**TFM-Crimes-in-Madrid**

A Data Science Project about Crimes in Madrid and its districts.

# Kschool

# Master Data Science ed. VIII

# Isabel Flores Cánovas

# 2019, January

Introduction

In one of the classes we had, we saw an analysis about Crimes in Chicago, and It made me think in a similar exercise here, in Spain. I was being robbed in the Center of Madrid, so I though It could be a good exercise to test the Data Analysis with real Data.

Purpose: Analyse information about crimes in Madrid Districts, to see if there is any association between different types of crimes, districts crimes, and look for any pattern if it is posible (I do not have many data…)

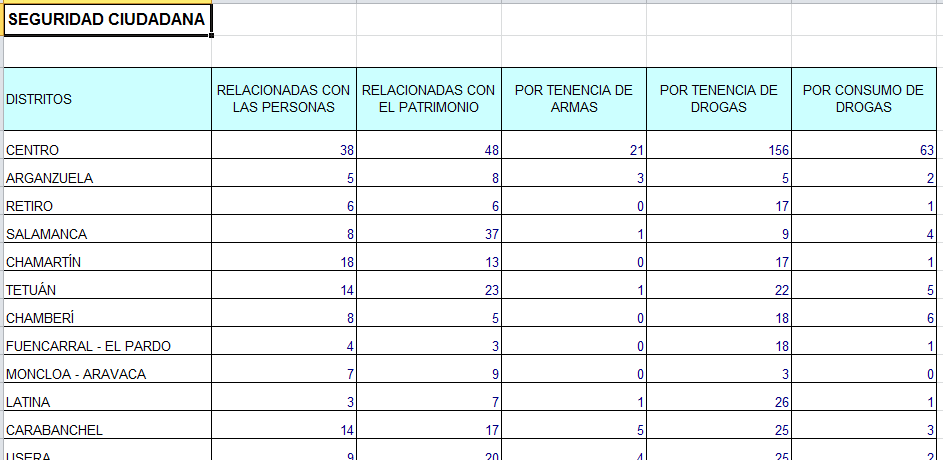
Data

In the city of Madrid, Spain, Local Police has been picking all kind of crimes in the city and writing in an excel file every month from 2014 until now. Then, data are suministrated by Madrid Town Hall Police from this link: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=bffff1d2a9fdb410VgnVCM2000000c205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default>

Every file from the Data (every Excel) contains few sheets (16) about different types of Crimes. I focus on the first sheet, ‘Seguridad’ which contains security crimes as:

* Related to people
* Related to patrimony
* About having weapons
* About having drugs
* Drugs Abuse

I think it is a good summary about the crimes a city as Madrid has, and it allows me to analyse the information in a good way:



Except the field “Distrito” (which is a string) the others fields are numerical data.

Another important point is the dates (month and year) are not in the data but in the file’s name, so I have to integrate that in the Data.

To get that, I have created a dataframe, as a dictionary, which associates file name with the date. I have done this exercise in a R script, ‘Auxiliar.R’, to use it in the main R script. Auxiliar.R exports a csv file with this manual association.

I have to integrate in the main data a code district (numeric one) too, in order to connect my districts with another data briefs. I get this data directly from the Excel, the order how they show the districts in the main table in Excel has the right order to get them sorted.

And, what for? I got some graphical data (shapes from Esri) from:

[https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=46b55cde99be2410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aR](https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=46b55cde99be2410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD)

|  |  |
| --- | --- |
| They content the poligonal form for every district, and the code, so I am able to associate this geografical information to our data. I will show the results at the end, and will allow us to see if there is any neighbour relation between districts, and the only way to find out this is watching graphical data. |  |

Work Methodology

The process has been done in 3 steps:

1. R 🡪Joinning files, clearing them, and addin other data in R script
2. Python 🡪Getting information from the Data in jupiter Notebook (PYTHON)
3. Tableau 🡪Showing and associating geographical information to our data
4. Web site design 🡪 Sharing the conclusions
5. **R.**

The main file I have used is named **1\_0\_GettingTheCrimes.R**. It takes all the Excel files from a directory, ‘/Delitos’, with an especial function. This function removes the first 2 rows in every Excel (they do not have data) and join them adding the name of the original file in a dataframe.

After that, It generates another dataframe which has the association between district and code number.

And last but not list, it imports the association between the name file and the date the file has from a csv file (generated in Auxiliar.R) in a dataframe .

Finally, it merges the dataframes (3 of them) building an only dataframe sorted by the dates. I export it to a csv file, Delitos.csv.

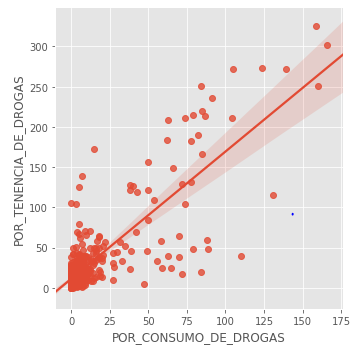
1. **Python.**

Once I have the data, I analyse them in Jupiter Notebook, **2\_CrimeAnalysis.ipynb**. (data are in the same directory as the main script, 2\_CrimeAnalysis.iptnb). So, I get few libraries that help me to visualize graphics in order to get information from the data.

I see the types of my data, the media, the standard deviation, just to have it in case I need it… and also I separate the total information from the districts. With this exercise, I can compare districts, and the total information in the different crimes we have.

At first, I try to get a relation time in the crimes, but it is not clear, so I decide to get the correlation. With this information, I can look for if there is any association between the crimes, and I find it. Drugs crimes are linked. So I work in this relation, and get some graphics (most of them are saved in html files)

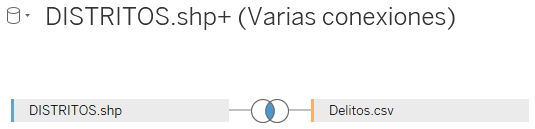
I get the linear regression between the drugs crimes (2 types) and the relation is 100 drugs abuse per 163 drugs possession…



I get also more information about other data in order to obtain any other conclusion.

1. **Tableau**

When I have some conclusions in Python, I analyse Delitos.csv in Tableau **GraphicsAnalysis.twb**. This file is in the same directory as Delitos.csv. I join my data with the graphical ones I have in the directory (‘/Gráfico’) inside tableau.



With this option I can analyse also the relation between time stamp and districts, and get other conclusions added, using geographic data. In fact, I share the results in Tableau public servers.

I have done 7 sheets to get enough information to build 2 dashboards. And with this dashboards I have done 2 stories, and then, I have uploaded them.

1. **Web page**

Once I have analysed this, I have done a web page using pingendo, as we learned in class. I upload the project here: <https://ifcanovas.github.io>

So I can share the data easily and add the html that I have created in python.

Conclussions

1. The stronger correlation is between drugs crimes: as I show in fig.1.most of drugs data are related (when drugs possesion increases, also increases drugs abuse).

It has sense because for drugs abuse it is necessary that someone have them to sell them.

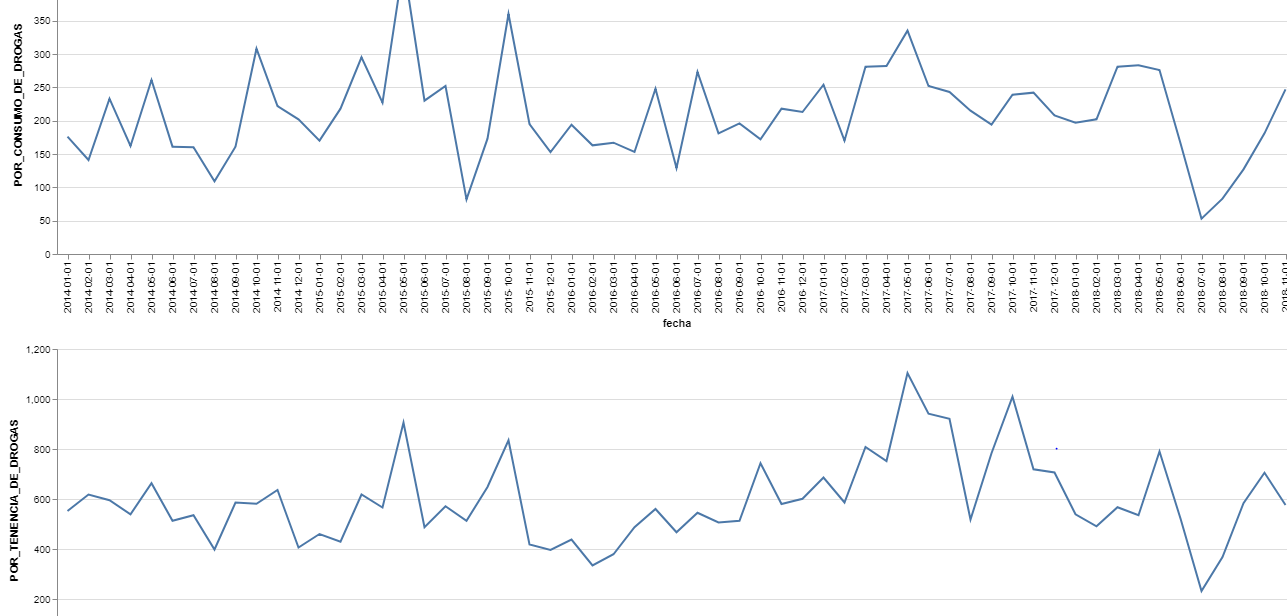
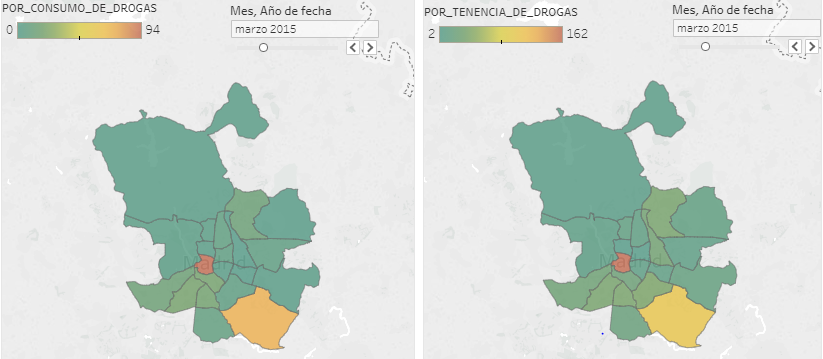


Fig.1

If we show it in a geographic way (2015 March as an example) we can see that there is a relation between both of drugs crimes, and also between districts. In this example, South and East Districs are associated:



Depends of the month and year, the relation changes… Fig.2

<https://public.tableau.com/profile/isabel1894#!/vizhome/GraphicsAnalysis/DrugsAssociationCrimes?publish=yes>

1. Police Raids in the center of the city, every Christmas time since 2015, to look for weapons (Crime weapons in date stamp)

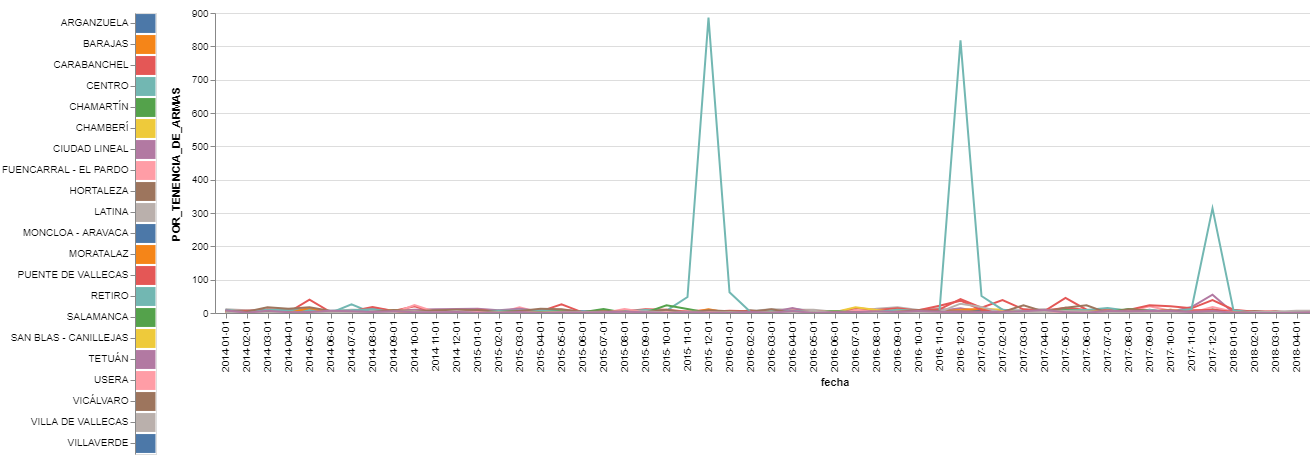
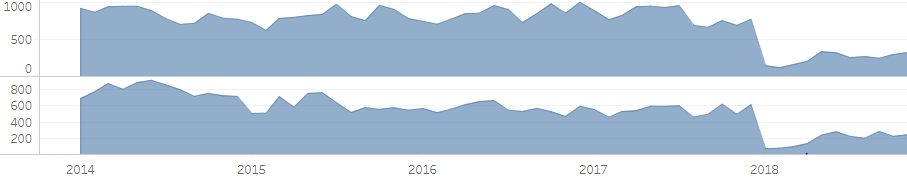


Fig3.

It has sense, because the center of the city is crowded every Christmas time, so, it is clear that Police has to work more to be sure they have a save center.

1. Something happended in January 2018, because patrimony crimes, as people crimes, decreased a lot:



Maybe there was a political decision, because Police were working a lot the year before, and the labour conditions weren’t as good as they wanted. They were negotiating labour conditions at the first of the year.

<https://autonomico.elconfidencialdigital.com/articulo/madrid/Indignacion-Policia-Municipal-Madrid-Navidad/20171222144556020139.html>

<https://public.tableau.com/profile/isabel1894#!/vizhome/GraphicsAnalysis/OtherCrimes?publish=yes>

Thank you!