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.

1. **How does crash frequency vary across different regions of Australia throughout the year?**

   Crash Frequency Across Australian Regions

   Visualization Type: Line Chart

   Data Attributes Needed:

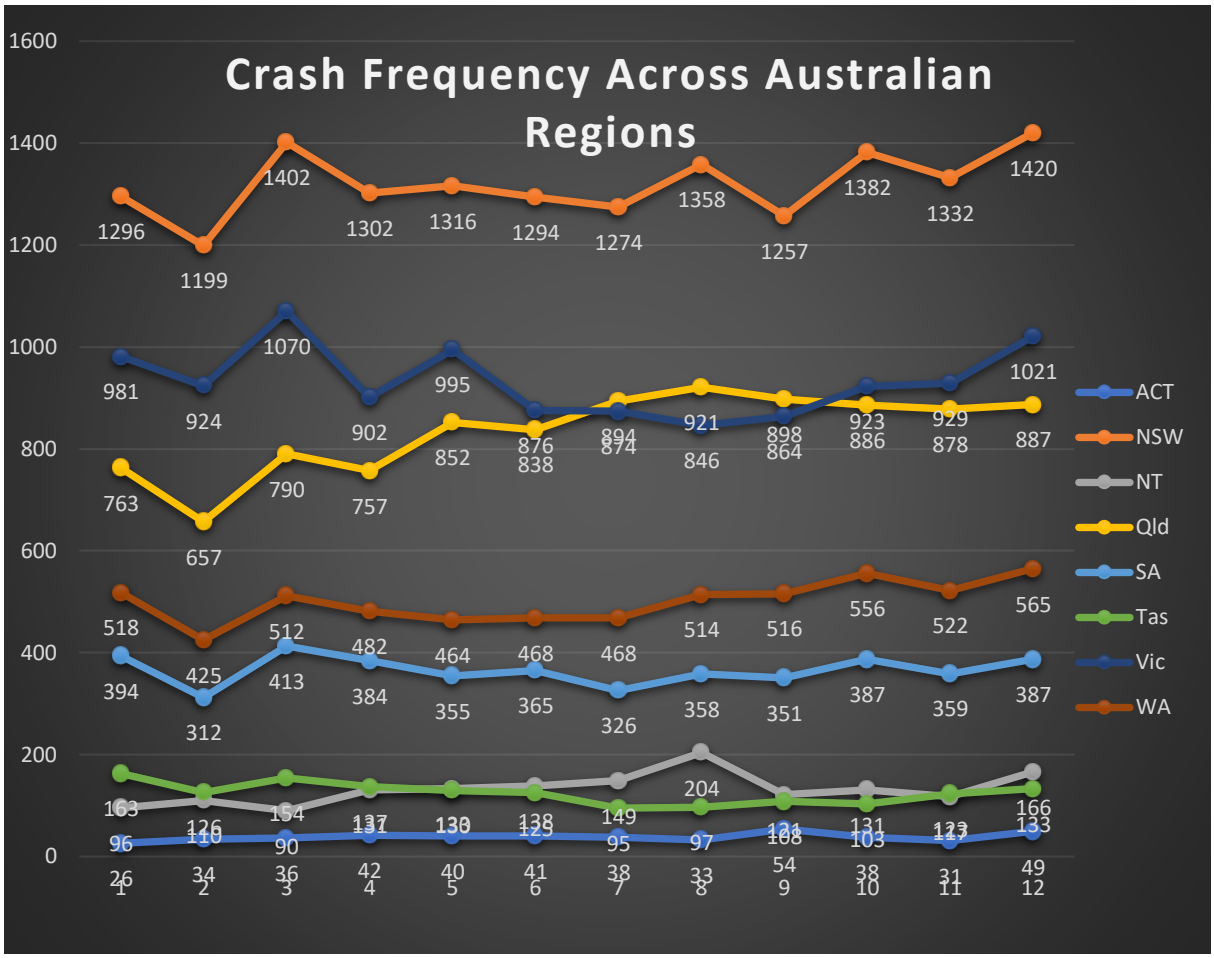
   - State: Represents the region or state in Australia where the crashes occurred (e.g., NSW, Qld, SA).
   - Month: Represents the month of the year in which the crashes occurred, used to analyze trends across the year.
   - Crash Count: Represents the total number of crashes that occurred in each state per month.

   Data Transformation Needed:

   - Grouping and Summing: The data needs to be grouped by both State and Month. For each state and month, you will sum the total number of crashes to match the line graph format.
   - Pivoting Data: The data needs to be pivoted so that each state becomes a separate line in the graph, with months on the x-axis and crash counts on the y-axis.

   Data Type Changes Due to Transformation:

- Crash Count: After grouping, individual crash records are aggregated into totals for each month and state, so they remain as numerical data.
- State and Month: These remain categorical (nominal) data types before and after the transformation. No changes are needed.
- The line graph shows crash frequency trends across Australian states over 12 months.



Visualization Insight:

- **NSW** has the highest crash rates, peaking in February (1402) and December (1420).
- **ACT** shows a steady rise, with a sharp increase in December (1021).
- **Qld** remains stable with slight peaks in March and December.
- **WA** gradually increases, ending with the second-highest count in December (565).
- **Tas** and **NT** have low and stable crash rates.
- **Vic** and **SA** show moderate fluctuations.
- **Seasonal Trends**: Several states see a rise in crashes toward the year's end, possibly due to holidays or increased travel.

2. **How does the distribution of crash types vary across different times of day?**

Distribution of Crash Types Across Different Times of Day

Visualization Type: Bar Chart

Data Attributes Needed:

- Time of day (categorized as Day/Night)
- Crash Type (Single, Multiple, Pedestrian)
- Crash ID (to count occurrences)

Data Transformation Needed:

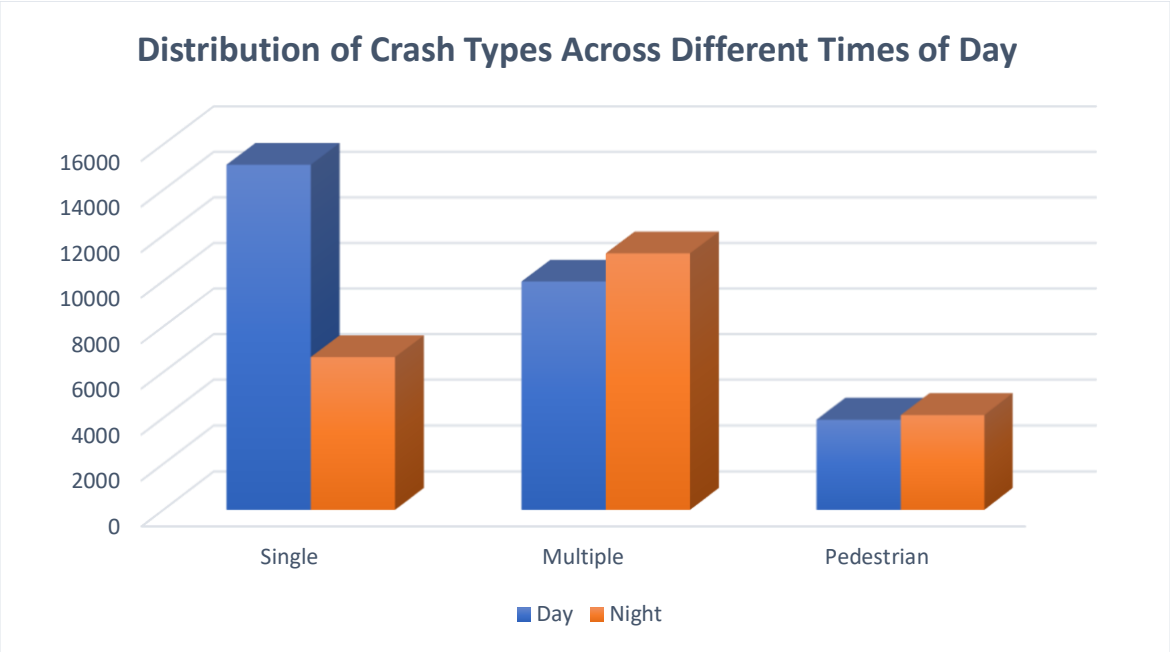
- Categorize Time into "Day" (06:00-17:59) and "Night" (18:00-05:59).
- Count the number of occurrences for each Crash Type (Single, Multiple, Pedestrian) within these time categories.

Original Data Type:

- Time is quantitative (interval) in HH format.
- Crash Type is categorical (nominal).
- Crash ID is categorical (nominal), serving as a unique identifier.

Transformed Data Type:

- Time becomes a categorical variable (nominal) with values "Day" and "Night".
- Crash Type remains categorical (nominal).
- Crash Count is a new numeric variable (quantitative) representing the number of crashes for each type in the given time categories.



Visualization Insight:

- The bar chart shows that **Single** crash types occur significantly more during the **Day** compared to the **Night**.
- For **Multiple** crash types, the occurrences are almost equal for both **Day** and **Night**.
- **Pedestrian** crashes show similar frequencies during both **Day** and **Night**, but with lower overall numbers compared to Single and Multiple crashes.

3. **What is the gender distribution of road users involved in crashes?**

Gender Distribution in Road User Crashes

Visualization Type: Pie Chart

Data Attributes Needed:

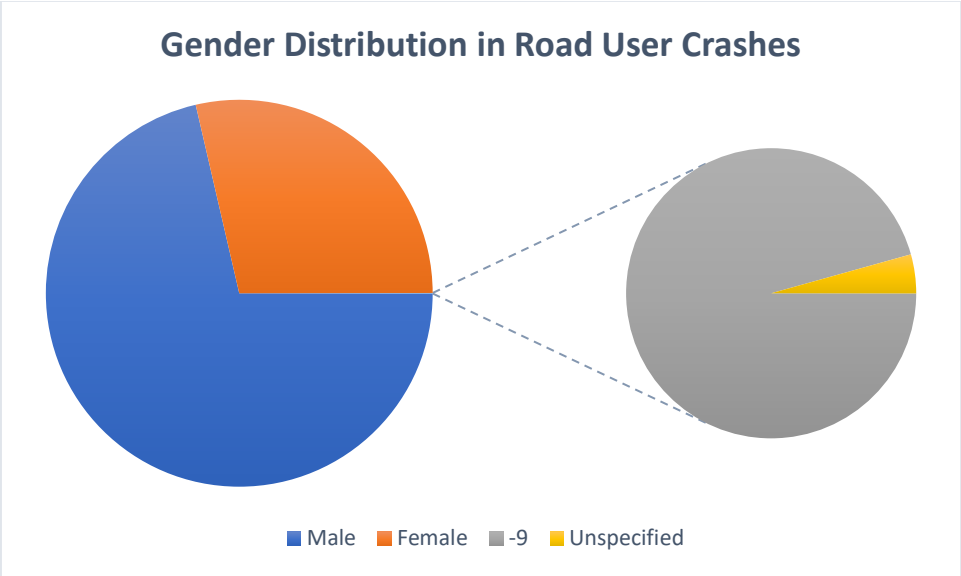
- **Gender:** Represents the gender of the road users involved in crashes.
- **Road User:** Represents the type of road user (e.g., driver, passenger, pedestrian).

Data Transformation Needed:

- No transformation is needed because the data in the Gender and Road User columns is already in a suitable format for analysis.

Data Type Changes Due to Transformation:

- There are no changes in data type as no transformation is required.
- Both Gender and Road User remain categorical (nominal) data types before and after the analysis.



Visualization Insight:

- The pie chart visually illustrates the distribution of road user crashes by gender.
- It helps identify which gender group is more frequently involved in road user crashes.
- It reveals the proportion of crashes where gender data is missing or unspecified.
- Provides a clear overview of gender-based patterns in road user incidents.