CA Summary

Overview description of your CA and what you are trying to do. What are your main objectives and what are you hoping to discover?

Background

Why and how you came up with this idea. Why do you have an interest doing data collection, analysis and visualisation in this area? Has this or a similar data visualisations / analysis been done before in this area? Etc.

Data sets

List the datasets you have acquired and where you got them from

Seven stages

Document the tasks you completed at each stage of the seven stages for each data set. Some of these stages are done on the data itself and some are done in Tableau. Document both and include a screenshot to illustrate your point.

Problems and solutions

What problems did you encounter and what solutions did you implement to resolve these problems?

Did you create any visualisations that did not work and why? Did you have to change the visualisation chart type to another, so as to better deliver your message (e.g. bar chart to bubble chart etc.). Include these visualisations that did not work in your documentation also.

Conclusion

Final conclusions and journey of your CA. This should be the most interesting section in your document that outlines the discoveries you have made after visualising and interacting with your data.

Specifically mention what you discovered and how your visualisation brings these findings to light. What makes your visualisations unique and clearly highlights this new information.

What story does your visualisation tell? What pattern / trend or item of discovery do you now see that you could not have seen before when just reviewing the raw data?

What are your recommendations and direction for further analysis and visualisations?

Visualization of the data

The data visualisation is to be built using Tableau. You should build one or two dashboards that tell the story of your data. The type of charts / tables you use is important as some work better than others or some simply make your message clearer.

Using captions, headings and appropriate labels / colour etc. is of paramount importance as it puts your visualisation into context. These contributing elements on the dashboard helps focus your reader’s attention and should increase their understanding of the data visualised.

You should also look to include some of the more advanced features of Tableau such as interactions, parameters, functions, calculated fields etc. This is required as you need to demonstrate your knowledge of Tableau. Implementing these advanced features also shows that you have added a certain level of complexity to your CA since Tableau allows you to create visualisations quite quickly.

The percentage of marks is significant for documentation as you need to show how you analysed, cleaned and formatted your data. It also highlights how you arrived at your conclusions and identified further areas for investigation. The documentation is your way of specifying what you did along the way and therefore is something you should be documenting during the CA as opposed to at the end.

40 doc

60 viz

## Deliverables

A zip file should be uploaded to Moodle, which contains the entire contents of your CA. This should include at least the following:

1. Word document
2. Datasets acquired and a separate tab demonstrating some of the seven stages
3. Datasets that were imported into Tableau
4. Tableau file

You must ensure that your zipped file contains all these files and your visualisations work. Failing to do this will mean your CA will incur a 10% penalty as it dramatically slows down results being published.

**NOTES**

Overview description of your CA and what you are trying to do. What are your main objectives and what are you hoping to discover?

Summary of CA

The goal of this data visualization and analysis project is to determine and display a number of things. The safest counties, the most dangerous counties, which counties are the worst for particular types of offences and last but not least, whether or not there has been an increase or decrease in crime rate over the period of the dataset. I plan to portray all of this in a highly consumable manor, It will be very easy to make sense of what you’re looking at as I will use industry best practices and standards for the visualization of the data.

Background

Why and how you came up with this idea. Why do you have an interest doing data collection, analysis and visualisation in this area? Has this or a similar data visualisations / analysis been done before in this area? Etc.

When I initially decided on visualizing irish crime rates, I wanted to maintain a focus on bank and post office robberies, as I find the topic of bank robberies interesting. Although searching for that kind of data proved to be fruitless, as I couldn’t acquire a dataset of such things, however I was able to find irish crime statistics that took place from 2003 - 2017 per garda station, this piqued my interest as it was close to what I wanted, so I decided it would be a good idea to use this dataset, in conjunction with a population census dataset for each county in those years, to accurately determine the safest and most dangerous counties, by measuring the crime stats for a county against its population I can get a good representation of the crime per person living there.

Crime statistics dataset from cso.ie

Had to clean the dataset to make sure each row (police department) was belonging to a specific county, then i wrote a script to create a new CSV from the original dataset that counted up each row and summed it up as each county, making there 26 rows per crime, one for each county. We now have condensed 5000 rows of data into roughly 200 rows, as each row is now a county, instead of only a town in that county. This is much easier for us

Population dataset:

Since the population census is only every 5 years, i’ve had to make concessions for my 2003-2017 crime dataset, as i only have population data for 2002, 2006, 2011, 2016. So to combat this, i’ve forward filled the population dataset to give the crime set years population values. Ex.

Years 2003 - 2005 of the crime set all contain the 2002 population data.

Years 2006 - 2010 of the crime set, contains the 2006 population data.

Years 2011 - 2015 of the crime set, contains the 2011 population data.

Years 2016 & 2017 of the crime set, contains the 2016 population data.

Hopefully this is an understandable concession to make as I still wanted to use population data to create ratios between the counties i am visualizing. It is as accurate as it can be using the data available.