Proyecto Ecobici-AutoML

# Descripción del flujo operativo Ecobici-AutoML

Este documento resume el flujo completo del proyecto Ecobici-AutoML hasta la versión con dashboard simplificado (25 octubre 2025).  
Incluye las etapas principales del pipeline, descripción de los archivos generados y observaciones finales sobre el despliegue.

## 1. Etapas del pipeline

1️⃣ Ingesta → Recolección de datos desde la API Ecobici y guardado en parquet.  
2️⃣ Preprocesamiento → Limpieza, imputación y creación de variables lag, rolling y estacionales.  
3️⃣ AutoML → Entrenamiento y comparación de modelos con PyCaret, H2O y FLAML.  
4️⃣ Evaluación → Métricas por horizonte (1,3,6,12h) y reporte en /reports.  
5️⃣ Despliegue → Generación batch (predict\_batch.py) y tiles para el mapa de calor del dashboard.

## 2. Reportes generados

- backtest\_metrics\_by\_horizon.csv → desempeño por horizonte.  
- error\_by\_hour\_and\_horizon.csv → análisis de errores temporales.  
- holdout\_month\_metrics.csv → validación externa.

## 3. Archivos de salida batch

Predicciones: /predictions/\*.parquet  
Columnas:  
 • station\_id → ID numérico de la estación.  
 • timestamp\_pred → hora de la predicción.  
 • h → horizonte en horas.  
 • yhat → valor predicho (bicis disponibles).  
 • level → nivel de confianza o estado del recurso.

## 4. Observaciones finales

La versión del 25/10/2025 marca el cierre de la etapa de modelado y despliegue básico.  
El dashboard simplificado permite visualizar la disponibilidad horaria de bicicletas por estación, filtrando por horizonte.  
Siguientes pasos potenciales: integración de caché, métricas dinámicas y API pública.