NYCU 2023 Autumn

Data Vislualization
Final Project Report Team23

Bo-Han Chen (陳柏翰)
Student ID: 312551074
bhchen312551074.cs12@nycu.edu.tw

Xu Lin (林煦)
Student ID: 312553027
f94061042.cs12@nycu.edu.tw

Abstract

In this project, we develop a data vislualization system for raffic accident data in Taiwan in 2022. Our goal is to find out several insights related to the cause and trend of traffic accident, which can help the experts and government to improve traffic safety. With our system, the user can easily find out the temporal and spatial trend of the accident, additionally, the driver-contributed cause and vehicle information are also included, which can be used for further analysis.

Motivation

Since there are many people injured or killed in traffic accidents every year, so the analysis of the traffic accident data is important and necessary to prevent the accident. Some of the accidents are caused by the driver's behavior, and some are caused by the road condition and the circumstance such as weather condition at the time that the accident happened. Therefore, an easy-to-understand data vislualization can help the experts to analyze the cause of the accident and promote the right policy to drivers and pedestrians. By analyzing the circumstance of the accident, we can also find the dangerous road condition and time period, which can help the government to decrease the accident rate by improving the road condition and employing more traffic police at the certain time period.

Questions to Answer

The following questions are what we want to answer by visualizing the dataset.

1. Which time period has the highest accident rate?
2. Which weather condition mainly causes the accident?
3. Is there any relationship between the driver-contributed causes and the type of vehicle?
4. Which part of the vehicle is most likely to be hit and fragile during the crash?

Dataset

The dataset we choose is the Taiwan road traffic accident statistics in 2022 (*111 年傷亡道路交通事故資料*), which contains the details of the accidents that cause death or injury in Taiwan. The dataset is public and can be accessed from Open Data Website of Taiwan Government [1].

Data Description

The dataset contains 845547 records, including 4544 A1 level records and 841003 A2 records. The A1 level means the accident causes death in within 24 hours, and the A2 level means the accident causes injury or death in more than 24 hours. Each record contains 51 attributes, including the time, location, weather, road type, vehicle type, and the number of death/injury. Some details such as whether the driver is drunk or not, which part of the vehicle hit in the accident, and the driver-contributed cause of the accident are also included.

[1] `https://data.gov.tw/dataset/161199`.

Methodology

Figure 1 shows the overview of our vislualization system. The system is devided into 6 parts, including Taiwan traffic accident map, line chart of death / injury trend, stream chart of death / injury trend on different weather condition, heatmap of accident rate, stacked bar chart of driver-contributed cause and vehicle type, and the crash position distribution of vehicle.

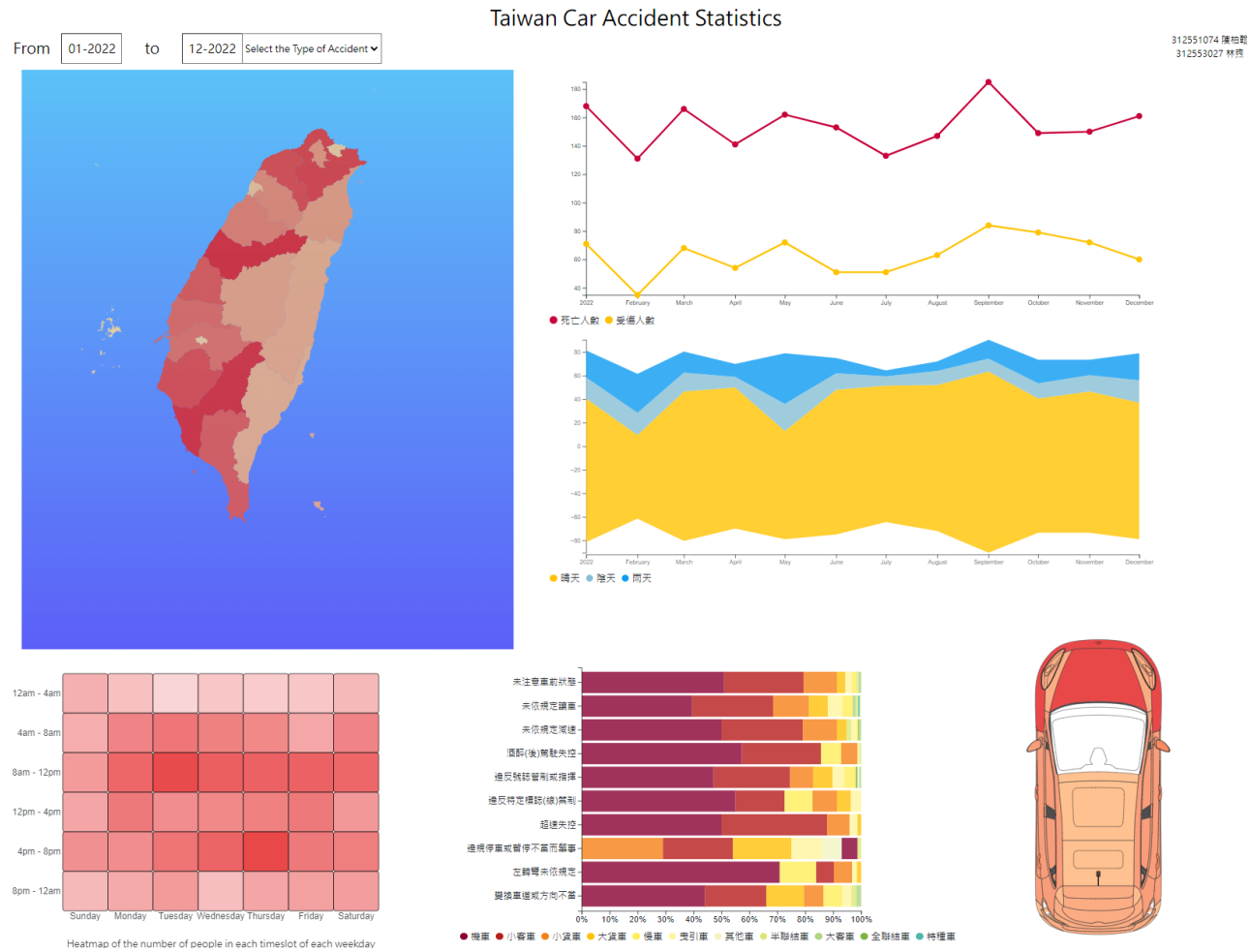


Figure 1: System Overview

Taiwan Traffic Accident Map

This map shows the death / injury count of each cities in Taiwan. The user can directly indicate which city has high fatality counts in the selected time period with the ordinal color scheme. For detailed information, the user can move the mouse to the city and the tooltip will show the death / injury count, drunk driver count, and the mobile phone usage count, which shows in Figure 2.

Line Chart of Death / Injury Trend

This line chart shows the death / injury trend in the selected time period, which can help the user to find out the temporal trend of the accident. The chart will also show fine-grained information when the selected time period is small, For example, if the user selects the time period of 1 month, the chart will show the trend of each day in the selected month, if the selected period is larger than 1 month, the chart will show the weekly trend. The tooltip shows the time period and death / injury count of each data point, which shows in Figure 3.

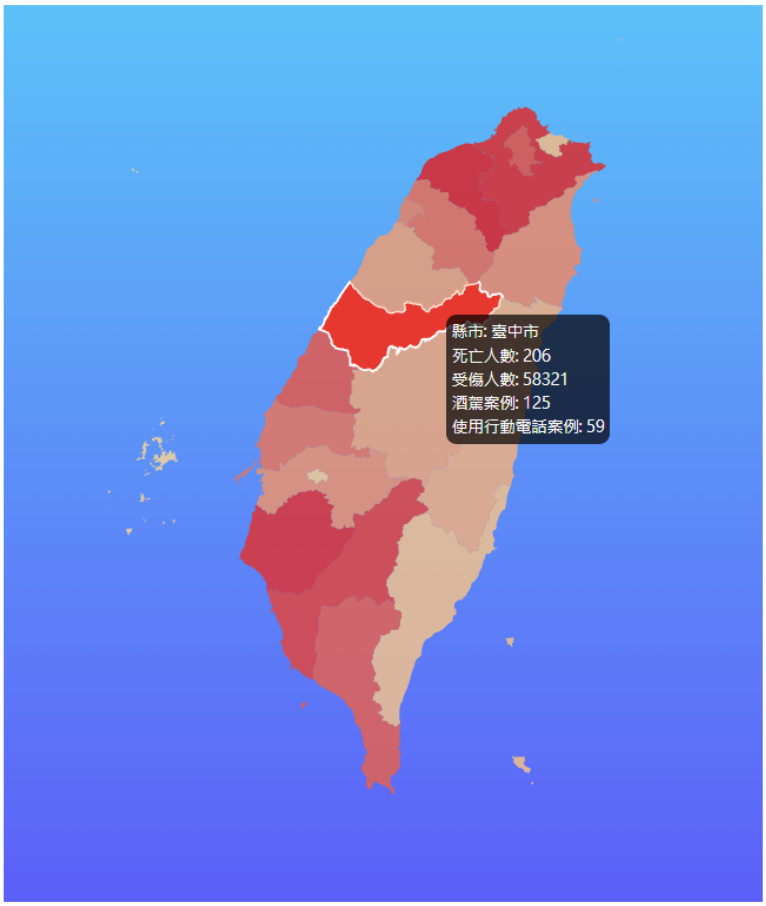


Figure 2: Map Tooltip

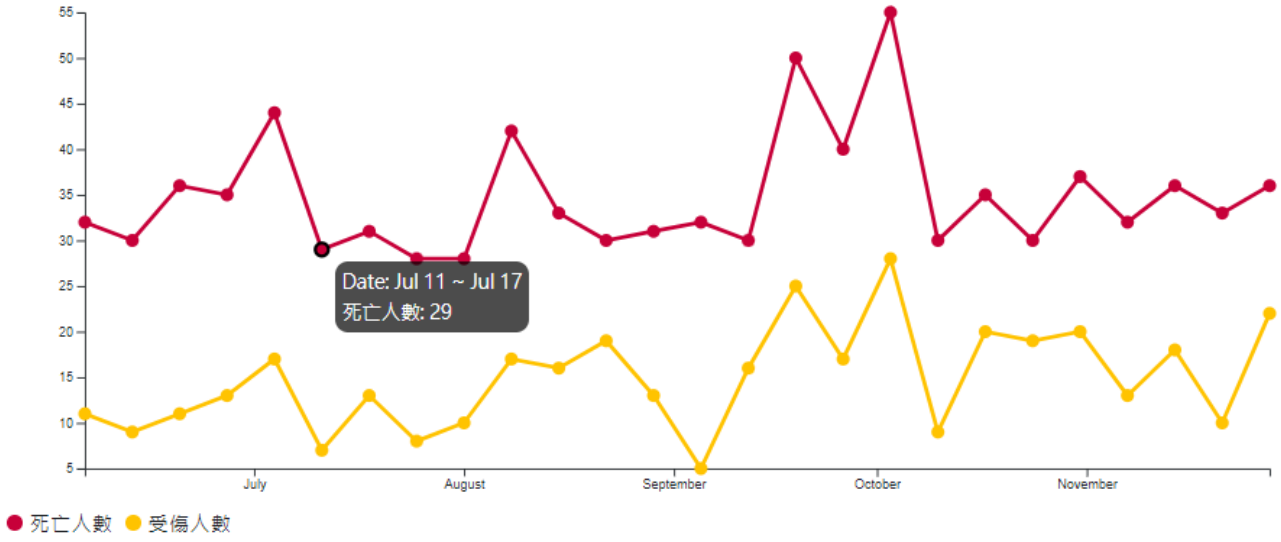


Figure 3: Line Chart Tooltip

Stream Chart of Accident Count on Different Weather Condition

This stream chart shows the accident count on different weather condition. User can find out the accident trend on different weather condition. The weather condition is divided into 3 categories, including sunny, cloudy, and rainy. Figure 4 shows the stream chart in our system, the tooltip shows the time period and accident count on different weather condition.

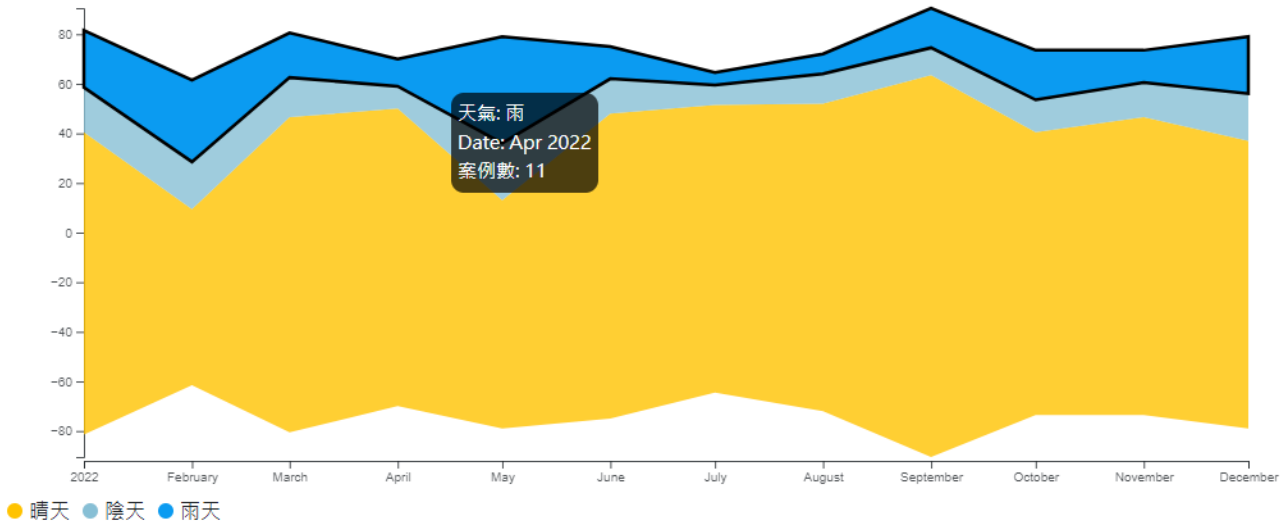


Figure 4: Stream Chart Tooltip

Grid Heatmap of Accident Count

The heatmap shows the accident count by time of day and day of week. This helps user to find out the cyclical trend of the accident, such as comparing the accident count between weekday and weekend, and the accident count in the morning and afternoon. Figure 5 shows the heatmap in our system, the tooltip shows the time period and accident count of each cell.

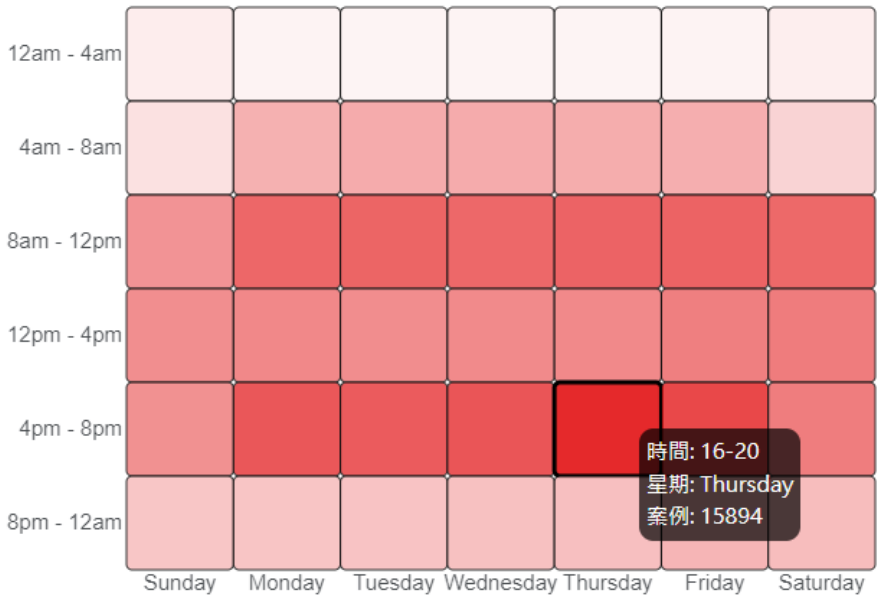


Figure 5: Heatmap Tooltip

Stacked Bar Chart of Driver-Contributed Cause and Vehicle Type

This stacked bar chart shows the ratio of vehicle type for each driver-contributed cause. The causes of accidents are listed from top to bottom in descending order according to the number of accidents

they caused. User can find out the vehicle type that is most likely to be involved in the corresponding cause. Figure 6 shows the stacked bar chart in our system, on this chart, we can find out that the most frequent cause is ” 未注意車前狀態”, and the tooltip shows that scooter is the most likely vehicle type to be involved in this cause, which is 50.9% of the total accident count.

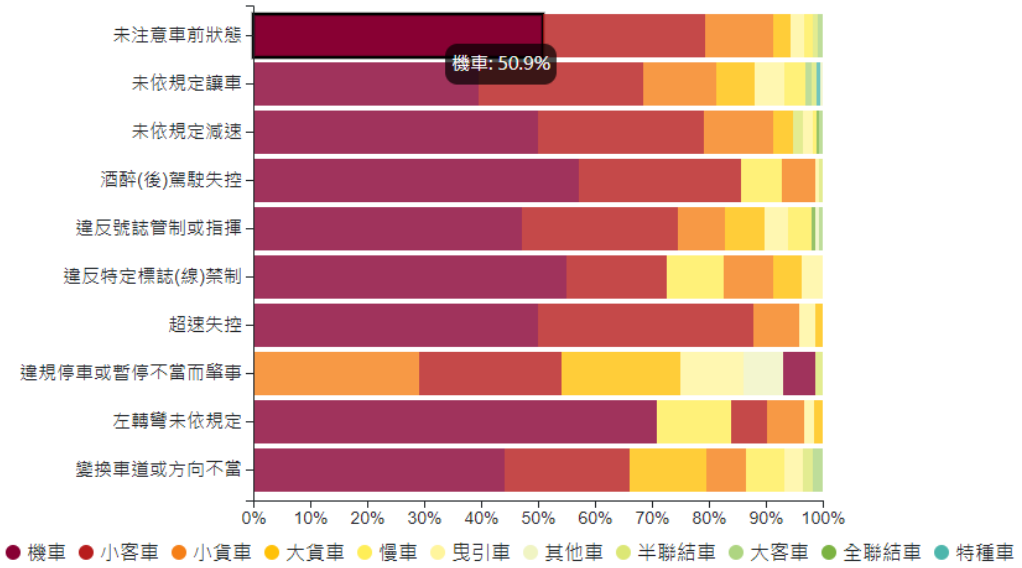


Figure 6: Stacked Bar Chart Tooltip

Crash Position Distribution of Vehicle

This chart shows the crash position distribution of vehicle. In our dataset, the crash position is divided into 8 categories, including front, rear, left, right, front-left, front-right, rear-left, and rear-right. User can find out which part of the vehicle is most likely to be hit and fragile during the crash by the ordinal color scheme. Figure 7 shows the crash position distribution of vehicle in our system, the tooltip shows the crash position and the accident count.

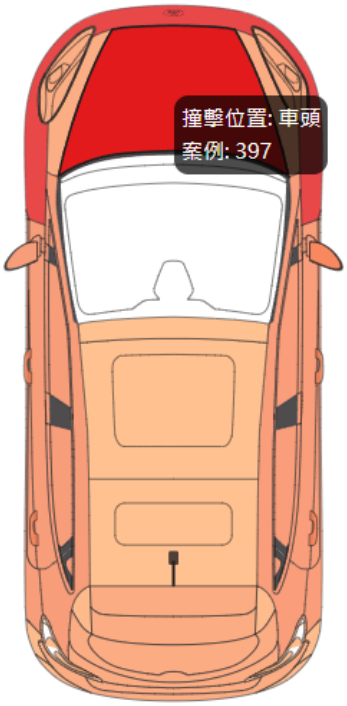


Figure 7: Crash Position Distribution of Vehicle Tooltip

Interaction

The user can select the time period and type of accident data (A1, A2, or both) in the control panel on top of the system (Figure 8). Furthermore, city selection is also available in the map, by clicking the city, system will show the corresponding information.

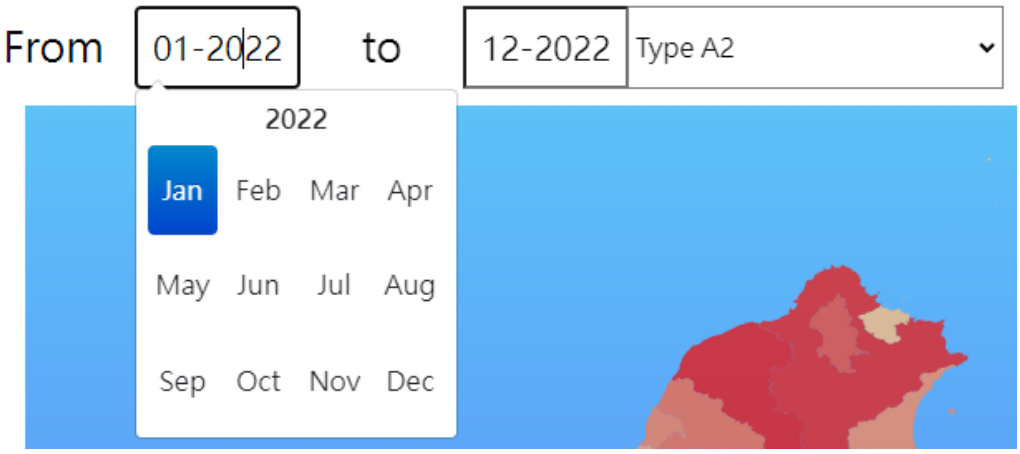


Figure 8: Control Panel

Insights

Temporal & Spatial Trend

By selecting both A1 and A2 data, we can first find out the city that has the high fatality and injury count in 2022 is New Taipei City, TaoYuan City, Taichung City, and Tainan City in Figure 9. each of them has more that 58000 injury count and 160 death count. Comparing the casualty count between different cities in Taiwan, we can find out that the count in eastern Taiwan is much lower than the other cities, which we can infer is because the traffic density is lower in eastern Taiwan.

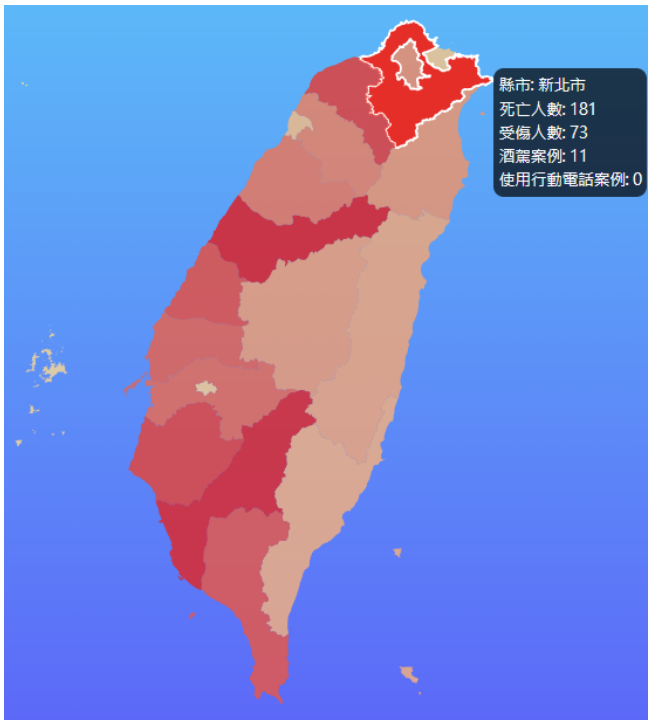
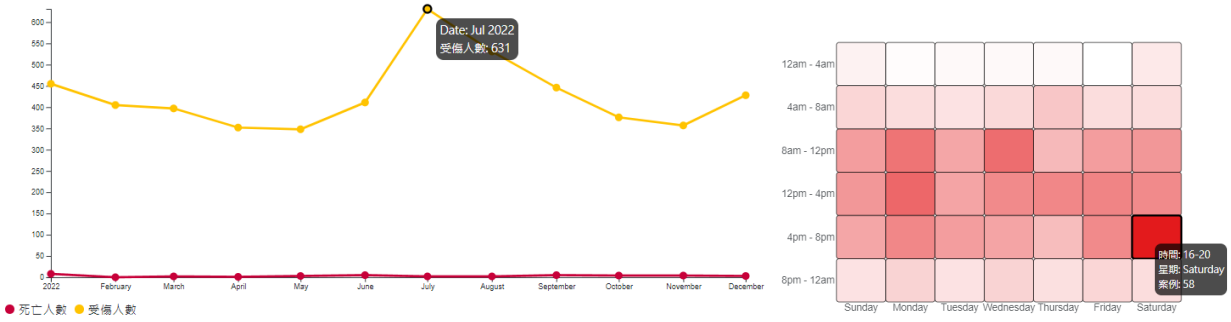


Figure 9: Accident Trend in Different City

We can further find out the accident trend in eastern / western Taiwan. By selecting eastern cities such as Taitung, we can find out that the peak of the accident count is in july and August from Figure

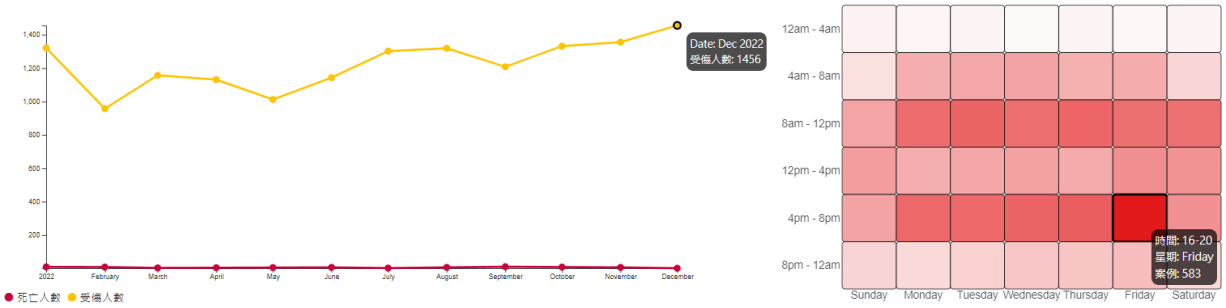
10a, and the accident count on weekend is abnormaly high shown in Figure 10b. We infer that the reason is there are many tourists in eastern Taiwan during summer vacation, and the traffic density will increase during the weekend, which causes more accident than other time period.



(a) Accident Trend in Taitung (b) Accident Count in Taitung

Figure 10: Accident Trend in Taitung

For western cities such as Hsinchu, the trend can be found in Figure 11. We can find out that the casualty count becomes higher in winter from Figure 11a, this may cause by the early sunset and rainy weather in winter. From Figure 11b, we can notice that the accident count is higher in the 8:00-12:00 and 16:00-20:00, which is the time period that people go to work and go home, so the traffic density is higher than other time period, which causes more accident. Furthermore, the accident count on Friday in period of 16:00-20:00 is higher than other weekday, we think the reason is that people may go to other city for travel or go back to their hometown during the weekend, so the traffic density is higher than other weekday.



(a) Accident Trend in Hsinchu (b) Accident Count in Hsinchu

Figure 11: Accident Trend in Hsinchu

These observation can help the government in different cities to improve the traffic safety by employing more traffic police in the time period or place that has high accident count. Traffic education can also be promoted to the drivers and pedestrians to reduce the accident count in the future.

Vehicle Type & Driver-Contributed Cause

Regarding the proportion of each vehicle type in the causes of accidents throughout the entire year's whole data, we can find out some interesting insights in Figure 12.

Motorcycles occupy a larger proportion in all major causes of accidents

Since motorcycles have the highest number of accidents among all vehicle types, it is reasonable that they also have a larger proportion in all major causes of accidents.

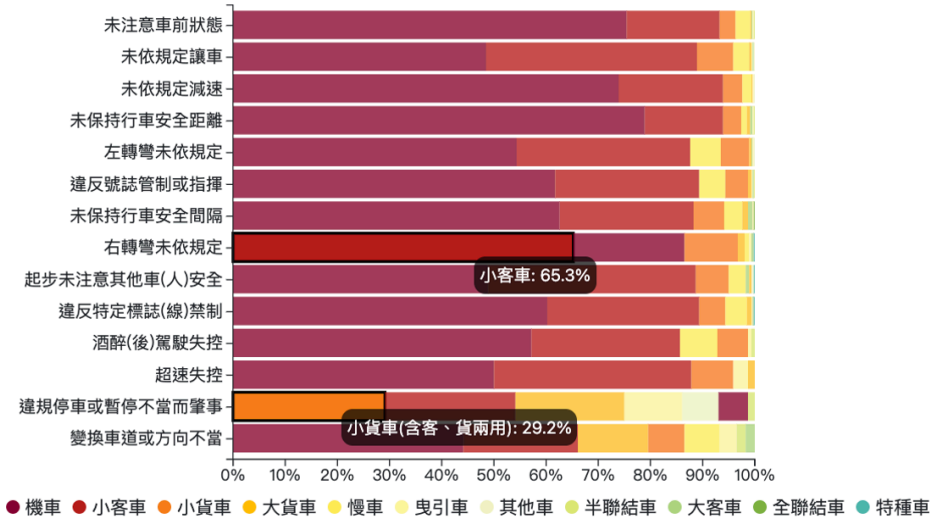


Figure 12: Proportion of Each Vehicle Type in the Causes of Accidents

Compared to other vehicle types, small cars are more likely to have accidents due to ”improper right turns,” while motorcycles still have a higher proportion for ”improper left turns”

In the cause of ”improper right turns,” small cars have a much higher percentage of accidents than motorcycles (accounting for 65.3% of the total number of accidents caused by this reason). One possible explanation is that small cars have a blind spot when turning right, especially for vehicles coming from the right rear, making it easier to overlook potential dangers. However, motorcycles have a higher proportion for ”improper left turns,” this may be due to many riders ignoring the ”2 stage left turn” signals that are designated for motorcycles at most intersections, resulting in a high number of accidents. From this analysis, it is important to emphasize that small car drivers need to be particularly aware of pedestrians and vehicles on the right, especially when turning right, to prevent collisions due to blind spots. It is also crucial for motorcycle riders to be more aware and compliant with the ”2 stage left turn” rules.

Small trucks have a higher proportion of accidents due to “illegal parking or improper temporary stopping” compared to other vehicle types

There are two main reasons for this: first, small trucks often need to load and unload goods on streets without dedicated areas, forcing drivers to park or stop in inappropriate places; second, the larger size of small trucks compared to other vehicles means that even temporary improper parking can cause more significant traffic blockages or visibility issues, increasing the risk of accidents. Understanding why small trucks have a higher proportion of accidents due to ”illegal parking or improper stopping” can help develop targeted traffic management strategies and driver training to reduce the occurrence of such accidents. Measures might include providing more parking spaces for trucks, enhancing driver training, and optimizing delivery schedules.

Crash Position Distribution of Vehicle

Figure 13 shows the crash position distribution of vehicle with all data in 2022.
In the dataset, the concentration of impact locations on the front and sides of vehicles suggests a correlation with the causes of accidents. For instance, frontal collisions, which are most prevalent (indicated by the deep red color in the chart), are likely related to ”not paying attention to the condition of the car in front.” This could be due to drivers being distracted, leading to rear-end collisions.
Side impacts, on the other hand, could be associated with causes such as ”improper left or right turns,” ”improper lane changes or directional changes,” or ”failing to yield as required.” These types of accidents might occur when a vehicle is turning and either misjudges the distance to an

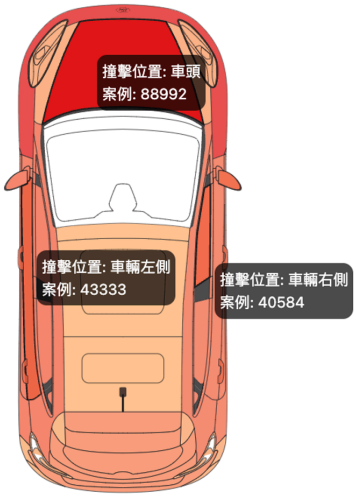


Figure 13: Crash Position Distribution of Vehicle in 2022

approaching vehicle or does not notice it. Improper lane changes can lead to side collisions if drivers do not properly check their blind spots or misjudge the speed and distance of vehicles in adjacent lanes.

Additionally, the distribution pattern of impacts could be influenced by traffic flow and road design. Intersections, for instance, are common sites for both frontal and side impacts due to the crossing paths of vehicles and the complexity of driving maneuvers involved.

Understanding the relationship between impact locations and accident causes can be crucial for developing targeted safety measures and driver education programs. For example, enhancing awareness about maintaining safe distances, proper lane-changing techniques, and the importance of attentiveness at intersections can help reduce the occurrence of both frontal and side collisions.

Challenge and Future Work

During the preprocessing of the dataset and the selection of features, many potentially valuable pieces of information were discarded due to missing data, imbalanced distributions, or the inability to effectively compare and integrate certain features with others. Additionally, the multifaceted nature of traffic accidents, influenced by factors like temperature, holidays, road conditions, and human behavior, makes it challenging to definitively identify specific causes or predict patterns based solely on our dataset.

This difficulty in data analysis highlights the complexity of traffic safety as a subject. Traffic incidents are not isolated events but are the result of a confluence of various factors, some of which may be outside the scope of the available data. For future work, it would be beneficial to integrate additional datasets that include more comprehensive environmental, temporal, and behavioral factors. This could provide a more holistic view of the circumstances leading to traffic accidents and aid in developing more effective prevention strategies.

Moreover, the challenge of dealing with incomplete or unbalanced data underscores the need for more robust data collection methods in the field of traffic safety. Improving data quality and ensuring a more representative sample would significantly enhance the reliability of the analysis. It also opens up possibilities for applying advanced data analysis techniques, such as machine learning, to uncover deeper insights and more accurately predict accident risks under various conditions.

Conclusion

The comprehensive analysis of Taiwan's 2022 traffic accident data has illuminated the intricate and multifaceted nature of traffic incidents, revealing how they are influenced by regional dynamics, temporal trends, and specific vehicle-related factors. This study has provided valuable insights, addressing the questions raised in our objectives. We observed notably higher accident rates in major

urban areas compared to Eastern Taiwan, discerned seasonal variations in accident occurrences, and identified a disproportionate involvement of motorcycles and small cars in certain accident types. These findings answer the questions initially posed, offering a clearer understanding of the variables affecting road safety. The insights derived from this study are invaluable for the development of targeted traffic management strategies and the formulation of effective driver education programs.