

A photograph showing a person's hands working on a financial calculator. One hand is pressing buttons on the calculator, while the other hand holds a pen over a document with various charts and graphs. The background is dark, and the overall atmosphere suggests a professional or academic setting.

DATA SCIENCE PROJECT

CREDIT SCORING PREDICTION

MACHINE LEARNING
by Kelly Viana

Problema de negocio

En el dinámico mundo financiero, la gestión efectiva del riesgo crediticio es crucial para garantizar la estabilidad y el éxito de las instituciones financieras.

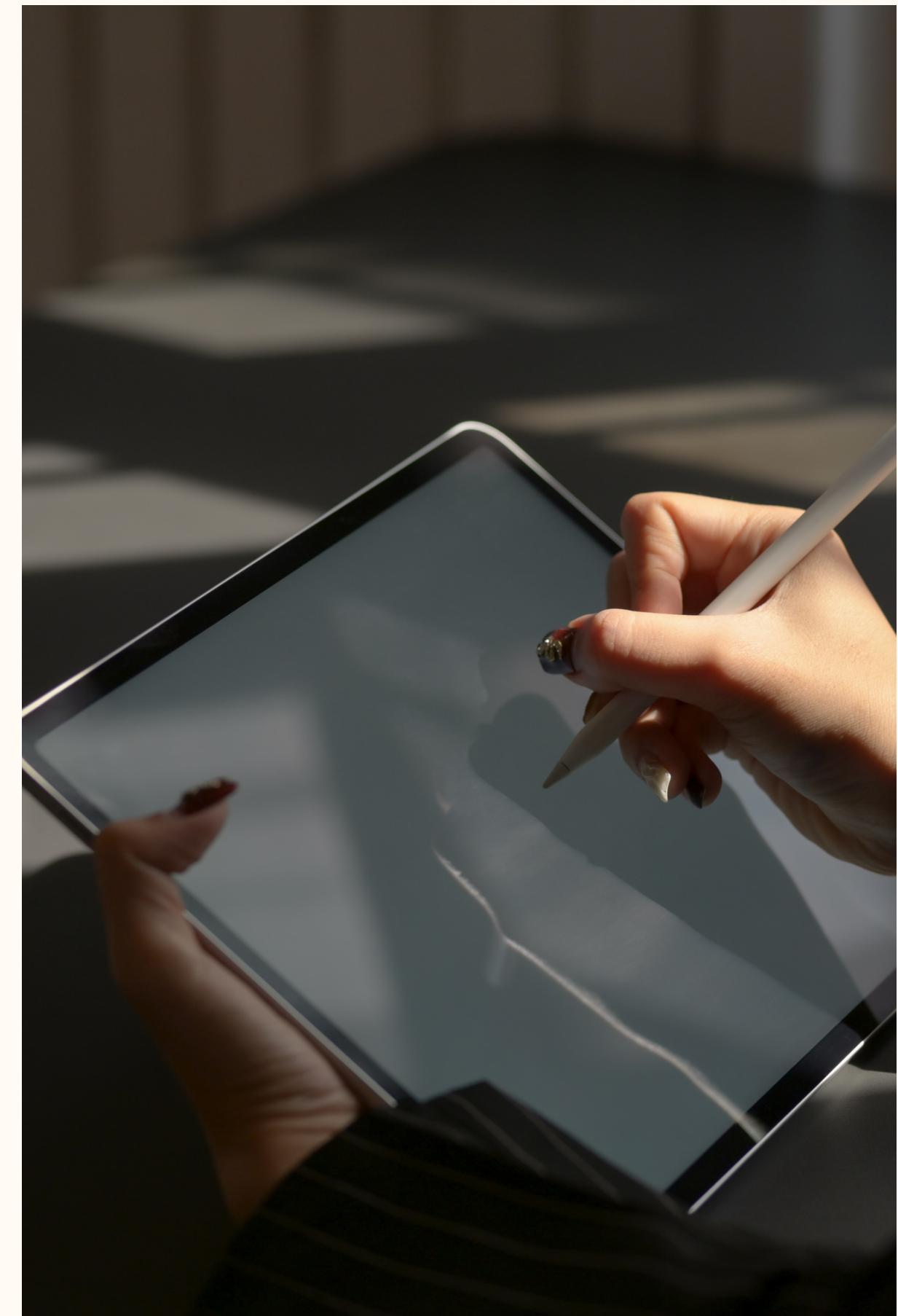
En este contexto, una institución financiera alemana ha reconocido la necesidad imperante de elevar los estándares de evaluación de solvencia crediticia.



Optimización de Modelos Predictivos para la Evaluación de Solvencia Crediticia mediante Análisis de Datos y Machine Learning

En este proyecto, el enfoque principal será el manejo eficiente de datos financieros para la evaluación de solvencia crediticia. Desde la limpieza y preprocesamiento de datos hasta la construcción y evaluación de modelos de machine learning, el objetivo es identificar patrones clave y relaciones en el conjunto de datos para desarrollar un modelo predictivo robusto. Se utilizarán técnicas avanzadas, como la Regresión Logística, Árboles de Decisión, Random Forest y Naive Bayes, para mejorar la precisión y confiabilidad en la predicción de la solvencia crediticia.

La exploración detallada de datos, junto con la selección cuidadosa de características, permitirá un análisis profundo y proporcionará una base sólida para la toma de decisiones crediticias. La evaluación de modelos se llevará a cabo mediante métricas clave como precisión, recall, área bajo la curva ROC y F1-score, asegurando la elección del modelo más efectivo para el contexto de evaluación de solvencia crediticia.

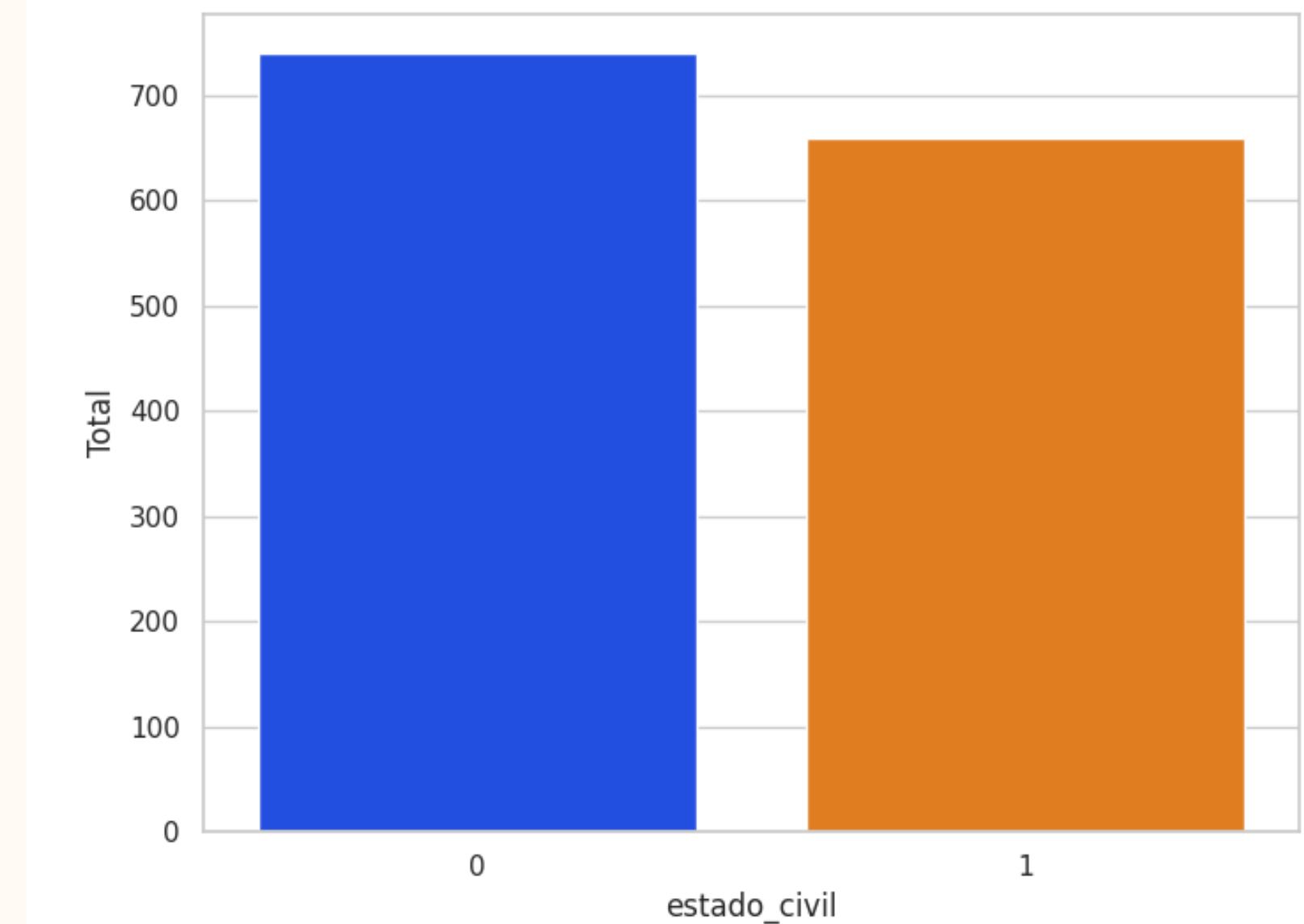
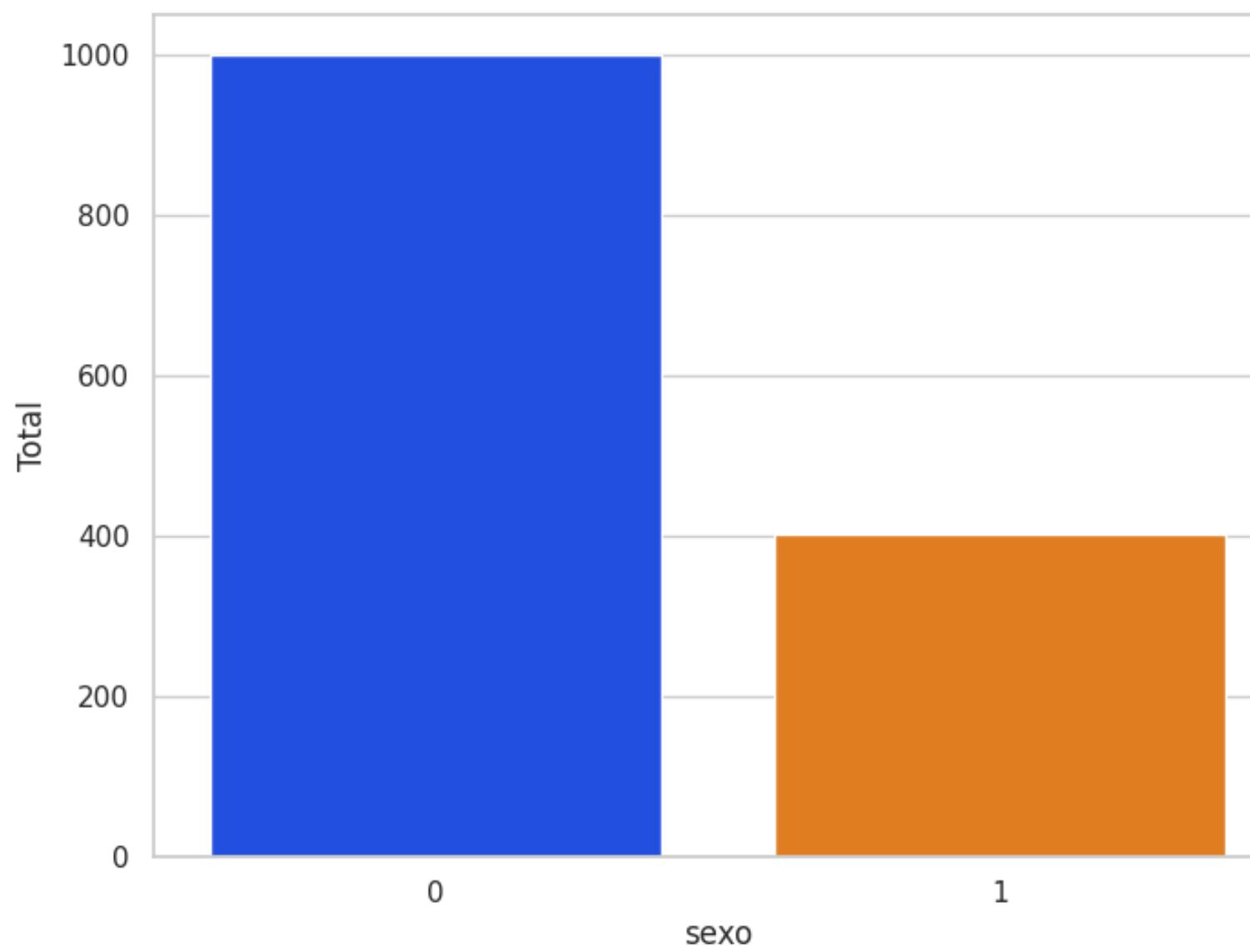


Exploración de datos

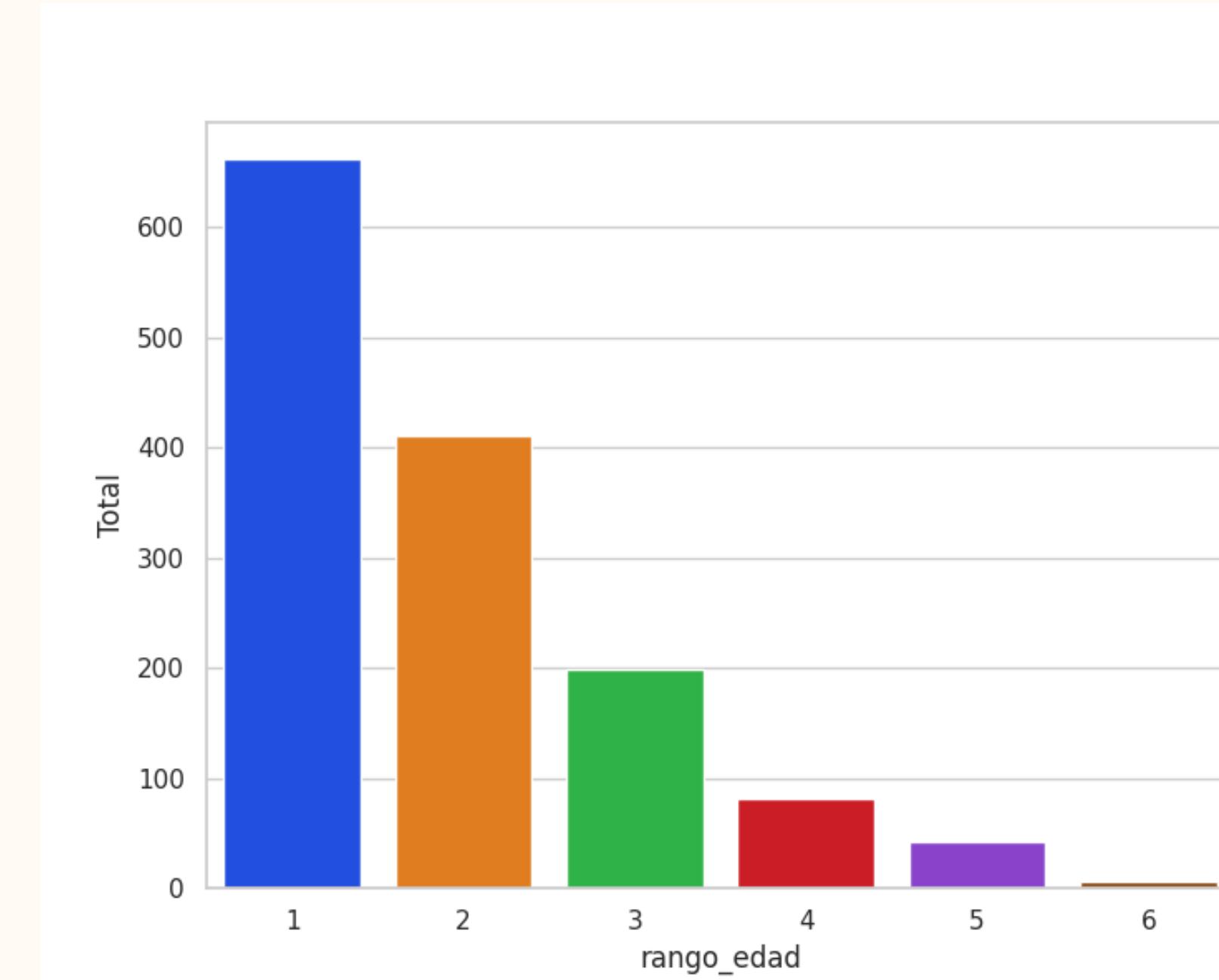
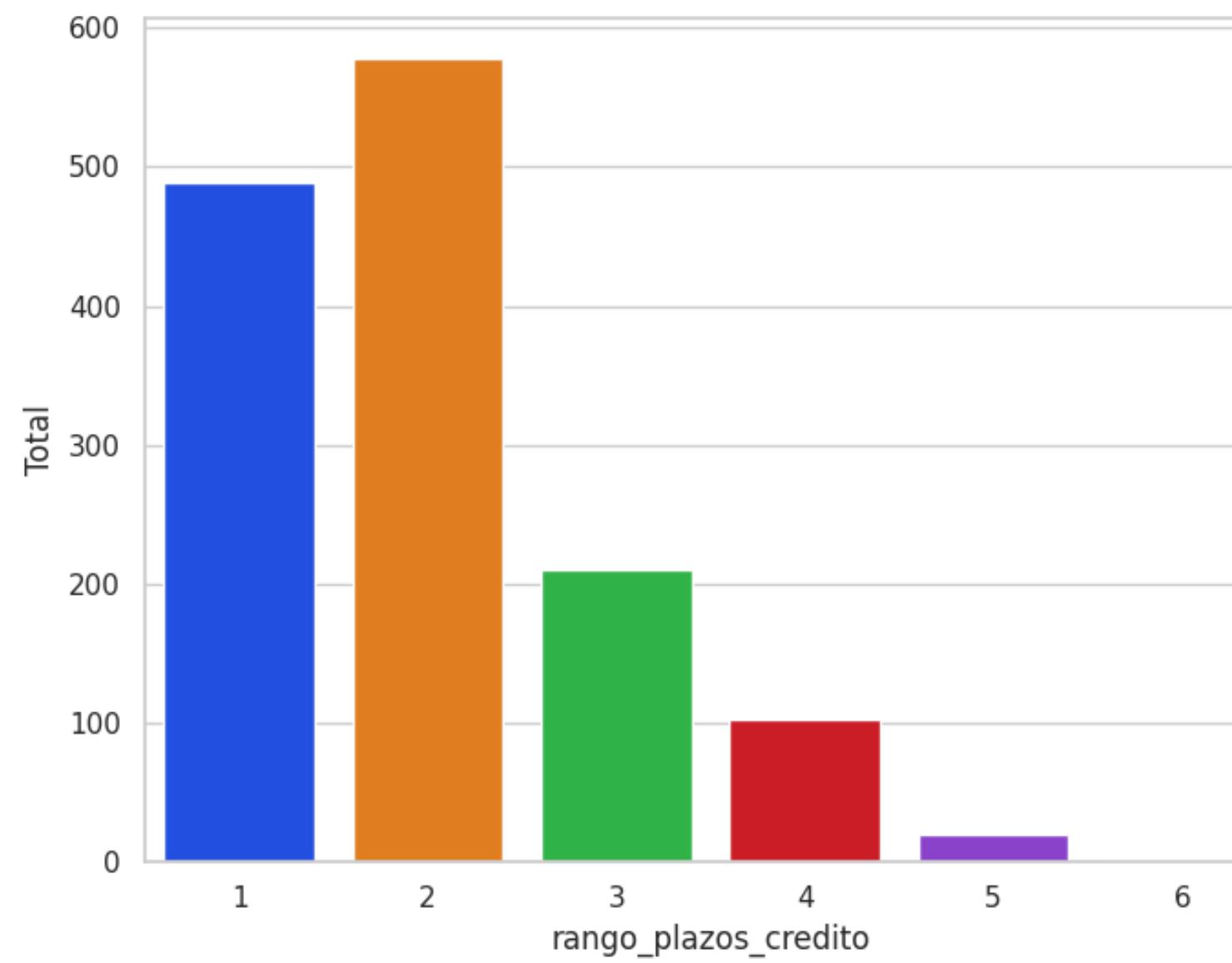
Para la estrategia de gestión de riesgos, se ha potenciado la evaluación crediticia mediante un modelo de machine learning. A través de un proceso llamado 'feature engineering', se optimizaron variables cruciales como sexo, estado civil, edad y plazos de crédito. Este enfoque avanzado, respaldado por la eliminación selectiva de variables y la conversión de datos, nos permite tomar decisiones más fundamentadas y proactivas en la administración de la cartera crediticia, fortaleciendo así la posición corporativa en la gestión del riesgo.



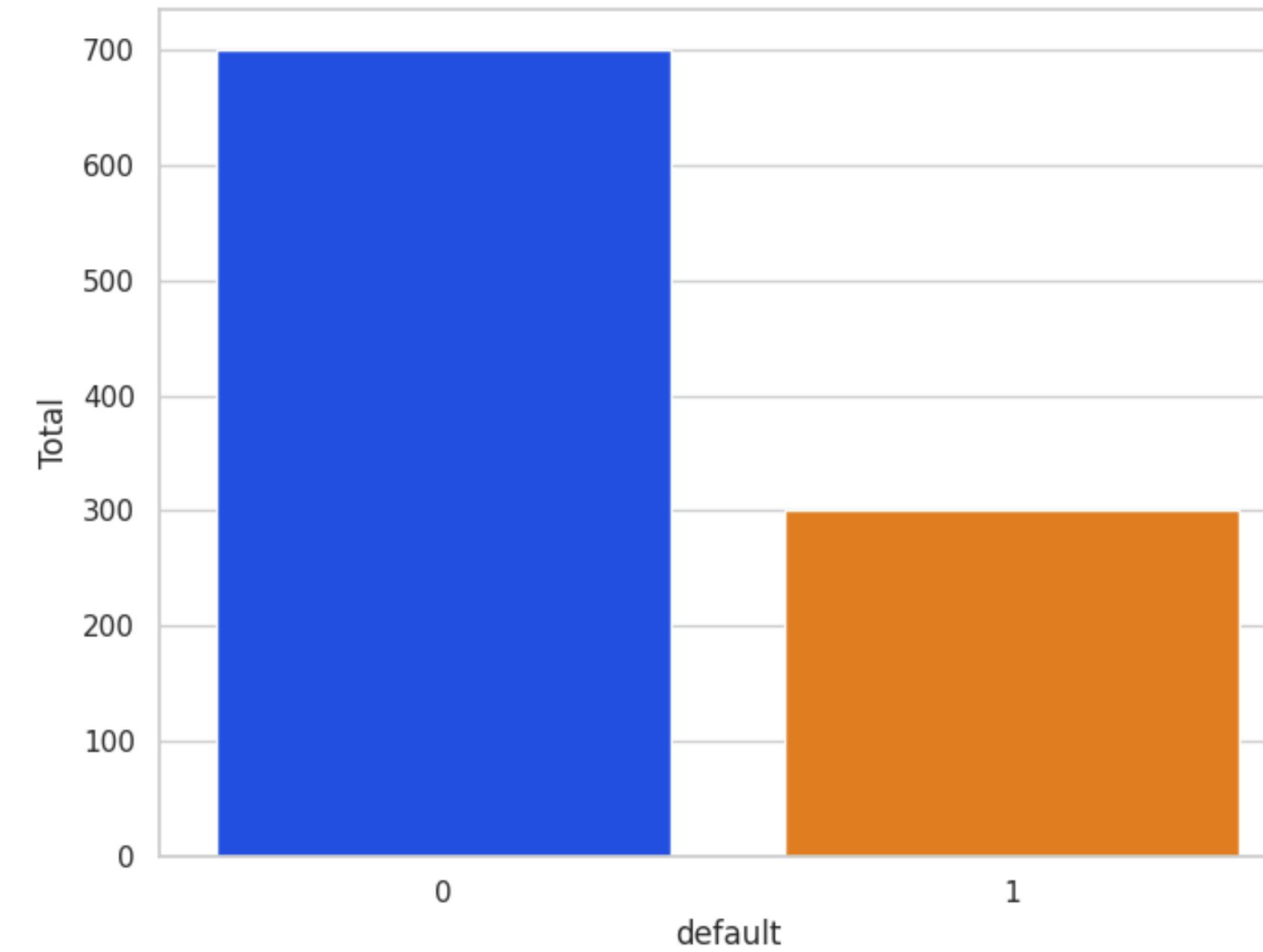
Análisis exploratorio



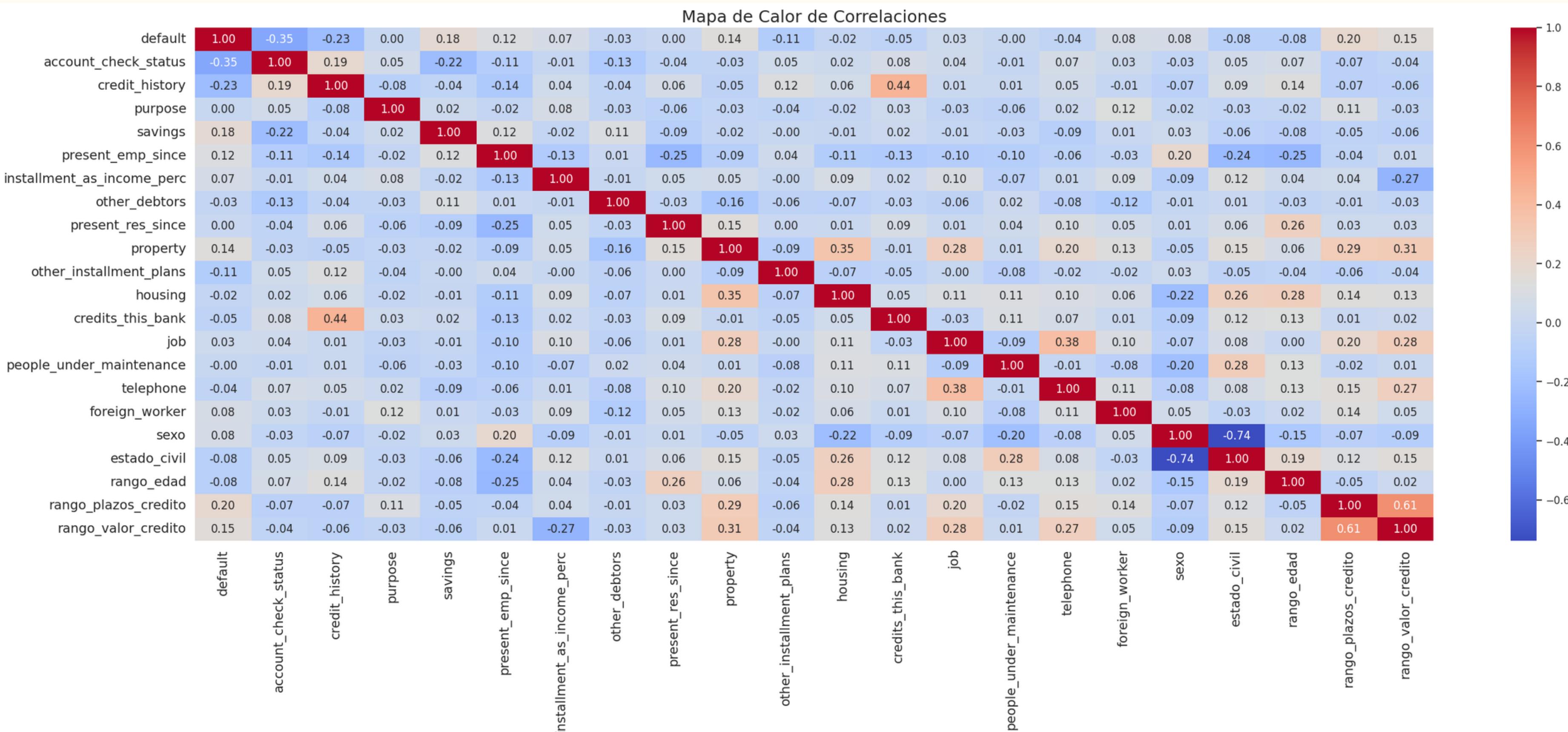
Análisis exploratorio



Análisis exploratorio

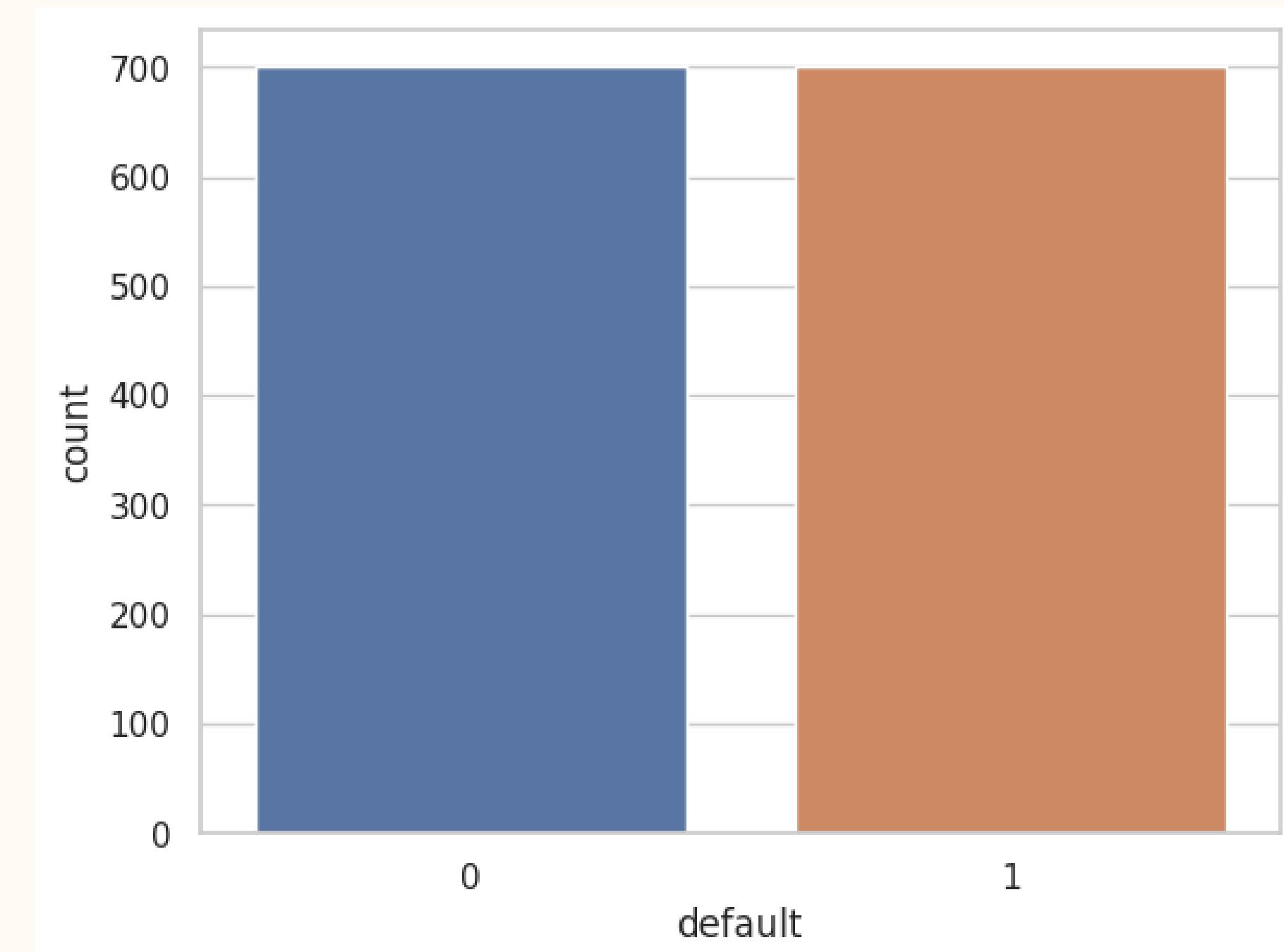


Análisis exploratorio - Mapa de Calor - Correlaciones entre variables.



Exploración de datos

Se ha implementado la función 'balancear_datos' para equilibrar proporciones en la base, utilizando la técnica SMOTE. Ahora, tras el proceso, se observa una distribución más uniforme de las categorías 'default', fortaleciendo la capacidad predictiva y equidad en la evaluación crediticia.



Construcción de modelos

Se ha desarrollado la función 'crea_modelos', que entrena y evalúa diversos modelos de machine learning, como Regresión Logística, Árbol de Decisión, Random Forest y Naive Bayes. Los resultados detallados, incluyendo precisión, recall y puntuación F1, se almacenan para facilitar la elección del modelo más efectivo en la predicción de solvencia crediticia.

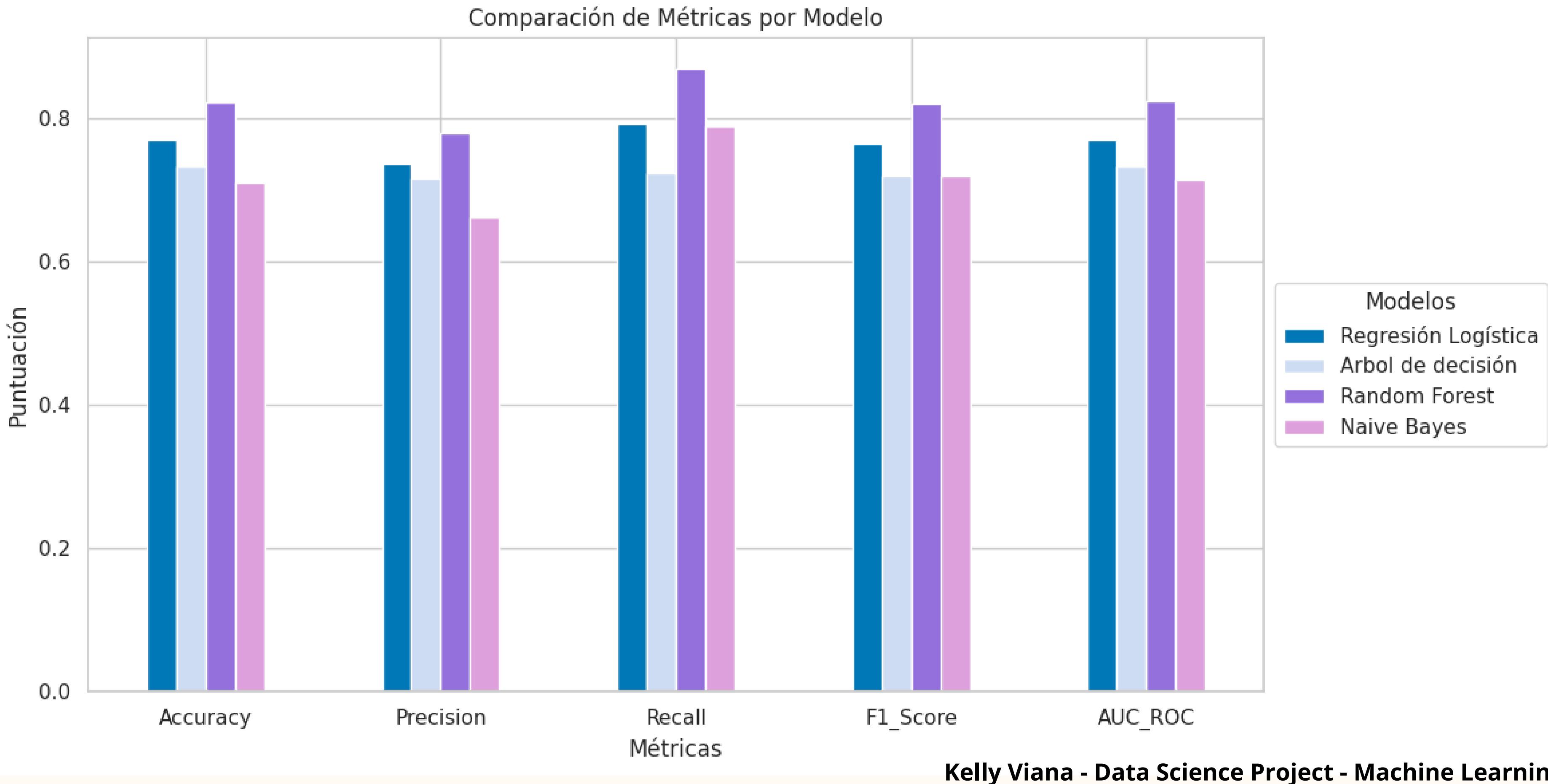
Modelo: Regresión Logística
Accuracy: 0.7690476190476191
Precisión: 0.7370892018779343
Recall: 0.7929292929292929
F1-score: 0.7639902676399025
AUC-ROC: 0.7703385203385203

Modelo: Árbol de decisión
Accuracy: 0.7333333333333333
Precisión: 0.715
Recall: 0.7222222222222222
F1-score: 0.7185929648241206
AUC-ROC: 0.7327327327327328

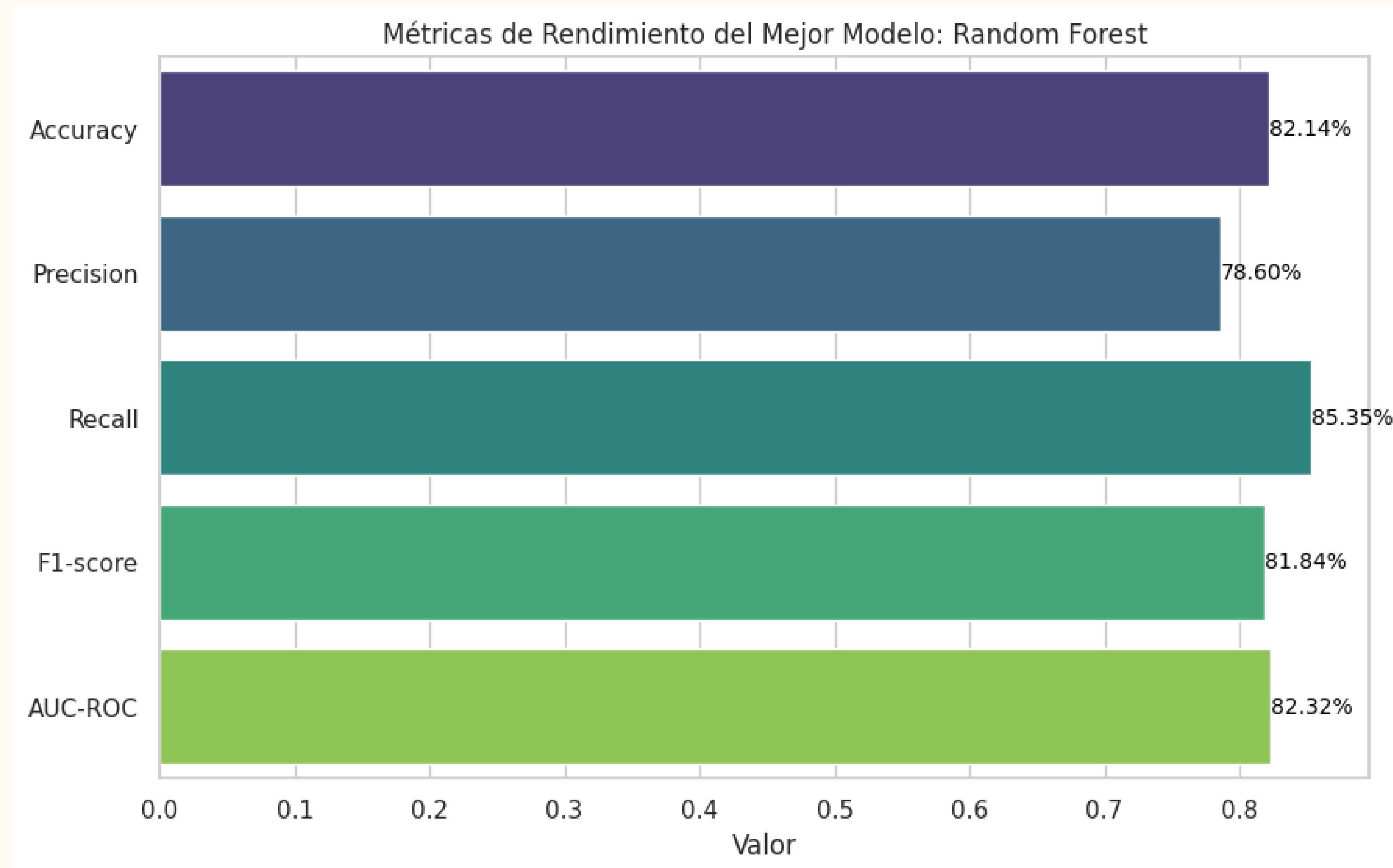
Modelo: Random Forest
Accuracy: 0.8214285714285714
Precisión: 0.7782805429864253
Recall: 0.8686868686868687
F1-score: 0.8210023866348449
AUC-ROC: 0.823983073983074

Modelo: Naive Bayes
Accuracy: 0.7095238095238096
Precisión: 0.6610169491525424
Recall: 0.7878787878787878
F1-score: 0.7188940092165897
AUC-ROC: 0.7137592137592138

Comparación de Métricas por Modelo

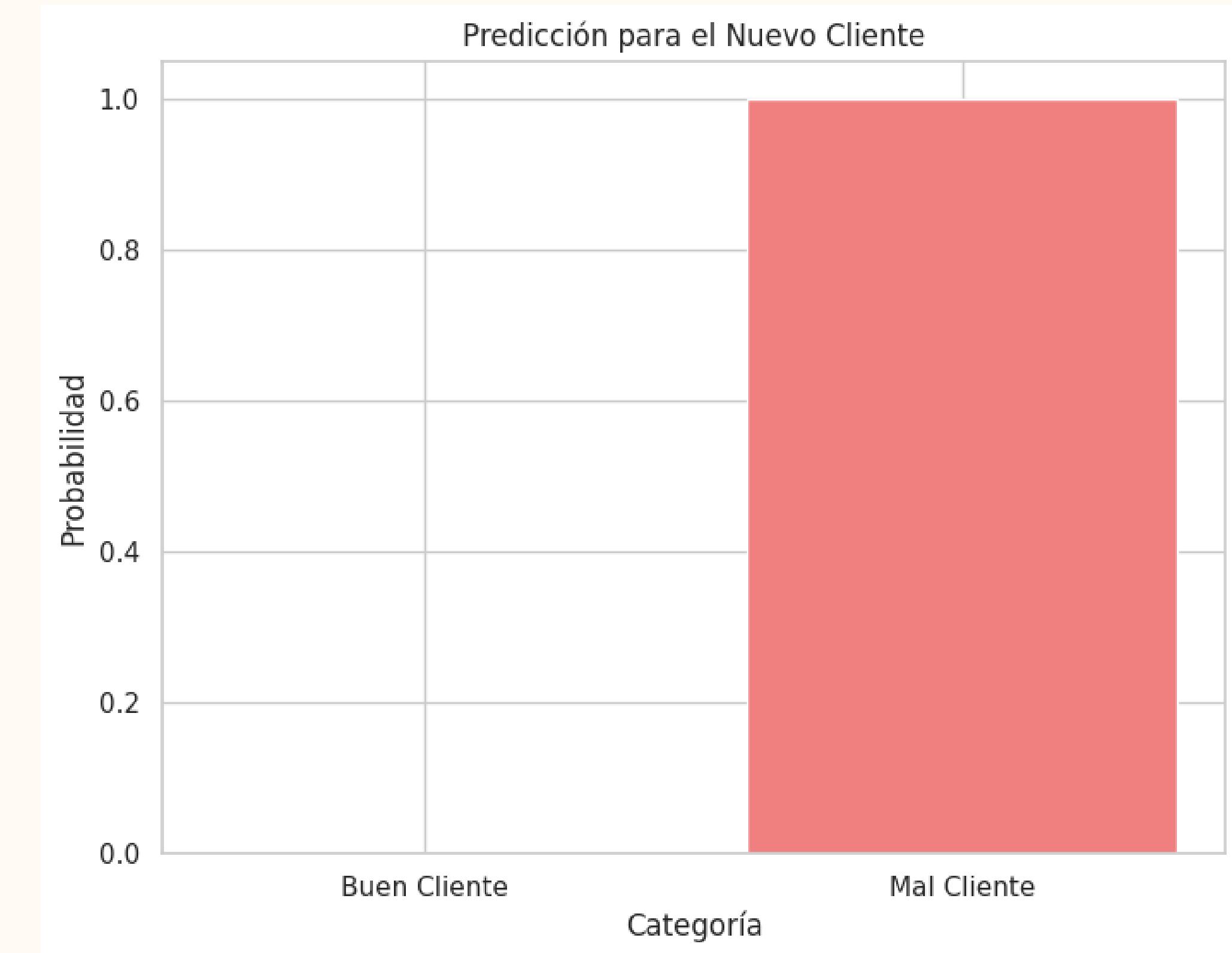


Escogencia del mejor modelo



Probando el modelo con nuevos datos

```
nuevo_registro = {  
    'account_check_status': 1,  
    'credit_history': 2,  
    'purpose': 3,  
    'savings': 4,  
    'present_emp_since': 5,  
    'installment_as_income_perc': 4,  
    'other_debtors': 1,  
    'present_res_since': 4,  
    'property': 2,  
    'other_installment_plans': 1,  
    'housing': 3,  
    'credits_this_bank': 2,  
    'job': 2,  
    'people_under_maintenance': 1,  
    'telephone': 1,  
    'foreign_worker': 1,  
    'sexo': 1,  
    'estado_civil': 0,  
    'rango_edad': 3,  
    'rango_plazos_credito': 2,  
    'rango_valor_credito': 10  
}
```



Conclusiones

La implementación exitosa del modelo Random Forest, con una destacada precisión del 82.38%, resalta la eficacia de las técnicas de machine learning en la evaluación de solvencia crediticia. La predicción de riesgo para un nuevo cliente subraya la utilidad práctica de este enfoque avanzado en la toma de decisiones financieras. Este proyecto no solo optimiza la eficiencia operativa, sino que también establece un marco sólido para la toma de decisiones proactiva y precisa en el sector financiero. La capacidad del modelo para identificar riesgos potenciales refuerza la seguridad y la solidez en la gestión crediticia, mejorando así la robustez del sistema.