TECHNICAL REPORT

PROJECT 1

UNIT: CITS5504

*Date: 08 April 2025*

Student 1: Cedrus Dang - ID: 24190901

Student 2: Cedrus Dang - ID: 24190901

1. **Introduction**

This report details the design, implementation, and usage of the data warehouse to store and analyze the Australian road crash fatality data in 2023, collected by Australia Bureau of Statistics.

In order to help the government and public understand the importance of road safety and lower traffic risks, the project focus on the key objectives as follow:

1. Build a data warehouse to store historical data on fatal crashes.
2. Use the created data warehouse to present key insights with a dashboard.
3. Support decision-making with data mining techniques.
4. List and describe a few suggestions for improving road safety for the government.

The tech stack that is implemented into this project are:   
Jupyter python notebook for EDA pre-developement, ETL and Data Mining.

PostgresSQL for Data base server with SQL as query language.

Tableau/PowerBI data analysing.

Library for ETL is pandas and for Data mining is ….. (!later)

1. **Design, implementation, and usage of the data warehouse:**

The data warehouse conceptual model and dimensions are designed as follows:

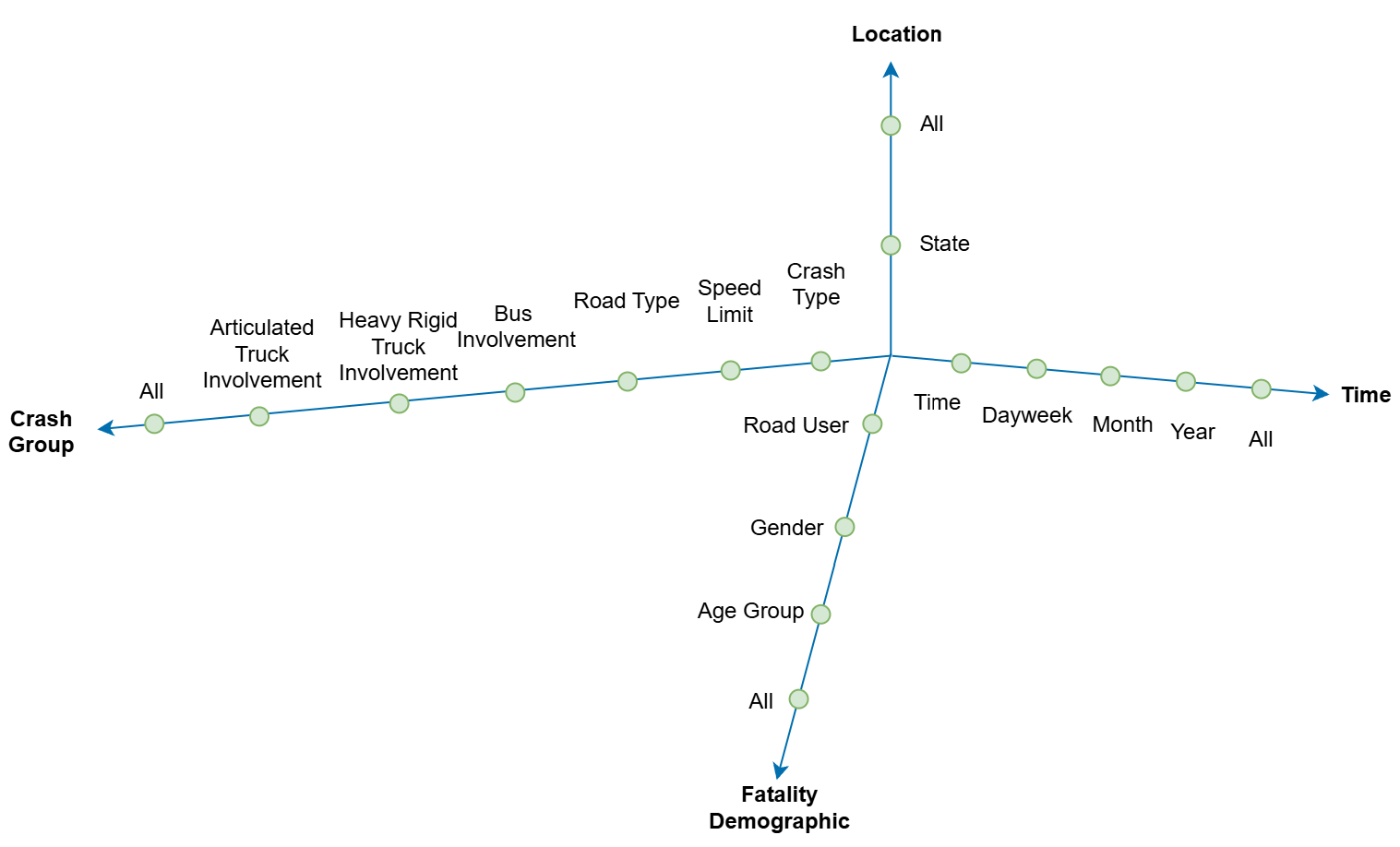


Diagram 2: Starnet Conceptual Model Diagram

The design’s dimensions were chosen by grouping attributes in fatality cases, following its topic, that can answer high-level questions as listed:

1. What are the fatalities in that period?

2. What are the fatalities in that location?

3. What are the fatalities with the same crash case scenario?

4. What are the fatalities with the same demographic attributes?

Some location dimensions that can gain high-value insights but are unstable due to a high percentage of missing data (> 30%) and can be replaced with state dimensions have been dropped after the exploratory stage to save resources and stabilize the data warehouse structure, such as LGA code, RA code, etc. This ensures the balance between maintaining and richness required for OLAP analysis.

Add that you build rules from Kimball -> Lecs but now need for reference :P

For the logical model of the data warehouse, the surrogate keys are used to replace all natural id to create stability and effectiveness. Population of state in 2023 also been added into the dim\_location to provide data for the OLAP.

The data warehouse logical model presented in the below diagram:

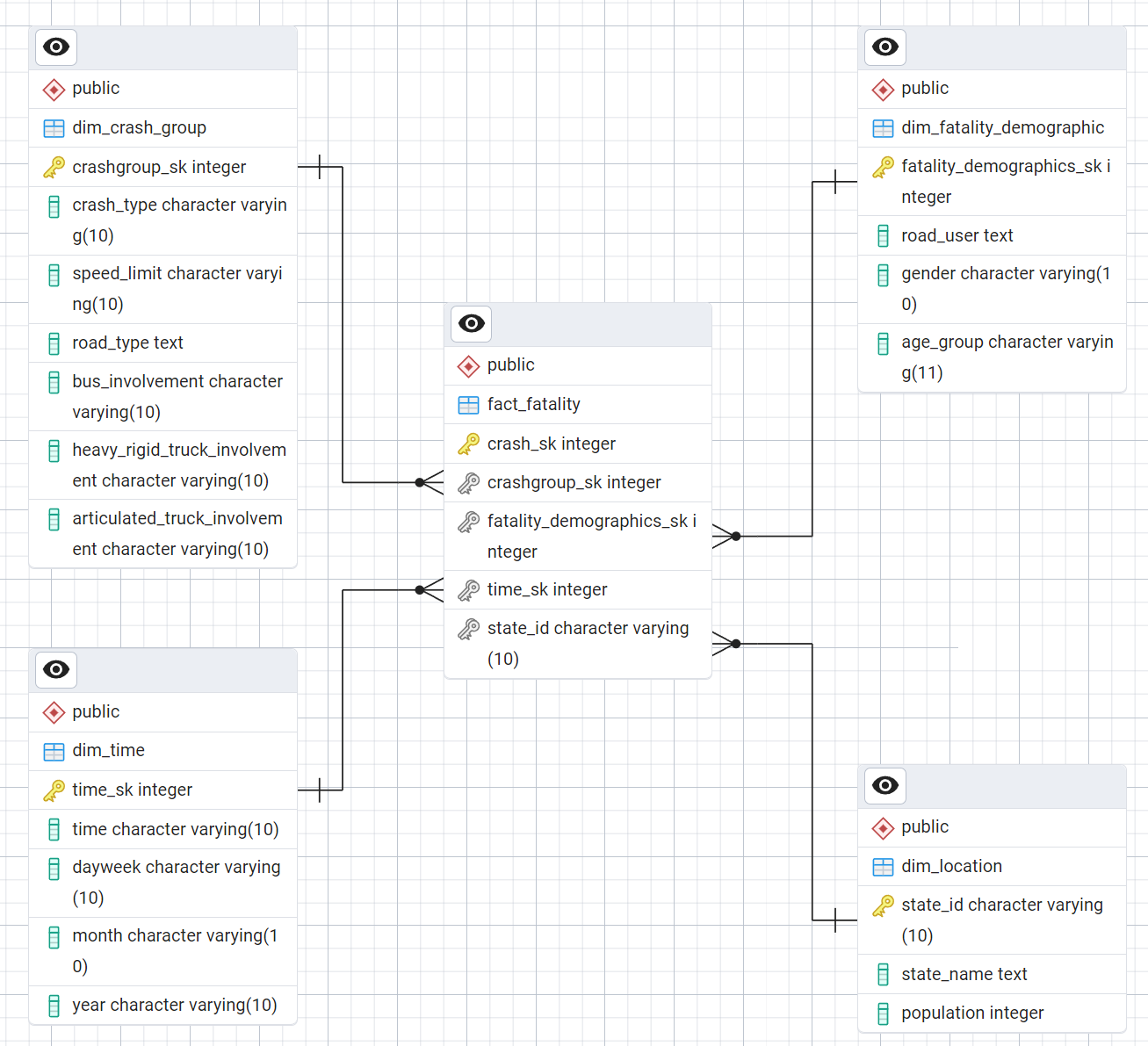


Diagram 2: Star Schema Diagram

One ETL already done

1. **Schema, Starnet and query footprints:**
2. **Data cleaning, preprocessing, and ETL process:**
3. **Visualisation of query results**
4. **Association rules of data mining**