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.

What are the most common road type of car accidents?

The data attributes that is required are “National road type”, because the government area can be used to differentiate the area. The data was originally in nominal, but then change to ordinal data. The data was originally in word format, but was modify to two column (National Road type & the field count). The new table display the count of each road types.

The area where accident occurs the most is at National or State Highway which takes up almost 30% of the pie chart. The second comes from Arterial Road, which also takes up 28% of the pie chart. The third most accident occurs road type is Local road which is 20% of the pie chart. The total of the top 3 road types where accident tends to occurs the most occupied 78% of the pie chart. The other 22% are made up from other 5 road types.

Does Christman or Easter tends to have car accident?

The data I need is Christmas period and Easter period. I need to transform the yes and no data for the periods into the count of yes and no into another table. After transforming the data into the count of yes and no table.

As you can see, The orange bar represents the Easter Period and the blue bar represents Christmas Period. As you can see the accident happen in Christmas is more than accident in Easter. Accident happen at Christmas is at least 6 times of accident happen in Easter. This result shows that during Christmas Period, accident will occurs more than Easter Period.

What is the age range that tends to cause accident?

I need to extract column Age Group. It is a categorical data, but after transformation of data it change to ordinal data. The new table shows the count for the categories

The above pie chart show the percentage age group that is in the car accident. The total of car accident is 50917. The age group 17 to 25 consist 26%, which is the most percentage in the pie chart. Seconds most percentage is 25%, for age 40 to 64. The category are not synchronized, so some data have a bigger range compare to others. The age group 40 to 64 have a much bigger range compare to 17 to 25, this prove that the age group 15 to 25 did have many accident. The third most percentage is age group 26 to 39, which contain 23%. This age group(26 to 39) have a lesser range compares to age group 40 to 64, which might be why age group 26 to 39 is lesser than age group 26 to 39. The top three of the pie chart already taken 74% of the graph. The other three age group are 0 to 16 - 8% , 65 to 74 - 8%, and 75 or older – 10%. The age group 75 or older maybe be cause by the oldness of the driver, and therefore the accident is quite high.