Autores: Addiel Paz Perez

Reinaldo Mursuli Tavio  
  
  
  
Primeramente se carga el csv para el preprocesamiento de datos , donde al utilizar : describe(), info() y head() Se obtienen estos valores...  
  
  
  
Información general:

Total de filas: 4521

Total de columnas: 17

# Column Non-Null Count Dtype

0 age 4521 non-null int64

1 job 4521 non-null object

2 marital 4521 non-null object

3 education 4521 non-null object

4 default 4521 non-null object

5 balance 4521 non-null int64

6 housing 4521 non-null object

7 loan 4521 non-null object

8 contact 4521 non-null object

9 day 4521 non-null int64

10 month 4521 non-null object

11 duration 4521 non-null int64

12 campaign 4521 non-null int64

13 pdays 4521 non-null int64

14 previous 4521 non-null int64

15 poutcome 4521 non-null object

16 y 4521 non-null object

dtypes: int64(7), object(10)

Estadísticas descriptivas:

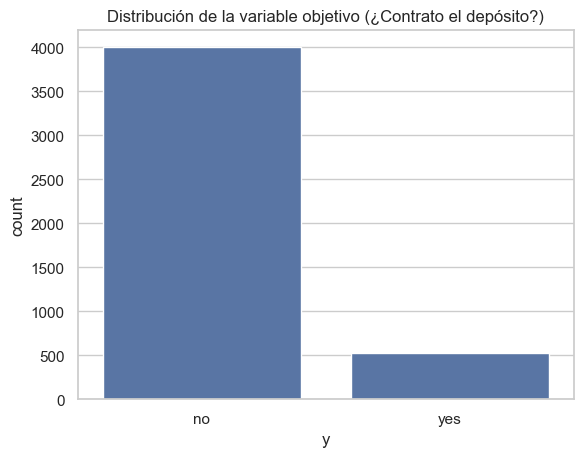
|  | **age** | **balance** | **day** | **duration** | **campaign** | **pdays** | **previous** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| count | 4521.000000 | 4521.000000 | 4521.000000 | 4521.000000 | 4521.000000 | 4521.000000 | 4521.000000 |
| mean | 41.170095 | 1422.657819 | 15.915284 | 263.961292 | 2.793630 | 39.766645 | 0.542579 |
| std | 10.576211 | 3009.638142 | 8.247667 | 259.856633 | 3.109807 | 100.121124 | 1.693562 |
| min | 19.000000 | 3313.000000 | 1.000000 | 4.000000 | 1.000000 | -1.000000 | 0.000000 |
| 25% | 33.000000 | 69.000000 | 9.000000 | 104.000000 | 1.000000 | -1.000000 | 0.000000 |
| 50% | 39.000000 | 444.000000 | 9.000000 | 185.000000 | 2.000000 | -1.000000 | 0.000000 |
| 75% | 49.000000 | 1480.000000 | 21.000000 | 329.000000 | 3.000000 | -1.000000 | 0.000000 |
| max | 87.000000 | 71188.000000 | 31.000000 | 3025.000000 | 50.000000 | 871.000000 | 25.000000 |

Se analiza la Distribución de la variable objetivo 'y':

no 0.88476

yes 0.11524

Name: proportion, dtype: float64



Se analizan valores nulos y unknown por columna:

Resumen de valores nulos y unknown:

Valores\_Nulos Valores\_Unknown Total\_Registros Porcentaje\_Nulos

job 0 38 4521 0.0

education 0 187 4521 0.0

contact 0 1324 4521 0.0

poutcome 0 3705 4521 0.0

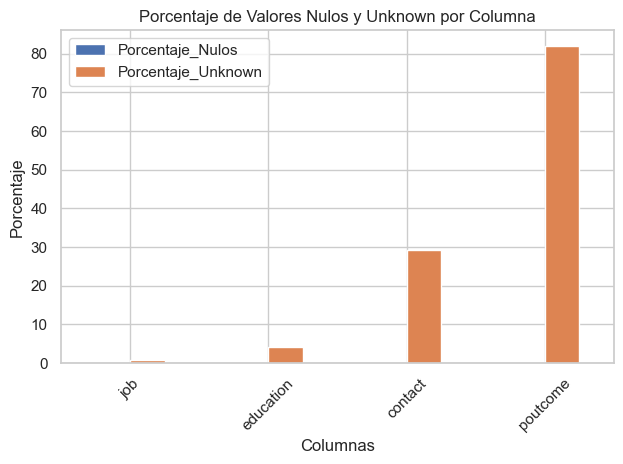
Porcentaje\_Unknown

job 0.840522

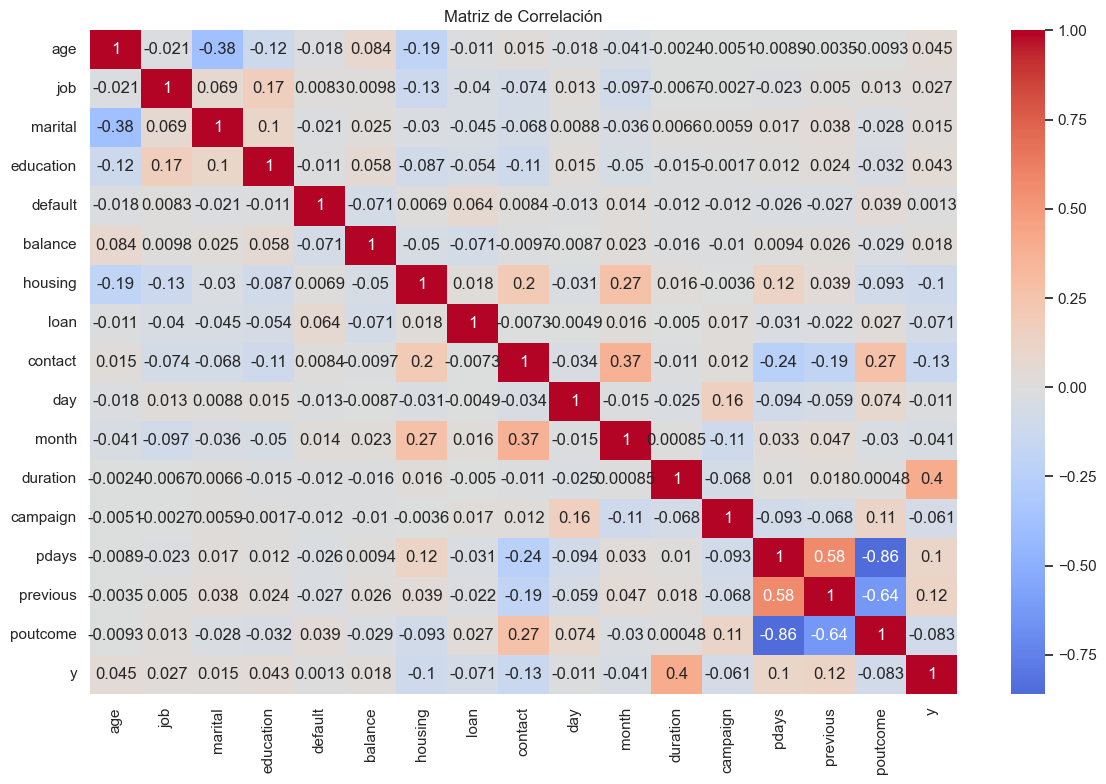
Education 4.136253

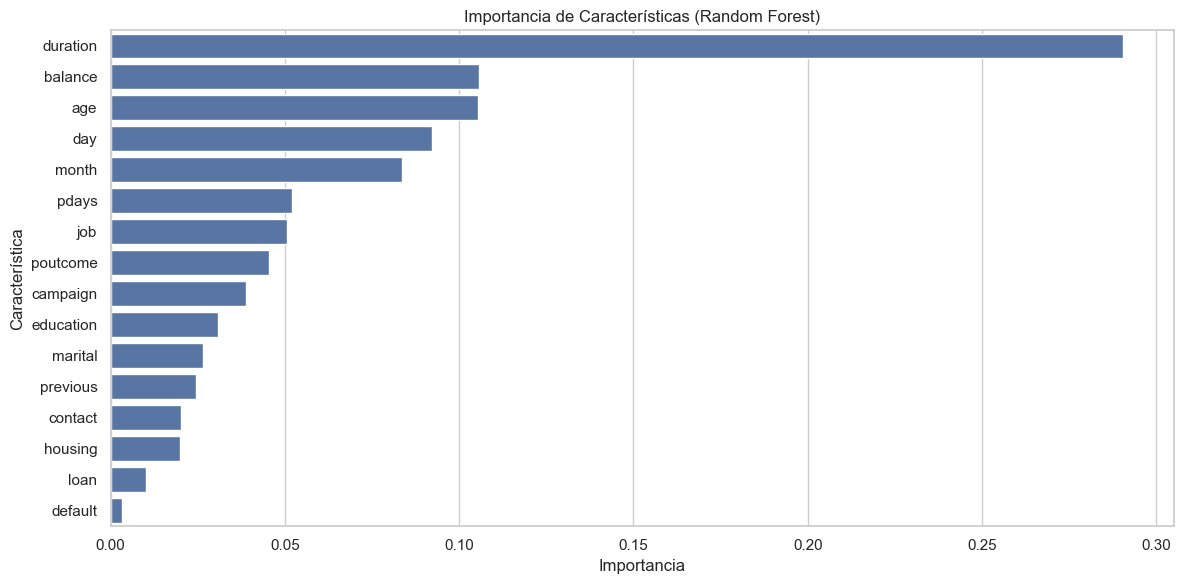
contact 29.285556

poutcome 81.950896



Aui la representacion de la matriz de correlacion para observar si las variables con valores unknonw influyen matematicamente en la variable objetivo:





Top 10 características más importantes para random forest:

Característica Importancia

11 duration 0.290557

5 balance 0.105800

0 age 0.105436

9 day 0.092174

10 month 0.083508

13 pdays 0.052059

1 job 0.050703

15 poutcome 0.045556

12 campaign 0.038802

3 education 0.030945

Se convierte los valores unknown a nulos

Se elimina la columna poutcome y contact,ademas de la filas job y education

Se procede a Entrenamiento y prueba

División completada:

Train: 3017 muestras

Test: 1294 muestras

Positivos en train: 349 (11.57%)

Entrenando: Árbol de Decisión

Reporte de clasificación:

precision recall f1-score support

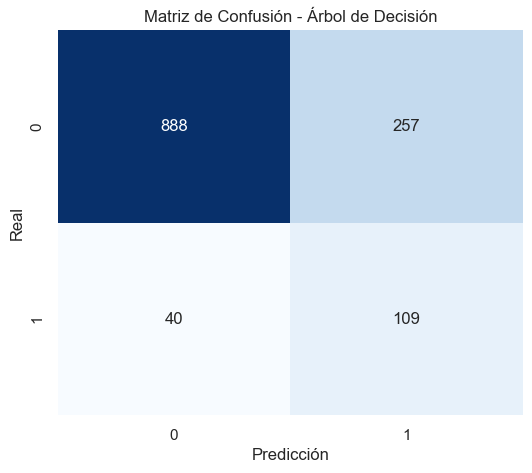
0 0.9569 0.7755 0.8567 1145

1 0.2978 0.7315 0.4233 149

accuracy 0.7705 1294

macro avg 0.6274 0.7535 0.6400 1294

weighted avg 0.8810 0.7705 0.8068 1294



Entrenando: Random Forest

Reporte de clasificación:

precision recall f1-score support

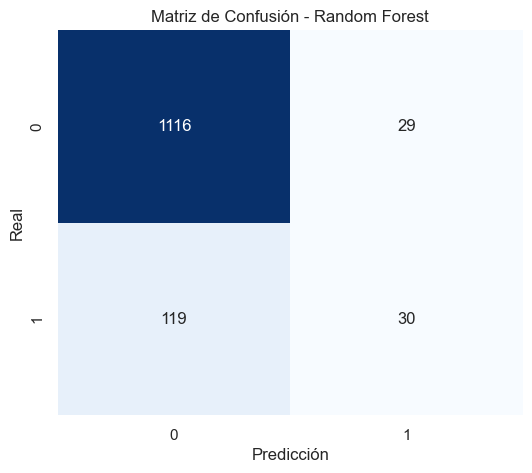
0 0.9036 0.9747 0.9378 1145

1 0.5085 0.2013 0.2885 149

accuracy 0.8856 1294

macro avg 0.7061 0.5880 0.6131 1294

weighted avg 0.8581 0.8856 0.8630 1294



Entrenando: SVM

Reporte de clasificación:

precision recall f1-score support

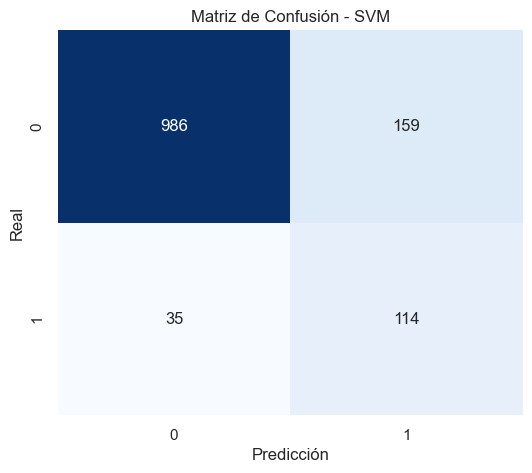
0 0.9657 0.8611 0.9104 1145

1 0.4176 0.7651 0.5403 149

accuracy 0.8501 1294

macro avg 0.6917 0.8131 0.7254 1294

weighted avg 0.9026 0.8501 0.8678 1294



Entrenando: Regresión Logística

Reporte de clasificación:

precision recall f1-score support

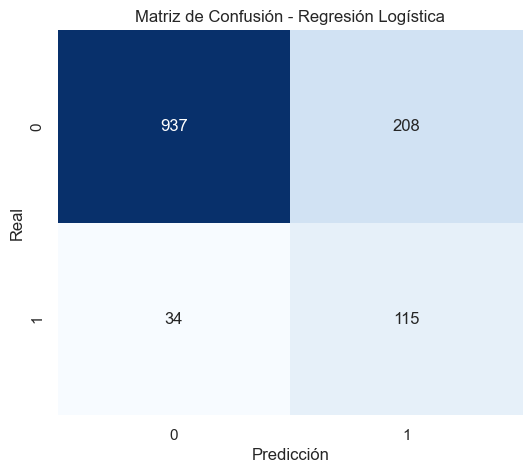
0 0.9650 0.8183 0.8856 1145

1 0.3560 0.7718 0.4873 149

Accuracy 0.8130 1294

macro avg 0.6605 0.7951 0.6865 1294

weighted avg 0.8949 0.8130 0.8398 1294



Entrenando: MLP

Reporte de clasificación:

precision recall f1-score support

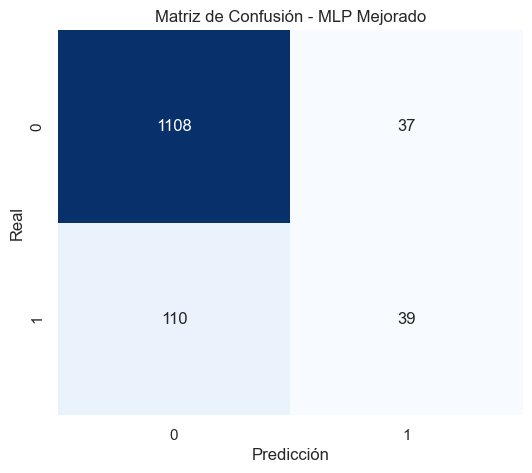
0 0.9097 0.9677 0.9378 1145

1 0.5132 0.2617 0.3467 149

accuracy 0.8864 1294

macro avg 0.7114 0.6147 0.6422 1294

weighted avg 0.8640 0.8864 0.8697 1294



Se selecciono bajo los resultados SVR  
  
  
  
Dados estos resultados , se elige como modelo final SVM y se entrena y despliega en streamlit