**Task Description**

**Background**

Recently, there was an accident at the end of the South Eastern Freeway on Wednesday 15th May, where a truck travelling towards Adelaide crashed into the gateway fountain wall (ABC News, 2024). This crash resembles an accident which happened at the same place 10 years ago in January 2014 where a truck driver died by slamming his truck into the fountain wall and there is a long history of crashes at this intersection. These accidents show that there is less awareness among the public about these accident-prone zones. So implementing accident-prone zone signs could inform drivers of the high risk and encourage them to choose alternative routes (Aguilera, Moysés & Moysés, 2014).

**Initial Questions**

Which intersections and roads in Australia are the most accident-prone and What are the common factors contributing to accidents in these areas?

Identifying these high-risk locations and the factors behind them can help in concentrating effort and money by officials for implementing warning signs, and other safety measures to reduce accidents. It also helps the public to avoid these intersections to evade the danger.

**Dataset Description**

So to answer these questions the information about the road crashes is collected from the five most populous states of Australia i.e., NSW, VIC, QL, SA, WA (Australian Bureau of Statistics, 2023). The SA Road Crash Dataset has 63,601 records with 34 columns collected over the years 2018 to 2022 (Government of South Australia, 2023). The WA Road Crash Data includes 127,173 records with 25 columns over the time period 2019 to 2023 (Main Roads Western Australia, 2023). The VIC Road Crash Data contains 168,360 records in a GeoJSON format with 50 columns over the period 2012 to 2023 (Victoria State Government, 2023). The NSW Road Crash Data includes 95,873 records with 50 columns over the period 2018 to 2022, (Transport for NSW, 2023). Lastly, the QLD Road Crash Data has 386,837 records with 53 columns from 2001 to 2023 (Queensland Government, 2023). So to avoid discrepancies in the dataflow and to maintain consistency between the various states datasets, Road crash data between the years 2019 to 2022 is chosen.

**Data processing**

The data collected from five major Australian states—South Australia, Western Australia, Victoria, New South Wales, and Queensland had fields which are common in all of them but with different notations and terms. Out of which we extracted the fields which are required for our model.

**Data Extraction and Standardization**

**Column Name (ID):** All the five datasets had an unique identifier for the crash record. In the final dataset all of these identifiers are grouped together under the field(ID). This field is used for backtracking the values to their original data set.

**Date, Time, Day**: In each dataset date, time, day was spread across different columns and formats. This feature is required for accurate temporal analysis. This date and time information were Standardized and formatted into ISO 8601 (YYYY-MM-DD) for dates and 24-hour format (HH-MM) for times, to maintain consistency.

**Statistical Area:** I categorized geographic areas into three types: City, Metropolitan, and Country for Spatial Analysis. South Australia, New South Wales, Victoria, and Queensland had corresponding attributes for these classifications, while Western Australia did not provide specific data for this categorization. I used another data set Special-Use from WA Government (Main Roads Western Australia, 2023) and connected this with the original data set to obtain the required statistical records.

**Latitude and Longitude:** This feature is essential for mapping accident hotspots and conducting geospatial analyses. This geospatial data helps us to create detailed maps that highlight high-risk areas, aiding in the visual presentation of our findings. In the road crash datasets NSW, VIC, WA and QLD provided specific latitude and longitude attributes South Australia had no latitude and longitude data so I used another dataset Road Crashes .geojson file (Government of South Australia, 2023) and used it to get the required records in the Latitude and longitude fields.

**Local Government Area (LGA) and Speed Limit:** This feature helps us to find the impact of local governance/councils and speed regulations on accident rates. This information is crucial for recommending specific interventions, such as altering speed limits or improving local infrastructure, to enhance road safety. In the 5 datasets SA, NSW, VIC, and QLD provided corresponding attributes, while WA did not provide any data for (LGA) and Speed limit. So I used Legal Speed Limit Data (Main Roads Western Australia, 2023) to get the missing records in the LGA and speed limit field.

**Road Name, Intersection Name, Intersection Type:** Understanding the specific locations and types of intersections (e.g., intersections, midblocks) where accidents occur most frequently allows us to pinpoint exact areas for targeted interventions. This detail is crucial for proposing the installation of warning signs and other safety measures at high-risk intersections. Although road and intersection names were not directly available in some datasets, I used complex rules to extract these feature details for NSW, WA, VIC, QLD but there was no data for SA. So, I used the Google Maps Geocoding API to get the required records for the final dataset.

**Light Conditions and Weather:** Lighting conditions were standardized to Day or Night across all datasets. For South Australia, New South Wales, Victoria, and Queensland, various fields were mapped to these common categories, while in Western Australia, lighting conditions were deduced from the date and time of the crashes. Similarly, weather conditions were classified as Raining or Not Raining, with fields from SA, NSW, VIC, and QLD being mapped to these categories, and for WA the required details are deduced from DPIRD Weather Stations dataset(Department of Primary Industries and Regional Development 2023). This standardization ensures consistency in analyzing environmental factors affecting road crashes.

**Crash Type:** Crash types were available in all the five data sets they were standardized using the Definition for classifying accidents (DCA)

**Crash Severity:** Severity of crashes was categorized uniformly across datasets into three levels: serious injury, minor injury, and property damage only. All the five datasets had the required fields they were grouped together into five categories.

**Refined Problem**

How can we identify and rate accident-prone zones for roads and intersections in Australia to enhance traffic safety?

The next steps involve conducting exploratory data analysis (EDA) to visualize accident frequency and severity, and analyse their distribution over time and various conditions. Model development will include building classification models to predict accident severity based on factors such as location, weather, and road conditions, and developing a scoring system to rate intersections and midblocks from 1 to 4 based on their accident-prone nature.

**Back-up Plan**

How do the number and type of traffic signs relate to road crash patterns in South Australia?

To address this question we have to analyse and compare the SA Road Crash Data (Government of South Australia 2023) from 2018 to 2022 with detailed information about crashes, the Traffic Volume Data (Government of South Australia, 2022) which provides the density of traffic across different roads and the Traffic Signals Data (Government of South Australia 2022) which gives us the details the locations and types of traffic signals at the intersections. These datasets can be used to identify correlations between density of traffic signs, traffic volume and their relation with crash rate. This will help us in proposing improvements to road safety by understanding the impact of signage and traffic signals on crash patterns.

**References**

ABC News 2024, Truck crashes into fountain wall on South Eastern Freeway, viewed 9 June 2024, <https://www.abc.net.au/news/2024-05-15/adelaide-south-eastern-freeway-truck-crash/103848622?utm_campaign=abc_news_web&utm_content=link&utm_medium=content_shared&utm_source=abc_news_web>.

Aguilera, SLVU, Moysés, ST & Moysés, SJ 2014, ‘Road safety measures and their effects on traffic injuries: a systematic review’, Revista panamericana de salud pública, vol. 36, no. 4, pp. 257–265.

Australian Bureau of Statistics 2023, National, state and territory population, viewed 9 June 2024, <https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/sep-2023#states-and-territories>.

Department of Primary Industries and Regional Development 2023, DPIRD Weather Stations, viewed 9 June 2024, <https://catalogue.data.wa.gov.au/dataset/dpird-weather-stations>.

Government of South Australia 2022, Traffic Signals, viewed 9 June 2024, <https://data.sa.gov.au/data/dataset/traffic-signals>.

Government of South Australia 2022, Traffic Volumes, viewed 9 June 2024, <https://data.sa.gov.au/data/dataset/traffic-volumes>.x

Government of South Australia 2023, Road Crashes in SA, viewed 9 June 2024, <https://data.sa.gov.au/data/dataset/road-crashes-in-sa/resource/6e2bb999-cc85-41a5-9ef2-3a892d1ffe81>.

Government of South Australia 2023, SA Road Crash Data, viewed 9 June 2024, <https://data.sa.gov.au/data/dataset/road-crash-data>.

Main Roads Western Australia 2023, Crash Information Last 5 Years, viewed 9 June 2024, <https://portal-mainroads.opendata.arcgis.com/datasets/mainroads::crash-information-last-5-years>.

Main Roads Western Australia 2023, Legal-Speed-Limit Data, viewed 9 June 2024, <https://portal-mainroads.opendata.arcgis.com/datasets/c30239d960df4eaf85890201068ea521_8/explore>.

Main Roads Western Australia 2023, Special Use Data, viewed 9 June 2024, <https://portal-mainroads.opendata.arcgis.com/datasets/f61dc8fbb1bd45258dbbcb6f1162d73c_31/explore>.

Queensland Government 2023, Crash Data from Queensland Roads, viewed 9 June 2024, <https://www.data.qld.gov.au/dataset/crash-data-from-queensland-roads>.

Transport for NSW 2023, NSW Crash Data, viewed 9 June 2024, <https://opendata.transport.nsw.gov.au/dataset/nsw-crash-data>.

Victoria State Government 2023, Crash Stats Data Extract, viewed 9 June 2024, <https://discover.data.vic.gov.au/dataset/crash-stats-data-extract>.