

Éxito en Google Play Store

Minoyetti
Casas Aguilar

El Desafío y la Oportunidad

Objetivo Principal (con Enfoque de Negocio): "Este proyecto se embarca en un análisis profundo de los datos de Google Play Store para:

Descubrir los factores clave que se correlacionan con la popularidad (instalaciones) y la satisfacción del usuario (ratings).

Construir modelos capaces de predecir el potencial de instalaciones de una app y de identificar características de apps con baja visibilidad (sin ratings).

Segmentar el mercado de aplicaciones para identificar arquetipos y nichos.

Proveer insights accionables para optimizar estrategias de desarrollo, monetización y marketing."

DataSet

- 24 Columnas
- Más de 2 millones de registros

Variables Protagonistas:

- Métricas de Rendimiento: **Maximum Installs** (nuestro Santo Grial para popularidad), **Rating, Rating Count**.
- Características de la App: **Category_General** (nuestra agrupación estratégica), **Size** (y su versión _log), **Min_Age** (derivada de Content Rating).
- Modelo de Negocio: **Price, In App Purchases, Ad Supported**.
- Ciclo de Vida: **Active_Life_log, Inactive_Time_log** (nuestras métricas de "frescura" y mantenimiento).

Nuestra Metodología Analítica

Exploración y Limpieza Profunda

Entendimiento inicial, tratamiento de faltantes, transformación de datos (ej. `Size` a numérico, `Price` a float).

Ingeniería de Características Estratégica

Creación de variables con mayor poder informativo (`Category_General`, `Active_Life_log`, `Min_Age`, transformaciones logarítmicas).

Segmentación del Mercado

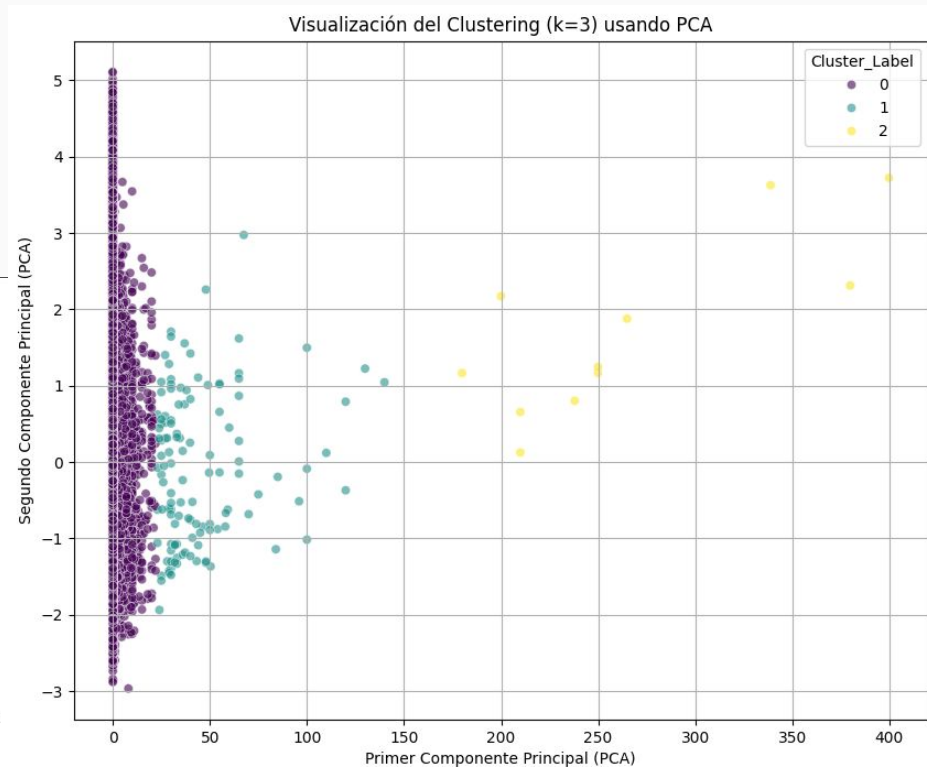
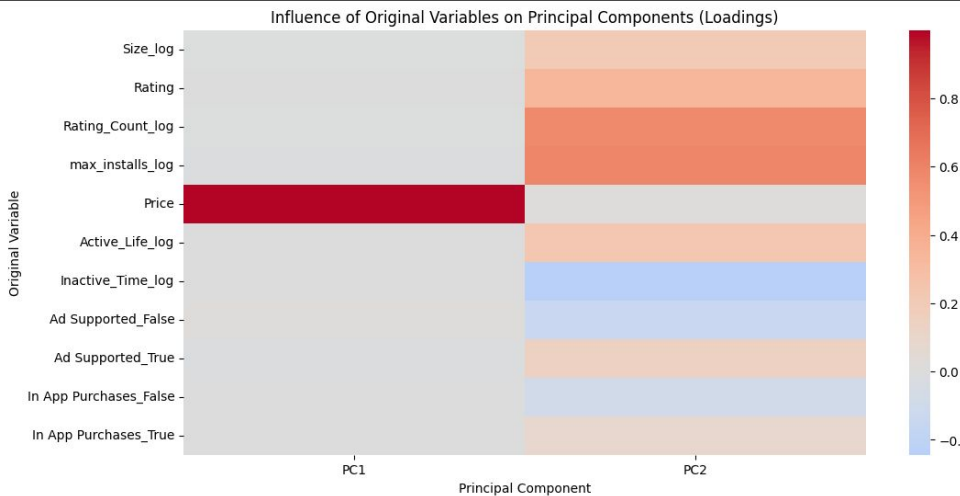
Uso de K-Means para descubrir grupos naturales de aplicaciones.

Modelado Predictivo

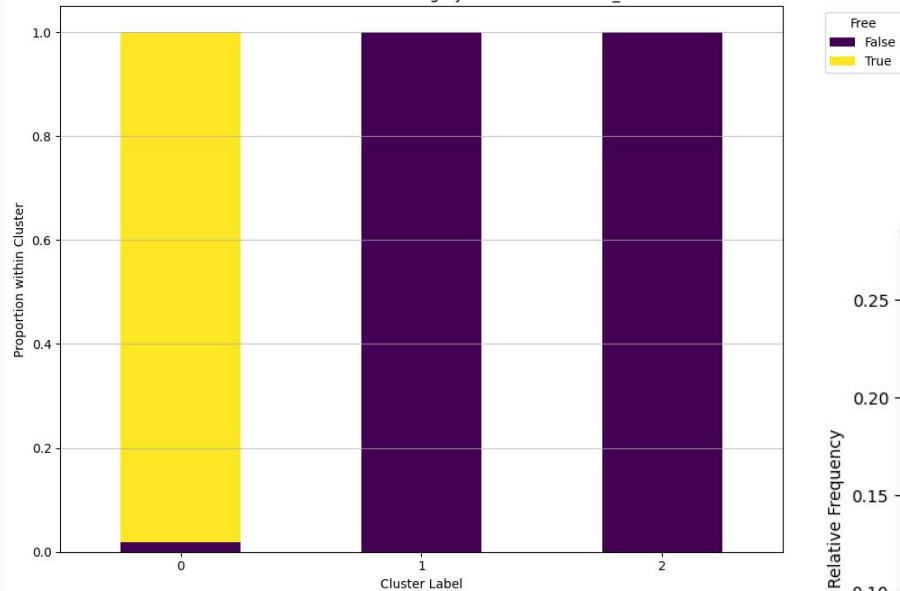
Predicción de `max_installs_log` con RandomForest y XGBoost.
Predicción de `is_No_Ratings` con Regresión Logística.

Extracción de Insights.

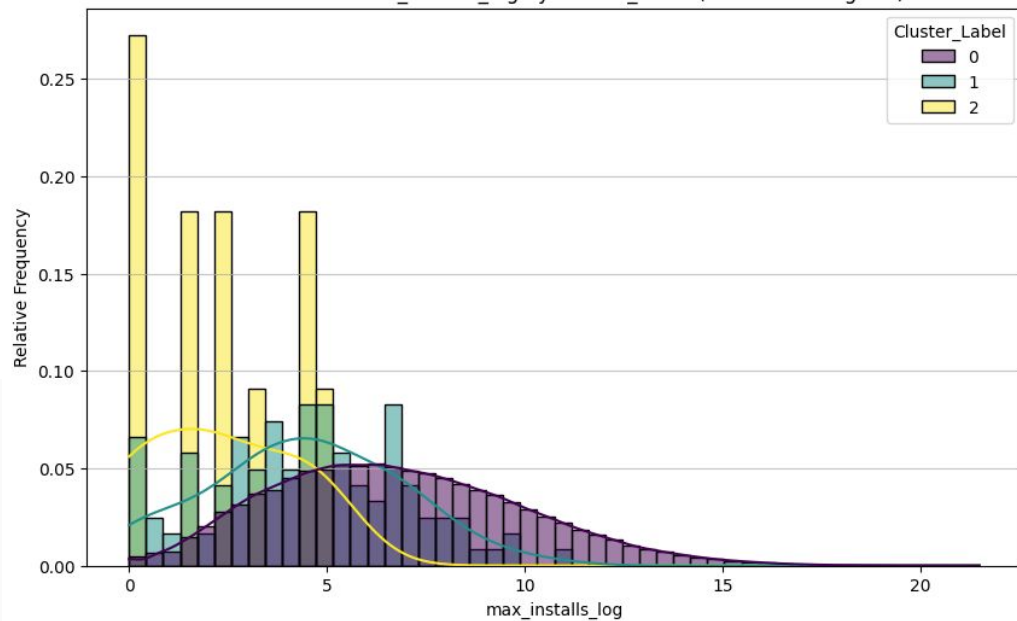
Insights del Clustering (K-Means)



Relative Distribution of Free Category within each Cluster_Label



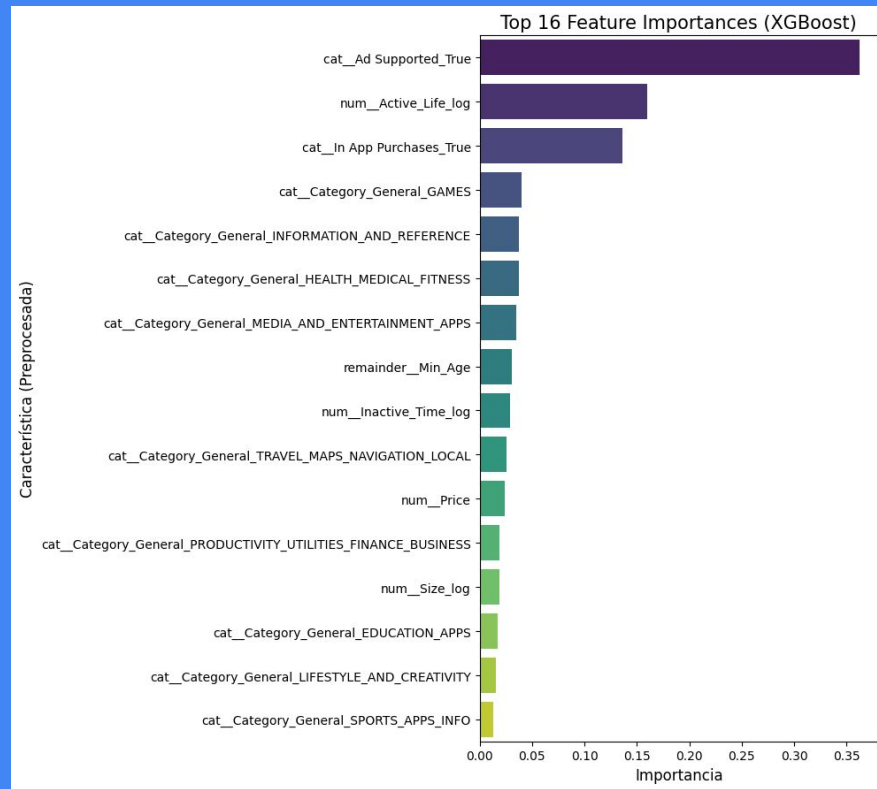
Distribution of max_installs_log by Cluster_Label (Relative Histogram)



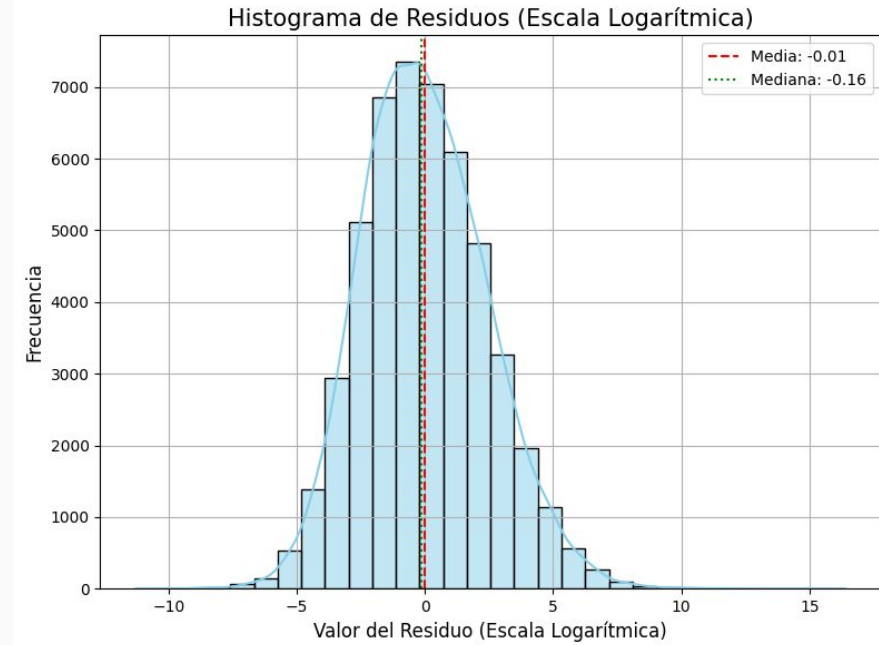
Modelos Predictivos

XGBOOST con target transformado

Se busca predecir el logaritmo
de 'Maximum Installs'

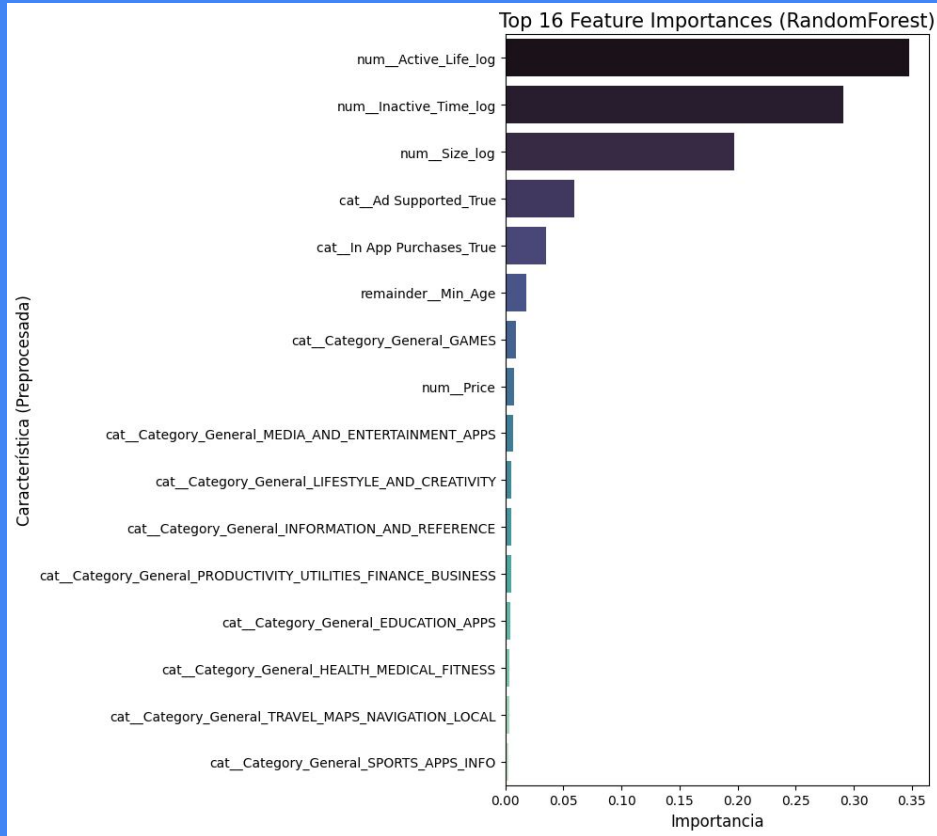



```
--- Métricas de Rendimiento del Modelo (XGBoost) ---  
Métricas en Escala Logarítmica (Target: max_installs_log):  
MAE (log): 1.96  
MSE (log): 5.99  
RMSE (log): 2.45  
R2 (log) : 0.4016
```

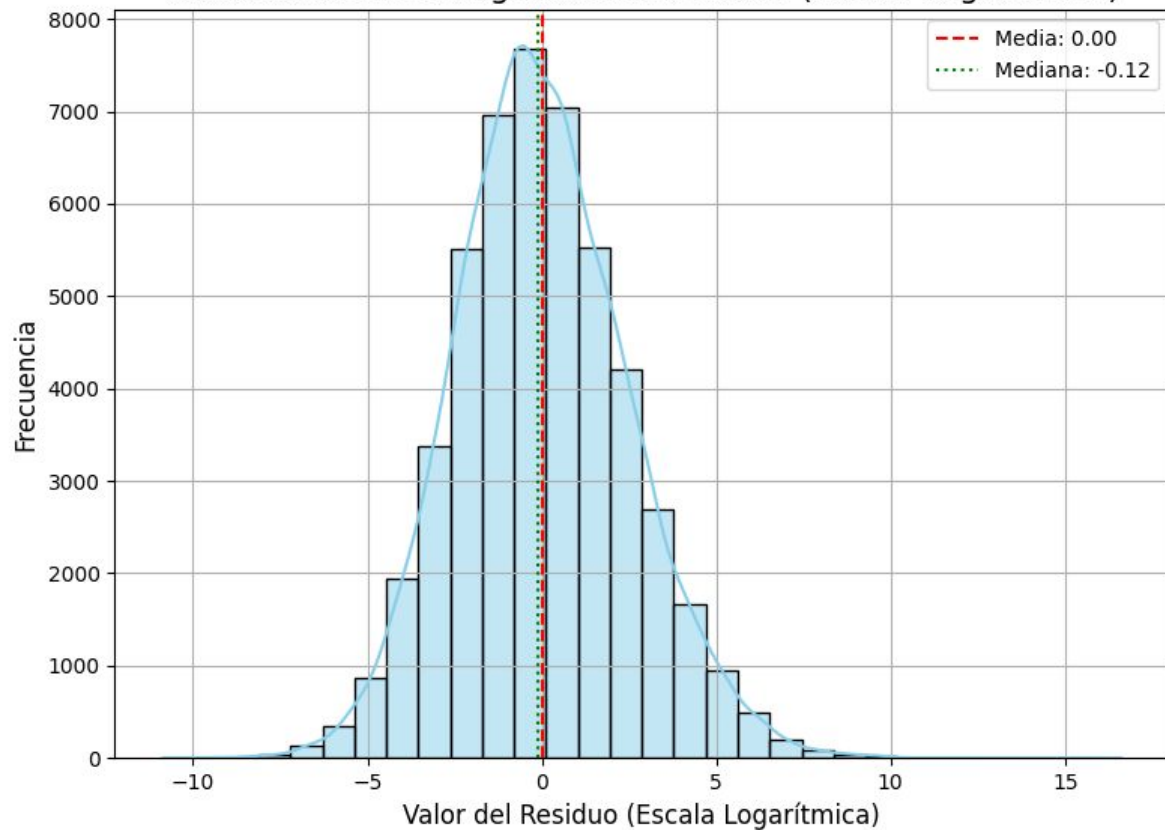


RandomForest con target transformado

Se busca predecir el logaritmo
de 'Maximum Installs'



RandomForest: Histograma de Residuos (Escala Logarítmica)



--- Métricas de Rendimiento del Modelo (RandomForestRegressor) ---

Métricas en Escala Logarítmica (Target: max_installs_log):

MAE (log): 1.9547

MSE (log): 6.1340

RMSE (log): 2.4767

R² (log) : 0.3871

Métricas en Escala Original (Target: Maximum Installs):

MAE (original): 282,047

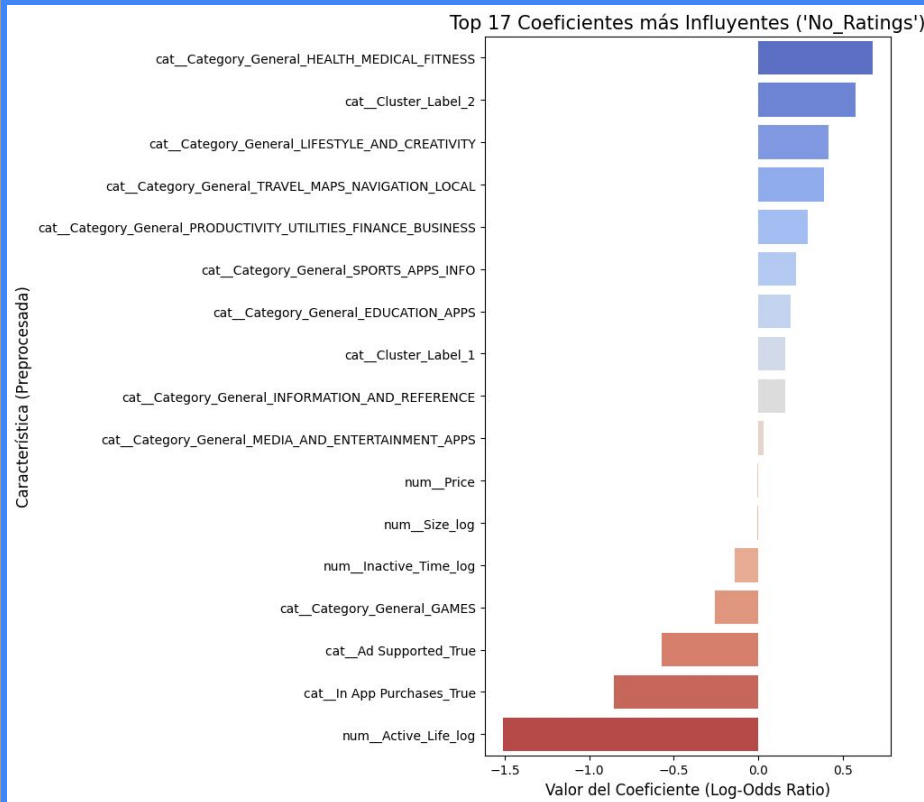
MSE (original): 144,496,531,790,899

RMSE (original): 12,020,671

R² (original) : 0.0008

Regresión Logística

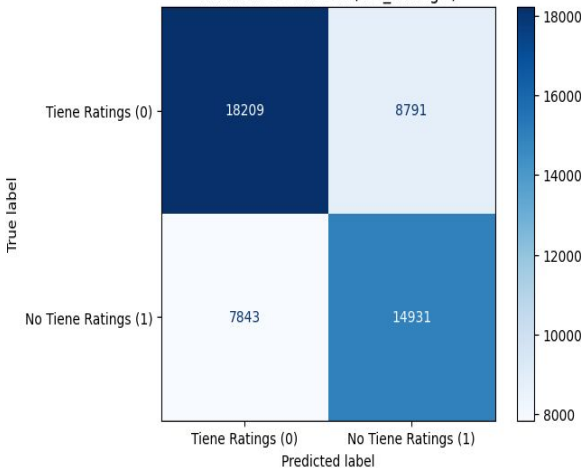
Busca predecir si habrá ratings o no.



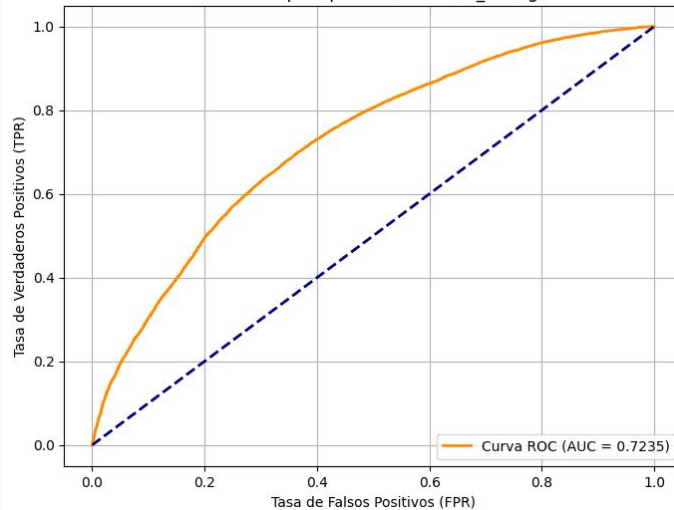
Reporte de Clasificación:

	precision	recall	f1-score	support
Tiene Ratings (0)	0.70	0.67	0.69	27000
No Tiene Ratings (1)	0.63	0.66	0.64	22774
accuracy			0.67	49774
macro avg	0.66	0.67	0.66	49774
weighted avg	0.67	0.67	0.67	49774

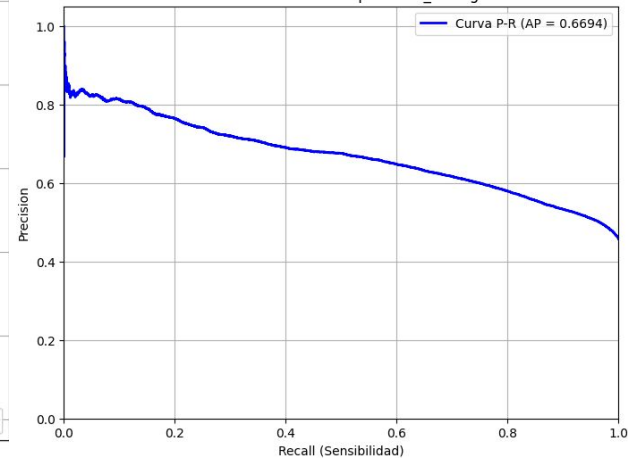
Matriz de Confusión ('No_Ratings')



Curva ROC para predicción de "No_Ratings"



Curva Precision-Recall para "No_Ratings"



Conclusiones

La Calidad es Reina (y Rey): "Un alto Rating no solo refleja satisfacción, sino que es un predictor e impulsor clave de la popularidad (Maximum Installs, Rating Count). Recomendación: Priorizar la experiencia del usuario y la calidad de la app es fundamental."

La Monetización Depende del Contexto: "Las apps gratuitas con anuncios dominan en volumen de instalaciones, pero ciertas categorías (PRODUCTIVITY_UTILITIES_FINANCE_BUSINESS) muestran mayor tolerancia a modelos de pago. Recomendación: Adaptar la estrategia de monetización a la categoría y al valor percibido por el usuario."

Visibilidad y Engagement Temprano Son Cruciales: "Las características de las apps sin ratings sugieren que la falta de visibilidad inicial o un bajo engagement temprano son barreras significativas. Recomendación: Implementar estrategias de ASO (App Store Optimization) y promoción inicial para fomentar las primeras valoraciones."

El Mantenimiento Activo Impulsa la Longevidad: "Las aplicaciones actualizadas y con un ciclo de vida activo (Active_Life_log) tienden a mantener ratings más altos. Recomendación: Planificar el mantenimiento y la evolución continua de la app como parte de la estrategia a largo plazo."

Conoce tu Nicho (Insights del Clustering): "Existen diferentes 'arquetipos' de apps con distintos perfiles de éxito. Recomendación: Entender en qué segmento compite tu app puede ayudar a definir estrategias de diferenciación y expectativas realistas."