Exploring the Impact of Weather and Climate on Bicycle sharing in London

Presented by:

Md Rifat Hasan
MSc Student in Artificial Intelligence
Advanced Data Engineering,
WS’23/24

Outline

A roadmap to guide our presentation journey

Project Introduction

Brief overview of the project, its purpose, and objectives

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

Objectives:

- Explore the variation in temperature throughout the years in London.
- Analyse the trends and patterns in bicycle traffic across different seasons and years in London 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 humidity with total bicycle traffic in London.
- Assess the influence of wind speed and sunshine duration on the total bicycle traffic.
- Assess how the different seasons influences bicycle sharing

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

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

Data Source Details

Information about the data sources used in the project

Datasource1: Bicycle Traffic Data in London

- Source: Kaggle
- Metadata URL: https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset
- Sample Data URL: https://cycling.data.tfl.gov.uk/
- Data Type: CSV

This data source contains London's bicycle traffic generated from several counting stations throughout the city from 2015.

Datasource2: Weather and Climate Data of London

- Source: Kaggle
- Metadata URL: https://www.kaggle.com/datasets/emmanuelfwerr/london-weather-data
- Sample Data URL: [https://www.ecad.eu/dailydata/index.php
- Data Type: CSV

This data source will provide weather and climate data in London, including date, cloud cover, sunshine, global radiation,max_temp,mean_temp, min_temp, precipitation, pressure, snow_depth.

Project Structure: ETL Pipeline

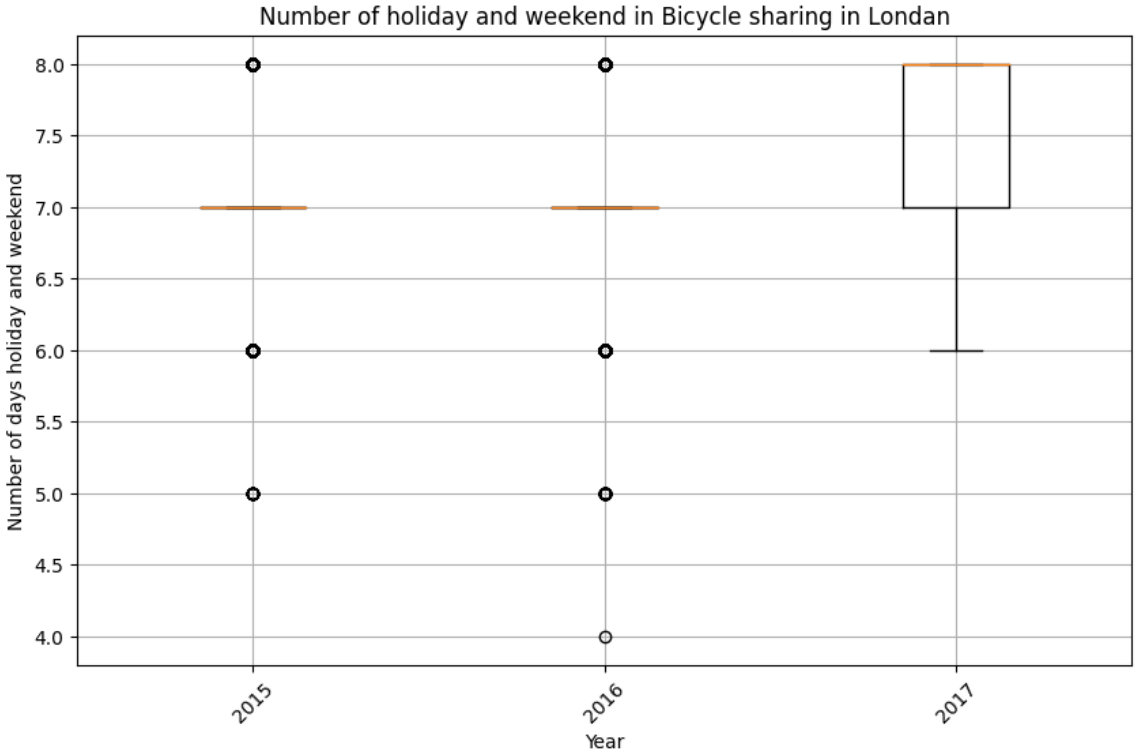
Explanation of the ETL process and overall project structure

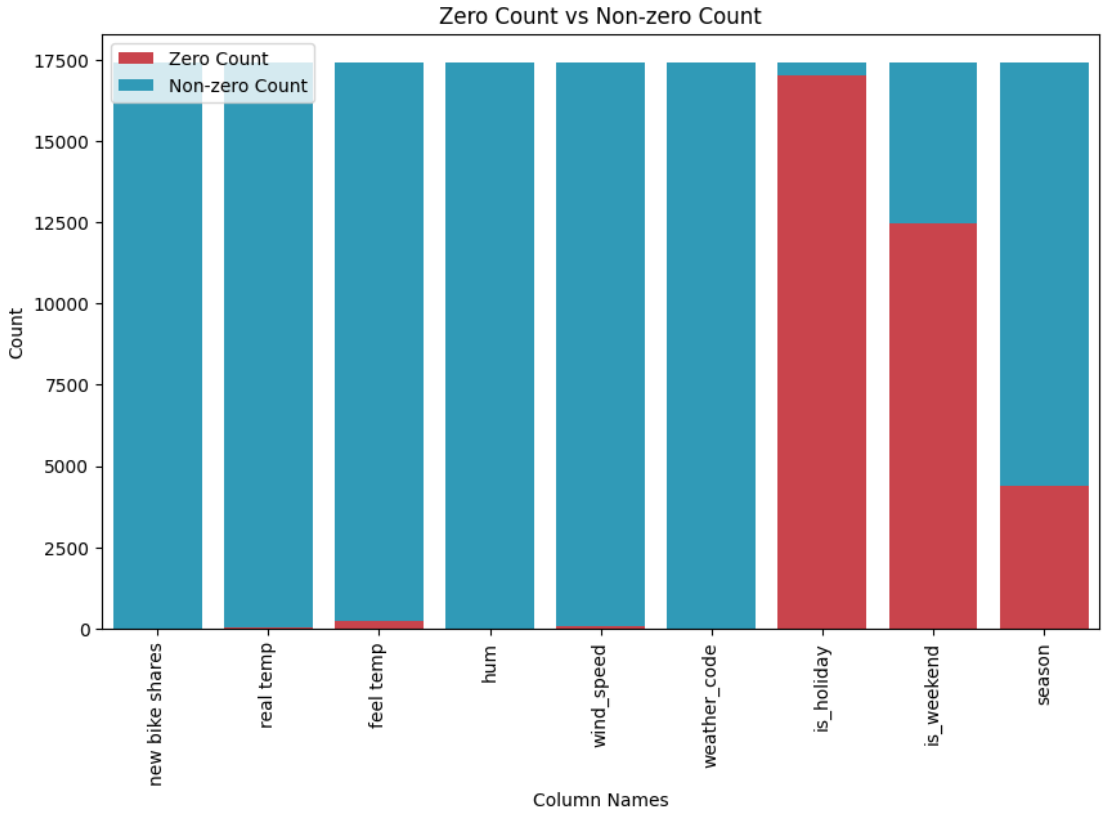
```bash
project/
├── config/                    # Configuration files and settings
│   ├── __init__.py
│   ├── basepipeline.py        # Configuration variables
│   └── kaggle.json            # Source information
├── data/                      # Data directory
│   ├── mian.sqlite            # Processed data
├── etl_pipeline.py            # ETL (Extract, Transform, Load) pipeline modules
├── tests_pipe.py              # all the unittests are wriiten here
├── 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 etl_pipeline.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 holiday and weekend in Bicycle sharing in London

Is_holiday, is_weekend has high number of null count in bike_ data dateset

Modifying and Combining Two Data Sources

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

```
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   date              15341 non-null  int64
 1   cloud_cover       15322 non-null  float64
 2   sunshine          15341 non-null  float64
 3   global_radiation  15322 non-null  float64
 4   max_temp          15335 non-null  float64
 5   mean_temp         15305 non-null  float64
 6   min_temp          15339 non-null  float64
 7   precipitation     15335 non-null  float64
 8   pressure          15337 non-null  float64
 9   snow_depth        13900 non-null  float64
dtypes: float64(9), int64(1)
```

- To analyze the overall bicycle traffic in London, we focus on first the weather dataset columns and sort the data

```
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   date              732 non-null    datetime64[ns]
 1   cloud_cover       732 non-null    float64
 2   sunshine          732 non-null    float64
 3   global_radiation  731 non-null    float64
 4   max_temp          732 non-null    float64
 5   mean_temp         732 non-null    float64
 6   min_temp          732 non-null    float64
 7   precipitation     732 non-null    float64
 8   pressure          732 non-null    float64
 9   snow_depth        730 non-null    float64
dtypes: datetime64[ns](1), float64(9)
```

- Keep only the entries from January 2015 to January 2017 to match the bike data’

Modifying and Combining Two Data Sources

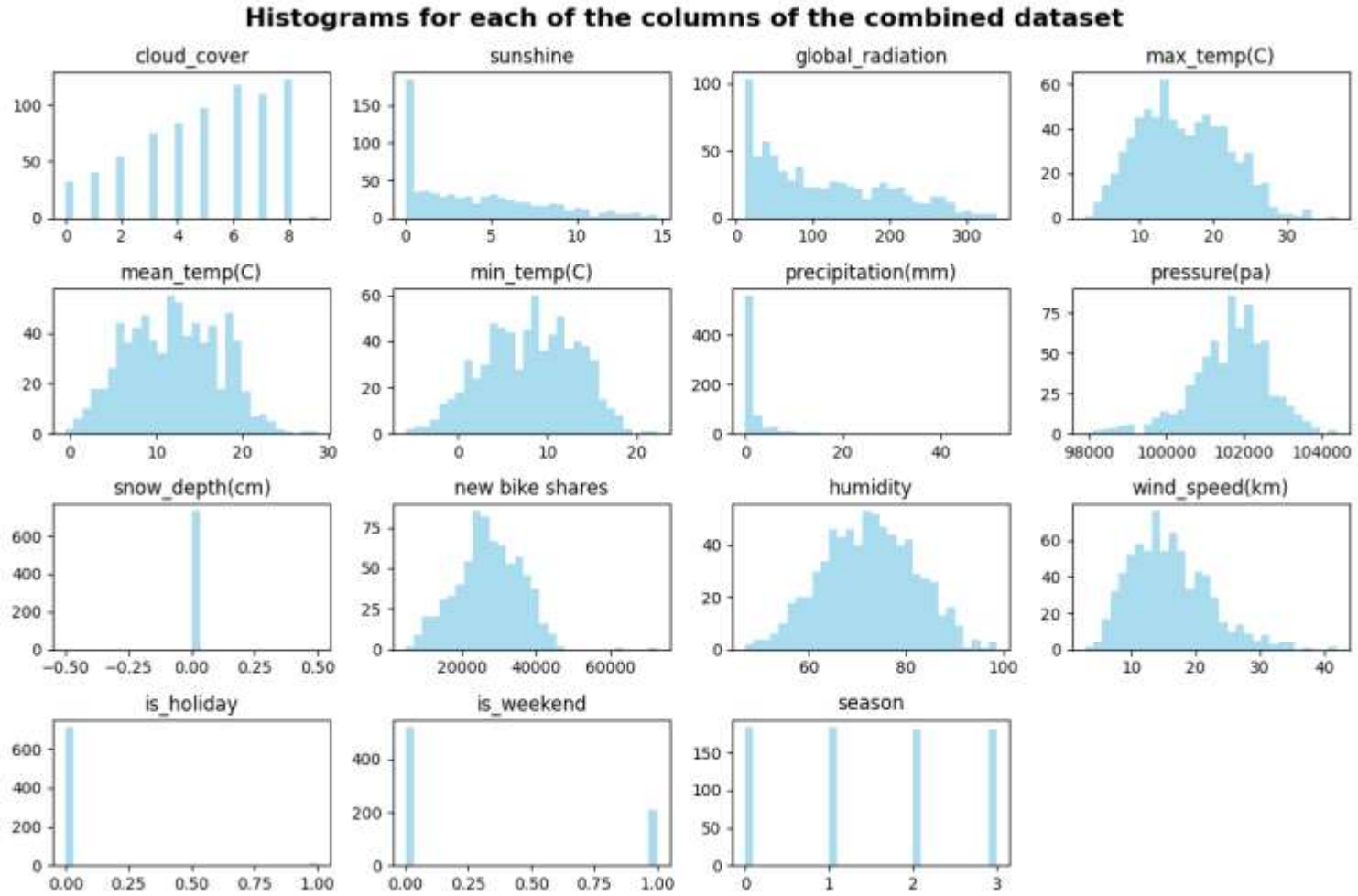
Description of how two data sources were modified and combined for analysis

	date	new bike shares	humidity	wind_speed	is_holiday	is_weekend	season
0	2015-01-04	9234.0	94.270833	7.500000	0.0	1.0	3.0
1	2015-01-05	20372.0	80.312500	8.854167	0.0	0.0	3.0
2	2015-01-06	20613.0	78.895833	16.000000	0.0	0.0	3.0
3	2015-01-07	21064.0	78.108696	19.760870	0.0	0.0	3.0
4	2015-01-08	15601.0	79.312500	20.479167	0.0	0.0	3.0

- Rename Timestamp to date
- date: The date in the month-year format.
- We will drop the real temp and feel temp column as we have the mean temp value in the weather dataset of London.

Exploratory Data Analysis (EDA)

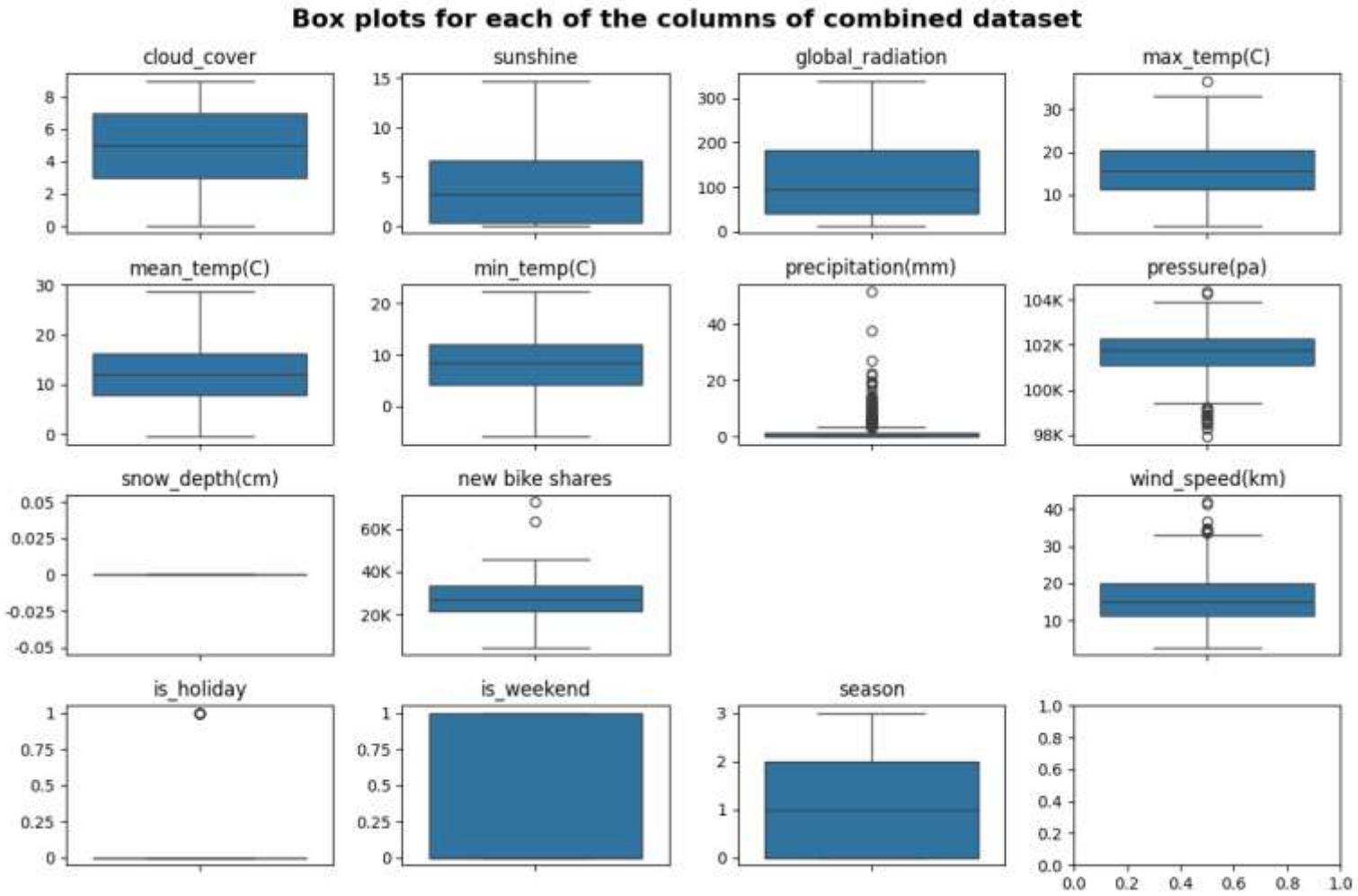
Highlights of the EDA process and insights gained from analyzing the data



- Histograms were created for each of the columns and it shows the distribution of values for columns such as cloud_cover', 'sunshine', 'global_radiation', 'max_temp', 'mean_temp', 'min_temp', 'precipitation', 'pressure', 'snow_depth', 'new bike shares', 'humidity', 'wind_speed', 'is_holiday', 'is_weekend', 'season'
- The number of bins was set to 30, and kernel density estimation (KDE) was enabled to visualize the underlying distribution.

Exploratory Data Analysis (EDA)

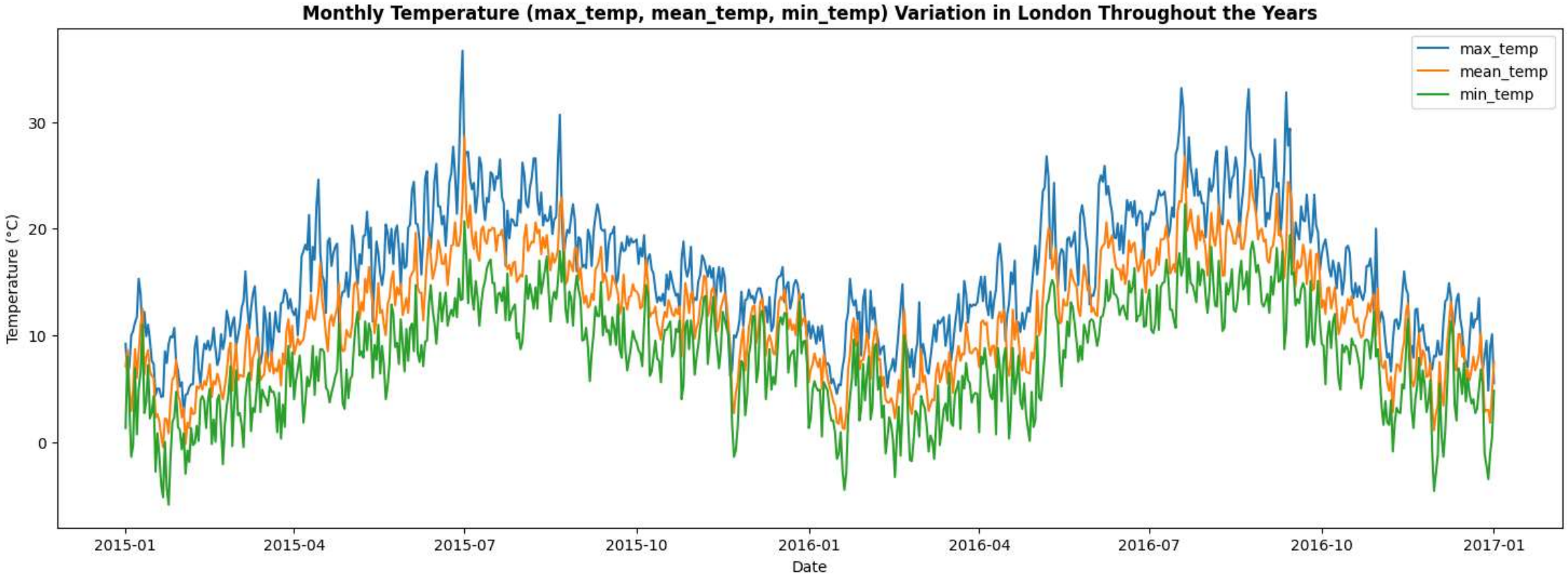
Highlights of the EDA process and insights gained from analyzing the data



Box plots 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.

Questions and Analysis

Detailed analysis of specific questions of interest and their findings

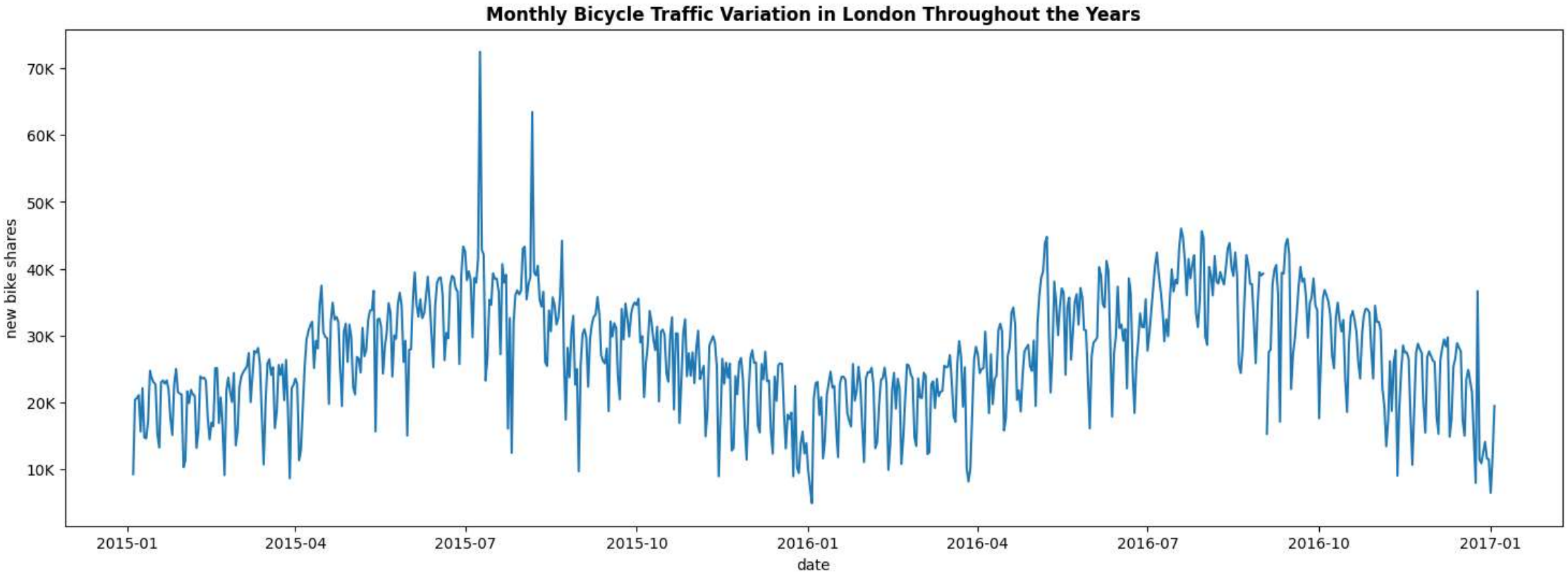


Q1. How does the temperature in London city vary throughout the years?

The temperature in London 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.

Questions and Analysis

Detailed analysis of specific questions of interest and their findings

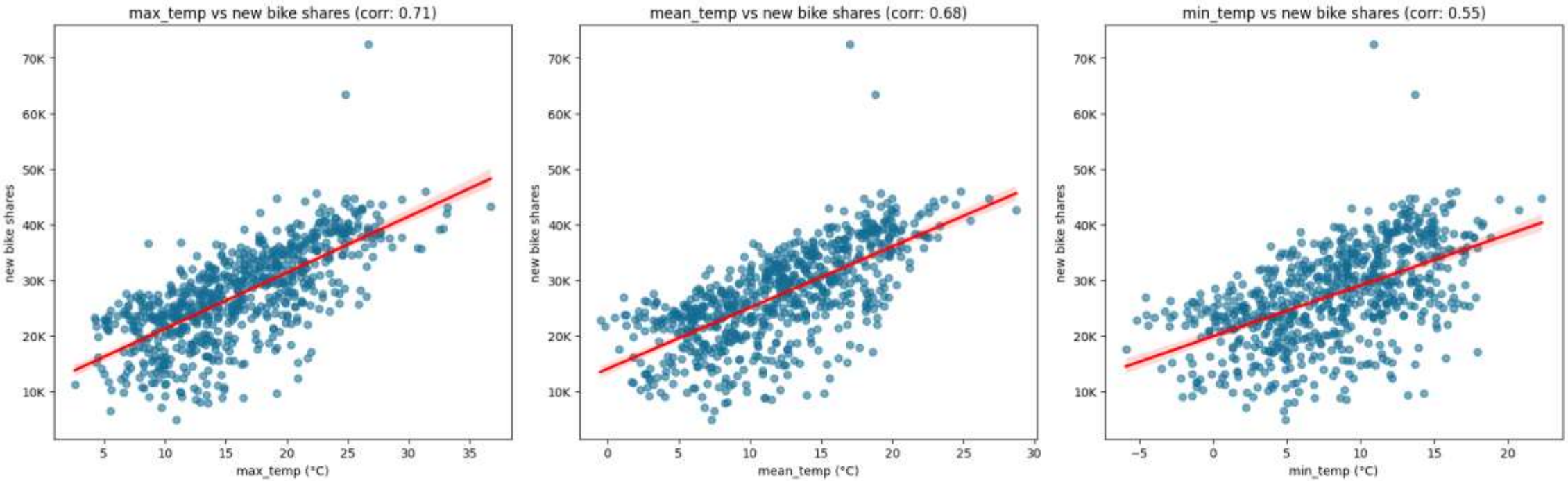


Q2. How does the bicycle traffic in London city vary throughout the years?

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

Questions and Analysis

Detailed analysis of specific questions of interest and their findings

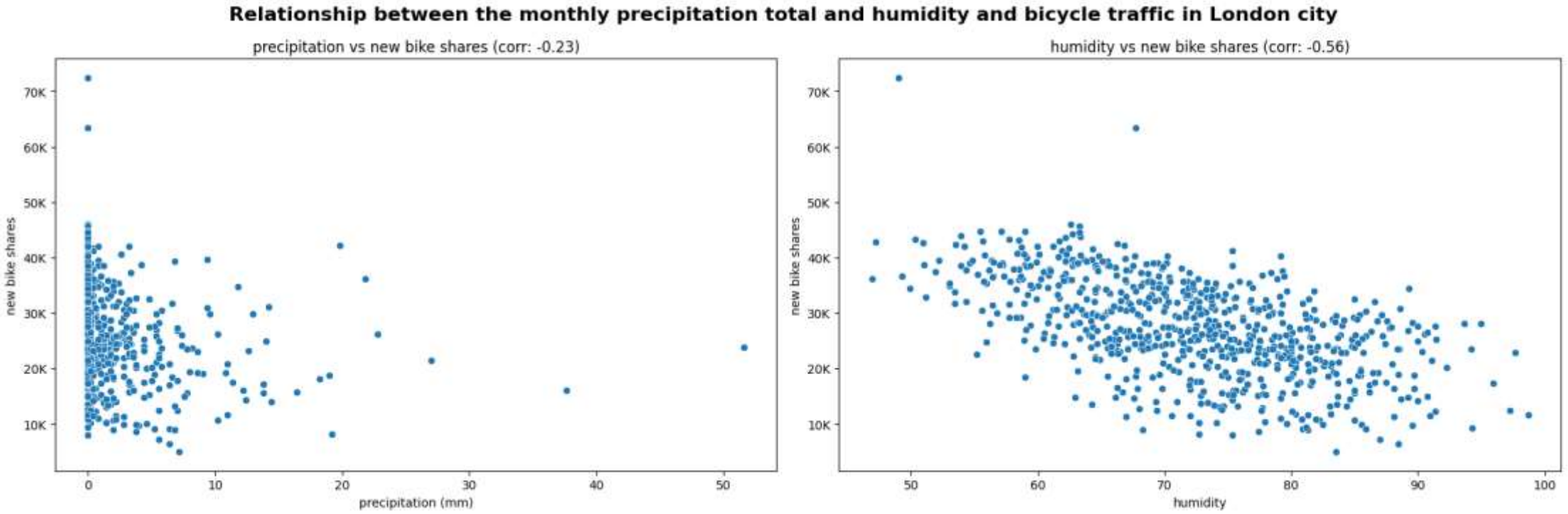


Q3. Does temperature affect bicycle traffic in London city?

Temperature has a significant impact on bicycle traffic in London city. Through the analysis of temperature data ('max_temp', 'mean_temp', 'min_temp') and bicycle traffic data, we observe a clear relationship between temperature and bicycle usage.

Questions and Analysis

Detailed analysis of specific questions of interest and their findings

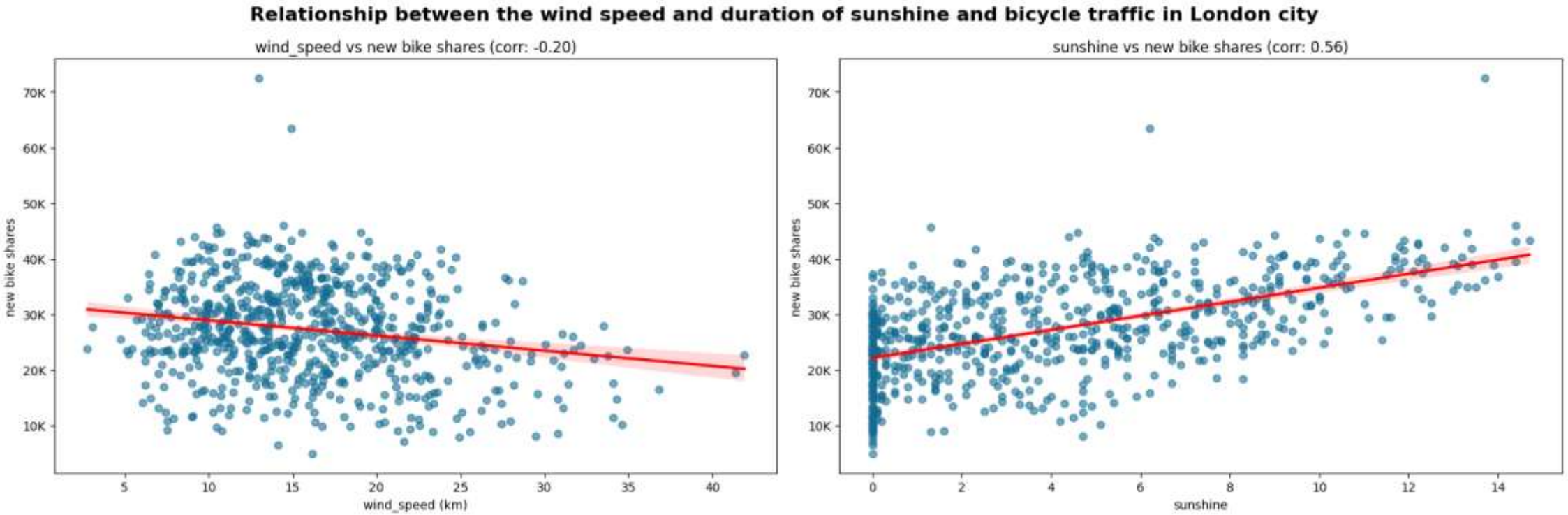


Q4. Is there any relation between the monthly precipitation total and humidity on the total bicycle traffic in London?

A reverse relation is found between the monthly precipitation total and bicycle traffic in London city. When there is less precipitation, bike sharing increases significantly. In precipitation levels throughout the year, there is s clear pattern or between the two variables(a negative correlation). Similarly, a more reverse correlation is observed between the humidity and total bicycle traffic..

Questions and Analysis

Detailed analysis of specific questions of interest and their findings

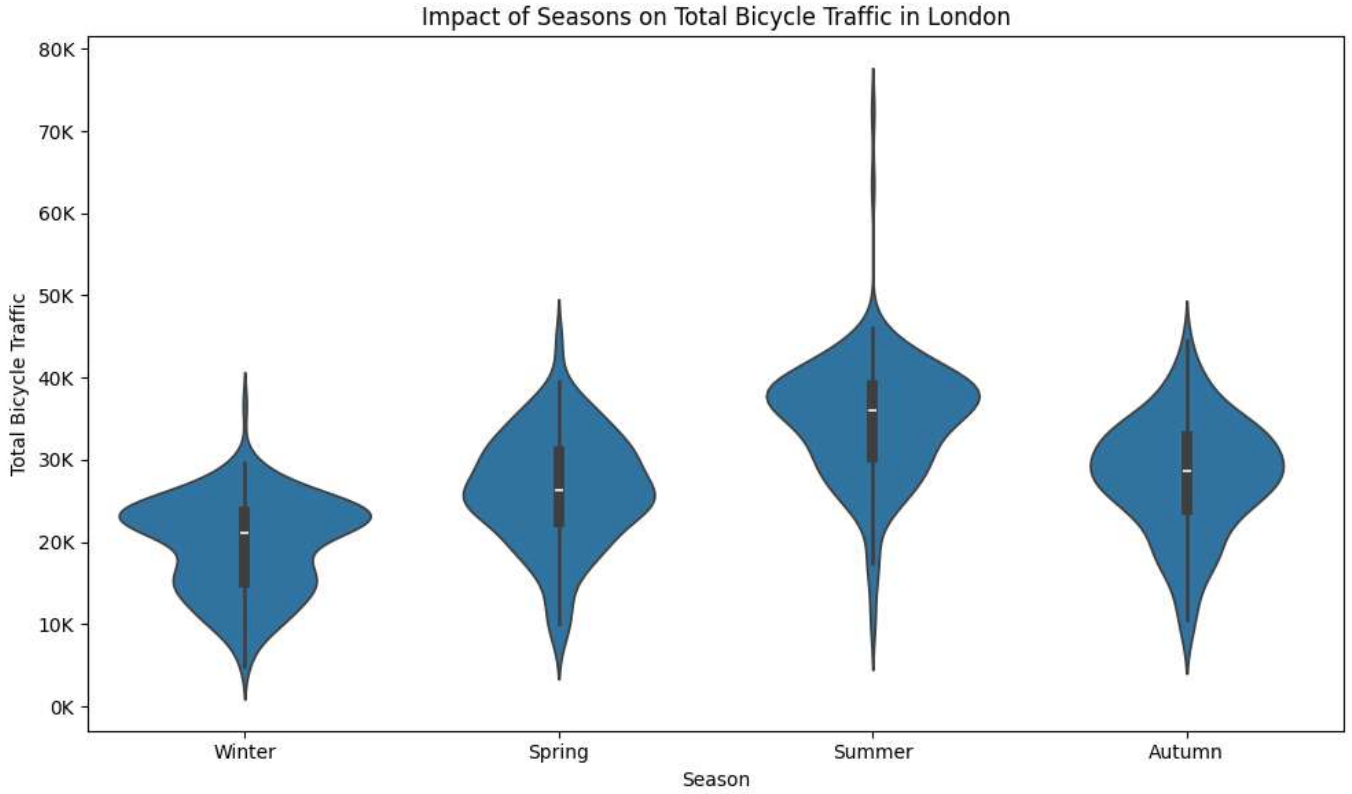


Q5. How do wind speed and sunshine duration impact the total bicycle traffic in London?

There is a reverse relationship between wind speed and bicycle traffic, indicating that as wind speed increases, bicycle traffic tends to decrease, and vice versa. This suggests that strong winds may deter people from cycling or affect their comfort and safety while cycling. On the other hand, there is a positive correlation between sunshine duration and bicycle traffic

Questions and Analysis

Detailed analysis of specific questions of interest and their findings



Q6. How do the different seasons impact the total bicycle traffic in London?

Summer sees a substantial increase in cyclists, likely due to better weather for biking. In contrast, winter has a marked decrease in bicycle use, possibly because of the cold weather. Both spring and autumn exhibit relatively higher bike traffic compared to winter, indicating a steady preference for cycling in these seasons.

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, humidity, 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 London.

Thank you!