

# Idea

Partía de la idea de hacer un análisis de los movimientos terrestres en la Comunidad Autónoma de Andalucía. Analizar los trayectos en coche, moto, autobuses, metro, etc... con el propósito de investigar si hay un alto uso de los transportes públicos frente al uso de los automóviles. No conseguí encontrar la información que requería y ligado a este punto de conciencia climatológica decidí buscar datos sobre las estaciones de calidad del aire.

Esta segunda idea si que la avancé bastante, llegando a recopilar la información a través de selenium. El impedimento que me llevó a rechazar esta segunda propuesta de EDA fue que los datos obtenidos no tenían ninguna relación medioambiental. Simplemente contenían datos de los cuales era difícil plantear un análisis coherente más allá de mirar números. Por eso pensé ¿Podría cruzar datos de AEMET con datos de estaciones medioambientales de Andalucía? Lamentablemente no he sido capaz. No sé si por la idea de abordarlo o por si realmente no he sabido utilizar las herramientas para conseguirlo.

Con los datos de AEMET en la mano, si que se me ocurrieron ideas de cómo trabajarlos y de hipótesis que contrastar. Conservando la idea de investigar sobre datos en Andalucía, he trabajado con datos meteorológicos del año 2022.

## Hipótesis

Hay dos hipótesis que he creado porque conozco por tradición oral y quiero comprobar si son reales o sólo dichos populares.

Las capitales de interior son más cálidas que las costeras tanto en verano como en invierno. Y que las capitales costeras tienen menos amplitud térmica, por aquello de que el mar suaviza las temperaturas.

La otra que planteo es: Las capitales con mayor altitud tienen temperaturas más extremas.

Con los datos de la API de AEMET puedo ver la altitud de las estaciones meteorológicas que analizaré, los registros diarios de temperaturas medias, máximas y mínimas. Este es el punto de partida para el EDA que quiero presentar para el bootcamp de Data Science de The Bridge.

# Metodología

Lo primero, son las herramientas usadas.

Para programar en Python he usado VisualStudio Code con las siguientes librerías pandas, numpy, requests, json, seaborn, plotly, plotly.graph\_objs, WordCloud, matplotlib.pyplot, sklearn.datasets.

La data la he obtenido de la API de AEMET. Primero tuve que registrarme para obtener el token de acceso y después buscar en la documentación para ver cómo obtener el volcado de datos. El token se introducía directamente en la web para realizar las llamadas. Es decir, el código nunca lo introducía en VisualStudio Code, sino que directamente desde la web realizaba la búsqueda y me devolvía un código que usaría en VS Code. Para mi sorpresa los datos no se volcaban directamente, sino al hacer la llamada obtiene un código json con una url a la que había que llamar para extraer ya sí los datos.

Una llamada tipo sería:

```
diciembre = "https://opendata.aemet.es/opendata/sh/4daaa7c9"
response_diciembre = requests.get(diciembre)
dic = response_diciembre.json()
dic
```

Recopile todos los meses y los concatené en un único documento de 87.375 filas. Desde ahí saqué los datos sólo correspondientes a las provincias andaluzas, quedando ya sólo 14.254 filas y 22 columnas. Eliminé columnas que no entrarían a formar parte del análisis. Y finalmente filtré los datos por las estaciones meteorológicas de las capitales de provincia.

El trabajo se realizó con 2.876 filas.

La limpieza de los datos que trabajé, fueron los siguientes:

- Pasar a formato datetime la columna fecha.
- Crear la columna mes con formato datetime.
- Re-ordenar las columnas en datos str que se convertirán a float.
- Sustituir las comas por puntos y eliminar datos lp. (*sigo sin comprender o en entender el dato lp que me ofrecía AEMET*)
- Convertir a float los datos de temperaturas.
- Eliminar los NaN.

La presentación de los resultados está hecha en google.sites

<https://sites.google.com/view/eda-aemet-andalucia-jfa/inicio?authuser=0>

# Resultados y conclusiones

El análisis y conclusiones está expuesto en la presentación que se puede consultar en el siguiente enlace: <https://sites.google.com/view/eda-aemet-andalucia-jfa/conclusiones>

En resumen para la memoria expongo:

**Comprobamos que las hipótesis eran correctas.**

- Las capitales de interior son más cálidas que las exteriores tanto en verano como en invierno.
- Las capitales costeras tienen menor amplitud térmica.

**Sin embargo, en esta hubo alguna discrepancia y no puedo certificarla.**

- Las capitales con mayor altitud tienen temperaturas más extremas.

## Bibliografía

Uso de API AEMET -> <https://opendata.aemet.es/centrodedescargas/inicio>

Uso de regex -> <https://www.geeksforgeeks.org/python-pandas-dataframe-replace/>

Notebooks The Bridge -> clase

Imágenes -> Google libre de derechos