Friedrich-Alexander-Universität
Technische Fakultät

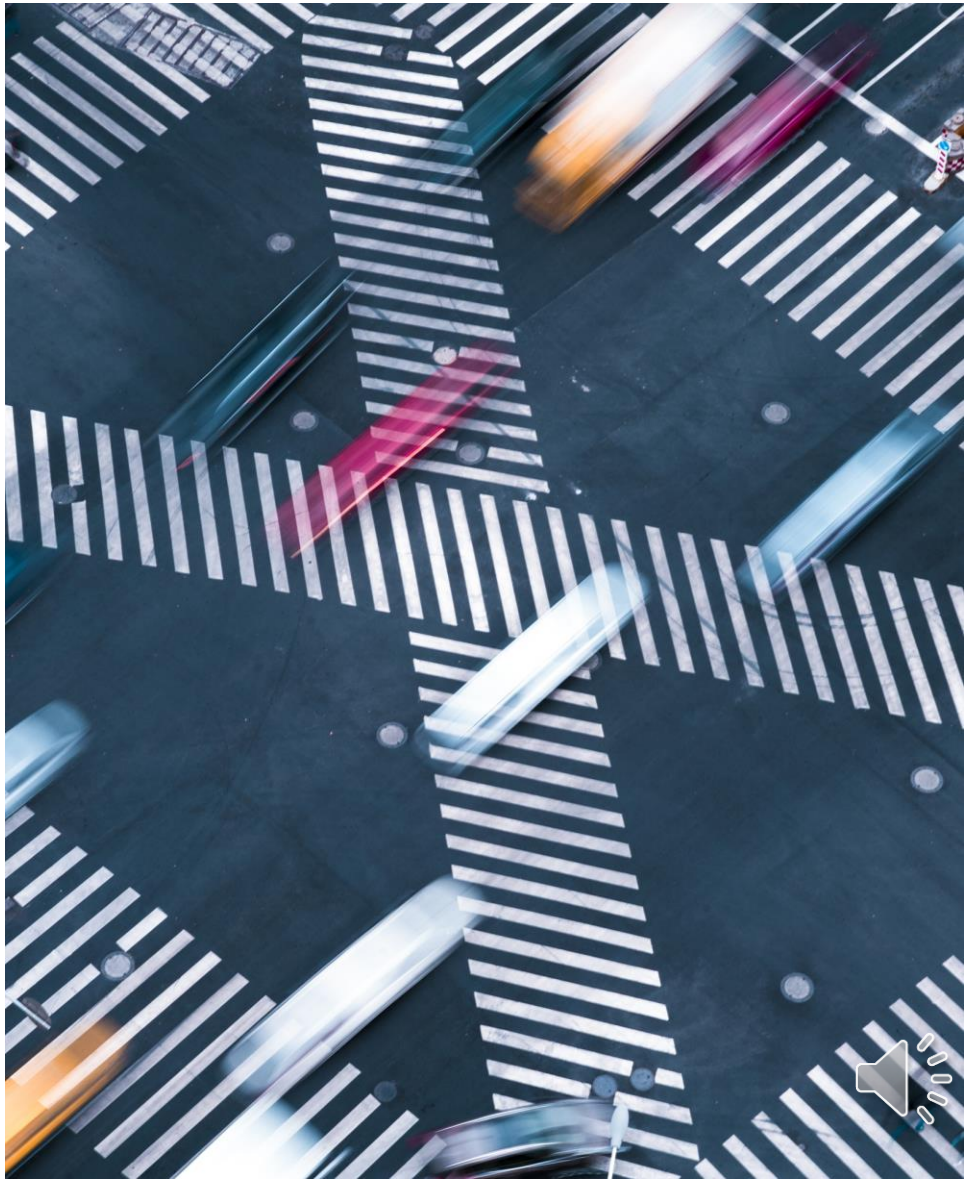
Analyze different weather conditions and their impact on different types of vehicles involved in road accidents

Methods of Advanced Data Engineering

Presented by: Sahil Sharma [ka32qewe]

Contents

- Introduction
- Project Structure
- Methods and Dataset
- Results
- Conclusion
- Limitations
- Outlook for Future Work



Introduction

Road accidents pose significant threats to public safety.

Understanding the factors influencing road accidents is crucial for developing effective safety measures.

One such factor of interest is weather conditions.

This study aims to investigate whether weather has an impact on the types of vehicles involved in road accidents.

Project Structure

- **project-plan.md:** Includes project plan, milestones, and relevant information about the datasets.
- **data_exploration.ipynb:** Contains exploratory data analysis and visualization for gaining insights into the dataset.
- **data_pipeline.py:** Implementing a data processing pipeline, for cleaning and transforming data.
- **pipeline.sh:** Shell script providing commands for running the data processing pipeline.
- **test.py:** Testing specific functionalities and aspects of the pipeline.
- **test.sh:** Shell script providing commands for running tests on the pipeline.
- **report.ipynb:** Containing the main report, presenting findings, analysis, and visualizations.

Methods and Dataset

Dataset 1: Road Accident Dataset

This dataset provides detailed information on accident types, vehicle types, month-to-month road conditions, and other relevant factors.

URL: https://www.statistik-berlin-brandenburg.de/opendata/AfSBBB_BE_LOR_Strasse_Strassenverkehrsunfaelle_2019_Datensatz.csv

Cleaning:

1. All the unnecessary columns were removed from the dataset.
2. Column names were changed to English for easy understanding.
3. Month and road condition were changed from numerical to string values for better readability.

Methods and Dataset

Dataset 2: Weather Dataset

The second dataset comprises weather data for all of Germany, which includes precision, air pressure, and other meteorological variables.

Data

URL: https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/precipitation/

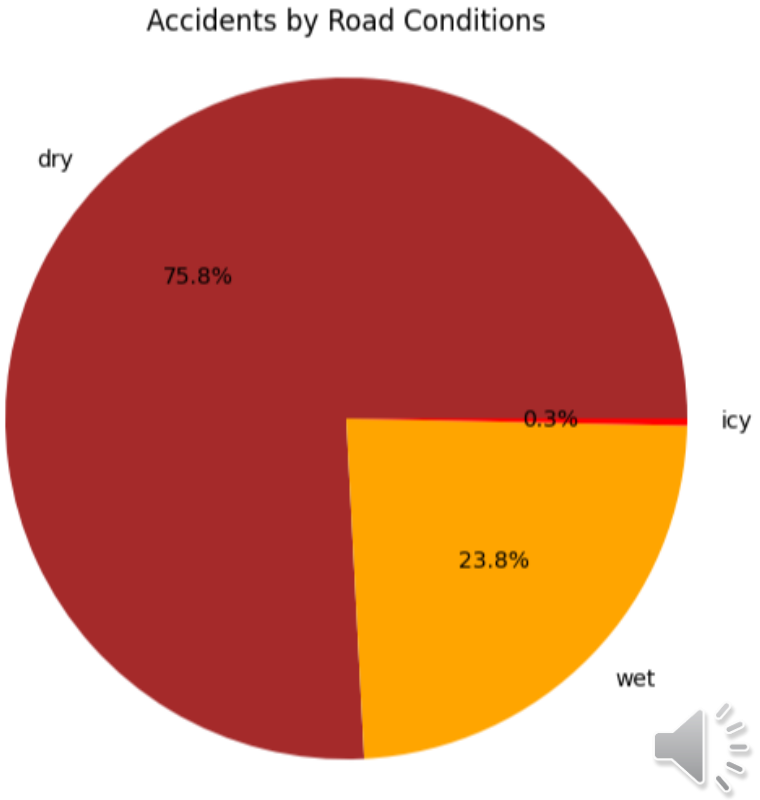
Cleaning

1. Data converted from a TXT file to a CSV.
2. Read the data only for Berlin in 2019.
3. Combine the data from precision, air pressure and sunshine duration in one data frame, referring to the month column.

Results

Q1. Which road condition has the most impact on road accidents?

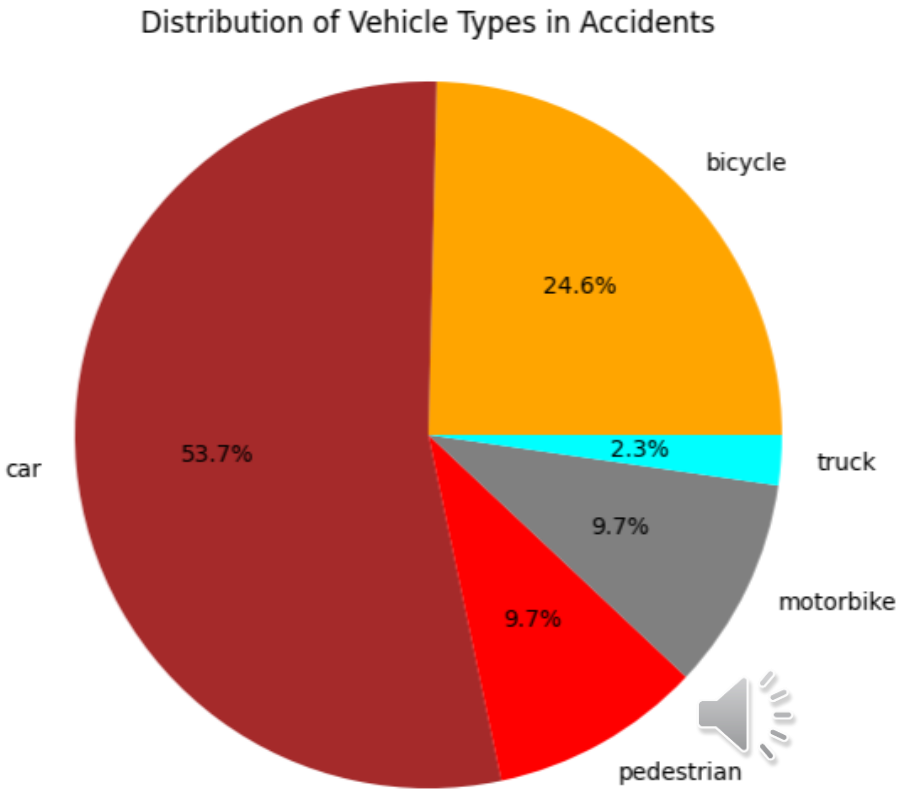
- The pie chart analysis reveals that dry road conditions constitute the majority, with a significant share of 75.8% in total accidents.
- Following dry conditions, wet road conditions contribute to 23.8% of accidents, emphasizing their considerable impact on road safety.
- Icy road conditions present the lowest risk, accounting for only 0.3% of the overall accident probability.



Results

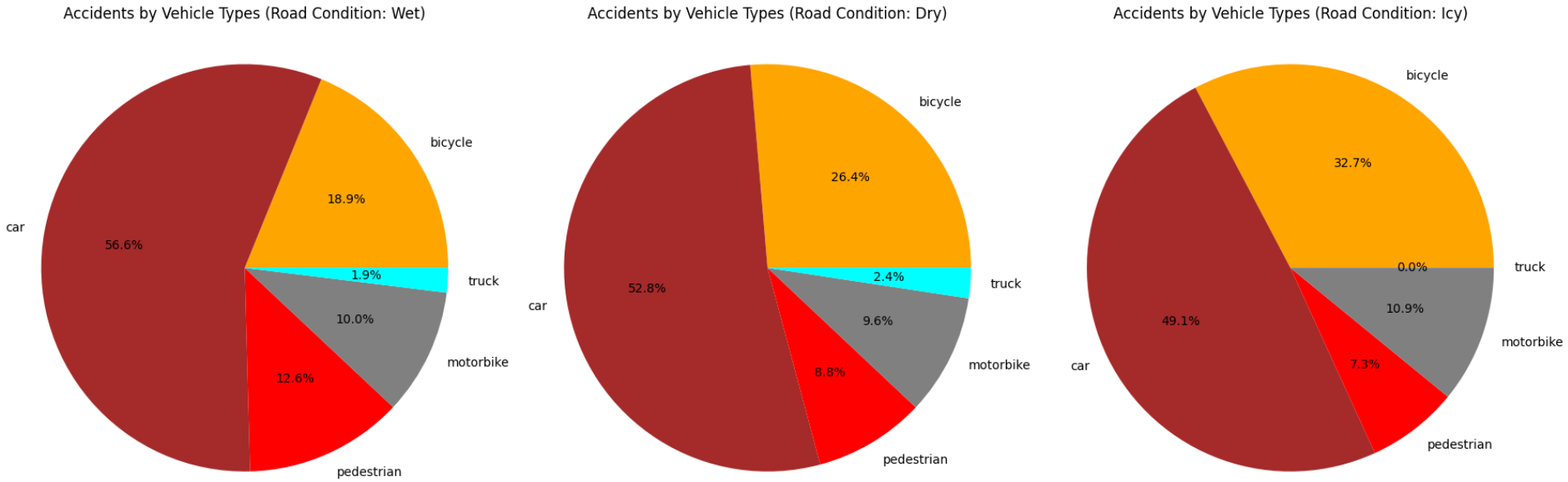
Q2. Which type of vehicle is most likely to be involved in an accident?

- The pie chart analysis indicates that cars are the most frequently involved vehicle type in road accidents, with a substantial probability of approximately 53.7%.
- Following closely, bicycles contribute significantly to accidents, representing a probability of 24.6%.
- Both pedestrians and motorbikes share a similar likelihood, each accounting for 9.7% of road accidents.
- Trucks have a lower probability of involvement, standing at 2.3%.



Results

Q3. How do different weather conditions impact the types of vehicles involved in accidents?

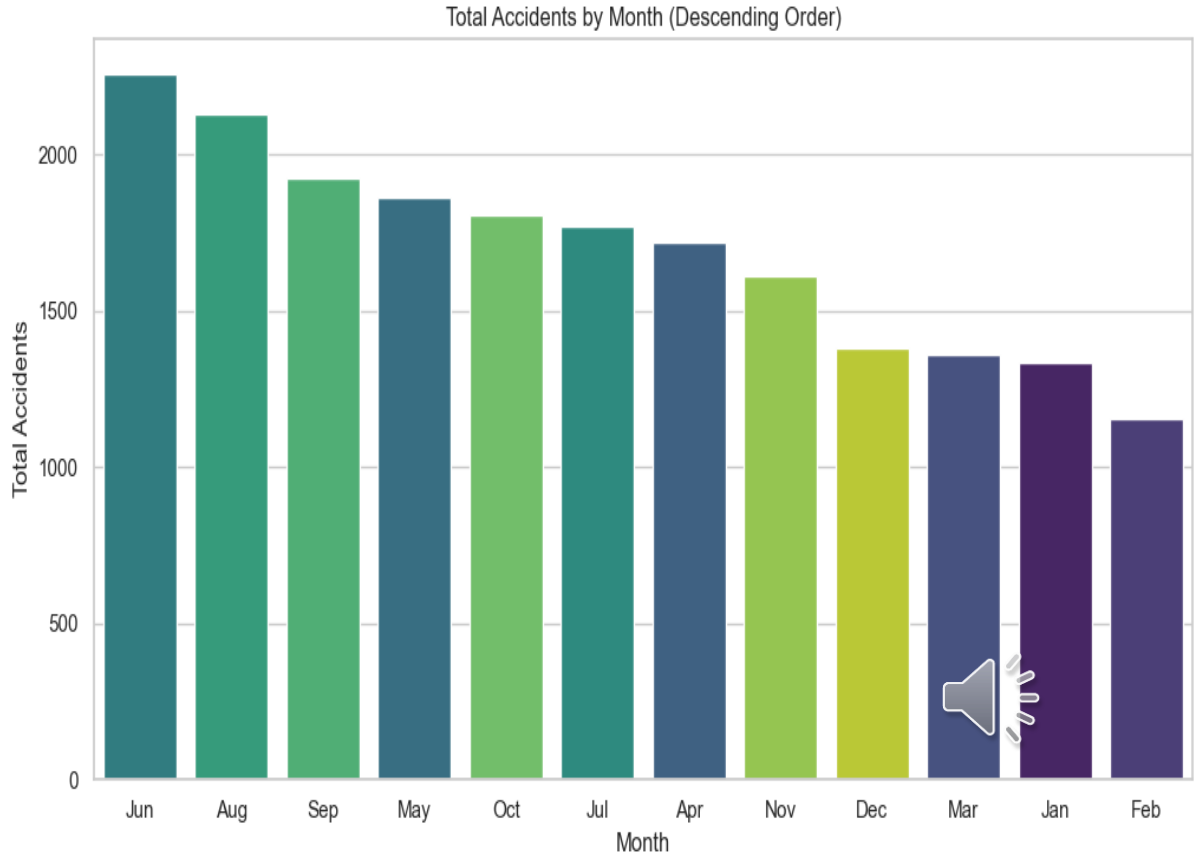


- Cars consistently lead in road accidents, with 56.6% occurring in wet conditions, 52.8% in dry, and 49.1% in icy conditions, emphasizing a higher likelihood in wet conditions
- Bicycles rank second in accidents, with the highest risk in icy conditions (32.7%), followed by dry (26.4%) and wet conditions (18.9%).
- Pedestrian accidents peak in wet conditions (12.6%), followed by dry (8.8%), and icy conditions (7.3%).

Results

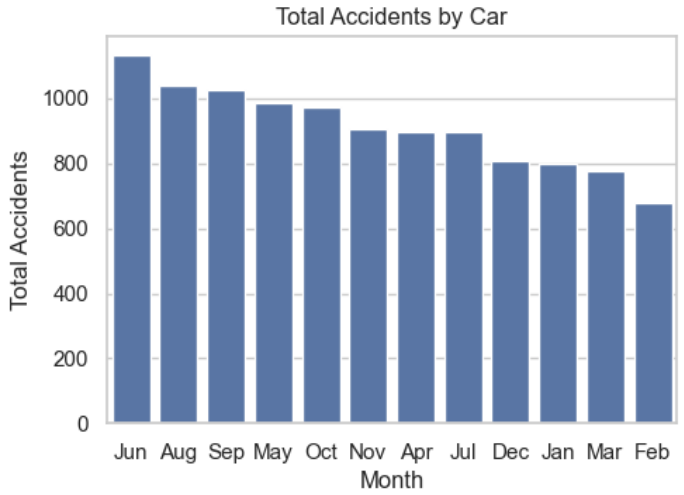
Q4. Is there any trend in vehicle accidents over the period of a year?

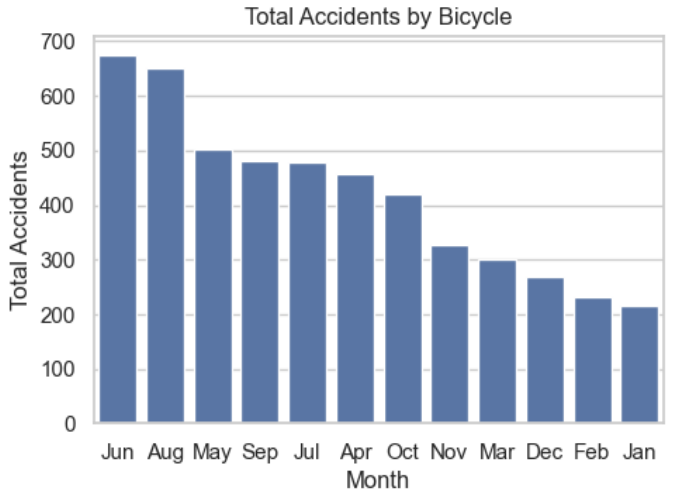
- The column chart analysis reveals that road accidents reach their highest frequency during the summer months, spanning from June to September.
- In contrast, the winter months, covering December to March, showcase the lowest occurrence of road accidents.
- This observation suggests a potential correlation between seasonal variations and road accident frequencies, highlighting summer as a period of higher risk for road accidents.

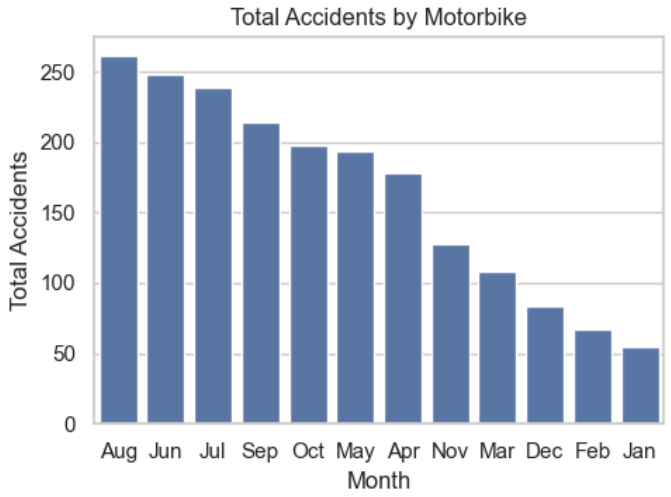


Results

- **Car Accident Patterns:** Car accidents consistently peak in summer (June to October) and decrease in winter (December to March).
- **Bicycle and Motorbike Trends:** Similar seasonal trends are observed for bicycle and motorbike accidents, with higher risks in summer and lower risks in winter.
- **Pedestrian Anomaly:** Pedestrian accidents deviate from the trend, with a higher risk in winter (October to January) and lower risks in summer (May to July).
- **Seasonal Variation Insight:** The findings highlight seasonal variations in accident frequencies across transportation modes, emphasizing the importance of mode-specific safety considerations.







Conclusion

- Dry road conditions has the highest risk of road accident.
- The accident distribution among vehicles, highlights the prominent roles of cars, bicycles, and pedestrians in incidents.
- Analysis reveals all vehicles has higher vulnerability to wet conditions but only bicycles' susceptibility to icy environments.
- Distinct patterns emerge in monthly trends, with consistently higher risks during summer months for various vehicle types.

Limitations

Single-City Limitation: Dataset's focus on a single city, limiting generalization to areas with distinct traffic dynamics.

Limited Time Frame: The one-year timeframe of the dataset hinders the identification of long-term trends.

Monthly Weather Granularity: The availability of weather data on a monthly basis, rather than daily or hourly, limits the analysis's granularity.

Absence of Traffic Data: Lack of traffic data restricts the analysis's depth, as it is crucial for identifying specific risk factors associated with traffic patterns.

Outlook for future work:

- Aggregate road accident data from multiple cities for a diverse and representative dataset.
- Extending the dataset timeframe beyond a year to reveal long-term trends and patterns in road accidents.
- Acquiring daily or hourly weather data for a robust analysis of the correlation between weather conditions and road accidents.
- Integrating comprehensive traffic-related information, would provide valuable context for understanding road accidents.

Thank you for your Attention