

# **Análisis de las precipitaciones de Barcelona**

## *Abstract*

La predicción del tiempo es un tema de interés general por su efecto sobre las vidas, tanto por temas de ocio como profesionales. Es una ciencia que se ha ido sofisticando con modelos matemáticos cada vez más precisos. A falta de tener acceso a estos modelos, se intentará analizar y utilizar los datos disponibles para encontrar un modelo de predicción de lluvia.

En este caso será una estudio sobre el clima de Barcelona con la dificultad de que es mediterráneo, cuyo régimen de lluvias es impredecible y que con el cambio climático todavía se está haciendo más complicada cualquier predicción.

## Índice

<b>1. Características climáticas de Barcelona</b>	<b>3</b>
Precipitaciones	3
<b>2. Estructura del proyecto</b>	<b>3</b>
Objetivo	3
Método	3
Diseño de la base de datos	4
<b>3. Análisis descriptivo</b>	<b>4</b>
Análisis histórico	4
Análisis de los factores	6
<b>4. Análisis predictivo</b>	<b>9</b>
Análisis no supervisado	9
Análisis supervisado. Clasificación	10
<b>5. Conclusiones</b>	<b>11</b>
<b>6. Referencias</b>	<b>12</b>

# 1. Características climáticas de Barcelona

La ciudad de Barcelona se caracteriza por un clima mediterráneo, un clima que se caracteriza por tener unas estaciones muy marcadas. Con una media de precipitaciones de entre los 400 y 1000 mm, es un clima seco sobre todo durante el verano. Las precipitaciones suelen concentrarse en primavera y otoño.

La temperatura media ronda los 18°C, con veranos secos y cálidos e inviernos moderados y poco húmedos.

De las diferentes estaciones meteorológicas se ha seleccionado la de l'Observatori Fabra cuyas mediciones serán ligeramente diferentes de las otras estaciones por estar a 411 metros sobre el nivel del mar (el resto se sitúan prácticamente a nivel del mar)

## Precipitaciones

Una característica típica del clima mediterráneo es la irregularidad de las lluvias, que las hace poco predecibles. Se puede pasar de un largo período de sequía a precipitaciones puntuales y muy intensas. El Cambio climático acentúa esta irregularidad de la cantidad, distribución e intensidad de las lluvias.

# 2. Estructura del proyecto

## Objetivo

El objetivo de este proyecto es encontrar patrones de precipitaciones para poder describir las características que los diferencian y también poder predecir con un poco de antelación si va a llover y cuánto.

## Método

Para la obtención de datos se han descargado las bases de datos que ofrece la AEMET (Agencia Estatal de Meteorología) en formato JSON y se ha realizado un web scarping (Python Selenium) de las páginas de TuTiempo para completar la información.

La base de datos resultante, que contiene los datos entre marzo de 1983 y agosto de 2021, ha sido la base para hacer un análisis descriptivo para entender mejor el patrón de precipitaciones.

Se han utilizado las librerías Pandas, Numpy, Seaborn, Matplotlib, Sklearn para Python para hacer los diferentes análisis y gestión de datos.

Finalmente se han usado dos métodos de aprendizaje, supervisado y no supervisado, para encontrar formas de predicción y patrones

## Diseño de la base de datos

Este proyecto ha requerido de una base de datos con una sola tabla que agrupa los atributos básicos para predecir y también describir los parámetros que pueden influir en las precipitaciones.

Para el análisis no supervisado se han utilizado los datos del día y para el análisis supervisado los datos de los tres días anteriores para poder intentar hacer una predicción.

Ne se ha precisado de una base de datos relacional.

## 3. Análisis descriptivo

### Análisis histórico

El régimen de lluvias en Barcelona se rige por un patrón irregular y se puede ver claramente en los gráficos temporales que muestran la cantidad de lluvia a lo largo de 38 años.

En los tres gráficos, tanto la suma de lluvia por día, mes y año (figuras 1, 2 y 3), no se ve un patrón claro de precipitaciones.

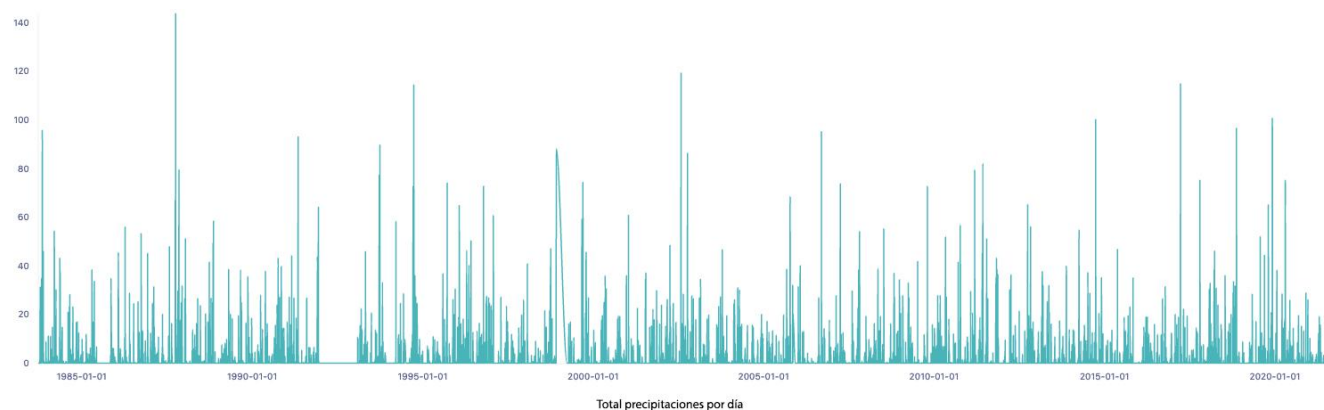


figura 1. Total diario de precipitaciones

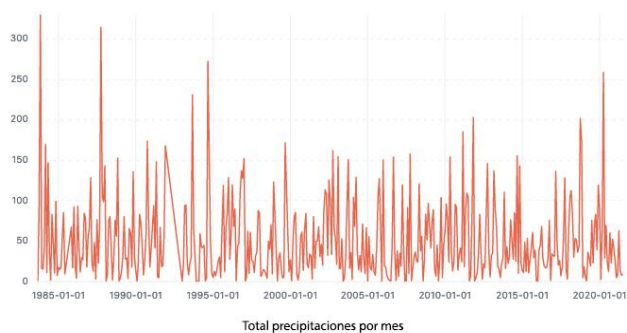


figura 2. Total mensual de precipitaciones

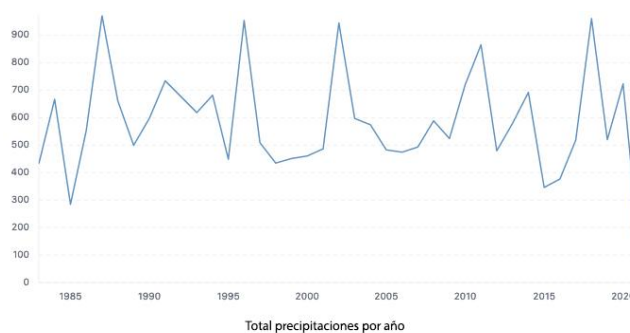
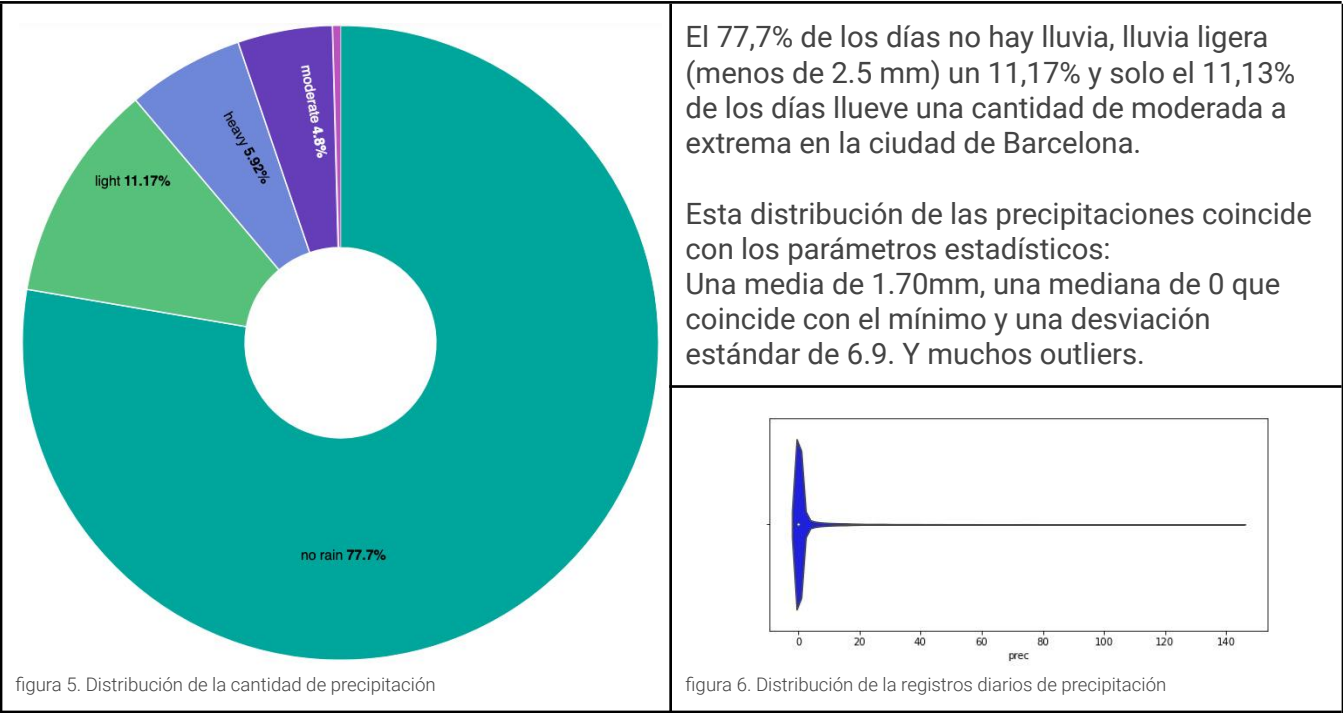


figura 3. Total anual de precipitaciones

La distribución de las lluvias a lo largo del año sí muestra que los episodios extremos coinciden en otoño como es propio de los climas mediterráneos, y en menor medida en primavera, aunque en ambas estaciones la media de lluvia por mes es mayor que el resto del año. El porcentaje de días secos (con referencia al año) se distribuye bastante uniformemente.

Mes	Media lluvia (mm)	Precipitaciones violentas		Días secos (%)
		Días	Porcentaje	
Enero	1,58	4	7,5	7,8
Febrero	1,27	0	0	7,6
Marzo	1,66	4	7,5	8,6
Abril	2,04	6	11,3	7,9
Mayo	1,75	2	3,7	8,4
Junio	0,91	2	3,7	8,9
Julio	0,87	4	7,5	9,8
Agosto	1,08	0	0	8,7
Septiembre	2,54	12	22,6	7,5
Octubre	2,9	11	20,7	7,8
Noviembre	2,44	5	9,4	7,8
Diciembre	1,45	3	5,6	8,6

figura 4. Distribución de la precipitación



## Análisis de los factores

Entre los muchos factores que afectan a la posibilidad y cantidad de lluvia se encuentran la presión atmosférica, la dirección y fuerza del viento y la humedad. En los siguientes gráficos se muestra la relación entre estos factores y la media o el total de lluvia.

Es significativa la relación entre la dirección del viento y la cantidad de lluvia (figuras 11 y 12) debido a que en un rango ancho de direcciones no hay precipitación. Para una predicción este factor puede ser relevante. (notar que es un atributo cíclico y que el valor 100 es inmediato al 1).

La formación de nubes de lluvia también está asociada a bajas presiones, y, como se aprecia en la figura 8, la cantidad de precipitación tiende a valores bajos y, en dos de los más bajos están las mayores precipitaciones.

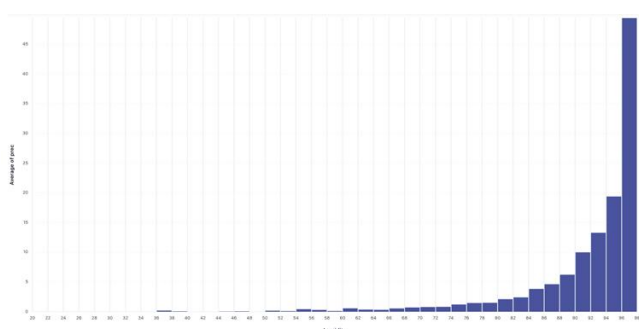


figura 7. Media de precipitación por humedad

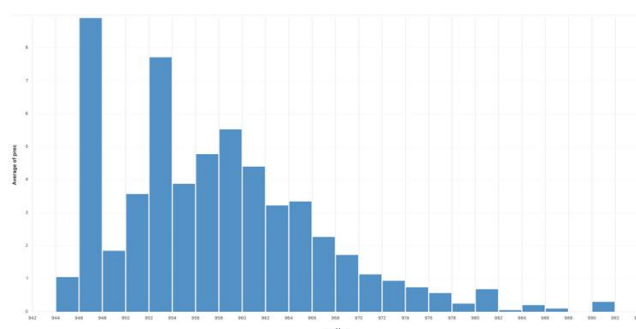


figura 8. Media de precipitación por presión atmosférica

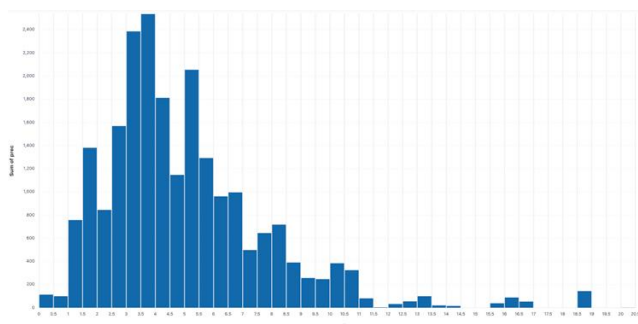


figura 9. Suma de precipitación por velocidad media del viento

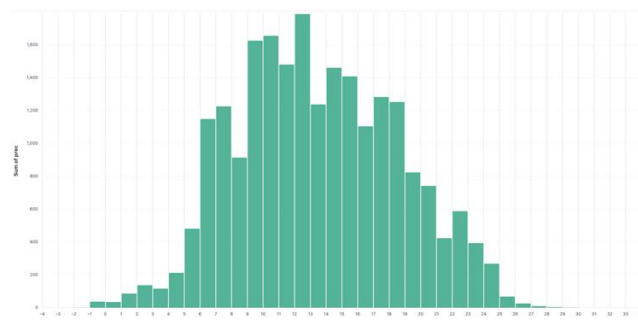


figura 10. Suma de precipitación por temperatura media

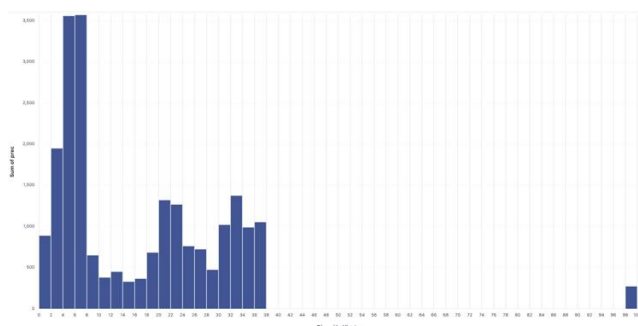


figura 11. Suma de precipitación por dirección del viento

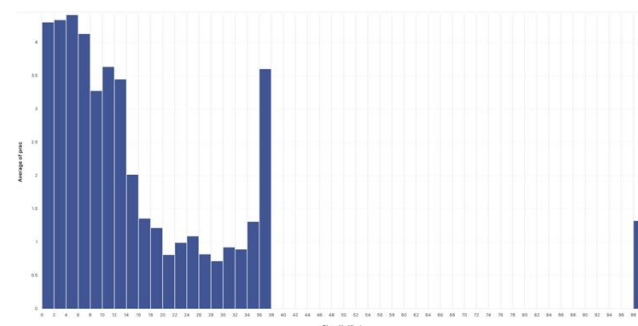


figura 12. Media de precipitación por dirección del viento

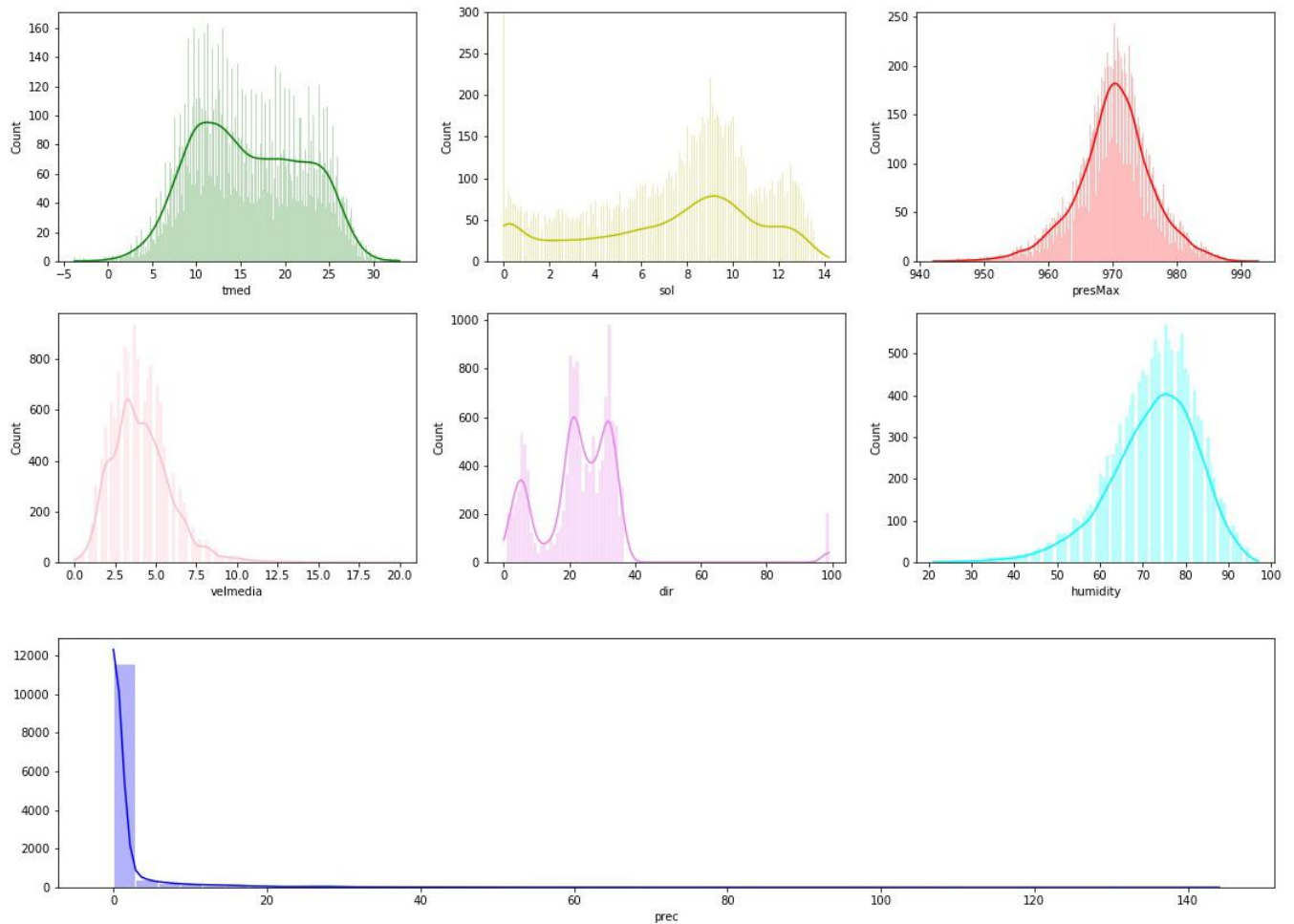


figura 13. Distribución de los atributos y de la precipitación

Todos los atributos muestran una distribución suficientemente normal, algunas un poco desviadas, menos las horas de sol, la dirección del viento que muestra tres picos muy marcados.

Para la predicción de las precipitaciones se tienen que considerar los valores de los días anteriores ya medidos. Estos gráficos intentan mostrar la relación entre estos atributos previos (mediciones del día anterior -1, dos días antes -2 y tres -3) y la precipitación.

La presión atmosférica de los días anteriores, sobretodo en el último día, tiene alguna correlación

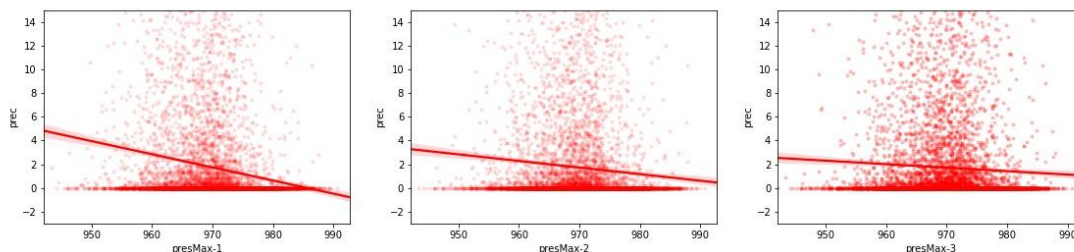


figura 14. Correlación tres días anteriores entre Presión atmosférica y lluvia

La dirección del viento de los días anteriores, sobretodo en el último día, tiene alguna correlación así como la humedad, más acentuado el día previo

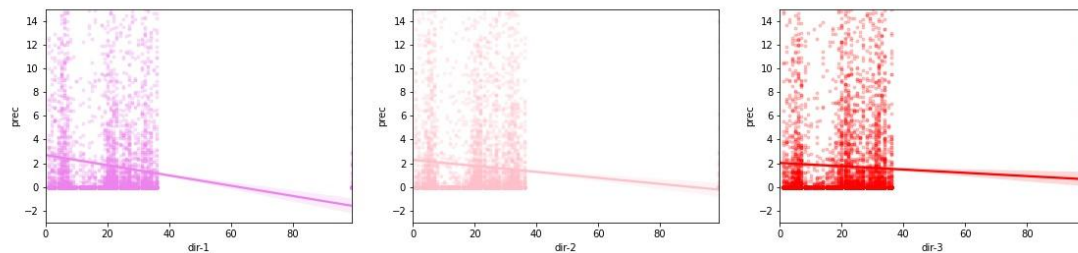


figura 15. Correlación tres días anteriores entre dirección del viento y lluvia

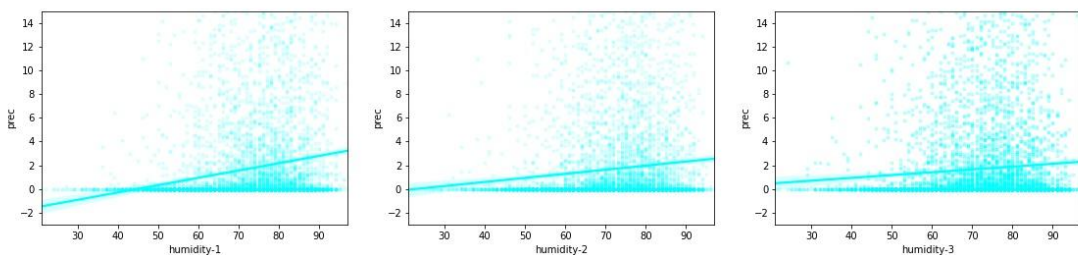


figura 16. Correlación tres días anteriores entre humedad y lluvia

El factor precipitación de los días anteriores también tiene una cierta correlación

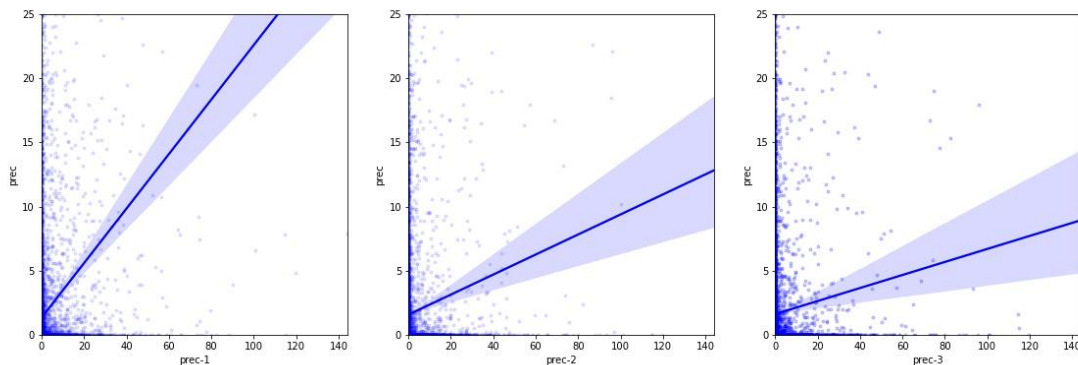


figura 17. Correlación entre la entre cantidad de lluvia tres días anteriores y lluvia



## 4. Análisis predictivo

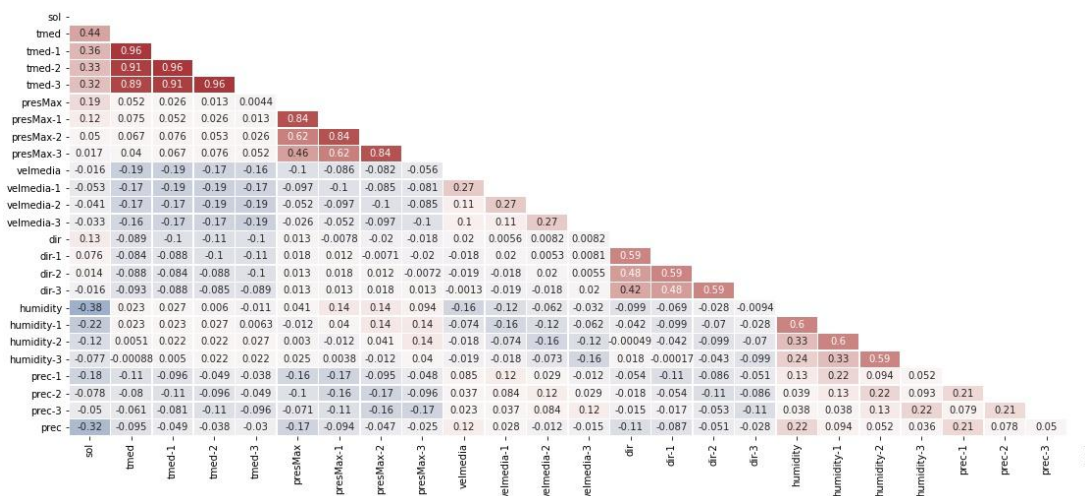


figura 18. Correlación entre la entre los atributos

## Análisis no supervisado

La base de datos para el análisis no supervisado es la que contiene los datos del mismo día y se han eliminado los atributos altamente correlacionados:

- fecha se ha simplificado a meses por la influencia que tienen en la distribución de la lluvia
- temperatura diaria media
- horas de sol
- presión atmosférica máxima'
- velocidad media del viento
- dirección del viento
- humedad
- cantidad de precipitación total diaria

Mediante el método K-means se han obtenido 4 clústers algo diferenciados con una silhouette baja de 0.333 que indica ejemplos que se superponen.

La influencia del gran porcentaje hace que los clústers tengan pocas características destacables.

Los clústers se han dividido por meses, 4 meses por clúster menos el n.2

Clúster 0	Clúster 1	Clúster 2	Clúster 3
marzo 910	julio 1034	meses muy repartidos	diciembre 907
mayo 896	junio 926		octubre 836
abril 858	agosto 915		enero 827
febrero 798	septiembre 797		noviembre 823

**Clúster 2:** Agrupa la mayoría de los días muy lluviosos con una media de 8.9 en comparación a la casi nula del resto y una mediana (3.8) también diferente a 0 como el resto

## Análisis supervisado. Clasificación

Para el análisis supervisado se han utilizado los atributos de los tres días anteriores, menos la temperatura media por estar altamente correlacionada.

La fecha también ha sido simplificada a meses.

Como valor target se ha utilizado la precipitación dividida en 5 valores, dry, light, moderate, heavy y violent.

Con LazyPredict (en kaggle) se ha encontrado el método con mejores niveles de accuracy y F1 Score. Con los resultados se ha escogido Random Forest Classifier como método de clasificación.

La gran desventaja de esta base de datos es el porcentaje altísimo, más de  $\frac{3}{4}$  partes, de un solo valor, dry, que se ha compensado sobrealimentando los otros valores con SMOTE.

Los resultados parecen muy positivos por su alta accuracy, pero es el resultado normal porque es muy probable acertar ya que el 76% de los valores corresponden a uno solo, como ya se ha comentado. Aunque se ha podido corregir, por el gráfico de confusión (figura.19) se una gran cantidad de falsos positivos y falsos negativos

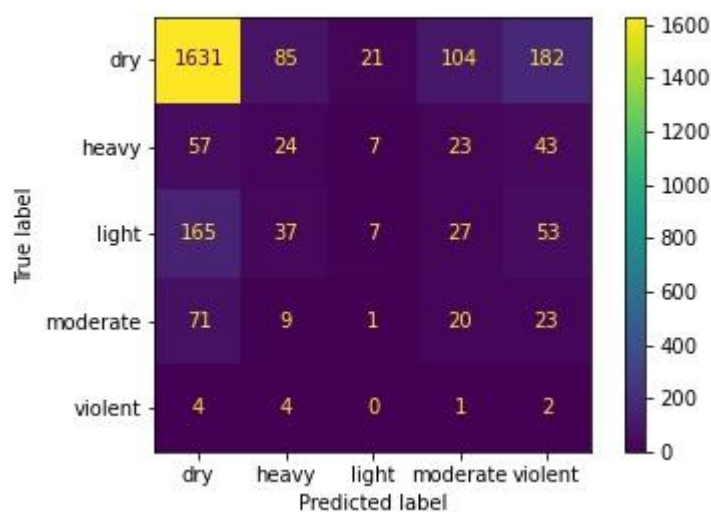


figura 19. Gráfico de confusión

## 5. Conclusiones

Barcelona tiene un clima mediterráneo y éste se destaca por tener un régimen de lluvias muy impredecible, con niveles de precipitación anual acumulada notable pero que está mal repartida a lo largo del año con pocos episodios de lluvias torrenciales.

Más de un tercio de los días son secos y un 10% más con menos de 2.5mm de lluvia como se ve en la figura 5.

Además, las precipitaciones se concentran en otoño especialmente y en menor grado en primavera.

El resto de indicadores, como humedad, temperatura y presión atmosférica son más estables con sus oscilaciones predecibles a lo largo del año.

Con el análisis descriptivo se ha hecho evidente la disparidad de datos de las cantidades de lluvia diaria, con muchísimos casos puntuales de grandes cantidades acumuladas (outliers).

Con los datos históricos obtenidos de AEMET, que se han ampliado con los datos de los niveles de humedad de Tutiempo, se han buscado métodos de análisis para describir y predecir las precipitaciones pero la climatología es una ciencia que requiere de datos mucho más sofisticados para hacer predicciones como se ha podido comprobar con el análisis de clasificación.

Los datos se han ampliado con las mediciones de los días anteriores pero aún así no es posible afirmar que se ha conseguido un resultado resaltable.

Con el análisis no supervisado tampoco se ha descubierto relaciones no conocidas previamente como que la distribución de lluvias está muy repartida a lo largo del año y solo que hay ciertos períodos con una mayor posibilidad de lluvias torrenciales.

Con las figuras 11,12 y 13 se podría concluir que hay dos direcciones del viento que influyen de manera significativa en las precipitaciones. De los mayores episodios de lluvia, más de un 50% de viento provenía de entre las direcciones 2 y 7, aunque el mayor registrado la dirección era 36.

## 6. Referencias

Clima mediterrani

<https://peremeteo.com/climes/clima-mediterrani>

El Clima de la Mediterrània

<http://www.xtec.cat/esc-cefax/Ciencies/clima/clima.htm>

AEMET

<http://www.aemet.es>

TuTiempo.net

[www.tutiempo.net](http://www.tutiempo.net)

Repositorio GitHub

[https://github.com/psansbcn/clima\\_barcelona](https://github.com/psansbcn/clima_barcelona)