
1. INFORMACIÓN GENERAL DEL DATASET

 Dimensiones totales: 1000 filas x 21 columnas

 Tipos de datos por columna:

- 'status_checking_account': object (4 valores únicos)
- 'month_credit_duration': object (33 valores únicos)
- 'credit_history': object (5 valores únicos)
- 'purpose': object (10 valores únicos)
- 'credit_amount': object (921 valores únicos)
- 'savings_type': object (5 valores únicos)
- 'years_of_employment': object (5 valores únicos)
- 'pct_fee_income': object (4 valores únicos)
- 'status_sex': object (4 valores únicos)
- 'debtors': object (3 valores únicos)
- 'years_of_residence': object (4 valores únicos)
- 'property': object (4 valores únicos)
- 'age': object (53 valores únicos)
- 'other_installments': object (3 valores únicos)
- 'housing': object (3 valores únicos)
- 'num_existing_credits': object (4 valores únicos)
- 'job': object (4 valores únicos)
- 'num_dependents': object (2 valores únicos)
- 'telephone': object (2 valores únicos)
- 'foreign_worker': object (2 valores únicos)
- 'target': object (2 valores únicos)

 Información detallada:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 21 columns):
 #  Column            Non-Null Count Dtype
 --- 
 0  status_checking_account    1000 non-null object
 1  month_credit_duration    1000 non-null object
 2  credit_history           1000 non-null object
 3  purpose                  1000 non-null object
 4  credit_amount             1000 non-null object
 5  savings_type              1000 non-null object
 6  years_of_employment       1000 non-null object
 7  pct_fee_income            1000 non-null object
 8  status_sex                1000 non-null object
 9  debtors                  1000 non-null object
 10 years_of_residence        1000 non-null object
 11 property                 1000 non-null object
```

```
12 age           1000 non-null object
13 other_installments 1000 non-null object
14 housing         1000 non-null object
15 num_existing_credits 1000 non-null object
16 job             1000 non-null object
17 num_dependents 1000 non-null object
18 telephone        1000 non-null object
19 foreign_worker   1000 non-null object
20 target           1000 non-null object
dtypes: object(21)
memory usage: 164.2+ KB
```

2.1 ESTADÍSTICAS DESCRIPTIVAS - VARIABLES NUMÉRICAS

- ✓ Se encontraron 7 variables numéricas

Variable: 'month_credit_duration'

- Media: 20.90
- Mediana: 18.00
- Desviación estándar: 12.06
- Mínimo: 4.00
- Percentil 25%: 12.00
- Percentil 50% (mediana): 18.00
- Percentil 75%: 24.00
- Máximo: 72.00

Variable: 'credit_amount'

- Media: 3271.26
- Mediana: 2319.50
- Desviación estándar: 2822.74
- Mínimo: 250.00
- Percentil 25%: 1365.50
- Percentil 50% (mediana): 2319.50
- Percentil 75%: 3972.25
- Máximo: 18424.00

Variable: 'pct_fee_income'

- Media: 2.97
- Mediana: 3.00
- Desviación estándar: 1.12
- Mínimo: 1.00
- Percentil 25%: 2.00

- Percentil 50% (mediana): 3.00
- Percentil 75%: 4.00
- Máximo: 4.00

 Variable: 'years_of_residence'

- Media: 2.85
- Mediana: 3.00
- Desviación estándar: 1.10
- Mínimo: 1.00
- Percentil 25%: 2.00
- Percentil 50% (mediana): 3.00
- Percentil 75%: 4.00
- Máximo: 4.00

 Variable: 'age'

- Media: 35.55
- Mediana: 33.00
- Desviación estándar: 11.38
- Mínimo: 19.00
- Percentil 25%: 27.00
- Percentil 50% (mediana): 33.00
- Percentil 75%: 42.00
- Máximo: 75.00

 Variable: 'num_existing_credits'

- Media: 1.41
- Mediana: 1.00
- Desviación estándar: 0.58
- Mínimo: 1.00
- Percentil 25%: 1.00
- Percentil 50% (mediana): 1.00
- Percentil 75%: 2.00
- Máximo: 4.00

 Variable: 'num_dependents'

- Media: 1.16
- Mediana: 1.00
- Desviación estándar: 0.36
- Mínimo: 1.00
- Percentil 25%: 1.00
- Percentil 50% (mediana): 1.00
- Percentil 75%: 1.00
- Máximo: 2.00

2.2 ESTADÍSTICAS DESCRIPTIVAS - VARIABLES CATEGÓRICAS

- ✓ Se encontraron 14 variables categóricas

Variable: 'status_checking_account'

- Total de categorías únicas: 4

 Top 5 categorías más frecuentes:

1. 'A14': 394 registros (39.40%)
2. 'A11': 274 registros (27.40%)
3. 'A12': 269 registros (26.90%)
4. 'A13': 63 registros (6.30%)

Variable: 'credit_history'

- Total de categorías únicas: 5

 Top 5 categorías más frecuentes:

1. 'A32': 530 registros (53.00%)
2. 'A34': 293 registros (29.30%)
3. 'A33': 88 registros (8.80%)
4. 'A31': 49 registros (4.90%)
5. 'A30': 40 registros (4.00%)

Variable: 'purpose'

- Total de categorías únicas: 10

 Top 5 categorías más frecuentes:

1. 'A43': 280 registros (28.00%)
2. 'A40': 234 registros (23.40%)
3. 'A42': 181 registros (18.10%)
4. 'A41': 103 registros (10.30%)
5. 'A49': 97 registros (9.70%)

Variable: 'savings_type'

- Total de categorías únicas: 5

 Top 5 categorías más frecuentes:

1. 'A61': 603 registros (60.30%)
2. 'A65': 183 registros (18.30%)
3. 'A62': 103 registros (10.30%)
4. 'A63': 63 registros (6.30%)

5. 'A64': 48 registros (4.80%)

 Variable: 'years_of_employment'

- Total de categorías únicas: 5

 Top 5 categorías más frecuentes:

1. 'A73': 339 registros (33.90%)
2. 'A75': 253 registros (25.30%)
3. 'A74': 174 registros (17.40%)
4. 'A72': 172 registros (17.20%)
5. 'A71': 62 registros (6.20%)

 Variable: 'status_sex'

- Total de categorías únicas: 4

 Top 5 categorías más frecuentes:

1. 'A93': 548 registros (54.80%)
2. 'A92': 310 registros (31.00%)
3. 'A94': 92 registros (9.20%)
4. 'A91': 50 registros (5.00%)

 Variable: 'debtors'

- Total de categorías únicas: 3

 Top 5 categorías más frecuentes:

1. 'A101': 907 registros (90.70%)
2. 'A103': 52 registros (5.20%)
3. 'A102': 41 registros (4.10%)

 Variable: 'property'

- Total de categorías únicas: 4

 Top 5 categorías más frecuentes:

1. 'A123': 332 registros (33.20%)
2. 'A121': 282 registros (28.20%)
3. 'A122': 232 registros (23.20%)
4. 'A124': 154 registros (15.40%)

 Variable: 'other_installments'

- Total de categorías únicas: 3

 Top 5 categorías más frecuentes:

1. 'A143': 814 registros (81.40%)

2. 'A141': 139 registros (13.90%)
3. 'A142': 47 registros (4.70%)

 Variable: 'housing'

- Total de categorías únicas: 3

 Top 5 categorías más frecuentes:

1. 'A152': 713 registros (71.30%)
2. 'A151': 179 registros (17.90%)
3. 'A153': 108 registros (10.80%)

 Variable: 'job'

- Total de categorías únicas: 4

 Top 5 categorías más frecuentes:

1. 'A173': 630 registros (63.00%)
2. 'A172': 200 registros (20.00%)
3. 'A174': 148 registros (14.80%)
4. 'A171': 22 registros (2.20%)

 Variable: 'telephone'

- Total de categorías únicas: 2

 Top 5 categorías más frecuentes:

1. 'A191': 596 registros (59.60%)
2. 'A192': 404 registros (40.40%)

 Variable: 'foreign_worker'

- Total de categorías únicas: 2

 Top 5 categorías más frecuentes:

1. 'A201': 963 registros (96.30%)
2. 'A202': 37 registros (3.70%)

 Variable: 'target'

- Total de categorías únicas: 2

 Top 5 categorías más frecuentes:

1. '1': 700 registros (70.00%)
2. '2': 300 registros (30.00%)

3. ANÁLISIS DE VALORES FALTANTES

=====

 Total de valores faltantes en el dataset: 0
 Porcentaje total de valores faltantes: 0.00%

 Valores faltantes por columna:

 No hay valores faltantes en ninguna columna.

=====

4. ANÁLISIS DE REGISTROS DUPLICADOS

 Número de filas duplicadas: 0
 Porcentaje de filas duplicadas: 0.00%

 No se encontraron registros duplicados en el dataset.

=====

5. ANÁLISIS DE LA VARIABLE OBJETIVO ('target')

 Valores únicos en 'target': [1', '2']

 Distribución de clases:

- Clase '1': 700 registros (70.00%)
- Clase '2': 300 registros (30.00%)

 Ratio de desbalance (clase mayoritaria/minoritaria): 2.33:1

 El dataset presenta desbalance de clases. Considere técnicas de balanceo.

=====

6. ANÁLISIS DE CORRELACIONES ENTRE VARIABLES NUMÉRICAS

 Correlaciones significativas ($|correlación| > 0.5$):

 'month_credit_duration' ↔ 'credit_amount': 0.625

 Resumen: Se encontraron 1 pares de variables con correlación significativa.

=====

7. DETECCIÓN DE OUTLIERS (MÉTODO IQR)

 Método: Rango Intercuartílico (IQR)
Outliers = valores < Q1 - 1.5*IQR o valores > Q3 + 1.5*IQR

 Variable: 'month_credit_duration'

-
- Q1 (Percentil 25%): 12.00
 - Q3 (Percentil 75%): 24.00
 - IQR: 12.00
 - Límite inferior: -6.00
 - Límite superior: 42.00

 Outliers detectados: 70 (7.00%)

 Ejemplos de outliers (máximo 5):

1. 48.00
2. 48.00
3. 60.00
4. 45.00
5. 48.00

 Variable: 'credit_amount'

-
- Q1 (Percentil 25%): 1365.50
 - Q3 (Percentil 75%): 3972.25
 - IQR: 2606.75
 - Límite inferior: -2544.62
 - Límite superior: 7882.38

 Outliers detectados: 72 (7.20%)

 Ejemplos de outliers (máximo 5):

1. 9055.00
2. 8072.00
3. 12579.00
4. 9566.00
5. 14421.00

 Variable: 'pct_fee_income'

-
- Q1 (Percentil 25%): 2.00
 - Q3 (Percentil 75%): 4.00
 - IQR: 2.00
 - Límite inferior: -1.00
 - Límite superior: 7.00

 No se detectaron outliers

 Variable: 'years_of_residence'

-
- Q1 (Percentil 25%): 2.00
 - Q3 (Percentil 75%): 4.00
 - IQR: 2.00

- Límite inferior: -1.00
 - Límite superior: 7.00
-  No se detectaron outliers

 Variable: 'age'

- Q1 (Percentil 25%): 27.00
 - Q3 (Percentil 75%): 42.00
 - IQR: 15.00
 - Límite inferior: 4.50
 - Límite superior: 64.50
-  Outliers detectados: 23 (2.30%)

 Ejemplos de outliers (máximo 5):

1. 67.00
2. 66.00
3. 66.00
4. 70.00
5. 65.00

 Variable: 'num_existing_credits'

- Q1 (Percentil 25%): 1.00
 - Q3 (Percentil 75%): 2.00
 - IQR: 1.00
 - Límite inferior: -0.50
 - Límite superior: 3.50
-  Outliers detectados: 6 (0.60%)

 Ejemplos de outliers (máximo 5):

1. 4.00
2. 4.00
3. 4.00
4. 4.00
5. 4.00

 Variable: 'num_dependents'

- Q1 (Percentil 25%): 1.00
 - Q3 (Percentil 75%): 1.00
 - IQR: 0.00
 - Límite inferior: 1.00
 - Límite superior: 1.00
-  Outliers detectados: 155 (15.50%)

 Ejemplos de outliers (máximo 5):

1. 2.00
2. 2.00

- 3. 2.00
 - 4. 2.00
 - 5. 2.00
-

===== 8. ANÁLISIS BIVARIADO: RELACIÓN ENTRE VARIABLES Y TARGET =====

8.1 VARIABLES NUMÉRICAS vs TARGET

 Variable: 'month_credit_duration'

Estadísticas por clase:

- Clase '1': media = 19.21, std = 11.08
- Clase '2': media = 24.86, std = 13.28

 Diferencia porcentual entre clases: 29.43%

 Alta diferencia - Variable potencialmente relevante para predicción

 Variable: 'credit_amount'

Estadísticas por clase:

- Clase '1': media = 2985.46, std = 2401.47
- Clase '2': media = 3938.13, std = 3535.82

 Diferencia porcentual entre clases: 31.91%

 Alta diferencia - Variable potencialmente relevante para predicción

 Variable: 'pct_fee_income'

Estadísticas por clase:

- Clase '1': media = 2.92, std = 1.13
- Clase '2': media = 3.10, std = 1.09

 Diferencia porcentual entre clases: 6.05%

 Baja diferencia - Variable con poca relevancia

 Variable: 'years_of_residence'

Estadísticas por clase:

- Clase '1': media = 2.84, std = 1.11
- Clase '2': media = 2.85, std = 1.09

 Diferencia porcentual entre clases: 0.25%

 Baja diferencia - Variable con poca relevancia

 Variable: 'age'

Estadísticas por clase:

- Clase '1': media = 36.22, std = 11.38
- Clase '2': media = 33.96, std = 11.22

 Diferencia porcentual entre clases: 6.24%

 Baja diferencia - Variable con poca relevancia

 Variable: 'num_existing_credits'

Estadísticas por clase:

- Clase '1': media = 1.42, std = 0.58
 - Clase '2': media = 1.37, std = 0.56
-  Diferencia porcentual entre clases: 4.05%
-  Baja diferencia - Variable con poca relevancia

 Variable: 'num_dependents'

Estadísticas por clase:

- Clase '1': media = 1.16, std = 0.36
 - Clase '2': media = 1.15, std = 0.36
-  Diferencia porcentual entre clases: 0.21%
-  Baja diferencia - Variable con poca relevancia

8.2 VARIABLES CATEGÓRICAS vs TARGET

 Variable: 'status_checking_account'

Distribución porcentual de target por categoría:

Categoría 'A11':

- Clase '1': 50.73%
- Clase '2': 49.27%

Categoría 'A12':

- Clase '1': 60.97%
- Clase '2': 39.03%

Categoría 'A13':

- Clase '1': 77.78%
- Clase '2': 22.22%

... (1 categorías adicionales no mostradas)

 Variable: 'credit_history'

Distribución porcentual de target por categoría:

Categoría 'A30':

- Clase '1': 37.50%
- Clase '2': 62.50%

Categoría 'A31':

- Clase '1': 42.86%
- Clase '2': 57.14%

Categoría 'A32':

- Clase '1': 68.11%
- Clase '2': 31.89%

... (2 categorías adicionales no mostradas)

 Variable: 'purpose'

Distribución porcentual de target por categoría:

Categoría 'A40':

- Clase '1': 61.97%
- Clase '2': 38.03%

Categoría 'A41':

- Clase '1': 83.50%
- Clase '2': 16.50%

Categoría 'A410':

- Clase '1': 58.33%
- Clase '2': 41.67%

... (7 categorías adicionales no mostradas)

 Variable: 'savings_type'

Distribución porcentual de target por categoría:

Categoría 'A61':

- Clase '1': 64.01%
- Clase '2': 35.99%

Categoría 'A62':

- Clase '1': 66.99%
- Clase '2': 33.01%

Categoría 'A63':

- Clase '1': 82.54%
- Clase '2': 17.46%

... (2 categorías adicionales no mostradas)

 Variable: 'years_of_employment'

Distribución porcentual de target por categoría:

Categoría 'A71':

- Clase '1': 62.90%
- Clase '2': 37.10%

Categoría 'A72':

- Clase '1': 59.30%
- Clase '2': 40.70%

Categoría 'A73':

- Clase '1': 69.32%
- Clase '2': 30.68%

... (2 categorías adicionales no mostradas)

 Variable: 'status_sex'

Distribución porcentual de target por categoría:

Categoría 'A91':

- Clase '1': 60.00%
- Clase '2': 40.00%

Categoría 'A92':

- Clase '1': 64.84%
- Clase '2': 35.16%

Categoría 'A93':

- Clase '1': 73.36%
- Clase '2': 26.64%

... (1 categorías adicionales no mostradas)

 Variable: 'debtors'

Distribución porcentual de target por categoría:

Categoría 'A101':

- Clase '1': 70.01%
- Clase '2': 29.99%

Categoría 'A102':

- Clase '1': 56.10%
- Clase '2': 43.90%

Categoría 'A103':

- Clase '1': 80.77%
- Clase '2': 19.23%

 Variable: 'property'

Distribución porcentual de target por categoría:

Categoría 'A121':

- Clase '1': 78.72%
- Clase '2': 21.28%

Categoría 'A122':

- Clase '1': 69.40%
- Clase '2': 30.60%

Categoría 'A123':

- Clase '1': 69.28%
- Clase '2': 30.72%

... (1 categorías adicionales no mostradas)

 Variable: 'other_installments'

Distribución porcentual de target por categoría:

Categoría 'A141':

- Clase '1': 58.99%
- Clase '2': 41.01%

Categoría 'A142':

- Clase '1': 59.57%
- Clase '2': 40.43%

Categoría 'A143':

- Clase '1': 72.48%
- Clase '2': 27.52%

 Variable: 'housing'

Distribución porcentual de target por categoría:

Categoría 'A151':

- Clase '1': 60.89%
- Clase '2': 39.11%

Categoría 'A152':

- Clase '1': 73.91%
- Clase '2': 26.09%

Categoría 'A153':

- Clase '1': 59.26%
- Clase '2': 40.74%

 Variable: 'job'

Distribución porcentual de target por categoría:

Categoría 'A171':

- Clase '1': 68.18%
- Clase '2': 31.82%

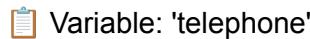
Categoría 'A172':

- Clase '1': 72.00%
- Clase '2': 28.00%

Categoría 'A173':

- Clase '1': 70.48%
- Clase '2': 29.52%

... (1 categorías adicionales no mostradas)



Distribución porcentual de target por categoría:

Categoría 'A191':

- Clase '1': 68.62%
- Clase '2': 31.38%

Categoría 'A192':

- Clase '1': 72.03%
- Clase '2': 27.97%



Distribución porcentual de target por categoría:

Categoría 'A201':

- Clase '1': 69.26%
- Clase '2': 30.74%

Categoría 'A202':

- Clase '1': 89.19%
- Clase '2': 10.81%

=====

9. RESUMEN EJECUTIVO DEL ANÁLISIS

=====



CARACTERÍSTICAS DEL DATASET

- Total de registros: 1,000
- Total de variables: 21
 - Variables numéricas: 7
 - Variables categóricas: 14



CALIDAD DE LOS DATOS

- Valores faltantes: 0 (0.00% del total)
- Registros duplicados: 0 (0.00%)
- Completitud de los datos: 100.00%



VARIABLE OBJETIVO

- Clases: 2
- Ratio de desbalance: 2.33:1

 Dataset significativamente desbalanceado

HALLAZGOS PRINCIPALES

- Se detectaron 326 outliers en variables numéricas
 - Variable con más outliers: 'num_dependents' (15.50%)
- Se encontraron 1 pares de variables con correlación significativa ($|r| > 0.5$)
 - Correlación más fuerte: 'month_credit_duration' ↔ 'credit_amount' ($r=0.625$)

RECOMENDACIONES

1. Considerar técnicas de balanceo (SMOTE, undersampling, class weights)
 2. Analizar outliers y decidir estrategia de tratamiento
 3. Aplicar normalización/estandarización a variables numéricas
 4. Codificar variables categóricas (One-Hot, Label Encoding)
-

ANÁLISIS EXPLORATORIO DE DATOS COMPLETADO
