

# Datathon Allianz

# El Problema

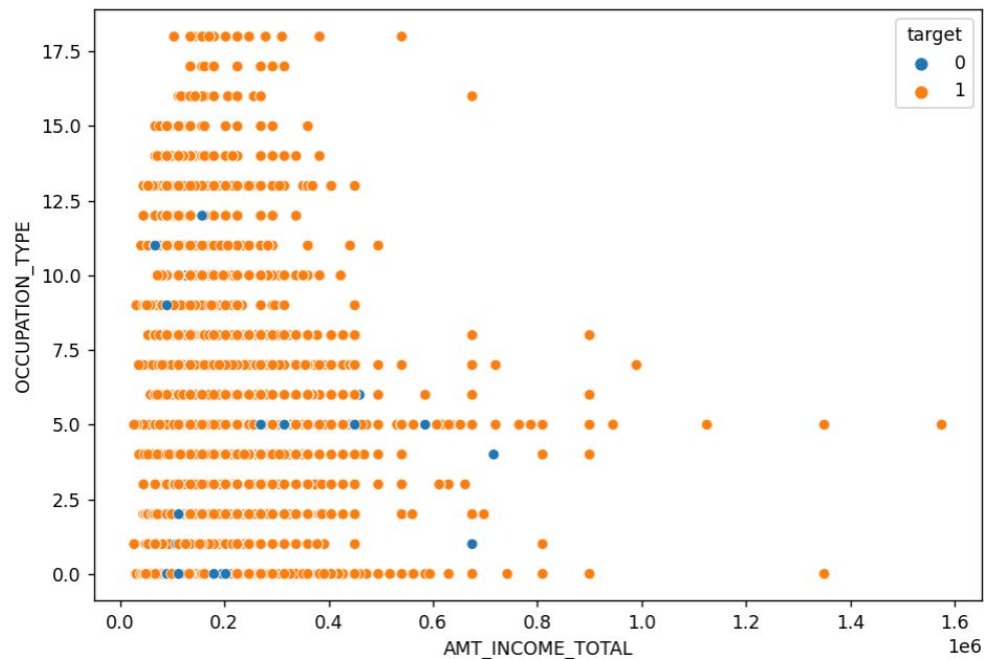
¿A qué clientes nuevos se le puede otorgar un fraccionamiento de prima gratuito?

- Tres tablas
- Proceso de resolución
- Planteamiento de variable target (2 meses de retraso)
- Selección de modelos según métricas obtenidas

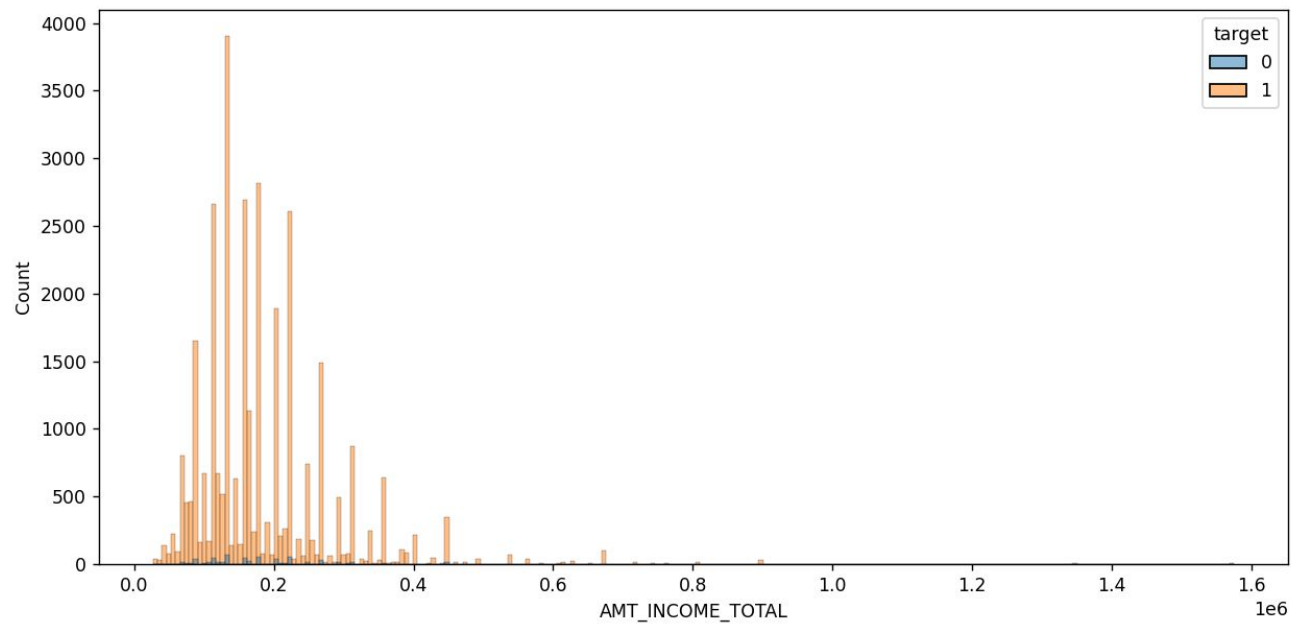
# Data Treatment & Data Cleaning

- Python
- Eliminado registros con solo status X de ambos conjuntos de datos
- Variable “Target” calculada  $>1$  mes impagado = 1
- Valores nulos
- Unión entre cuadros ID con variable Target con la información cualitativa

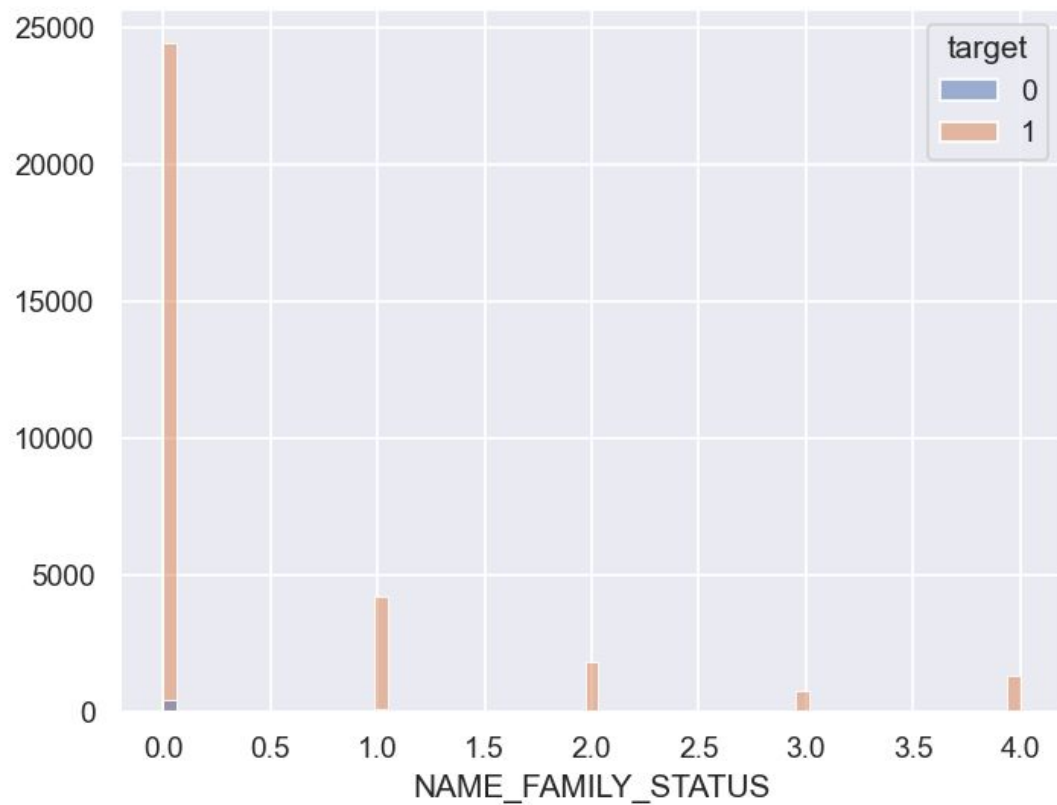
# Tipo de trabajo vs Ingresos totales



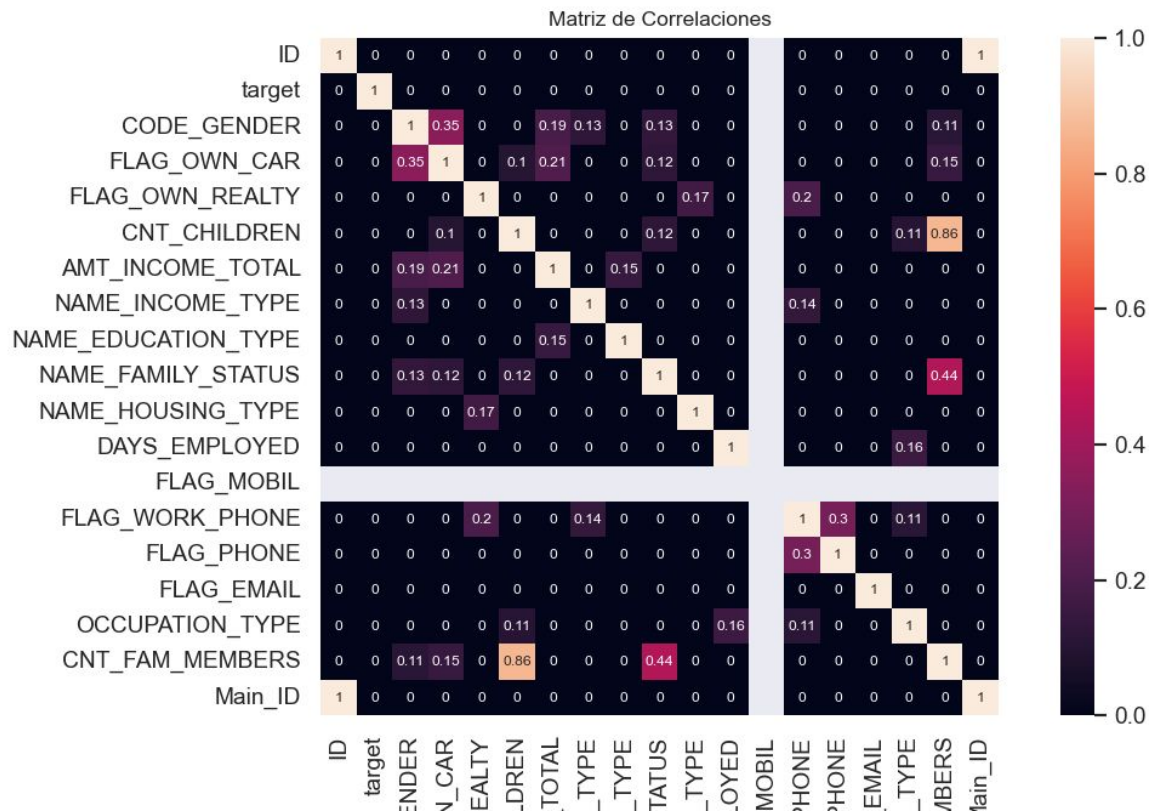
# Ingresos Totales



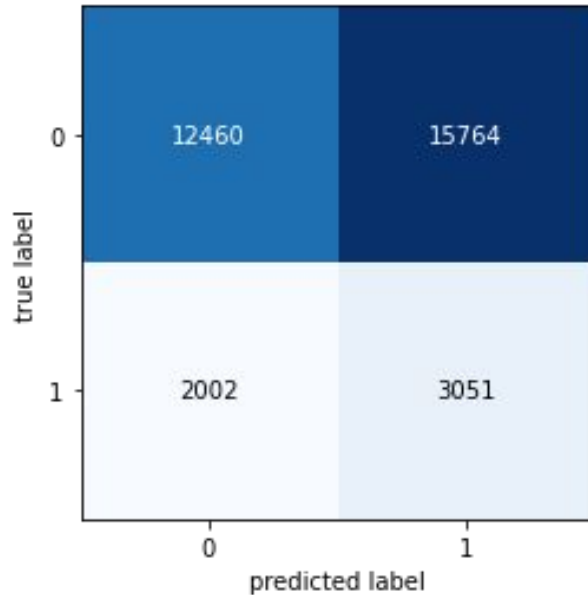
# Family Status vs Target



# Matriz de Correlaciones



# Logistic Regression



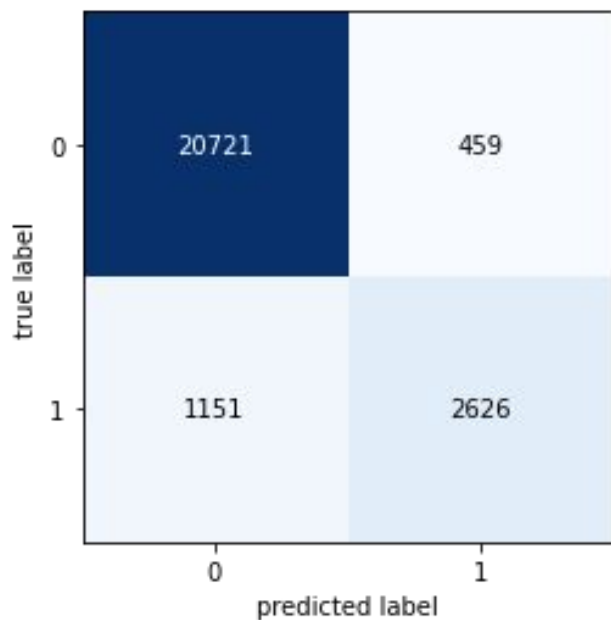
Cross-validation  
L2-regularization

AUC score: 0.54  
F1-score: 0.256



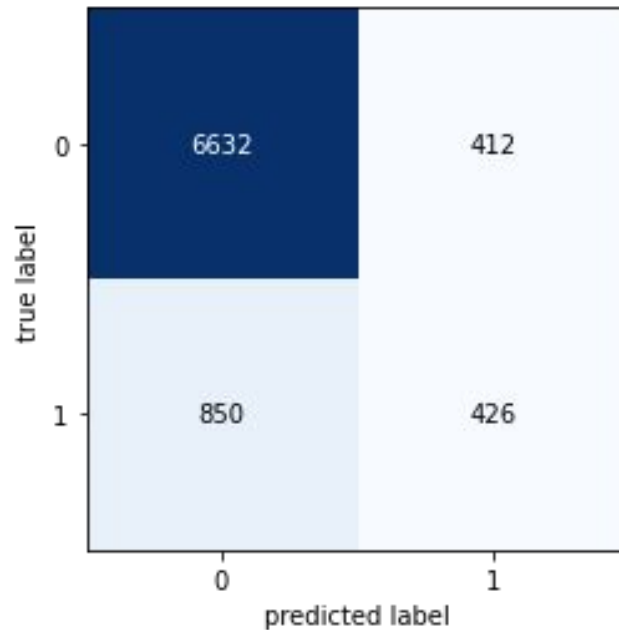
# AdaBoost

Train



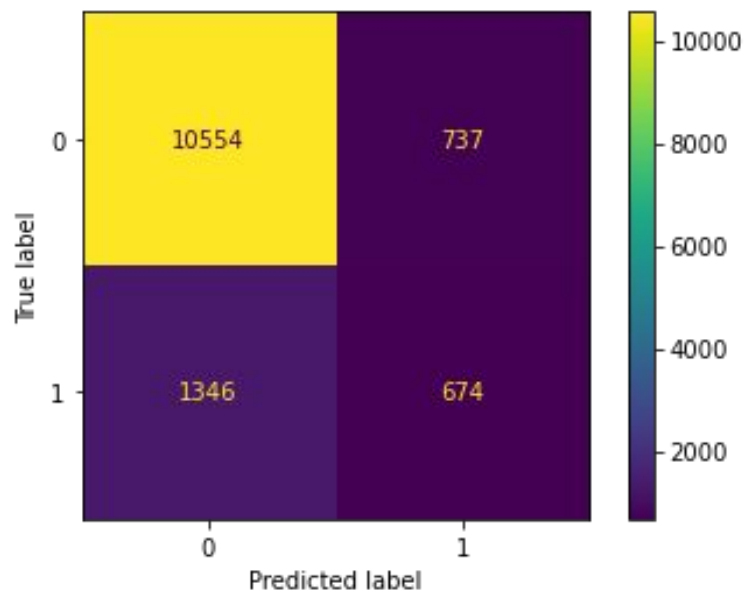
AUC score: 0.91  
F1-score: 0.76

Test

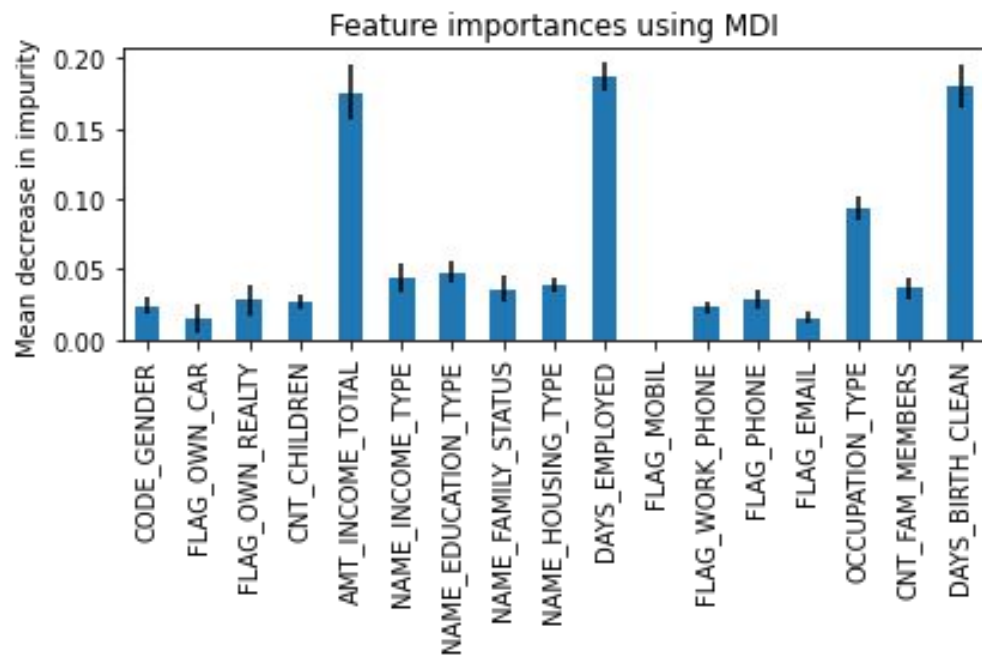


AUC score: 0.71  
F1-score: 0.4

# Random Forest



AUC score: 0.73



F1-score: 0.38

# Conclusiones

Seleccionamos modelos Random Forest y AdaBoost

Al tener una variable target desbalanceada nos fijamos en AUC

Es posible saber qué clientes son mejores para otorgar el fraccionamiento de prima gratuito

# Propuestas para el futuro

Usar una clasificación en los addresses

Detectar Outliers en mayor profundidad

Agrupar las edades por rangos

# Propuesta de automatización

- Usando una plataforma nube
- Data pipeline
- Actualizar el modelo
- Mantenimiento de los pasos de limpieza

