Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

Analysing Crime Data in Ireland with focus on changes during the Recession

Table of Contents

Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

Introduction

I've chosen this data set based on my interest in crime and crime statistics. When I first moved to Ireland, the first thing I googled was how high the crime rates are, so it feels appropriate to use a dataset that focusses on crime statistics in Ireland in this report. I've found the dataset on the official website of the Irish Central Statistics Office[i].

The dataset includes all crimes committed in the Republic of Ireland from 2003 up to 2019. The column headers are Garda Region, Type of Offence, Quarter and the number of incidences (named DATA in the set). The dataset has 19,596 entries, due to the Type of Offence section being quite vast, and being split into quarters instead of years covering 4 separate regions.
The types of offences in the set include the following main categories and sub-categories:

1. 01 Homicide offences'
   a. '0111 Murder'
   b. '0112 Manslaughter'
   c. '0113 Infanticide'
   d. '012 Dangerous driving leading to death'
2. '02 Sexual offences'
   a. '021 Rape and sexual assault'
   b. '022 Other sexual offences'
3. '03 Attempts/threats to murder, assaults, harassments and related offences'
   a. '0311 Murder-attempt'
   b. '0312 Murder-threat'
   c. '033 Harassment and related offences'
   d. '034 Assault causing harm, poisoning'
   e. '035 Other assault'
4. '04 Dangerous or negligent acts'
   a. '0411 Dangerous driving causing serious bodily harm'
   b. '0412 Driving/in charge of a vehicle while over legal alcohol limit'
   c. '0413 Driving/in charge of a vehicle under the influence of drugs'
   d. '0421 Endangerment with potential for serious harm/death'
   e. '0422 Abandoning a child, child neglect and cruelty'
   f. '0423 Unseaworthy/dangerous use of boat or ship'
   g. '0424 False alarm/interference with aircraft or air transport facilities'
   h. '0425 Endangering traffic offences'
5. '05 Kidnapping and related offences'
   a. '0511 False imprisonment'
   b. '0512 Abduction of person under 16 years of age'
   c. '0513 Human trafficking offences'
6. '06 Robbery, extortion and hijacking offences'
   a. '0611 Robbery of an establishment or institution'
   b. '0612 Robbery of cash or goods in transit'
   c. '0613 Robbery from the person'
   d. '0621 Blackmail or extortion'
   e. '0631 Carjacking, hijacking/unlawful seizure of aircraft/vessel'
7. '07 Burglary and related offences'
   a. '0711 Aggravated burglary'
   b. '0712 Burglary (not aggravated)'
   c. '0713 Possession of an article (with intent to burgle, steal, demand)'
8. '08 Theft and related offences'
   a. '081 Theft/taking of vehicle and related offences'
   b. '0821 Theft from person'
   c. '0822 Theft from shop'
   d. '084 Other thefts, handling stolen property'
9. '09 Fraud, deception and related offences'
10. '10 Controlled drug offences'

    a. '1011 Importation of drugs'
    b. '1012 Cultivation or manufacture of drugs'
    c. '1021 Possession of drugs for sale or supply'
    d. '1022 Possession of drugs for personal use'
    e. '103 Other drug offences'
11. '11 Weapons and Explosives Offences'
    a. '111 Explosives, chemical weapons offences'
    b. '1121 Discharging a firearm'
    c. '1122 Possession of a firearm'
    d. '113 Offensive weapons offences (n.e.c.)'
    e. '114 Fireworks offences'
12. '12 Damage to property and to the environment'
    a. '1211 Arson'
    b. '1212 Criminal damage (not arson)'
    c. '1221 Litter offences'
13. '13 Public order and other social code offences'
    a. '131 Disorderly conduct'
    b. '132 Trespass offences'
    c. '133 Liquor licensing offences'
    d. '134 Prostitution offences'
    e. '135 Regulated betting/money, collection/trading offences'
    f. '136 Social code offences (n.e.c.)'
14. '15 Offences against government, justice procedures and organisation of crime'
    a. '151 Offences against government and its agents'
    b. '152 Organisation of crime and conspiracy to commit crime'
    c. '153 Perverting the course of justice'
    d. '157 Offences while in custody, breach of court orders'

Additionally, it is split into 4 regions, which I've found a visualisation of here:

ii

In order to analyse this set fully, I will partially apply the CRISP-DM model[iii], which includes the following six phases:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Data Modelling
5. Evaluation
6. Deployment

While I will follow the CRISP-DM model to the best of my ability, seeing as my report focusses more on analysis instead of predictions, I will likely omit the Modelling and Deployment phase.

Due to the data set including the data from 2003 until 2019, I thought it would be interesting to take a closer look at which, if any, type of offences increased during the recession, which took place approximately from 2008 to 2011.

For guidance, I will attempt to answer the following questions:

1. Which types of offences spiked during the recession?
2. Are these types of offences actually only higher during the years of recession or do they appear to be fluctuating overall?

In order to answer these questions, I will import the data set into the Jupyterlab and analyse it using Python.

Data Wrangling

IV. INITIAL IMPORT AND FINDINGS

For my first findings, I've imported pandas, pyaxis, altair and the dataset using the URL into the Jupyter notebook. Importing pyaxis[iv] actually was a first challenge for me, because I had never used this before, but it was necessary to read the .px file type that the data set is saved as.

For my initial findings, I've used the *data.info()*, *data.describe()*, *data.dtypes*, *print(data)* operators[v] as well as a *for-loop* to extract all unique values of Garda Region, Type of Offence and Quarter.

- There are no missing values in the dataset
- All data types are objects
- The Garda Region with the highest incident count is the Eastern Region
- The highest Type of Offence is Sexual Offences
- The Quarter with the highest incident amount is 2006Q2
- The Types of Offences appear to be split into categories and sub-categories

TRANSFORMING THE DATA AND CREATING LISTS

Necessary steps:
The following changes are necessary to move forward with the analysis:

- DATA column will have to be renamed to Count to avoid confusion, because I've named the dataset itself data
- The data type of that same column will have to be changed into integer to allow calculations to be made

Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

- A check of whether the incident count of the categories is the sum of the sub-categories

Additionally, it is imperative to create lists to make analysing the desired data easier. This includes lists of years, overarching categories of offence types, regions.

EXECUTION

Renaming the Column from DATA to Count was done using the *data.columns* operator that we were taught in class[vi]. In addition, I've changed the data type of the same column using *data['Count'].astype(int)*. The *astype* operator[vii] is used to convert data types in pandas.

After changing the name and data type of the Column, I did a test to see if the sums of the overarching categories and their sub-categories are identical. For this, I used the first offence type, homicide offences and the sub-categories murder, manslaughter, infanticide and dangerous driving leading to death.

I created a list of the sub-categories (titled *heinous_crimes* in the notebook), and used the *.sum()* operator on both the homicide offences and heinous crimes.
The outcome was 1763 for both, thus proving that the overarching categories are the sums of the individual sub-categories.

```
# Are the '01 Homicide offences' inclusive of the sub-categories '0111 Murder' '0112 Manslaughter' '0113 Infanticide'

homicide_num = data['Count'][data['Type of Offence']== '01 Homicide offences'].sum()

heinous_crimes = ['0111 Murder', '0112 Manslaughter', '0113 Infanticide', '012 Dangerous driving leading to death']

hc_num = data['Count'][data['Type of Offence'].isin(heinous_crimes)].sum()

print(homicide_num)
print(hc_num)
```

```
1763
1763
```

Knowing this makes it easier to further delve into the data, seeing as I was able to create a list of just the main categories, thus omitting the sub-categories for calculations where possible, and avoiding duplication.

I created lists using the metadata (*regions, offence_types, quarters*) to begin with, but the desire to be able to use purely the main categories of offence types as well as the option to use years instead of quarters meant a need for further listing. I achieved both using the for-loop operator. Please see the list and for-loop for years as an example below.

```
# combining quarters to be able to use years; using lists to create this with for loop

quarter_names = ['Q1', 'Q2', 'Q3', 'Q4']
years = ['2003', '2004', '2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016'
year_lists = {}

for year in years:
    year_lists[year] = [year + q for q in quarter_names]

print(year_lists)
```

```
{'2003': ['2003Q1', '2003Q2', '2003Q3', '2003Q4'], '2004': ['2004Q1', '2004Q2', '2004Q3', '2004Q4'], '2005': ['2005Q
1', '2005Q2', '2005Q3', '2005Q4'], '2006': ['2006Q1', '2006Q2', '2006Q3', '2006Q4'], '2007': ['2007Q1', '2007Q2', '20
07Q3', '2007Q4'], '2008': ['2008Q1', '2008Q2', '2008Q3', '2008Q4'], '2009': ['2009Q1', '2009Q2', '2009Q3', '2009Q4'],
'2010': ['2010Q1', '2010Q2', '2010Q3', '2010Q4'], '2011': ['2011Q1', '2011Q2', '2011Q3', '2011Q4'], '2012': ['2012Q
1', '2012Q2', '2012Q3', '2012Q4'], '2013': ['2013Q1', '2013Q2', '2013Q3', '2013Q4'], '2014': ['2014Q1', '2014Q2', '20
14Q3', '2014Q4'], '2015': ['2015Q1', '2015Q2', '2015Q3', '2015Q4'], '2016': ['2016Q1', '2016Q2', '2016Q3', '2016Q4'],
'2017': ['2017Q1', '2017Q2', '2017Q3', '2017Q4'], '2018': ['2018Q1', '2018Q2', '2018Q3', '2018Q4'], '2019': ['2019Q
1', '2019Q2', '2019Q3', '2019Q4']}
```

The for-loop for the offence type categories proved to be more difficult, seeing as I had to separate the overarching categories from the sub-categories, and find a way to filter them from each other. In order to achieve this, I've used the *.startswith()* operator[viii], because the categories all have numbers in the front, which the sub-categories and main categories share. Additionally, to exclude the main categories from the sub-categories, I used the != (not equal) Boolean operator[ix].

```
crime_categories = ['01 Homicide offences',  '02 Sexual offences', '03 Attempts/threats to murder, assaults, harassment
crime_by_cat = {}

for category in crime_categories:
    crime_by_cat[category] = [crime for crime in offence_types if crime.startswith(category[0:2]) and crime != category
```

Recession Analysis

After this adaption of the dataset, I started the Recession analysis by getting overall sums averages, including all regions and the main offence categories, from 2003 to 2019. I got the averages using the *.mean()* operator and also sorted the sums and averages with the .sort_values() operator[x]. Additionally, I used the *.groupby()* operator to sum up by the *Type of Offence* column[xi].

```
# avg of all main crime categories between 2003 and 2019
# create subtable of group by function
mean_of_crimecats = data_by_crimecat.groupby('Type of Offence').mean()
# sort the group subtable by count
mean_of_crimecats.sort_values('Count')
```

	Count
Type of Offence	
01 Homicide offences	6.387681
05 Kidnapping and related offences	7.018116
02 Sexual offences	127.655797
06 Robbery, extortion and hijacking offences	160.192029
11 Weapons and Explosives Offences	181.134058
09 Fraud, deception and related offences	329.967391
15 Offences against government, justice procedures and organisation of crime	687.402174
04 Dangerous or negligent acts	728.246377
10 Controlled drug offences	1044.750000
03 Attempts/threats to murder, assaults, harassments and related offences	1060.286232
07 Burglary and related offences	1511.333333
12 Damage to property and to the environment	2049.405797
13 Public order and other social code offences	2777.855072
08 Theft and related offences	4591.039855

I repeated this process for the years of the Recession, meaning I only included data from 2008 to 2011. This already gave me an idea of which type of offences may have peaked during the recession. However, seeing as the overall sums and means included the recession, I created another process, excluding the recession years to allow for more accuracy.

In order to achieve this, I excluded all the Recession years using the not (~) operator[xii], and repeated the previous process.

After creating all these processes, and running them successfully, I've compared the recession and none-recession processes using basic calculation operators. I've multiplied it by 100 to get the percentage.

```
difference_absolute = (mean_of_crimecats_recession - mean_of_crimecats_not_recession)

percentage_diff_nr = (difference_absolute/mean_of_crimecats_not_recession)*100

percentage_diff_nr
```

```
INFO:numexpr.utils:NumExpr defaulting to 8 threads.
```

Type of Offence	Count
01 Homicide offences	-2.473385
02 Sexual offences	-16.799316
03 Attempts/threats to murder, assaults, harassments and related offences	12.158650
04 Dangerous or negligent acts	26.813068
05 Kidnapping and related offences	4.344874
06 Robbery, extortion and hijacking offences	10.253565
07 Burglary and related offences	14.445573
08 Theft and related offences	5.934736
09 Fraud, deception and related offences	5.349351
10 Controlled drug offences	32.928198
11 Weapons and Explosives Offences	51.934467
12 Damage to property and to the environment	30.867496
13 Public order and other social code offences	36.175570
15 Offences against government, justice procedures and organisation of crime	6.418838

The result of this comparison shows that during the recession years, 4 different types of offences have increased by over 30 %. These offence types include Weapons and Explosive Offences (52 %), Public order and other social code offences (36 %), Controlled drug offences (33 %) and Damage to property and the environment (31 %).

Next, I plotted these top 4 types by region, in order to see if the increase during the recession was part of an ongoing trend or actually appeared to be influenced by the recession. When plotting the first offence type – Weapons and Explosive Offences – by region, I used the .drop() operator[xiii] to remove the offence type, as it would be repetitive to keep it.

Directly afterwards, I plotted the offence type by region to visualise the results in a bar graph, which we learned while analysing churn in Class on 12th May.
I repeated the previously mentioned process and the plotting for all top 4 types of offences.

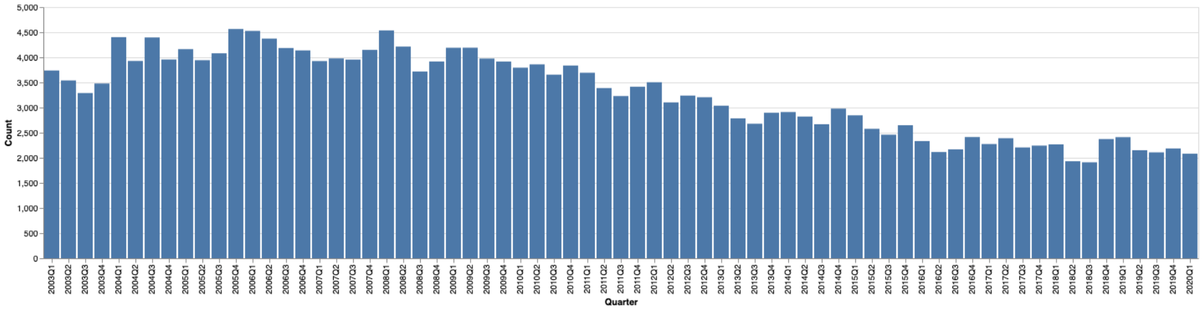
Conclusion

Having charts per region over all quarters for the offences that varied most between recession years and none-recession years, the assumptions that can be made highly differ between offences.

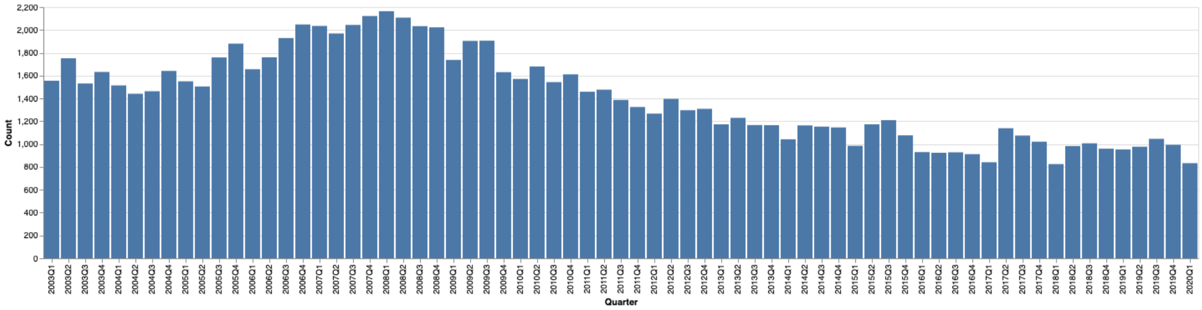
1. Damage to property and to the environment & Public order and other social code offences
    a. Both of these offence types appear to be trending downward over the years
    b. This allows the assumption that the variation from the recession years is due to the average being pulled from both before the recession, when the amount of incidences was comparably high, and after, where the incidences became lower

Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

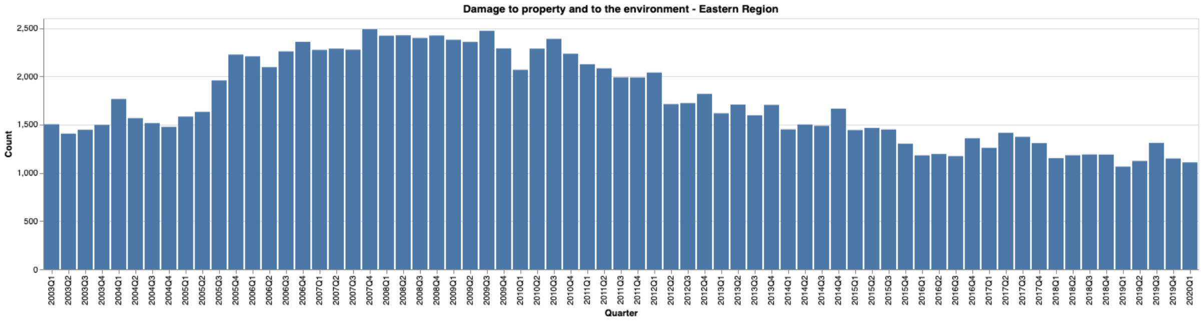
Damage to property and to the environment - Dublin



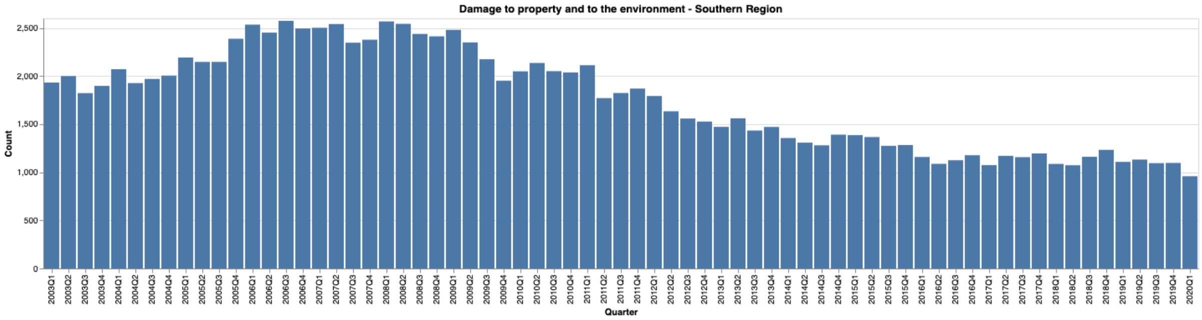
Damage to property and to the environment - North and Western Region



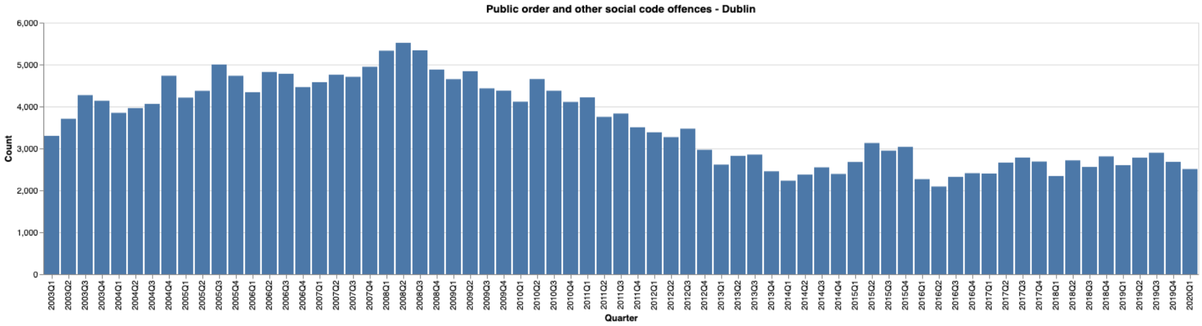
Damage to property and to the environment - Eastern Region

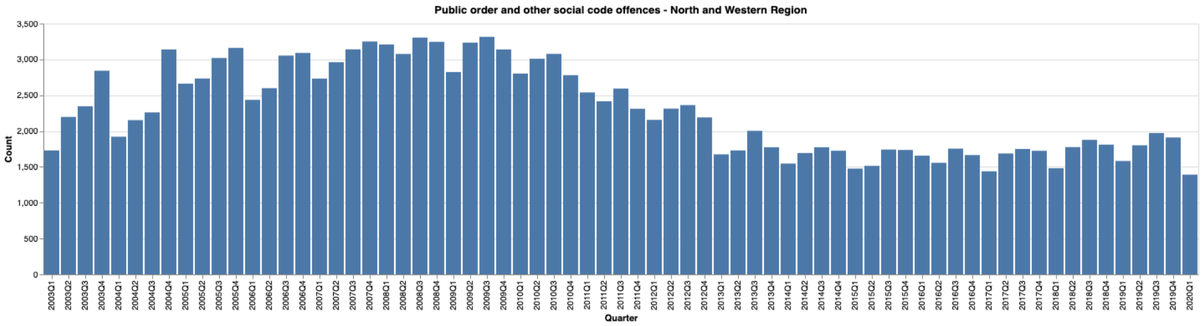


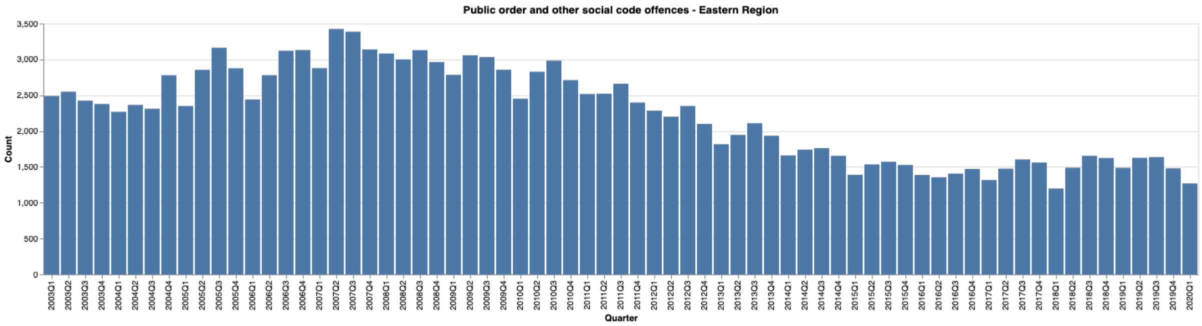
Damage to property and to the environment - Southern Region

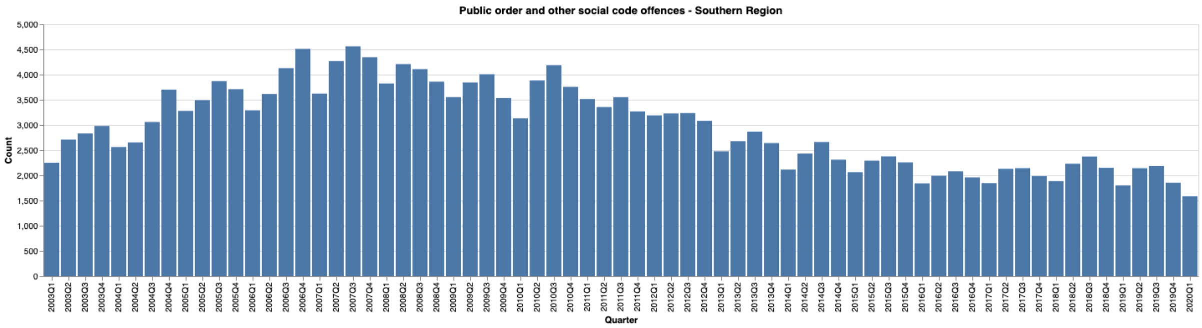


Public order and other social code offences - Dublin

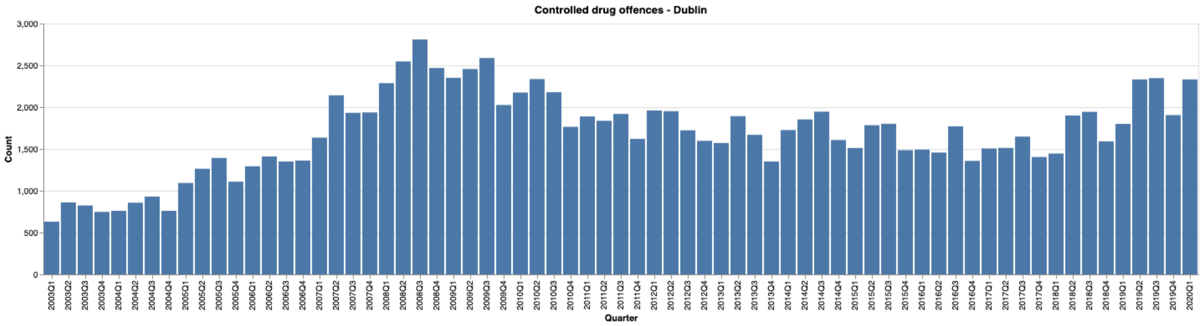


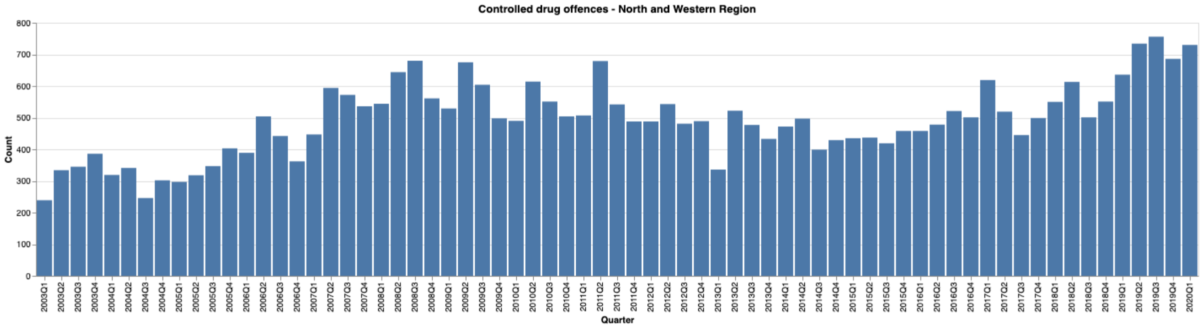


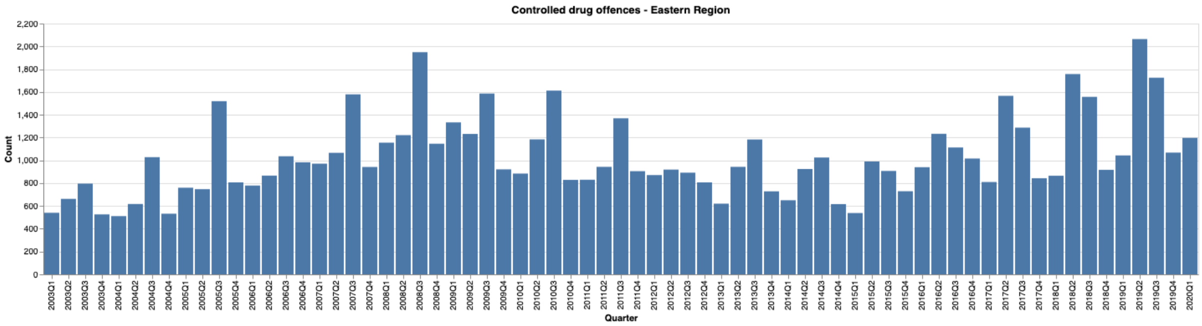


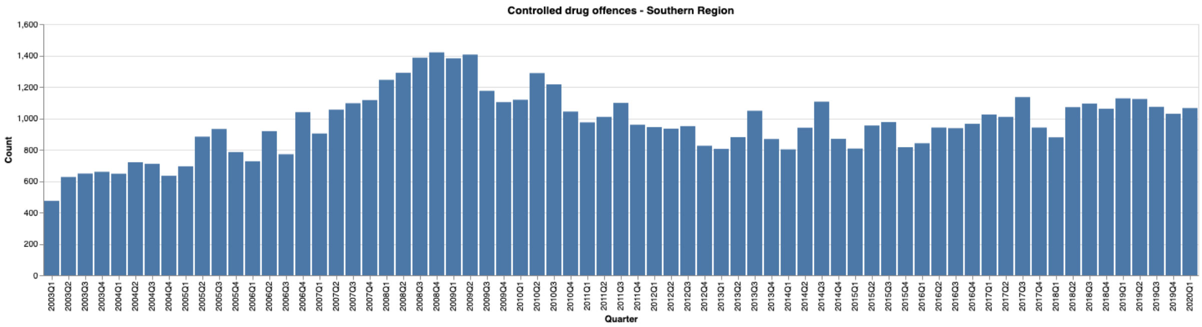


2. Controlled drug offences
    a. Where the previous two offence types appear to be declining over the years, this offence type seems to be fluctuating
    b. While it is higher during the recession, the difference is not as high as the average made it appear, seeing as this offence type appears to have dramatically increased generally since around 2006.



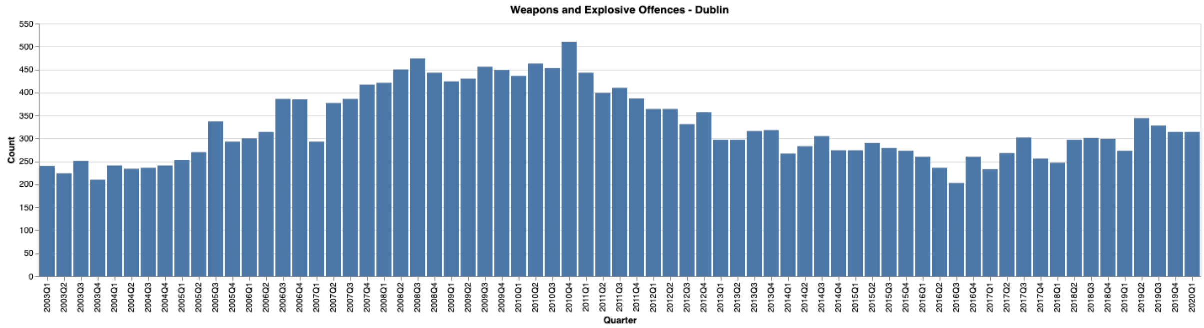


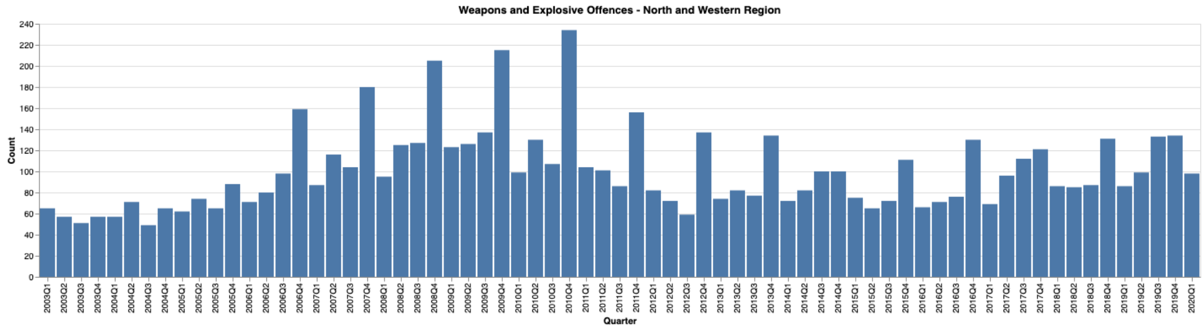


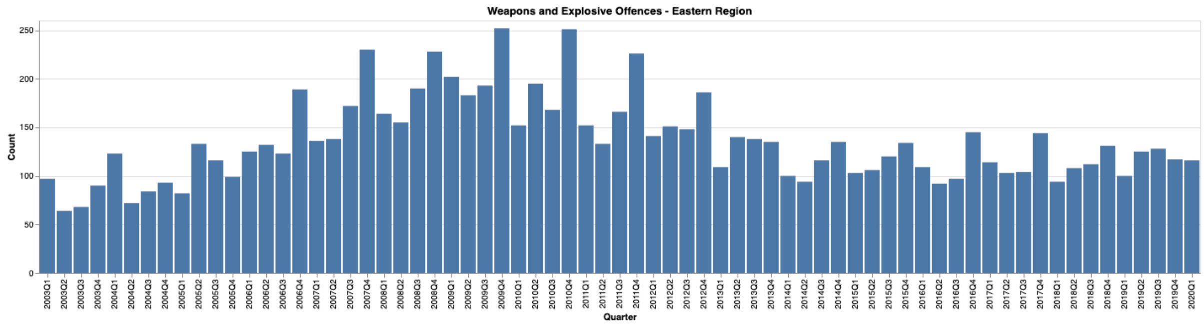


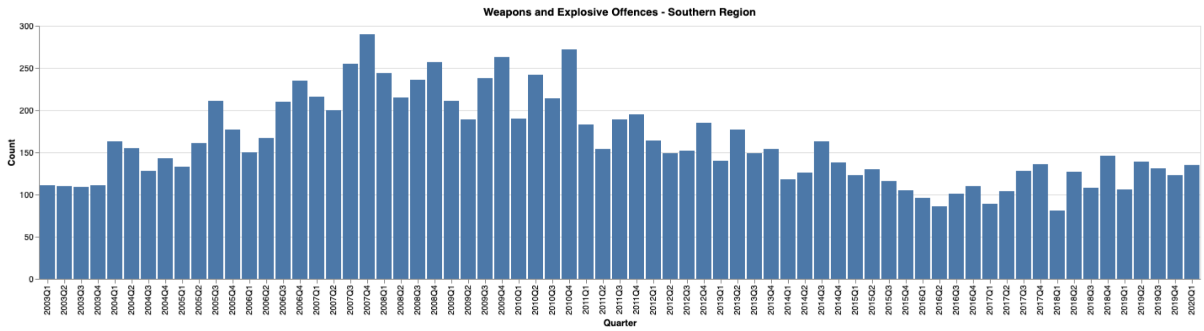
3. Weapons and Explosive Offences
    a. This offence type is the only one that actually looks to have been affected by the recession
    b. In the chart it is visible that weapons and explosive offences were low in the early 2000s and are similarly low again almost immediately after the beginning of 2012

Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics









Given these results, I decided to look more closely into the weapons and explosive offence type, especially given that the North and Western Region as well as the Eastern Region results have a high amount of spikes in the 4th quarter of each year during (and a little after) the recession, indicating seasonal variation.

In order to investigate this seasonal variation, I used the categories list I created to quickly look at the sub-categories included in weapons and explosive offences. This made the reason for the variation glaringly obvious: Fireworks offences are a sub-category, clearly explaining why the 4th quarter, which includes New Years' celebrations, would have an increase in this specific type of offence.

This new information made me exclude the Firework offences, to get a better idea if the recession still affected all other sub-categories of the weapons and explosives offences, so essentially removing the seasonal variation.

In order to compare the two, I used the *merge()* operator[xiv], which allowed me to combine columns from two separate tables into one table.

```
all_wae_offences = data[data['Type of Offence'] == '11 Weapons and Explosives Offences'].drop('Type of Offence',1)

fireworks_offences = data[data['Type of Offence'] == '114 Fireworks offences'].drop('Type of Offence',1)
```

```
merged_fw_all = pd.merge(all_wae_offences, fireworks_offences, on=['Garda Region', 'Quarter'], how='inner')
merged_fw_all
```

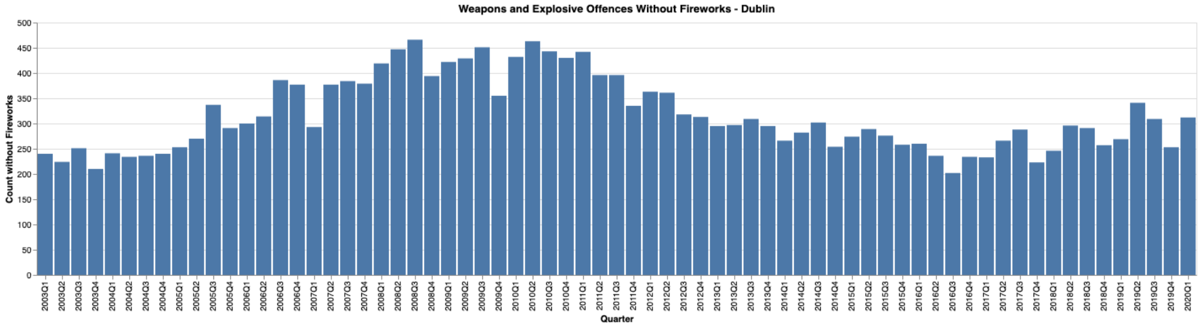
	Garda Region	Quarter	Count_x	Count_y
0	10 Dublin Metropolitan Region	2003Q1	240	0
1	10 Dublin Metropolitan Region	2003Q2	224	0
2	10 Dublin Metropolitan Region	2003Q3	251	0
3	10 Dublin Metropolitan Region	2003Q4	210	0
4	10 Dublin Metropolitan Region	2004Q1	241	0
...	...	...	...	...
271	40 Southern Region	2019Q1	106	0
272	40 Southern Region	2019Q2	139	2
273	40 Southern Region	2019Q3	131	9
274	40 Southern Region	2019Q4	123	22
275	40 Southern Region	2020Q1	135	2

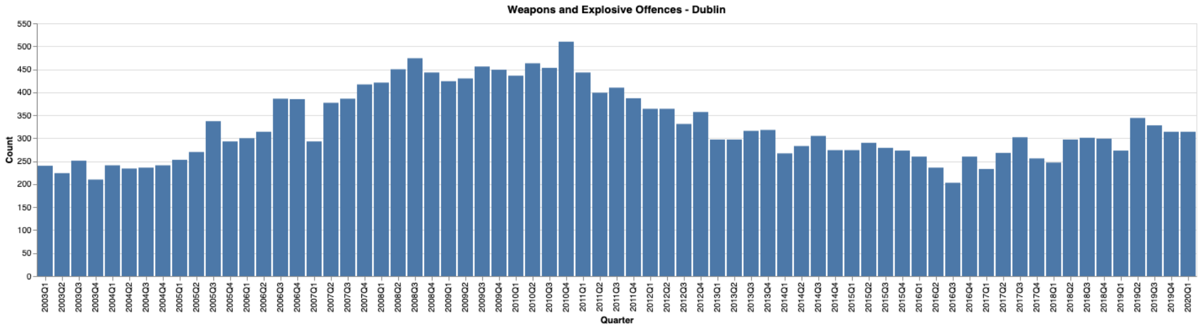
276 rows × 4 columns

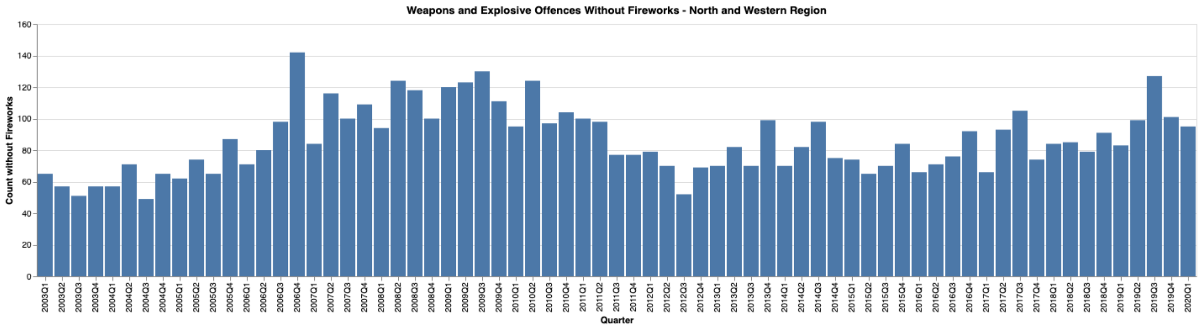
After combining the two, I created charts for the regional weapons and explosive offences without fireworks, and plotted them alongside the previously created charts which include all weapons and explosive offences.

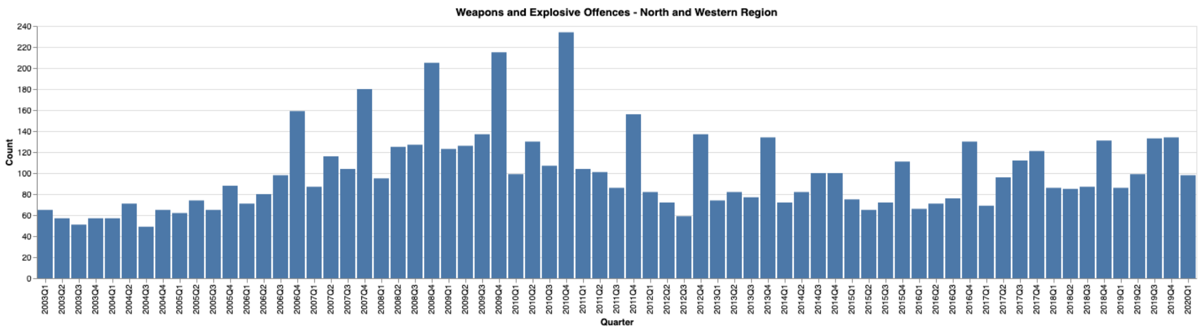
This proved my initial statement that weapons and explosive offences were higher during the recession, as visible in the charts below. Additionally, it definitely proves the seasonal variation very prominently.

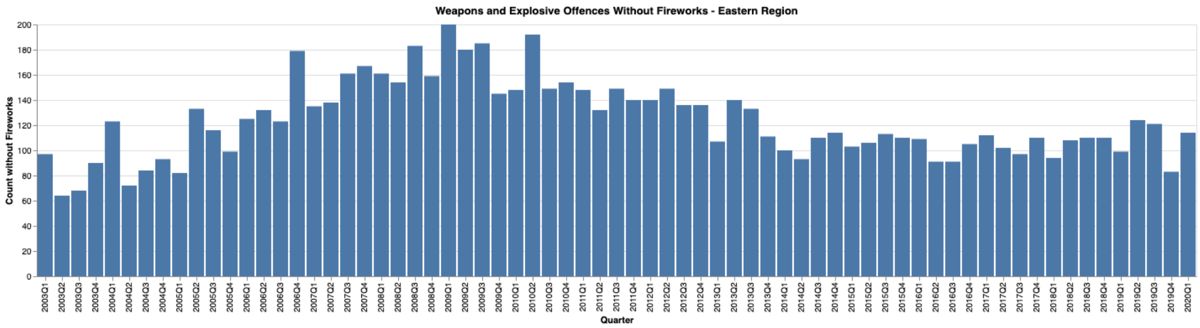
Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

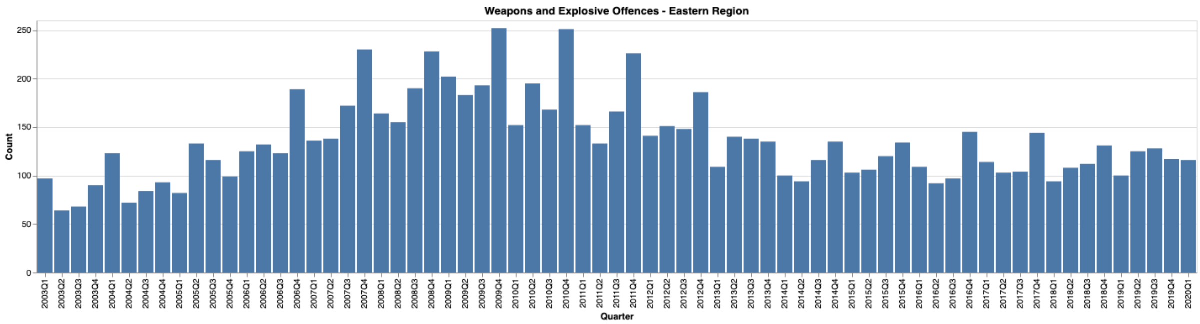


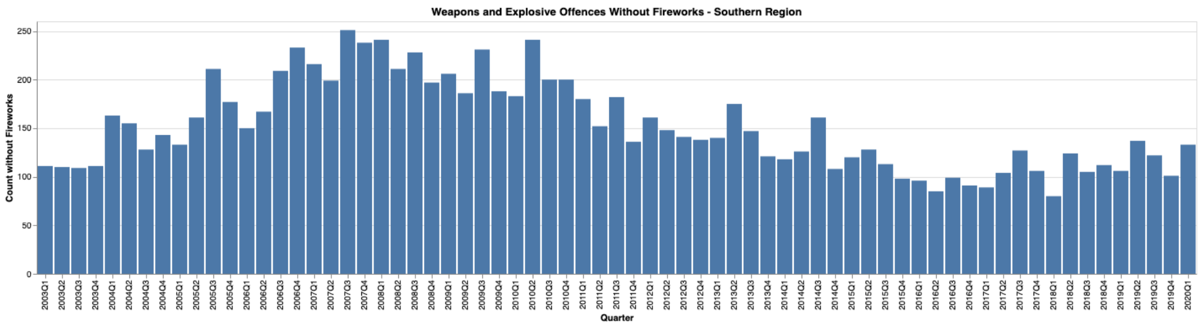


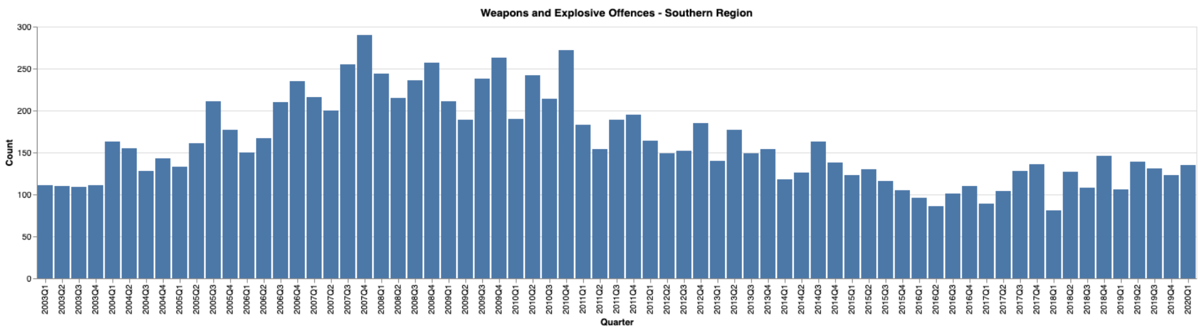












Conclusion

In conclusion, I will answer my guiding questions from the Introduction.

1. Which types of offences spiked during the recession?
    a. The four types which spikes during the recession were the following:
        i. Weapons and Explosive offences
        ii. Damage to property and to the environment
        iii. Public order and other social code offences
        iv. Controlled drug offences
2. Are these types of offences actually only higher during the years of recession or do they appear to be fluctuating overall?

   a. Except for the weapons and explosive offences, none of the other ones actually purely spiked during the recession. Two of them were on an upward trend from the early 2000s until 2019, and the controlled drug offences appear to be fluctuating.
3. Are there any other reasons the type of offences could go up during the recession other than the recession itself?

Final Thoughts

Overall, it was very interesting to deep dive into crimes committed in Ireland, and to see the differences in region. I was actually surprised about the weapons and explosives category being the only one that actually spiked during the recession, without much visible fluctuation outside of that timeframe. I had personally expected more Public order and other social code offences, seeing as frustration is a given during times such as the recession, and can easily lead to more disorderly conduct.

Additionally, on a purely data-based note, the dataset proved to be more difficult to work with than I initially expected, seeing as I had to create a massive amount of lists to be able to index into years, categories of crimes, and similar. The creation of those lists took a lot of time and patience.

Student: Lisa Tschoertner Student No: 22001500
Student Email: 22001500@ibat.ie Personal Email: lisa.tschoertner@googlemail.com
Course: Diploma in Predictive Data Analytics

References

[i] https://statbank.cso.ie/px/pxeirestat/DATABASE/Eirestat/Recorded%20Crime/Recorded%20Crime_statbank.asp?sp=Recorded%20Crime&Planguage=0
[ii] https://www.garda.ie/images_upload/en/About-Us/Organisational-structure/Regional-2019.png
[iii] http://www.datascience-pm.com/crisp-dm-2/
[iv] https://jakevdp.github.io/blog/2017/12/05/installing-python-packages-from-jupyter/
[v] https://drive.google.com/file/d/1mhFgY-fqE9mBk58Hgy8sIj1p-4N2Vbu-/view
[vi] https://drive.google.com/file/d/1mhFgY-fqE9mBk58Hgy8sIj1p-4N2Vbu-/view
[vii] https://www.geeksforgeeks.org/python-pandas-dataframe-astype/#:~:text=astype()%20method%20is%20used,type%20to%20another%20data%20type.
[viii] https://stackoverflow.com/questions/44517191/how-to-find-the-python-list-item-that-start-with
[ix] https://python-reference.readthedocs.io/en/latest/docs/operators/not_equal.html
[x] http://datacamp-community-prod.s3.amazonaws.com/dbed353d-2757-4617-8206-8767ab379ab3
[xi] https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.groupby.html
[xii] https://stackoverflow.com/questions/21415661/logical-operators-for-boolean-indexing-in-pandas
[xiii] http://datacamp-community-prod.s3.amazonaws.com/dbed353d-2757-4617-8206-8767ab379ab3
[xiv] https://datascience.stackexchange.com/questions/33053/how-do-i-compare-columns-in-different-data-frames.