

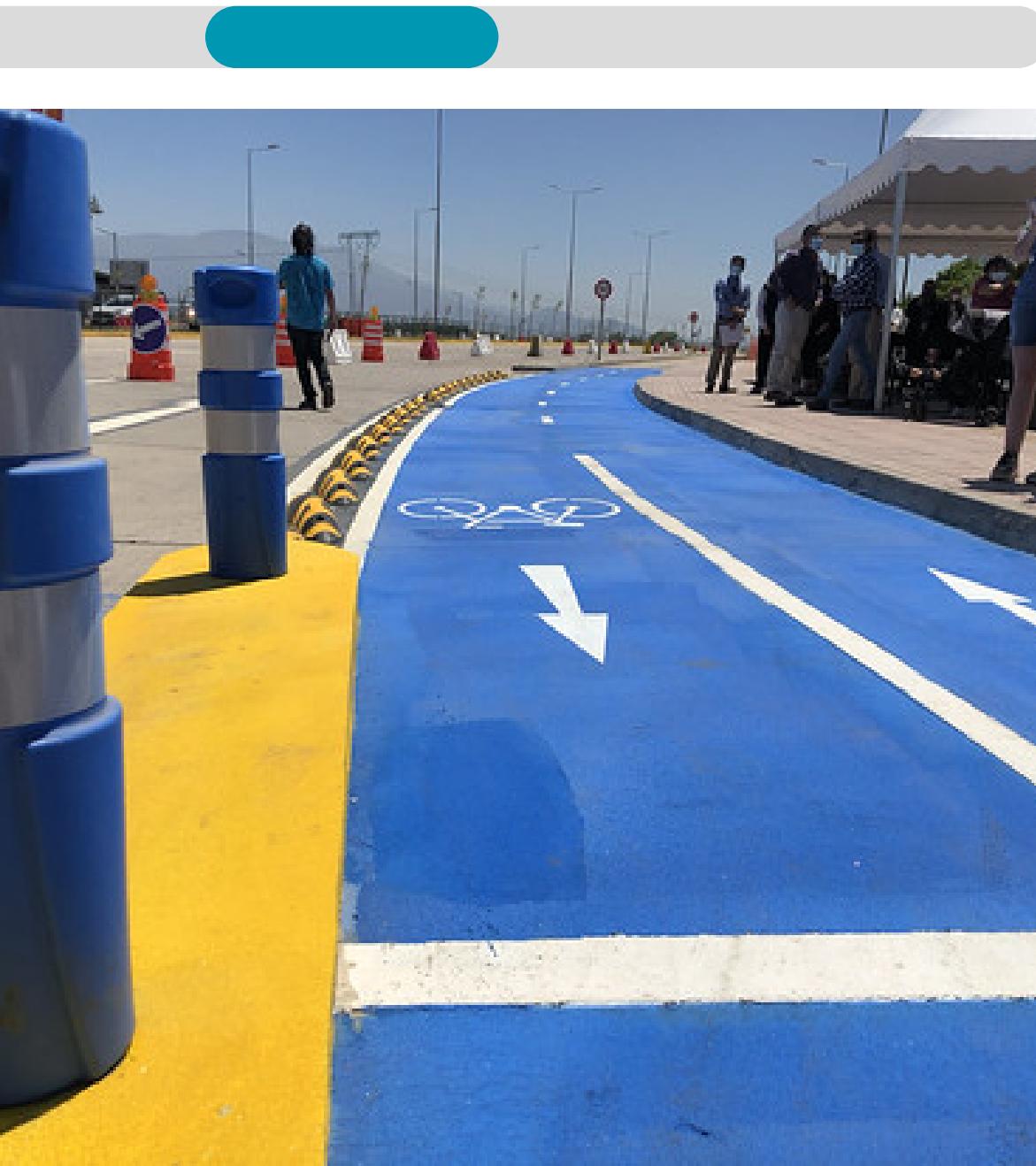


ANÁLISIS DE SERIES DE TIEMPO

Predicción del Flujo de Ciclistas

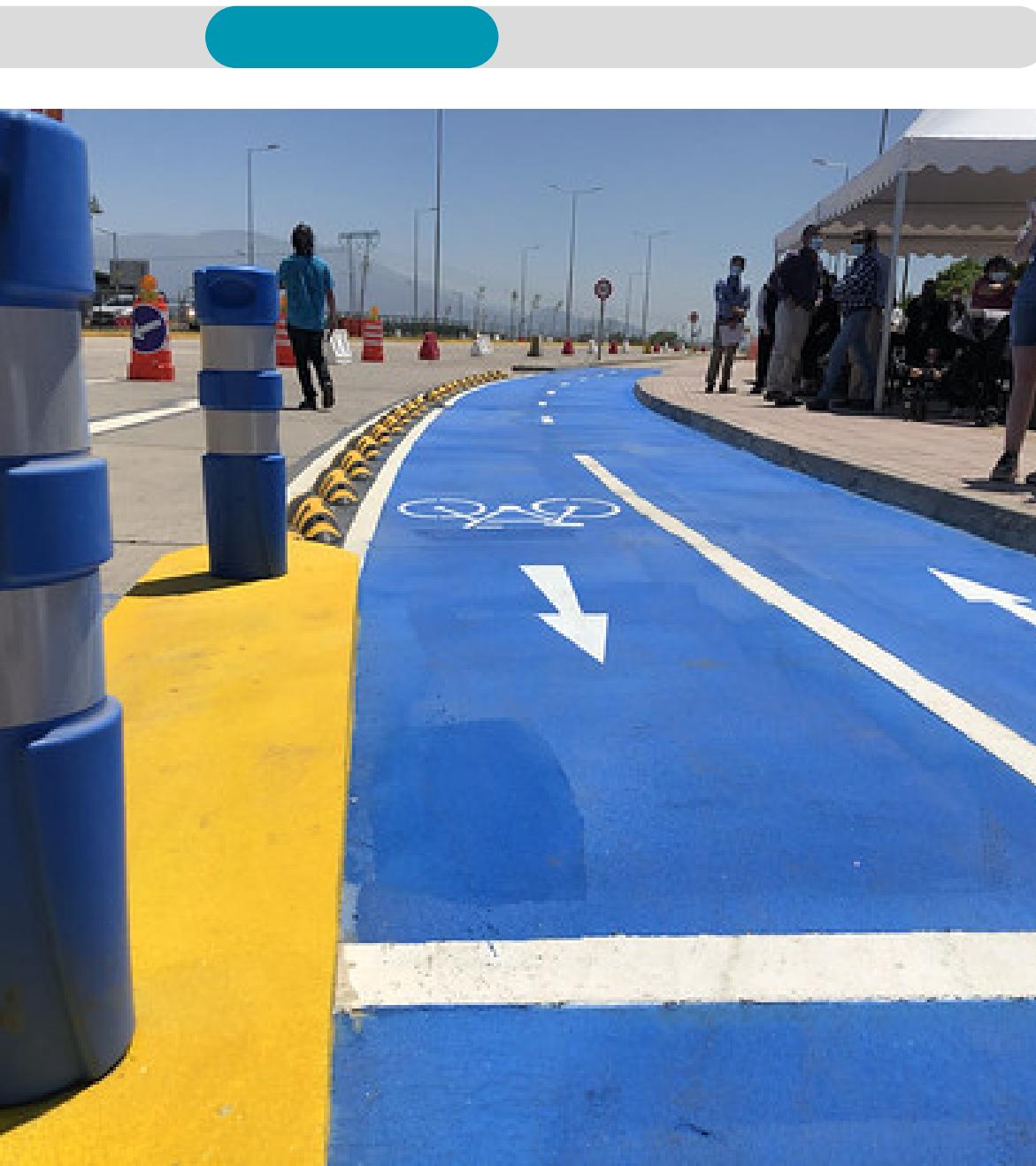
GABRIEL DÍAZ - MARIA DROGUETT - IGNACIO JIMÉNEZ

índice



- 01 Introducción
- 02 Motivación
- 03 Análisis exploratorio
- 04 Prophet
- 05 Arima
- 06 Conclusión

índice

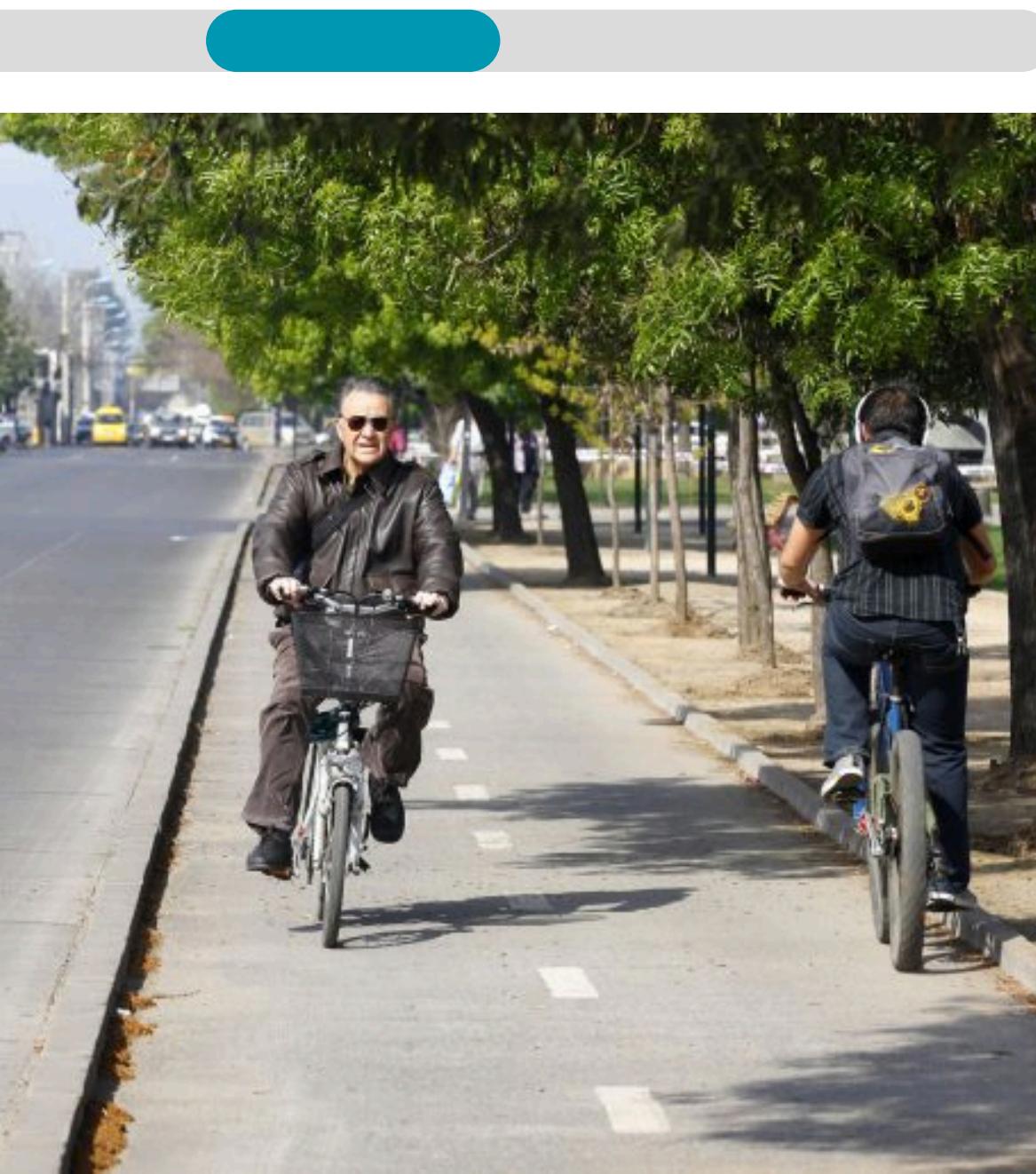


- 01 Introducción
- 02 Motivación
- 03 Análisis exploratorio
- 04 Prophet
- 05 Arima
- 06 Conclusión

Introducción



Conozcamos las ciclovías



Calle Alameda

- Inaugurada en junio de 2015
- Proyecto para promover el ciclismo urbano

Calle Cabello

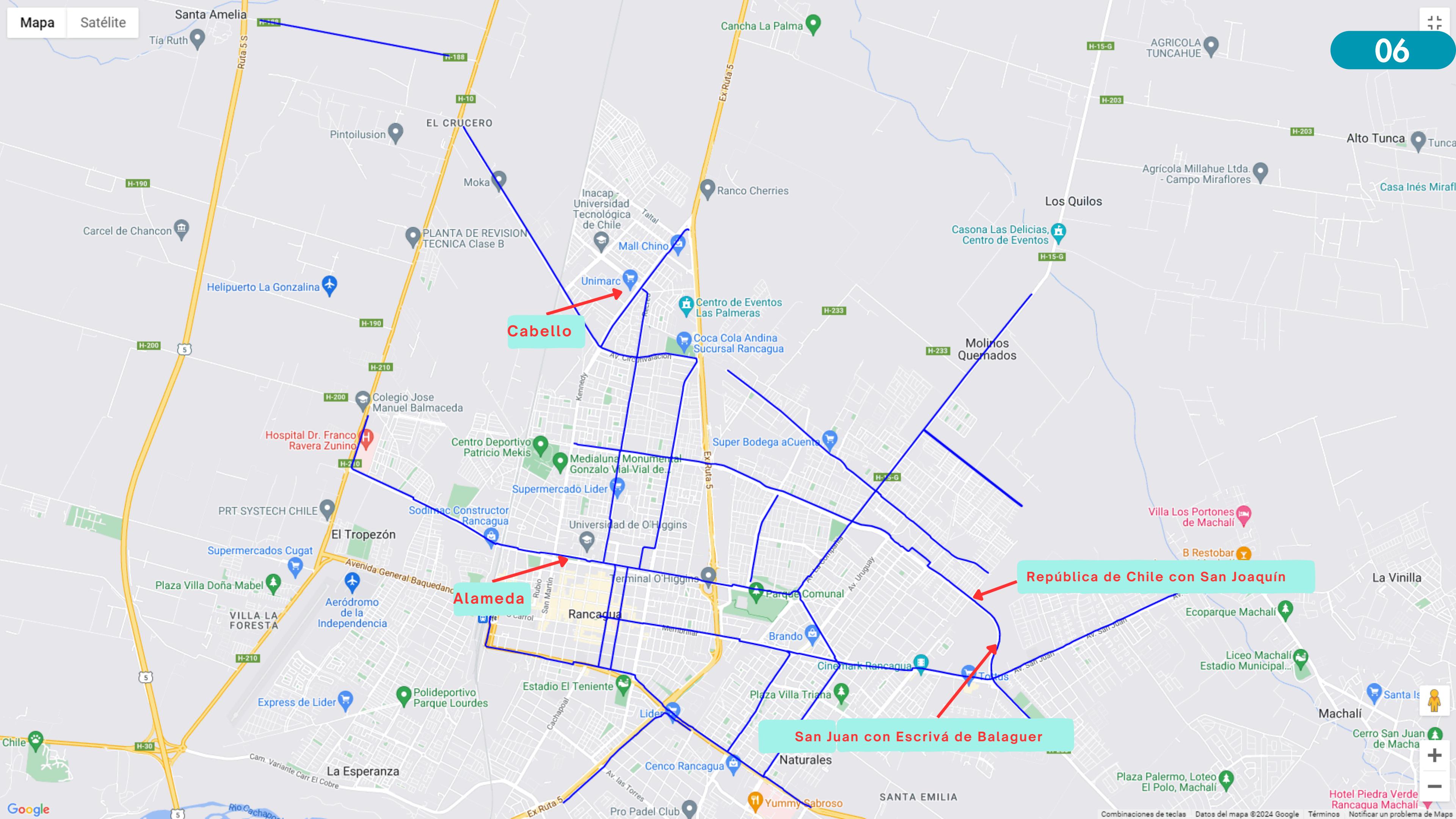
- Inaugurada en septiembre de 2018
- Proyecto de mejoramiento de la infraestructura para bicicletas en Rancagua

Calle República de Chile con San Joaquín

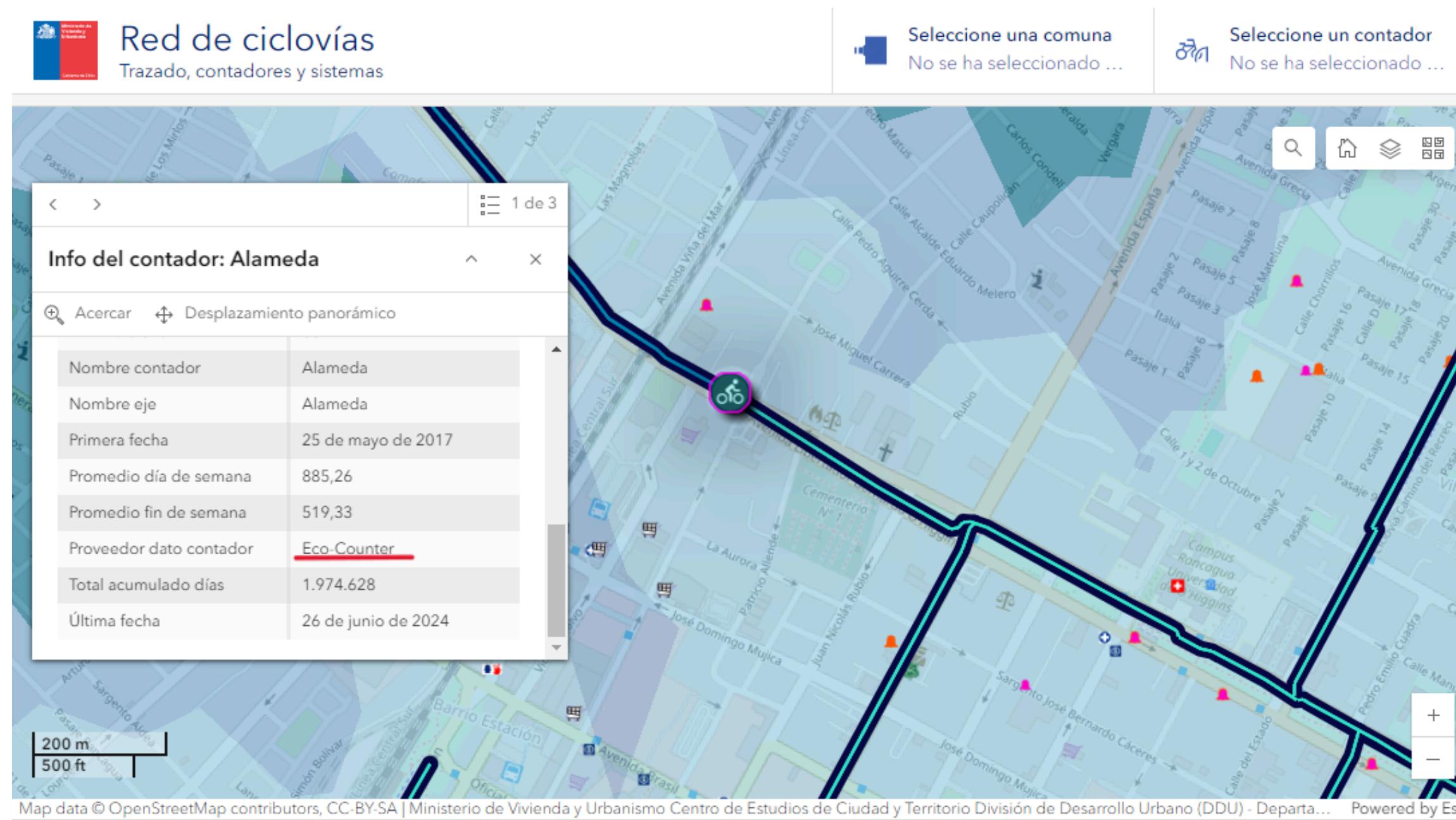
- Inaugurada en 2020
- Proyecto de mejoramiento de la infraestructura para bicicletas en Rancagua

Calle San Juan con Escrivá de Balaguer

- Inaugurada en 2020
- Proyecto de mejoramiento de la infraestructura para bicicletas en Rancagua



Sensor ciclovía calle Alameda



Fuente:<https://www.minvu.gob.cl/ciclovias-minvu/>

Sensor Gamma MULTI Urbano





Gama MULTI

Multi Urbano

Contador de peatones y bicicletas en entorno urbano

Movilidad Sostenible

Ventajas

- Tendencias fiables a largo plazo
- Detecta el sentido del paso
- Tecnología no intrusiva
- Alcance : hasta 6 m
- No requiere mantenimiento

Pedir un presupuesto

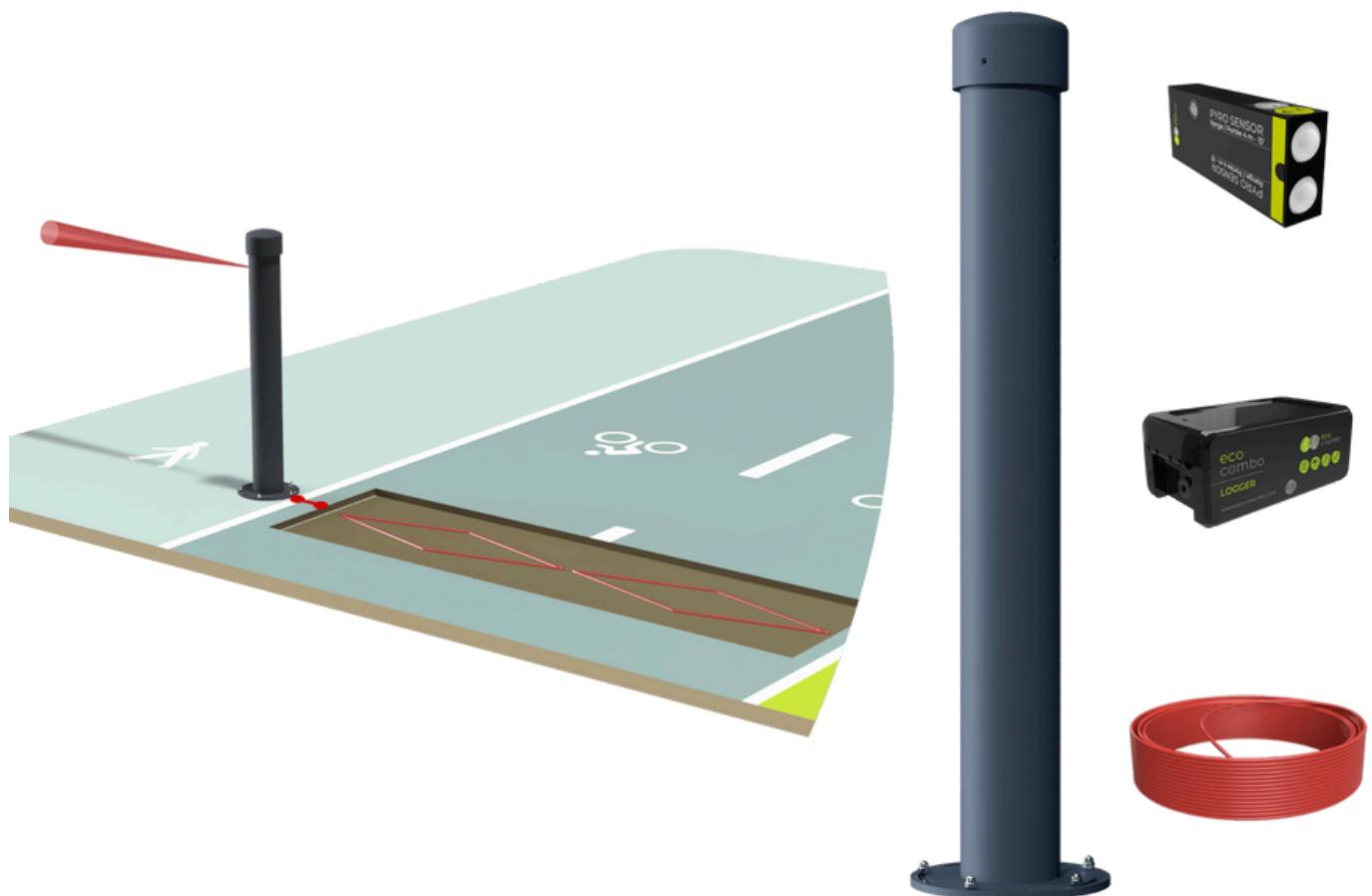


↑↓ 🔍 ✖️ ⚠️ IP 66 2 ans ⚡

co-Compteur

Fuente: <https://es.eco-counter.com/produits/gama-multi/multi-urbain-2/>

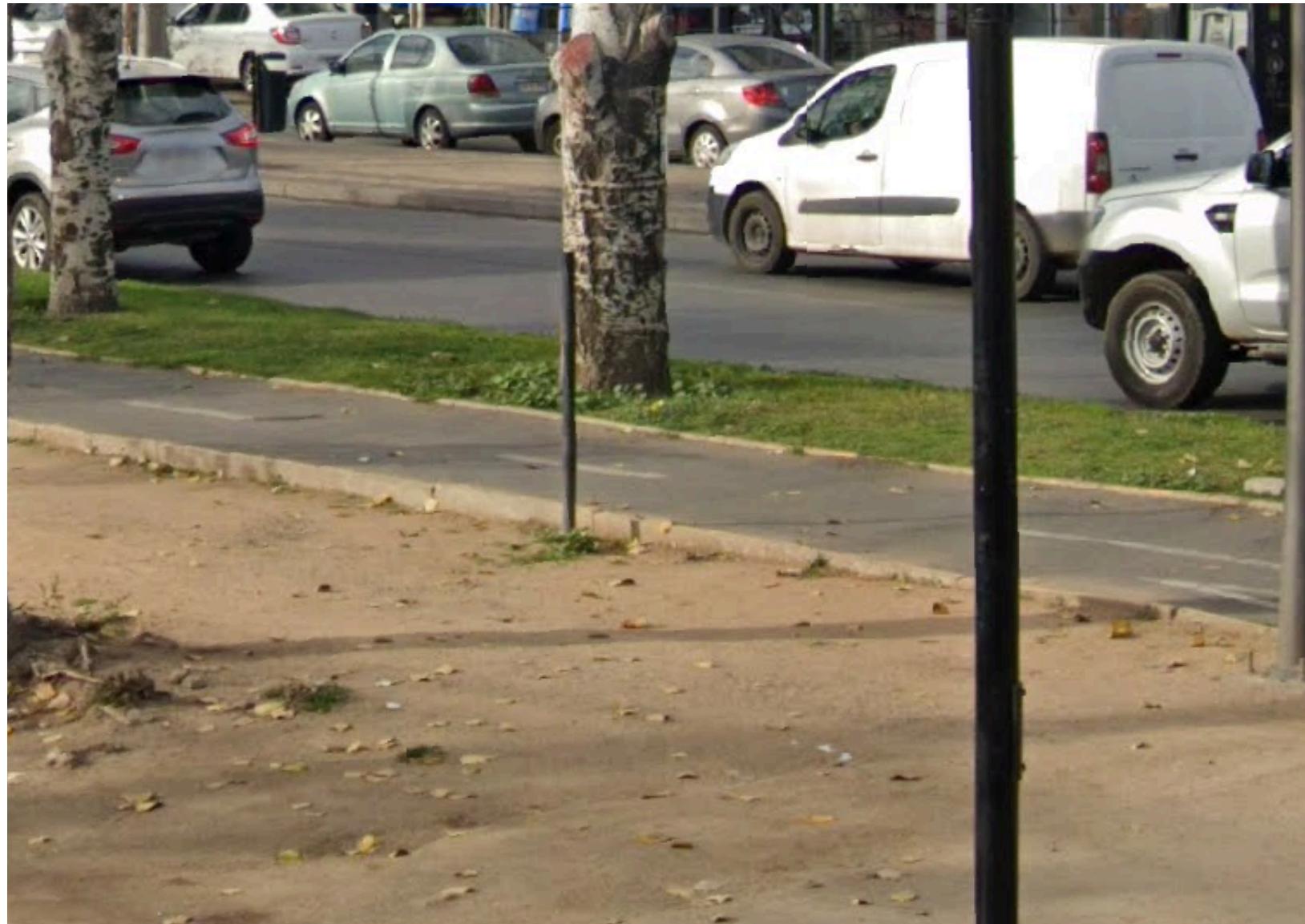
Sensor Gamma MULTI Urbano



¿Cómo funciona?

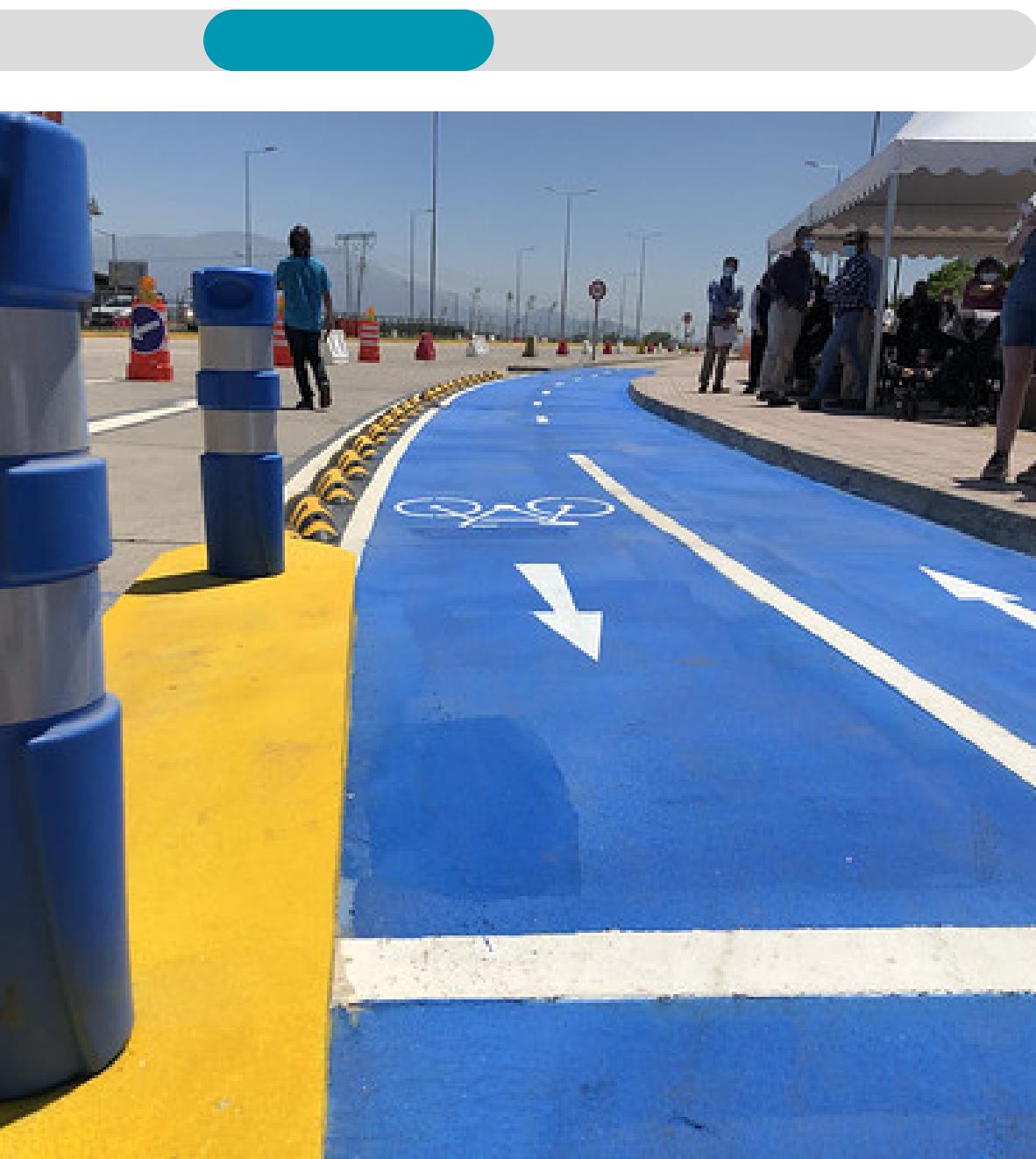
- El Multi es un contador innovador y flexible.
- Diferencia varias prácticas (ciclistas, peatones, jinetes, etc.) urbanas y rurales.
- El sistema se compone de sensores basados en diferentes tecnologías.
- Las diferentes prácticas son gestionadas por un sistema inteligente, el Smart Connect, capaz de tomar decisiones de prioridad. Los datos se recogen y almacenados por un Logger, llamado Eco-Combo, luego se analizan mediante el software en línea Eco-Visio.

Sensor ciclovía calle Alameda



Fuente: Elaboración Propia

índice



- 01 Introducción
- 02 Motivación
- 03 Análisis exploratorio
- 04 Prophet
- 05 Arima
- 06 Conclusión

Problemática



- A menudo, la capacidad de prever cambios en el uso de bicicletas y adaptar la infraestructura urbana en consecuencia se ve limitada por la falta de herramientas analíticas avanzadas y modelos predictivos adecuados.
- Esto dificulta la planificación estratégica a largo plazo, la optimización de recursos y la implementación oportuna de políticas públicas que promuevan un transporte urbano más sostenible y saludable.



Objetivo General y específicos

- Desarrollar un modelo predictivo basado en series de tiempo para estimar la cantidad de bicicletas.
- Proporcionar información para las autoridades gubernamentales, de esta manera comprender las tendencias en el uso de bicicletas y tomar decisiones sobre políticas públicas relacionadas con el transporte sostenible y la promoción del ciclismo como medio de transporte en la ciudad.

RECOPILACIÓN Y PREPARACIÓN DE DATOS:

01

- Recopilar datos históricos sobre la cantidad de bicicletas utilizadas en el área de interés.
- Realizar un proceso de limpieza de datos para prepararlos adecuadamente para el análisis.

ANÁLISIS Y MODELADO:

02

- Realizar un análisis exploratorio de los datos recopilados.
- Investigar diversos modelos de series de tiempo para seleccionar el más apropiado.
- Entrenar el modelo seleccionado.

EVALUACIÓN Y AJUSTE DEL MODELO:

03

- Evaluar el rendimiento del modelo entrenado utilizando métricas adecuadas.
- Realizar ajustes adicionales en el modelo según sea necesario para mejorar su rendimiento.

Preguntas de investigación



¿CUÁLES SON LAS TENDENCIAS HISTÓRICAS EN EL USO DE BICICLETAS EN LAS CALLES ESTUDIADAS? 01

02 ¿CÓMO VARÍA EL FLUJO DE BICICLETAS EN DIFERENTES MOMENTOS DEL AÑO?

¿EXISTEN PATRONES ESTACIONALES EN EL FLUJO DE BICICLETAS EN LA REGIÓN DE ESTUDIO? 03

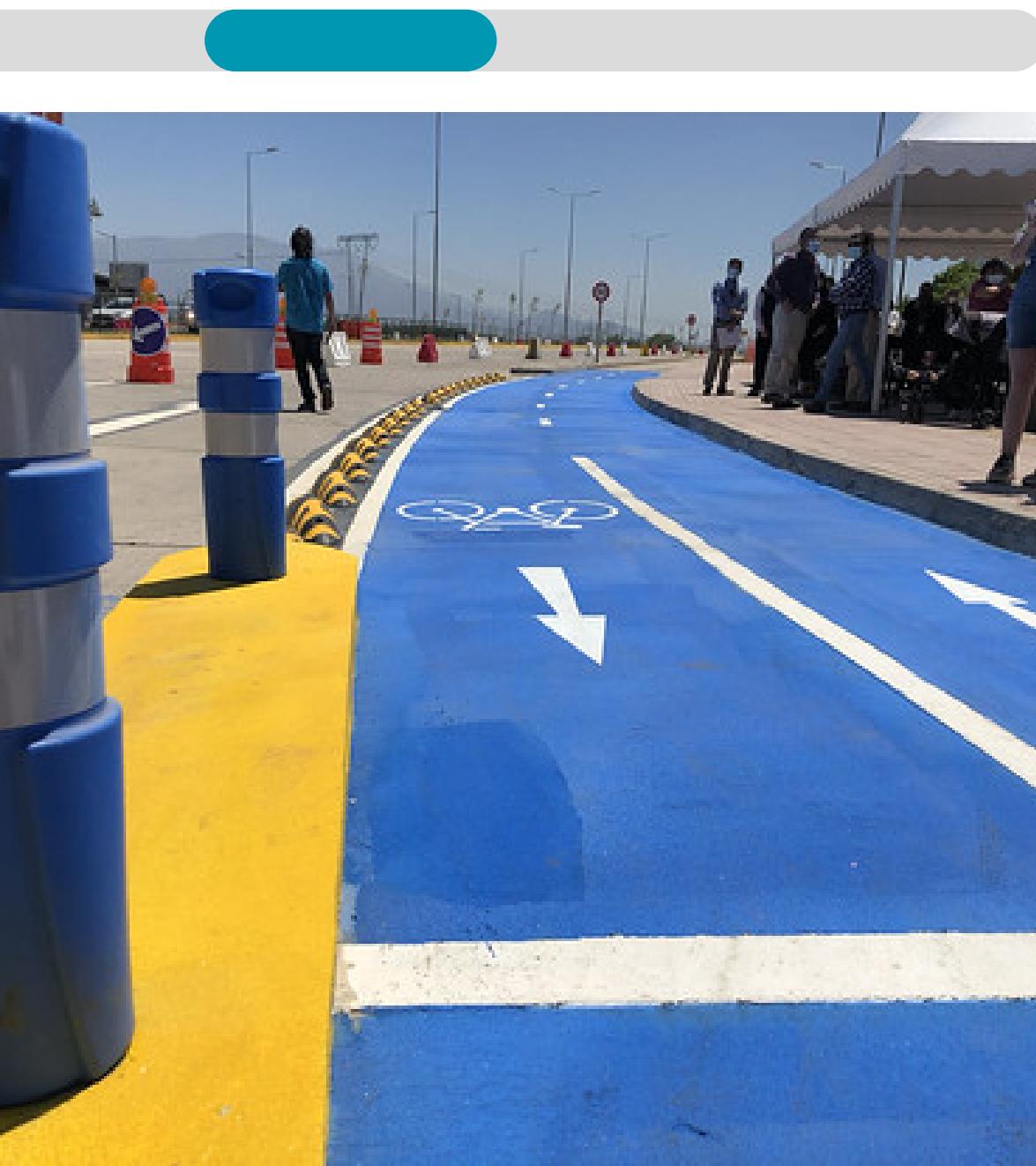
04 ¿HAY ALGÚN EVENTO O FACTOR EXTERNO QUE INFLUYA SIGNIFICATIVAMENTE EN EL FLUJO DE BICICLETAS?

Preguntas de investigación



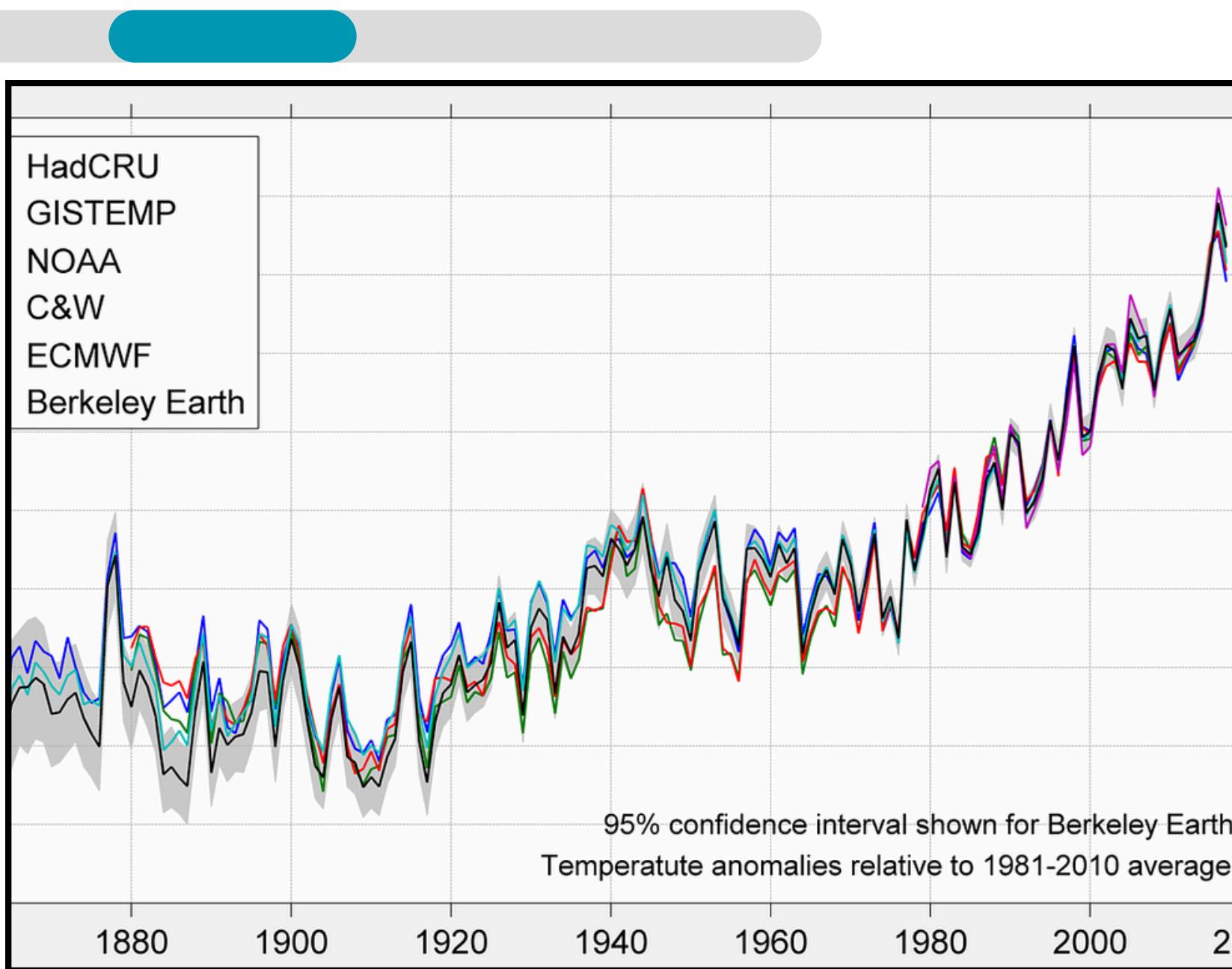
- 05 ¿EXISTE UNA RELACIÓN ENTRE EL FLUJO DE BICICLETAS Y OTROS FACTORES, COMO LA TEMPERATURA?
- 06 ¿CÓMO HA EVOLUCIONADO EL FLUJO DE BICICLETAS A LO LARGO DEL TIEMPO?
- 07 ¿QUÉ MODELO ES MAS PRECISO PARA GENERAR LAS PREDICCIONES?
- 08 ¿PUEDEN LOS DATOS HISTÓRICOS AYUDAR A PREDECIR FUTUROS PATRONES DEL USO DE ESTAS?

índice



- 01** Introducción
- 02** Motivación
- 03** Análisis exploratorio
- 04** Prophet
- 05** Arima
- 06** Conclusión

Time Series analysis



Algunos pasos cruciales para trabajar con series temporales:

- Importar el conjunto de datos.
- Limpieza de datos.
- Comprender las variables.
- Descubrir relaciones entre variables.
- Identificar tendencias y estacionalidad.
- Preprocesamiento.
- Entrenamiento de un modelo de aprendizaje automático.

Fuente: Time-Series Analysis with Python



RECOPILACIÓN DE DATOS

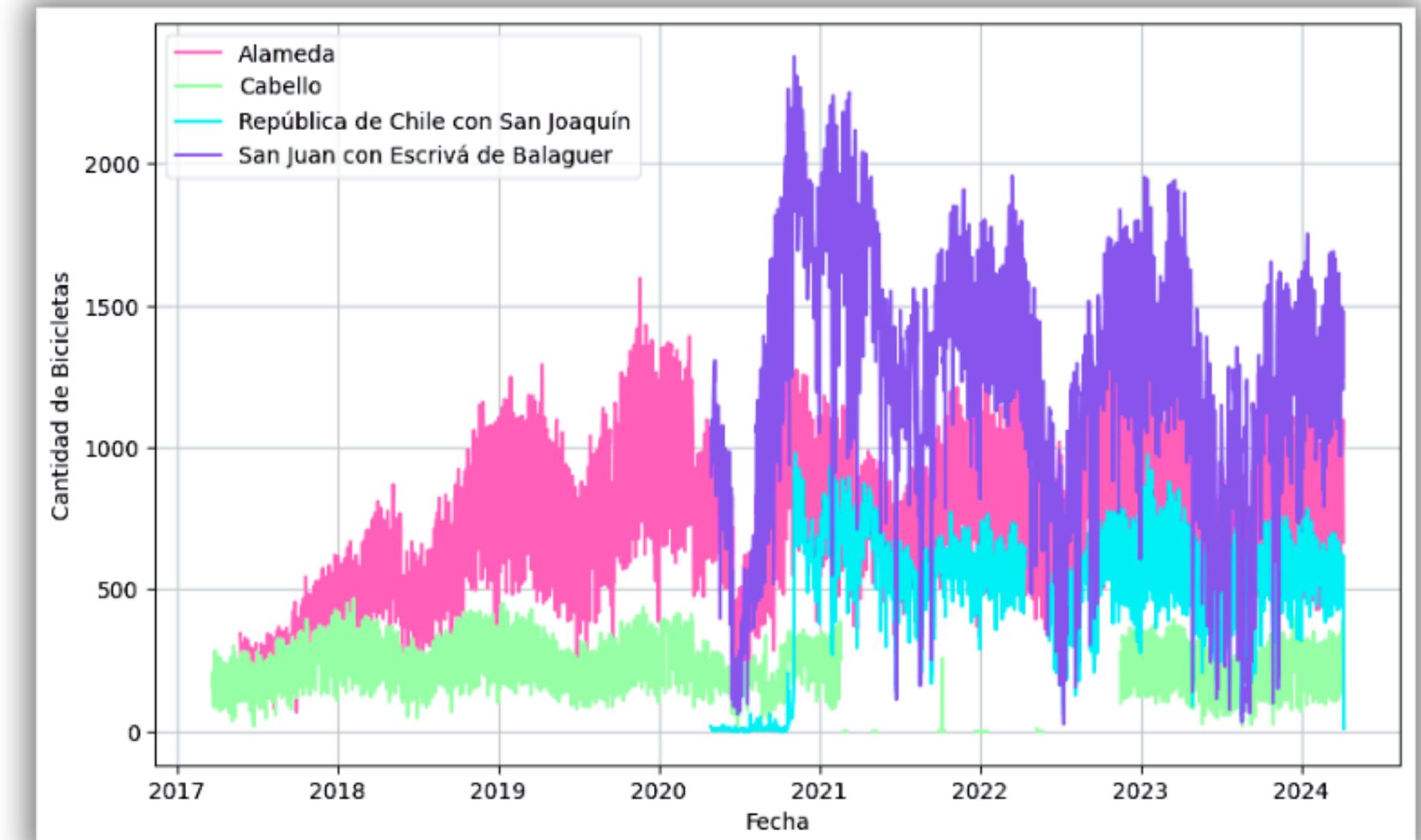
LIMPIEZA DE DATOS

IMPUTACIÓN DE DATOS

Metodología a seguir

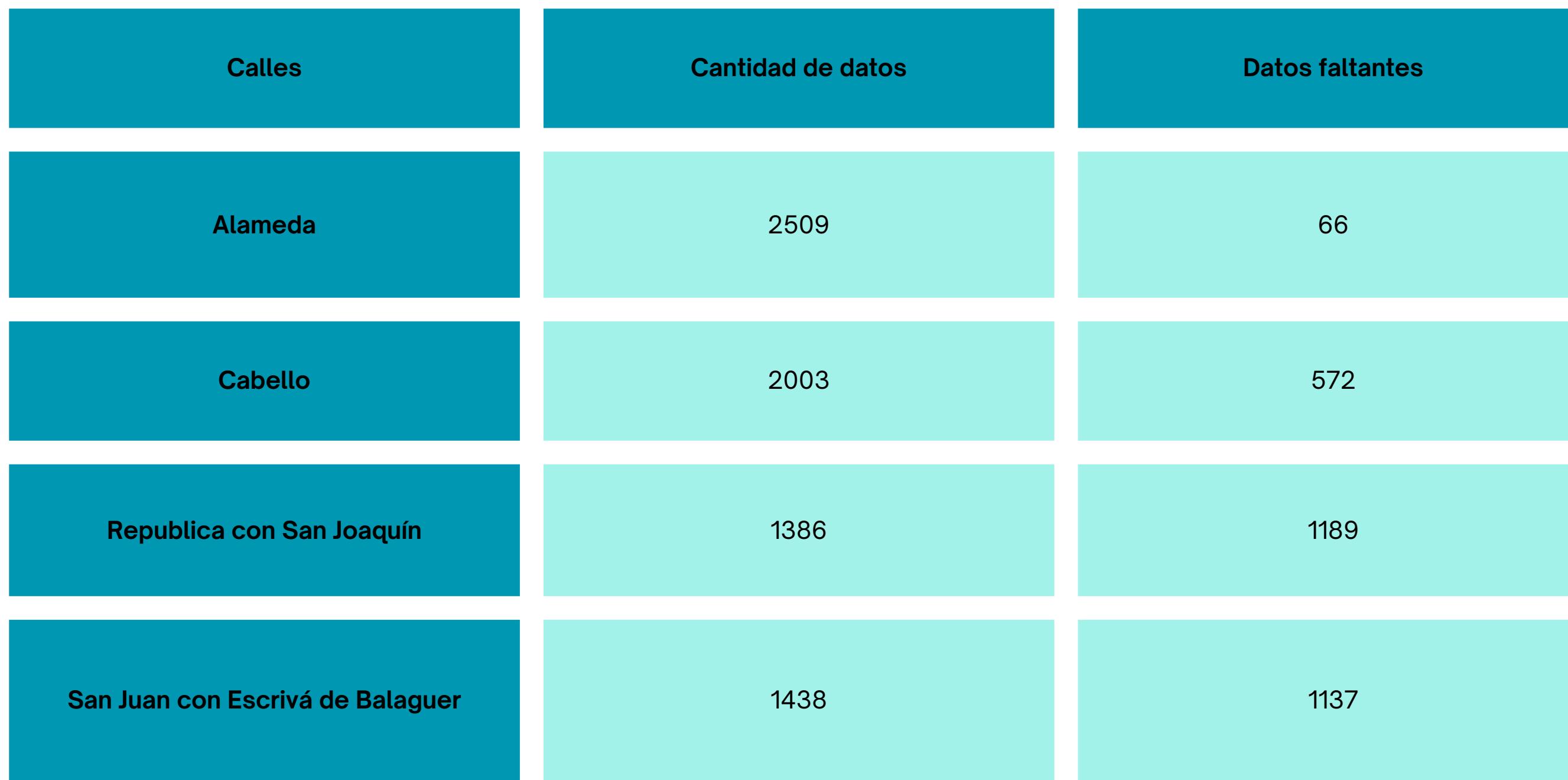
Recopilación de los datos (bike_rancagua.xlsx)

Time	Alameda	Cabello	República de Chile con San Joaquín	San Juan con Escrivá de Balaguer
2017-03-21 00:00:00	Nan	Nan	Nan	Nan
2017-03-22 00:00:00	Nan	159.0	Nan	Nan
2017-03-23 00:00:00	Nan	203.0	Nan	Nan
2017-03-24 00:00:00	Nan	207.0	Nan	Nan
2017-03-25 00:00:00	Nan	132.0	Nan	Nan
...
2024-04-03 00:00:00	1006.0	280.0	490.0	1249.0
2024-04-04 00:00:00	1088.0	323.0	584.0	1454.0
2024-04-05 00:00:00	1099.0	392.0	618.0	1481.0
2024-04-06 00:00:00	665.0	257.0	439.0	1207.0
2024-04-07 00:00:00	Nan	Nan	9.0	Nan



Limpieza de datos

Análisis exploratorio



Imputación de datos

Modelo de Regresión Lineal Múltiple

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i$$

Magnitud común a todos los sujetos.

Peso de cada una de las **k** variables independientes dentro de la ecuación de regresión.

Error para cada sujeto.

Regresión lineal

Técnica estadística utilizada para predecir el valor de una variable dependiente basada en dos o más variables independientes

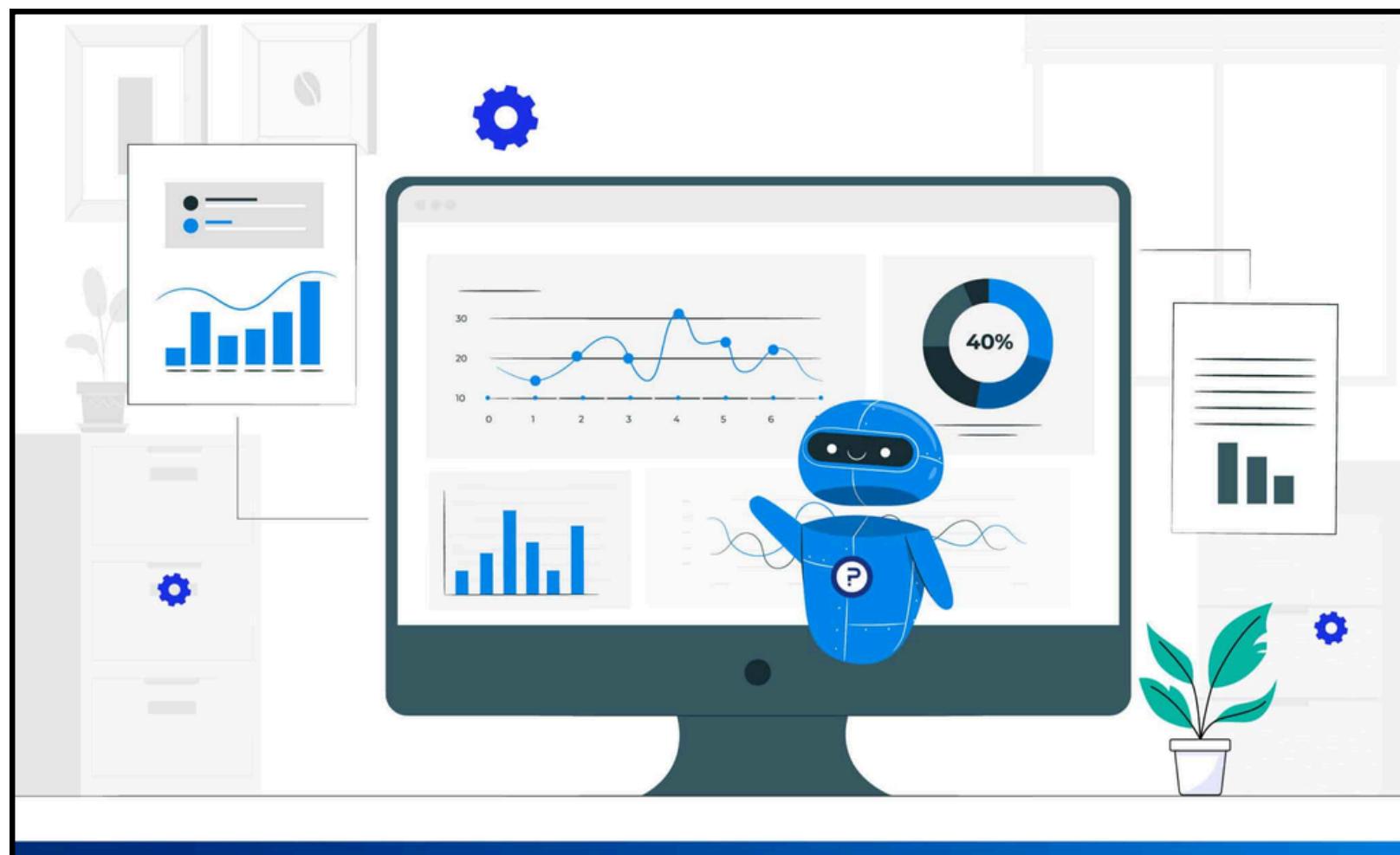
$$\hat{Cabello}_i = \beta_0 + \beta_1 \cdot Alameda_i + \beta_2 \cdot República_i + \beta_3 \cdot San\ Juan_i + \varepsilon_i$$

$$\hat{Alameda}_i = \beta_0 + \beta_1 \cdot Cabello_i + \beta_2 \cdot República_i + \beta_3 \cdot San\ Juan_i + \varepsilon_i$$

$$\hat{República}_i = \beta_0 + \beta_1 \cdot Alameda_i + \beta_2 \cdot Cabello_i + \beta_3 \cdot San\ Juan_i + \varepsilon_i$$

$$\hat{San\ Juan}_i = \beta_0 + \beta_1 \cdot Alameda_i + \beta_2 \cdot Cabello_i + \beta_3 \cdot República_i + \varepsilon_i$$

Regresión lineal



1. Cargar base de datos.
2. Cambio de variable en la columna “Time”.
3. Inicializamos el modelo con la librería `“sklearn.linear_model import LinearRegression”`.
4. Dividir los datos en características (x_i) y la variable a predecir (y_i), esto para cada calle y con todas las calles.
5. Reemplazar valores nulos en X e y con el promedio de cada calle.
6. Entrenar el modelo, `modelo.fit(x_i, y_i)`.
7. Predecir los valores faltantes en la calle i basados en las otras zonas.
8. Gráficos con `“matplotlib.pyplot”`.

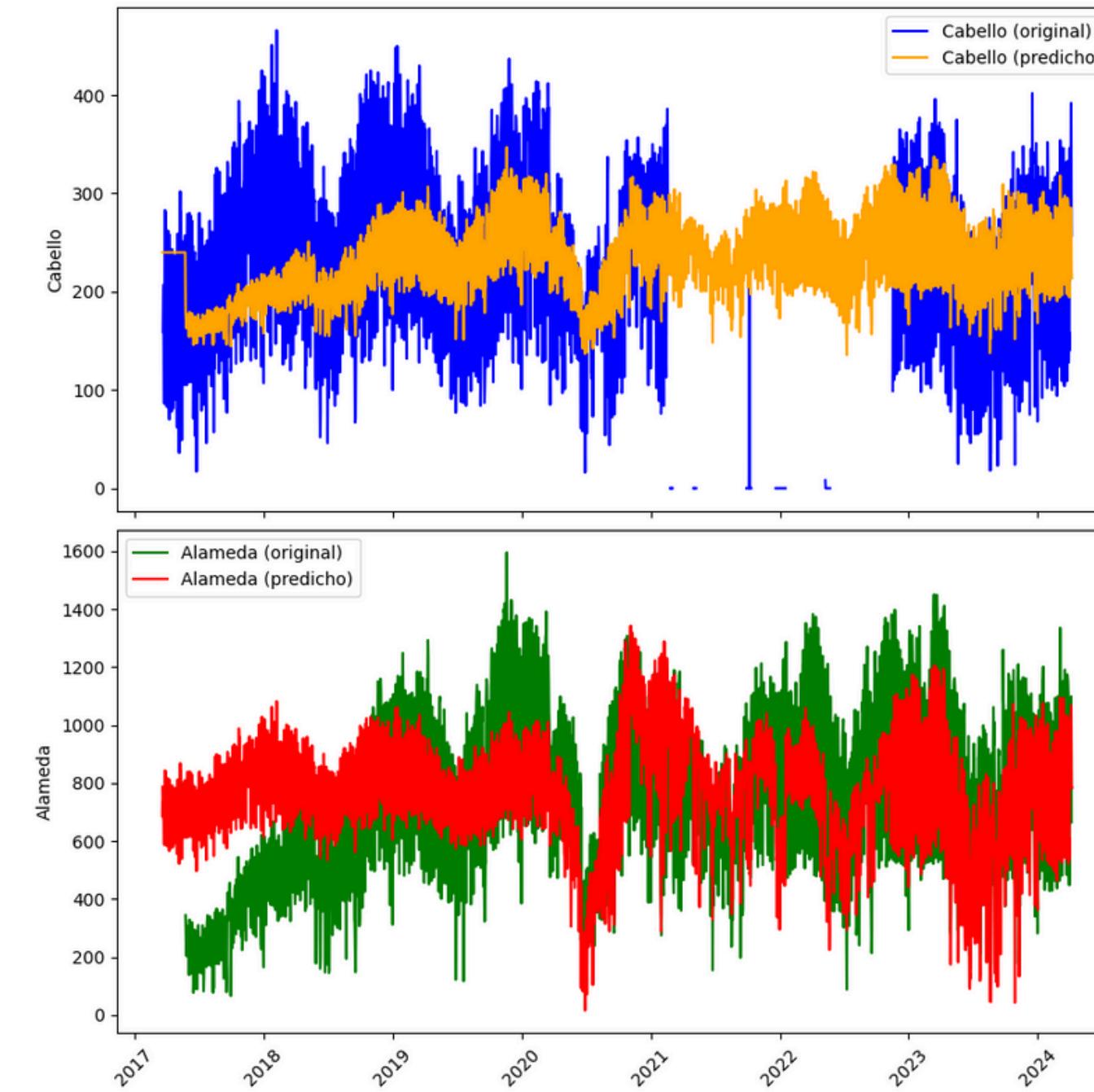
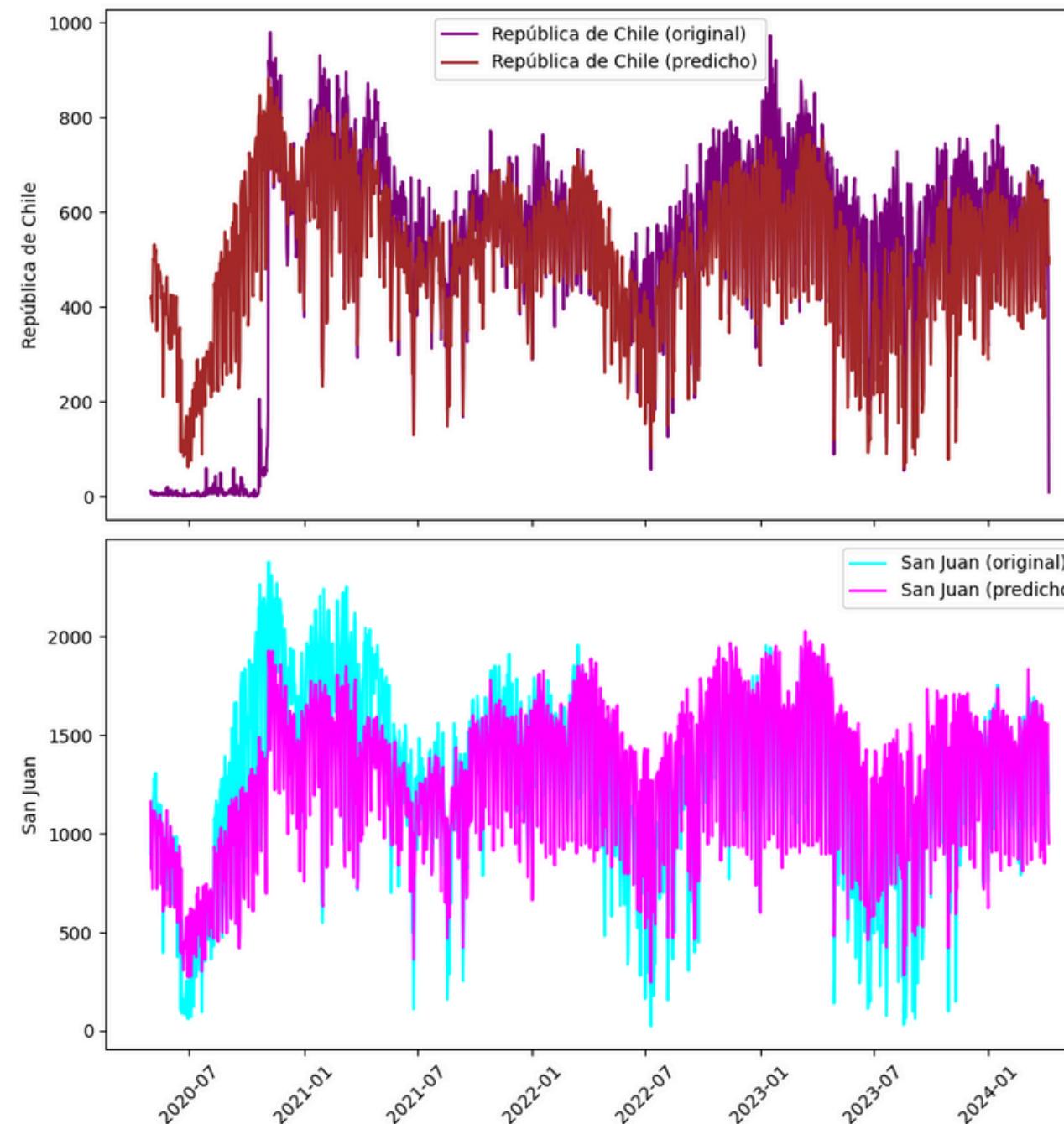
Regresión lineal

Time	Cabello_predicho	Alameda_predicho
2017-03-21	254.329876	862.012475
2017-03-22	262.965924	799.513999
2017-03-23	253.000571	831.313165
2017-03-24	252.289652	800.960075
2017-03-25	219.241334	594.288990
...
2024-04-03	280.365178	885.500379
2024-04-04	280.904299	989.551327
2024-04-05	282.635561	1019.467525
2024-04-06	206.179636	668.231175
2024-04-07	155.437809	581.031255

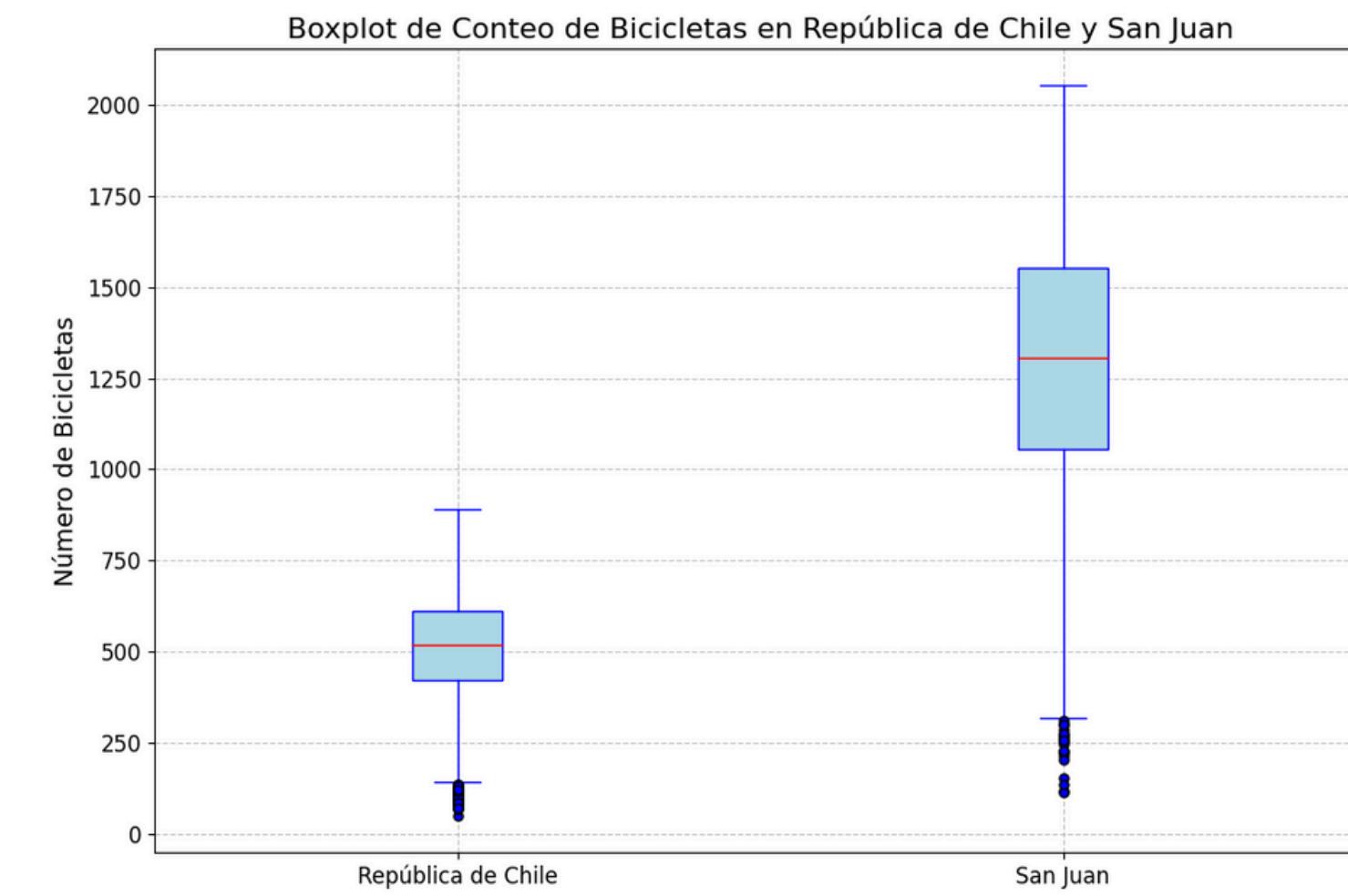
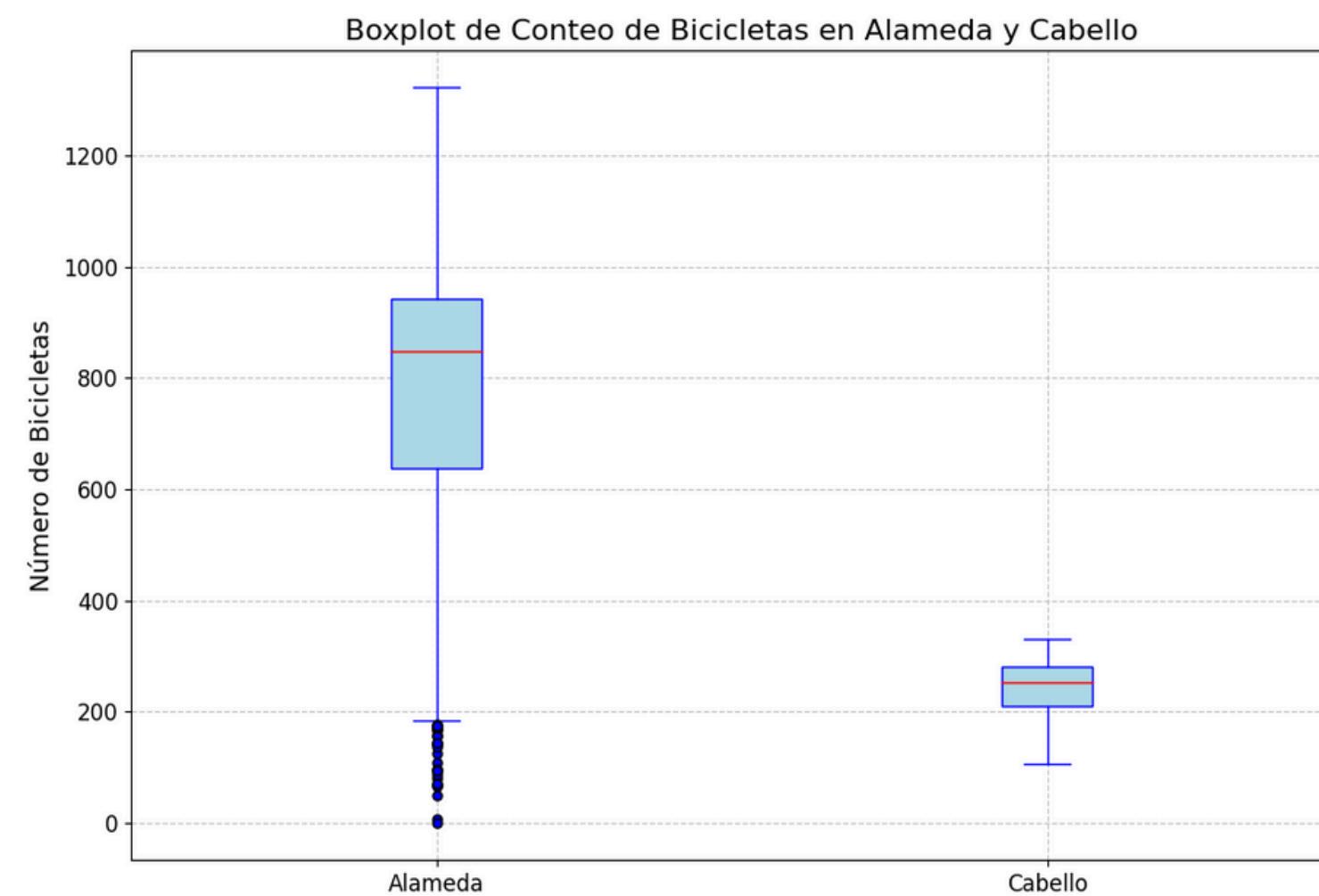
Time	República de Chile_predicho	San Juan_predicho
2020-01-01	466.409807	685.700597
2020-01-02	538.490299	1531.181407
2020-01-03	549.147862	1594.290090
2020-01-04	499.179669	1364.446444
2020-01-05	481.123468	1477.989589
...
2024-04-03	515.033265	1336.652617
2024-04-04	600.638080	1466.704609
2024-04-05	626.652337	1488.473994
2024-04-06	492.573109	1062.033165
2024-04-07	496.576311	1186.613822

Regresión lineal

Visualización de los datos

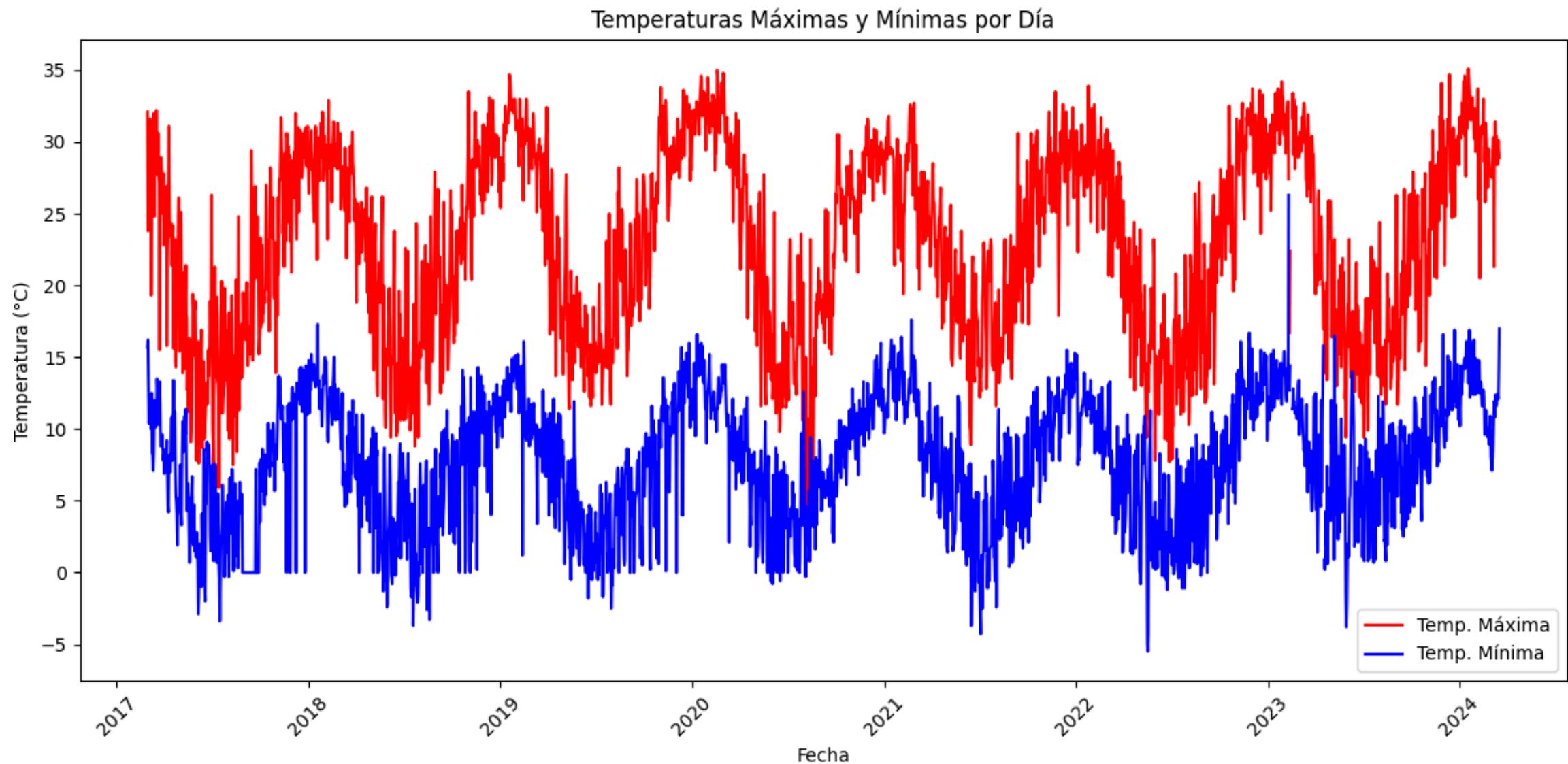


Visualización de los datos



Variable exógena

Fecha	Temp. Máxima	Temp. Mínima
2017-03-21	30.6	11.8
2017-03-22	27.8	10.3
2017-03-23	27.2	10.7
2017-03-24	15.5	13.3
2017-03-25	26.4	11.8
...
2024-03-13	28.4	12.3
2024-03-14	29.6	12.4
2024-03-15	30.1	12.1
2024-03-16	29.4	14.2
2024-03-17	28.9	17.0



índice



- 01** Introducción
- 02** Motivación
- 03** Análisis exploratorio
- 04** Prophet
- 05** Arima
- 06** Conclusión

Prophet

PROPHET

Prophet es un modelo para pronosticar datos de series temporales basado en un modelo aditivo en el que las tendencias no lineales se ajustan a la estacionalidad anual, semanal y diaria, además de los efectos de los días festivos.

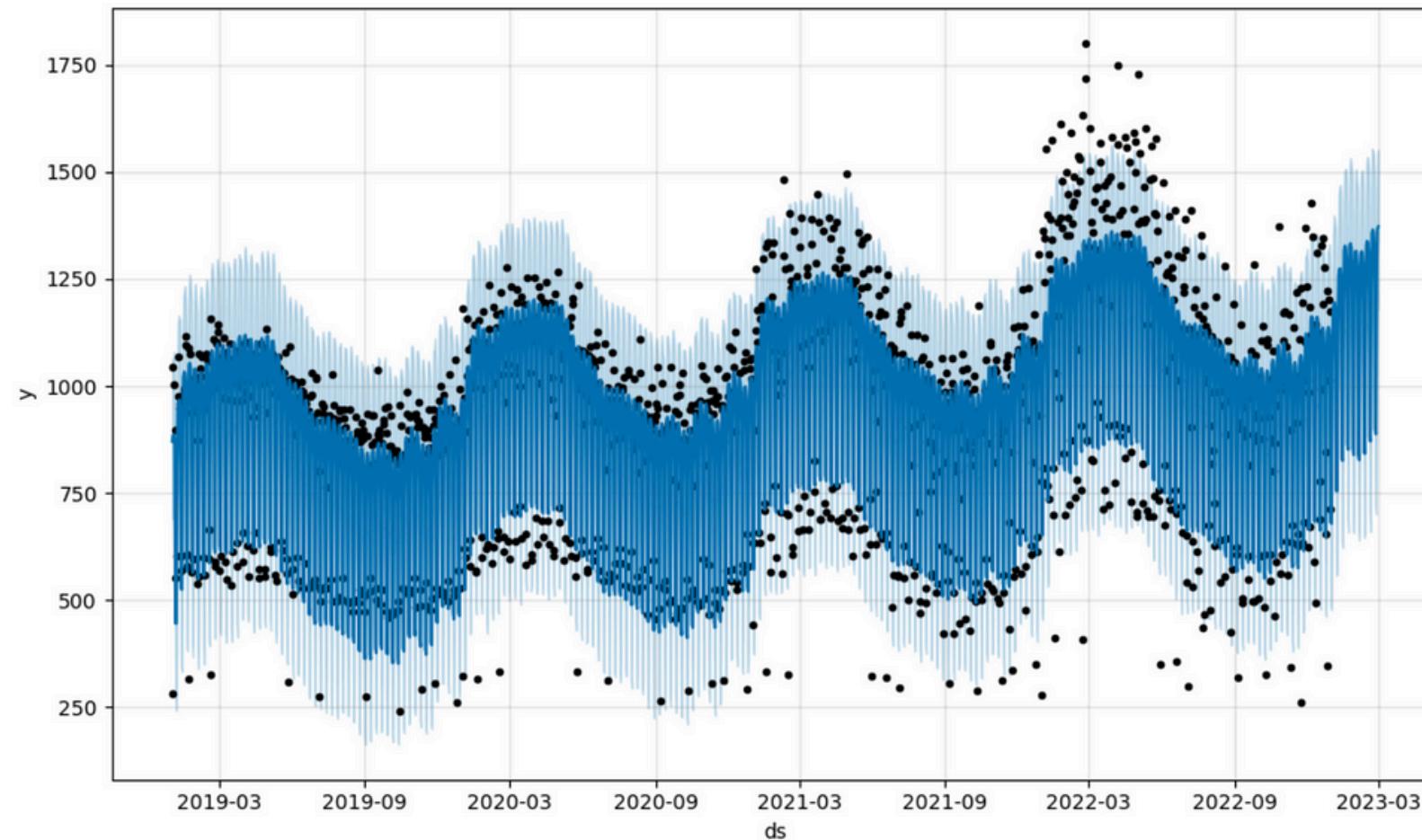
Funciona mejor con series temporales que tienen fuertes efectos estacionales y varias temporadas de datos históricos. Prophet es robusto ante datos faltantes y cambios en la tendencia, y normalmente maneja bien los valores atípicos.

Prophet

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

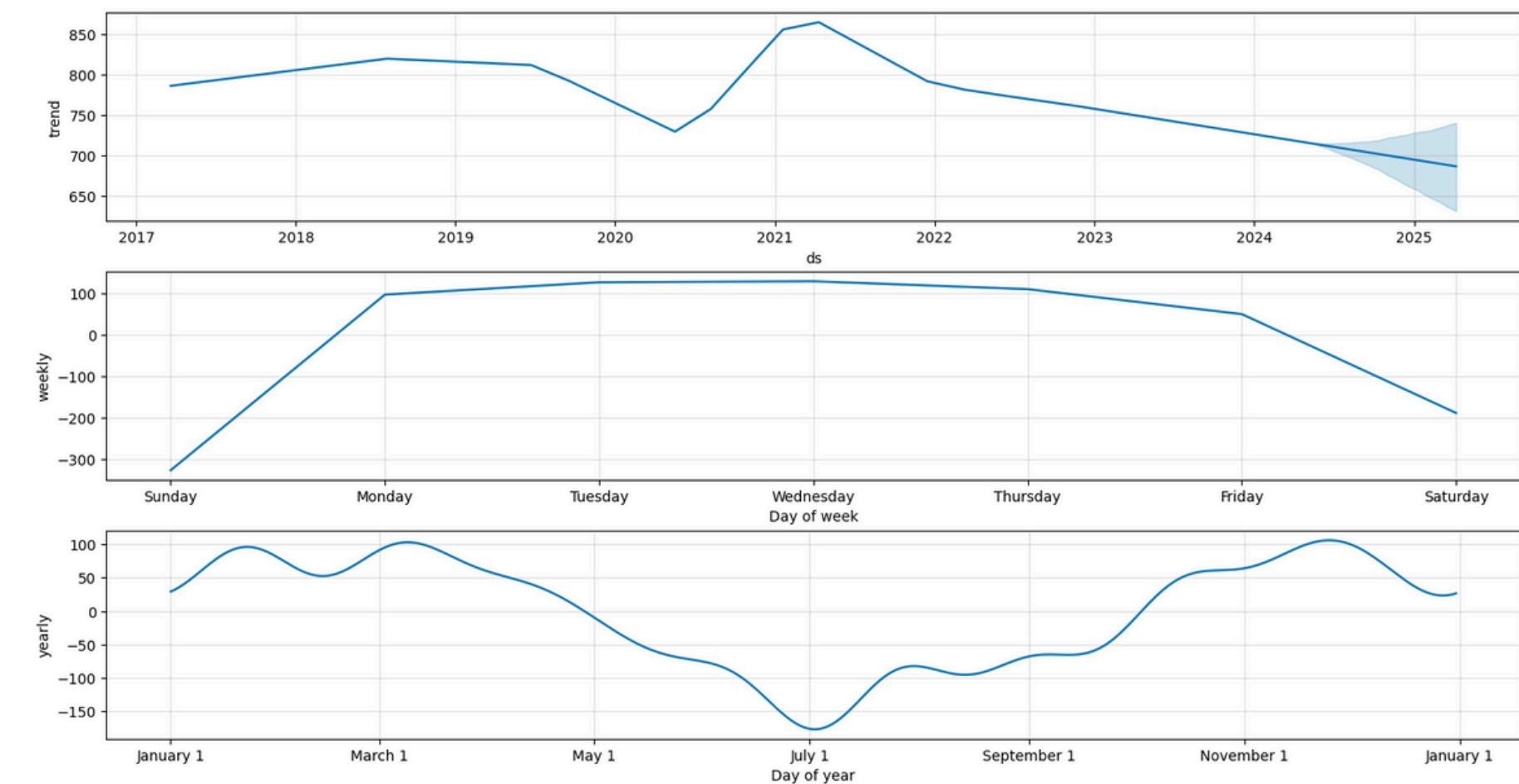
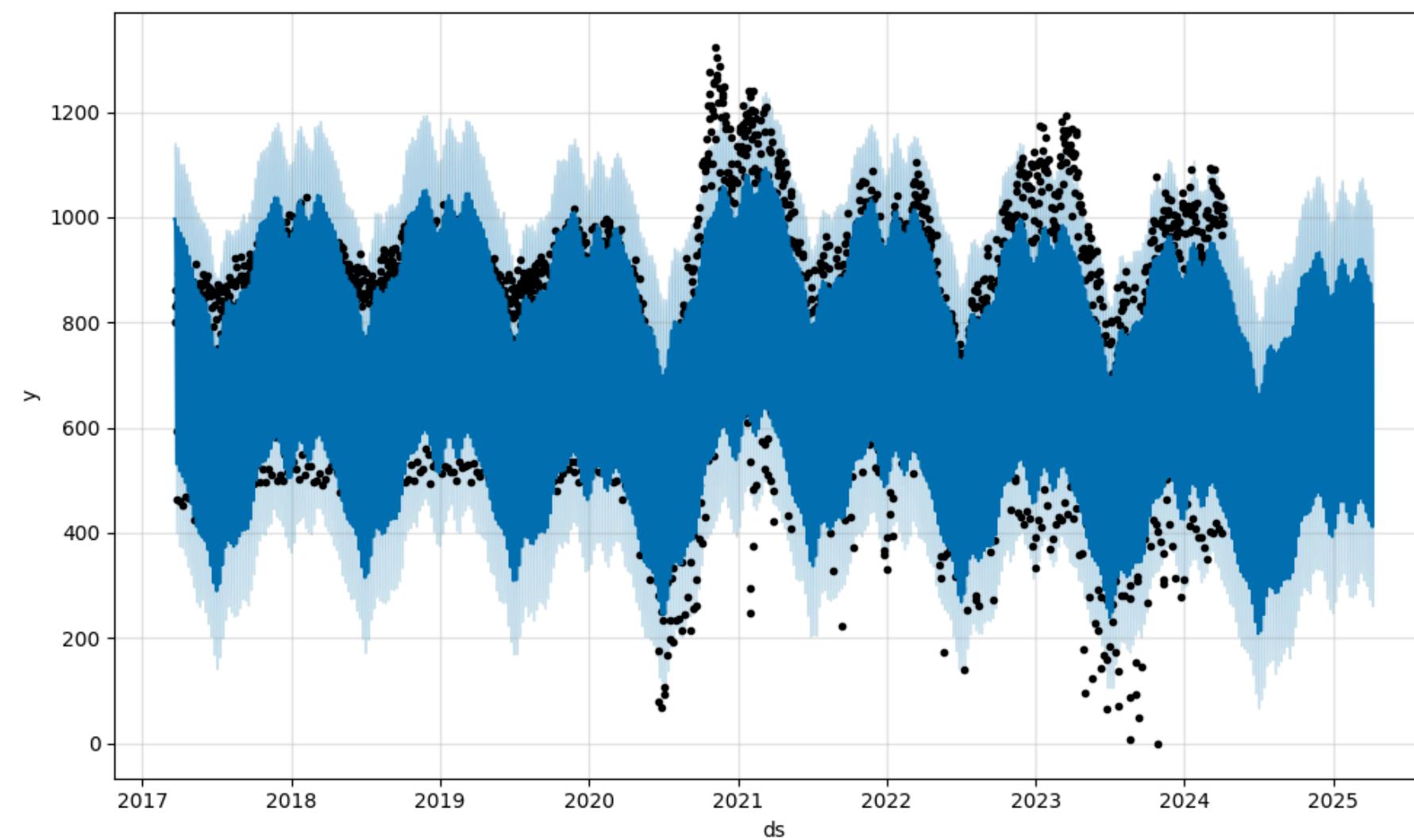
- $y(t)$ es la serie temporal observada en el tiempo t.
- $g(t)$ representa la **tendencia**, que modela los cambios a largo plazo en el nivel de la serie.
- $s(t)$ es el componente **estacional**, que captura las fluctuaciones periódicas.
- $h(t)$ representa los efectos de días festivos o eventos especiales.
- $\epsilon_{\{t\}}$ es el término de **error**, que representa las variaciones no explicadas por los otros componentes.

Prophet

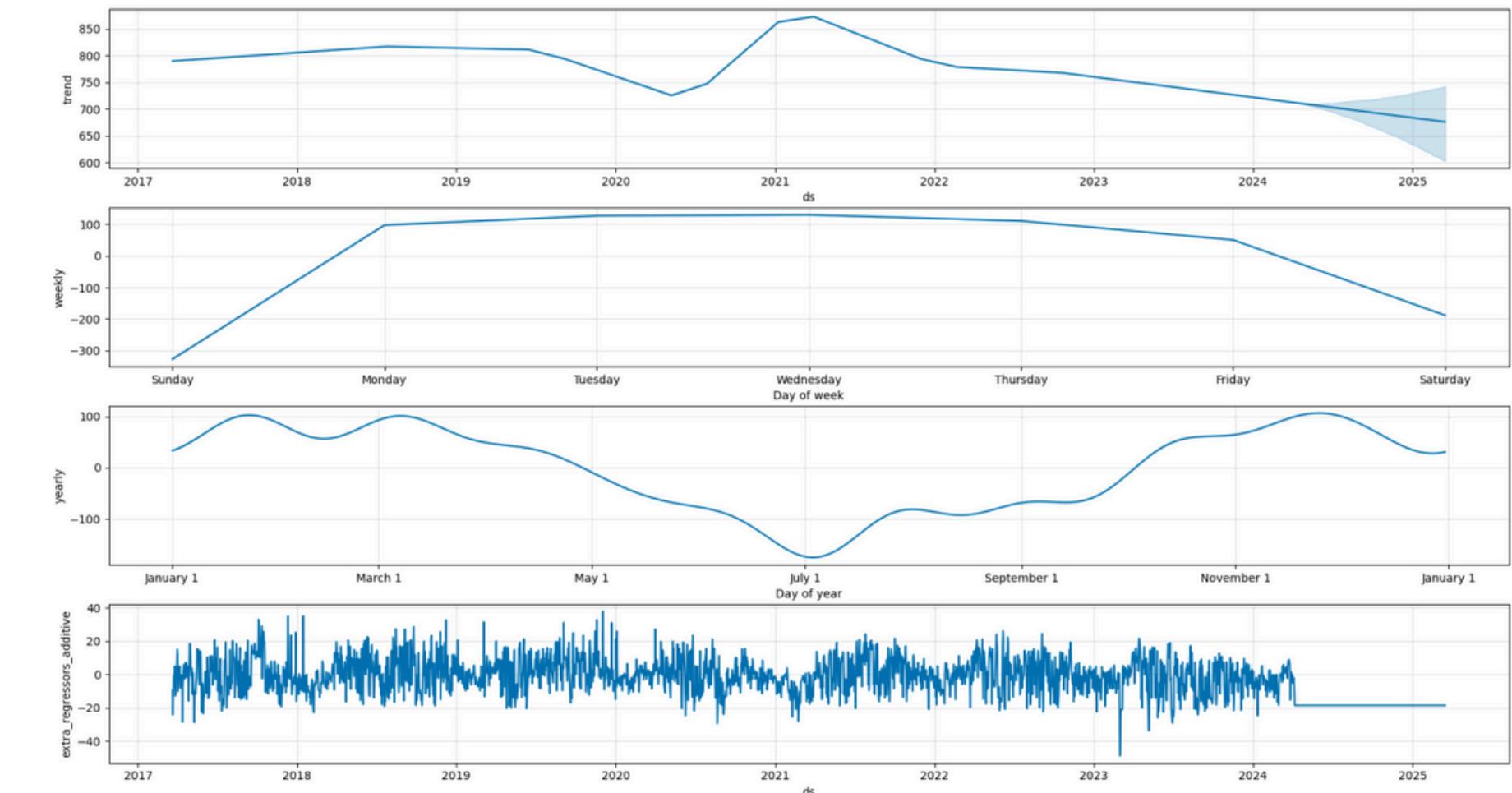
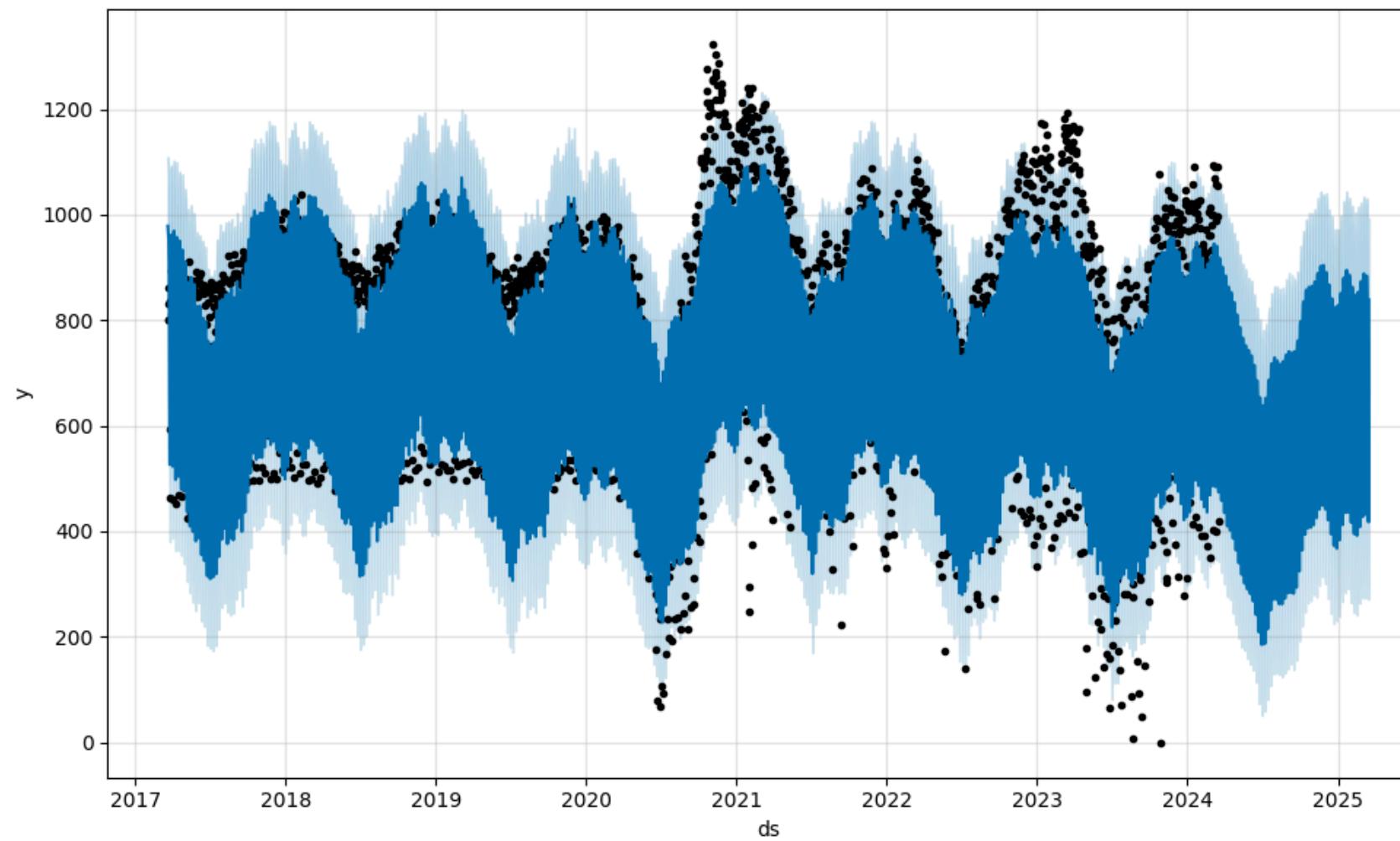


- *Cargar los datos.*
- *Renombrar las columnas del DataFrame, 'ds' para las fechas e 'y' para los valores.*
- *Inicializar el modelo Prophet.*
- *Entrenar el modelo `m.fit(df)`.*
- *Realizar predicciones “`forecast = m.predict(future)`”.*
- *Calcular el MAPE (Mean Absolute Percentage Error) y Calcular el MAD (Mean Absolute Deviation).*
- *Visualizar las últimas predicciones.*
- *Visualizar los resultados.*
- *Visualizar los componentes del pronóstico.*

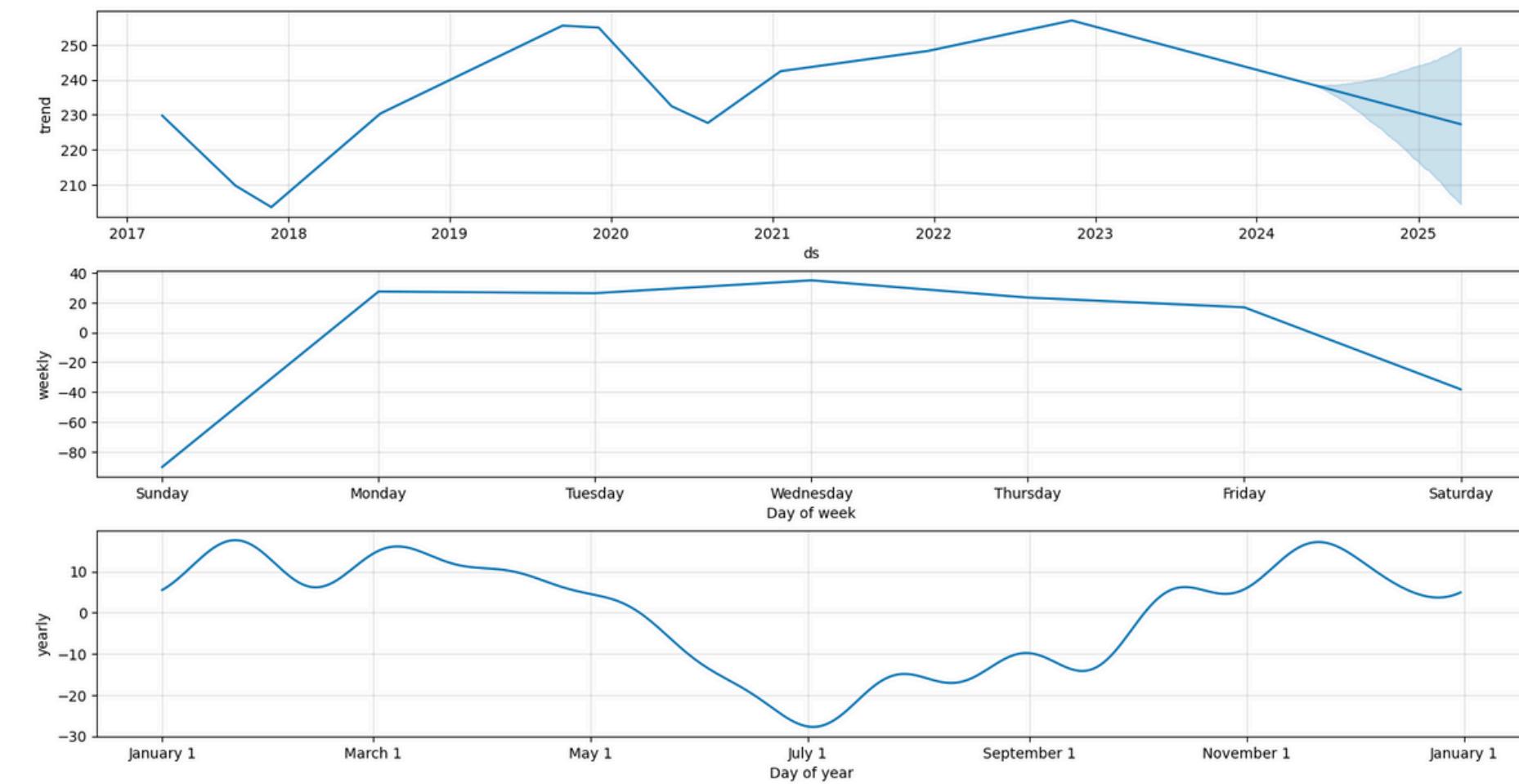
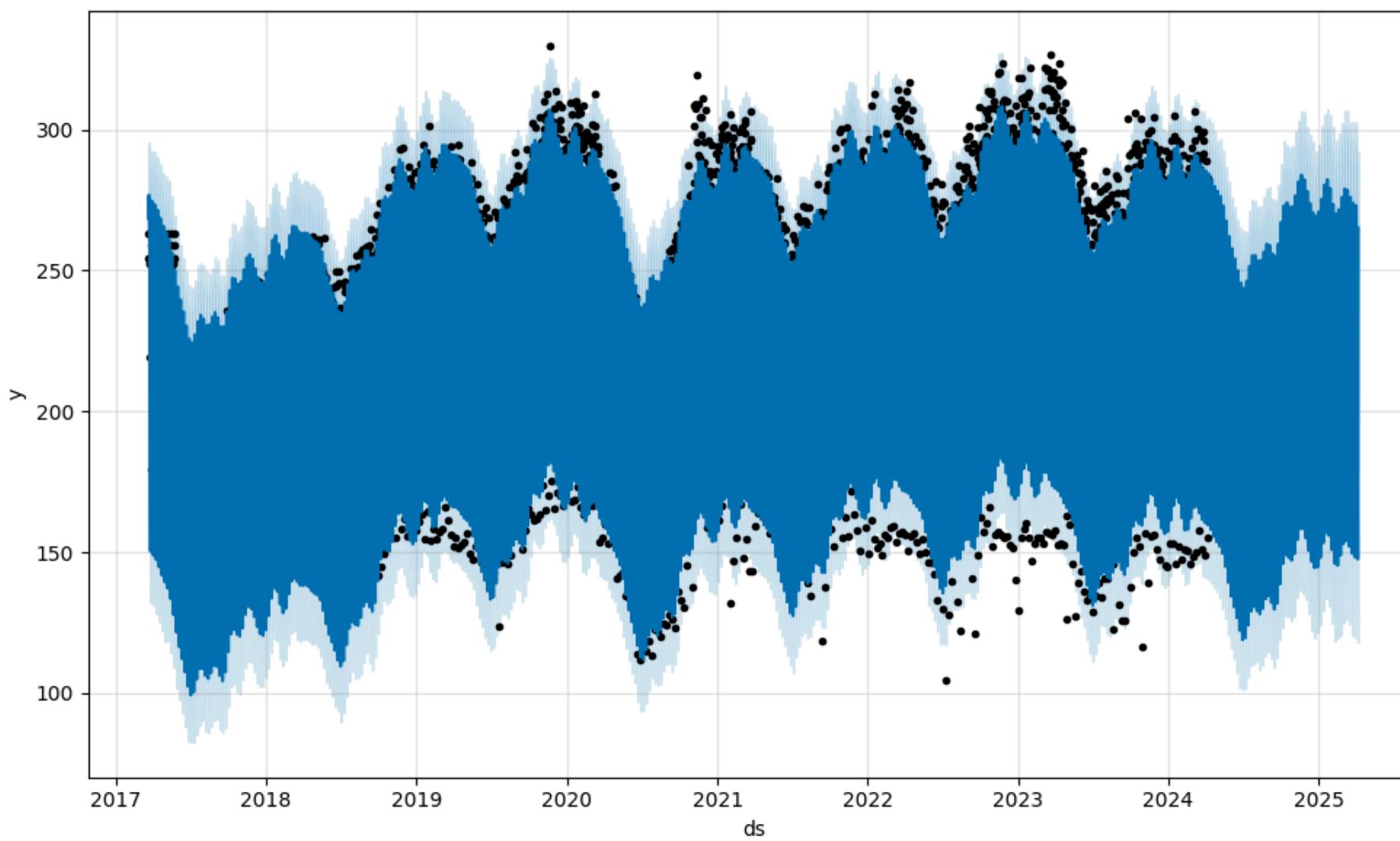
Modelo Prophet para Alameda



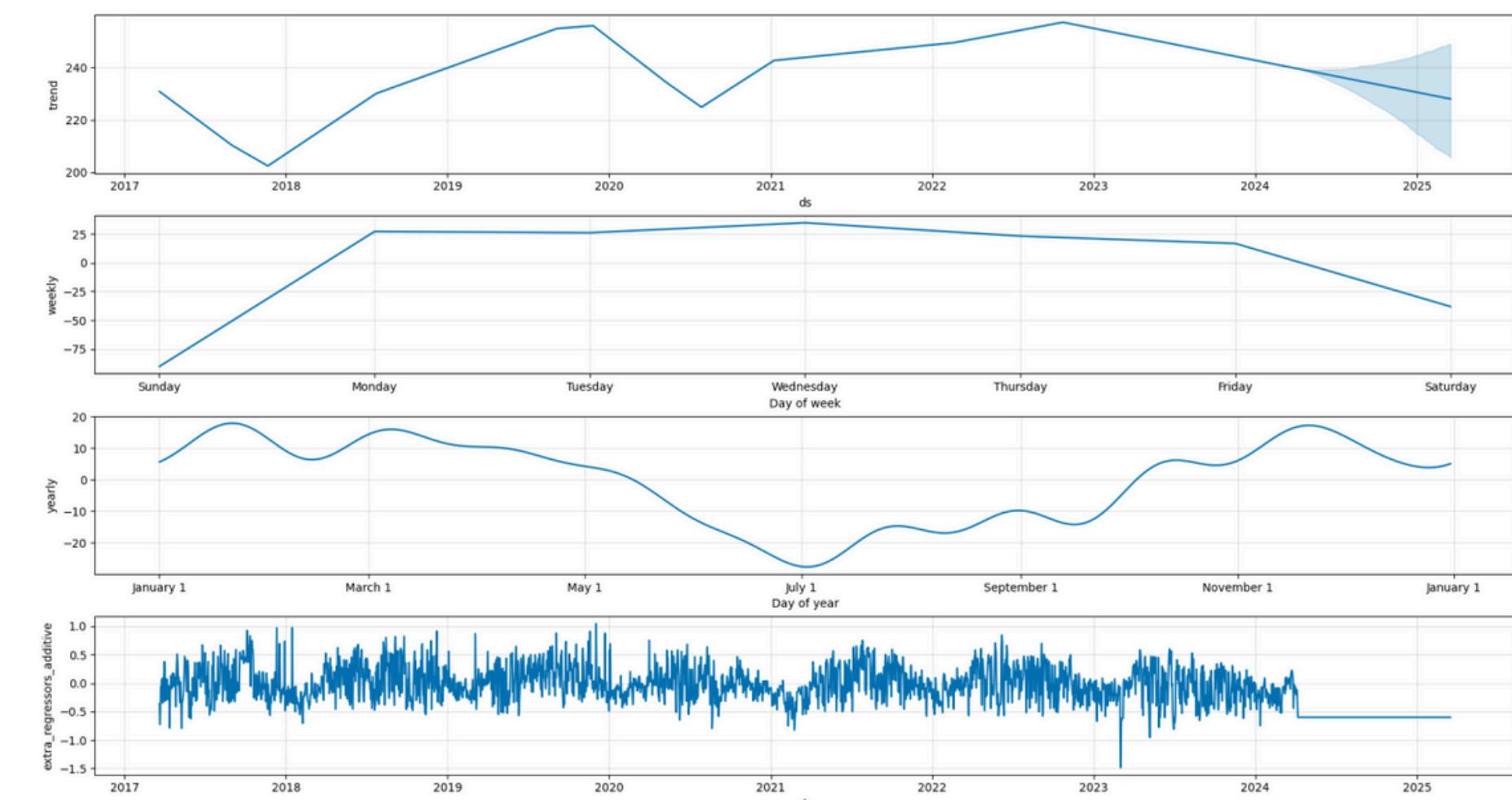
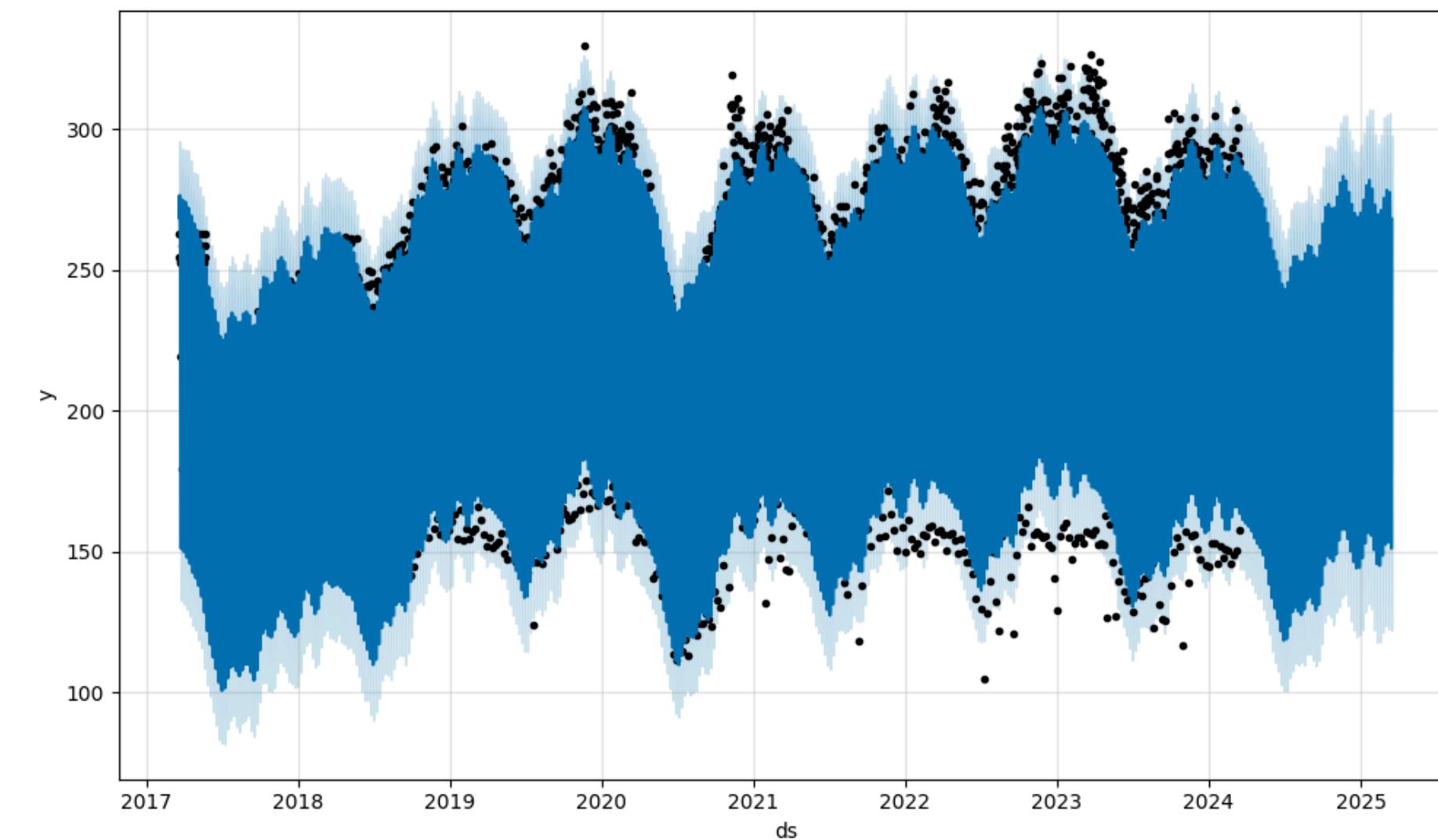
Modelo Prophet para Alameda con variables exógenas



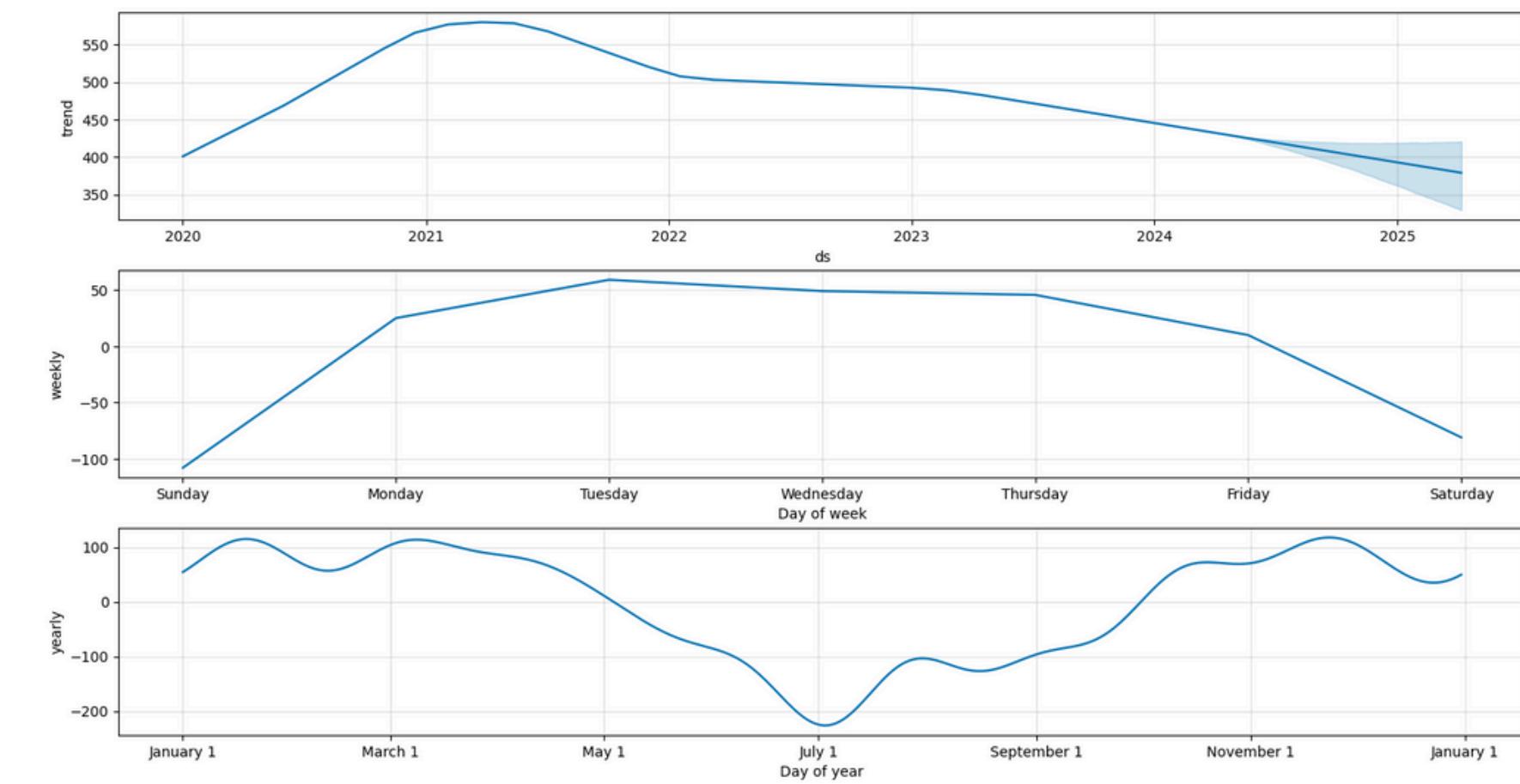
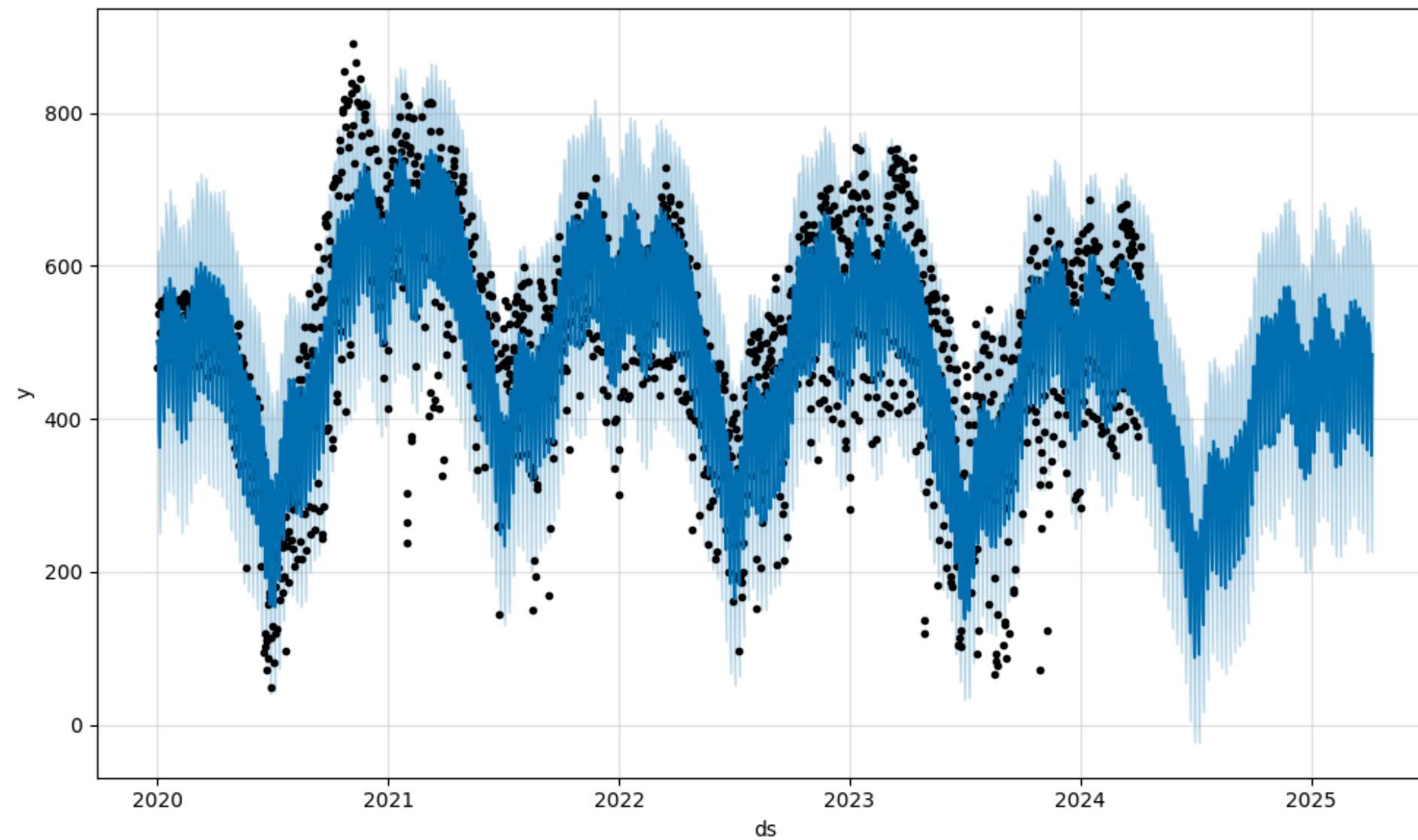
Modelo Prophet para Cabello



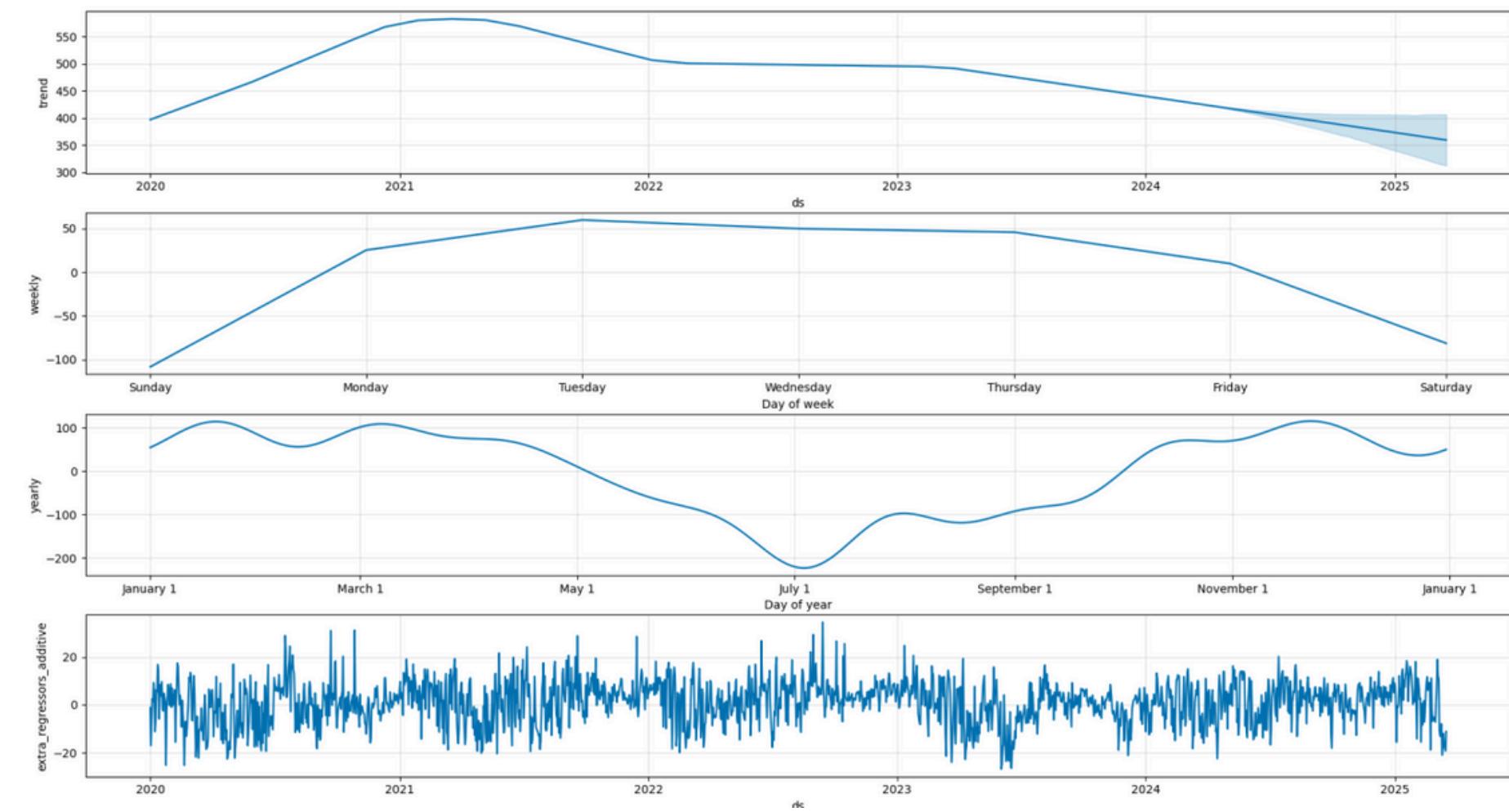
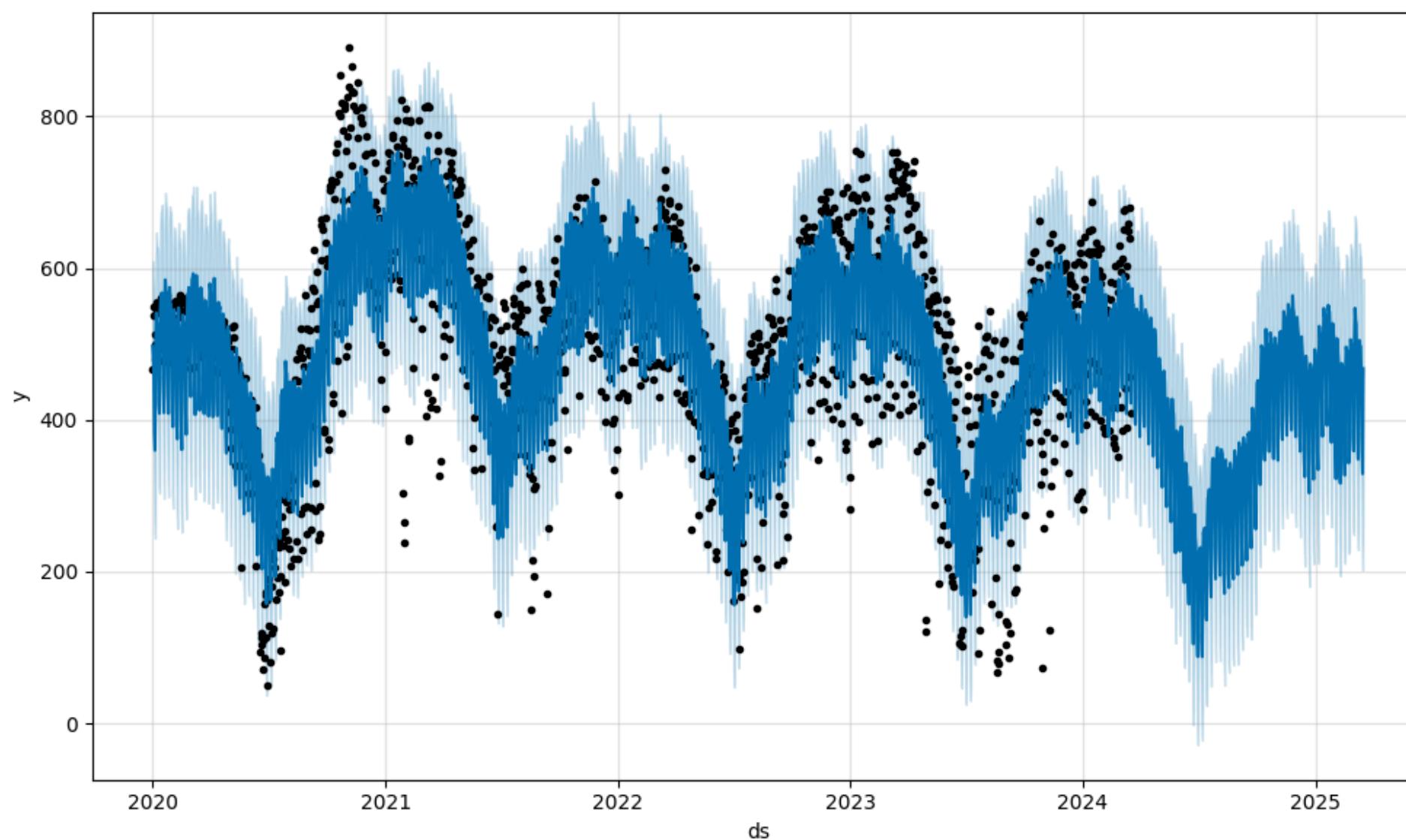
Modelo Prophet para Cabello con variables exógenas



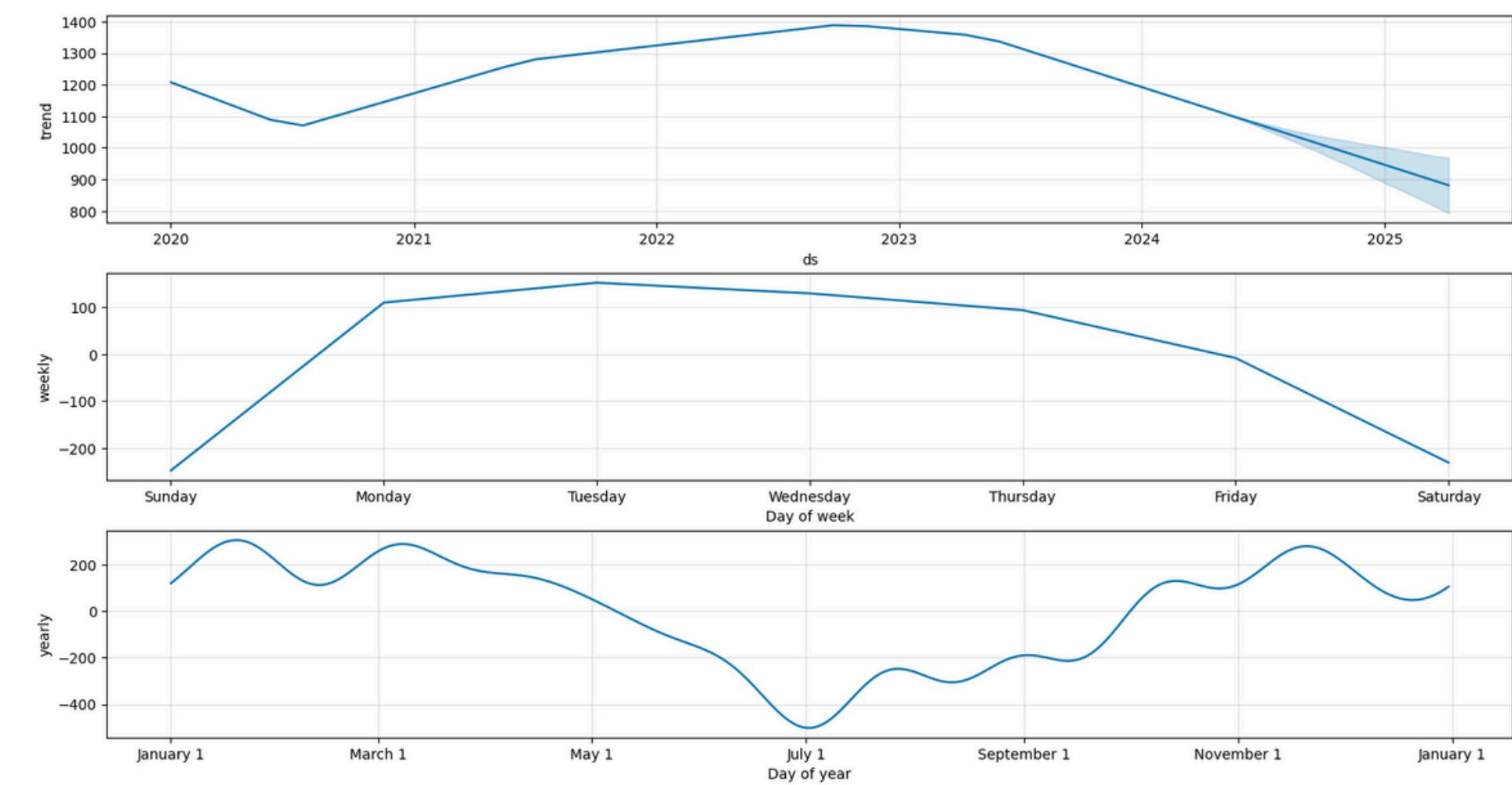
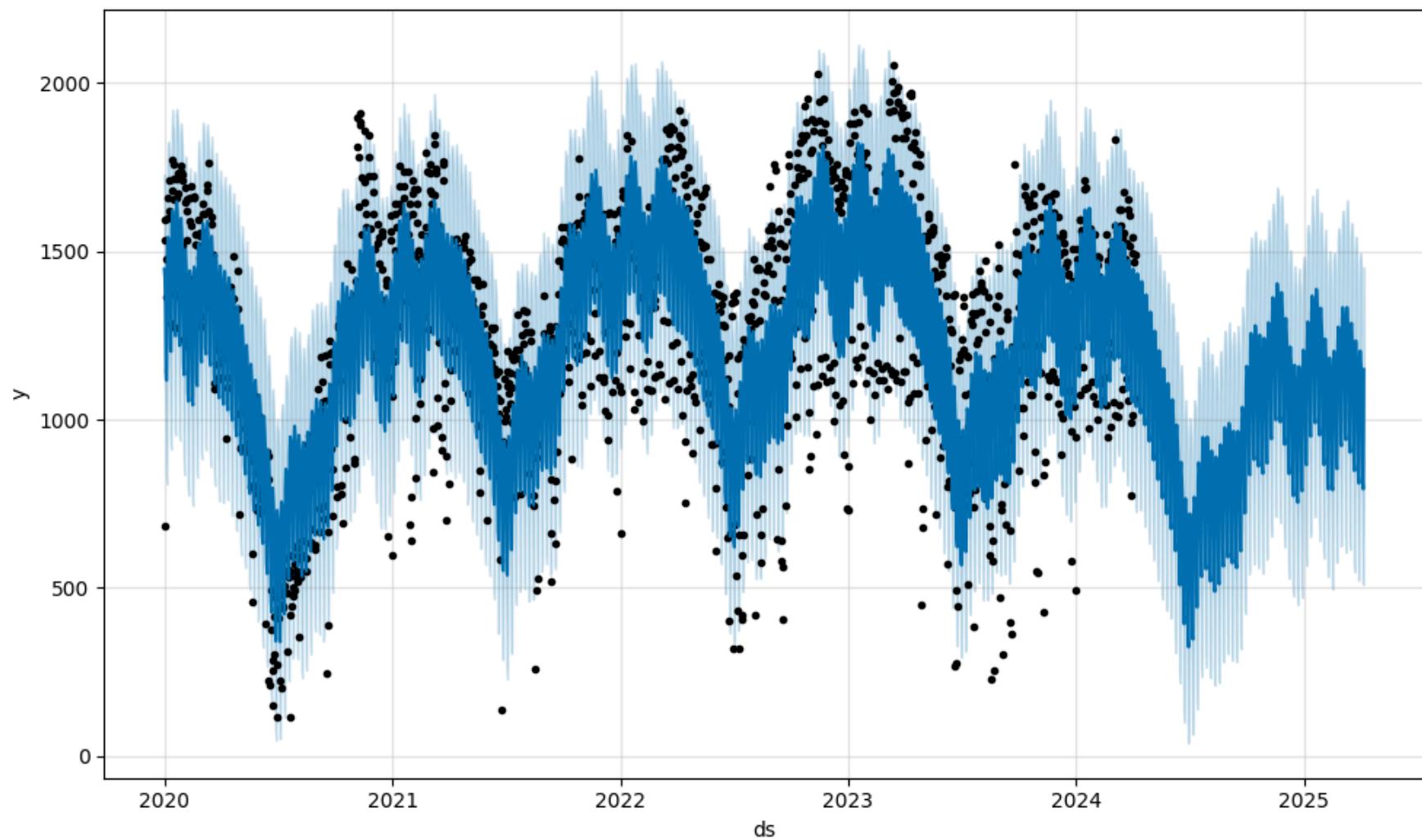
Modelo Prophet para República



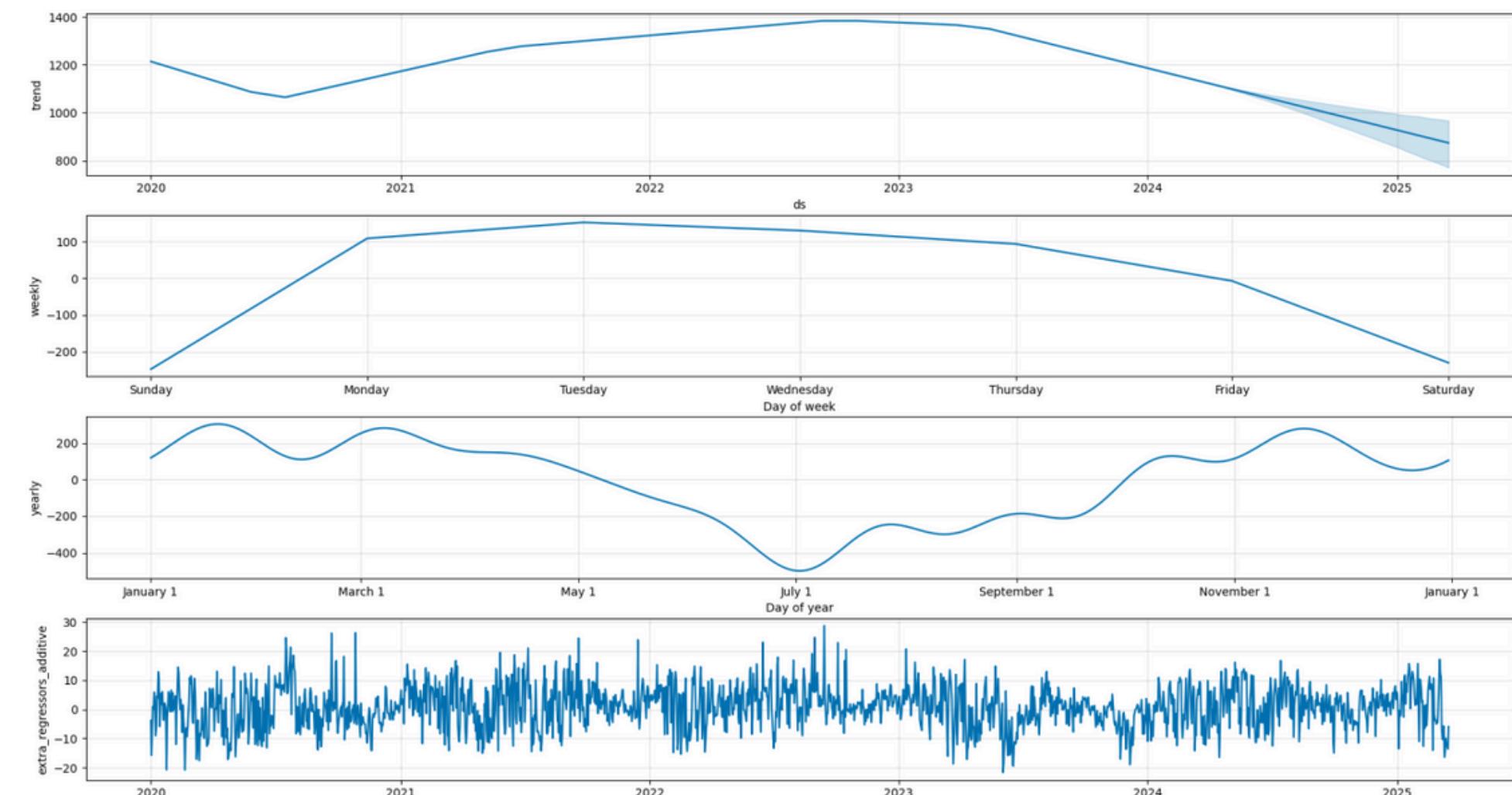
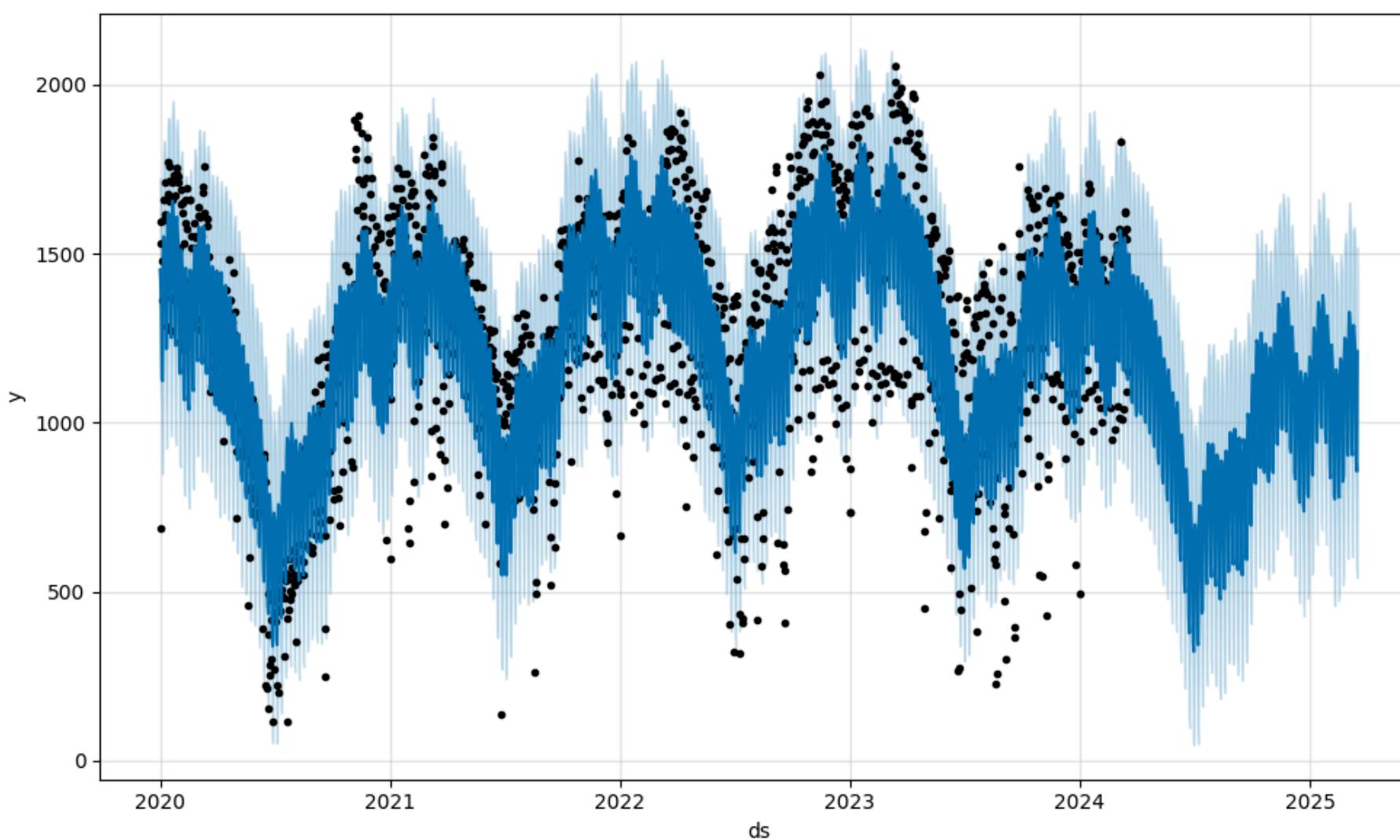
Modelo Prophet para República con variables exógena



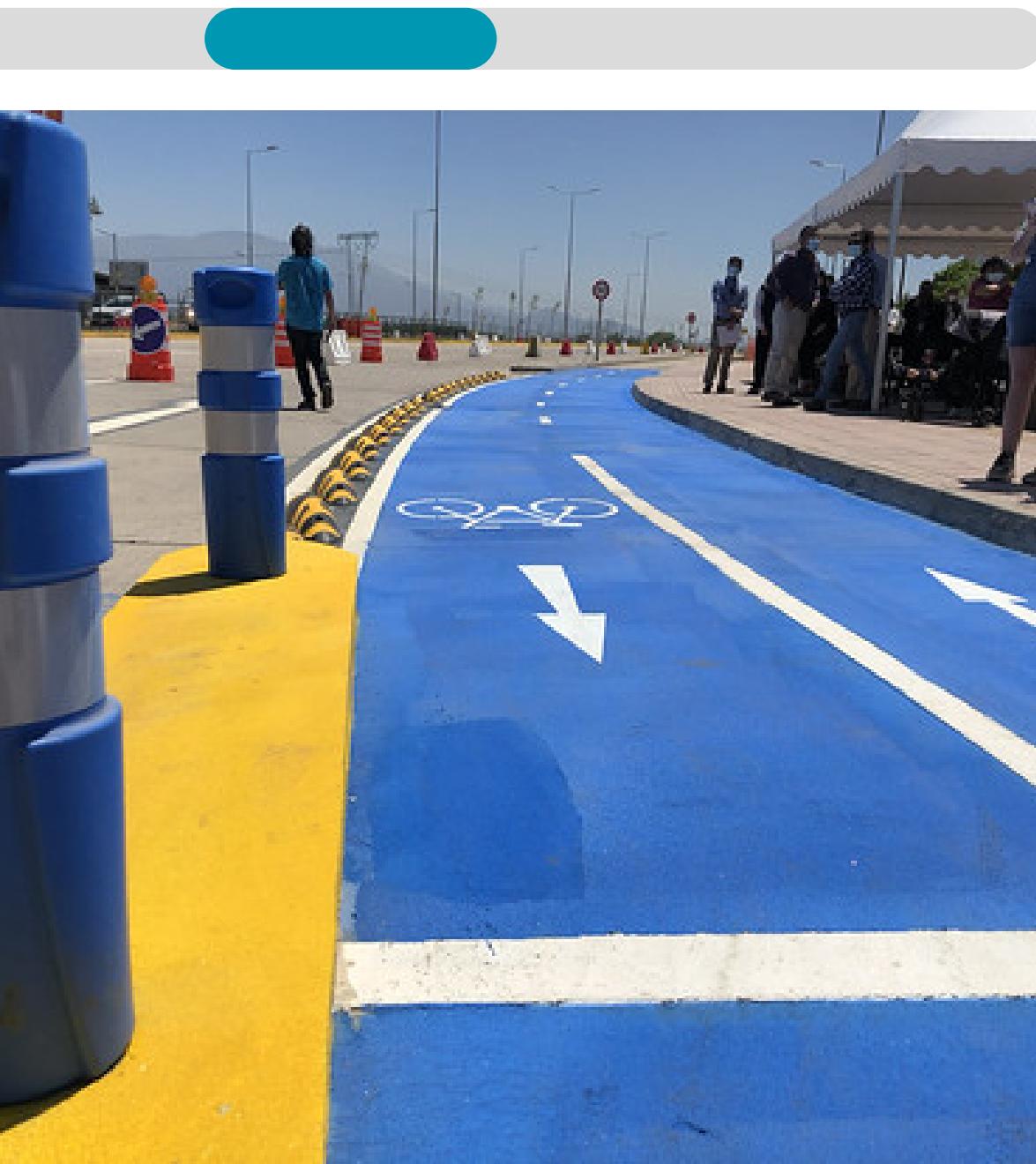
Modelo Prophet para San Juan



Modelo Prophet para San Juan con variables exógena



índice



- 01** Introducción
- 02** Motivación
- 03** Análisis exploratorio
- 04** Prophet
- 05** Arima
- 06** Conclusión



Promedio móvil integrado autorregresivo

Es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro.

Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado.



Promedio móvil integrado autorregresivo

ARIMA(p,d,q)

Cada una de las tres partes del acrónimo se le denomina una componente y modela un comportamiento distinto de la serie (p,d,q) .

- p: Autorregresión de la serie
- d: Integración o diferenciación de la serie
- q: Media móvil de la serie

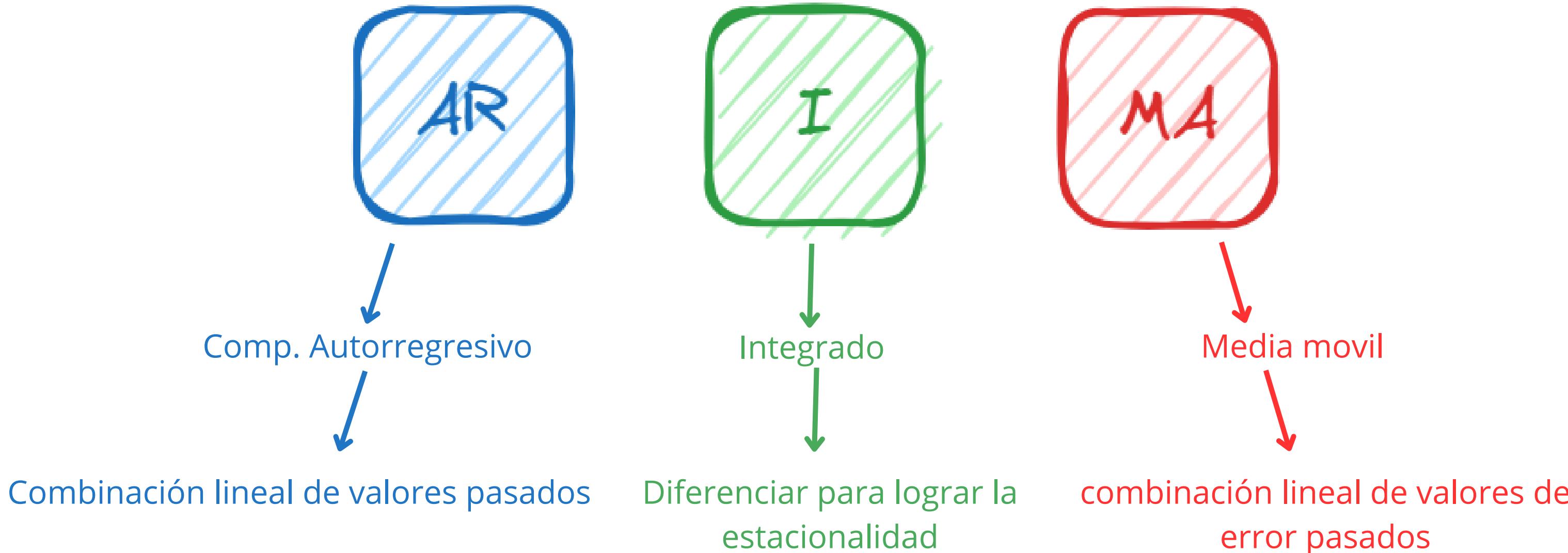
Ecuación general ARMA

Promedio móvil integrado autorregresivo

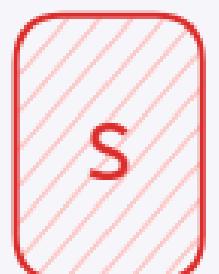
$$y_t = C + \phi_1 \cdot y_{t-1} + \cdots + \phi_p \cdot y_{t-p} + \theta_1 \cdot \varepsilon_{t-1} + \cdots + \theta_q \cdot \varepsilon_{t-q} + \varepsilon_t$$

↔ Comp. Autorregresiva ↔ Comp. de Media Móvil

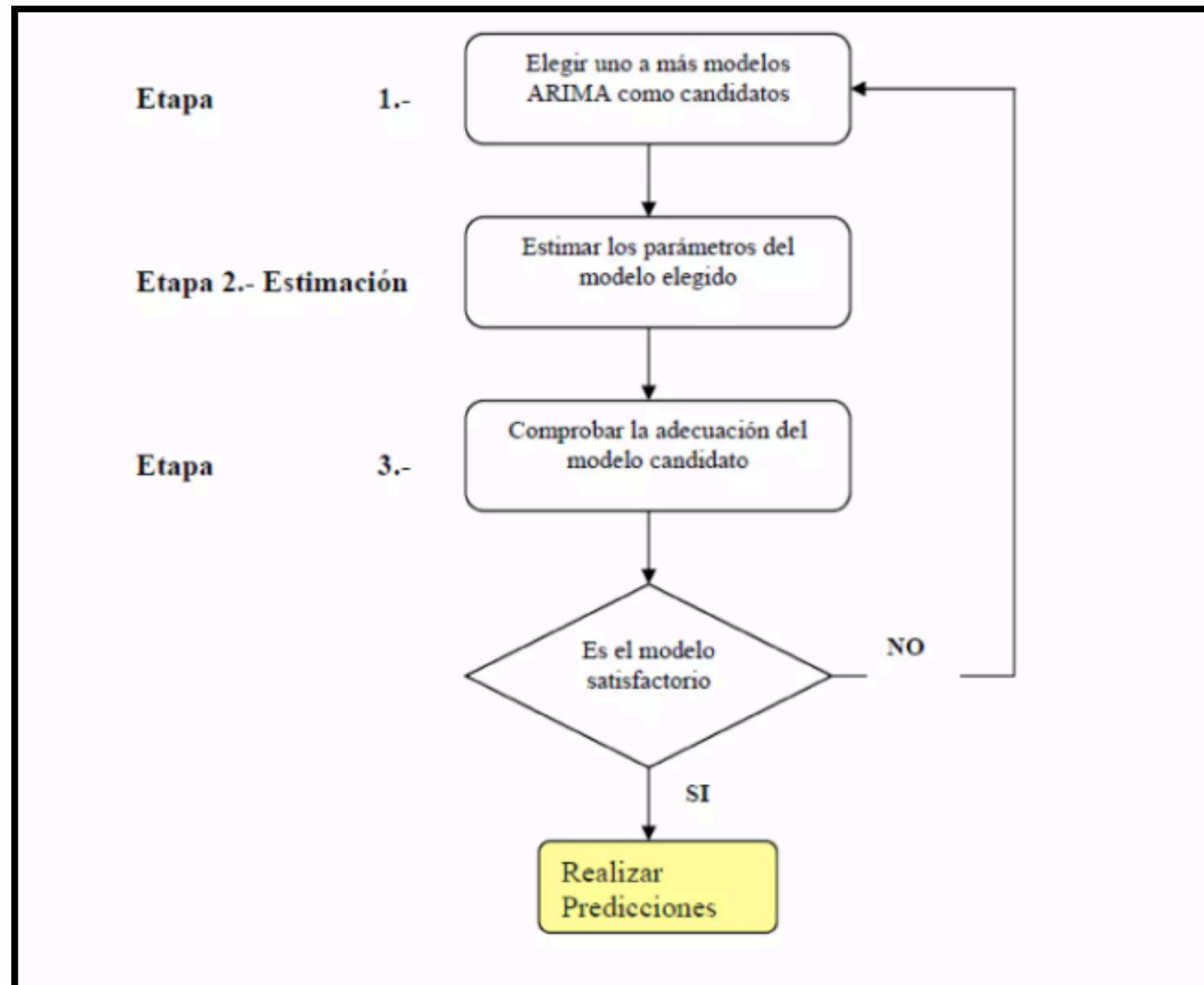
Promedio móvil integrado autorregresivo



Promedio móvil integrado autorregresivo

	S → Estacionalidad → Patrón recurrente o variación en los datos a intervalos fijos.
	AR → Comp. Autorregresivo → valor actual de la serie temporal influenciado por sus valores pasados.
	I → Integrado → Hacer que la serie temporal sea estacionaria cuando hay una tendencia o estacionalidad.
	MA → Media móvil → El valor actual de la serie temporal es influenciada por sus residuos pasados.
	X → Variables exogenas → Factores externos que impactan la serie temporal.

Promedio móvil integrado autorregresivo



Carga y preparación de datos:

- Se importan las librerías necesarias.
- Se cargan los datos desde un archivo Excel y se convierten las fechas al formato correcto.
- Se visualizan los primeros datos y la serie temporal.

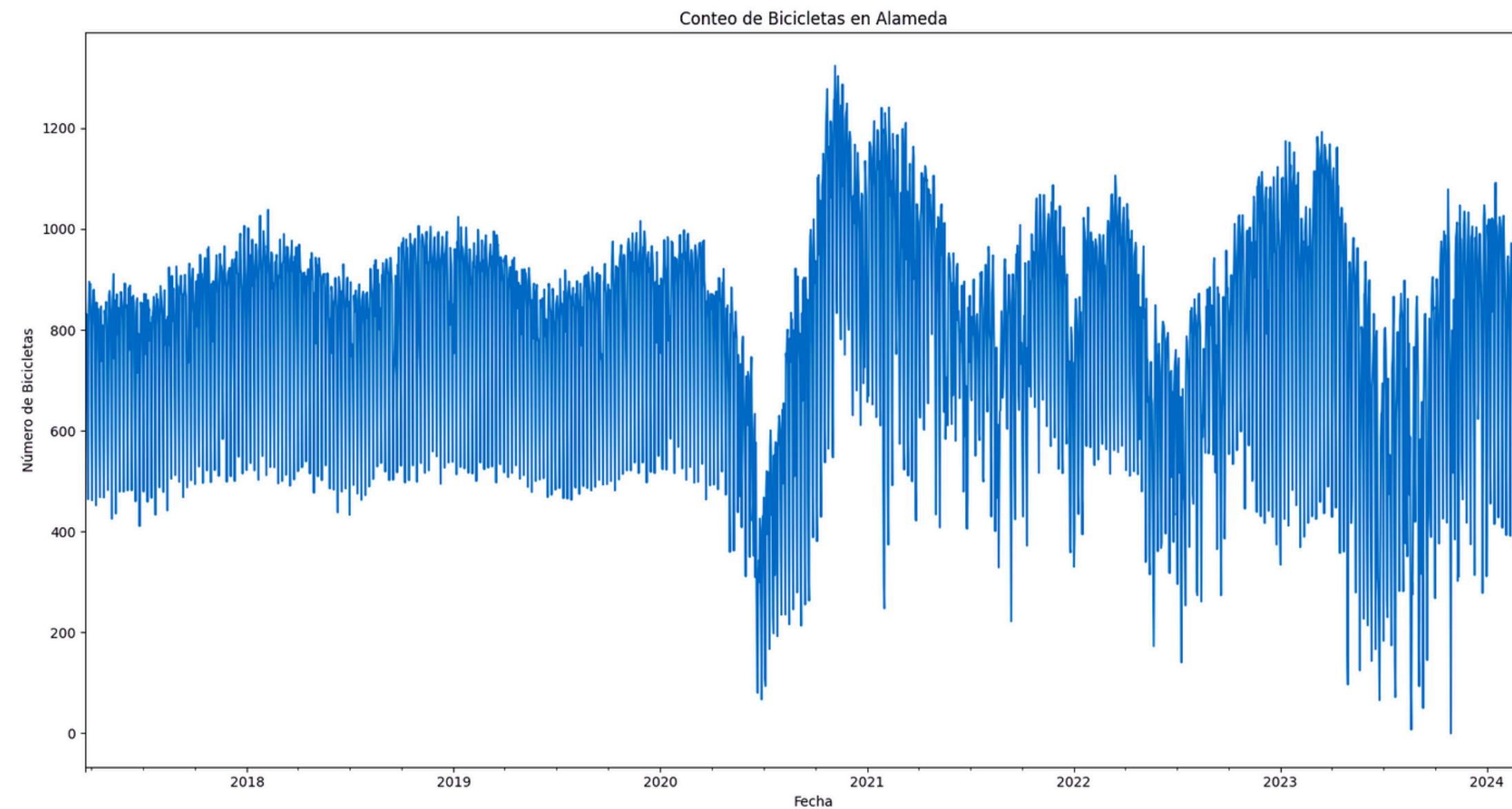
Análisis de la serie temporal:

- Se descompone la serie temporal para verificar componentes estacionales mediante `seasonal_decompose`.
- Se utiliza `auto_arima` para determinar los mejores parámetros (p, d, q) del modelo, incluyendo la estacionalidad.

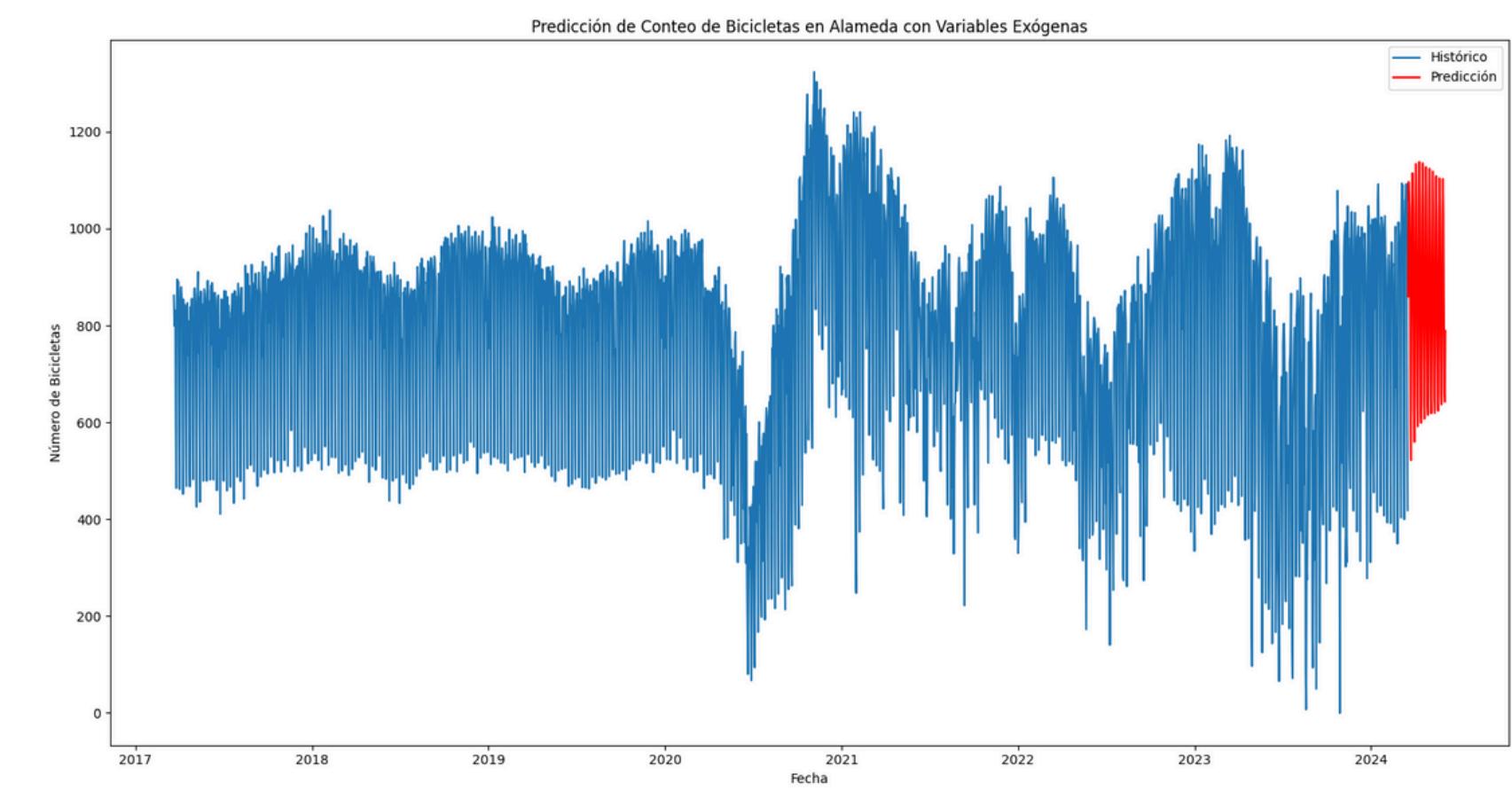
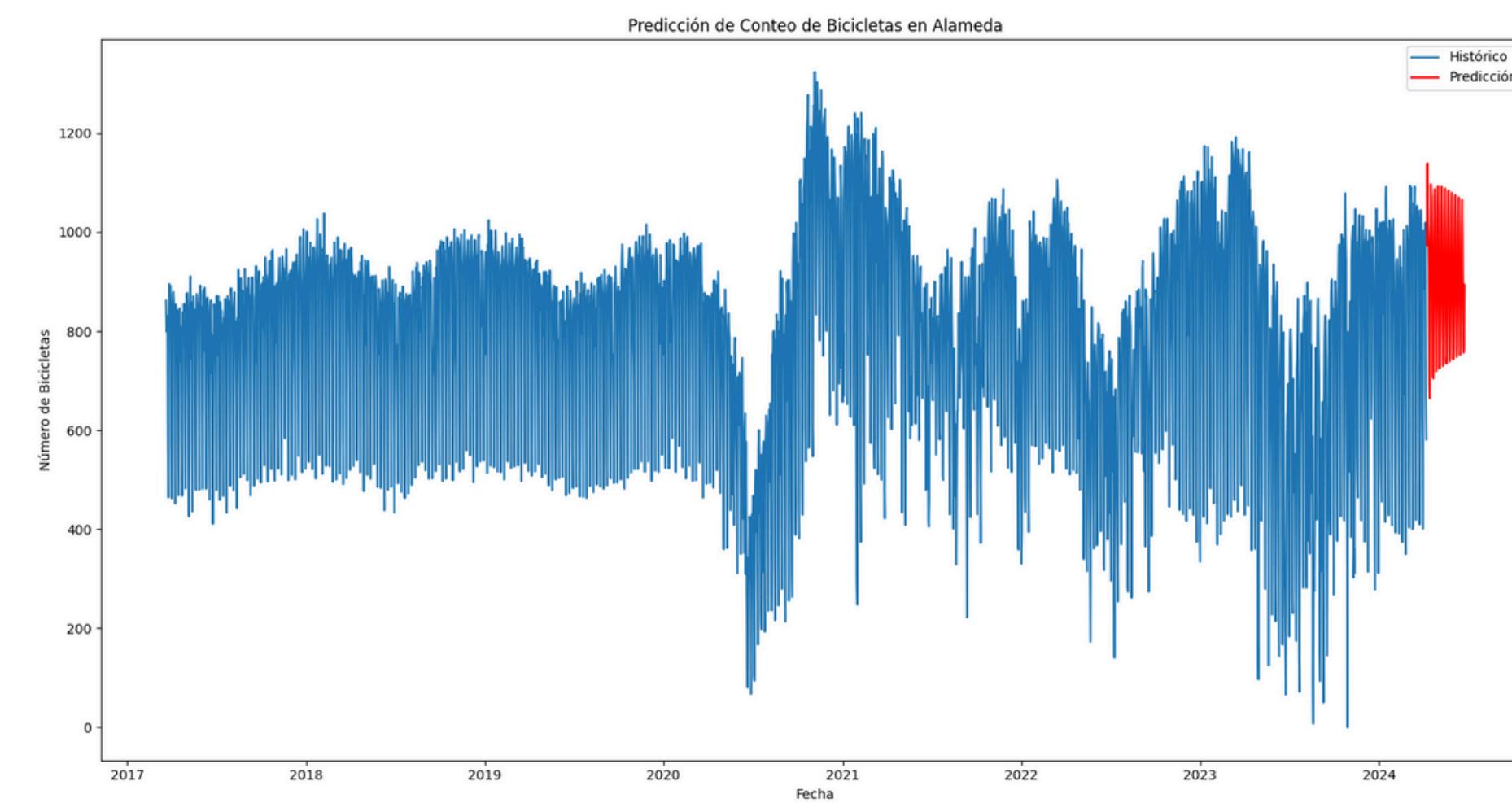
Modelado y predicción:

- Se ajusta un modelo SARIMA con los parámetros obtenidos.
- Se realizan predicciones para un período futuro específico.
- Las predicciones se convierten a un DataFrame con índice de fecha y se visualizan junto con los datos históricos.

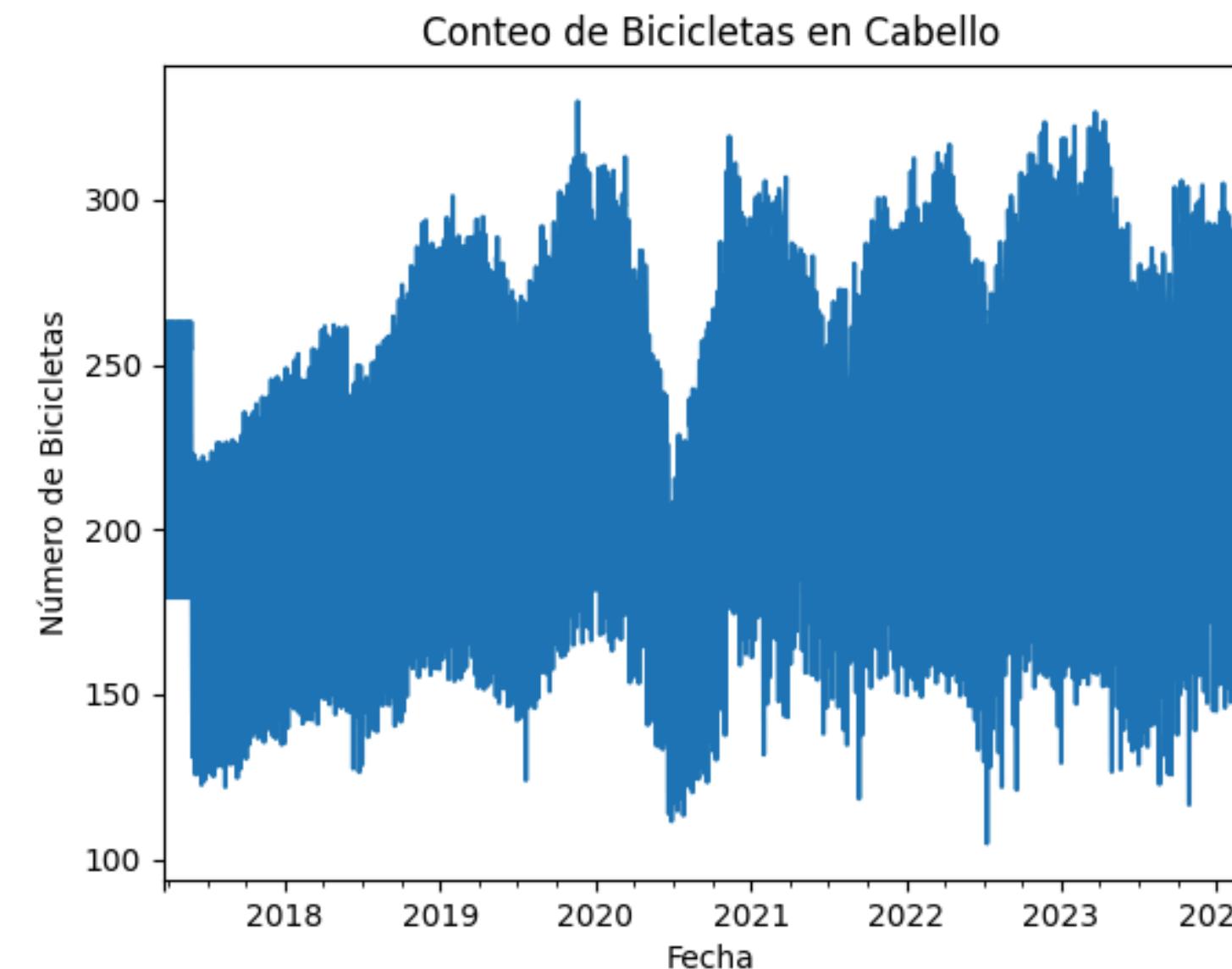
Modelo Arima para Alameda



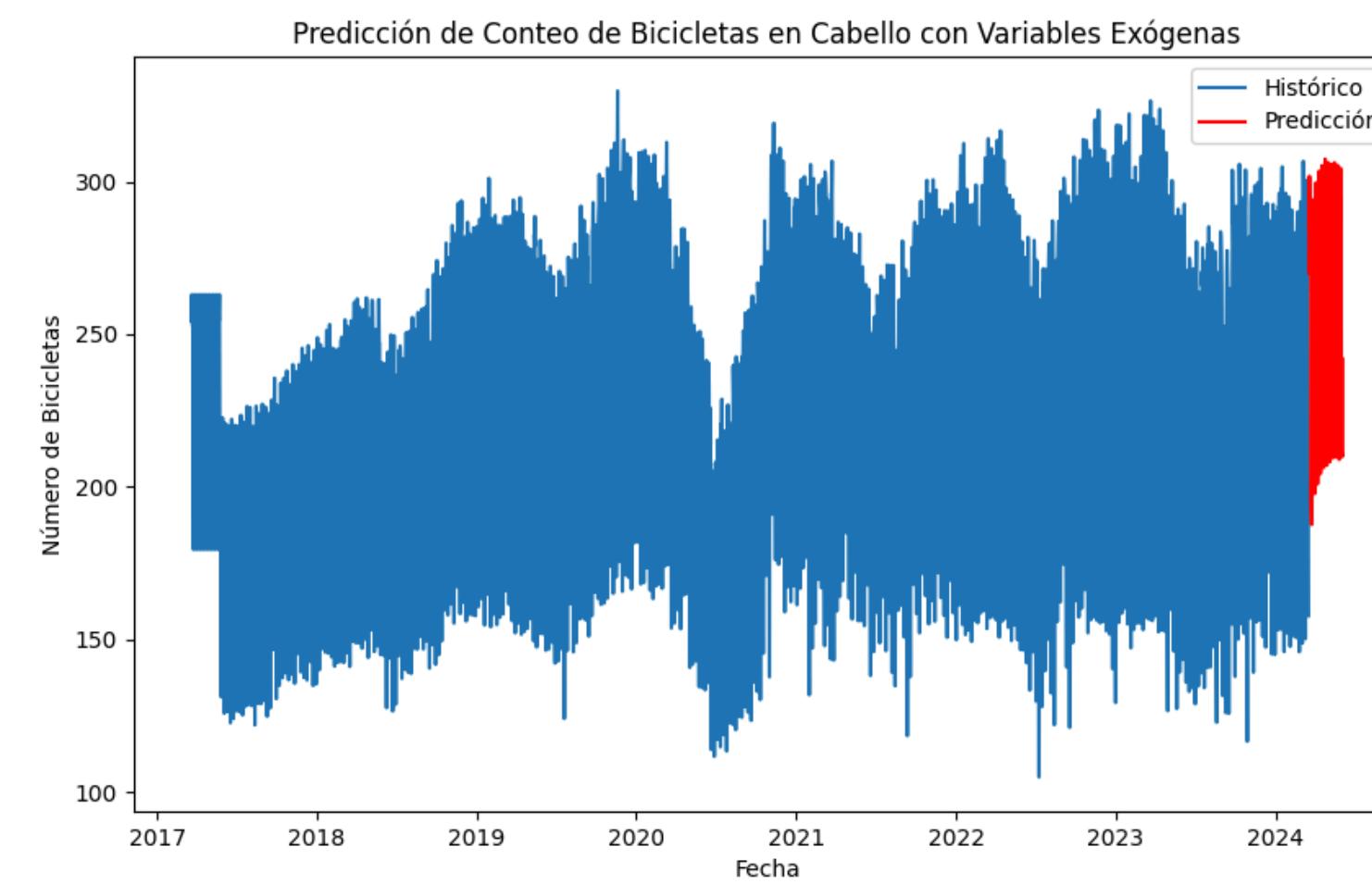
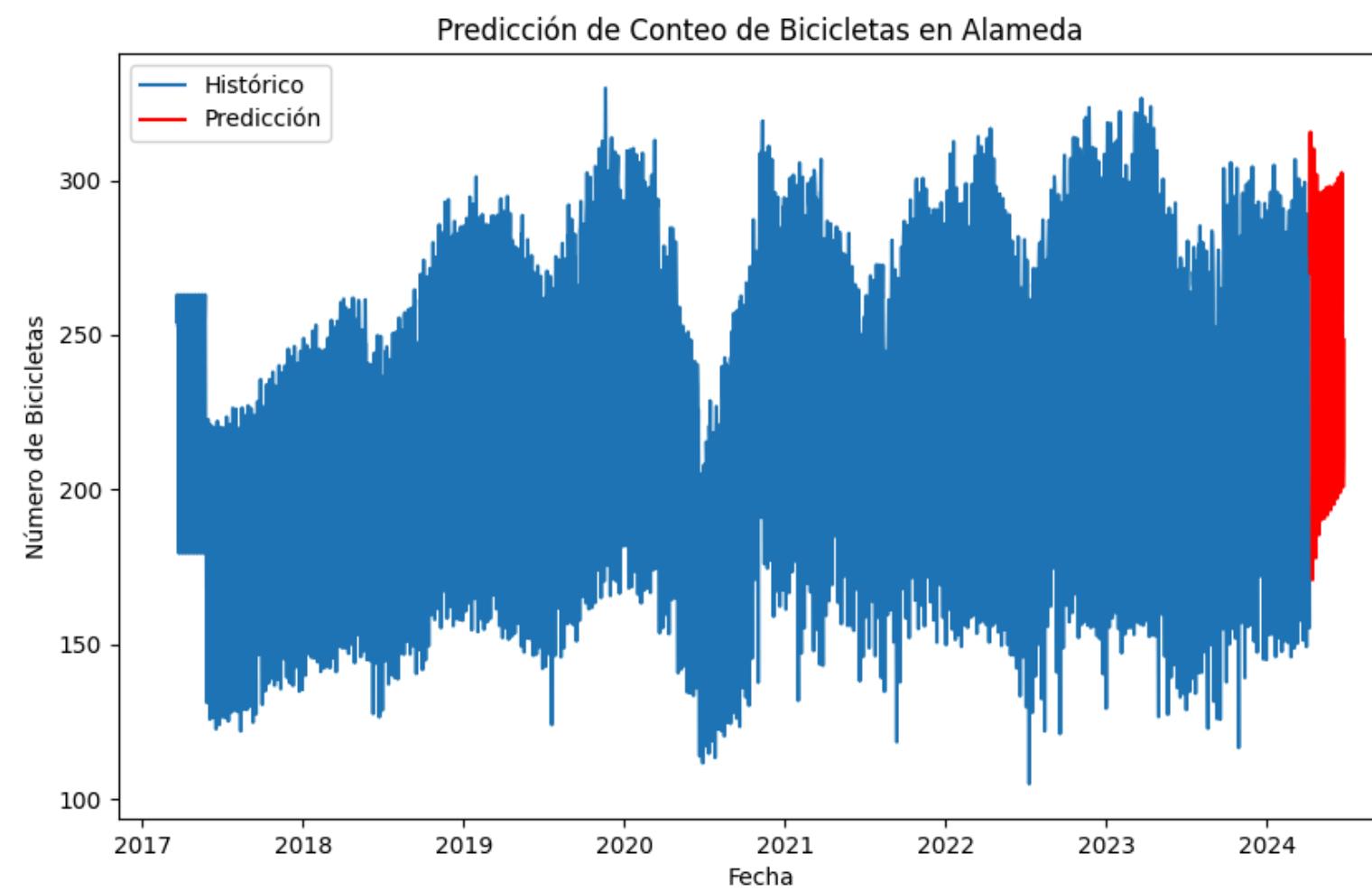
Modelo Arima para Alameda



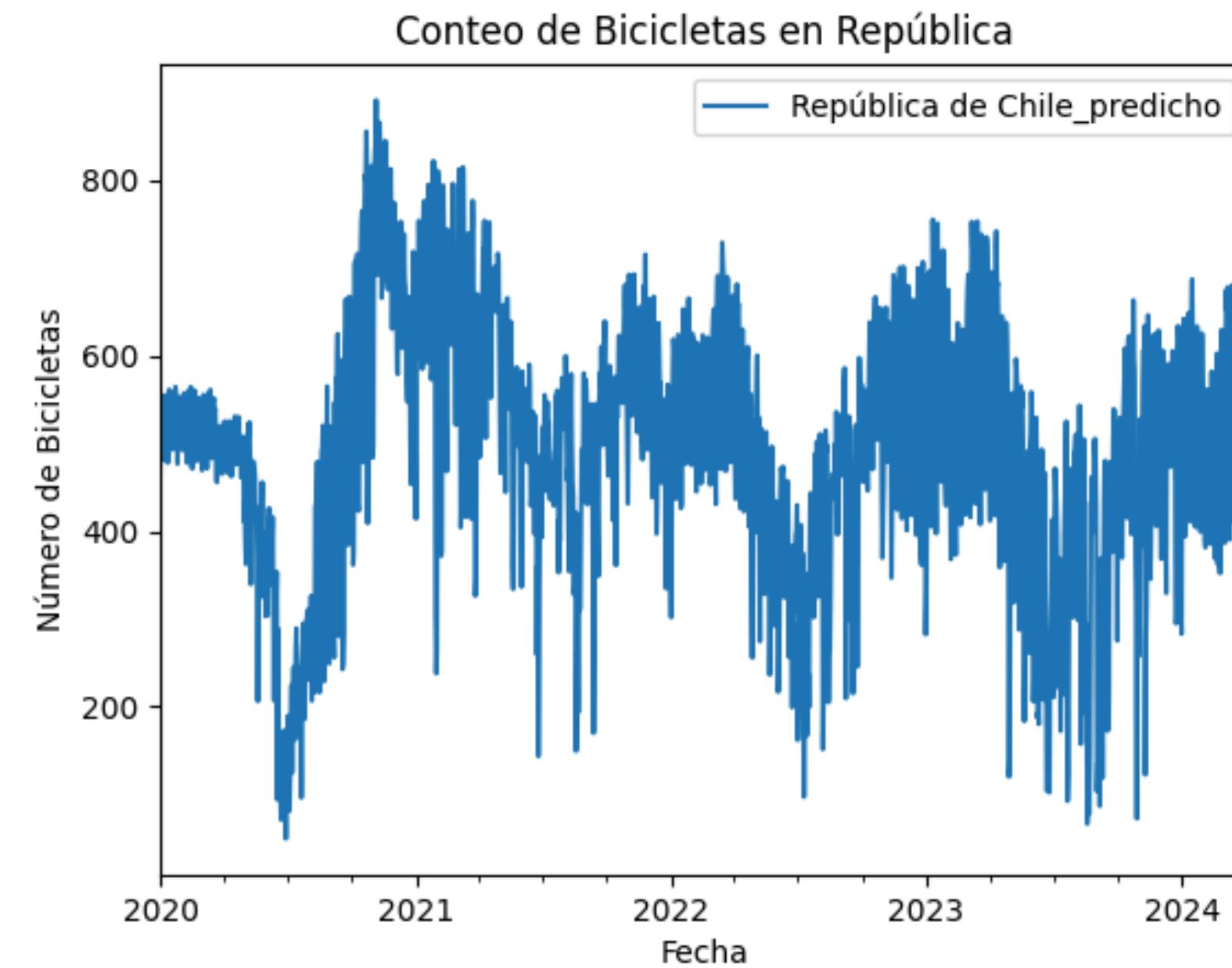
Modelo Arima para Cabello



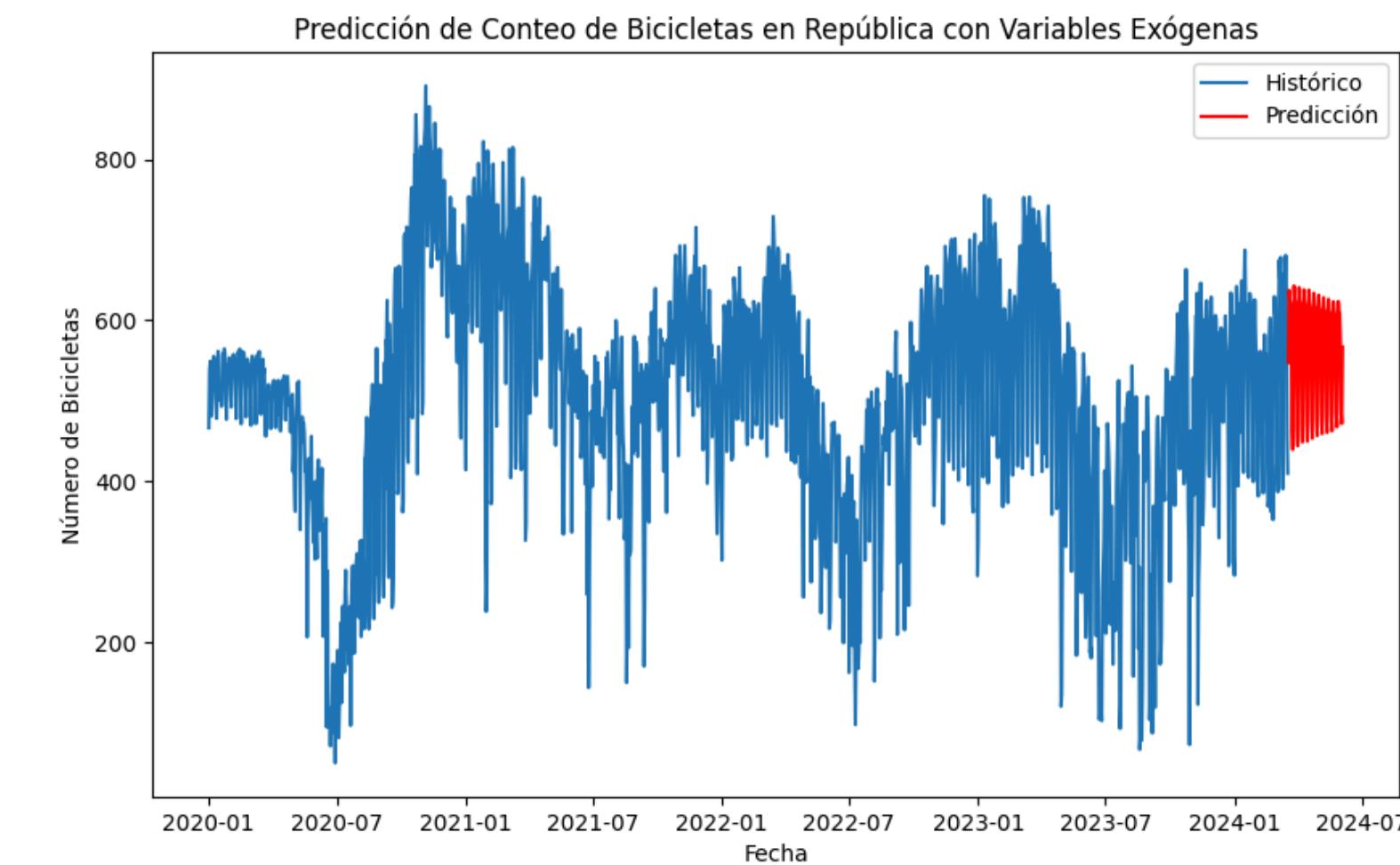
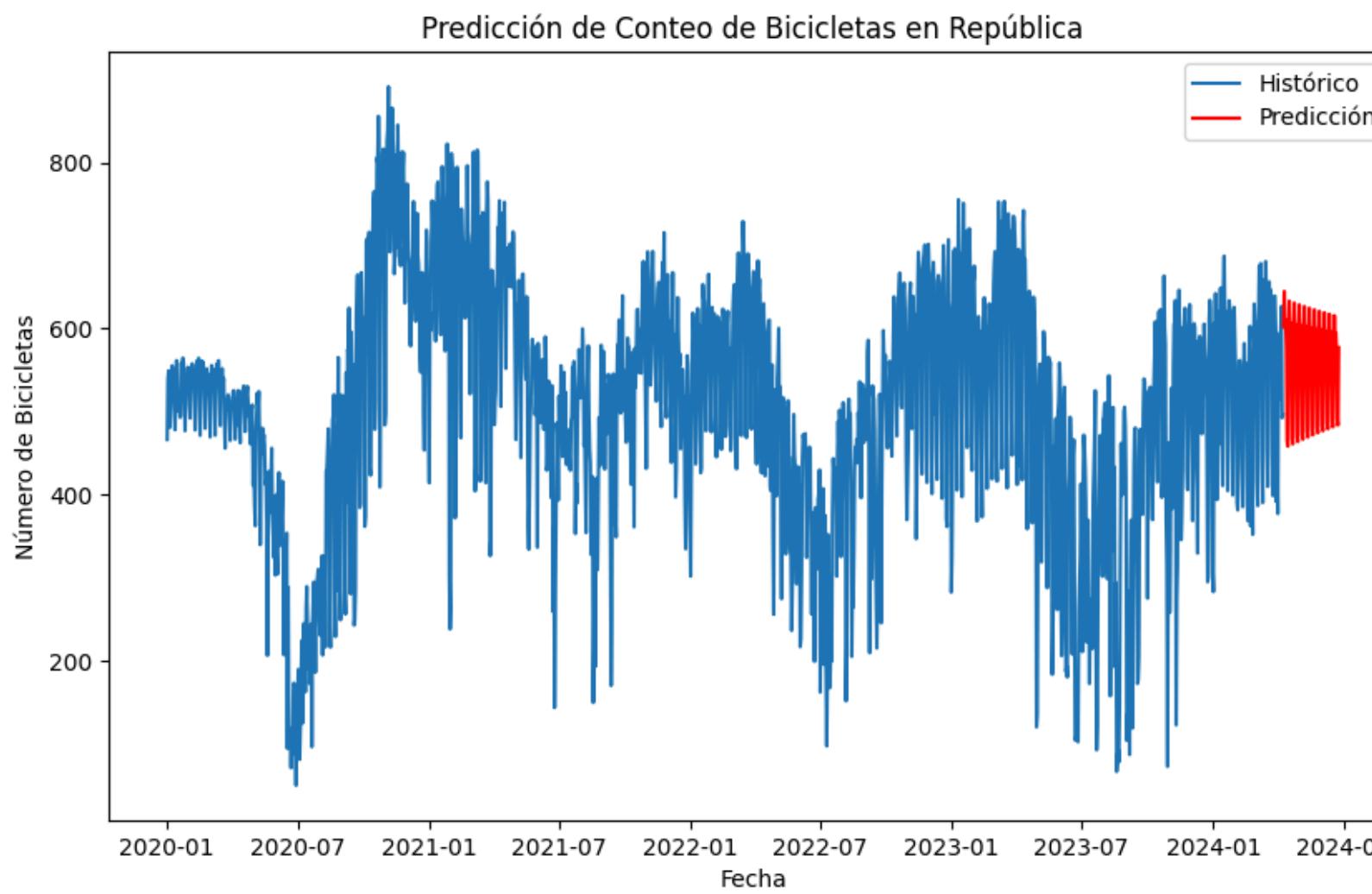
Modelo Arima para Cabello



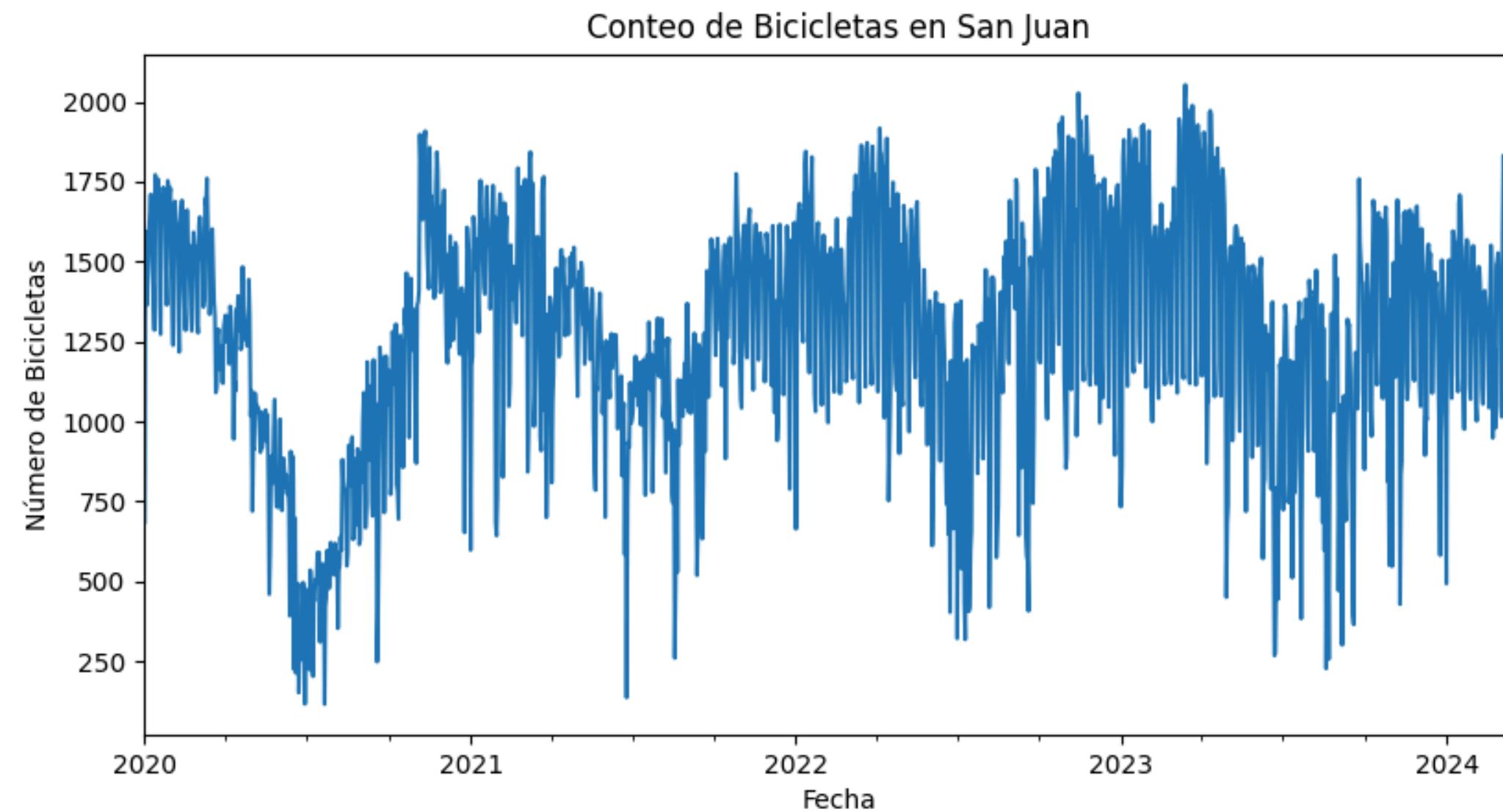
Modelo Arima para República de Chile con San Joaquín



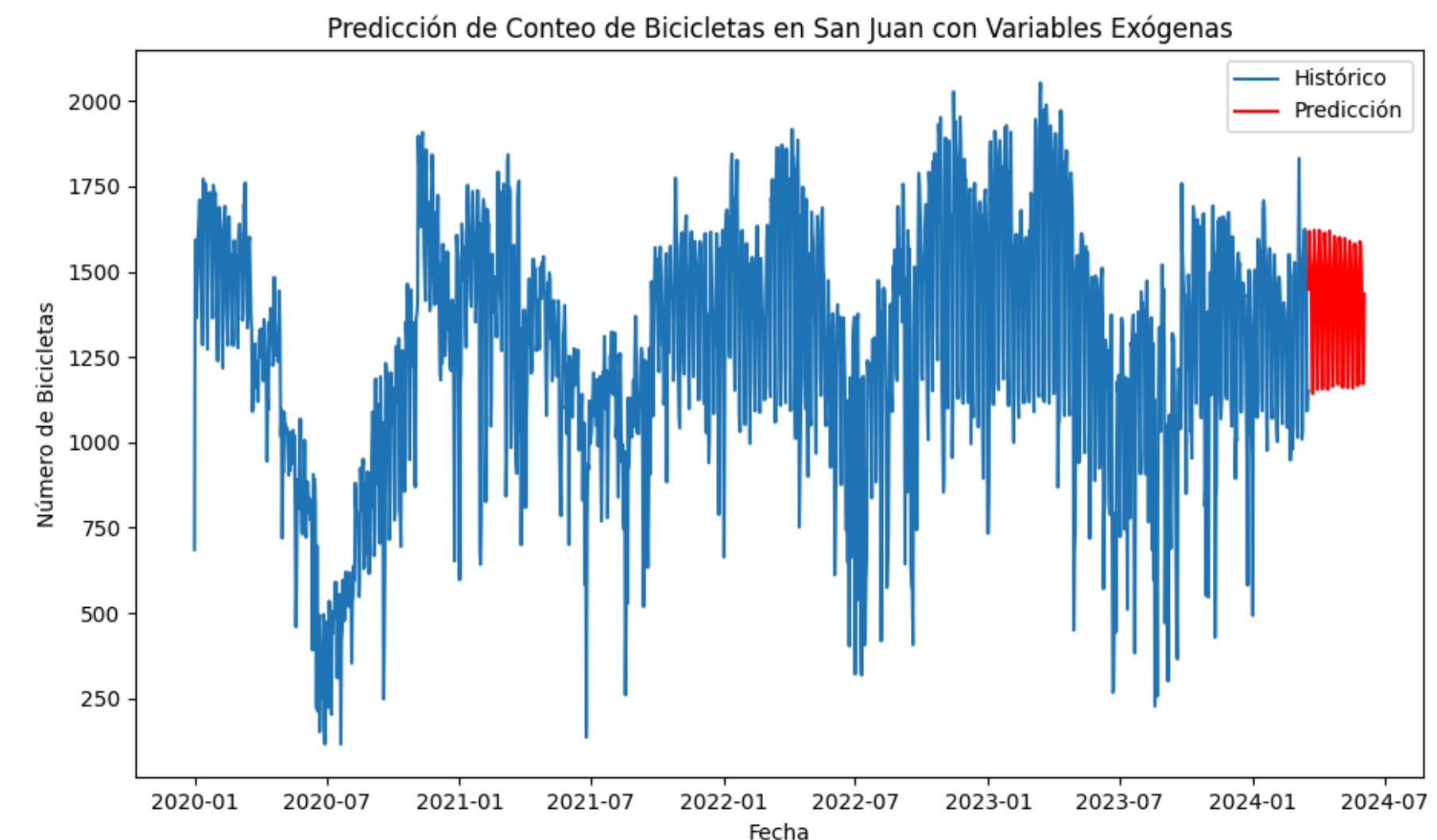
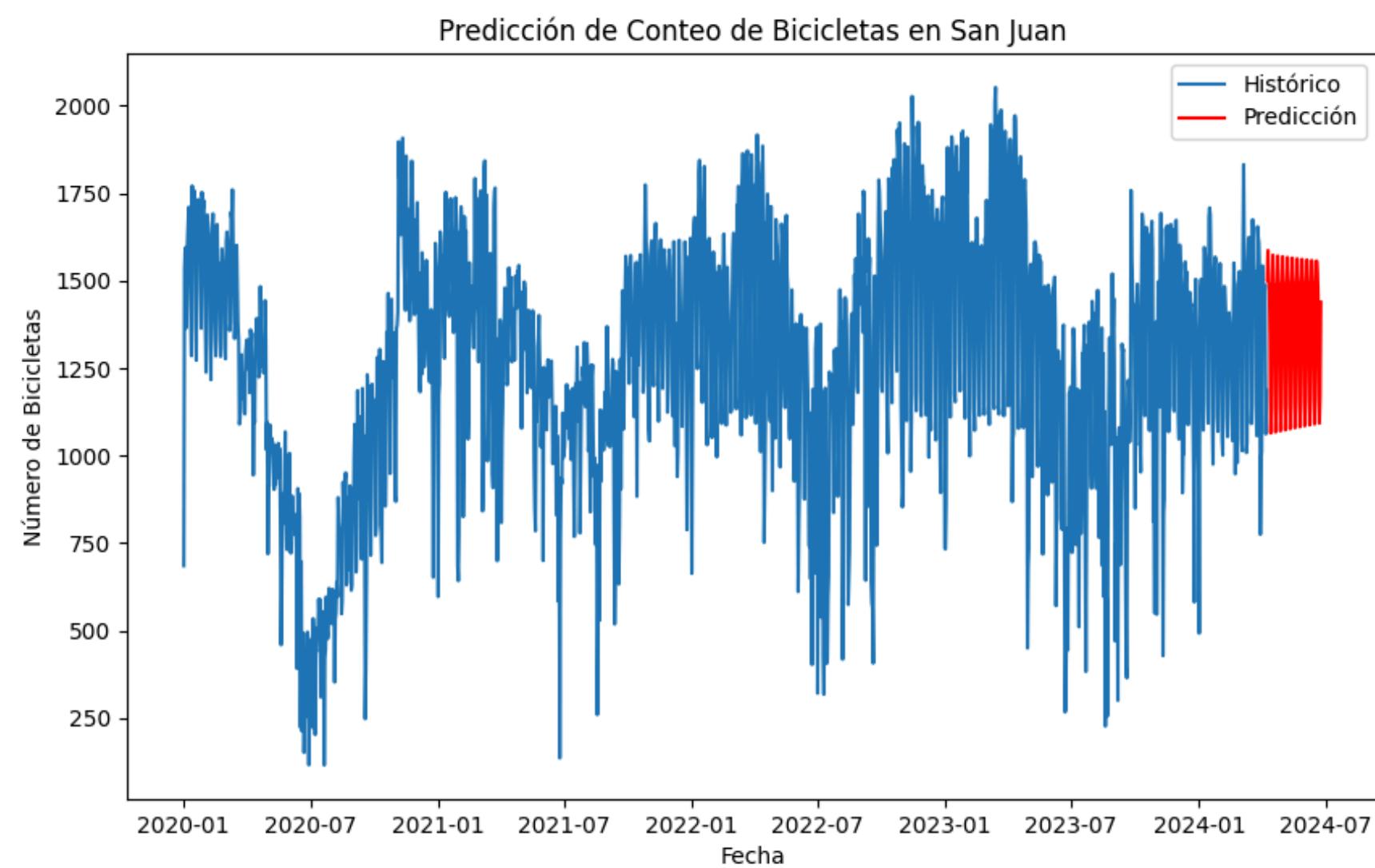
Modelo Arima para República de Chile con San Joaquín



Modelo Arima para San Juan con Escrivá de Balaguer



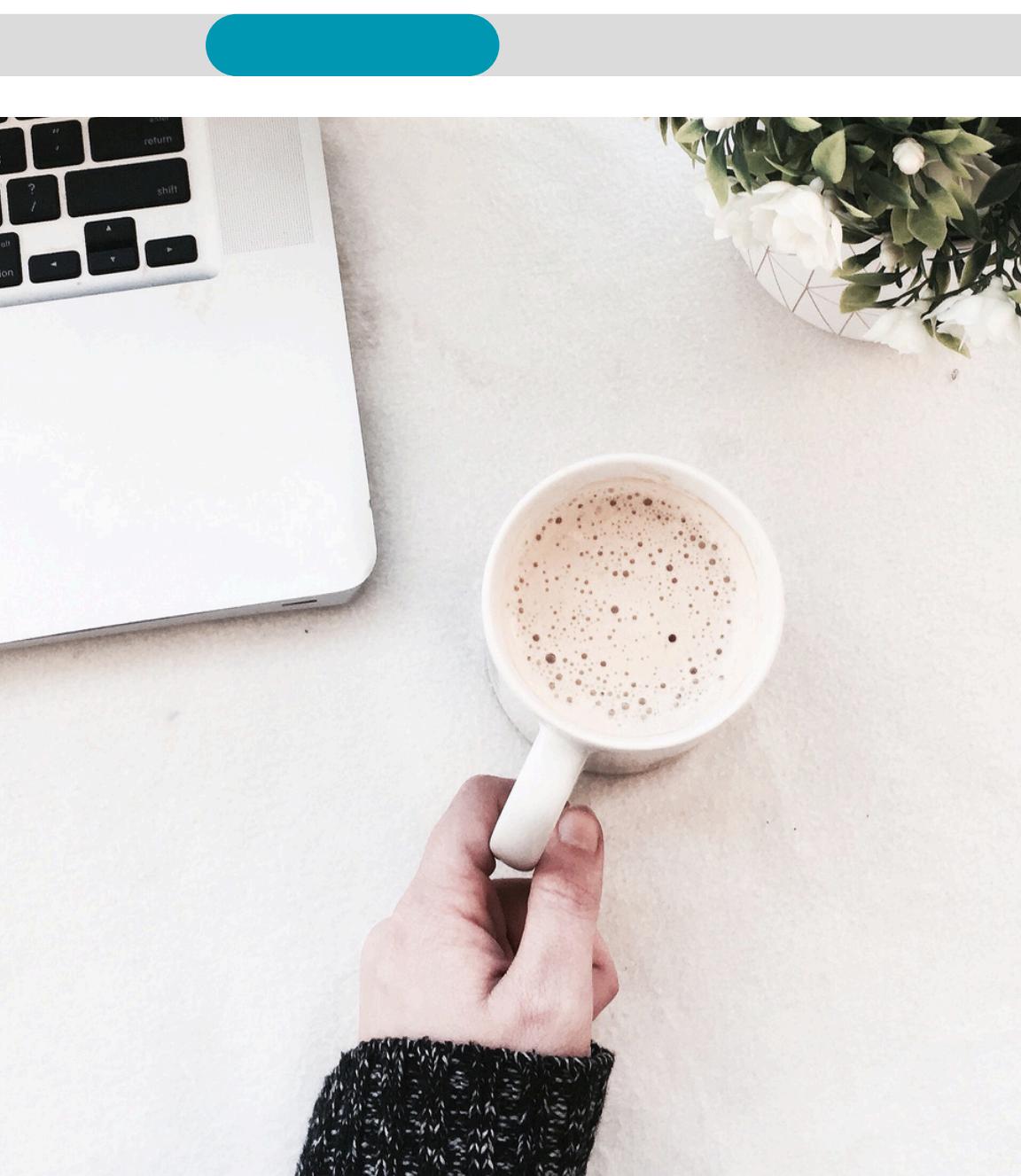
Modelo Arima para San Juan con Escrivá de Balaguer



Variable exógena	Metodo	Alameda	Republica	San juan	Cabello
Sin	Arima	397.12%	<u>49.23%</u>	<u>26.44%</u>	<u>12.97%</u>
Con	Arima	<u>313.52%</u>	51.4%	28.25%	23.93%
Sin	Prohet	57.49%	<u>19.40%</u>	<u>19.68%</u>	5.05%
Con	Prohet	<u>55.59%</u>	19.63%	19.80%	<u>5.01%</u>

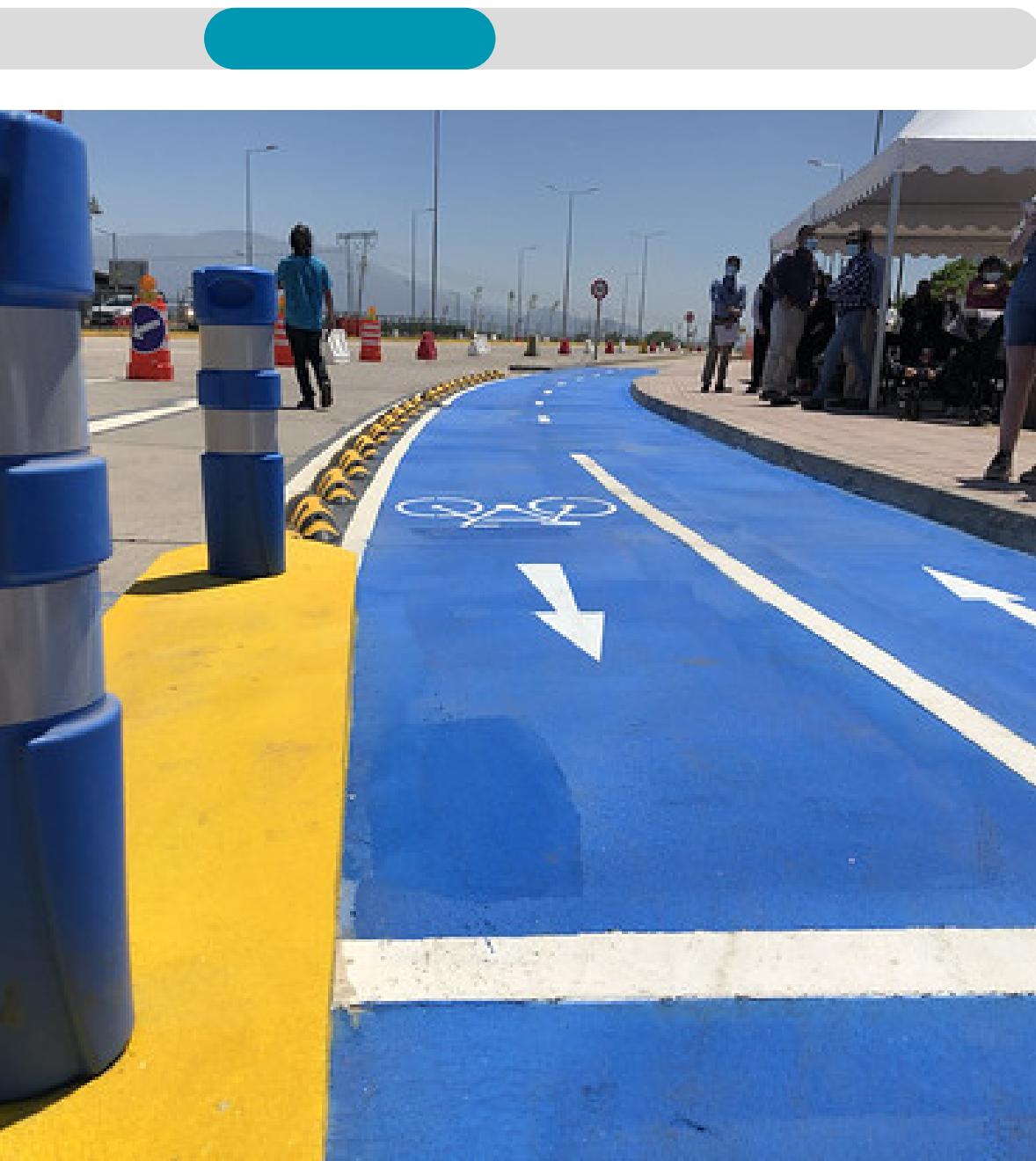
Variable exógena	Metodo	Alameda	Republica	San juan	Cabello
Sin	Arima	184.72	<u>121.29</u>	<u>193.65</u>	<u>28.4</u>
Con	Arima	<u>154.6</u>	122.47	196.15	47.76
Sin	Prohet	<u>79.67</u>	<u>65.24</u>	<u>167.42</u>	11.09
Con	Prohet	78.48	65.61	167.61	<u>10.99</u>

Cosas que podrían mejorar el modelo



- Excluir los días feriados
- Ajuste de la estacionalidad
- Transformaciones de datos
- Detección y tratamiento de outliers

índice



- 01** Introducción
- 02** Motivación
- 03** Análisis exploratorio
- 04** Prophet
- 05** Arima
- 06** Conclusión

Conclusiones

- Se realizó un desarrollo intensivo de limpieza y manejo de datos.
- Aplicación de modelos de análisis de series temporales.
- Se implementaron dos modelos: ARIMA y Prophet, ambos con la inclusión de variables exógenas, específicamente la temperatura máxima y mínima.
- Al comparar las métricas de error MAPE y MAD, se observó que el modelo Prophet proporciona valores más razonables que el modelo ARIMA.
- Esta diferencia puede deberse a las distintas formas en que funcionan ambos modelos. Sin embargo, al considerar posibles mejoras, es posible que el modelo ARIMA pueda llegar a ser más efectivo.

MUCHAS GRACIAS



¿Preguntas?

