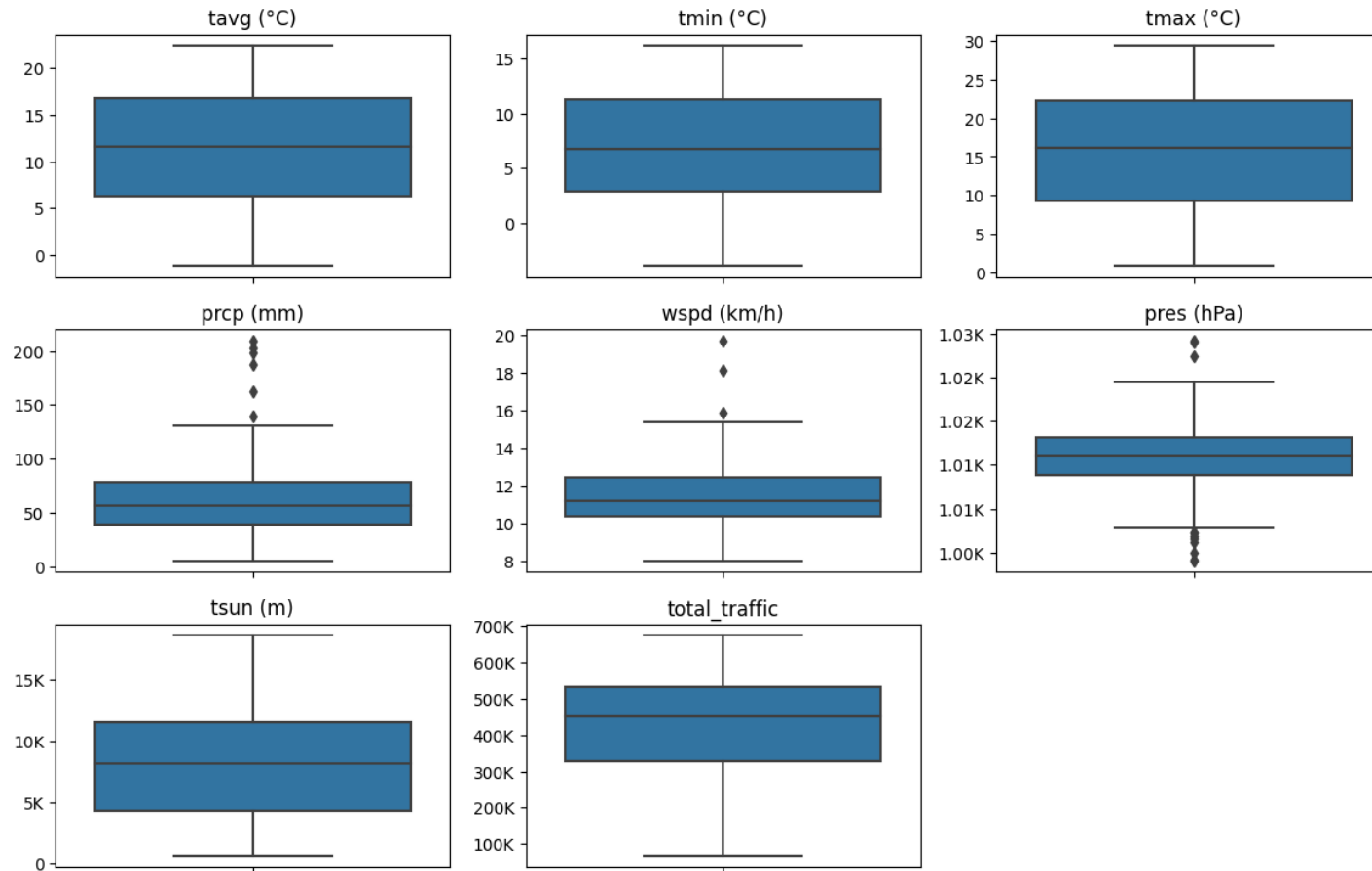


- Histograms were created for each of the columns and it shows the distribution of values for columns such as tavg, tmin, tmax, prcp, wspd, pres, tsun, and total_traffic.
- The number of bins was set to 25, and kernel density estimation (KDE) was enabled to visualize the underlying distribution.

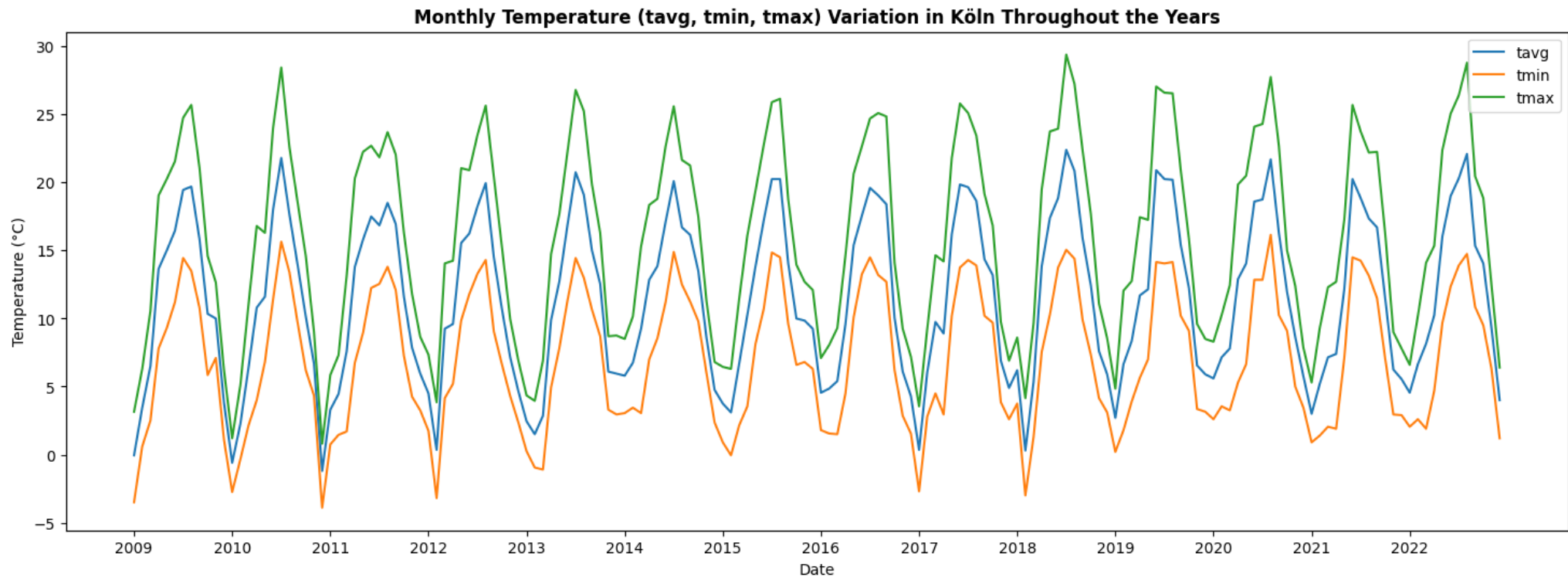
Exploratory Data Analysis (EDA)

Highlights of the EDA process and insights gained from analyzing the data

Box plots for each of the columns of combined dataset

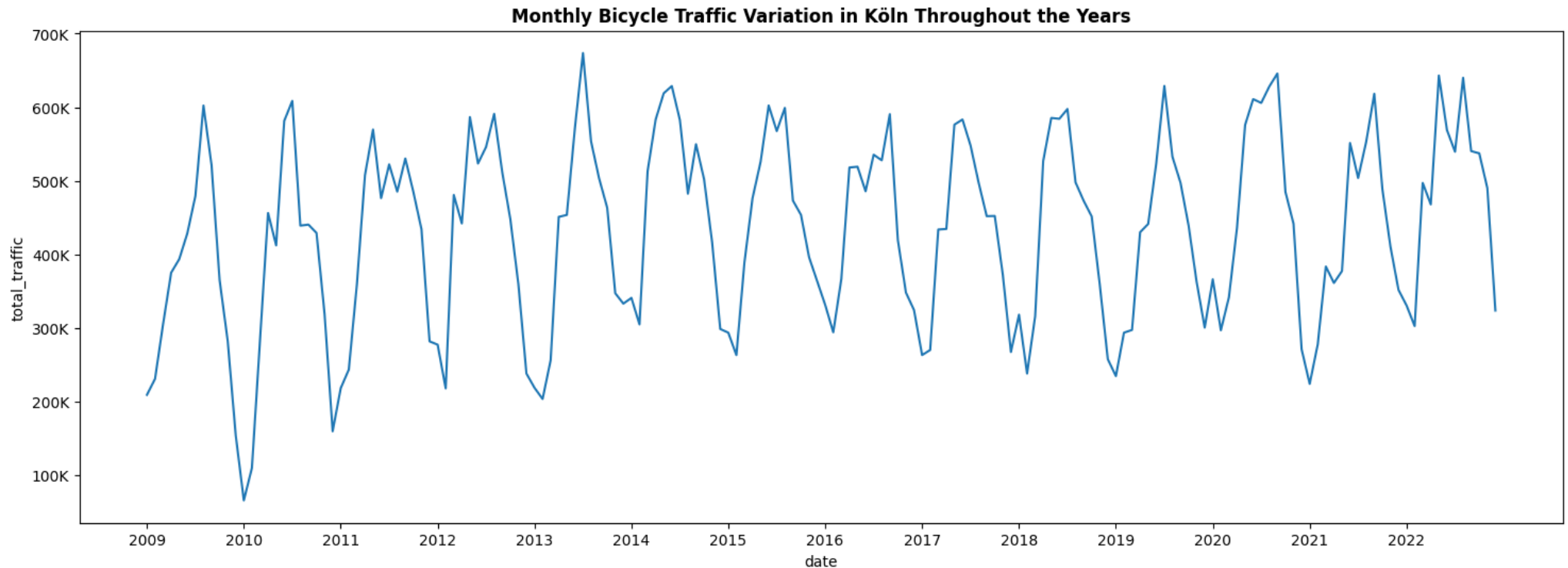


Box plots were created for each of the columns and each box plot provide information about the distribution, central tendency, and presence of outliers in the variables.



Q1. How does the temperature in Köln city vary throughout the years?

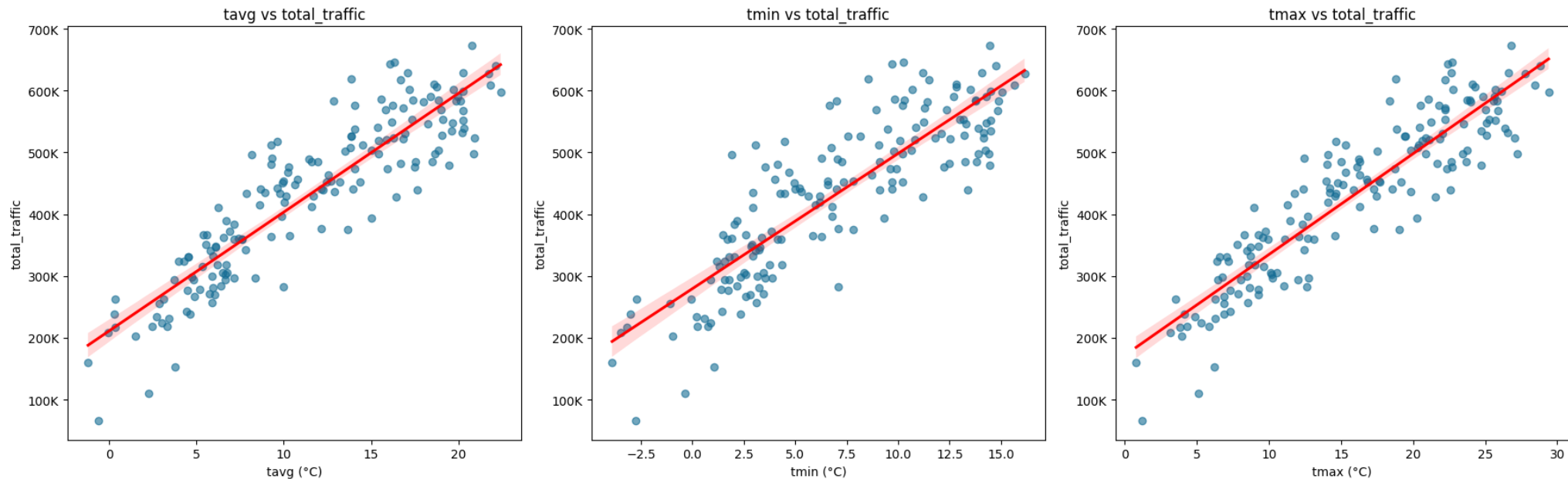
The temperature in Köln city exhibits significant variation throughout the years, following a seasonal pattern. It typically shows higher values during the summer months and lower values during the winter months.



Q2. How does the bicycle traffic in Köln city vary throughout the years?

Bicycle traffic in Köln city also displays variation throughout the years, reflecting the seasonal changes and potential influences of weather conditions.

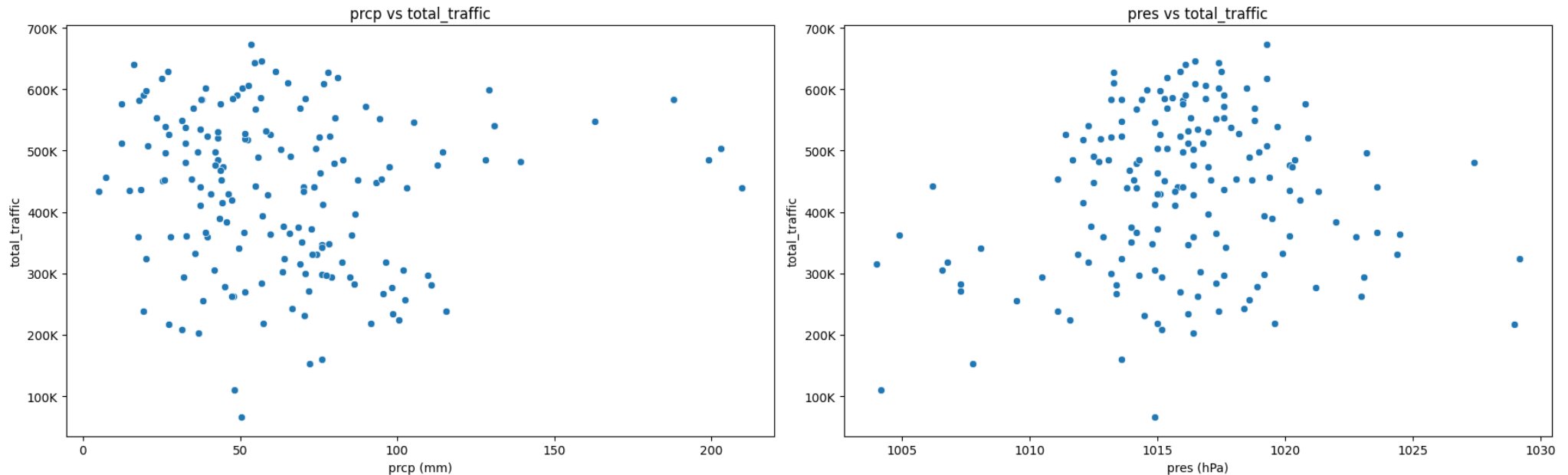
Relationship between Temperature and Total Bicycle Traffic in Köln city



Q3. Does temperature affect bicycle traffic in Köln city?

Temperature has a significant impact on bicycle traffic in Köln city. Through the analysis of temperature data (tavg, tmin, tmax) and bicycle traffic data, we observe a clear relationship between temperature and bicycle usage.

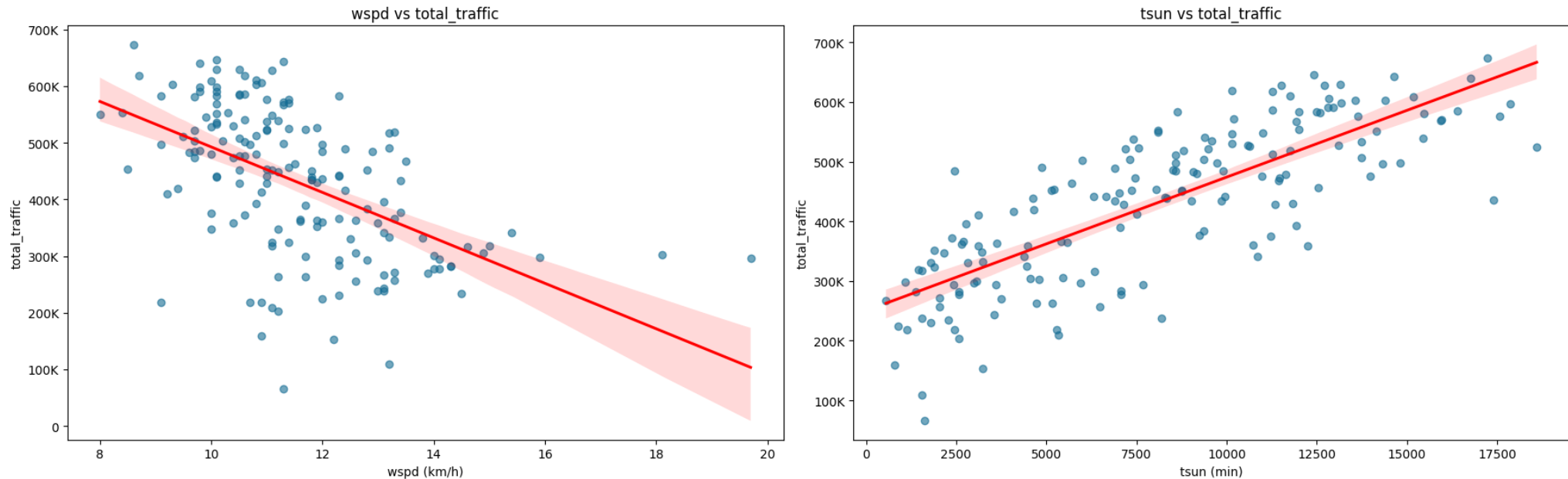
Relationship between the monthly precipitation total and average sea-level air pressure and bicycle traffic in Köln city



Q4. Is there any relation between the monthly precipitation total and average sea-level air pressure on the total bicycle traffic in Köln?

No significant relation is found between the monthly precipitation total and bicycle traffic in Köln city. Similarly, no strong relationship is observed between the average sea-level air pressure and total bicycle traffic.

Relationship between the wind speed and duration of sunshine and bicycle traffic in Köln city



Q5. How do wind speed and sunshine duration impact the total bicycle traffic in Köln?

Wind speed and sunshine duration have distinct effects on the total bicycle traffic in Köln. There is a reverse relationship between wind speed and bicycle traffic. On the other hand, there is a positive correlation between sunshine duration and bicycle traffic.

Conclusion

Summary of the project findings, key takeaways, and recommendations



Key findings:

- Temperature plays a crucial role in influencing bicycle traffic, with higher temperatures leading to increased usage and vice versa.
- No significant correlation was found between monthly precipitation total, average sea-level air pressure, and bicycle traffic.
- Wind speed negatively impacts bicycle traffic, while longer sunshine durations positively affect it.

Recommendations:

- These insights provide valuable information for policymakers and urban planners seeking to promote cycling as a sustainable transportation option.
- By considering the impact of weather conditions on bicycle usage, city officials can make informed decisions to create a more bicycle-friendly environment in Köln.

**Thanks very much
for your attention!**

The background of the slide features a series of concentric, wavy lines in shades of blue, creating a sense of motion and depth. The lines are more pronounced in the lower half of the slide.