School of Science, Computing and Engineering Technologies

COS30045

LAB 4.1 Design Studio

Overview

In this lab you will be given a sample data set and asked to identify the different data and attribute types. You will also think about some questions about this data set that might be answered by a visualisation.

ardd_fatalities_Jan2020_0.xlsx (download from Canvas)

Download and review this data set before attempting this exercise.

1 Interpreting the data set

Complete the LAB 4.1 Quiz.

2 Visualisation Design

Think of three questions you would like to answer with that require a data visualistion.

For each data question you will need to consider the following:

Which data attributes (columns) do you need to answer this question?

Do you need to transform any of the data?

Does the data type change when you transform the data? If so how.

Make a sketch of how you think your visualisation might look and add to this document.

Question 1 - What is the distribution of fatalities by age group?

-Data Attributes Needed: Age group, Crash ID.

-Data Transformation: Grouping the data by age group.

-Potential Data Type Change: No change in data type because there are predefined age groups (categorical), otherwise may need to aggregate data if using raw age values


Distribution of Fatalities by Age Group
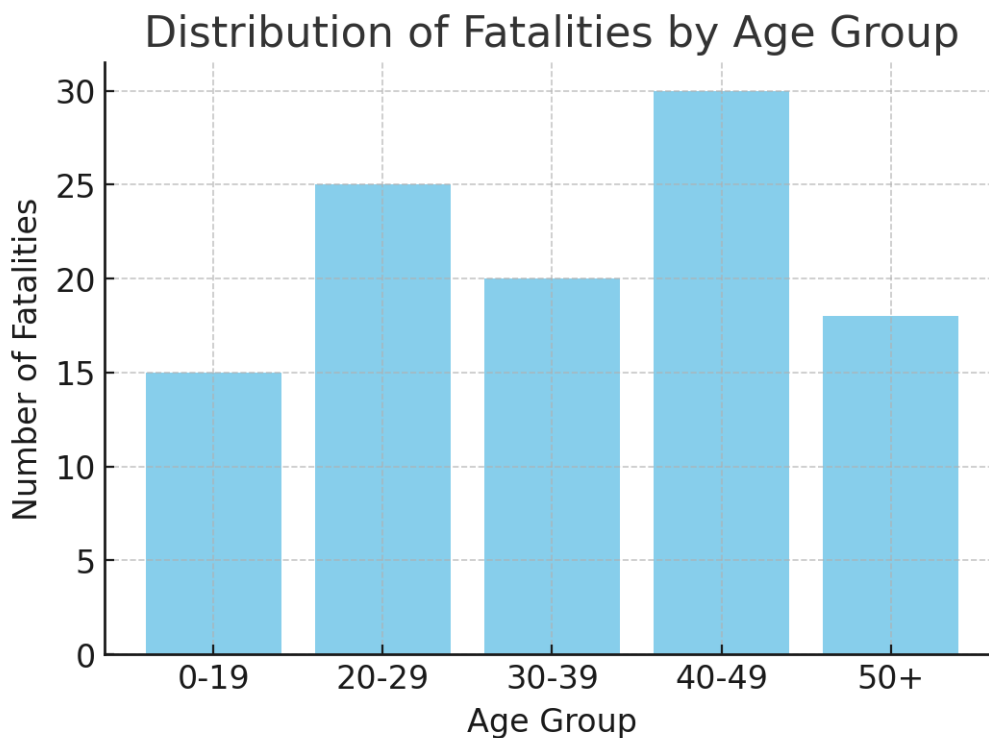
Question 2 - How do fatalities vary by time of day across different states?

Data Attributes: Columns to use: Time, State, Crash ID

Transformation: The Time column can be grouped into hourly ranges (e.g., morning, afternoon, evening, night). The data needs to be filtered by State and grouped by Time.

Data Type Transformation: Time can be transformed into a categorical type representing time periods, while Crash ID is counted per state and per time.

Visualization Sketch: Stacked bar chart
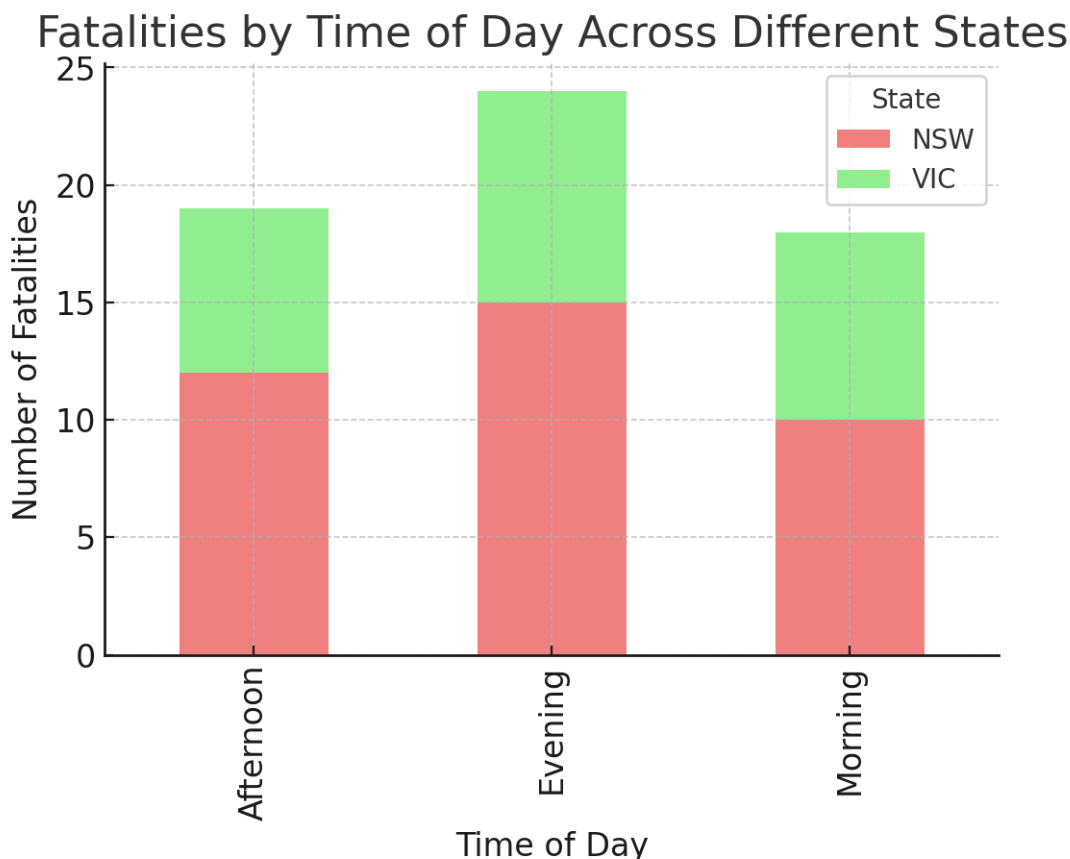
Question 3 - What is the correlation between speed limit and crash fatalities involving heavy rigid trucks?

Data Attributes: Columns to use: Speed Limit, Heavy Rigid Truck Involvement, Crash ID

Transformation: Filter the data where Heavy Rigid Truck Involvement is marked as "Yes". Group the data by Speed Limit and count the number of crashes.

Data Type Transformation: Speed Limit would remain a continuous numeric variable, but the crash count (from Crash ID) would be aggregated into bins of speed limits (e.g., ranges of 10-20 km/h).

Visualization Sketch: Scatter plot or line chart

Correlation between Speed Limit and Fatalities Involving Heavy Rigid Trucks
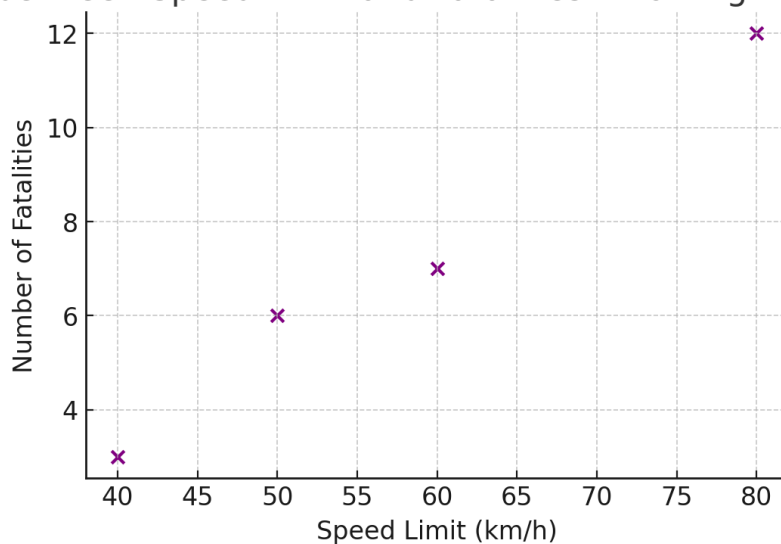
Include this file as evidence for your Demonstration 2

LAB 3.1

LAB 3.2