# 📊 MobilePandaAnalytic – Final Project (Ironhack Data Analytics)

## 🎯 Objetivo

Analizar y predecir la dinámica entre \*\*inflación\*\* y \*\*rendimientos de bonos soberanos a 10 años (yield 10Y)\*\* en varios países, usando dos enfoques:

- \*\*Econometría\*\*: modelos VAR.

- \*\*Machine Learning\*\*: ElasticNet, Random Forest, Gradient Boosting.

Comparar resultados entre países y evaluar qué método predice mejor.

---

## 🗂️ Datos

- \*\*Yields (bonos 10Y)\*\*: dataset `yields.csv` (varios países, anual/mensual).

- \*\*World Bank\*\*: indicadores macro (`world\_bank\_data\_2025.csv`), incluyendo inflación.

- \*\*Complemento FRED (USA, DE)\*\*: para años recientes.

Países incluidos en este análisis:

- 🇺🇸 USA

- 🇩🇪 Alemania

- 🇬🇧 Reino Unido

- 🇯🇵 Japón

- 🇪🇸 España

Periodo: \*\*2010–2024 (anual)\*\*

---

## ⚙️ Metodología

1. \*\*Limpieza y alineación de datos\*\*

- Resample a anual.

- Emparejar inflación (WB) y yields (10Y).

2. \*\*Modelos evaluados\*\*

- Naive (baseline)

- VAR (econometría)

- ElasticNet (ML lineal regularizado)

- Random Forest

- Gradient Boosting

3. \*\*Backtesting\*\*

- Ventana expandida.

- Horizonte h=1 año.

- Métricas: MAE y RMSE.

4. \*\*Comparaciones estadísticas\*\*

- Test de \*\*Diebold–Mariano (DM)\*\* para evaluar si las mejoras de ML son significativas.

---

## 📈 Resultados principales

### Ranking por país

👉 [`best\_model\_per\_country.csv`](reports/tables/best\_model\_per\_country.csv)

### Métricas detalladas (MAE / RMSE)

👉 [`rmse\_by\_country\_model.csv`](reports/tables/rmse\_by\_country\_model.csv)

### Test de Diebold–Mariano

👉 [`diebold\_mariano\_results.csv`](reports/tables/diebold\_mariano\_results.csv)

---

## 📊 Visualizaciones

### Ganador por países

![](reports/figures/model\_wins\_barplot.png)

### Real vs Predicho (mejor modelo por país)

Ejemplos (ver carpeta `reports/figures/` para todos):

![](reports/figures/USA\_real\_vs\_pred\_best.png)

![](reports/figures/Germany\_real\_vs\_pred\_best.png)

![](reports/figures/Japan\_real\_vs\_pred\_best.png)

---

## 📌 Conclusiones

- El \*\*modelo Naive\*\* (yield = último valor) es muy competitivo en frecuencia anual.

- \*\*Random Forest\*\* mejora significativamente en USA (DM p≈0.000).

- La muestra corta (10–15 obs) y la alta inercia del 10Y explican la fuerza del Naive.

- Shocks recientes (COVID, inflación 2021–2022) son difíciles de capturar.

---

## 🔮 Trabajo futuro

- Usar datos \*\*mensuales/trimestrales\*\* para más observaciones.

- Añadir más variables: política monetaria, expectativas de inflación, GDP growth.

- Probar modelos más avanzados: VARMAX, LSTM, XGBoost con tuning.

- Comparar horizontes h>1 (multi-step forecasts).

---

## 🧑‍💻 Estructura del repositorio