

Prueba técnica

Curso: "Data Science con Python"

Análisis del histórico de precipitaciones en Barcelona.

Autor:

Julio Abril Cuesta

Tutor:

Kevin Mamaqi

| | |
|-------------------------|----------|
| Resumen: | 3 |
| Enunciado: | 4 |
| Plan de trabajo: | 5 |
| Conclusiones: | 6 |

1. Resumen:

Este documento representa el plan de trabajo de la prueba técnica del curso de data science con python impartido por IT Academy.

El objeto de la prueba técnica es realizar un análisis de los datos históricos de precipitaciones en la ciudad de Barcelona, llevando a cabo las fases de web scraping de los datos, creación de una base de datos relacional con MySQL, análisis exploratorio y aplicación de métodos de aprendizaje automático.

2. Enunciado:

Objetivo:

Hacer un análisis y recomendación sobre las precipitaciones de la ciudad de Barcelona basados en sus datos históricos.

Guión:

Obtener los datos necesarios y estructurarlos. (<https://datosclima.es/>).

Guardar los datos de manera estructurada en formato SQL.

Realizar un plan sobre (máximo una página) sobre el tipo de análisis a realizar y cómo se presentarán los resultados.

Utilizar PANDAS, NUMPY, MATPLOTLIB (u otras librerías) junto a un cuaderno JUPYTER para la presentación del trabajo realizado. Subir el resultado final a un repositorio privado de Github y compartirlo una vez finalizado con el tutor.

Tiempo estimado:

72 horas.

Debe tener:

El plan sobre el análisis a realizar.

Incluir un análisis supervisado y uno no supervisado.

3. Plan de trabajo:

El análisis de las precipitaciones tendrá los siguientes documentos.

Notebook 1. Obtención de los datos y creación de una base de datos MySQL.

En este documento se hace web scraping de la página <https://datosclima.es/> con la librería Selenium y una vez extraídos los datos se ingestan en una base de datos relacional MySQL.

Notebook 2. Análisis exploratorio de los datos.

El análisis exploratorio de los datos tendrá la siguiente estructura:

1. Análisis univariable. Exploración de las distribuciones y parámetros estadísticos de las variables del dataset.
2. Análisis multivariable. En este apartado voy a explorar cómo se comportan las variables del dataset en función de otras variables. Aquí analizaremos las siguientes:
 - a. Correlaciones de las variables con ayuda de la matriz de correlaciones.
 - b. Visualizar de manera global la relación entre variables con la ayuda de gráficos de dispersión.
 - c. Visualizaremos como se distribuyen las variables respecto a si hubo lluvia o no hubo lluvia.
3. Análisis de la serie temporal. Puesto que estoy trabajando con una serie temporal voy a agregar los datos por meses, teniendo total de precipitaciones acumuladas mensuales y horas de insolación, temperaturas, presiones y velocidad medias mensuales.

Notebook 3. Modelos de aprendizaje supervisado y no supervisado aplicados a los datos meteorológicos.

El siguiente paso será aplicar un modelo de aprendizaje no supervisado (KMeans) a nuestro dataset.

A continuación construiré un modelo de clasificación que ayude a predecir si habrá lluvia o no en función de los datos meteorológicos.

Notebook 4. Predicción de la serie temporal con la librería Prophet.

Por último voy a construir un modelo de predicción de la cantidad de lluvia mensual con la librería Prophet.

Todos los documentos se pueden encontrar en el repositorio de github:

[JuAC-ITAcademy/Precipitaciones \(github.com\)](https://github.com/JuAC-ITAcademy/Precipitaciones)

Conclusiones:

Se ha analizado el histórico de precipitaciones en Barcelona de los últimos treinta años (desde enero de 1992 a junio de 2021).

Junto con las precipitaciones diarias (litros / m²) se han analizado los valores registrados de horas de insolación, temperaturas, presiones y velocidad media del viento.

Del análisis de los datos se han recogido las siguientes conclusiones.

- En los **últimos 30 años en Barcelona ha llovido 2.234 días**, lo que supone un 20,8% de los días o un expresado de otra manera **un día de cada 5 de promedio**.
- Barcelona ha tenido una mediana de **horas de sol** en los últimos treinta años de **7,4 horas**.
- Las **temperaturas** en barcelona son **suaves** con un valor mediano de **entre 20 y 12 °C**.
- La **precipitación diaria promedio es de 1,5 litros / m²**.

En la figura 1 se puede observar las distribuciones de estas variables en Barcelona en los últimos treinta años.

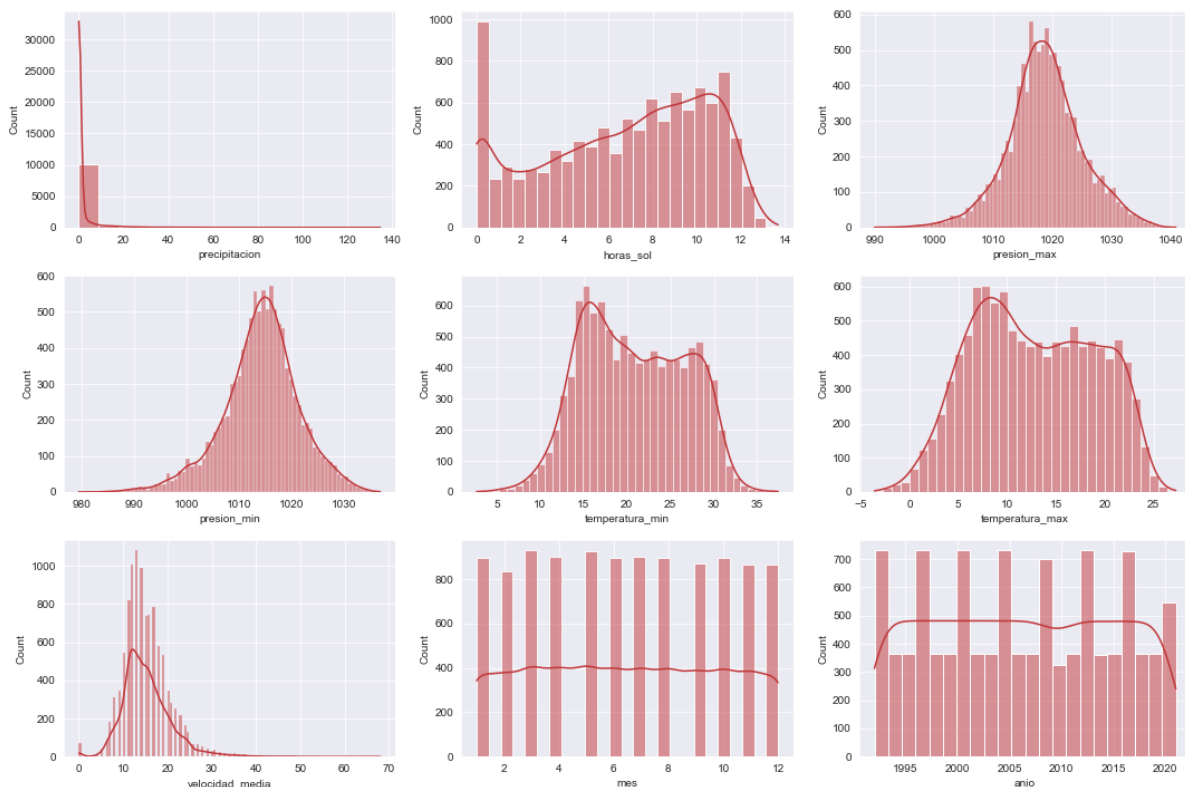


Figura 1

Si tenemos en cuenta las relaciones entre variables:

- Las temperaturas y presiones máxima y mínima diaria están muy correlacionadas (0.95).
- Las horas de insolación tienen una correlación media positiva con la temperatura. Es decir que **cuando aumentan las horas de sol tienden a aumentar las temperaturas.**

Se pueden ver tendencias como que las temperaturas tienden a ser mayores cuando hay más horas de insolación y que Junio, julio, agosto y septiembre son los meses más calurosos.

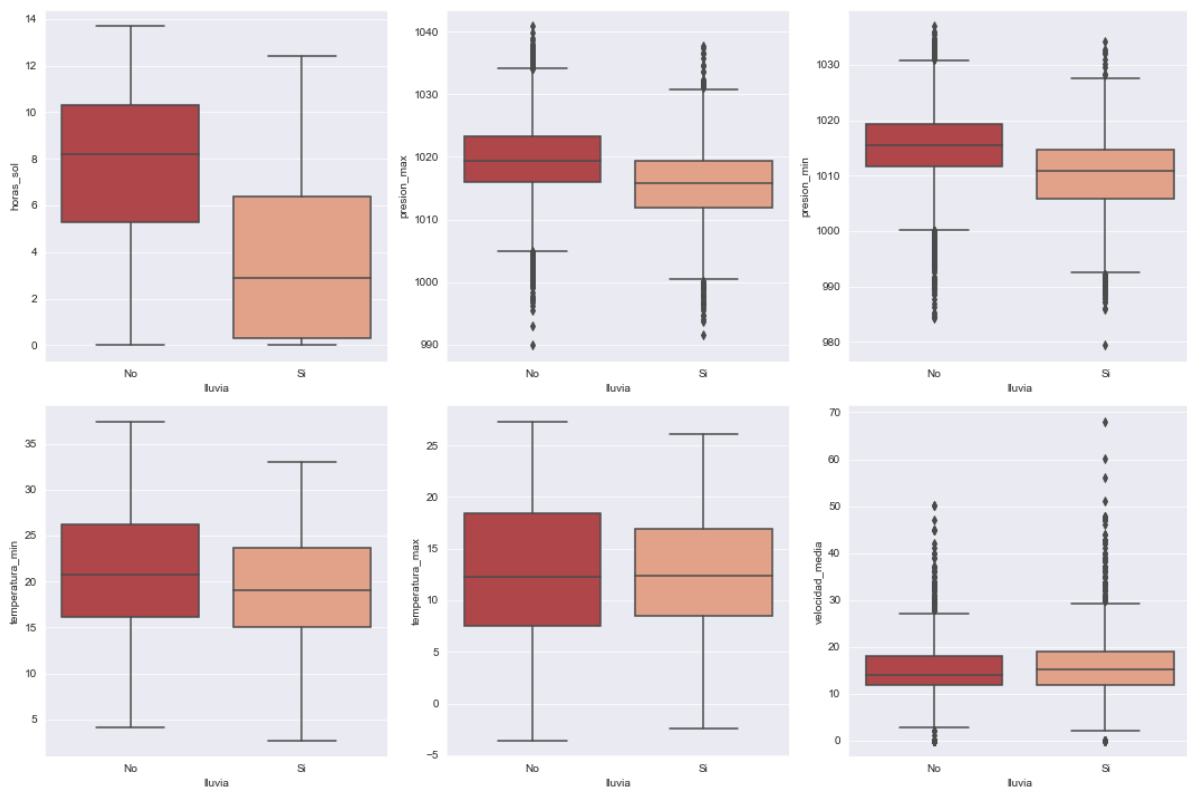


Figura 2

De la figura 2 podemos observar como en general se aprecia mucha variabilidad en los datos. A pesar de ello se aprecian los siguientes patrones:

- Existe una tendencia a **menor número de horas de insolación los días que hay lluvia.**
- Las **presiones tienden a ser menores en días con lluvia.**
- Respecto a temperaturas y velocidad del viento no se observan patrones claros.

Analizando las precipitaciones mensuales promedio durante el periodo podemos ver:

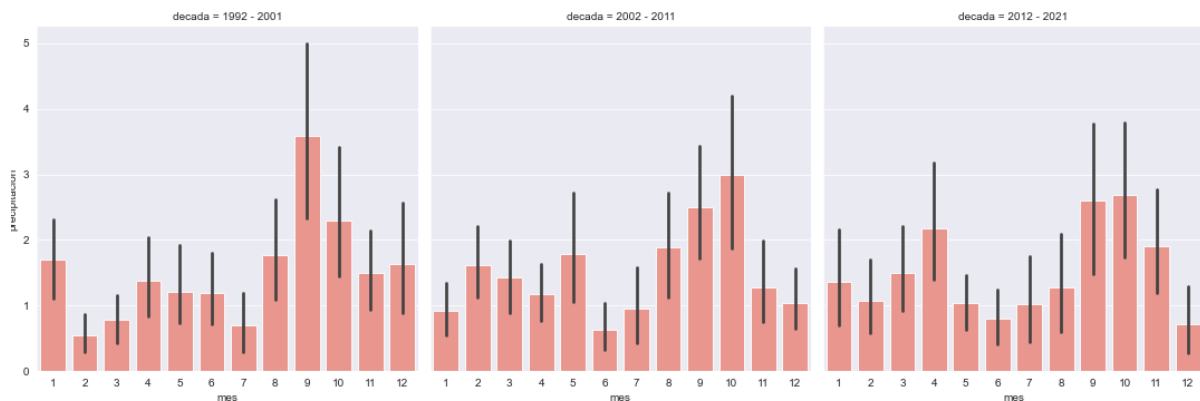


Figura 3

Durante este periodo **septiembre y octubre** han sido los **meses más lluviosos del año**. En la década de los 90 estos meses eran con diferencia los más lluviosos. Sin embargo se puede ver cómo **a partir de los 2000 las lluvias se han ido repartiendo más a lo largo del año** y marzo - abril han ido aumentando su promedio de precipitaciones mientras que septiembre - octubre han ido disminuyendo las precipitaciones.

Como resultado podemos afirmar que **se ha modificado la estacionalidad de las lluvias** durante los últimos treinta años.

Serie temporal

Si tratamos los datos como una serie temporal agregando los datos por meses, podemos analizar sus componentes:

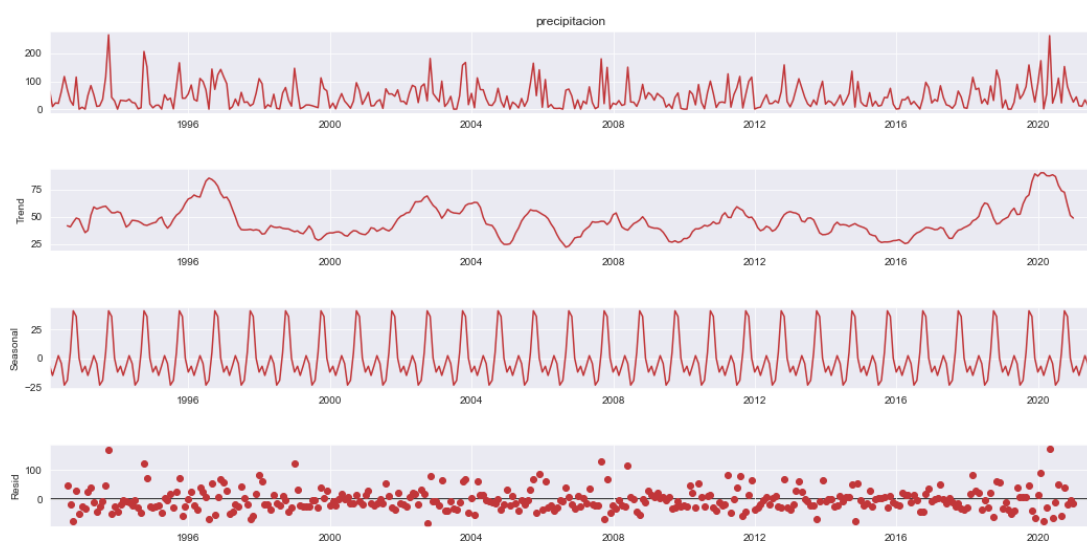


Figura 4

La **tendencia de la serie oscila** a lo largo de los años, observándose periodos donde hay un **aumento** de las precipitaciones acumuladas (de 1996 a 1997, de 2001 a 2003 o de 2019 a 2020) y otros donde la tendencia es **descendente** (de 1997 a 1998 o el actual periodo de 2020 a 2021).

El **componente estacional es muy constante**, observándose un máximo de lluvias en el último trimestre del año y un mínimo sobre la mitad del año.

La componente residual presenta mucha homogeneidad, no se observan desviaciones o tendencias dentro de la serie.

Del análisis de la tendencia se observa que **entre 2017 y 2020** hubo un periodo de tendencia **ascendente** y a partir de **mediados del año 2020** estamos iniciando un **periodo** de tendencia de **menos lluvias**.