



# DRUG & BURGLARY OFFENSE IN IRELAND LARGEST THREE COUNTIES BETWEEN 2003 & 2015

Tom Lyons – [tom@scopeav.ie](mailto:tom@scopeav.ie)

## GitHub URL

[https://github.com/TommyLyons1989/UCDPA\\_TommyLyons](https://github.com/TommyLyons1989/UCDPA_TommyLyons)

## Introduction

I was interested in examining if there were any trends in two primary anti-social behaviour indicator crimes; controlled drugs use and burglary. During my undergraduate degree in sociology a correlation between these two acts was a phenomenon that was claims as true. I was able to attain a dataset which contains original variables from the CSO regarding criminal offences recorded in the 563 Garda Stations in the Republic of Ireland is available on an annual basis from 2003 – 2016 (link states 2010-2016 but data available back to 2003)

<https://www.kaggle.com/abhijithchandradas/ireland-crimes-at-garda-stations-level-20102016>

I decided to focus on the three most populous counties in the Republic of Ireland – Dublin, Cork & Galway. These counties also have the three largest cities. I examined Drug offenses and Burglary offense for the years 2003, 2009 and 2015. I did this to investigate over even 6 year intervals. I was also interested to determine if there were any noticeable changes in the during the economic crash of 2009 i.e. would Crime spike noticeable in 2009

## Dataset

As outlined previously I was able I was able to attain a dataset which contains original variables from the CSO regarding criminal offences recorded in the 563 Garda Stations in the Republic of Ireland is available on an annual basis from 2003to 2016. This dataset listed by name each Garda Station and also showed which Division they related to. For example, the following Stations are shown to be within the Clare Division:

- Ballyvaughan
- Crusheen
- Lisdoonvarna
- Corofin
- Ennis
- Ennistymon
- Kildysart
- Kilkee
- Killaloe
- Kilrush
- Lissycasey
- Miltown Malbay
- Newmarket on Fergus
- Scariff
- Shannon
- Sixmilebridge
- Tulla.

There is no column in the dataset specifying which county each station is in, but as shown above, most of the Divisions denote a specific county and include all the relevant stations. There are a few exception to this.

- Dublin Stations : Due to the scale of the population of Dublin, the stations are broken down into six divisions. Eastern, Northern, North Central, Western, Southern and Southern Central
- Cork Stations: These are broken down into 3 divisions. Cork City, Cork West, Cork North,
- There are also a number of counties where the stations are combined into a single division. This includes: Kilkenny/Carlow, Cavan/Monaghan, Laois/Offaly, Roscommon/Longford, Sligo/Leitrim,

The crimes recorded in this dataset range from the year 2003 to 2016. There is an asterisk all 2016 columns which indicates to me that these are not complete year totals. Therefore, I decided to disregard these columns of any analysis.

The list of tracked offences are:

- Attempts or threats to murder, assaults, harassments and related offences
- Dangerous or negligent acts
- Kidnapping and related offences
- Robbery, extortion and hijacking offences
- Burglary and related offences
- Theft and related offences
- Fraud, deception and related offences
- Controlled drug offences
- Weapons and Explosives Offences
- Damage to property and to the environment
- Public order and other social code offences
- Offences against government, justice procedures and organisation of crime

Each of these offenses is tracked per year with a nominal figure attached. There are two columns labelled (X) & (Y) which appear to have co-ordinate values. These are of no use to me and will be removed.

## Implementation Process

My first step was to import all the required packages and denote them with the relevant shortcuts.

- import matplotlib.pyplot as plt
- import numpy as np
- import pandas as pd
- import seaborn as sns
- import requests

I used the pandas to import my CSV file and labelled it `Garda_Station_Stats`. I had to add `r` to be beginning of the `pd.read_csv()` function in order to get my file to import.

*(I could not find a SQL, JSON or API relevant to this dataset. There at this point I perform a request gets on [newsapi.org](https://newsapi.org) to show my understanding of performing this task)*

I wanted to confirm all the column names within my imported CSV. I used a for loop linking key word `col` to the `.columns` function then passing `'print(col)'`.

I determined that the columns "x" & "y" were not useful columns to have in the data set so I removed them and saved the data frame as `GSS`. I then ran my for-loop line again to print the columns for `GSS`. "x" & "y" had been removed.

I wrote code to show my understanding of the sorting and grouping features of python. I created a sub data frame `GSS_Sorted_Divisions` by sorting `GSS` by the values within the `Divisions` column first, then the `Station` column.

I then created a sub data frame `GSS_GROUPBY_Division` to group `GSS` by the `Divisions` column

I used `drop_duplicates` on a subset of `GSS` which only had the "Divisions" column. After dropping duplicates I was able to get a list of the `Divisions` which helped in my next step.

Since my goal was to investigate Dublin, Galway, and Cork, I created list for each county with all the relevant divisions included (a dictionary would not have be of use in this instance)

1. Dublin = DMR Eastern Division, DMR North Central. DMR Northern Division, DMR South Central Division, DMR Southern Division, DMR Western Division

2. Galway = Galway Division
3. Cork = Cork City Division, Cork North Division, Cork West Division

My next step was creating a sub data frame as the subset of GSS for all rows where the Division was included in a my above created lists. I did this with the line:

```
GSS[GSS["Divisions"].isin(Listname)]
```

I now had a three sub data frames (Dublin\_Crime, Galway\_Crime and Cork\_Crime) which I knew included all the data relating to all the Stations for each region.

I then created two re-usable functions: Drug\_Crime\_peryear & Burg\_Crime\_peryear.

These functions had four steps (Step 2 varied based on the offences I was searching):

- Step 1: Subset data frame to retrieve contents of the Station and Divisions columns using .loc
- Step 2: Subset data frame to retrieve columns index 100:113 using .iloc (Drug Offenses 2003-2015). For the Burglary 2003-2015 this function line was changed to index 58:71
- Step 3: Next I used the pd.concat function to join these two pandas subsets horizontally (this shows understand of join dataframes) . I added a sort\_values function to the end of this command. I sorted by Divisions then Stations. This meant my function would be ordered by the divisions with all stations alphabetical within that divisions. On my charts as a result it is easy to identify when each division ends by the alphabetical order
- Step 4: command return 'County'\_crime\_'Drug/Burg'\_Final

The result of the above is two functions which I re-used three times each

*(I am aware the above function could have been streamlined, but I wanted to show use of loc,iloc and data joining)*

I passed my three lists separately into both functions which then allowed me to have my final subset pandas data frames for creating my charts:

- Dublin\_Crime\_Drugs\_Final
- Galway\_Crime\_Drugs\_Final
- Cork\_Crime\_Drugs\_Final
- Dublin\_Crime\_Burg\_Final
- Galway\_Crime\_Burg\_Final
- Cork\_Crime\_Burg\_Final

*(I then used the iterrow command on all 6 of the above to demonstrate my understand of using this)*

## Charts

At this point I had to decide which way to show the data. Since I was comparing the same data across multiple locations, I decided a bar chart was the most appropriate. I was also comparing the same stations data over three-time intervals, so I then stacked this data on the bar charts. This way I could easily see Drug and Burglary offenses over time per station. This allowed me to identify any trends.

## Results

Figure 1 – Burglary Offense Dublin 2003-2009-2015

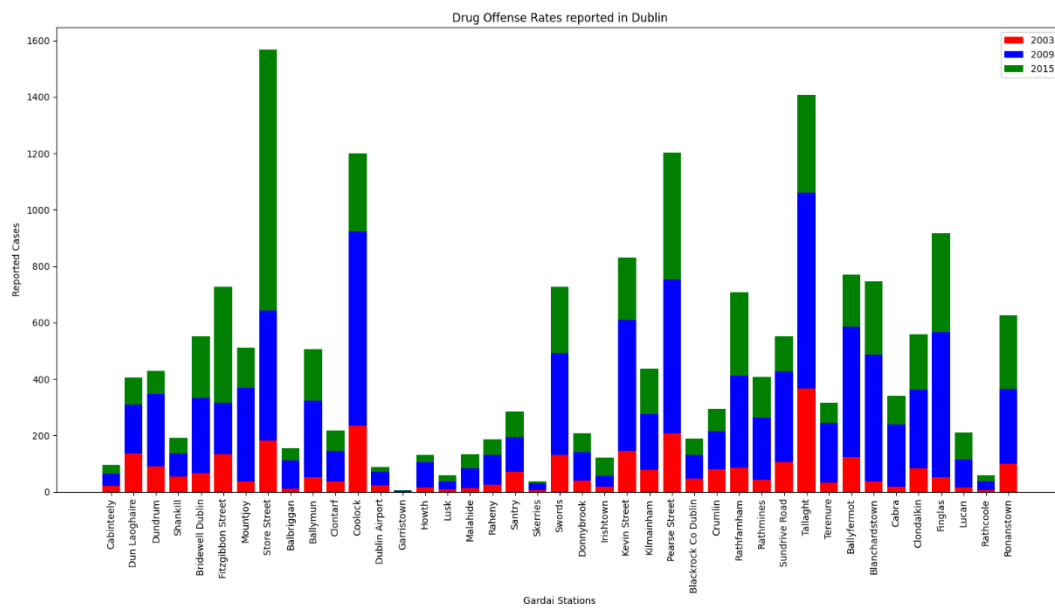


Figure 2- Drug Crime Offenses Dublin 2003-2009-2015

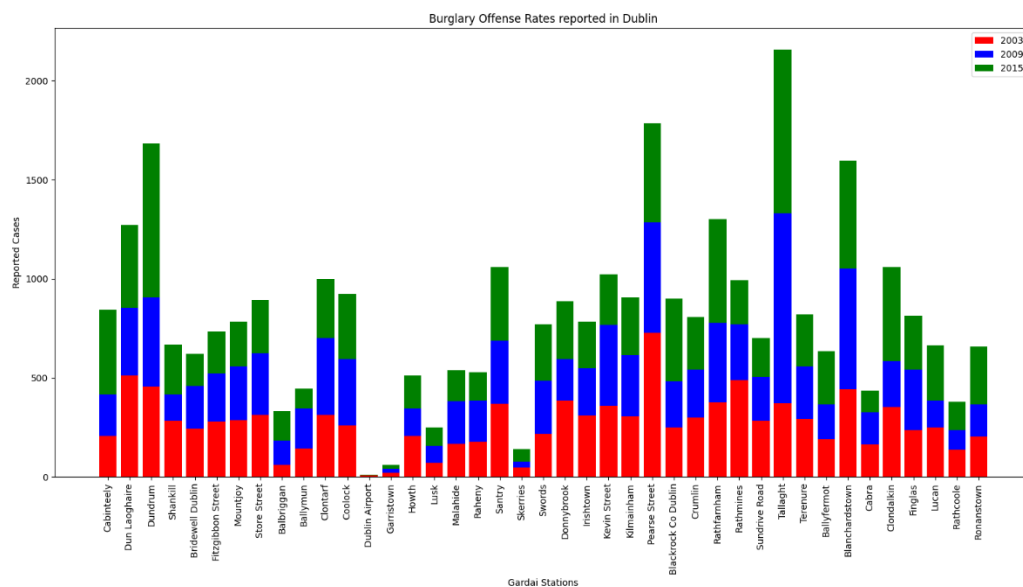


Figure 3 – Burglary Offenses Galway 2003-2009-2015

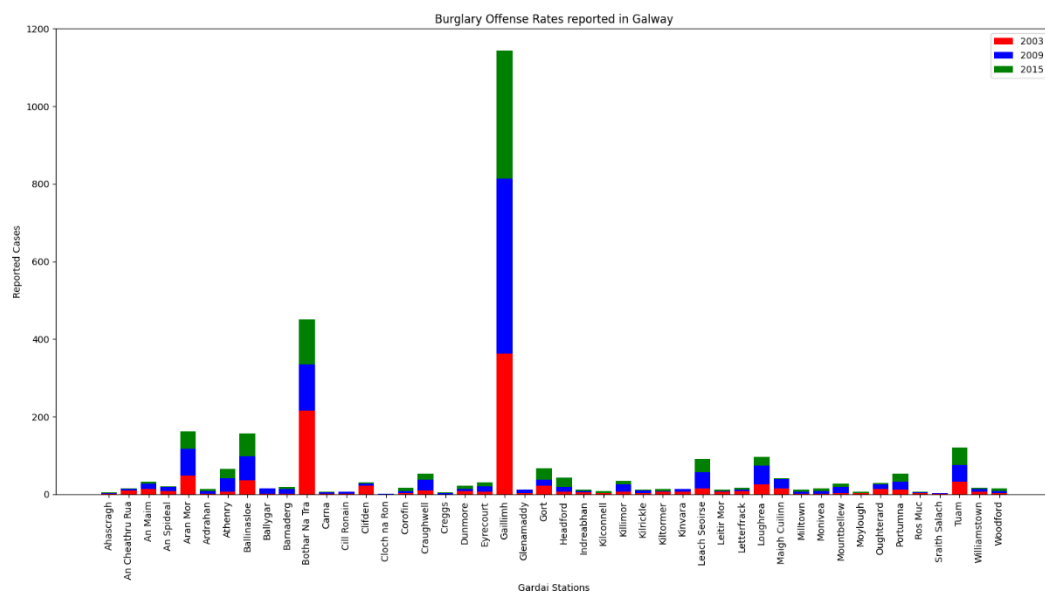


Figure 4 – Drug Crime Galway 2003-2009-2015

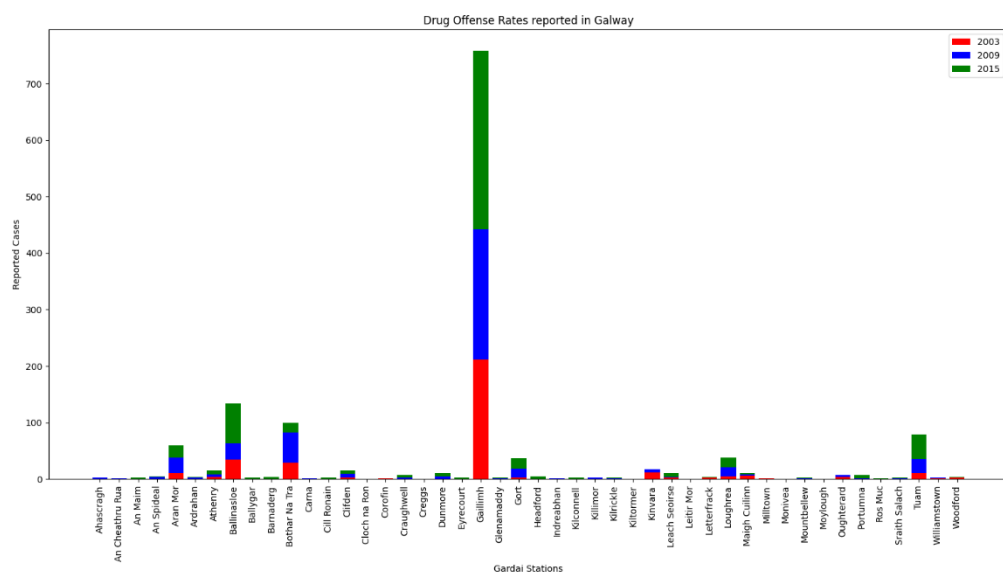


Figure 5 – Burglary Cork 2003-2009-2015

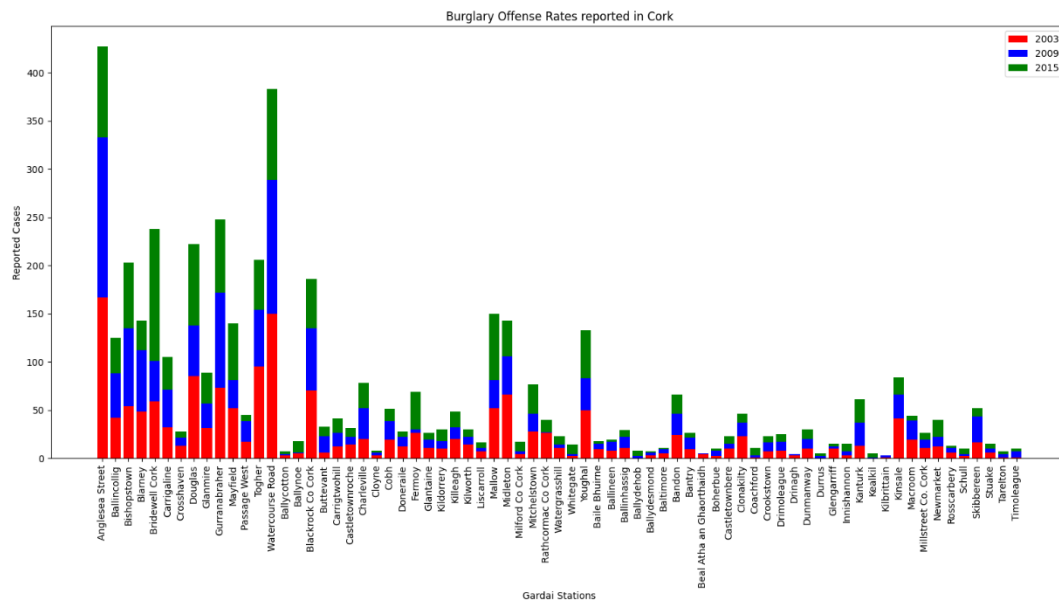


Figure 6 – Drug Crime Cork 2003-2009-2015

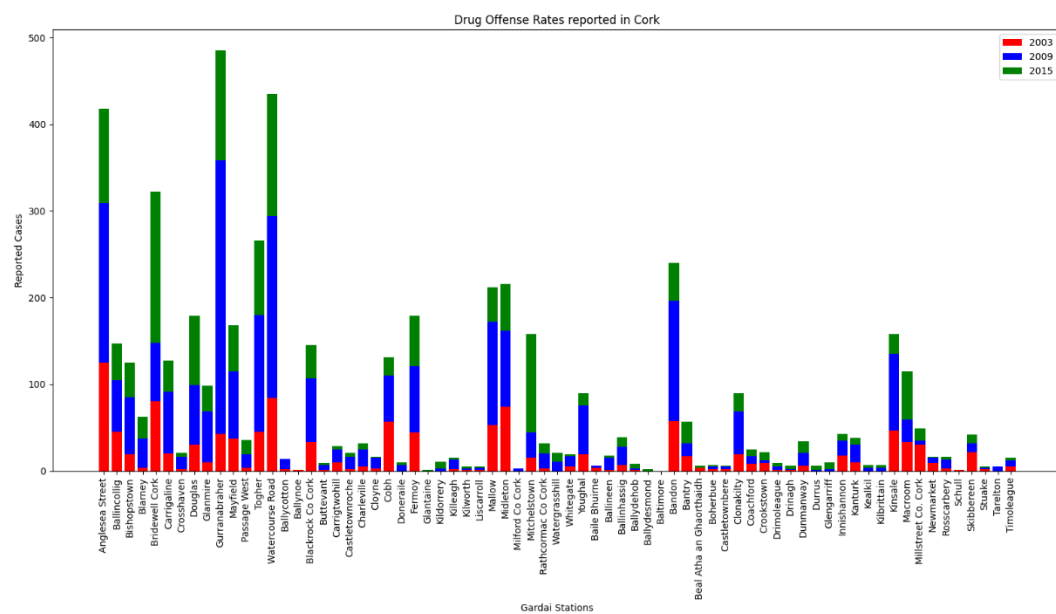


Figure 7 – Drug Crime Dublin 2015 (Seaborn)

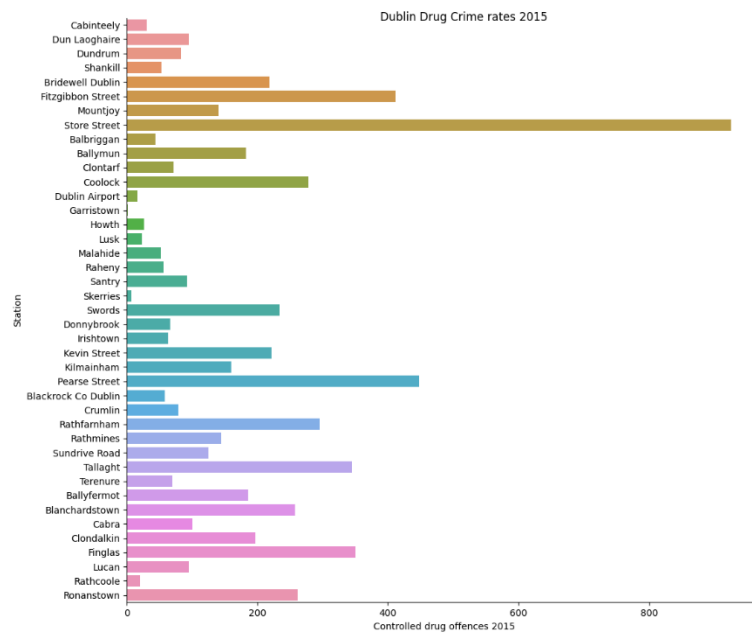
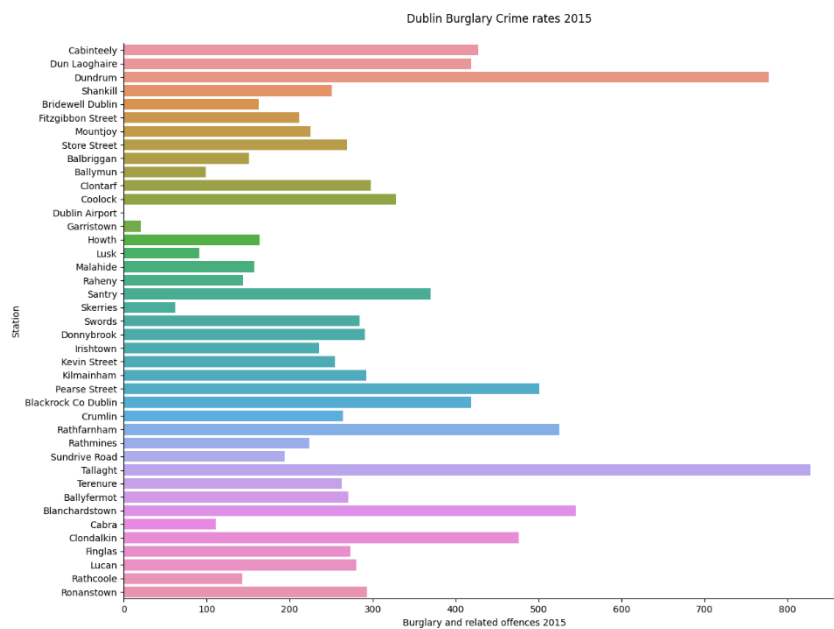


Figure 8 – Burglary Dublin 2015 (Seaborn)





## Insights

- **Galway Crime** : for both Drugs and Burglary you can clearly see from figure 3 & 4 that the vast majority of offenses occurring the Gaillimh Garda Station. This would make sense as this is the primary Garda Station in Galway City. However, I would find it interesting that none of the surrounding areas have any meaningfully comparable figures.
- **Coolock Dublin (Figure 2)** :Drug related offenses in grew significantly between 2003 and 2009, but then receded back to 2003 levels by 2015. It would be worthwhile investigating what the route cause for this is. Perhaps 2009 is an exceptional year.
- **Growth of Drug Crime reported from Store Street Garda Station (Figure 2)** : reported offenses between 2003 and 2009 more than doubled. The 2015 total was then more than double the 2009 figure. This level of growth is alarming and is worth further investigation
- **Cork Crime Dispersion:** based on how I have sorted my final data frames, Cork City related to the first 14 bars in the chart. Although Burglary offense are reasonably flat in terms of their increases across all stations, there has been huge growth in drug offenses in the Cork City divisions since 2003 (Figure 6). There has also been a higher level of growth of drug offenses than burglary offense in the remaining two divisions
- **No obvious link between Burglary rates and drug offense rates (Figure 7 & 8):** Based on the data available to me and what has been used in the above charts, I cannot see any correlation between high levels of drug offenses and burglary rates.