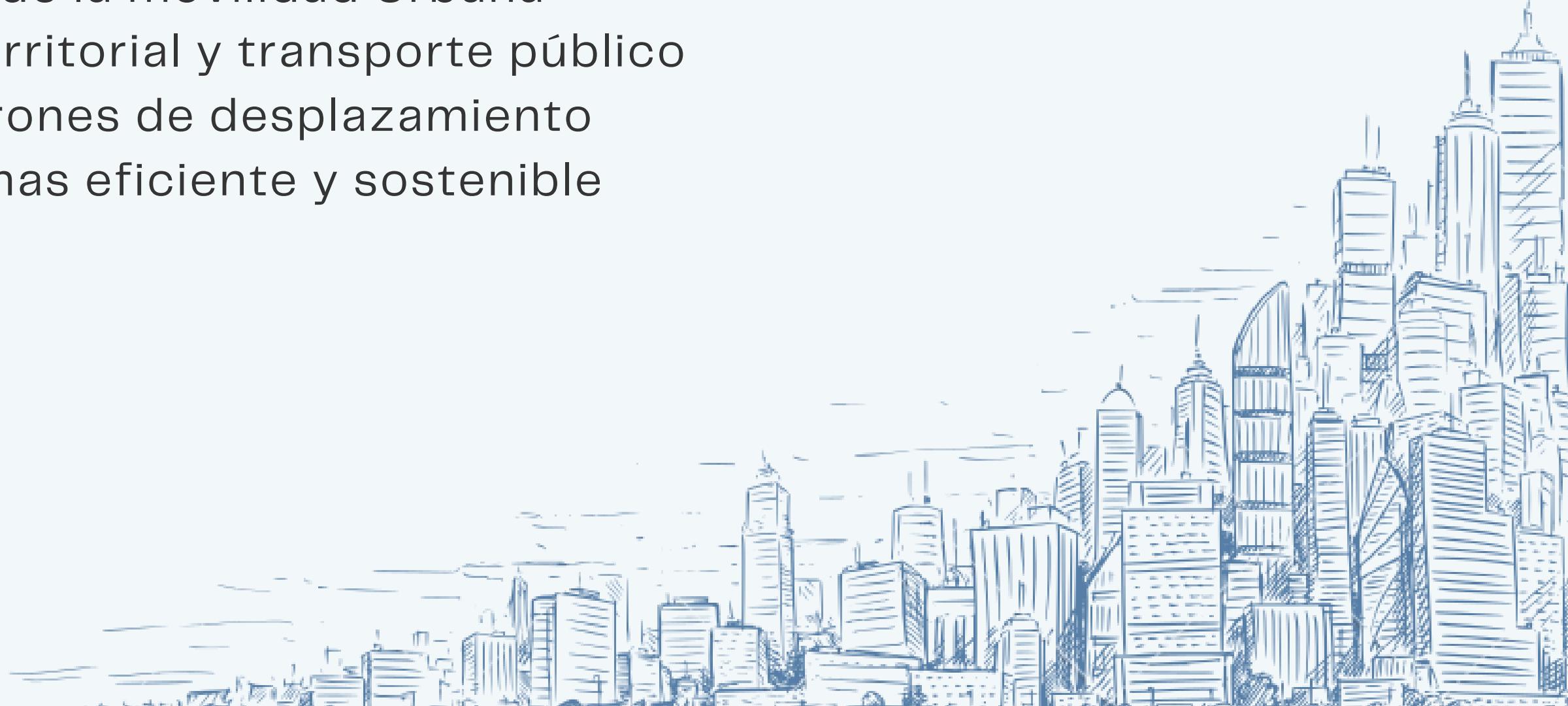


PROYECTO DATA SCIENCE

Integrantes: Lucas Abello, Samuel Angulo,
Tomás León, Natalia Ortega, Hugo Rojas

CONTEXTO Y MOTIVACIÓN

- Crecimiento y complejidad de la movilidad Urbana
- Desafíos en planificación territorial y transporte público
- Necesidad de anticipar patrones de desplazamiento
- Busqueda de una ciudad mas eficiente y sostenible

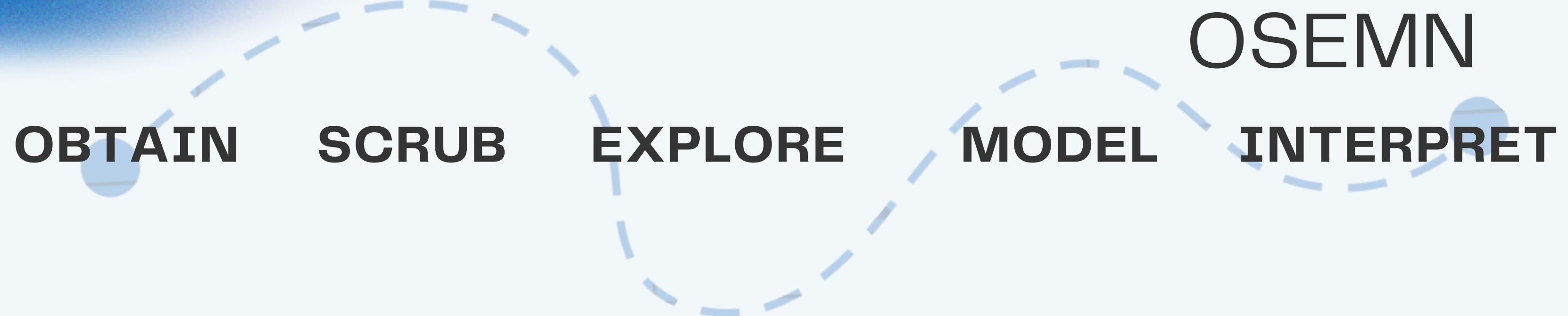


OBJETIVOS



- Predecir movilidad por comuna con un modelo de Machine Learning
- Anticipar flujos de personas a lo largo del tiempo
- Facilitar la planificación urbana para que sea mas precisa y eficiente

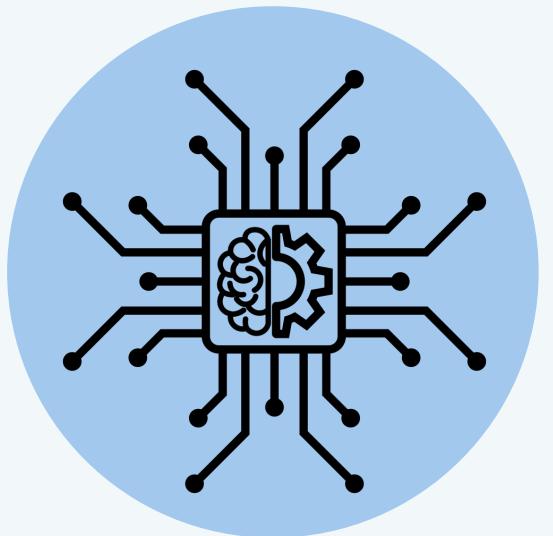
FRAMEWORK



Permite abordar integralmente el proceso de análisis y predicción de la expansión urbana, desde la obtención de datos hasta la interpretación de resultados.

ROLES

MACHINE LEARNING
ENGINEER



LUCAS ABELLO

DATA ANALYTICS
MANAGER



DATA ANALYST



NATALIA ORTEGA

DATA SCIENTIST



HUGO ROJAS

DATA ENGINEER



TOMÁS LEÓN



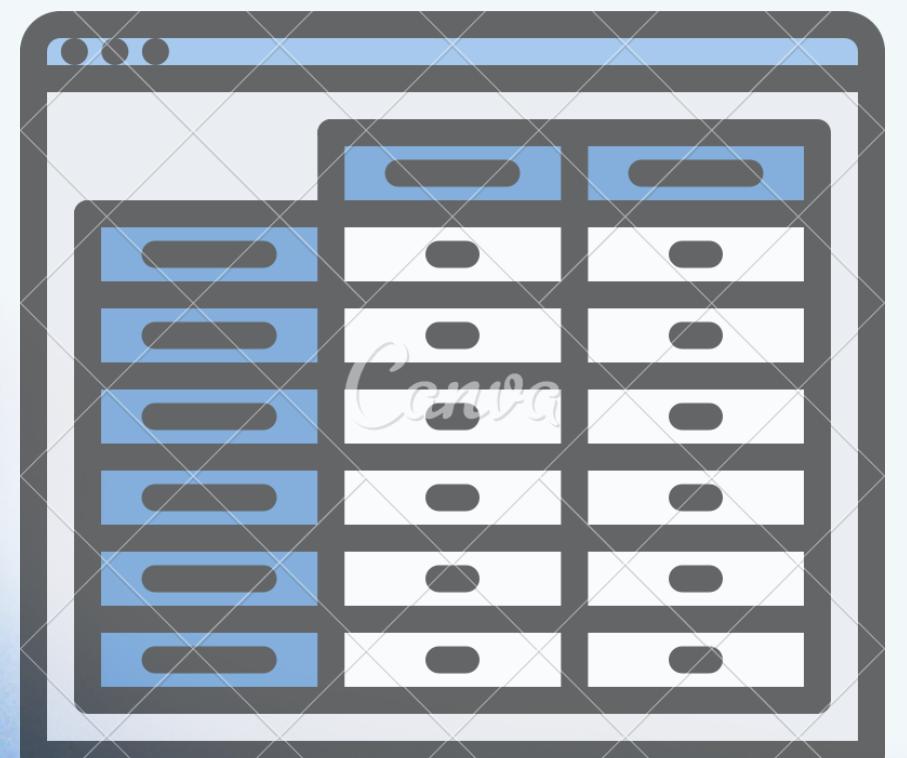
OBTENCIÓN DE DATOS

Basado en investigación respecto al comportamiento de la migración en la población en Chile



LIMPIEZA

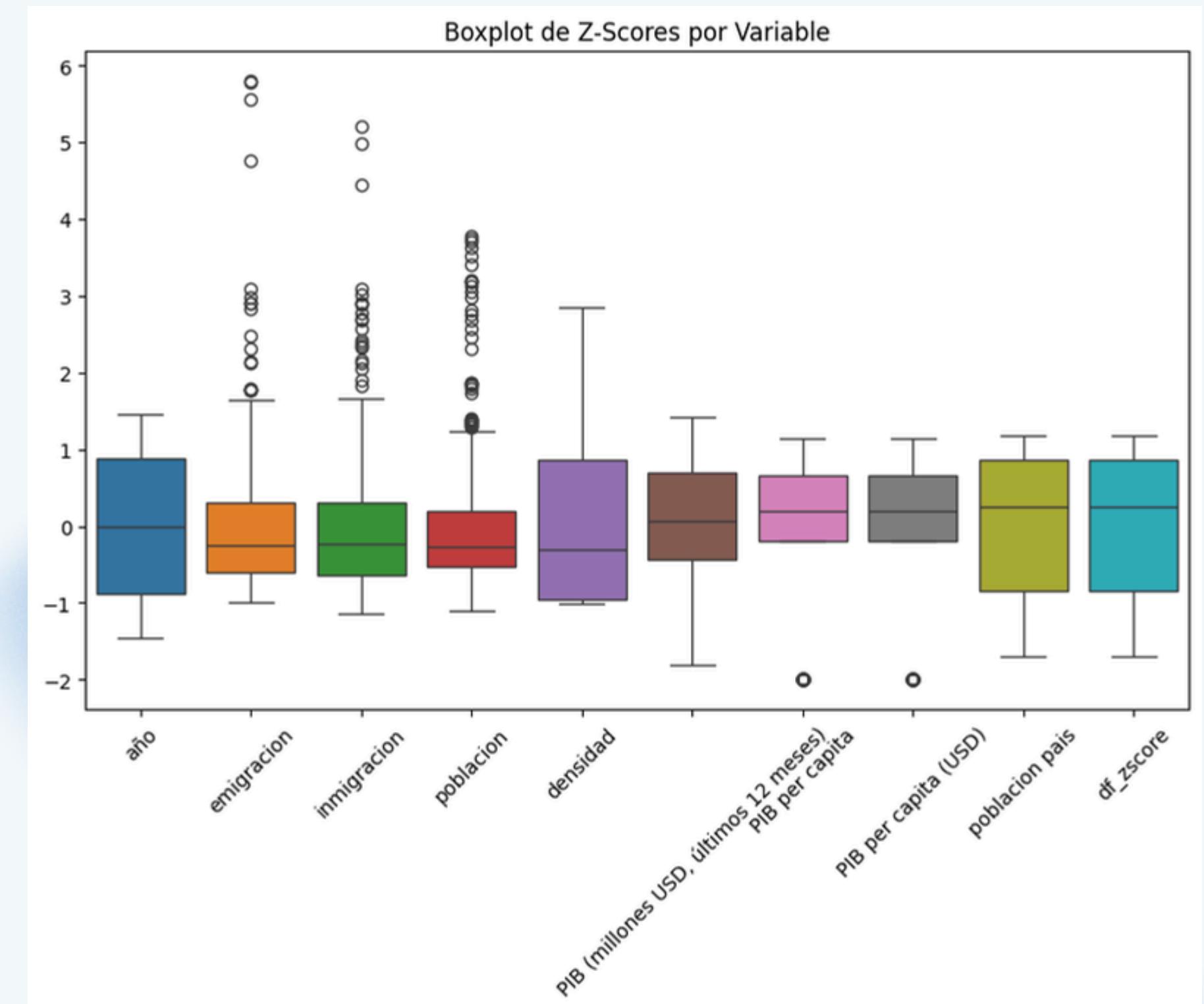
- Datos presentes no requeridos para el modelo
- Información comunas dentro de la Región Metropolitana
- Información provincias y regiones fuera de la Region
Metropolitana
- Información parcial y poco relevante 2002-2024
- Información de alta importancia 2018-2024



ANÁLISIS EXPLORATORIO

Outliers

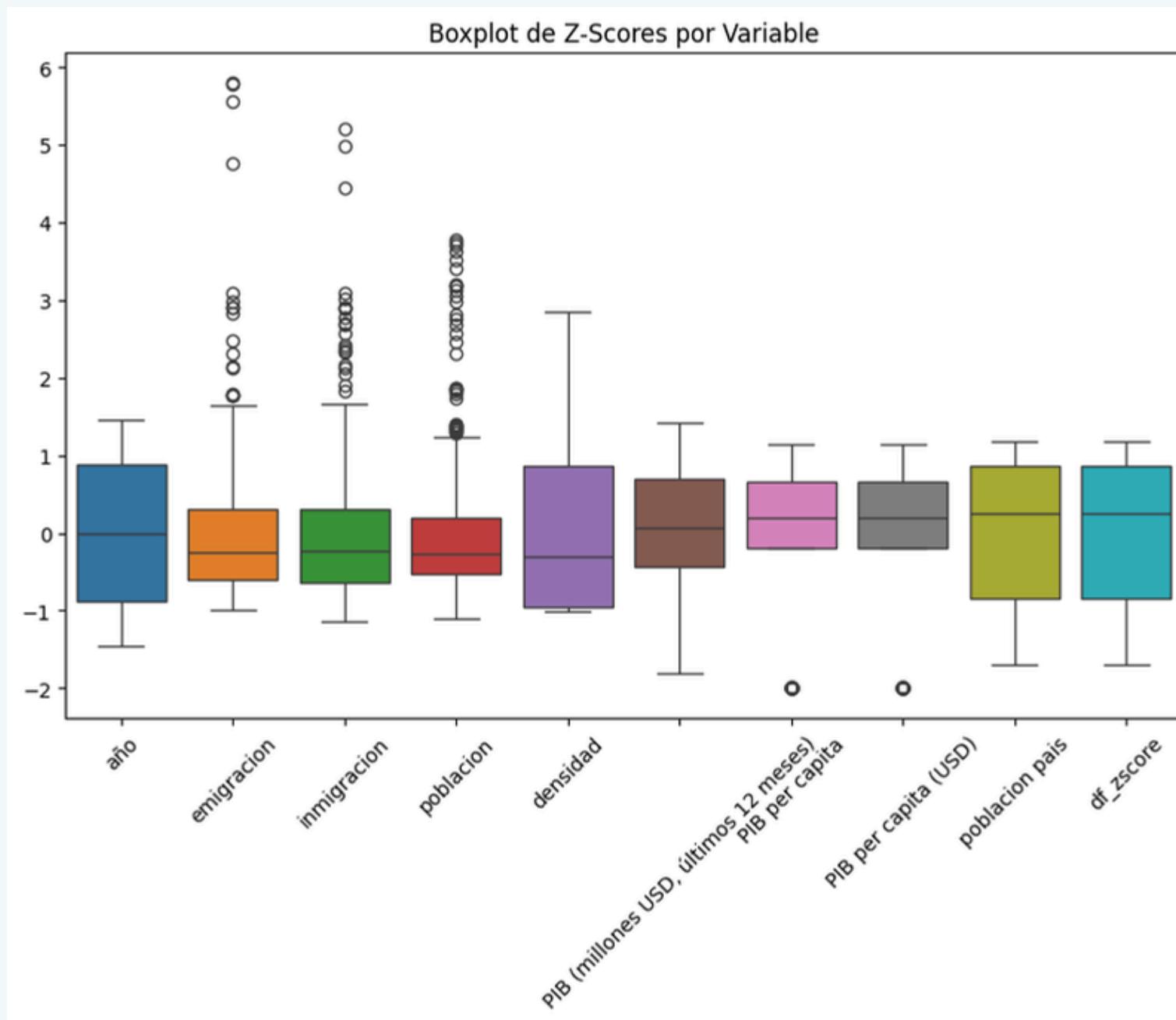
Se aplico una estandarización de los datos con Z-Score para analizar la presencia de outliers



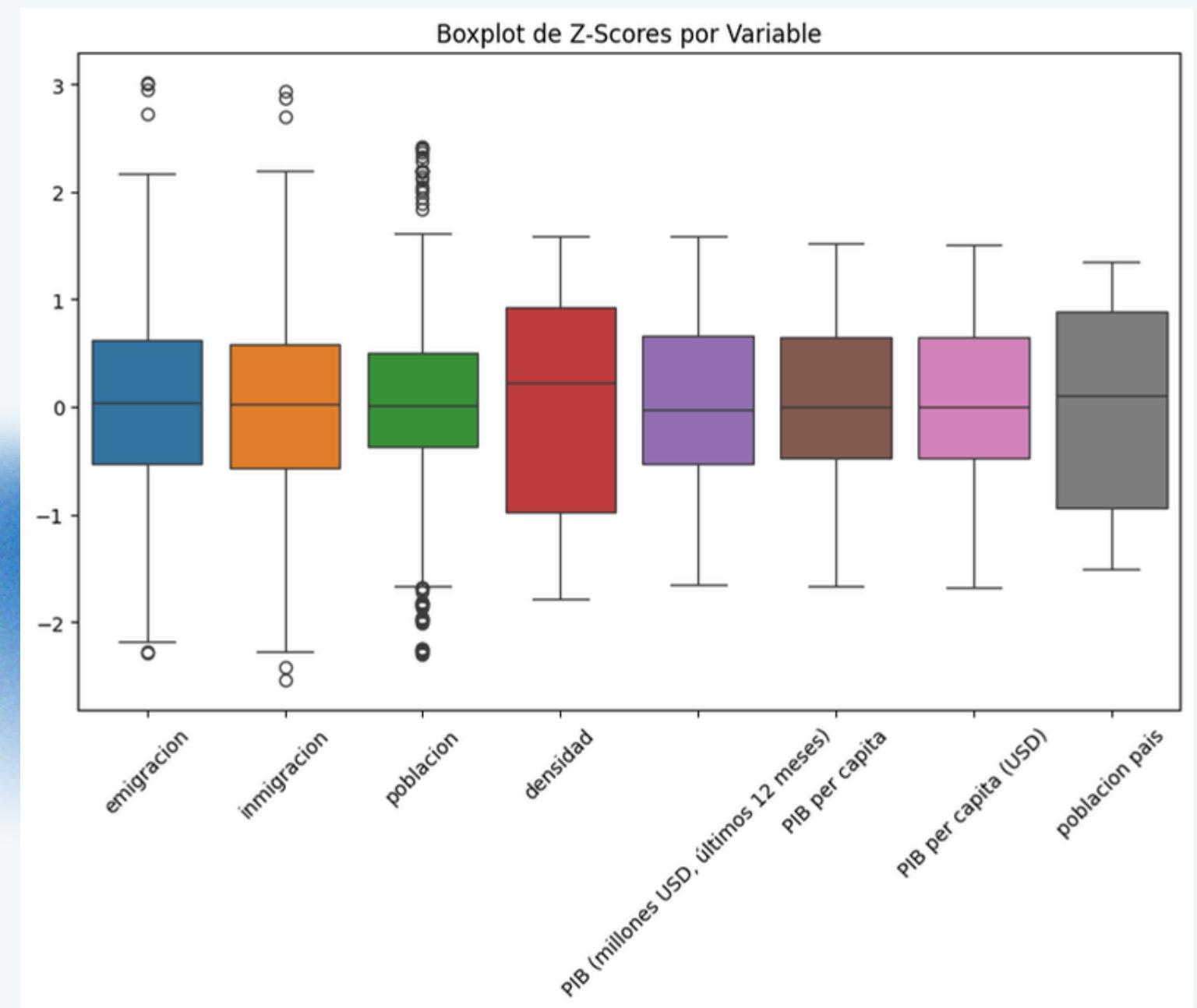
ANÁLISIS EXPLORATORIO

Outliers

Z-score



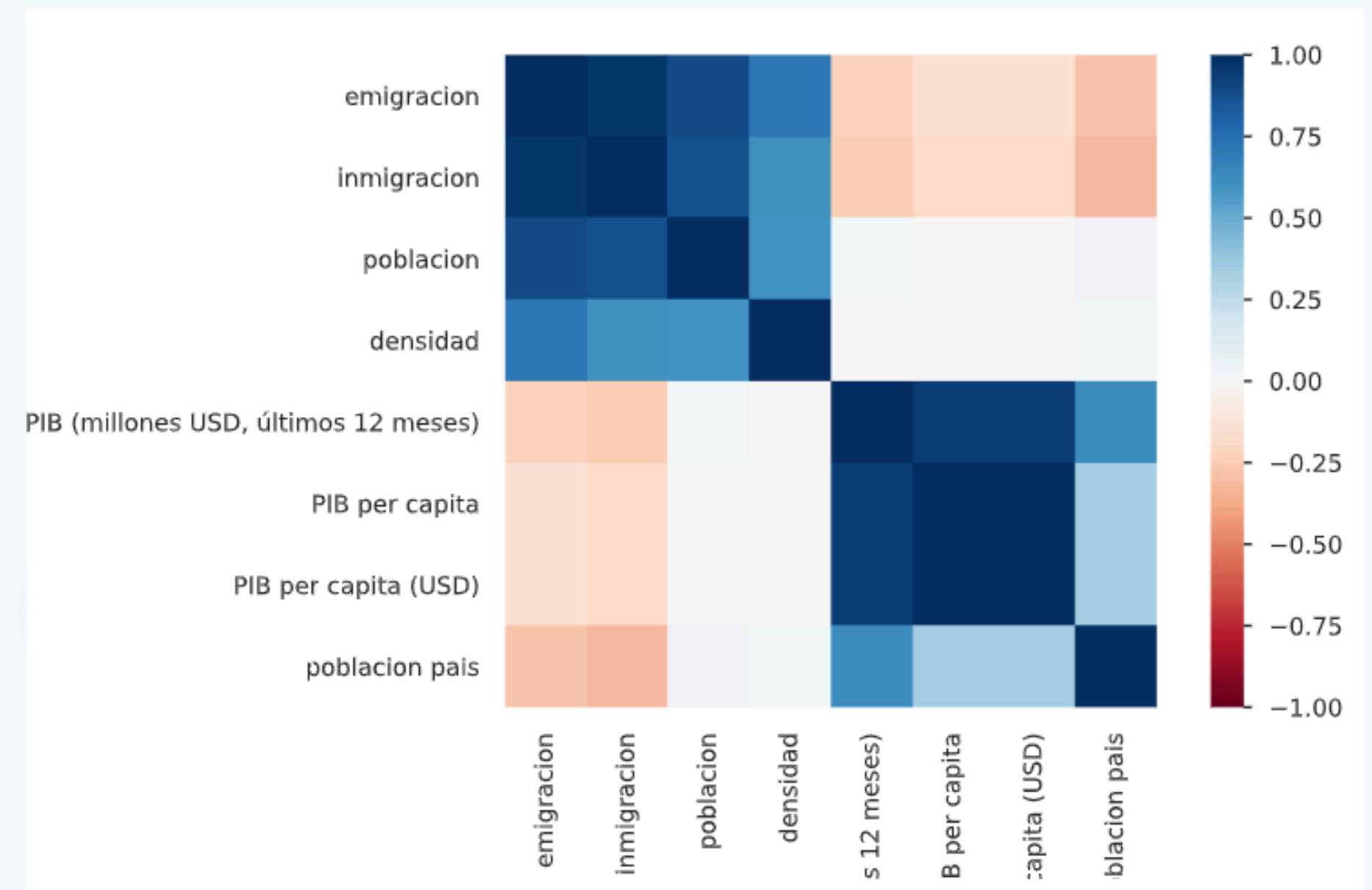
Z-Score + Yeo Johnson



ANÁLISIS EXPLORATORIO

Correlación

Se encontró alta correlación positiva entre 8 variables en total posterior



PCA

PCA Variables

Predicción Emigración

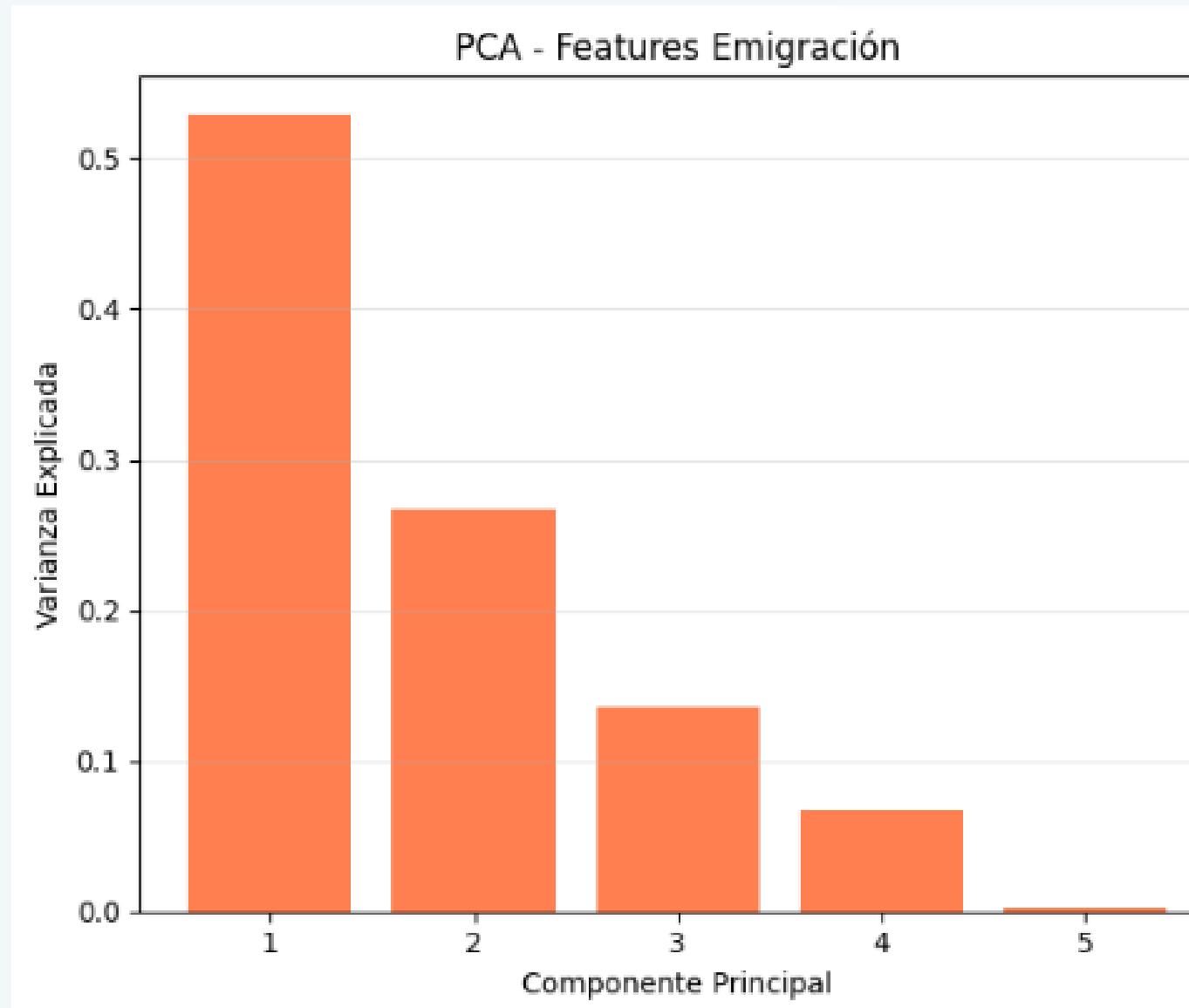
	PCA1	PCA2	PCA3
poblacion	0.013156	0.706689	-0.009169
densidad	0.008347	0.706463	-0.041199
PIB (millones USD, últimos 12 meses)	0.558231	-0.006467	0.055406
PIB per capita	0.540366	-0.018826	-0.296794
PIB per capita (USD)	0.539982	-0.018954	-0.300587
poblacion pais	0.323357	0.027286	0.903720

Predicción Inmigración

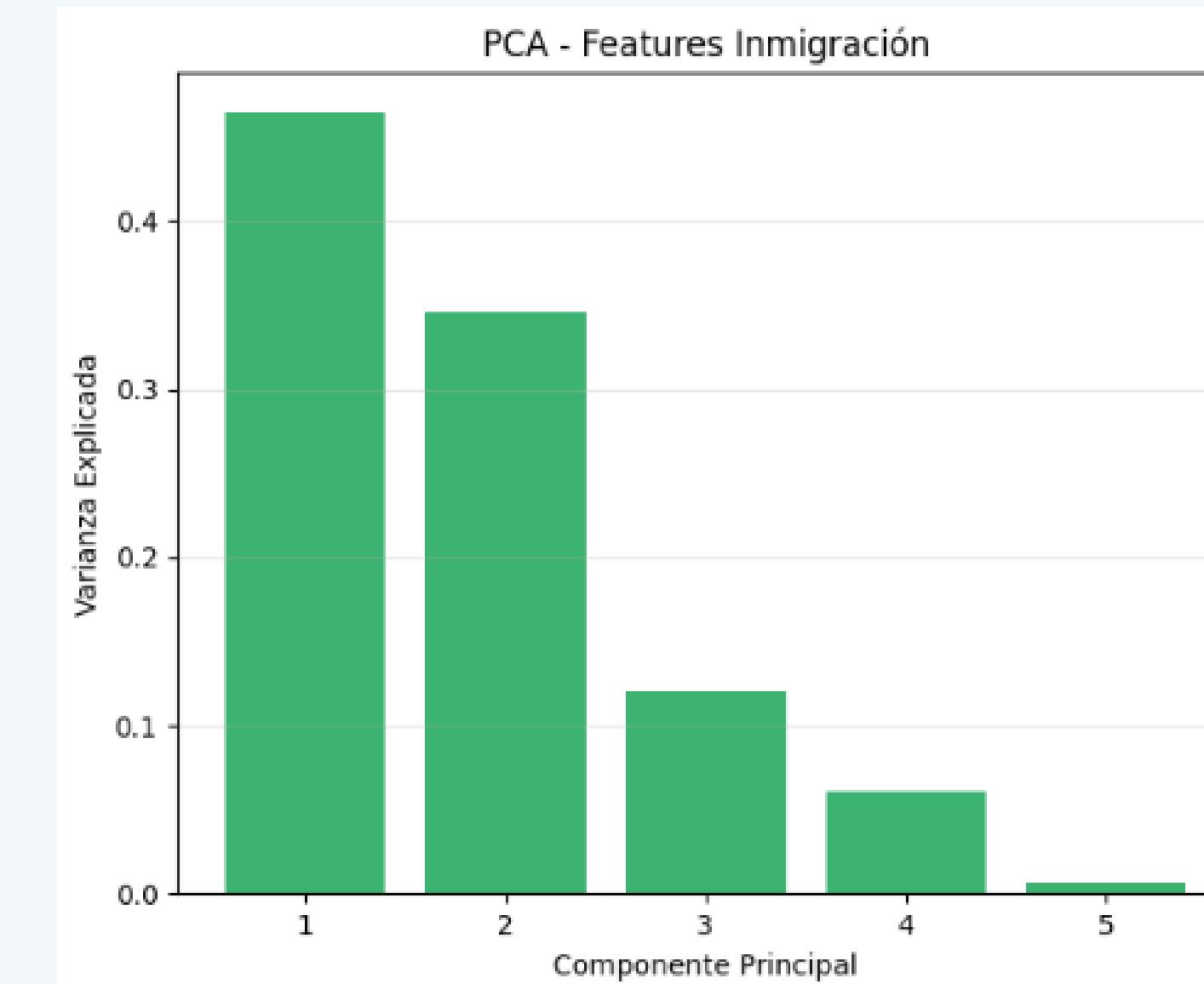
	PCA1	PCA2	PCA3	PCA4
emigracion	-0.229874	0.565683	-0.129019	-0.185763
poblacion	-0.105809	0.578563	0.100397	-0.552972
densidad	-0.096413	0.528739	0.122156	0.808940
PIB (millones USD, últimos 12 meses)	0.537990	0.138919	0.041574	-0.017731
PIB per capita	0.516116	0.149535	-0.302561	0.008555
PIB per capita (USD)	0.515678	0.149661	-0.306301	0.008733
poblacion pais	0.323936	0.040558	0.878215	-0.069763

PCA

Emigración

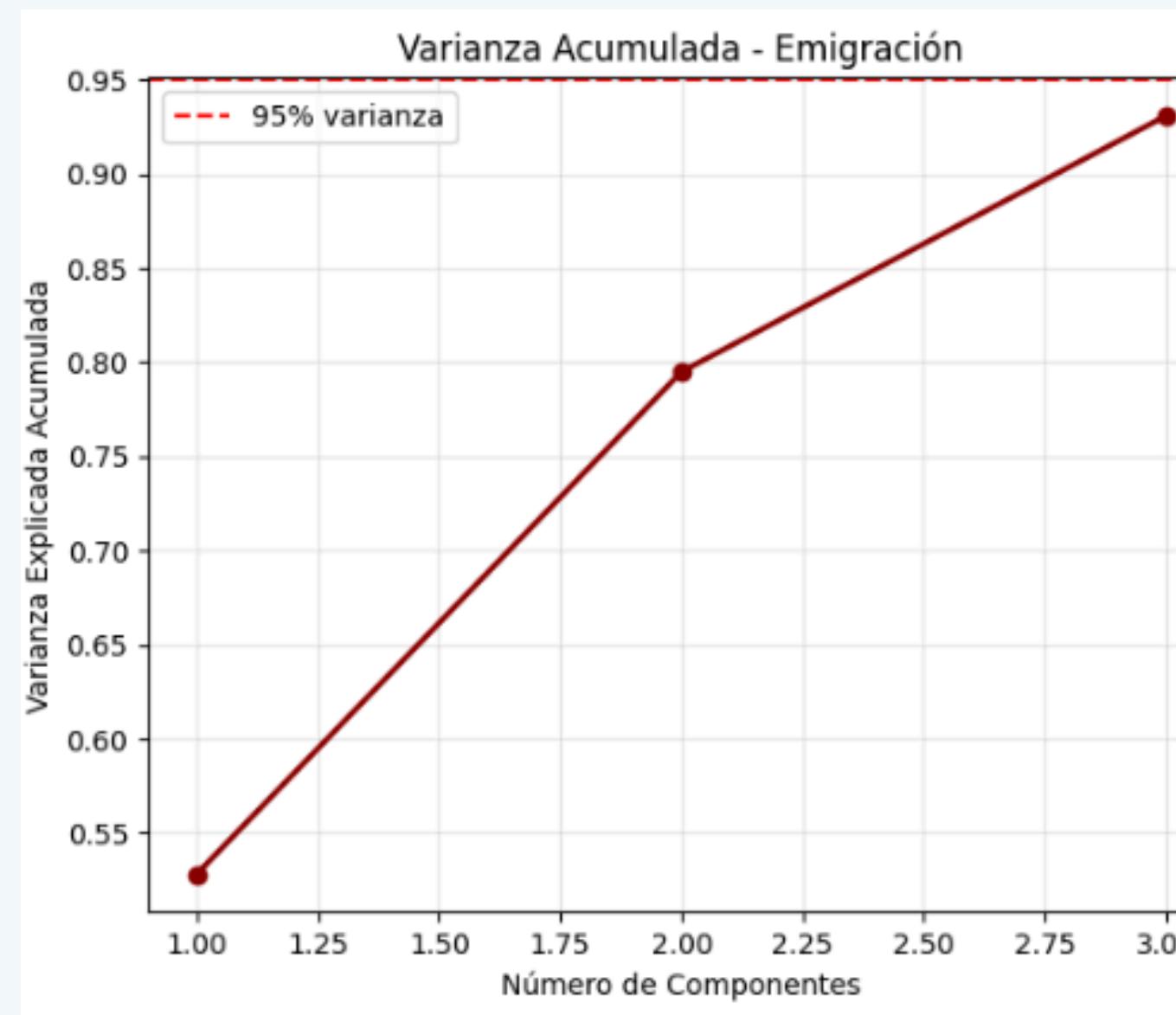


Inmigración

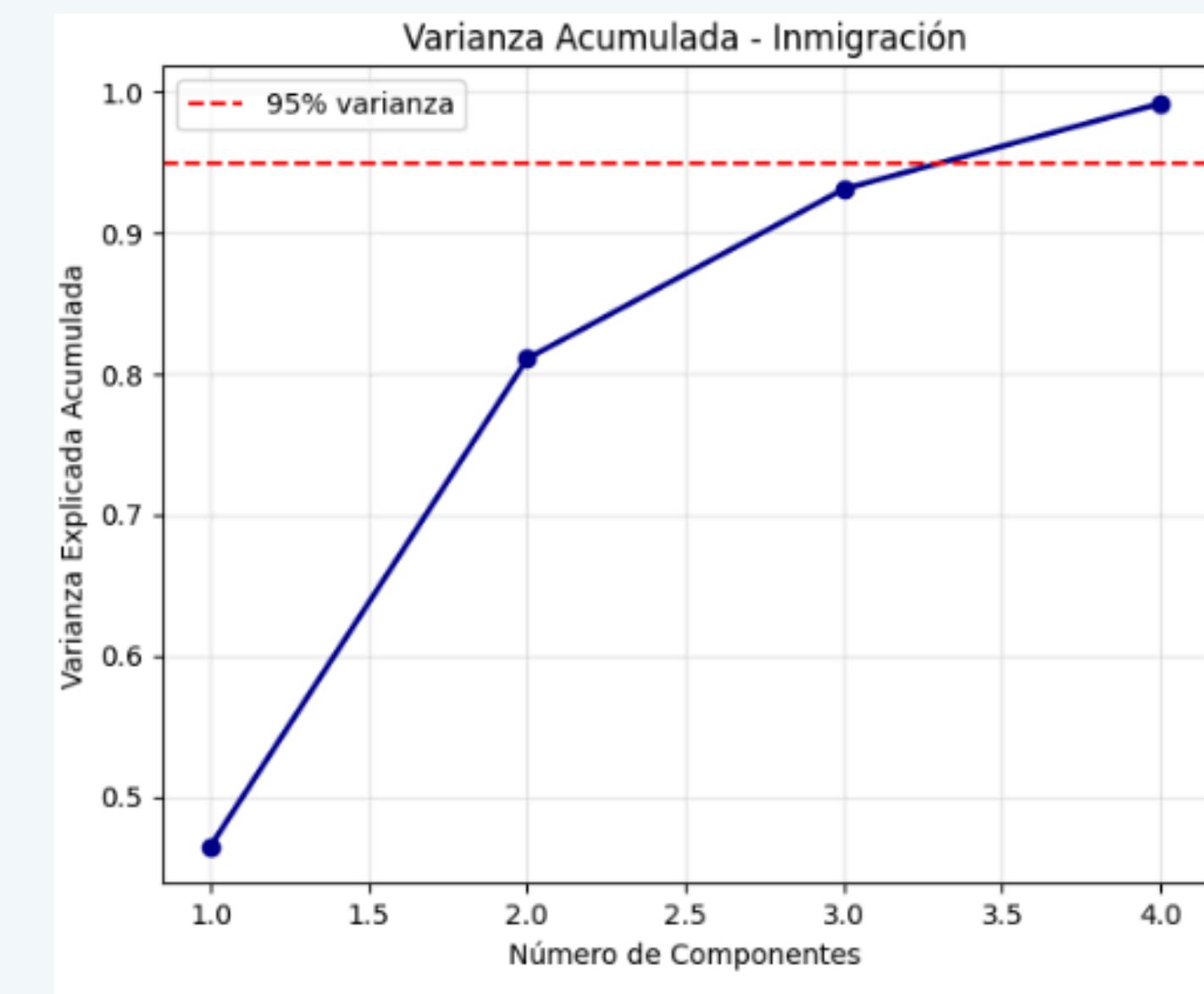


PCA

Emigración (Acumulada)



Inmigración (Acumulada)



PCA

Emigración

- El **Componente 1 es aún más dominante**, explicando por sí solo más del **52%** de la información.
- El **Componente 2** añade un **27%**.
- Para superar el umbral del **95% de información**, en este caso, se necesitarían los **3 componentes** del análisis.

Inmigración

- La distribución es muy similar a la del "Dataset Completo".
- El **Componente 1** explica cerca del **46%**** y el **Componente 2** un **35%**.
- Para alcanzar el **95%** de la información, se necesitan **3 componentes**.

MODELOS

Las predicciones se realizarán en dos fases contando con un modelo para cada una.

Predicción Emigración

Basado en:

- Poblacion
- Densidad
- PIB (millones USD, últimos 12 meses) *
- PIB per capita *
- PIB per capita (USD)*
- Poblacion pais

Predicción inmigración

Basado en:

- Emigracion
- Poblacion
- Densidad
- PIB (millones USD, últimos 12 meses)*
- PIB per capita*
- PIB per capita (USD)*
- Poblacion pais

MODELOS

Regresión Lineal

Se utilizó un 80% de la data para entrenamiento y un 20% de test.

Predictión Emigración

- MAPE = 195
- $R^2 = 0.9513$
- Asociación positiva: PIB per cápita, población
- Asociación negativa: población país, PIB per cápita

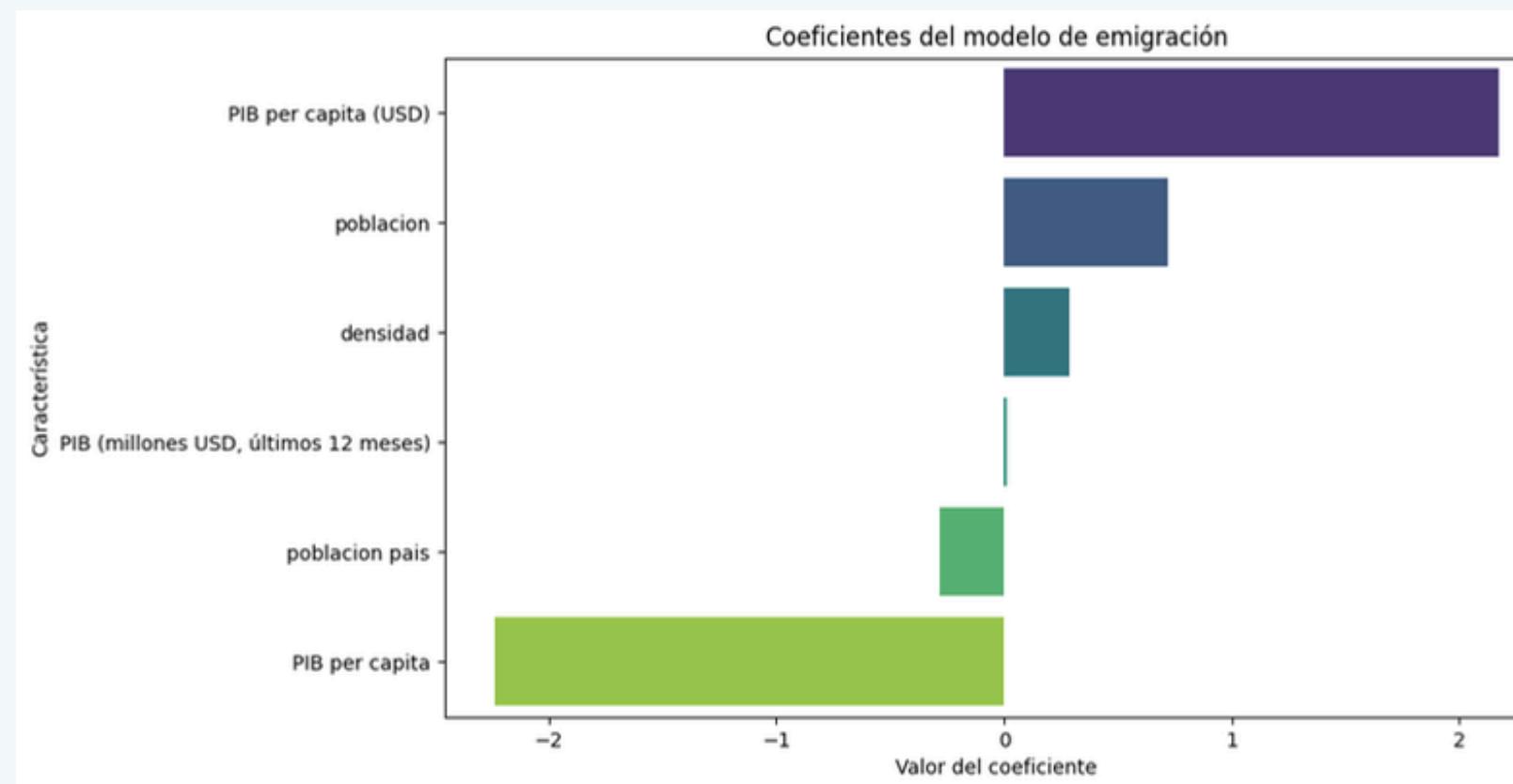
Predictión inmigración

- MAPE = 151
- $R^2 = 0.9011$
- Asociación positiva: población, PIB per cápita
- Asociación negativa: población país, PIB per cápita

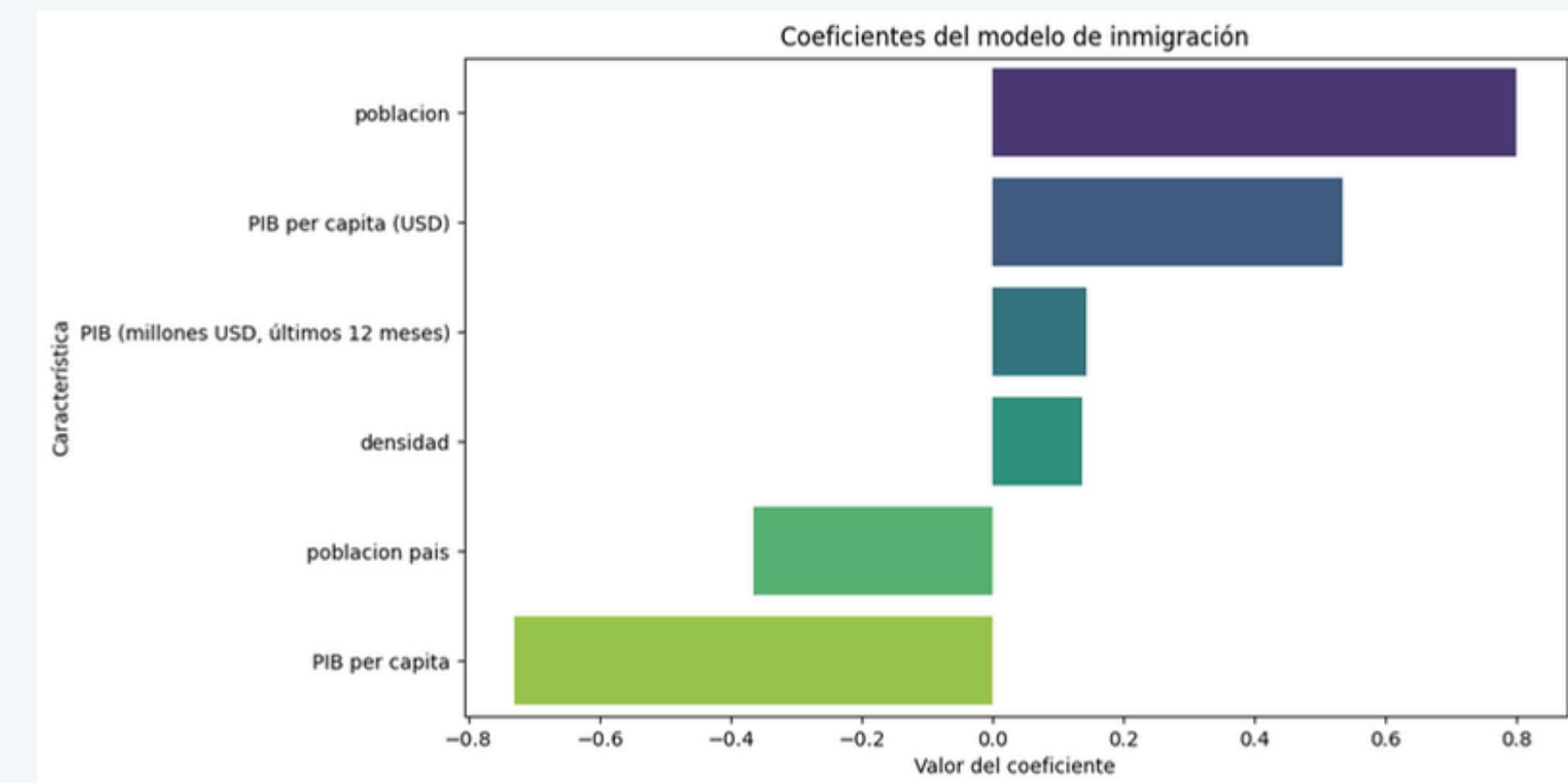
MODELOS

Regresión Lineal- Estudio coeficientes

Predicción Emigración



Predicción Inmigración



MODELOS

Gradient Boosting

Predicción Emigración

R2_Score: 0.9830

MAPE: 84,0041

RMSE: 0,1177

MAE: 0,0885

Predicción Inmigración

R2_Score: 0.9011

MAPE: 52,0342

RMSE: 0,1550

MAE: 0,1111

MODELOS

Random Forest

Predicción Emigración

R2_Score: 0,9533

MAPE: 143,6986

RMSE: 0,1952

MAE: 0,1493

Predicción Inmigración

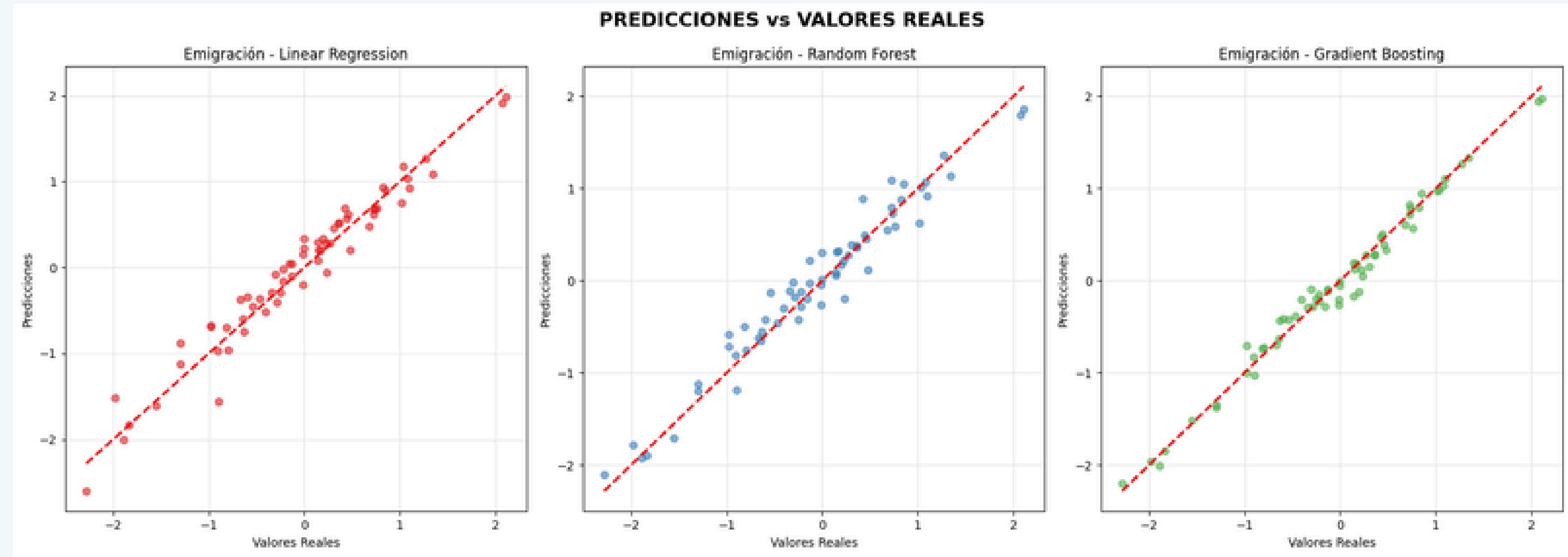
R2_Score: 0,8929

MAPE: 206,7810

RMSE: 0,2898

MAE: 0,2271

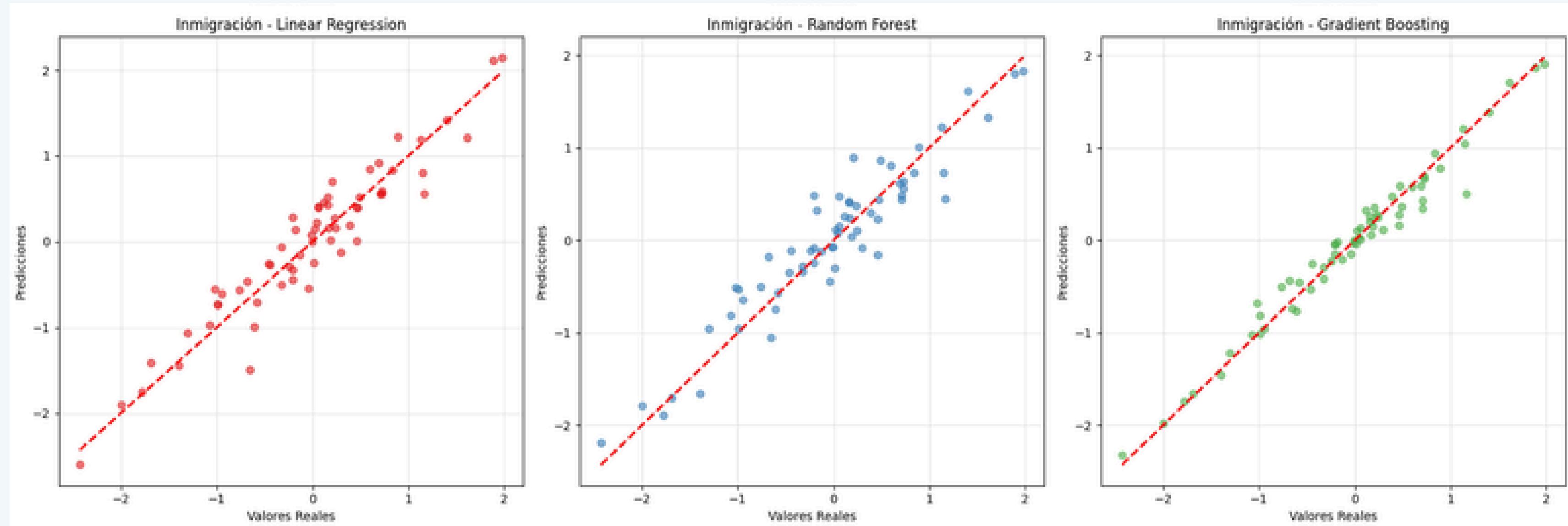
MODELOS (EMIGRACIÓN)



MODELOS (EMIGRACIÓN)

- **LR:** la nube sigue la diagonal pero con dispersión notable; errores en los extremos (valores muy negativos o muy positivos) son más visibles.
- **RF:** dispersión similar o algo mayor que LR en algunos puntos; se aprecia tendencia a subestimar algunos valores altos (puntos por debajo de la diagonal en el extremo derecho).
- **GB:** puntos muy alineados con la diagonal; errores pequeños y simétricos – explica por qué RMSE/MAE son claramente menores.

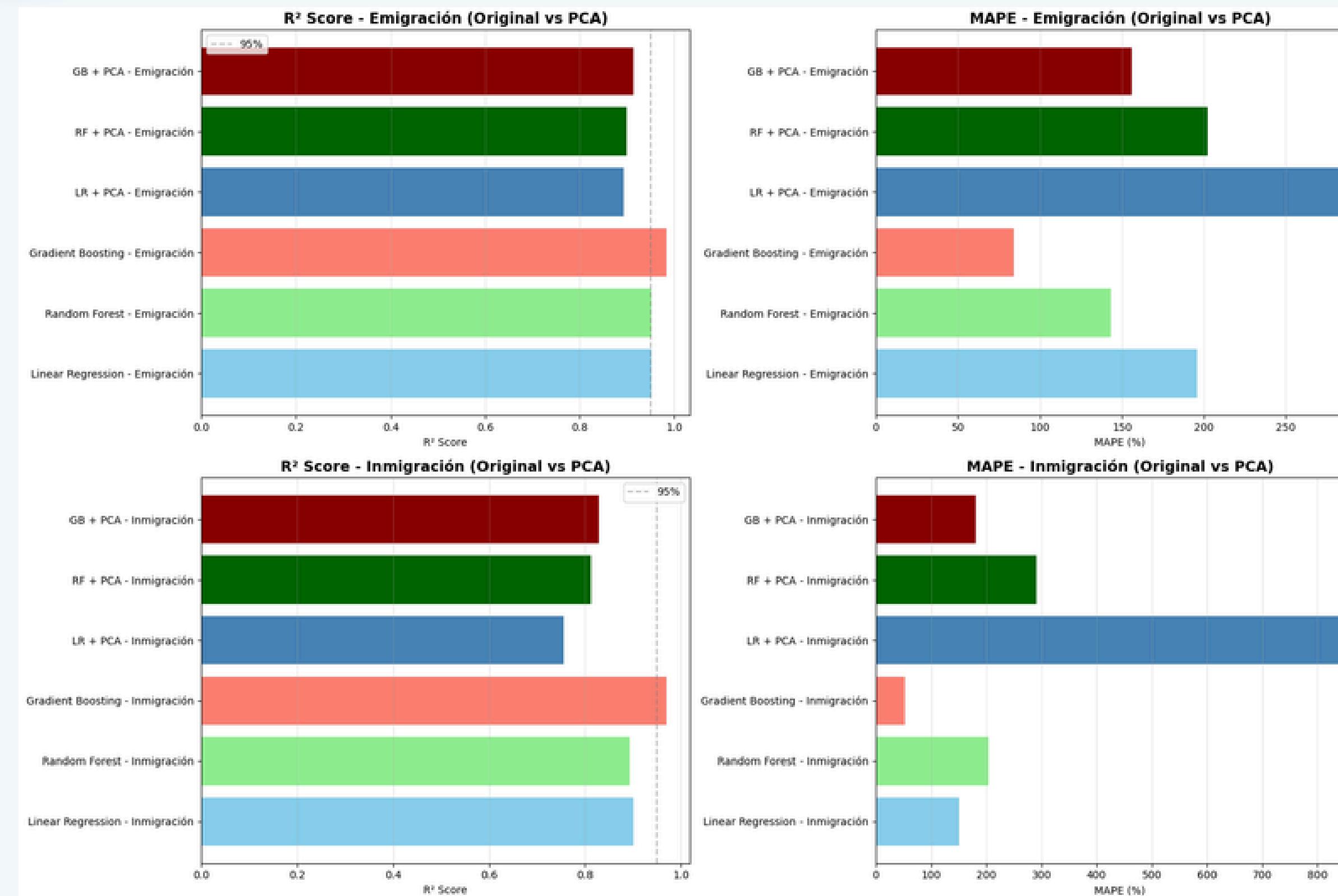
MODELOS (INMIGRACIÓN)



MODELOS (INMIGRACIÓN)

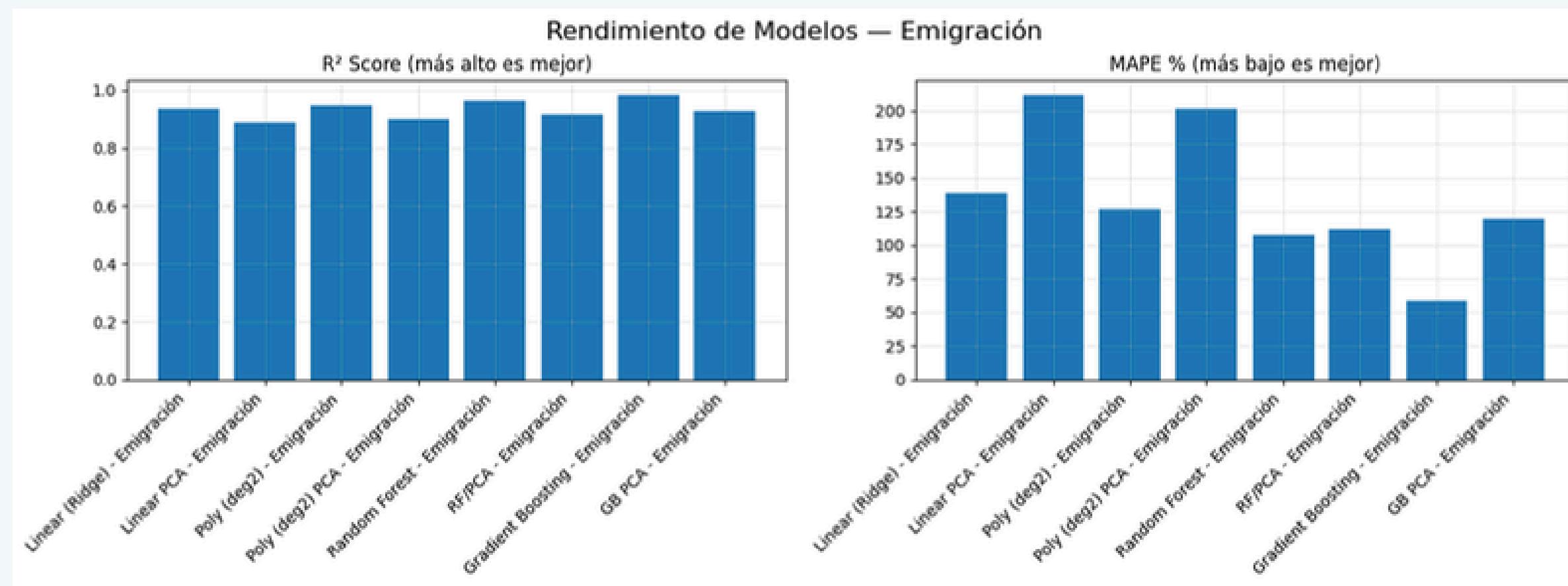
- **LR:** más dispersión que en emigración; varios puntos alejados en ambos extremos.
- **RF:** dispersión amplia y algunos puntos devueltos demasiado altos o bajos (menos consistencia que LR en este caso).
- **GB:** de nuevo el más consistente, aunque todavía hay algún outlier en el extremo positivo; en conjunto es el modelo más fiable.

MODELOS CATEGORIAS ORIGINALES VS PCA



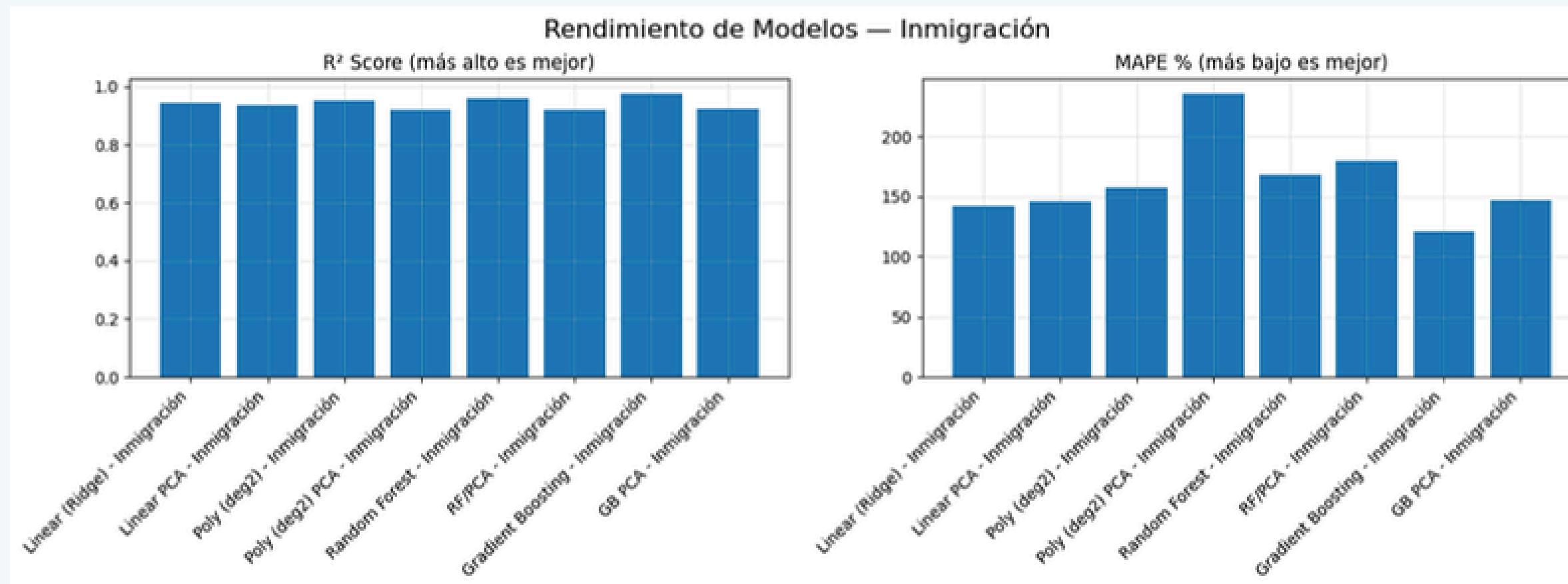
MODELO FINAL

- Métricas Emigración



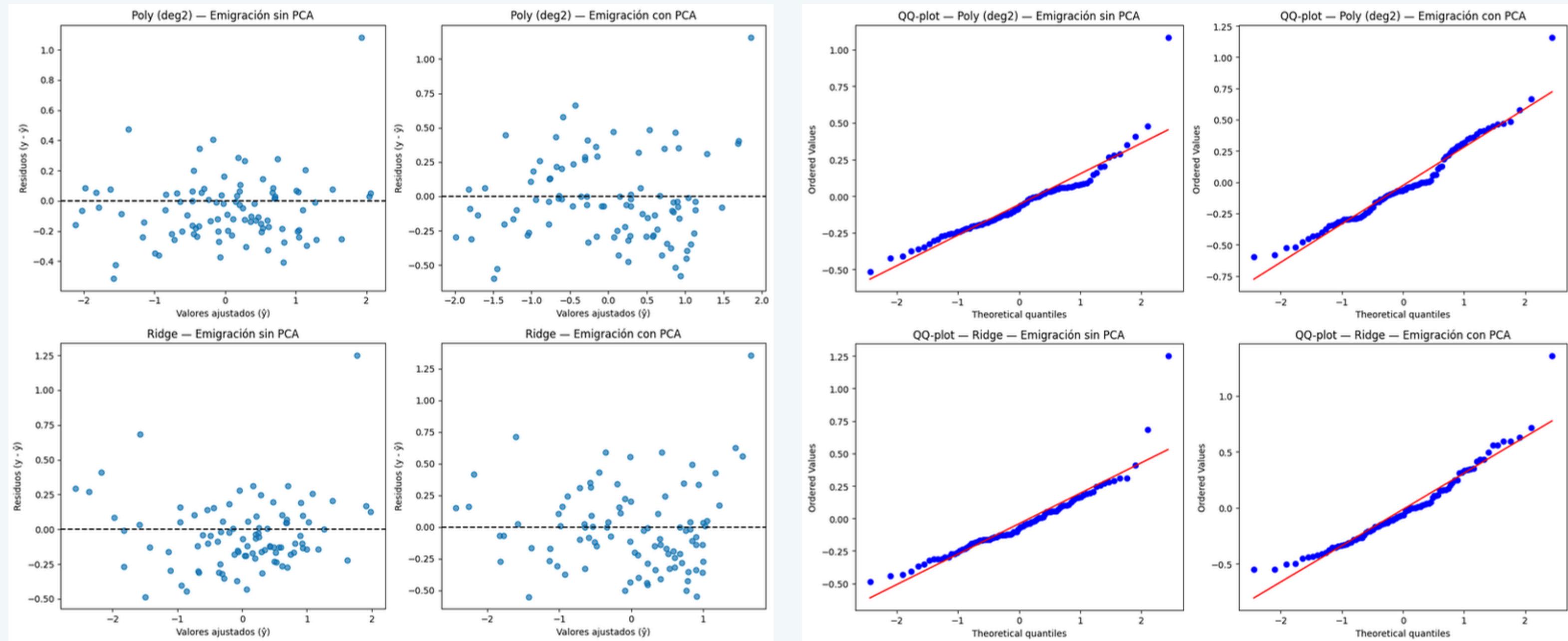
MODELO FINAL

- Métricas Inmigración



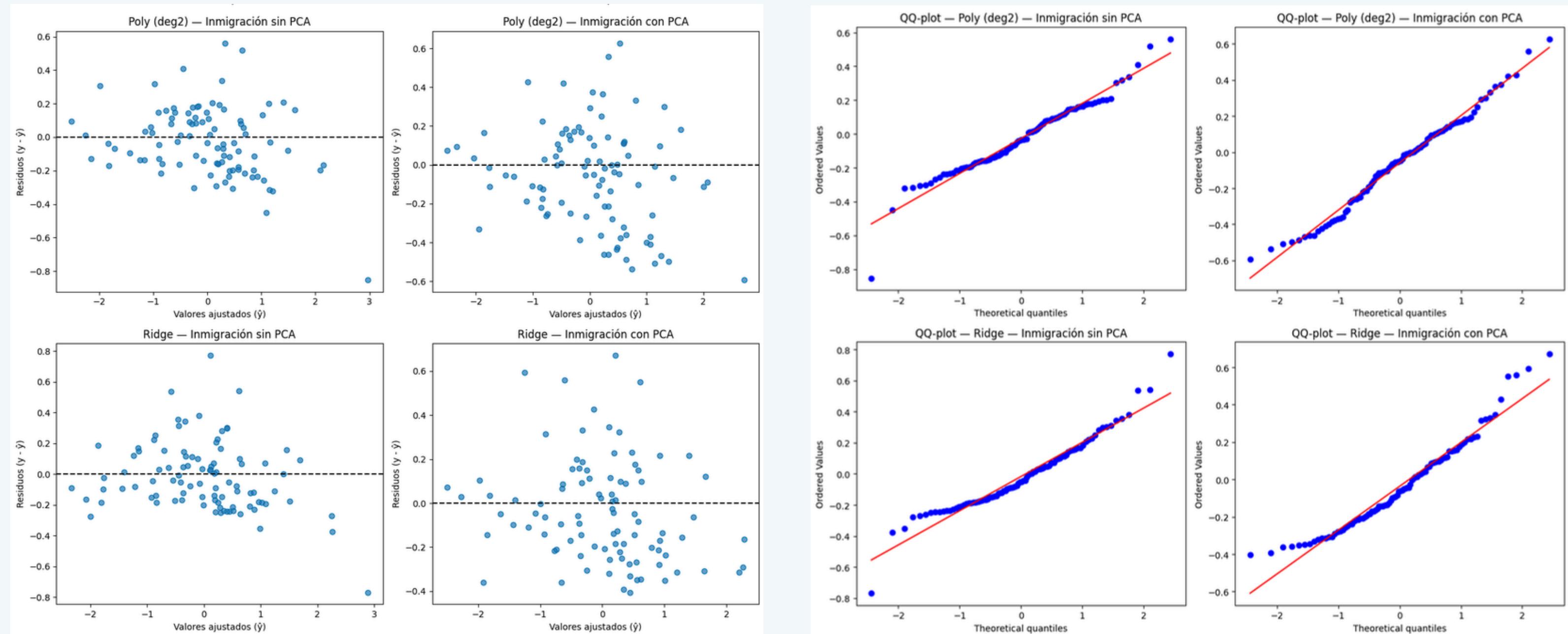
MODELO FINAL

- Residuos Regresion Emigracion



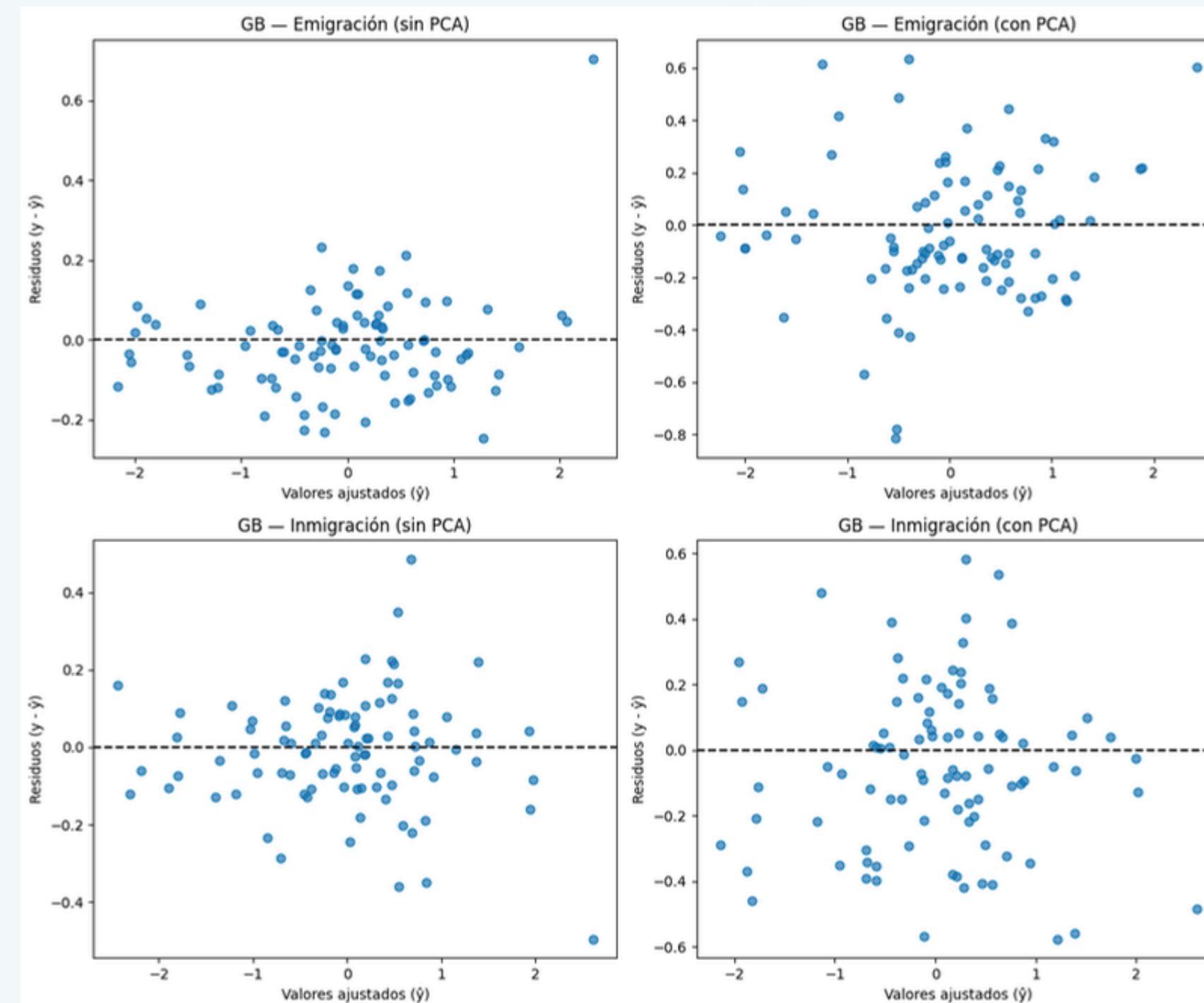
MODELO FINAL

- Residuos Regresion Inmigracion



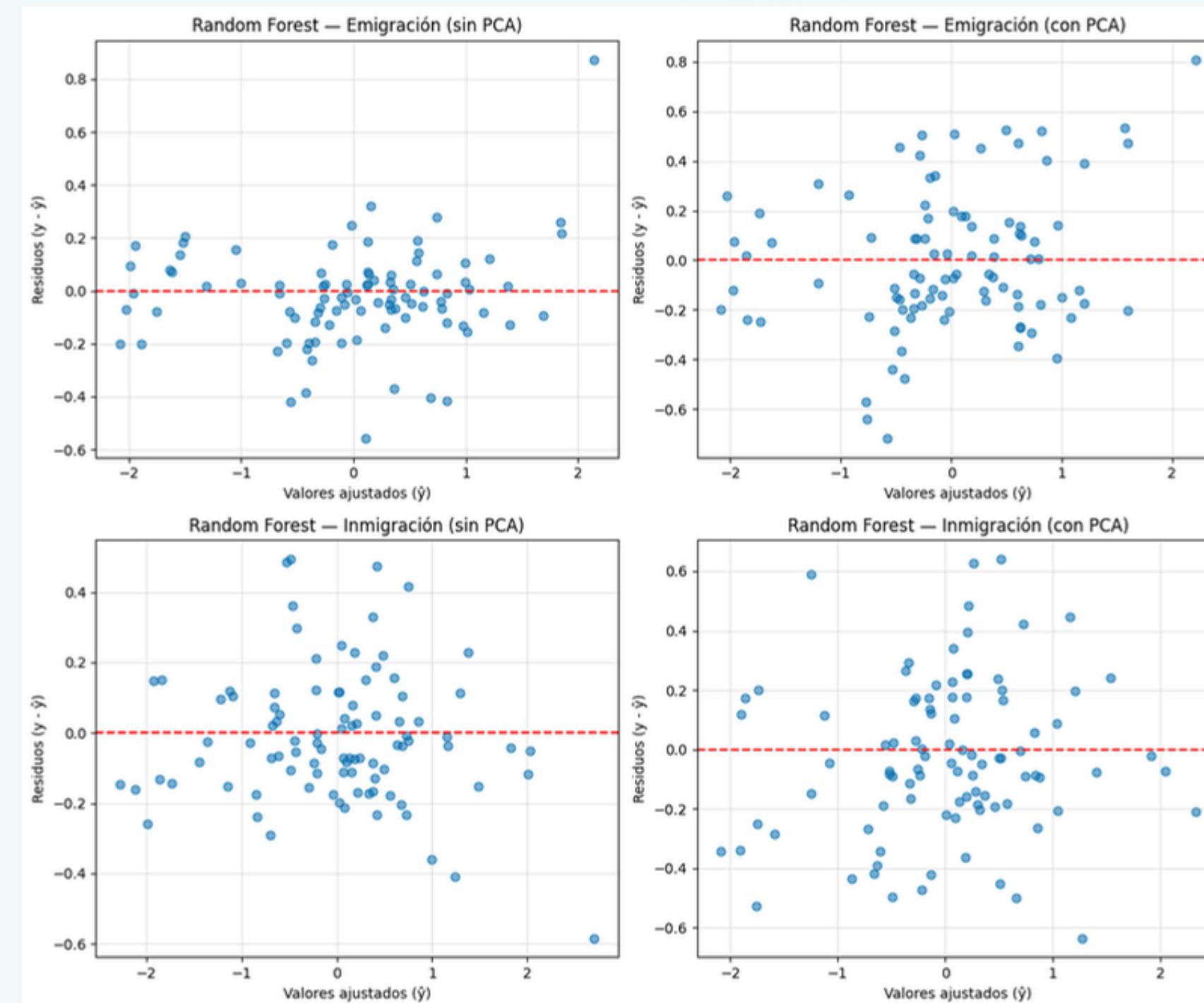
MODELO FINAL

- Residuos GB



MODELO FINAL

- Residuos Random Forest



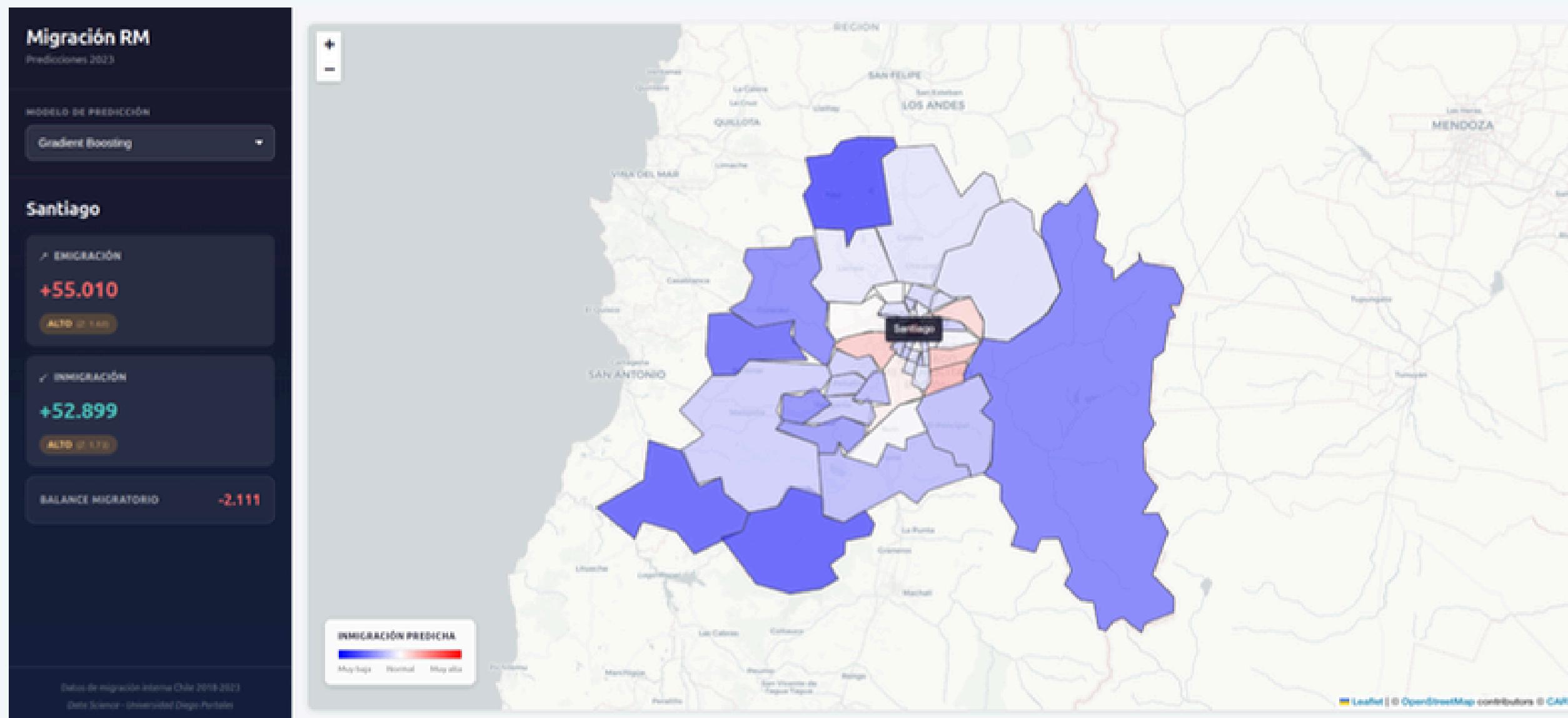
RESULTADOS FINALES

- Ejemplo de retorno:

comuna	modelo	emigracion_pred	inmigracion_pred
Santiago	gradient_boosting	55009.889714	52898.65588
Maipú	random_forest	40069.469275	34323.303891
Estación Central	gradient_boosting	26120.678217	22295.81766

RESULTADOS FINALES

- FrontEnd





GRACIAS