Methods of Advanced Data Engineering

Final Report
Ajharali Shaikh
23072460

Analysing climate conditions and their involvement on road accidents.

Data Sources:
There are two primary datasets that were utilized in this study. The first dataset contains road accident data in Berlin, and the second dataset contains weather data for different cities all over Germany.

Dataset 1: Road Accident Dataset

This dataset provides detailed information on accident types, vehicle types, month-to-month road conditions, and other relevant factors. This dataset was chosen because it provides all the necessary values that are needed to accurately analyze the problem statement.
URL: https://www.statistik-berlinbrandenburg.de/opendata/AfSBBB_BE_LOR_Strasse_Strassenverkehrsunfaelle_2019_Datensatz.csv

Dataset 2: Weather Dataset

The second dataset comprises weather data for all of Germany, which includes precision, air pressure, and
other meteorological variables.
URL:
https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/precipitation/

Data Pipeline

The data pipeline process involves loading the two datasets from their respective URLs into a Python script, performing data cleaning and transformation, and merging both the road accident and weather datasets based on month. After that, save the final dataset in SQLite. There are two primary datasets that were utilized in this study. The first dataset contains road accident data in Berlin, and the second dataset contains weather data for different cities all over Germany.

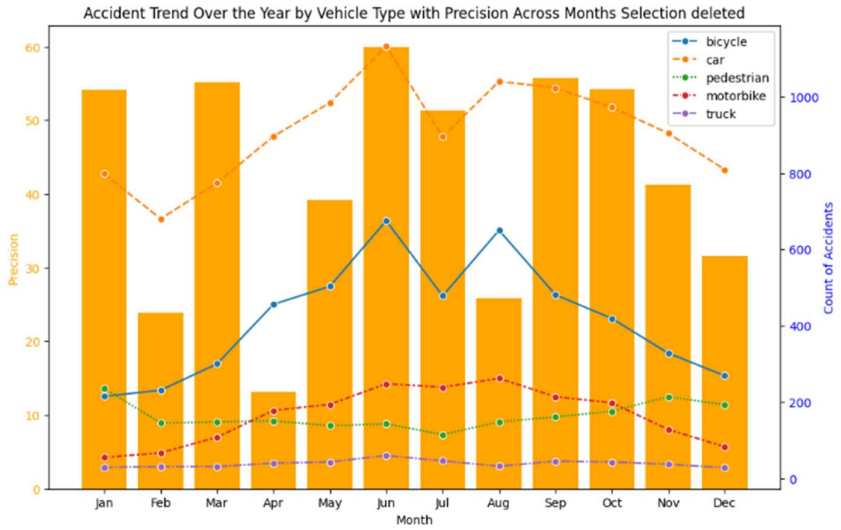
Data ingestion: fetching data from URLs
Data Cleaning:
Data Source 1:
1. All the unnecessary columns were removed from the dataset.
2. Column names were changed to English for easy understanding.
3. Month and road condition were changed from numerical to string values for better readability.

Data Source 2:
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:

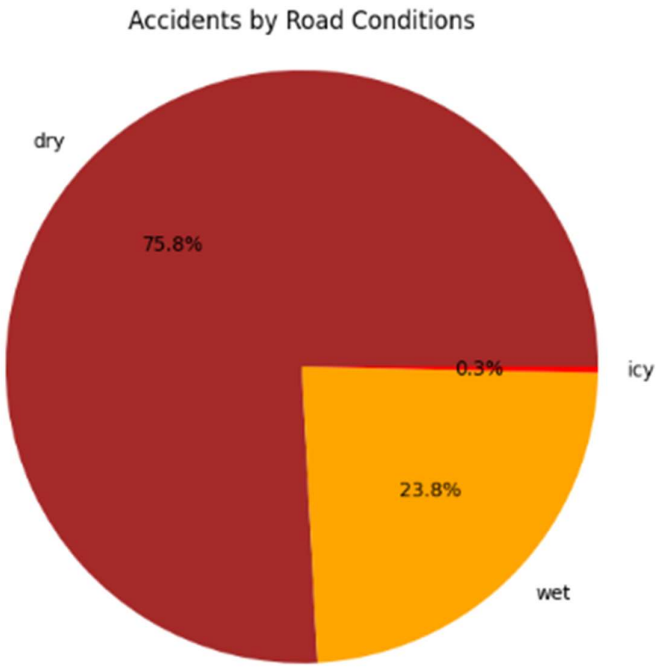
Output Data:
Upon examining the graph below, it's clear that there is no direct correlation between the precision values
from the weather dataset and the total number of road accidents involving different types of vehicles. This
lack of correlation developed from the limited precision values available, as we only have 12 values for

each month, making them insufficient for robust analysis. Consequently, I've opted to utilize road accidents as a reference point for analyzing the impact of weather conditions on the types of vehicles involved in road accidents.



Q1. <u>Which road condition has the most impact on road accidents?</u>

To address this inquiry, I visualized a pie chart depicting the total number of accidents and their corresponding road conditions. The road conditions considered were 'dry' for sunny weather, 'icy' for snowy conditions, and 'wet' for rainy weather.

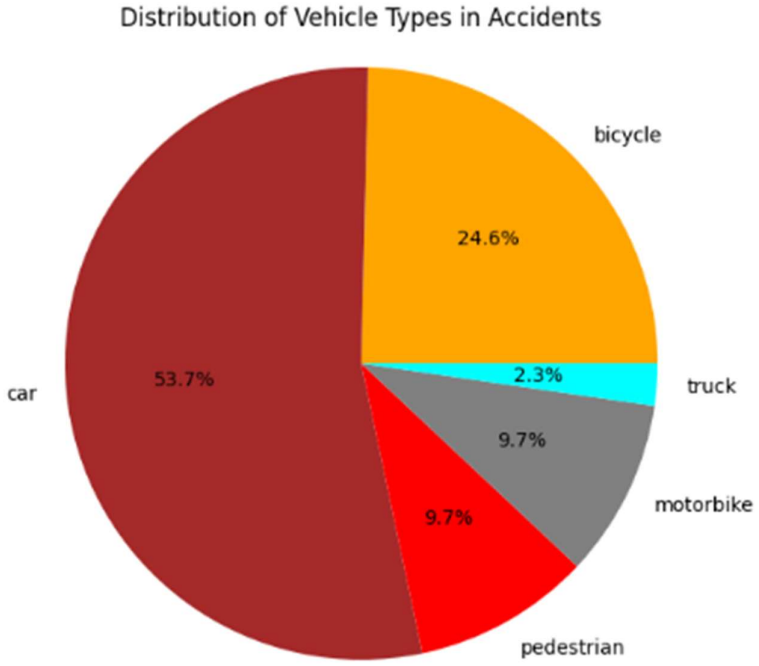


Inference from Question 1:

Analyzing the pie chart, it's evident that dry road conditions account for the majority, comprising 75.8% of the total accidents. Wet road conditions follow, contributing to 23.8% of the accidents, while icy conditions pose the least risk, constituting only 0.3% of the overall accident probability. This highlights the significant impact of weather conditions on road safety, with dry conditions being the most prevalent but also associated with a higher accident rate.

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

In addressing this question, I utilized a pie chart to visually represent the distribution of road accidents involving various types of vehicles. The chart provides a comprehensive overview by depicting the total number of accidents attributed to each vehicle category. This analysis offers valuable insights into the relative contribution of different vehicle types to overall road accidents, helping to identify potential areas of focus for safety measures and interventions.



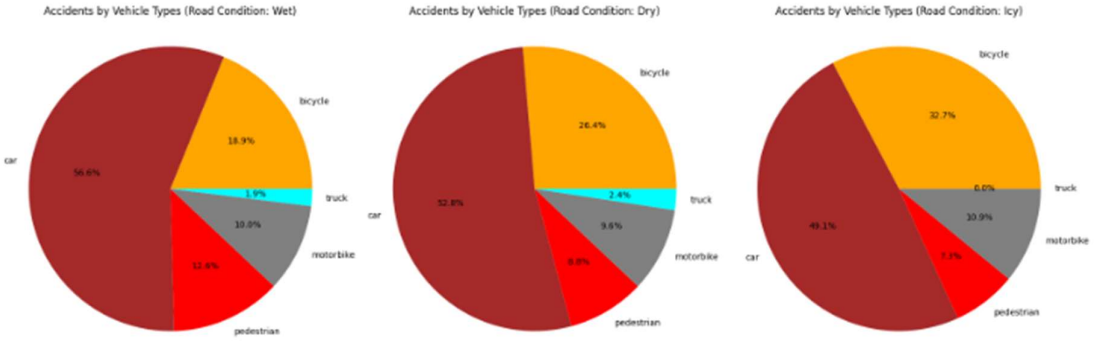
Inference from Question 2:

Analyzing the pie chart, it becomes evident that cars are the most frequently involved vehicle type in road accidents, constituting a significant probability of approximately 53.7%. Following closely, bicycles contribute to accidents with a probability of 24.6%. Pedestrians and motorbikes share a similar likelihood, both accounting for 9.7% of road accidents. Trucks have a lower probability of involvement, standing at 2.3%. This breakdown provides a clear understanding of the distribution of road accidents among different vehicle types, emphasizing that the cars have the highest chances of getting involved in a road accident.

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

To address this query, I generated three separate pie charts, each representing a distinct road condition. In each chart, I calculated the total occurrences of various vehicles involved in road accidents corresponding to the respective road conditions. This visual representation offers valuable insights into the impact of different road conditions on the distribution of accidents across various vehicle types.

By individually examining the pie charts for 'wet,' 'dry,' and 'icy' road conditions, one can understand the specific contributions of different vehicles in accidents under each condition. This analysis aids in understanding the relative impact of road conditions on various vehicle accidents, providing a better perspective for potential safety measures and interventions in specific scenarios.



Inference from Question 3:

Analyzing the information conveyed by the three pie charts, it is evident that cars consistently emerge as the most prevalent vehicle involved in road accidents across different road conditions. The distribution reveals that wet road conditions contribute to 56.6% of car accidents, followed by 52.8% in dry conditions and 49.1% in icy conditions. This highlights that car accidents are more likely to happen in wet road conditions.

Bicycles represent the second-most common vehicle type in road accidents. Notably, icy road conditions pose the highest risk for bicycle accidents, accounting for approximately 32.7%, followed by dry conditions at 26.4% and wet conditions at 18.9%.

For pedestrians, the analysis indicates that wet road conditions present the highest probability of accidents, constituting approximately 12.6%. Dry conditions follow with 8.8%, and icy conditions exhibit a 7.3% likelihood of pedestrian accidents. This detailed plot enhances our understanding of the distinctive patterns in road accidents involving various vehicles across different road conditions.

Limitations:

Single-City Dataset: The limitation of having the accident dataset for a single city implies that the findings
and patterns observed may not be universally applicable to other locations with different traffic dynamics,
infrastructure, and socio-economic factors. Road safety issues can vary significantly between cities, so
generalizing findings to a broader context may not be accurate.
Limited Time Frame: The dataset covering only one year may restrict the ability to identify long-term trends or seasonal variations that could influence road accidents. A more extended timeframe would provide a more comprehensive understanding of patterns and fluctuations, allowing for better-informed
decision-making and the planning of safety measures.
Monthly Weather Data: The availability of weather data only monthly rather than daily or hourly restricts
the granularity of the analysis. Accidents may be influenced by specific weather conditions at different times of the day or week, and having more detailed weather data would enhance the accuracy of identifying correlations between road accidents and weather.