

UNSE

Universidad Nacional
de Santiago del Estero

51
AÑOS

Universidad Nacional de Santiago del Estero
FACULTAD DE CIENCIAS EXACTAS Y TECNOLÓGICAS

ANÁLISIS Y PREDICCIÓN DEL CLIMA DE SANTIAGO DEL ESTERO

Taller de Probabilidad y Estadística

Autor: Pablo Rojas

Legajo: 129/18

Carrera: Licenciatura en sistemas de información

Profesores:

Pedro González

Walter Torres

Índice general

1. Introducción	2
2. Análisis de las variables climáticas	4
Perspectiva general anual de promedios	4
Relaciones entre variables	10
Temperatura y precipitaciones	10
Nubosidad y precipitaciones	11
Presión atmosférica y precipitaciones	12
Presión atmosférica y temperatura	13
Presión atmosférica y velocidad del viento	14
Precipitación y humedad	15
Temperatura y humedad	16
Valores atípicos	17
3. Machine Learning aplicado al clima y análisis de series temporales	18
Predicciones: elección de variables objetivo	18
Precisión de las predicciones de precipitación	19
Precisión de las predicciones de temperatura	20
Precisión de las predicciones de humedad	21
Precisión de las predicciones de viento	22
Predicciones y análisis de tendencias con series temporales realizadas para el resto del 2024	23
Predicción y serie temporal de la precipitación	24
Predicción y serie temporal de la temperatura	25
Predicción y serie temporal de la humedad	26
Predicción y serie temporal de la velocidad del viento	27
Anomalías en las predicciones	28
Anomalías en las precipitaciones	28
Anomalías en la temperatura	29
Anomalías en la humedad	29
Anomalías en la velocidad del viento	30
4. Conclusión	31

1. Introducción

El objetivo de este proyecto es analizar por medio de varios tipos de gráficas, las distintas variables referidas al clima de Santiago del Estero, cuyos datos se pudieron recolectar en el último año. Los datos son recolectados por una empresa del clima llamada OpenWeather que proporciona una API que proporciona los datos recolectados dada una ubicación geográfica expresada en latitud y longitud. Particularmente la latitud y longitud de Santiago del Estero son Latitud:-27.784444, Longitud:-64.266944.

Una vez analizados los datos el siguiente objetivo es realizar una predicción de las variables climáticas de lo que queda del año.

Herramientas usadas:

- El corazón del proyecto es el language de programación **Python**, en él se basó todo el proyecto desde la creación de las gráficas hasta finalmente las predicciones realizadas.
- Para recopilar los datos usé la API de datos estadísticos de **OpenWeather**, la cual ofrece un servicio de datos históricos del clima.
- Para la representación de las gráficas usé las librerías **Matplotlib** y **Seaborn**.
- Para los modelos de machine learning para las predicciones usé la librería **scikit-learn**.
- Y finalmente para manejar los datos usé **Pandas**, librería dedicada al análisis de datos

Las **variables** que se manejan en este caso son las siguientes:

- La **temperatura** es qué tan caliente o fría está la atmósfera en un lugar dado, cuántos grados está por encima o bajo el punto de congelación, o por encima o por debajo de los valores promedio de dicha temperatura. La temperatura es un factor muy importante en determinar el tiempo porque influye o controla las características de otros elementos del tiempo, como precipitación, humedad, nubes y presión atmosférica.
- La **humedad** es la cantidad de vapor de agua en la atmósfera. Se indica en valores absolutos (gramos por m³, por ejemplo) o en valores relativos o porcentuales, siendo un valor del 100 %, evidentemente, el valor máximo que puede tener una masa de aire que se denomina el punto de rocío, cuando la condensación del vapor de agua se acelera y se producen nubes y, eventualmente, lluvias.
- La **precipitación** es el producto de un proceso de condensación rápido (si este proceso es lento, solo produce cielos nubosos). Puede incluir nieve, aguanieve, granizo, llovizna y lluvia.

- La **presión atmosférica** (o presión del aire), es el peso del aire que descansa sobre la superficie de la Tierra en un lugar dado. Se mide generalmente con un barómetro aneroide, que nos indica las variaciones de presión traducidas en la estabilidad de la atmósfera (alta presión o anticiclón, con cielos despejados) o la inestabilidad de la misma con vientos y nubes cargadas que pueden originar lluvias o tempestades, de carácter más o menos violento, según sea la intensidad en el valor de descenso o disminución de la presión atmosférica en un lugar dado.
- El **viento** es el movimiento de las masas de aire, especialmente en la superficie de la Tierra.
- La **nubosidad** es la fracción de cielo cubierto con nubes, en un lugar en particular. Según las normas meteorológicas actuales la nubosidad se expresa en octas u octavos de la bóveda celeste

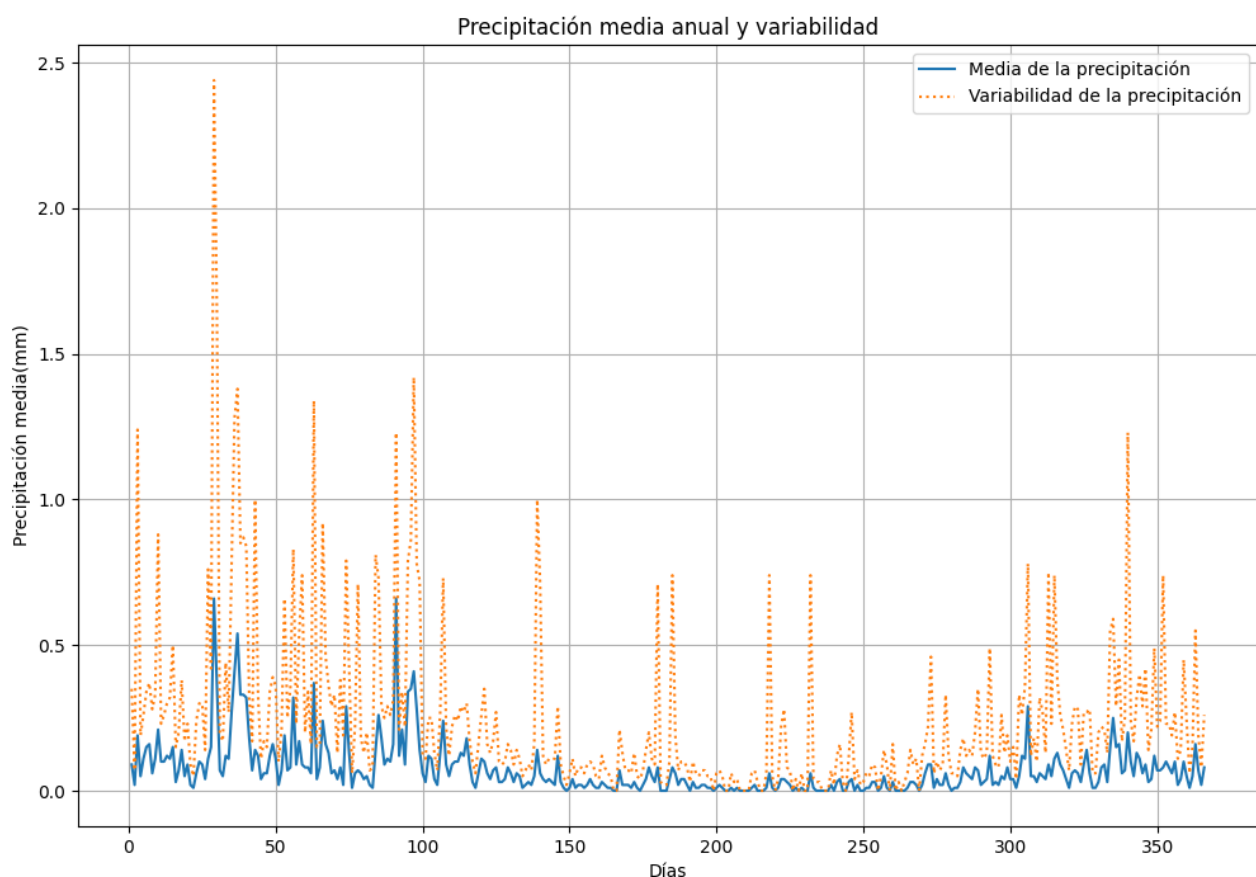
En los siguientes capítulos se explorará los temas centrales, gráficas y conclusiones de las mismas comenzando por el análisis del clima del año pasado hasta llegar al tema más importante que es la predicción de las variables del resto del año por medio de regresión lineal y Poisson.

Debido a la falta de datos históricos mirando muchos más años atrás la precisión de las predicciones podría verse afectada, sin embargo creo que es el acercamiento más adecuado el tratar con datos actuales y del año pasado para demostrar como funciona la aplicación de técnicas de Machine Learning en estadística y el clima.

2. Análisis de las variables climáticas

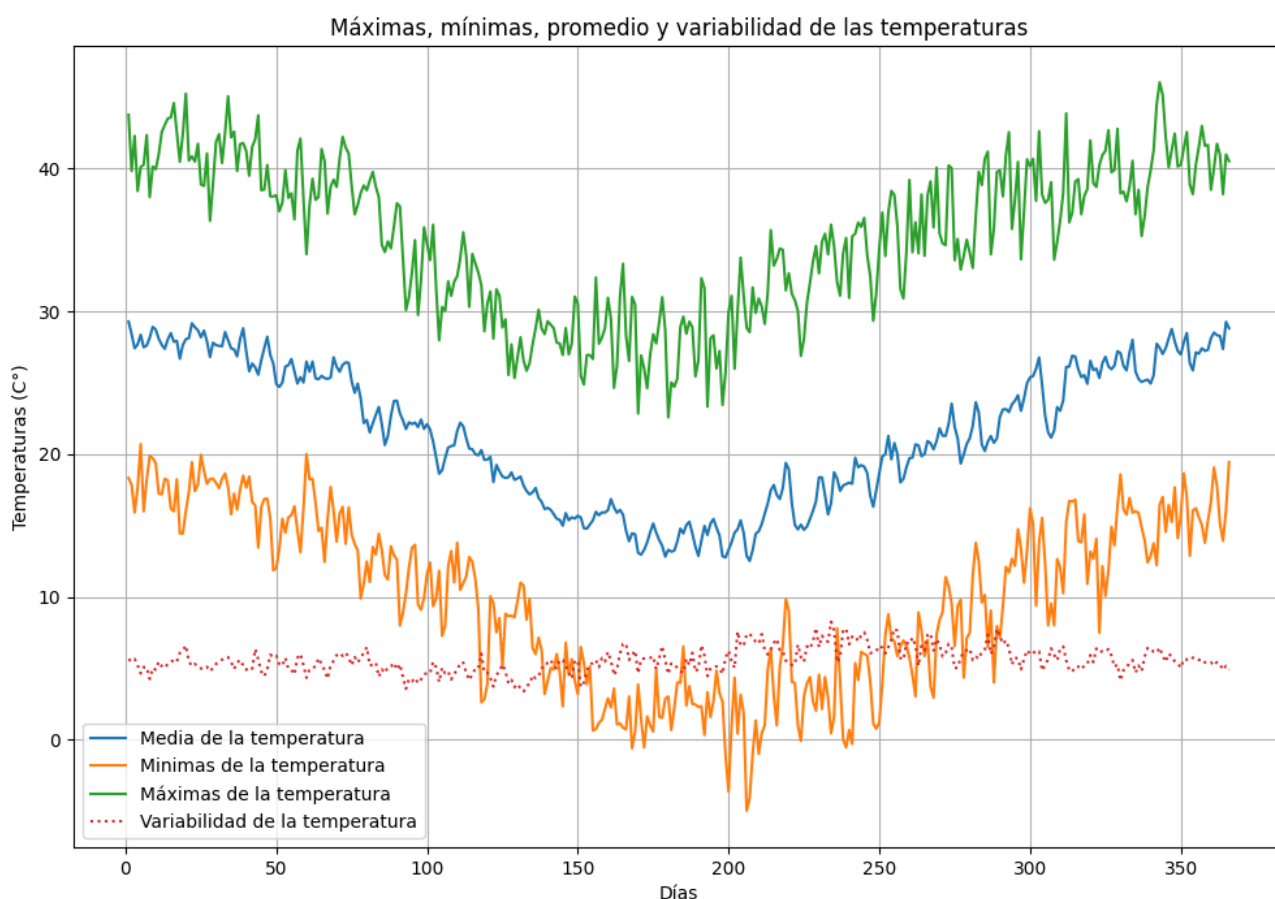
Perspectiva general anual de promedios

Una vez definidas conceptualmente las variables y sus distintas medidas podemos proceder a ver las gráficas donde se muestra el como evolucionaron las variables el año pasado.



Los meses en los que la línea azul está mas alta son los días donde hay mayor precipitación media que los demás, indicando una temporada de lluvias mientras que cuando la línea está baja es una temporada de sequías o de menor precipitación. Para complementar esta línea tenemos la punteada naranja que representa la variabilidad, (desviación estándar) la cual muestra cuánto varía la precipitación cada día en comparación con su media. Si la desviación estándar es alta en un día específico, significa que la precipitación es muy variable en ese día. Si es baja, la precipitación es más constante y predecible.

En este caso podemos ver como en febrero y finales de marzo las precipitaciones medias fueron las mayores de todo el año, sin embargo la variabilidad fue la mayor también es decir, así como hubieron días que llovió gran cantidad hubieron días donde no llovió o llovió muy poco. Sin embargo podemos observar como de octubre a diciembre las precipitaciones eran moderadas, sin embargo la variabilidad es alta.



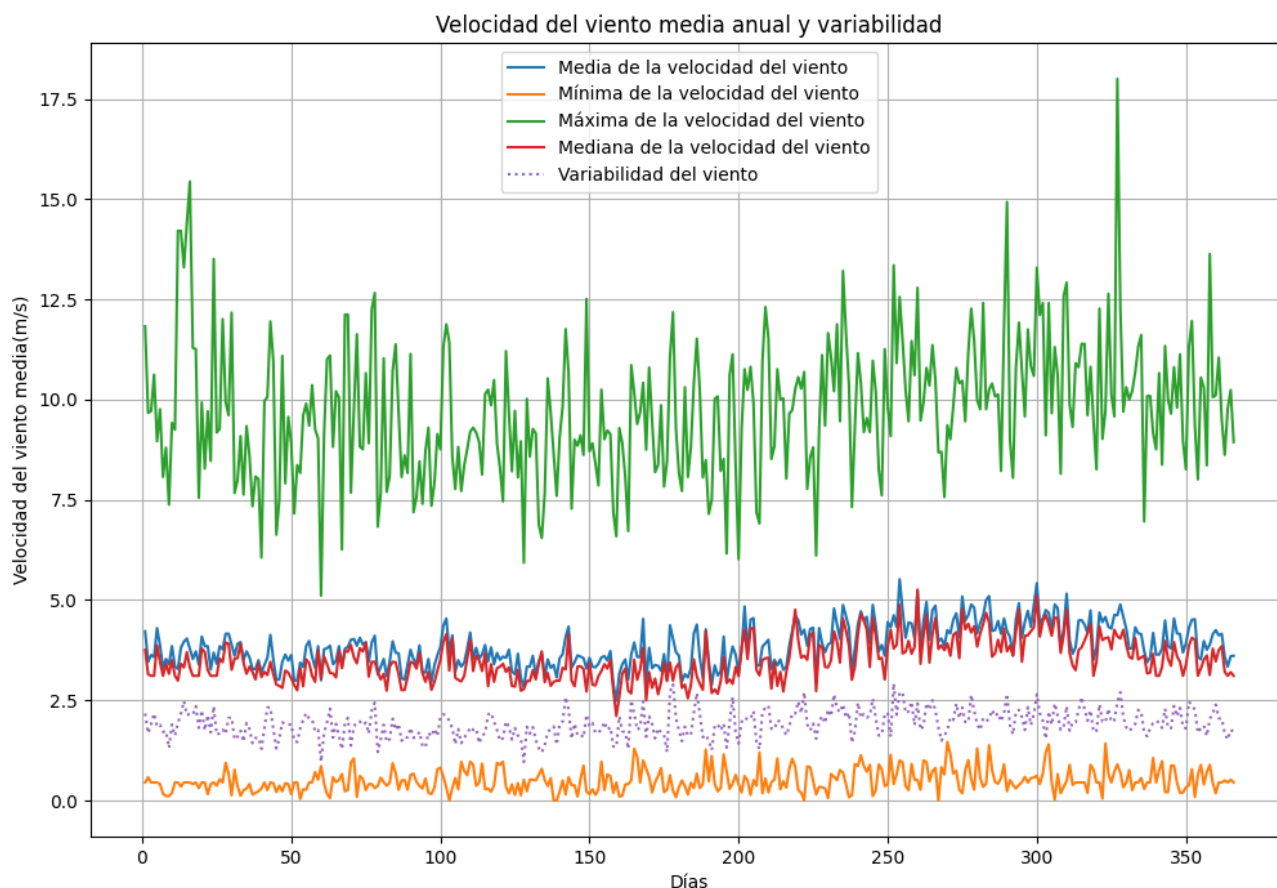
La línea naranja y verde indican la temperatura mínima y máxima respectivamente de ese día, mientras que la azul es la media entre los dos, y finalmente la línea punteada roja indica la variabilidad.

En general se puede observar como los meses de mayor calor en Santiago del Estero son en verano, es decir enero y diciembre, y luego comienza a bajar en los meses de mayo, junio y julio, para luego comenzar a subir nuevamente formando una estructura de V o valle.

Mientras tanto la variabilidad se mantiene en un rango de aproximadamente 4 a 8 grados centígrados en todos los meses sin que cambie demasiado, y los meses en donde más variable es en julio y agosto que es donde más baja es la temperatura.

Algo más que destacar es que las temperaturas máximas al inicio y final del año son superiores a 40 grados, mientras que en el medio del año bajan por debajo de 30, esto se debe a un patrón estacional claro.

Finalmente resulta interesante mencionar que en los meses de más frío que son los del medio del año, es decir junio y julio, la temperatura mínima llega incluso a tocar los bajo cero, siendo la segunda semana de junio la más fría, sin embargo manteniendo una variabilidad relativamente alta tocando aproximadamente los 8 grados.



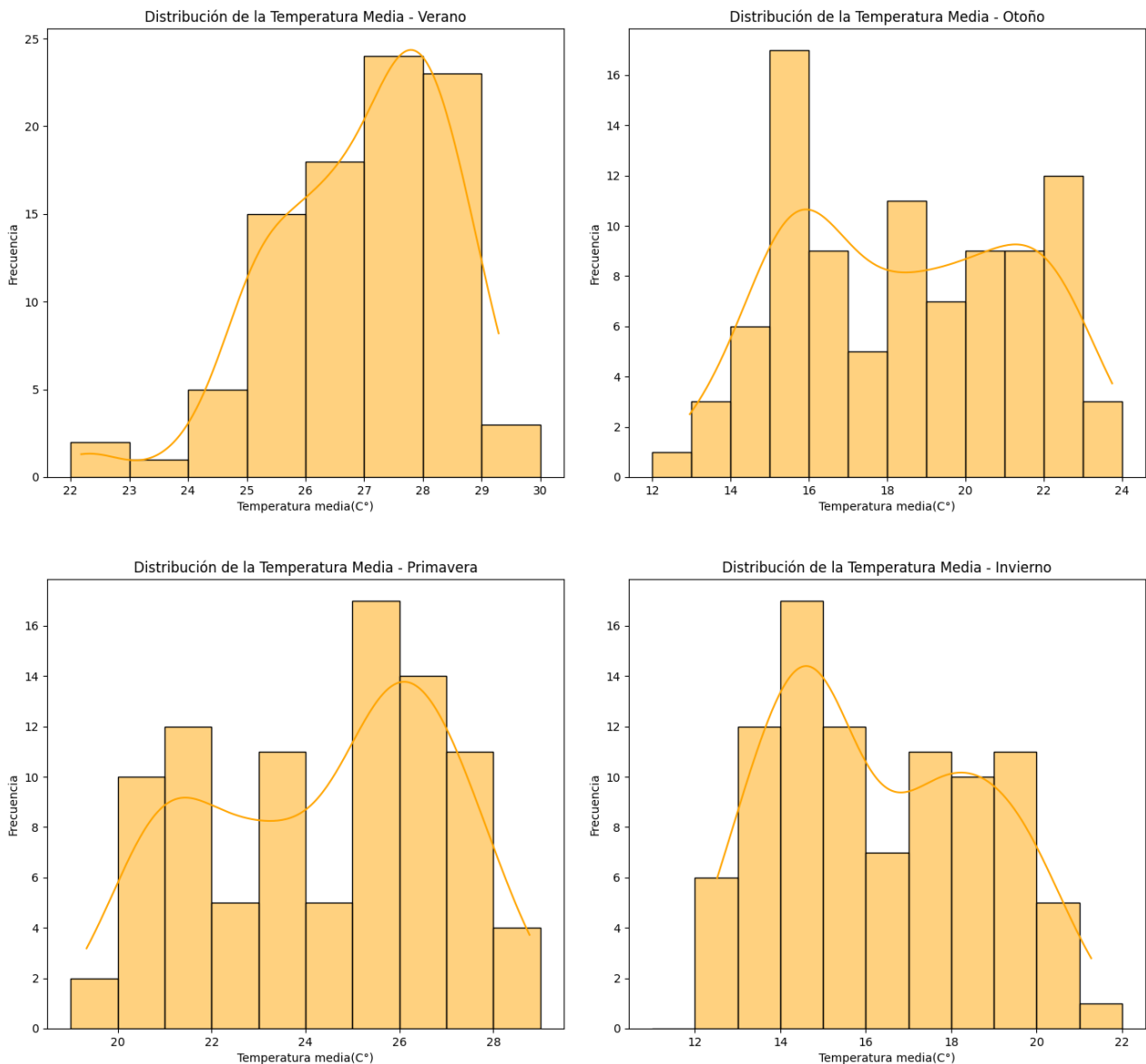
Se observa que la máxima (línea verde) de la velocidad del viento muestra picos significativos que varían considerablemente a lo largo del año, indicando días con velocidades de viento muy altas.

La mínima de la velocidad del viento (línea naranja) está muy cerca del eje x, indicando que hay muchos días en el año donde la velocidad del viento es muy baja o cercana a cero. Esto sugiere periodos de calma frecuentes.

La media y la mediana de la velocidad del viento (líneas azul y roja, respectivamente) están mucho más cercanas entre sí y son relativamente constantes en comparación con la línea verde. Esto sugiere que, en promedio, la velocidad del viento es bastante constante a lo largo del año. Es decir que a pesar de los picos ocasionales causados por la máxima la mayoría de los días tienen vientos moderados.

La variabilidad del viento (línea punteada púrpura) muestra cómo varía la velocidad del viento en diferentes días. La variabilidad parece ser más o menos constante a lo largo del año. Sin embargo presenta una gran variabilidad a lo largo del año, con algunos días experimentando velocidad de viento muy altas.

A partir de organizar las distribuciones en base a la estacionalidad saqué las siguientes gráficas con respecto a la temperatura media.



Las gráficas muestran la distribución de la temperatura media para cada una de las cuatro estaciones del año: **verano**, **otoño**, **primavera** e **invierno**. Cada gráfico es un histograma que muestra la frecuencia de ocurrencia de diferentes rangos de temperatura media, acompañado de una línea suave que representa la densidad de probabilidad de la distribución.

En el **verano** el rango de temperatura es aproximadamente de 22°C a 30°C, mientras que la mayoría de los días tienen una temperatura media entre 26°C y 28°C, siendo este el rango con mayor frecuencia. Finalmente la forma de la distribución es asimétrica a la izquierda (sesgada a la izquierda), con una cola más larga hacia las temperaturas más bajas.

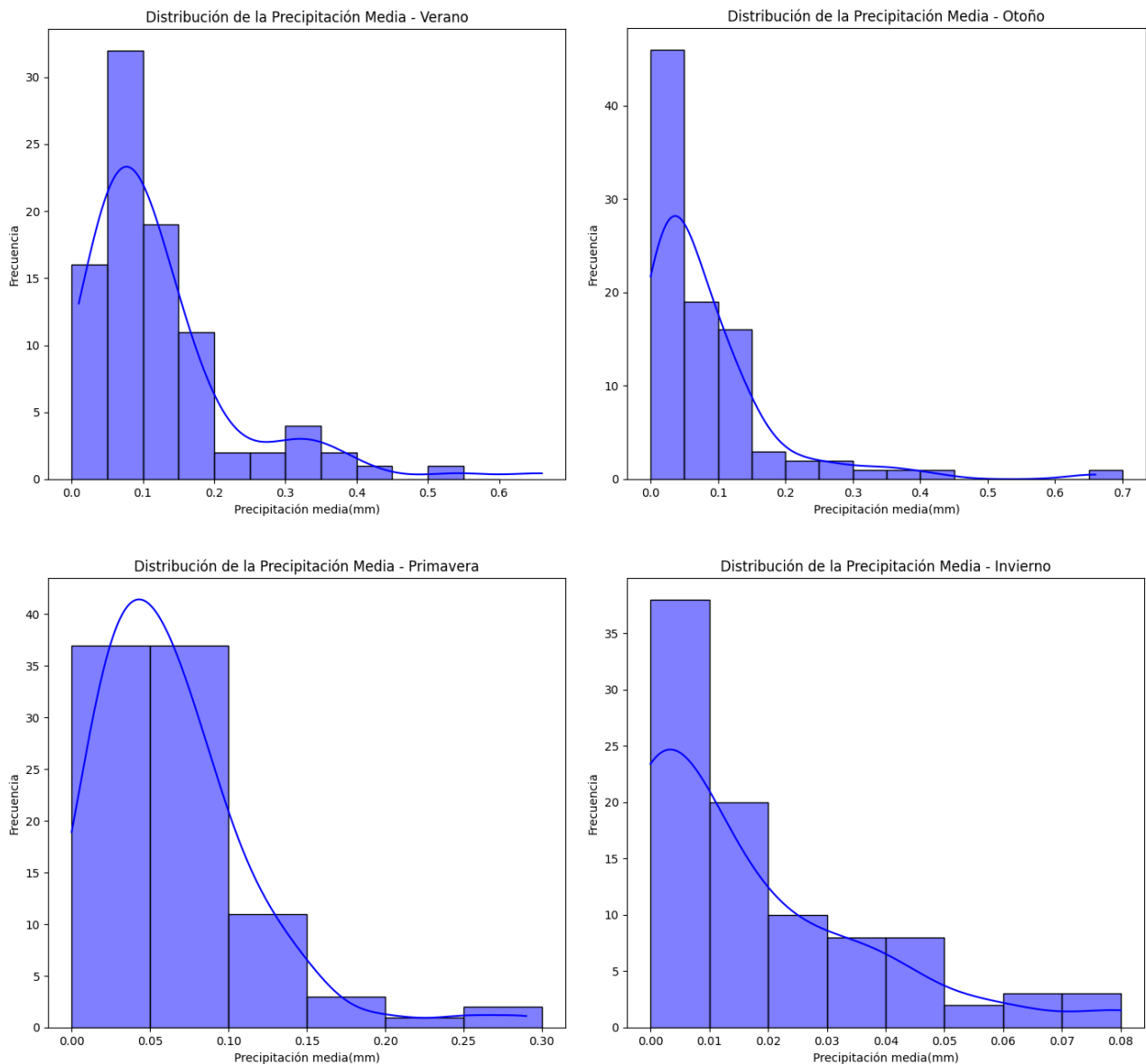
En **otoño** el rango de temperatura es aproximadamente de 12°C a 24°C, mientras que la mayor frecuencia de días se encuentra en el rango de 16°C a 18°C. Esta distribución es multimodal, con varios picos de frecuencia en diferentes rangos de temperatura (alrededor de 14°C, 18°C y 22°C).

En la **primavera** el rango de temperatura es aproximadamente de 20°C a 28°C, mientras que la mayoría de los días tienen una temperatura media entre 24°C y 26°C. La distribución tiene dos picos principales (bimodal), uno alrededor de 22°C y otro alrededor de 26°C.

Finalmente en el **invierno** el rango de temperatura es aproximadamente de 12°C a 22°C. La mayor frecuencia de días se encuentra en el rango de 14°C a 16°C y la distribución es sesgada a la derecha, con una cola más larga hacia las temperaturas más altas.

A manera de sacar conclusiones generales, las gráficas ayudan a ver que verano y primavera tienen distribuciones de temperatura media más altas en comparación con otoño e invierno, como es de esperar. Otoño y primavera muestran distribuciones multimodales, indicando una mayor variabilidad en la temperatura media diaria. Invierno tiene una distribución más concentrada en temperaturas más bajas, mientras que verano tiene temperaturas más altas, con una menor variabilidad comparativa. La forma y la dispersión de las distribuciones proporcionan información sobre la variabilidad y consistencia de las temperaturas medias en cada estación cuando se las observa junto con la gráfica de más arriba sobre la distribución en todo el año.

De manera similar a las gráficas anteriores con respecto a la temperatura media, saqué gráficas con respecto a las precipitaciones medias.



En verano la distribución de la precipitación media es sesgada hacia la derecha, con la mayoría de los valores de precipitación media por debajo de 0.2 mm.

En otoño la distribución de la precipitación media es también sesgada hacia la derecha, con la mayoría de los valores de precipitación media por debajo de 0.15 mm.

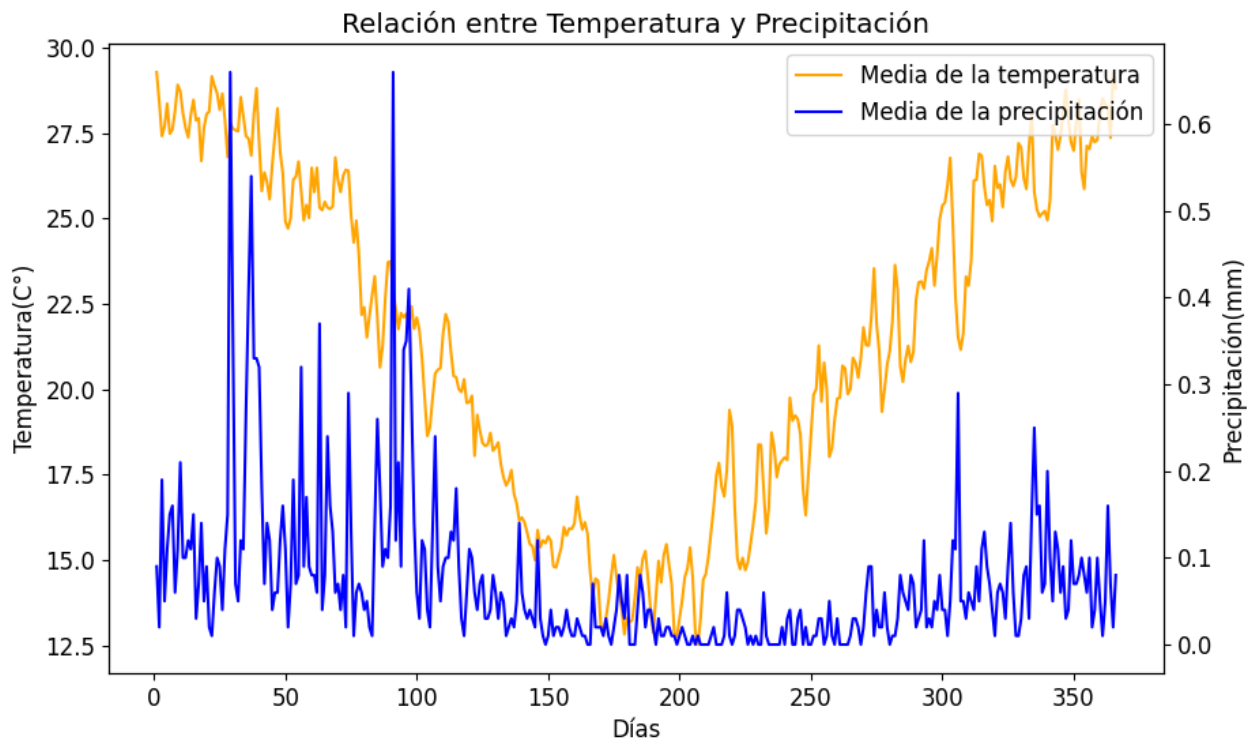
En la primavera la distribución de la precipitación media es más simétrica que en verano y otoño, con un pico alrededor de 0.05 mm.

Finalmente en invierno la distribución de la precipitación media es sesgada hacia la izquierda, con la mayoría de los valores de precipitación media por debajo de 0.02 mm.

En general, la precipitación media es mayor en invierno y menor en verano, con valores intermedios en primavera y otoño. Esto es consistente con los patrones climáticos mostrados en la gráfica anual.

Relaciones entre variables

Temperatura y precipitaciones



La gráfica muestra la relación entre la temperatura y la precipitación durante el período de 365 días del año pasado. La temperatura se representa en el eje izquierdo y la precipitación en el eje derecho. La línea superior naranja representa la temperatura media diaria, mientras que la línea azul representa la precipitación media diaria.

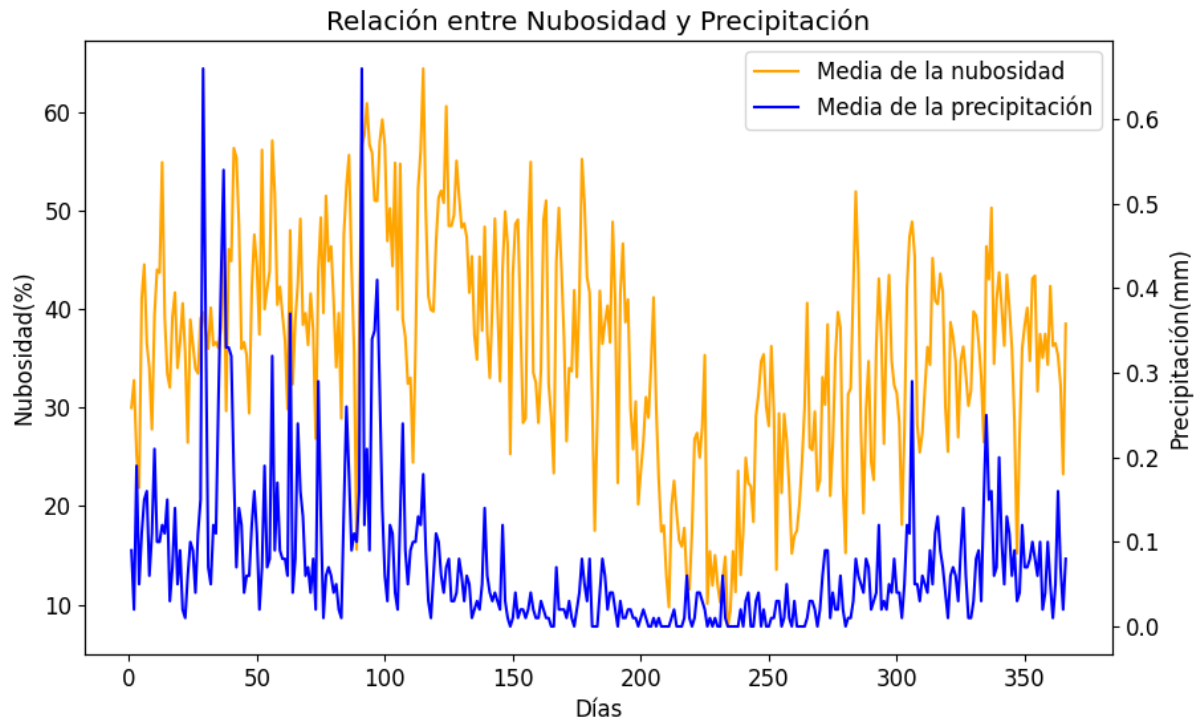
Se puede observar que la temperatura y la precipitación muestran una relación estrecha:

La temperatura aumenta a medida que aumenta la precipitación. Aunque no es una relación directa, la tendencia es clara. Se observa que los periodos de mayor precipitación coinciden con los de mayor temperatura.

La temperatura y la precipitación muestran un ciclo anual. Se observa un patrón estacional, con la temperatura y la precipitación más altas en verano y más bajas en invierno.

En general, la gráfica sugiere que existe una relación positiva entre la temperatura y la precipitación, lo que significa que cuando la temperatura es alta, es probable que haya más precipitación. Sin embargo, es importante destacar que la relación no es perfecta y que hay otras variables que pueden influir en la temperatura y la precipitación.

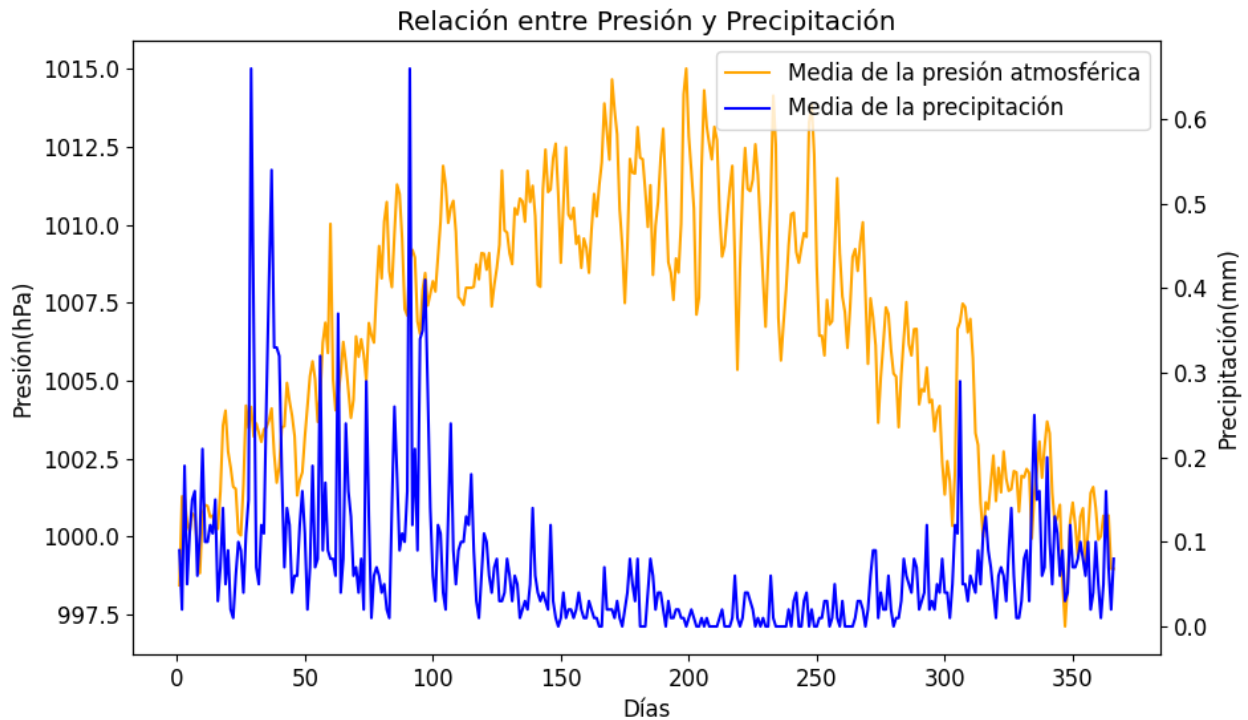
Nubosidad y precipitaciones



La gráfica muestra la relación entre la nubosidad y la precipitación. La línea superior es la nubosidad y la inferior la precipitación. Podemos observar que hay una correlación positiva entre ambas variables, es decir, que a mayor nubosidad, mayor posibilidades de precipitación.

En los días en que la nubosidad es alta, es más probable que haya precipitaciones. Sin embargo, la relación no es perfecta, ya que hay días con mucha nubosidad pero sin precipitaciones y viceversa. Esto indica que hay otros factores que influyen en la precipitación, como la humedad, la temperatura y la presión atmosférica. Sin embargo podemos destacar que a menor porcentaje de nubosidad las precipitaciones son exageradamente mínimas.

Presión atmosférica y precipitaciones



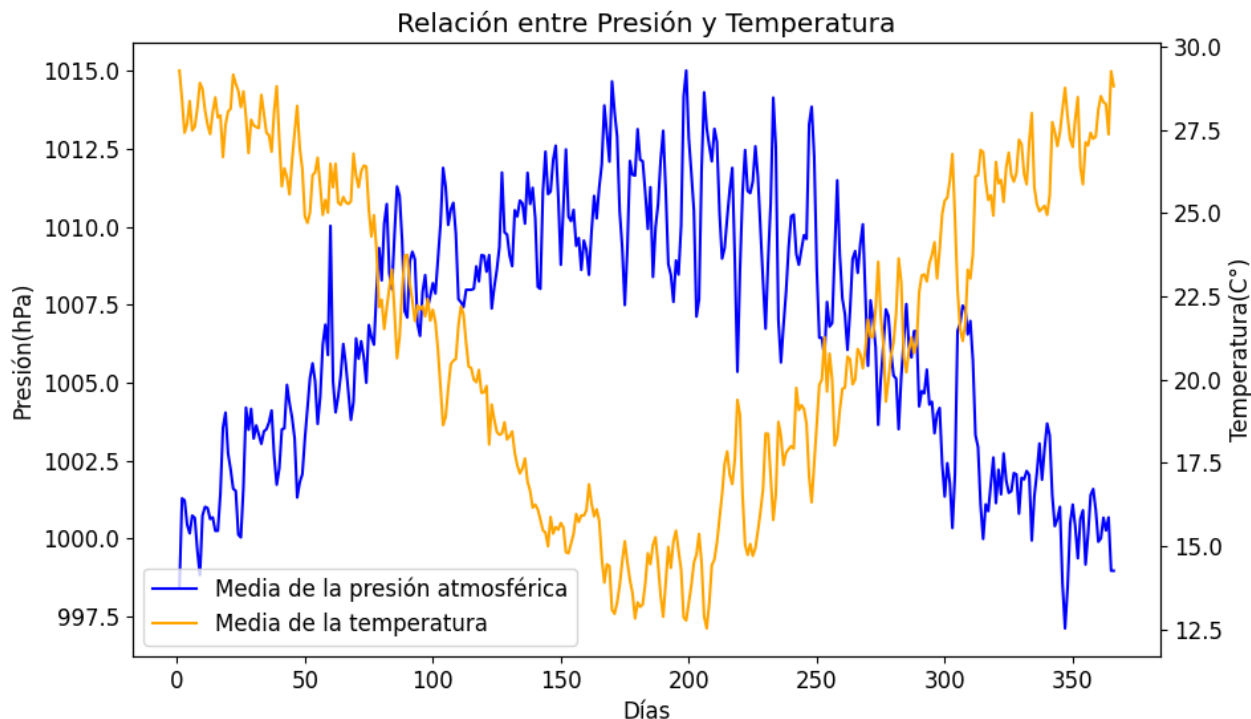
La línea superior es la presión atmosférica, y la línea inferior son las precipitaciones. La gráfica muestra que la presión atmosférica fluctúa en un rango de aproximadamente 1000 hPa a 1015 hPa. Se puede observar una tendencia general de disminución de la presión con el paso del tiempo.

La precipitación muestra picos de intensidad en diferentes momentos del año. Las variaciones en la precipitación son significativamente más abruptas que las variaciones en la presión atmosférica.

La relación entre ambas variables no es clara. Si bien podemos identificar algunos picos de precipitación que coinciden con momentos de baja presión, no se observa una relación lineal o directa entre ambos parámetros.

Es importante mencionar que esta gráfica es solo una muestra de datos y no representa una relación causal. Para comprender mejor la relación entre la presión atmosférica y la precipitación, se necesitaría un análisis más completo que incluya variables adicionales y un periodo de tiempo más amplio dado que teóricamente una menor presión demuestra la aparición de precipitaciones.

Presión atmosférica y temperatura



La línea que forma una V invertida representa la presión atmosférica, mientras que la línea que representa una V normal como un valle representa la temperatura.

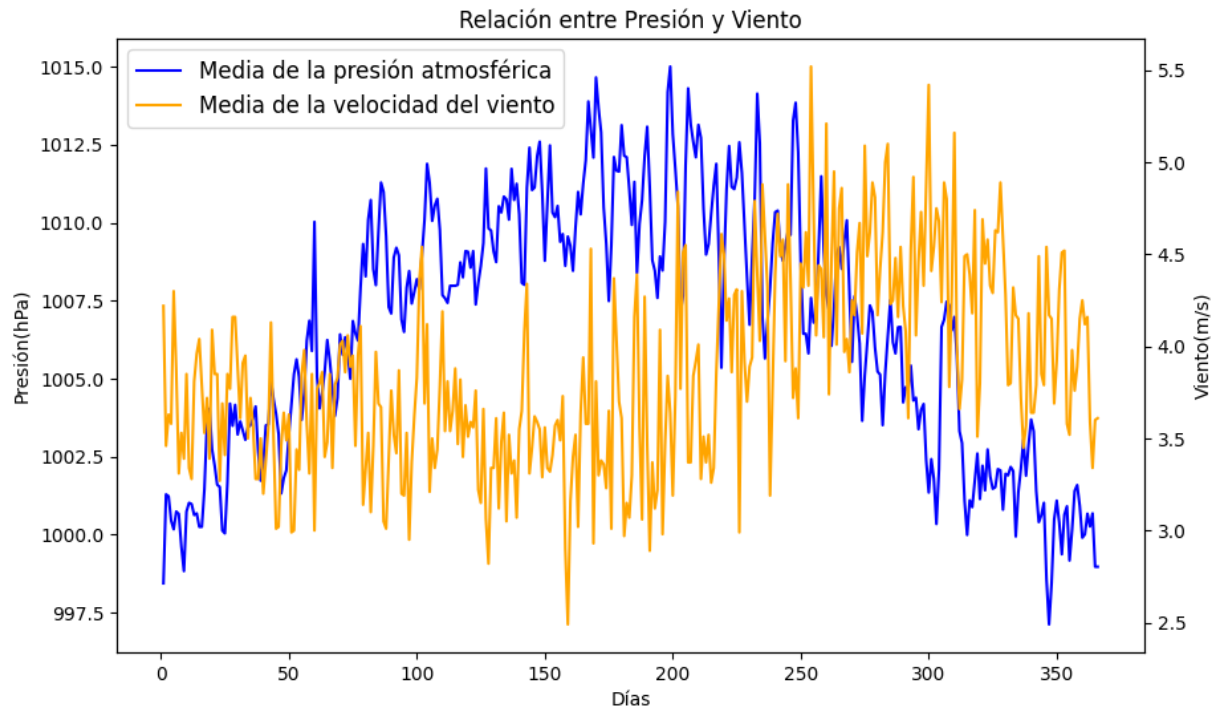
Se observa que hay una relación inversa entre las dos variables, es decir, cuando la presión atmosférica aumenta, la temperatura tiende a disminuir y viceversa. Sin embargo, la relación no es lineal, ya que se aprecian períodos en los que ambas variables se mueven en la misma dirección.

Es importante tener en cuenta que la gráfica no representa una relación de causa y efecto, sino que simplemente muestra una correlación entre las dos variables. Es posible que otras variables, como la humedad, el viento o la nubosidad, influyan en la presión atmosférica y la temperatura.

La gráfica también permite observar que la presión atmosférica es más variable que la temperatura. Se aprecia una mayor fluctuación de la presión, especialmente en los primeros meses del año.

En resumen, la gráfica muestra que existe una relación inversa y no lineal entre la presión atmosférica y la temperatura, pero no se puede afirmar que una variable cause la otra. Para obtener una comprensión más completa de la relación entre las dos variables, sería necesario analizar otros factores que puedan estar influyendo en ellas.

Presión atmosférica y velocidad del viento



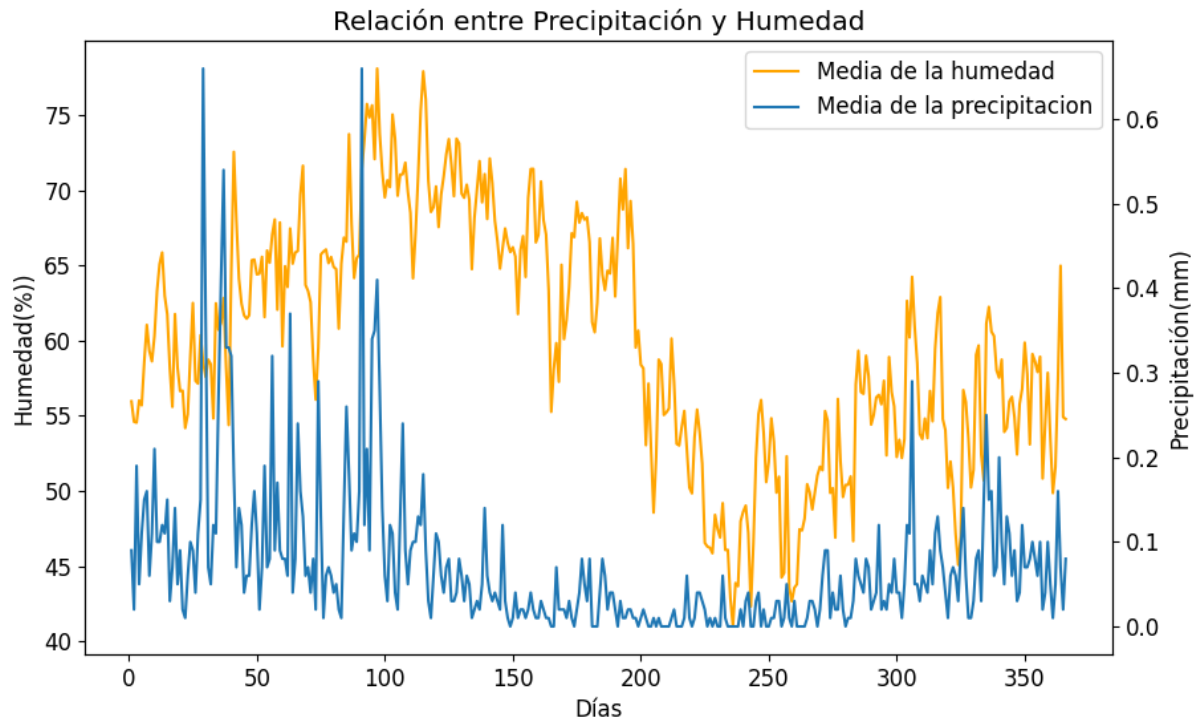
La línea que parece una V invertida representa la presión atmosférica, mientras que la otra que es similar a una onda representa la velocidad del viento a lo largo de los días del año pasado.

Se puede observar que la presión atmosférica y la velocidad del viento están correlacionadas, es decir, que tienden a variar juntas. Cuando la presión atmosférica es alta, la velocidad del viento tiende a ser baja, y viceversa.

La presión atmosférica parece oscilar entre 1000 hPa y 1015 hPa y la velocidad del viento parece oscilar entre 2.5 m/s y 5.5 m/s.

La gráfica no muestra una relación lineal perfecta entre la presión atmosférica y la velocidad del viento, pero sí una tendencia general.

Precipitación y humedad



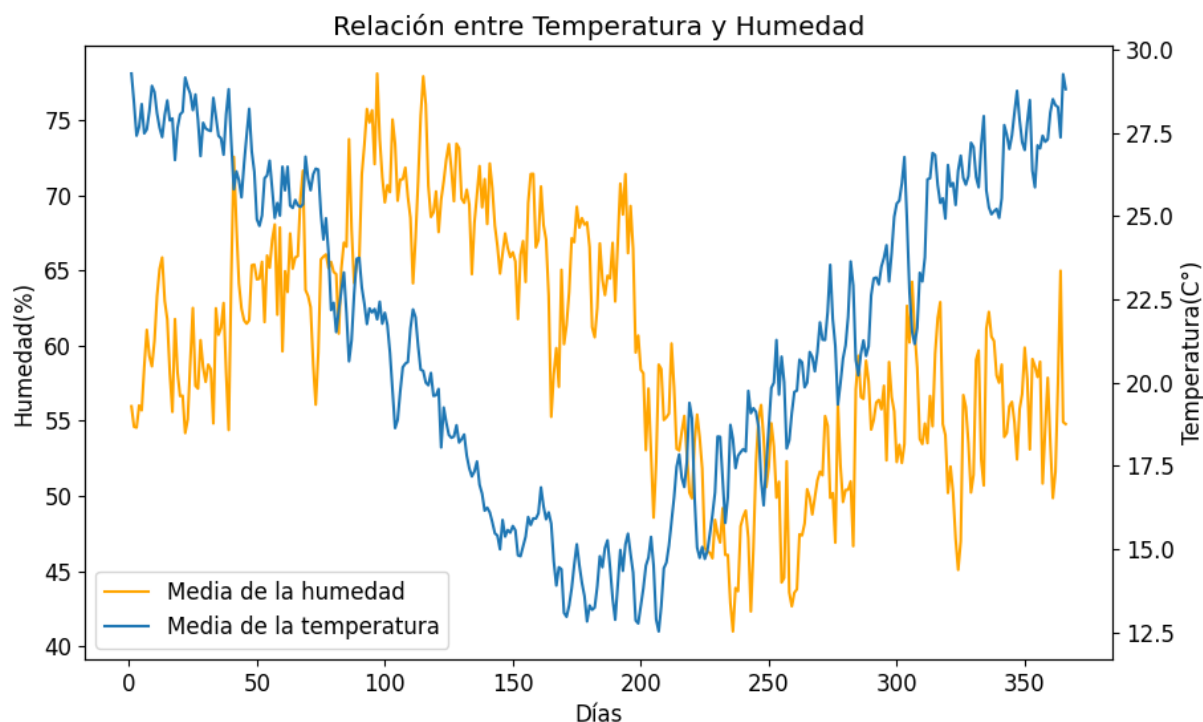
La línea que se mantiene arriba es la línea que representa la media de la humedad mientras que la otra línea representa a las precipitaciones.

Se puede observar que, en general, la humedad tiende a ser más alta cuando la precipitación es más alta. Esto es lógico, ya que la precipitación aumenta la humedad en el aire.

Sin embargo, también se puede observar que hay algunos casos en los que la humedad es alta sin que haya mucha precipitación. Esto puede deberse a la presencia de nubes que mantienen la humedad alta y que no permiten al vapor del agua evaporarse.

En general, la gráfica muestra una correlación positiva entre la humedad y la precipitación, pero también destaca que hay otros factores que pueden influir en la humedad.

Temperatura y humedad



La línea que forma una V representa la media de temperatura, y la otra línea que forma una especie de onda representa la humedad.

La gráfica muestra que la temperatura y la humedad tienen una relación inversa, es decir, cuando la temperatura aumenta, la humedad disminuye y viceversa. La temperatura alcanza su punto máximo en los meses de verano, mientras que la humedad alcanza su punto mínimo en los meses de invierno. Esto se debe a que el aire caliente puede contener más humedad que el aire frío.

La gráfica también muestra que hay un período de tiempo en el que la temperatura es más estable, pero la humedad tiene un rango más amplio de valores, esto puede ser debido a diferentes factores, como la lluvia, la evaporación, etc.

La gráfica también muestra que hay ciertos patrones en la relación entre la temperatura y la humedad a lo largo del año. Por ejemplo, en los meses de invierno, la humedad es relativamente alta y la temperatura es baja, mientras que en los meses de verano, la temperatura es alta y la humedad es baja.

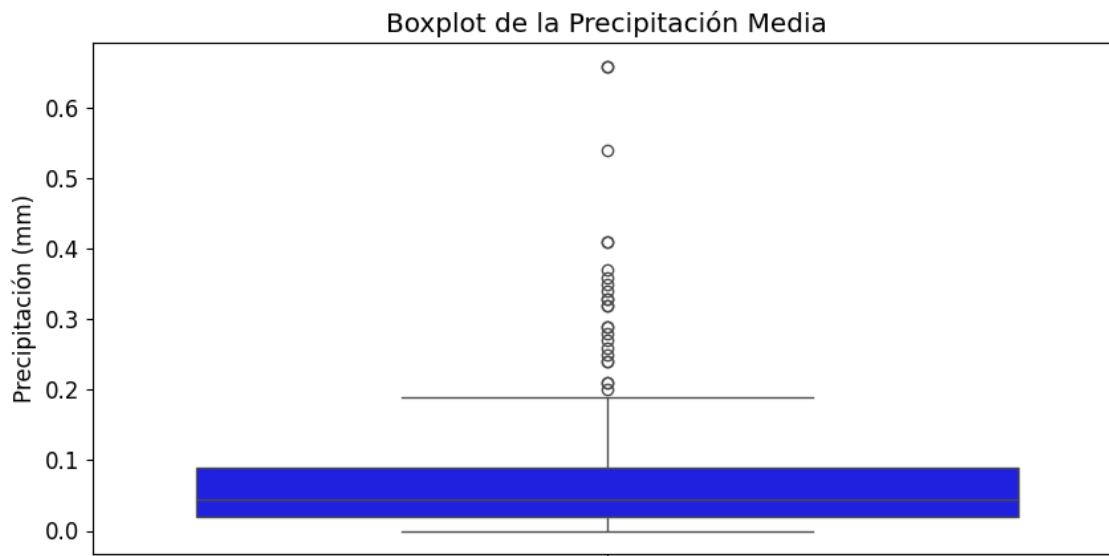
Además, se puede observar que hay ciertos puntos en los que la humedad aumenta repentinamente, a pesar de que la temperatura sigue siendo alta. Esto puede deberse a factores como la lluvia o la niebla, que aumentan la humedad en el aire.

También se puede ver que hay ciertos puntos en los que la humedad disminuye repentinamente, a pesar de que la temperatura sigue siendo baja. Esto puede deberse a factores como el viento, que puede secar el aire y disminuir la humedad.

En general, la gráfica muestra que la relación entre la temperatura y la humedad es compleja y está influenciada por una variedad de factores, incluyendo la estación del año, la lluvia, el viento y la evaporación.

Valores atípicos

En el análisis de datos tenemos herramientas para detectar valores atípicos dentro de nuestros datos, también llamados outliers. En el caso de la temperatura no se muestran valores atípicos, sin embargo en las precipitaciones si.



Este gráfico se llama boxplot, la altura de la parte coloreada está creada con una resta, la diferencia entre el percentil 75 y el percentil 25, estos dos son llamados cuartiles dado que dividen en cuatro partes iguales los datos, en este caso sería el primer y el tercer cuartil. Además esta diferencia se le llama **rango intercuartíl (RIQ)**.

La línea del medio muestra la mediana de los datos. Si los datos son simétricos la mitad de los datos estaría por encima de esa línea y la otra mitad por debajo, dado que en este caso no lo son la caja no es simétrica.

Las líneas que se extienden de la caja reciben el nombre de bigotes, estos se extienden exactamente 1.5 veces el rango intercuartílico y pueden no extenderse a lo largo de todos los datos. Si algunos datos salen fuera de los bigotes, estos datos se les llama outliers.

Si observamos la caja de las precipitaciones podremos observar que hay parámetros que se ven en forma de puntos fuera de los bigotes de la caja, estos son outliers que representan lluvias anormales altas.

En el tratamiento de datos, estos valores representan un problema y se debe tomar una decisión, de si eliminarlos o tratarlos de otra manera, en este caso particular decidí no tratarlos ni eliminarlos dado que representan valores útiles para saber los momentos donde hubo altas precipitaciones.

Con esto concluye el análisis del clima de Santiago del Estero y pasaremos a la parte final que tiene que ver con la predicción y las series temporales.

3. Machine Learning aplicado al clima y análisis de series temporales

Predicciones: elección de variables objetivo

Para predecir variables se debe elegir las como objetivo, es decir debemos definir variables objetivo, y definir con qué variables se predecirán, es decir definir variables predictoras o características.

Como se quiere predecir las variables climáticas de lo que queda del año, considero que las variables más importantes para predecir, es decir las variables objetivo son el promedio de las siguientes:

- **Precipitaciones:** Porque la media de precipitaciones nos servirá para observar durante las fechas como va creciendo o decreciendo la cantidad de lluvia en los días y para observar si hay alguna predicción de anomalía o desastre como inundaciones, ciclones o huracanes.
- **Temperatura:** La media de temperaturas es muy útil a la hora de ver tendencias y observar la estacionalidad
- **Humedad:** La media de humedad con el pasar de los días es muy importante dado que tiene correlación con las otras dos variables, para saber que tan agradable es un clima hace falta saber la humedad, y la media de la humedad nos proporciona una buena medida dado que si el aire es demasiado húmedo, existe el riesgo de que se forme moho.
- **Velocidad del viento:** El viento actúa como agente de transporte, en efecto, interviene en la polinización anemófila, en el desplazamiento de las semillas. Es también un poderoso agente erosivo, en especial en las zonas de clima seco o desértico, donde los granos de arena arrastrados por el viento pueden llegar a la transformación y hasta la denudación (es decir, la completa remoción) de las formas del relieve. También actúa como agente de sedimentación, ya que cuando el viento pierde velocidad, deposita los materiales que transporta. Por lo que saber la media de la velocidad del viento es muy importante para muchas áreas.

Todas las otras variables funcionarán como variables predictoras, esto es así para que se pueda predecir mediante correlaciones el valor del siguiente día, y así sucesivamente. Además de que muchos datos presentan correlación con el tiempo debido a la estacionalidad.

Para realizar el entrenamiento del modelo de regresión se dividen los datos en dos grandes conjuntos, el de pruebas y el conjunto de datos de entrenamiento. El conjunto de pruebas es necesario para evaluar el rendimiento del modelo elegido.

El proceso de establecer y elegir los modelos se da en base al problema, en este caso particular se usaron cuatro modelos de regresión.

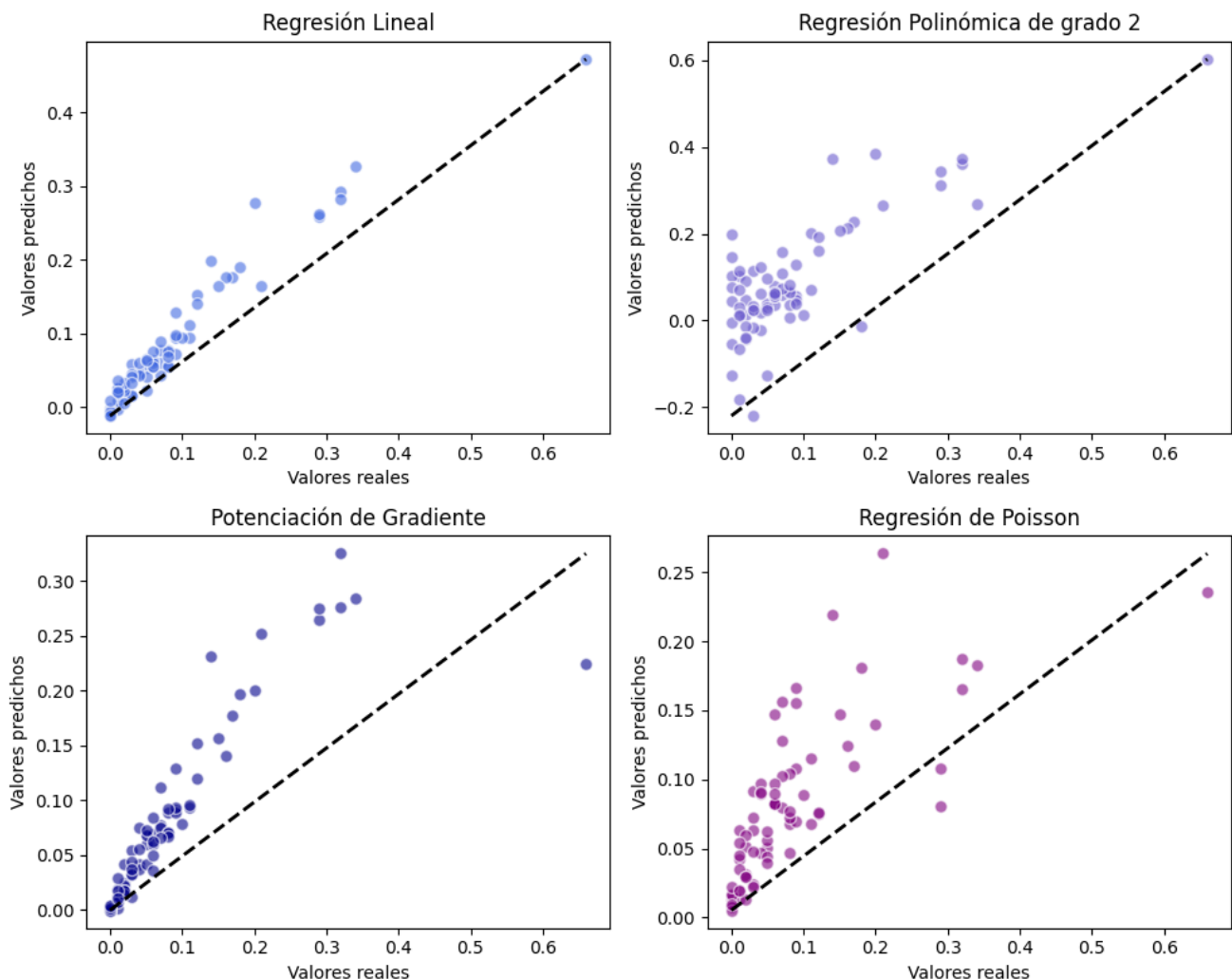
La regresión lineal, la regresión polinómica de grado 2, la regresión de Poisson y la regresión de potenciación de gradiente.

Los modelos de regresión son modelos matemáticos que buscan determinar y evaluar la relación entre una variable dependiente (Y), con respecto a otras variables, llamadas explicativas o independientes (X).

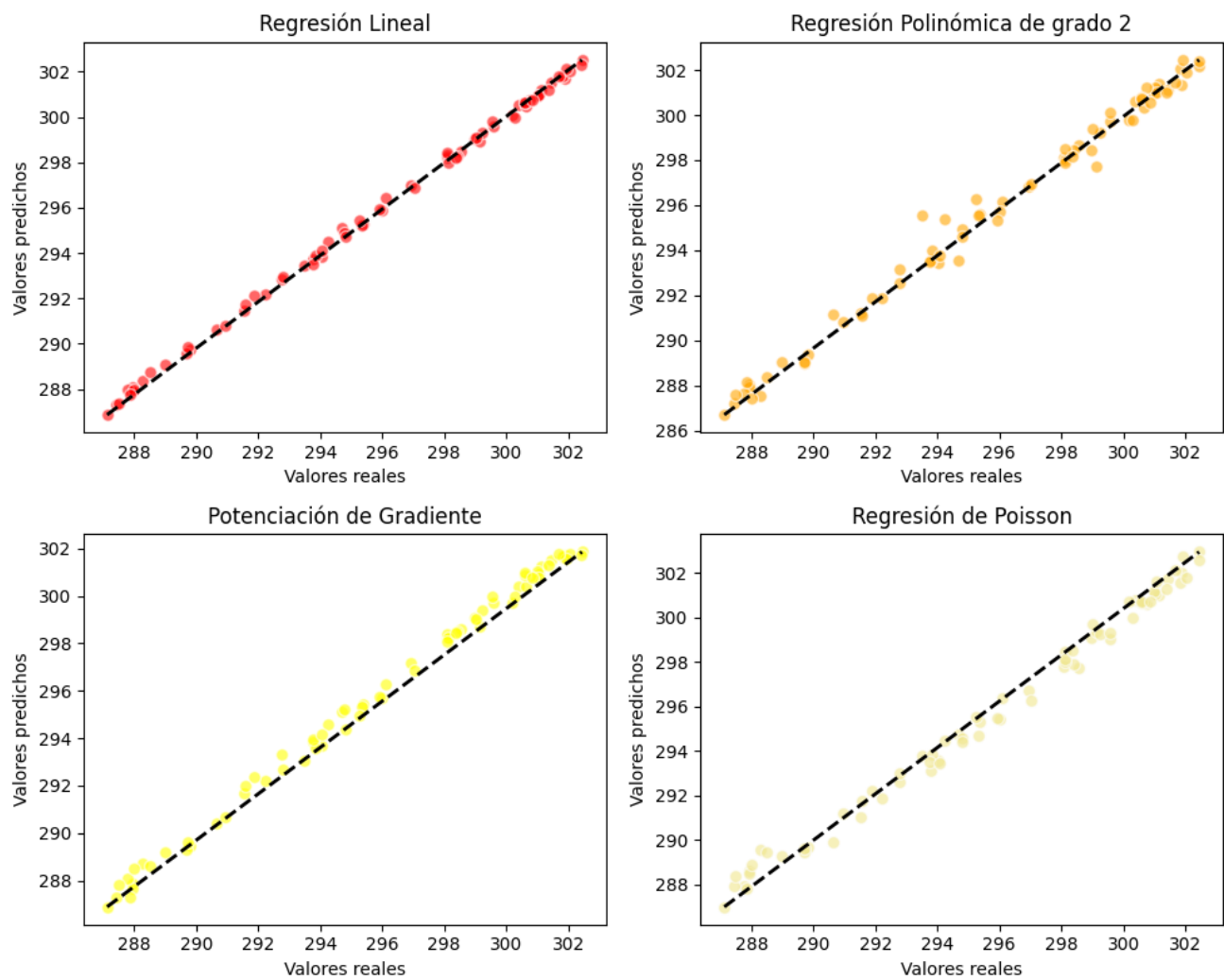
No todos se pueden aplicar de cualquier manera a cualquier variable, por ejemplo el modelo de regresión lineal no se puede aplicar cuando mi variable de interés es una variable categórica. Como en este caso como se intenta predecir la media de las precipitaciones, temperatura, humedad y viento se observarán cuales regresiones son las mejores para cada una.

Para evaluar los modelos podemos observar su precisión mediante estas gráficas.

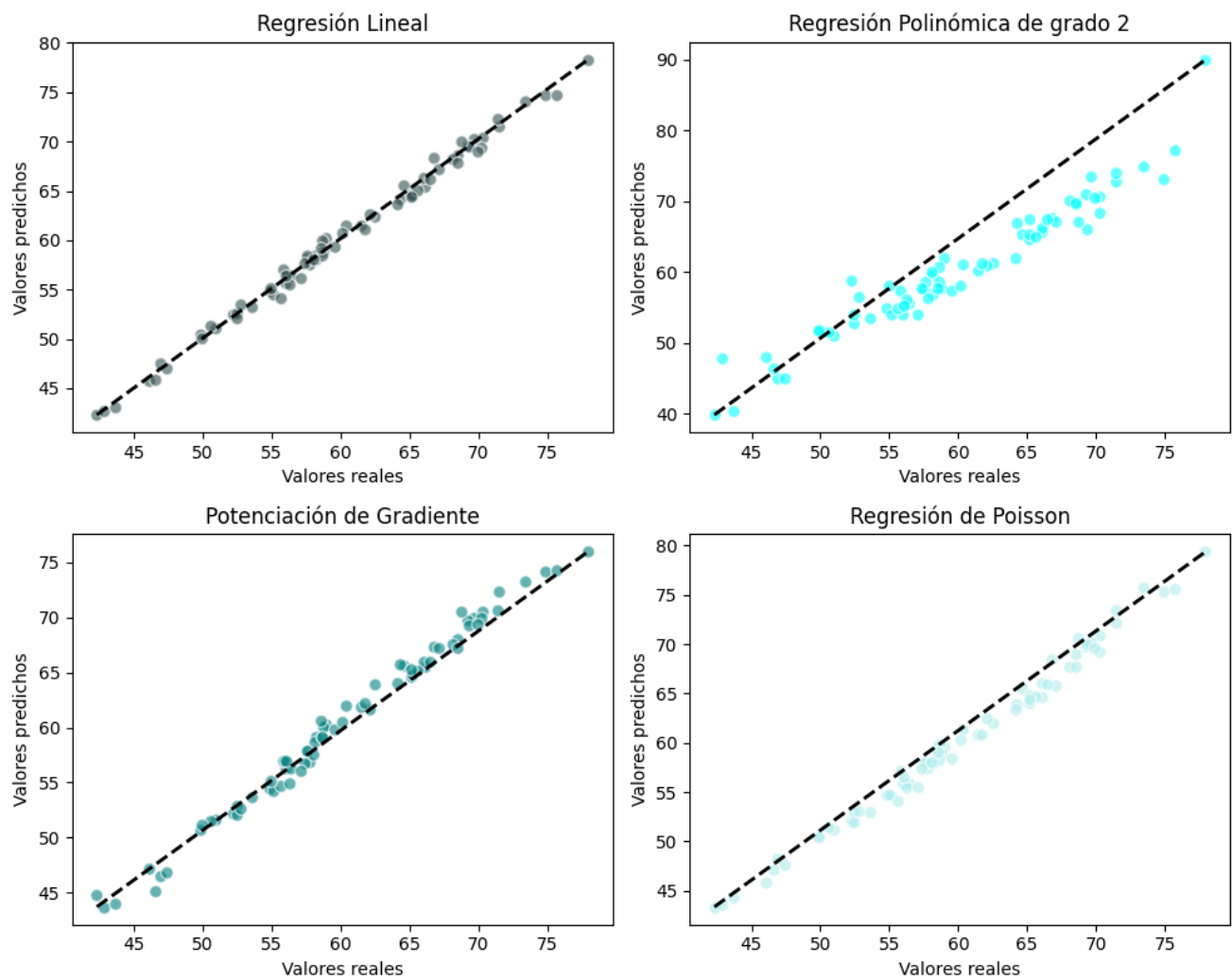
Precisión de las predicciones de precipitación



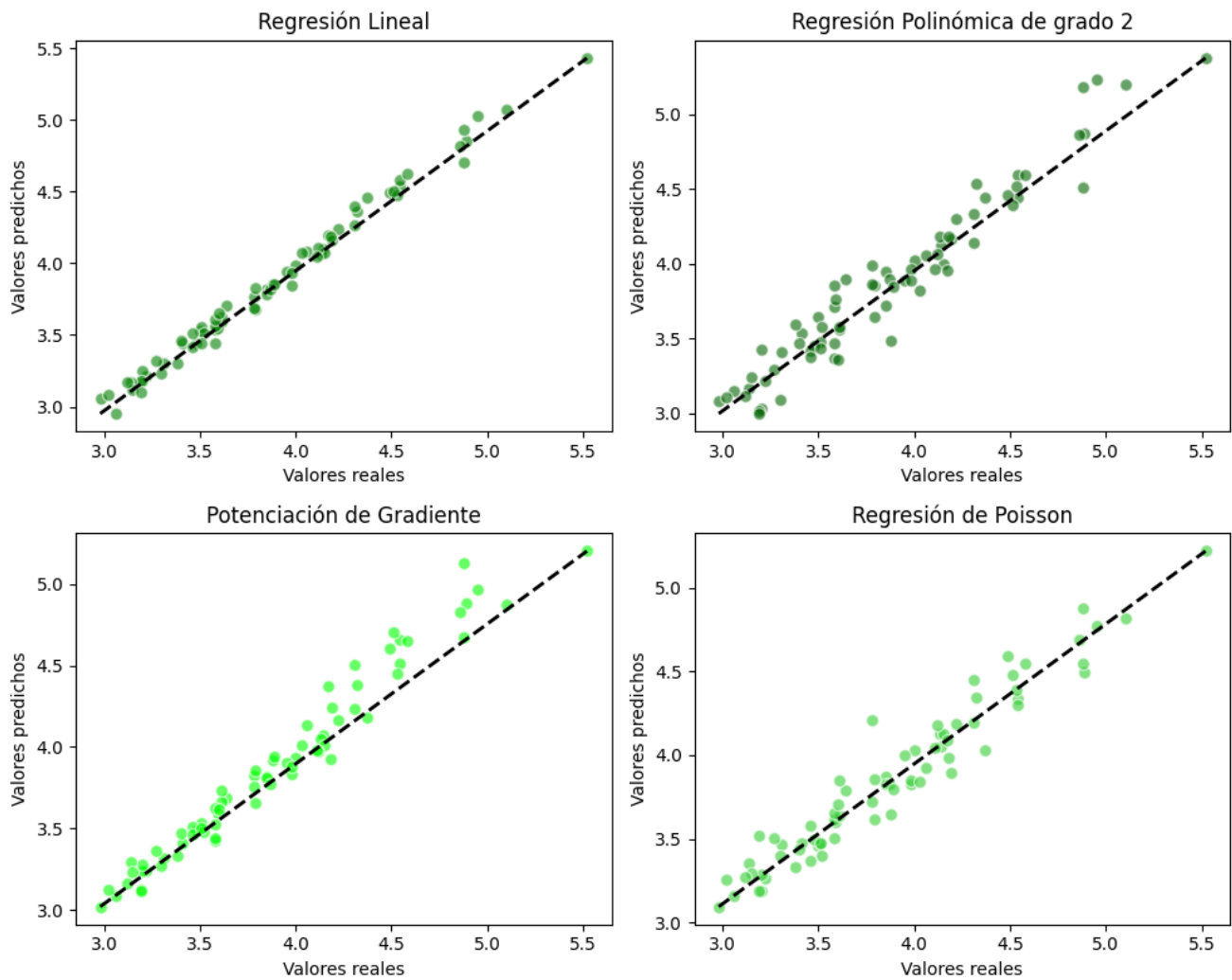
Precisión de las predicciones de temperatura



Precisión de las predicciones de humedad



Precisión de las predicciones de viento



En base a los gráficos de precisión se decidió usar el modelo de regresión lineal y entrenarlo para predecir las variables del promedio del viento, humedad y temperatura. Para la variable restante es decir la precipitación, se decidió usar el modelo de regresión de Poisson.

Predicciones y análisis de tendencias con series temporales realizadas para el resto del 2024

Los siguientes gráficos muestran series de tiempo divididas en cuatro componentes: observada, tendencia, estacional y residual.

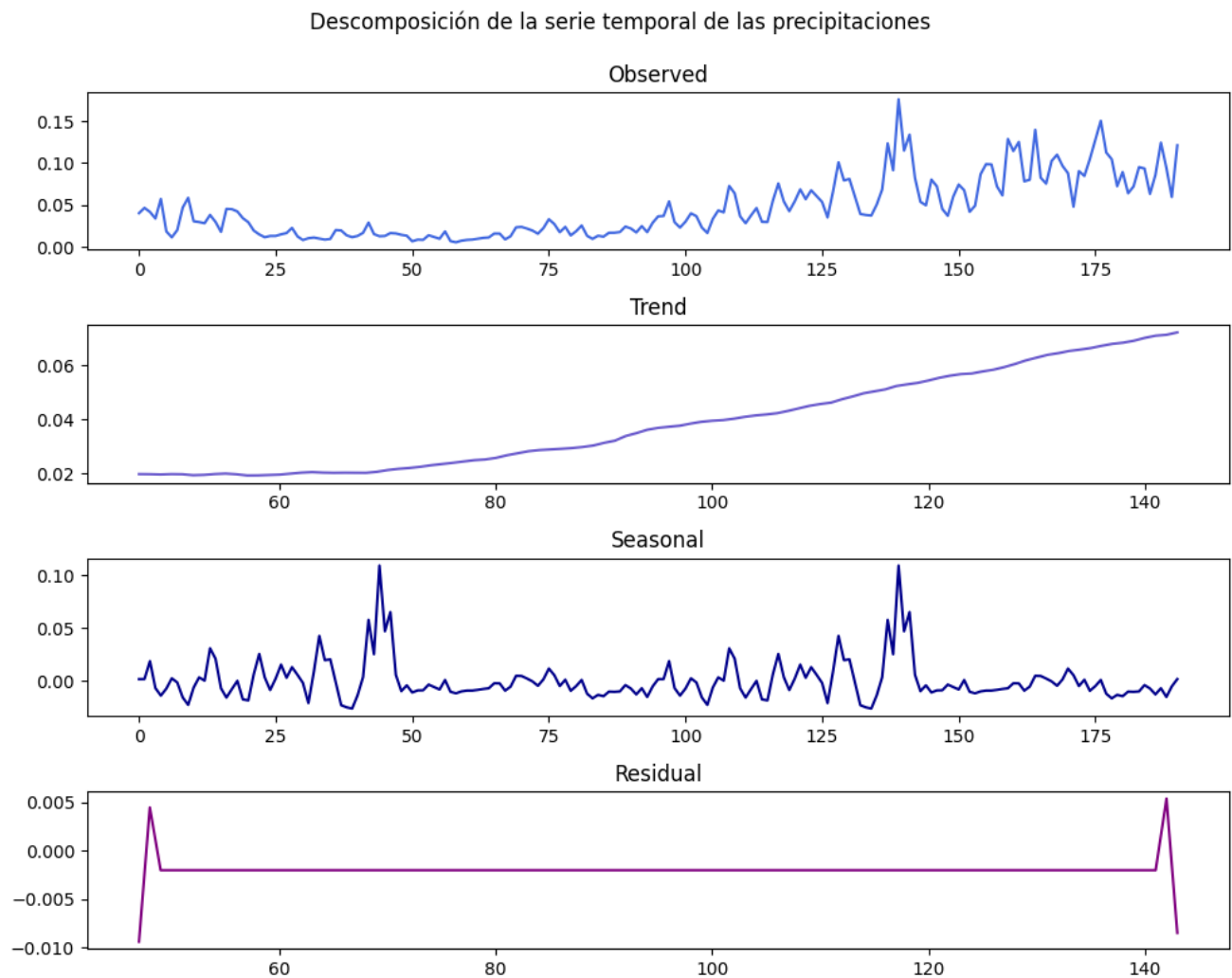
- **Observada:** La primer línea representa el valor observado de la serie de tiempo, en este caso es el valor predicho a futuro.
- **Tendencia:** La segunda línea representa la tendencia de la serie de tiempo.
- **Estacional:** La tercer azul representa la variación estacional de la serie de tiempo.
- **Residual:** La cuarta y última lineal representa la parte residual de la serie de tiempo, es decir, la variación que no se explica por la tendencia ni por la estacionalidad. Muestra que hay una parte de la variación de la serie de tiempo que no se explica por las componentes anteriores.

Una serie temporal o cronológica es una sucesión de datos medidos en determinados momentos y ordenados cronológicamente. Los datos pueden estar espaciados a intervalos iguales (como la temperatura en un observatorio meteorológico en días sucesivos al mediodía) o desiguales (como el peso de una persona en sucesivas mediciones en el consultorio médico, la farmacia, etc.).

Para el análisis de las series temporales se usan métodos que ayudan a interpretarlas y que permiten extraer información representativa sobre las relaciones subyacentes entre los datos de la serie o de diversas series y que permiten en diferente medida y con distinta confianza extrapolar o interpolar los datos y así predecir el comportamiento de la serie en momentos no observados, sean en el futuro (extrapolación pronóstica), en el pasado (extrapolación retrógrada) o en momentos intermedios (interpolación).

Las series temporales se usan para estudiar la relación causal entre diversas variables que cambian con el tiempo y se influyen entre sí. Desde el punto de vista probabilístico una serie temporal es una sucesión de variables aleatorias indexadas según parámetro creciente con el tiempo. Cuando la esperanza matemática de dichas variables aleatorias es constante o varía de manera cíclica, se dice que la serie es estacionaria y no tiene tendencia secular. Muchas series temporales tienen una tendencia creciente (por ejemplo, el número de automóviles en uso en casi todos los países durante los últimos cincuenta años) o decreciente (por ejemplo, el número de personas que trabajan en la agricultura); otras no tienen tendencia (la luminosidad a horas sucesivas, que varía cíclicamente a lo largo de las 24 horas del día) y son estacionarias.

Predicción y serie temporal de la precipitación



La línea de las precipitaciones muestra que la serie de tiempo presenta variaciones, con algunos picos y valles que se repiten de manera regular.

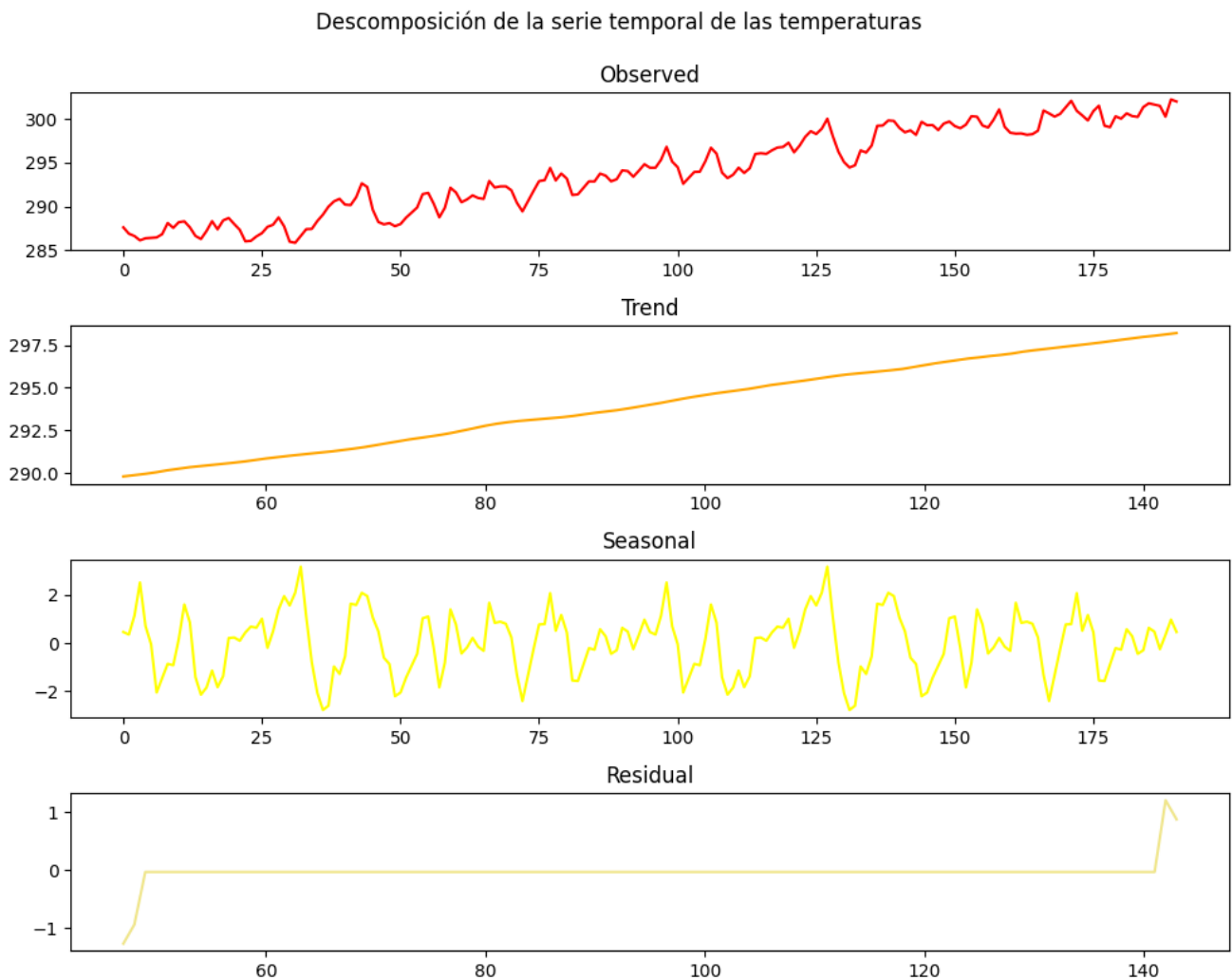
La línea de tendencia muestra una tendencia general al alza, lo que indica un incremento gradual en la variable de interés.

La línea de estacionalidad muestra un patrón regular de variación, con algunos picos y valles que se repiten a lo largo del año.

En general, el gráfico muestra que la serie de tiempo presenta una tendencia al alza, lo que podría indicar un cambio en el clima de la región, con variaciones estacionales significativas y una parte residual no explicada.

Esta parte residual sin embargo es relativamente pequeña, lo que sugiere que el modelo de machine learning tiene una buena precisión en la predicción de las precipitaciones medias.

Predicción y serie temporal de la temperatura



Cabe destacar que la temperatura predicha está en grados Kelvin.

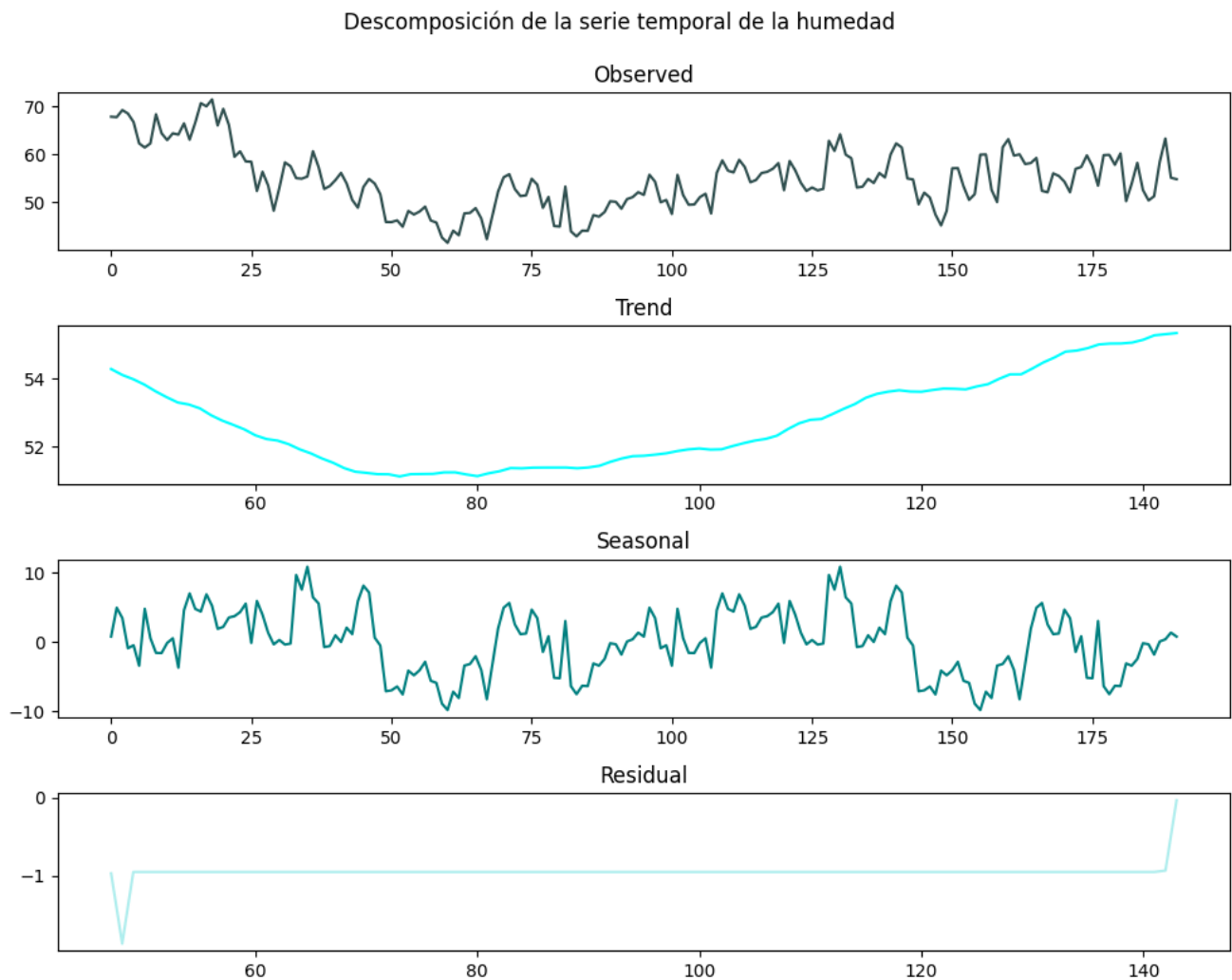
La tendencia muestra un claro aumento a lo largo del tiempo, lo que indica que la temperatura promedio ha aumentado durante el período considerado y aumentará a la larga.

La estacionalidad, por otro lado, muestra patrones regulares de variación a lo largo del año, con picos en verano y valles en invierno.

La variación a corto plazo se refiere a las fluctuaciones en la temperatura que no siguen un patrón regular y pueden ser causadas por diversos factores, como la variabilidad climática natural o la influencia humana.

Por último, los residuos representan el error de la descomposición y pueden ser el resultado de factores no considerados en el análisis. Esto se puede interpretar como el ruido aleatorio o las variaciones impredecibles de la temperatura.

Predicción y serie temporal de la humedad

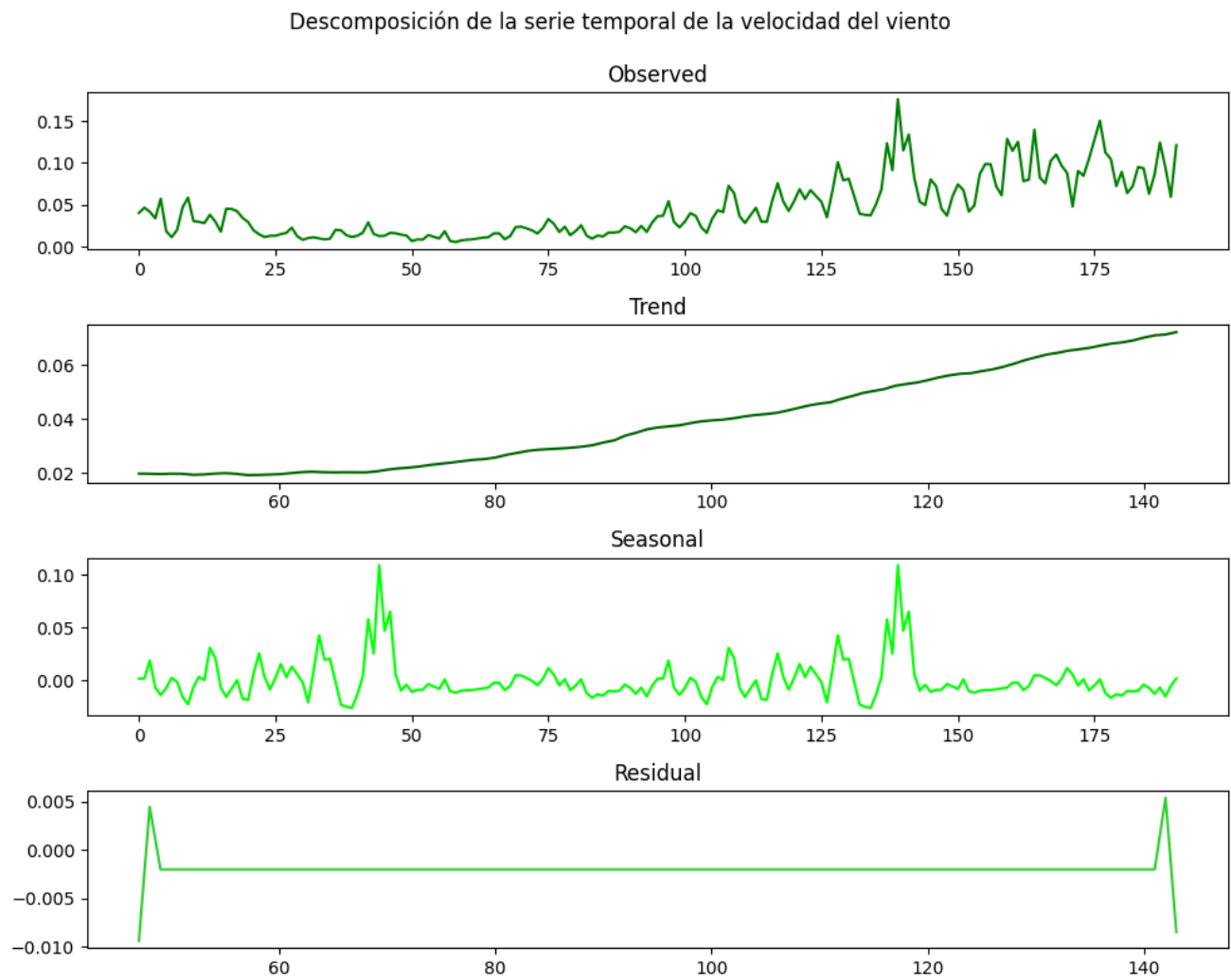


Se puede ver que la serie temporal tiene una tendencia decreciente en un principio, es decir que la humedad disminuirá con el tiempo y luego una tendencia creciente que se debe al componente estacional.

También se puede observar una variabilidad estacional, con picos de humedad en ciertas épocas del año y valles en otras. La componente estacional es cíclica, con picos y valles que se repiten cada cierto tiempo.

Se puede observar que el componente residual tiene una variabilidad aleatoria y no muestra ninguna tendencia o patrón estacional.

Predicción y serie temporal de la velocidad del viento



La tendencia es una línea creciente suave, que indica que la velocidad del viento ha aumentado a lo largo del tiempo.

La estacionalidad es un patrón repetitivo, que muestra picos en la velocidad del viento en ciertos momentos del año.

El residuo es la parte de la serie temporal que no se explica por la tendencia y la estacionalidad, y es aleatorio.

Anomalías en las predicciones

Durante los entrenamientos de los modelos para realizar predicciones, en algún momento puede haberse producido o, un valor raro o una anomalía.

Para encontrar este valor raro se puede usar por ejemplo la desviación estándar en el cálculo de la predicción, sin embargo esto no es del todo preciso dado que se trata de un valor raro, y para diferenciarlo de una anomalía haría falta más precisión

En este punto se entrenó un modelo de machine learning llamado Isolation Forest.

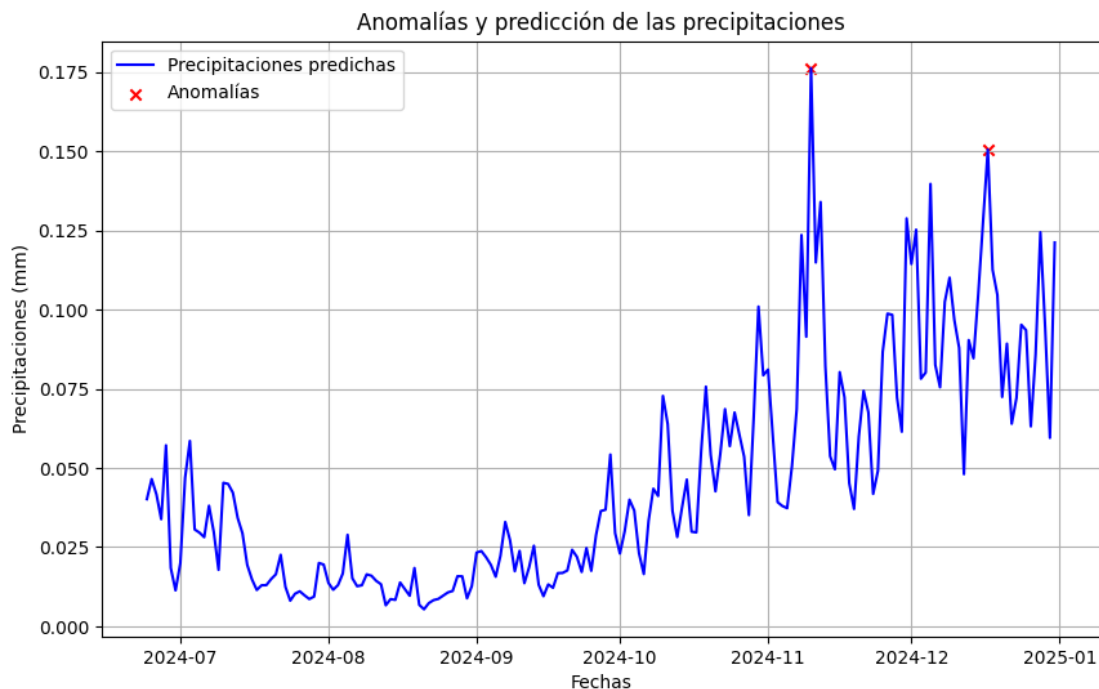
Isolation Forest es un algoritmo para la detección de anomalías en los datos desarrollado inicialmente por Fei Tony Liu en 2008. Isolation Forest detecta anomalías utilizando árboles binarios.

El algoritmo tiene una complejidad temporal lineal y requiere poca memoria, por lo que funciona bien con grandes volúmenes de datos, aunque en este caso particular no se están usando muchos datos funciona muy bien.

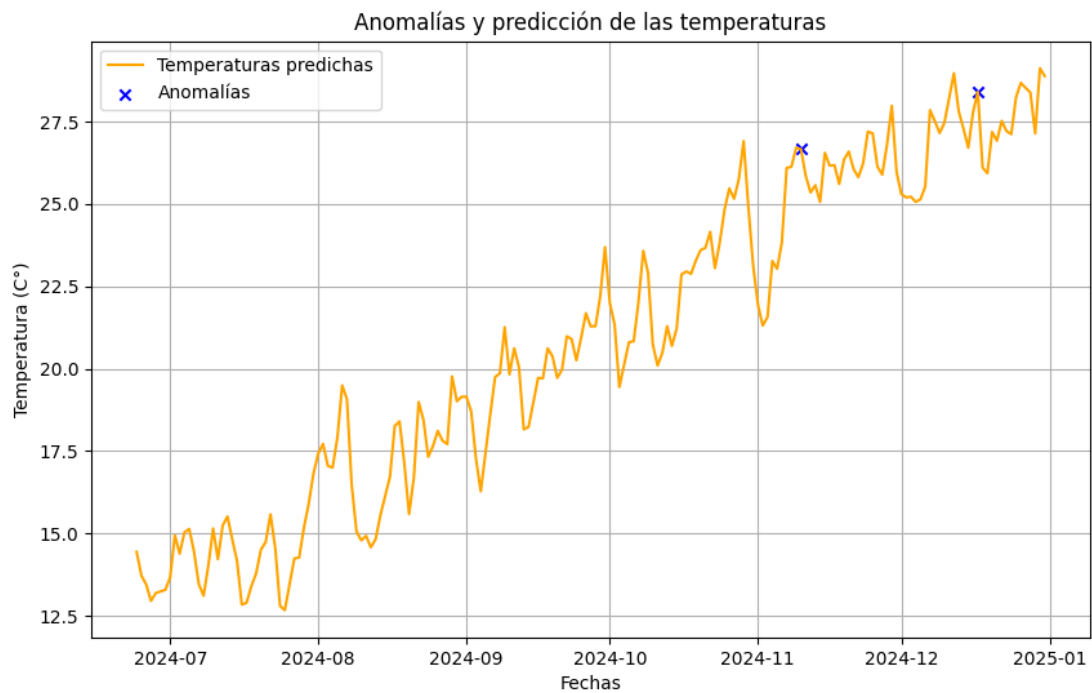
En esencia, el algoritmo se basa en las características de las anomalías, es decir, que sean pocas y diferentes, para detectarlas. En el algoritmo no se realiza ninguna estimación de la densidad. El algoritmo se diferencia de los algoritmos de árbol de decisión en que sólo se utiliza la medida o aproximación de la longitud del camino para generar la puntuación de la anomalía, no se necesitan estadísticas de los nodos hoja sobre la distribución de clases o el valor objetivo

En este caso particular el algoritmo detectó dos anomalías en las distintas gráficas.

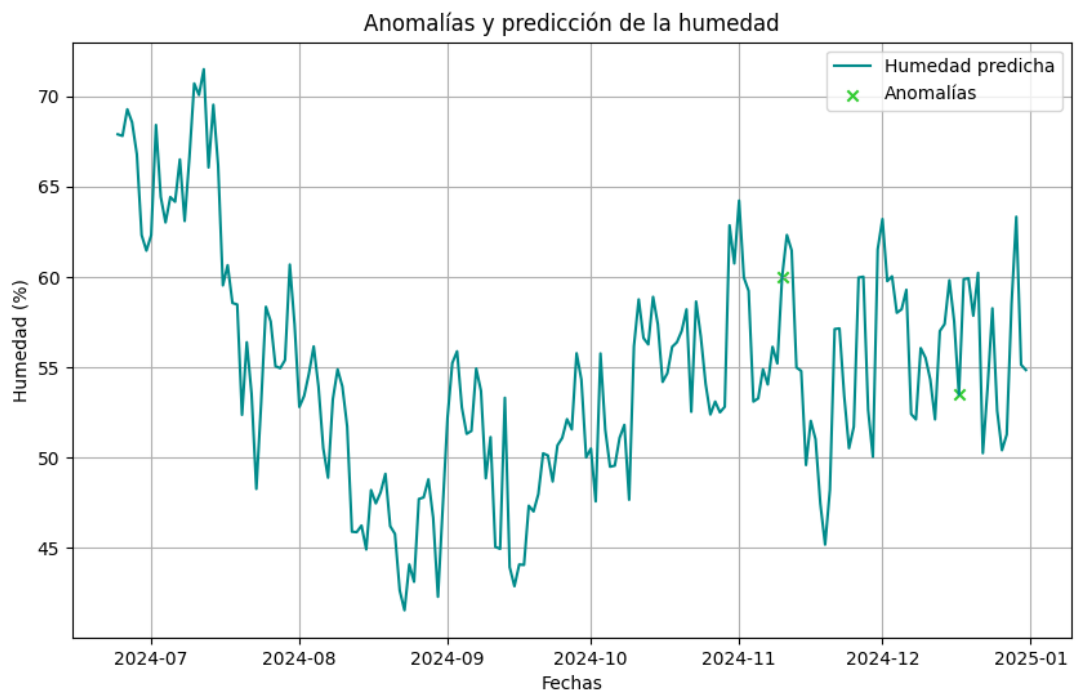
Anomalías en las precipitaciones



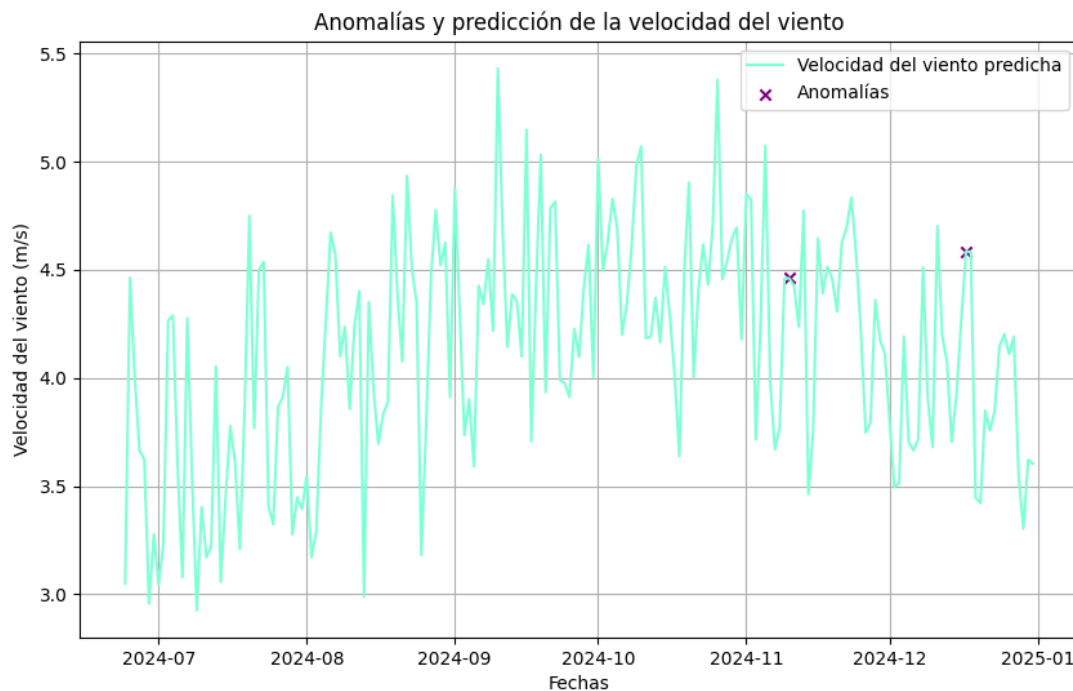
Anomalías en la temperatura



Anomalías en la humedad



Anomalías en la velocidad del viento



En este contexto, una anomalía es un valor que se desvía significativamente de los valores esperados o normales. Las anomalías pueden ser causadas por errores de medición, eventos inusuales o cambios en el sistema que se está monitoreando.

Se puede observar que el modelo de regresión lineal ajusta bastante bien los datos en las gráficas de la temperatura, humedad y velocidad del viento, y el modelo de regresión de Poisson ajusta bien los datos de la precipitación, capturando la tendencia general de las variables. Sin embargo, el algoritmo Isolation Forest ha detectado algunas anomalías, que se representan como puntos X en las gráficas. Estas anomalías sugieren que las variables han presentado valores inusuales en comparación con el comportamiento general de la serie de tiempo.

Las posibles razones por las que se podrían estar presentando anomalías en los datos:

- Errores de medición: Los sensores que miden la humedad pueden fallar o presentar errores, lo que puede provocar valores inusuales. Aunque en este caso particular es descartable dado que se trata de una predicción hecha en Machine Learning por los modelos.
- Eventos climáticos: Eventos climáticos como lluvias intensas o sequías pueden provocar cambios bruscos en las variables, esta es la posibilidad a la que más me decanto.
- Cambios en el entorno: Cambios en el entorno, como la construcción de nuevos edificios o la deforestación, pueden afectar las variables, aunque en este caso es muy improbable debido a que no se prevee en los modelos ningún tipo de cambios en el entorno.

4. Conclusión

En este proyecto, se demostró que el uso de técnicas de machine learning, como modelos de regresión, puede ser una herramienta poderosa para predecir variables climáticas clave en Santiago del Estero. La detección de anomalías, la identificación de tendencias y la validación de predicciones con datos históricos son aspectos fundamentales para tomar decisiones informadas en diversas áreas como la agricultura, la gestión de recursos y la preparación ante desastres naturales e incluso en la planificación urbana. Este proyecto de taller resalta la importancia de la ciencia de datos aplicada al clima y su potencial para mejorar la resiliencia y la adaptación frente a los desafíos climáticos.

Estos hallazgos no sólo mejoran nuestra comprensión del clima local si no que también proporcionan información valiosa. A través del monitoreo continuo y la adaptación estratégica, Santiago del Estero pueden mejorar su capacidad para enfrentar los desafíos climáticos y proteger a sus habitantes y recursos además, al utilizar técnicas avanzadas de visualización y análisis de datos, hemos podido identificar y interpretar tendencias clave que pueden guiar acciones prácticas y estratégicas en la región.