

# Memory

## TFM-Kschool-Madrid-Traffic

### Introduction:

The objective of this project is to analyze the evolution of Madrid traffic after the implementation of Madrid Central, a restriction measure started on December 2018 that limits partially the access with private vehicles to a delimited area of Madrid.

I decided to perform this project because the following reasons:

- It is a current topic which is always interesting, and the final output can add value.
- It was a measure a bit controversial that was not always well accepted by all the citizens.
- Traffic congestion is a continues problem not solved in cities.
- It allows to use statistical methodologies learned during the master course.

The state of art: traffic analysis is becoming more and more common on the daily news, but it is more oriented usually to highway status. Nowadays, cities are transforming to smart cities thanks to the use of devices to get data that is ready for analysis. The idea is to take advantages more and more of these data we get every second to perform analysis and get conclusion that can make our cities better places to live in.

### Raw data:

Regarding data, the project is based mainly on historical data on traffic.

These are the main variables of the traffic data:

- Id: is the unique identifier of the device measure point.
- Fecha: is the time and date observations measure were taken
- Tipo\_elem: is the type of device. It may be M-30 an urban point.
- Intensidad: is the intensity of the device in vehicles per hour in the previous 15 minutes.
- Ocupacion: is the % occupation time of the point in the previous 15 minutes.
- Carga: road load of the device. It depends on several factors as intensity, occupation and the capacity of the road where the devices are located (the % of use indicates the level of use of the road)
- Vmed: average speed of the vehicles in the past 15 minutes in km/h. It only works for M30 points.
- Error: indicates if there has been any error in the measure. Possible values are N for no errors, E when the quality parameters of the measures are not optimum, and S if any of the measures taken is wrong and has not been integrated.
- Periodo\_integracion: is the number of measures taken for the integration period.

These are the main variables of the devices data:

- Id: is the unique identifier of the device measure point.
- Tipo\_elem: is the type of device. It may be M-30 an urban point.
- Nombre: denominación del punto de medida
- Latitud: latitude coordinates (from old files it is called y)
- Longitud: longitude coordinates (from old files called x)

The data used for the project has been downloaded from the open portal “Portal de datos abiertos del Ayuntamiento de Madrid”. See links below:

Devices location: files with the location of the devices per month since 2017

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vnextoid=ee941ce6ba6d3410VgnVCM1000000b205a0aRCRD&vnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

Historical Traffic Data since 2013: files with traffic information per month since 2013

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vnextoid=33cb30c367e78410VgnVCM1000000b205a0aRCRD&vnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vnextfmt=default>

Live Traffic: file with live traffic information

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vnextoid=02f2c23866b93410VgnVCM1000000b205a0aRCRD&vnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

Calendar: file with calendar information as working days, weekends and bank-holidays

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vnextoid=9f710c96da3f9510VgnVCM2000001f4a900aRCRD&vnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vnextfmt=default&reloadCaptcha=true>

## Methodology:

1. **Data Acquisition:** mentioned above, from Madrid Portal Open Data. Files with information about traffic variables on different locations
2. **Data Cleaning and Preparation:** analyzing and transforming the data by adding columns, dealing with Nan, transforming variables, dropping no relevant observations and plotting.
  - a. Programming languages: Python3 – Notebooks
  - b. Notebooks developed:
    - i. 0. Notebook 0 Data Acquisition: Downloading files with the data
    - ii. 1. Notebook 1 Data Cleaning and Preparation (1): Devices Location
    - iii. 2. Notebook 2 Data Cleaning and Preparation (2): Traffic Data
    - iv. 3. Notebook 3 Analysis (1): Time Series
    - v. 4. Notebook 4 Analysis (2): Time Series
    - vi. 5. Notebook 5 Visualization and Conclusions
3. **Analysis:** focus on the variable ‘intensity’
  - a. Statistical methodologies: for the Time Series analysis and forecasting process I have used the Python module **statsmodels** to apply seasonal ARIMA (SARIMA) models to make my time series stationarity. Deal with time series components as trend, seasonality and residuals. I have selected critical locations around Madrid Central and done 2 type of analysis:
    - i. Prediction of the variable traffic intensity before Madrid Central and comparison vs reality
    - ii. Forecast of the variable traffic intensity including in my models Madrid Central observations
4. **Visualization:**
  - a. In the Notebook 1 there is an interactive map with the location of the devices used for identifying the ones more relevant around Madrid Central

- b. in the Notebook 5 I insert a summary a visualization part composed of plots from the module **matplotlib** of the devices analyzed including the time series. Finally, some conclusion may be found at the end of the notebook.

### **Summary of main results:**

As main results we observe that in general the traffic intensity has decreased significantly in most of devices analyzed. Of course. it is still soon to provide clear results since Madrid Central is running from few months ago.

### **Conclusions, applications and future work:**

The goal of this project was to provide a detail of the traffic evolution in the surrounding areas of Madrid Central. It is clear the traffic intensity shows signs of slowing down tender, but we can't get final results since more time is needed to have more data to analyze for that purpose. As following steps, that I would like to do within this initial approach on the project, we can consider the following ones the most relevant and the most practical for the city Government:

- Traffic congestion prediction: be able to spot where and when is going to be traffic jam
- Public transport utilization: predict the demand of the public transport infrastructures

Other analysis as impact on the sales of the stores located in the area, price evolution or parking renting areas.

Author:

Luis López Gállego

Kschool

16/05/2019