

Autor: Lander Combarro Exposito
Fecha: 31 de julio de 2024

EDA: Climatología del País Vasco en las últimas décadas

1 Introducción

El presente documento detalla el análisis exploratorio de datos (EDA) que se ha realizado sobre un conjunto de datos meteorológicos del País Vasco.

1.1 Objetivo

El objetivo de este proyecto es investigar las tendencias climáticas y proporcionar una comprensión preliminar de los patrones meteorológicos en la región del País Vasco a lo largo de varias décadas.

1.2 Hipótesis

- Las regiones que experimentan mayores aumentos en la temperatura también muestran mayores niveles de sequía/disminución de lluvias.
- El calentamiento global provoca un aumento de la temperatura media del territorio pero también de los fenómenos adversos como frío extremo o precipitaciones.

2 Fuente de datos

AEMET [1] gestiona el Banco Nacional de Datos Climatológicos que atesora la memoria del clima de nuestro país. En él hay datos históricos de casi 12.000 observatorios, de los que están en funcionamiento aproximadamente 3.300. Los datos más antiguos se remontan a 1805 y corresponden a la estación de San Fernando, en Cádiz, con una serie de datos de precipitación de casi 200 años. Otras 29 estaciones tienen series centenarias que aportan una información de un innegable valor histórico y climatológico.

Los datos utilizados para este análisis provienen de estaciones meteorológicas automáticas (EMA) ubicadas en las diferentes provincias del País Vasco: Araba, Bizkaia y Gipuzkoa. Cada estación meteorológica registra datos diarios y mensuales de diversas variables climáticas, incluyendo temperatura, precipitación y otros eventos meteorológicos.

2.1 Obtención de datos

AEMET OpenData es una API REST desarrollado por AEMET [2], que permite la difusión y la reutilización de la información meteorológica y climatológica de la Agencia, en el sentido indicado en

la Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.

Para poder realizar peticiones, es necesario solicitar en API Key. En la descripción de API se recoge qué tipo de datos se pueden solicitar, con qué frecuencia, y de qué forma. Todos los datos empleados en este EDA han sido extraído desde la API REST de AEMET.

3 Técnicas utilizadas

3.1 Pre-procesamiento de datos

1. Carga y visualización inicial

- Estudio de la sintaxis de acceso en AEMET OpenData.
- Aplicación de los diferentes “*requests*” para la solicitud de datos climatológicos. Librería *requests* de Python.
- Cargar los datos en *DataFrames* de *pandas* para su manipulación.
- Revisar estructura y contenido de los *DataFrames*.

2. Limpieza de datos:

- Limitar los *dataset* a municipios en la comunidad del País Vasco.
- Eliminar filas o columnas con valores críticos para el análisis, así como renombrar algunas etiquetas para estandarizar la nomenclatura con otras bases de datos y con la sintaxis de la API REST.
- Reemplazar valores cero por *numpy.nan* para evitar sesgos en el análisis (*pandas* autogestiona los *numpy.nan* a la hora de calcular estadísticas).

3. Filtrado temporal:

- Estos *dataset* son en realidad datos temporales. Se analiza en cada caso la conveniencia de agrupar datos por mes, año, década o provincia, y obtener así nuevos *sub-datasets*.
- Filtrar los datos para asegurar que cada estación meteorológica tuviera una cantidad de datos suficiente para el análisis temporal. Esto puede significar datos de los últimos 25 o 50 años como mínimo.
- Trabajar únicamente con datos anteriores a enero de 2024 para el análisis.

4. Legibilidad:

- Se estandarizaran el identificador y el nombre de las EMA para mayor legibilidad.
- Se asocia un color a cada provincia: Bizkaia (rojo), Araba (verde) y Gipuzkoa (azul).
- Adecuada elección de los parámetros climáticos que se van a analizar, porque 50 variables son demasiadas.

3.2 Análisis exploratorio

1. Análisis temporal:

- Evaluar qué localidades tienen un mayor número de registros históricos
- Agrupar los datos por año, década o provincia para observar tendencias a largo plazo.
- Calcular la pendiente de las tendencias de los eventos climáticos a lo largo de los años para identificar cambios climáticos significativos. Regresión simple con la librería *numpy*.
- Calcular las medias anuales de magnitudes físicas: temperatura, precipitaciones y las sumas anuales de días con precipitaciones de distinta índole, días con frío o calor extremo.

2. Visualización:

- Analizar los fenómenos meteorológicos de manera independiente.
- Utilizar gráficos de barras y de dispersión para visualizar la relación entre variables categóricas (provincias o municipios) y numéricas.
- Análisis geográfico sobre el mapa del País Vasco para ubicar las estaciones meteorológicas y visualizar los datos en un contexto espacial.
- Análisis de correlación entre las diferentes magnitudes climáticas.

4 Resultados

El análisis exploratorio de datos realizado proporciona una visión inicial valiosa de los patrones climáticos en el País Vasco. Se observaron tendencias significativas en el territorio Vasco, siendo el resultado más claro un aumento significativo de la temperatura media anual y un incremento en los días de calor extremo a lo largo de las décadas.

Los resultados indican un aumento de hasta de 0,036 °C por año en las estaciones meteorológicas analizadas. Este incremento, aunque pueda parecer pequeño anualmente, se traduce en un aumento de aproximadamente 1.8°C en un período de 50 años.

Además, se ha registrado un preocupante incremento en la frecuencia de días con temperaturas extremas (más de 30 °C). Los datos sugieren que en un lapso de 100 años, se han añadido entre 30 y 40 días adicionales de calor extremo por año. Este cambio no solo tiene implicaciones directas en el clima local, sino que también afecta a la biodiversidad, la agricultura, y la calidad de vida de los habitantes de la región.

En resumen, estos hallazgos subrayan la importancia de monitorear y mitigar los efectos del cambio climático. La evidencia del aumento de temperaturas y de los días de calor extremo resalta la urgencia de implementar políticas y medidas que puedan contrarrestar estos efectos negativos y promover un desarrollo sostenible.

La variabilidad en las precipitaciones, como granizo, lluvia, nieve, requiere un análisis más detallado para identificar posibles patrones. A priori no hay tendencias, y los resultados son muy divergentes según el municipio, así que no se puede llegar a una generalización para todo el territorio.

La altitud de los municipios también es un factor importante en el estudio de temperaturas medias o precipitaciones. Y no solo esta, sino que la latitud del municipio, y cuán alejado esté del mar también afecta a un clima más oceánico o no.

Las técnicas utilizadas, incluyendo la limpieza de datos, el análisis temporal y espacial, y las visualizaciones, han permitido una comprensión preliminar de los datos. Sin embargo, se recomienda un análisis más profundo con modelos estadísticos avanzados para confirmar y expandir estos hallazgos iniciales.

5 Futuras líneas de proyecto

En línea con los resultados obtenidos se podrían plantear las siguiente líneas de investigación:

- Las precipitaciones y reservas hídricas del País Vasco no parecen haberse visto especialmente modificadas en las ultimas décadas. Sería interesante ampliar el estudio a otras regiones y contar con *datasets* sobre sequía [3], o sobre el suministro y saneamiento del agua por sectores y regiones [4, 5].
- Analizar el origen de estos cambios climáticos y su relación en tendencias como el aumento del consumo de combustibles fósiles, la industrialización, consumo de alimentos de origen animal o deforestación, entre otros.

6 Dificultades encontradas

- Datos Incompletos. Se encontraron estaciones meteorológicas con registros muy insuficientes, lo que requirió filtrar y limpiar extensivamente los datos, así como optar por limitar el estudio al territorio del País Vasco.
- El *dataset* es algo diferente a lo visto hasta el momento en el *Bootcamp* de *Data Science* por consistir completamente en un conjunto de datos temporales. Ello dificulta la obtención de variables categóricas claras y los análisis de frecuencia o funciones de distribución si no se hacen previamente varios filtrados.
- Aparecieron nuevos desafíos técnicos como implementar solicitudes recurrentes a una API que tiene un número limitado de solicitudes, transformación de coordenadas, o integrar correctamente una imagen en el fondo de una figura.

7 Referencias

- [1] <https://www.aemet.es/en/portada>
- [2] <https://opendata.aemet.es/dist/index.html>
- [3] <https://drought.emergency.copernicus.eu/>
- [4] <https://www.ine.es/>
- [5] <https://www.miteco.gob.es/>