

¿Cómo Predecir el Clima con Precisión?

Explorando los Factores Clave Detrás del Tiempo

Autor: Luis Ignacio Alvarez

Temas

- 01 | Contexto y Audiencia
- 02 | Hipótesis/Preguntas de Interés
- 03 | Metadata
- 04 | Análisis Exploratorio
- 05 | Insights y Recomendaciones

Motivación y audiencia

Motivación:

El clima es un tema que nos afecta a todos, desde la planificación de nuestras actividades diarias hasta decisiones más grandes en sectores como la agricultura, el transporte y la energía. Entender y predecir el clima de manera precisa se ha vuelto cada vez más importante, y los datos meteorológicos disponibles ofrecen una gran oportunidad para ello.

Este proyecto busca explorar cómo diferentes técnicas de Machine Learning pueden ayudarnos a identificar patrones y factores clave que determinan las condiciones climáticas.

Audiencia:

Este análisis está pensado para cualquier persona interesada en entender mejor cómo predecir el clima: desde científicos de datos y meteorólogos, hasta agricultores, planificadores urbanos y estudiantes que buscan aprender sobre técnicas avanzadas de análisis de datos aplicadas a problemas reales.

Limitaciones:

Como trabajamos con un dataset sintético, los resultados podrían no reflejar todas las complejidades del clima en el mundo real. Además, la precisión de los modelos dependerá de la calidad y la cantidad de datos que se utilicen en estudios futuros.

Hipótesis y preguntas guía

- ¿Existe una relación significativa entre la humedad y el tiempo?
- ¿Cómo afecta la estación del año a la clasificación del tiempo?
- ¿En qué medida el tipo de ubicación altera la influencia de las variables meteorológicas en la predicción del tiempo?

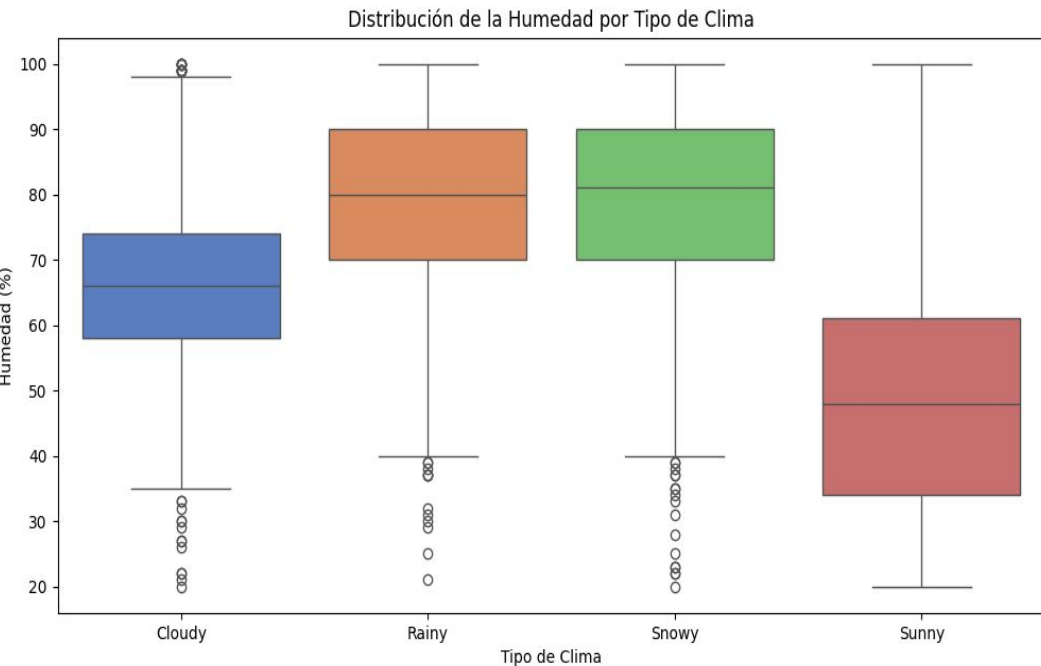


Resumen de Metadata

- **Total de Registros:** 13,201 filas
Representa observaciones individuales de condiciones climáticas registradas en distintas ubicaciones y momentos.
- **Total de Variables:** 10 columnas
Incluye tanto variables numéricas como categóricas que capturan diferentes aspectos del clima.
- **Variables Numéricas (7 columnas):**
 - **Temperature:** Valores en grados Celsius.
 - **Humidity:** Porcentajes de humedad (0 a 100%).
 - **Wind Speed:** Velocidad del viento en km/h.
 - **Precipitation (%):** Porcentaje de precipitación.
 - **Atmospheric Pressure:** Medidas en hPa.
 - **UV Index:** Valores del índice de radiación UV (generalmente de 0 a 11+).
 - **Visibility (km):** Distancia de visibilidad en kilómetros.
- **Variables Categóricas (3 columnas):**
 - **Cloud Cover:** Categorización del tipo de cobertura nubosa (e.g., "Clear", "Partly Cloudy", "Overcast").
 - **Season:** Estación del año (e.g., "Spring", "Summer", "Autumn", "Winter").
 - **Location:** Tipo de ubicación donde se realizaron las mediciones (e.g., "Urban", "Rural", "Coastal").
- **Variable Objetivo:**
 - **Weather Type:** Clasificación del clima (valores posibles: "Rainy", "Sunny", "Cloudy", "Snowy").

ANÁLISIS EXPLORATORIO

¿Existe una relación significativa entre la humedad y el tipo de clima?

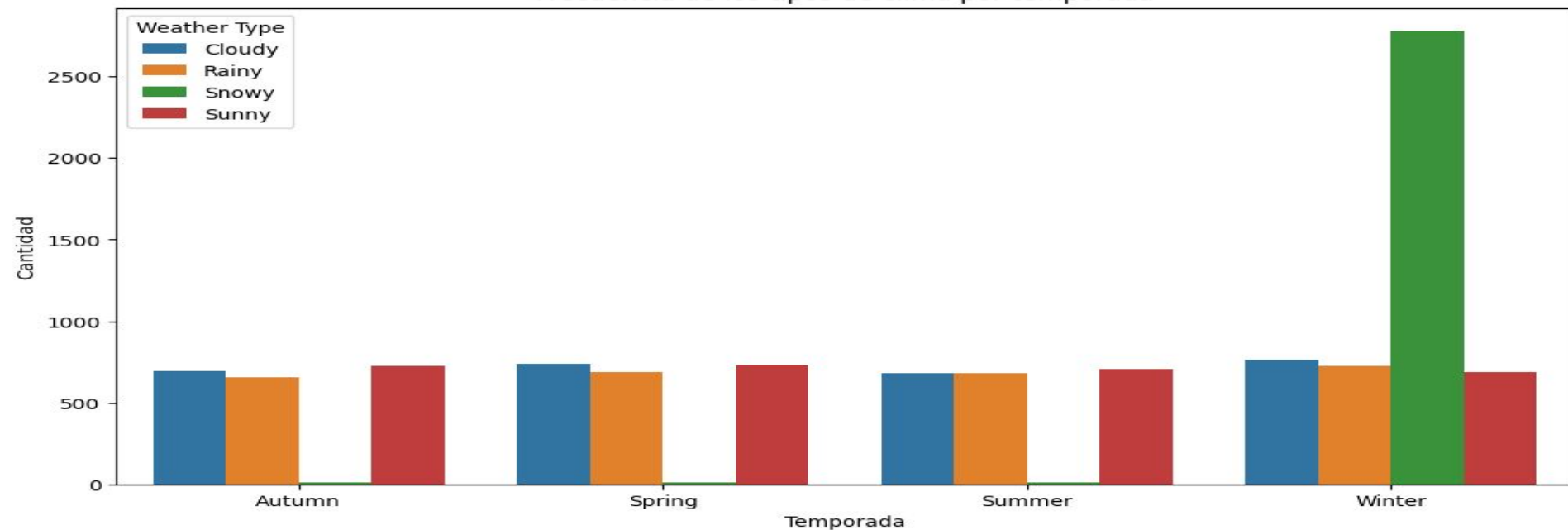


Como intuitivamente uno podría pensar, en este gráfico se puede visualizar que la humedad desempeña un papel clave en la distribución de las temperaturas. Aquellos que más humedad presentan tienden a ser lluviosos o nevados (naranja y verde), seguido por aquellos nublados (azul) y finalmente soleados (rojo).

¿Cómo afecta la estación del año a la clasificación del tipo de clima?

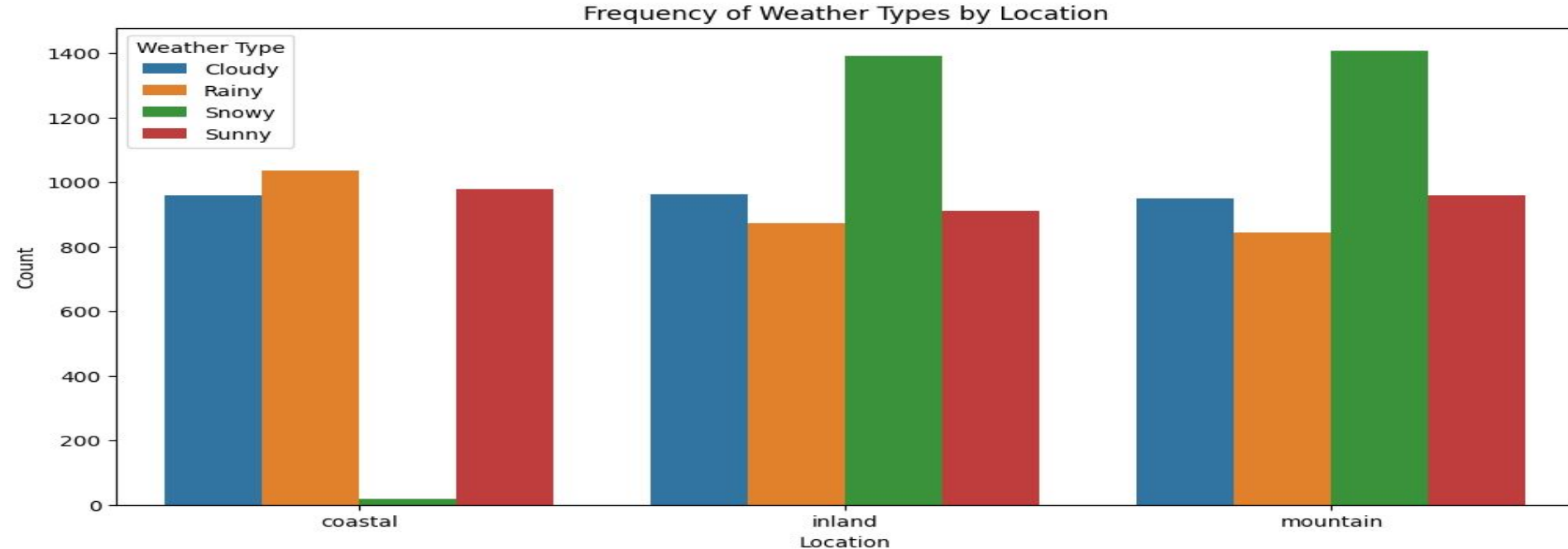
Como se puede observar en el gráfico, el tipo de estación no presenta cambios significativos en cuanto al tiempo soleado, nublado o lluvioso. Pero resulta determinante en cuanto los nevados, ya que es poco propenso a producirse a menos que sea en invierno.

Frecuencia de los tipos de clima por temporada



¿En qué medida el tipo de ubicación altera la influencia de las variables meteorológicas en la predicción del tiempo?

De forma similar al gráfico anterior, no representa ser una variable muy significativa a la hora de determinar entre el tiempo soleado, nublado o lluvioso, pero sí es determinante en cuanto al nevado. La probabilidad de nieve en áreas costeras es casi nula en comparación al resto.



INSIGHTS y Conclusiones

Insights del Proyecto Climático

1. **Temperatura y Presión Atmosférica:**
 - Son las variables más relevantes para clasificar el clima, según el modelo de Random Forest.
2. **Índice UV:**
 - Altos índices UV están correlacionados con días "Sunny" (soleados), mientras que índices bajos indican "Cloudy" (nublados) o "Snowy" (nevados).
3. **Precipitación y Visibilidad:**
 - Alta precipitación reduce la visibilidad. Días con más del 70% de precipitación suelen tener visibilidad menor a 5 km.
4. **Estacionalidad:**
 - Las estaciones afectan el clima: "Winter" (invierno) es más propenso a "Snowy" (nevado), y "Summer" (verano) a "Sunny" (soleado).
5. **Cobertura Nubosa:**
 - Días con alta cobertura nubosa se clasifican como "Cloudy" o "Rainy", mientras que días despejados son principalmente "Sunny".
6. **Ubicación Geográfica:**
 - Las áreas costeras tienen más días "Rainy", mientras que las zonas rurales experimentan más "Cloudy" y "Sunny".