**Common problems encountered**

1. Finding the right topic with the relevant data that is suitable or appropriate for machine learning purposes.

1. Once topic is found but data is all in summarised format like below and they are not related with other categories, e.g. Number of males who committed Domestic Violence related assault were not provided.

Graphical user interface, text, application, email

Description automatically generated

1. Luckily there is a web site <http://www.courtdata.com.au> where you can register and for a small fee, user can collect over 19.3 million criminal and 3 million civil records.

***Technical issues encountered***

1. ***URI = os.environ.get(‘DATABASE\_URL’, ‘’)***

*which must be replaced by:*

***URI = os.environ.get(‘DATABASE\_URL’).replace(“://“, “ql://“, 1)***

1. Not all tables are in tabular formats and some cells are in merged format, need to split them and transpose them individually.
2. The command **pip freeze > requirement.txt** has generated a lot of packages in the file and not a lot of them were meaningful to Heroku during deployment. Trial and error deployment processes which I spent a lot of time on.
3. When returning the json object to javascript, it must have the orient parameter set to ‘records’

Text

Description automatically generated

In order for the d3.select function to populate the drop down successfully:  
A screenshot of a computer

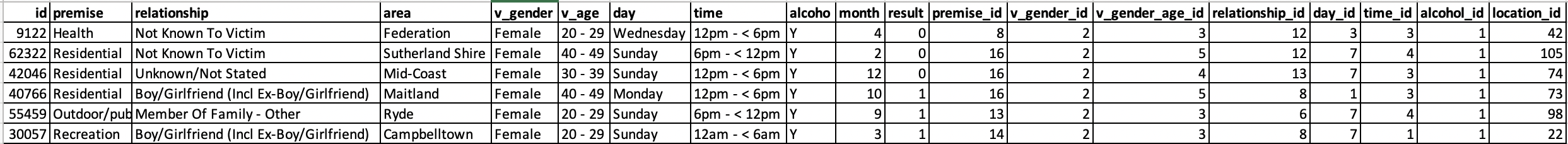
Description automatically generated with medium confidence

**Machine Learning discussion**

Here is the structure of the prediction table:  
Timeline

Description automatically generated with low confidence

Here is some data from the table:



Each of the category field has its own table and key value pair, e.g. premise:

Table

Description automatically generated

So we have in total 10 features for the domestic violence prediction namely:

Premise, relationship, area, victim gender, victim age, day, time range, alcohol, month and location.

Loaded the table to data frame.

A screen shot of a computer

Description automatically generated with low confidence

Import libraries and drop all the categorical columns from the data frame.

Text

Description automatically generated

Drop the result column and scale and train the data.Graphical user interface, text, application

Description automatically generated

Find out which model provides the highest degree of accuracy and SVM is the winner.

Text

Description automatically generated

Evaluate the prediction and produce the confusion matrix and classification report.

A picture containing text

Description automatically generated

Save the model and scaler for later use.

Text

Description automatically generated

Load the saved model and scaler files and make up two sample predictions Text

Description automatically generated

Prediction results for prediction 1 and 2.

Text

Description automatically generated

Conclusion:

- Should have included other states

- Should have made it a bit prettier

- Educate your loved ones or children so they won’t be either the victims or perpetrators of domestic violence crimes.