

# UK Crime Dashboard

Scalable analytical review tool for UK crime data analysis

Deniss Ustinovs

## Introduction

The Home Office became increasingly concerned by potential crime trends over the last few years. My task was to analyze the existing crime data or produce a tool which could present the initial data analysis in the simplest way possible. I have decided to produce an excel sheet which could aggregate the existing crime data after its initial processing in SQL. The purpose of this tool is to give the broad overview of the crime data for the whole UK and be a useful starting point for the initial presentation for higher management staff. I will go through the steps that were taken to obtain, cleanse and import the data as well as the method used to build this tool.

## Ingestion of Data

We have decided to use the crime data from the whole UK for the past 5 years. The data was sourced from: <https://data.police.uk/data/>, for all police forces, starting from January 2012 up to September 2016. The data was imported into the Visual studio for the initial profiling, where it was analyzed for the possible abnormalities, such as missing entries or inconsistencies. After obtaining the optimal sizes for the columns, which was done in order to minimize the size of the initial data, it was imported into the SQL database.

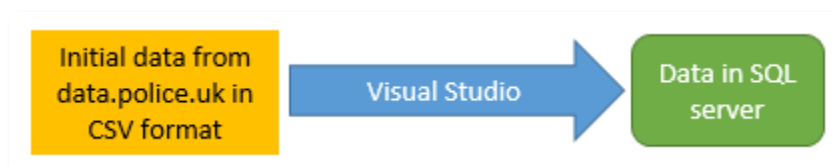


FIGURE 1 INITIAL STAGING OF DATA

## Munging & Wrangling of Data

After the analysis of the initial table imported via Visual Studio, the next step was to normalize the data in order to minimize the total size and make the queries run more efficient. The data was split in 10 different tables with 10 different primary keys which identified the main purpose of the given table. The data was further analyzed for the inconsistencies with blank entries. The SQL query was written which normalized the data and split it in the different tables. (The SQL code used is supplied within the project folder for future references)

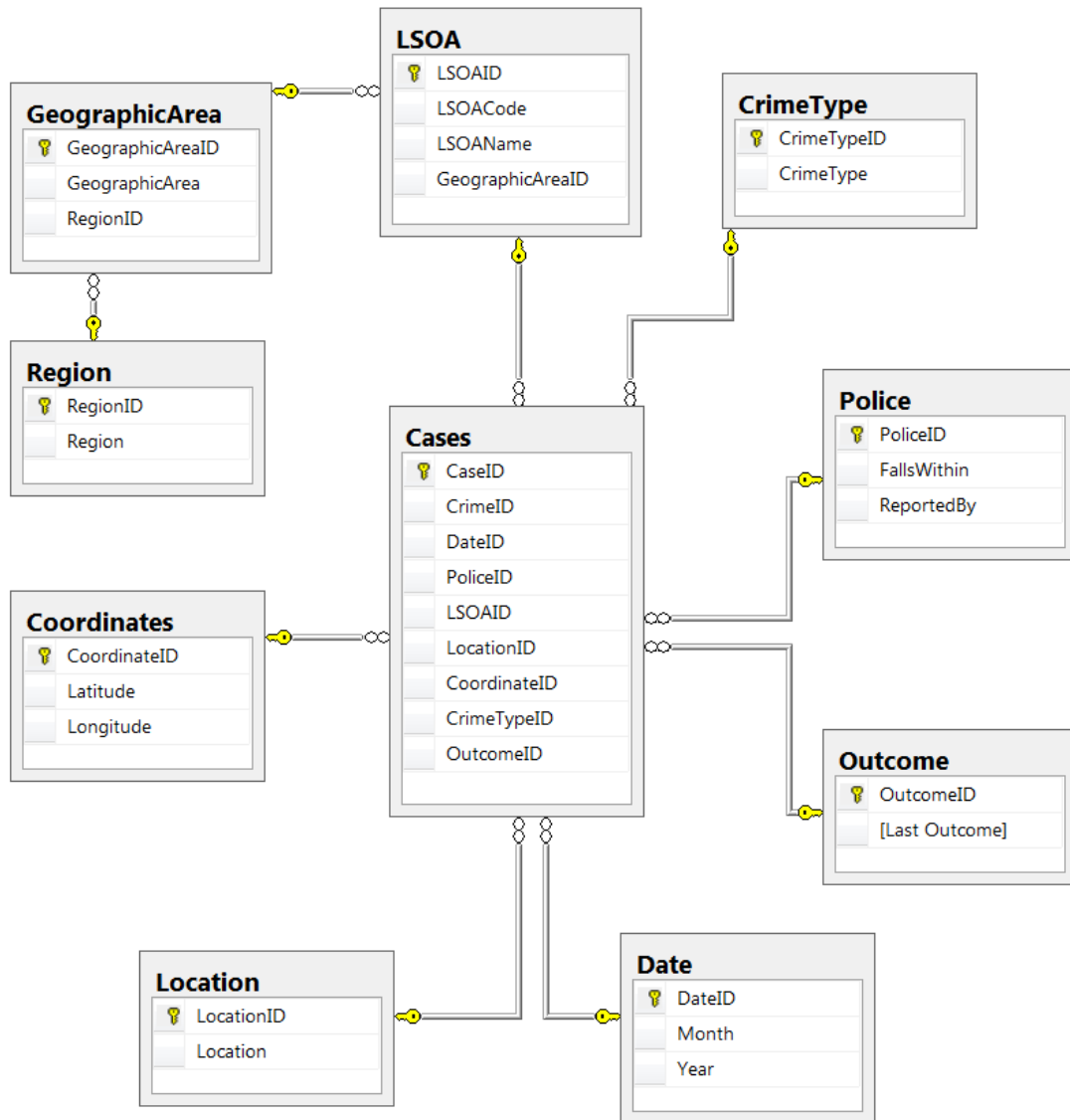


FIGURE 2 DATABASE DIAGRAM

The normalization process helped us to reduce the size of the stored data from 6GB to just under 2GB. It also made it more efficient to run queries because we were only querying the needed tables and entries. Furthermore, the UK population data has also been imported and joined with the initial dataset through the GeographicArea table via the GeographicArea column.

## Dashboard creation

After the analysis of the data, it was decided that only six columns from the crime dataset will be used in the creation of the dashboard:

- **Region** – A general name used to identify the region of the UK
- **County** – Subset of a region, used for more accurate identification of the area
- **CrimeType** – Name of the crime occurred at a certain place
- **Count** – Number of a particular crime cases that happened at certain time and location
- **Year** – Identifies the year when crime happened
- **Month** – Identifies the month when the crime happened

After the normalization, the imported data took only 340 thousand rows in Excel, which is a big difference comparing to 30 million rows which we initially had when we imported the raw data.

One of the ideas of the Dashboard was the possibility to analyze and compare the crime rate in different regions. It is not correct to compare the number of cases due to big difference in population between certain regions, hence a second data set had to be imported. The population data only consisted with GeographicArea and the average population between 2012 and 2015.

## Assumptions

Because this project was given in November, we were unable to get the crime data for the full year 2016, therefore it was decided that if the data from 2016 is being looked at, it will only be compared with the first 9 months of the second data set. For example, if you are comparing the crime rates in Hackney in 2016 with Pembrokeshire in 2015, it will only look at the first 9 months of 2015 in Pembrokeshire. Although different regions have different periodicities in different crimes, it was decided that this option is the best solution for the missing data, weather patterns still plays a big role in certain crime rates.

It was decided that the average population will be used due to the insignificant changes in the population across such a short time period. The Comparison sheet is using the proportion of crime occurrences and the present population in a certain region. It is known that the population density is also playing a big role in crime rates, but this data could only be used with the idea of the uniform distribution of population across the county or region, which is not the case in real world. Population density should be used at higher level of analysis for a smaller region rather than the whole UK.

## Dashboard Features

The Dashboard sheet produces the comparison between the selected year and the previous year for the selected region. This allows us to see the possible trends and the main differences between years on the selected region/county. It is useful to look at the distribution of crimes and their distribution thorough the year.

Year	2015
Region	South West
County	Christchurch

You can leave County field blank in order to look at the whole Region

The Comparison sheet allows you to compare the crime rates in different regions/counties. It can be used for the initial analytical review in the similar way like the Dashboard sheet, but with bigger flexibility due to the option of comparing different regions/counties across different years.



Buttons allow you to filter the crime rates which you want to be displayed for your analysis. It is useful to look into how different crime types correlate with each other on different graphs and gives you the basic idea of crime distribution. The circular buttons allow you to select and deselect all crime types

Due to excessive funds and time after the completion of the initial stage of the project, it was decided that the resources can be used to create the basic analyzing tool that would highlight some interesting points about the selected data. This feature is only presented at the Comparison sheet. The purpose of this feature was to explore the idea of a self-writing report that can automatically extract the needed information and present it in the text format rather than numbers. Unfortunately, it became clear that this feature is requiring much higher budget than the original project had. Therefore, the final result gives the hint of the usability of this feature, and possibly can influence shareholders in investing more resources on this tool.

Burglaries are three times more likely to happen in Ceredigion rather than Brent. If you own a shop in Ceredigion it is more likely to be robbed than in Brent. Your vehicle is many more times

## Conclusion

The main Idea behind this dashboard was to make a scalable and updatable product that can be used for the initial data analysis of UK crime data. The steps undertaken in SQL and Excel are recorded and can be easily implemented when the new crime data gets released, hence the Excel sheet can be easily updated. From the practical point of view, the fact that the dashboard is only 18 MB makes it a very mobile and accessible tool, which is a big improvement from the initial size of the data. The easy readable format and informative graphs used on this dashboard combined with practical reusability make it a useful tool to be presented for senior management officials on a very short notice.

With higher budget this tool can be scaled to perform even harder analytical reviews of the given information. Together with integration of SQL in Excel and addition of other datasets, this can potentially be a self-sufficient analysis tool which would allow a professional analyst to derive interesting insights.

The glimpse on the self-writing report tool makes it an interesting idea which is worth further discussion, although the high quality product would require a lot of resources, it could be a well worth investment which could substitute the analyst team.