alarma fraude

	t_id	mensaje_alarma
0	2736446	Detectado_fraude
1	3247297	Detectado_fraude
2	3760288	Detectado_fraude
3	5563713	Detectado_fraude
4	5996407	Detectado_fraude
5	5996409	Detectado_fraude
6	6168499	Detectado_fraude

```
dfaf.dtypes

t_id              int64
mensaje_alarma    object
dtype: object
```

Tipo

	t_id	tipo
0	0	PAYMENT
1	1	PAYMENT
2	4	PAYMENT
3	5	PAYMENT
4	6	PAYMENT
...	...	...
6362615	6362253	CASH_IN
6362616	6362271	CASH_IN
6362617	6362279	CASH_IN
6362618	6362298	CASH_IN
6362619	6362315	CASH_IN

```
dft.dtypes

t_id     int64
tipo     object
dtype: object
```

hacer una sola tabla con lo que queremos unir

Clientes

	t_id	origen	destino
0	0	C1231006815	M1979787155
1	1	C1666544295	M2044282225
2	2	C1305486145	C553264065
3	3	C840083671	C38997010
4	4	C2048537720	M1230701703
5	5	C90045638	M573487274
6	6	C154988899	M408069119

```
x=df.dtypes

x

t_id       int32
origen     object
destino    object
dtype: object
```

Cuantia

	t_id	monto
0	0	9839.64
1	1	1864.28
2	2	181.00
3	3	181.00
4	4	11668.14
5	5	7817.71
6	6	7107.77

```
dfc.dtypes

t_id     int32
monto    float64
dtype: object
```

Es Fraude

	False	True
0	1	0
1	1	0
2	0	1
3	0	1
4	1	0
...	...	...
6362615	0	1
6362616	0	1
6362517	0	1
8362518	0	1
6362519	0	1

6362620 rows × 2 columns

t_id poner

```
y.dtypes

False    uint8
True     uint8
dtype: object
```

Tiempo

	t_id	fecha y hora
0	0	2022-09-01 07:00:00
1	1	2022-09-01 07:00:00
2	2	2022-09-01 07:00:00
3	3	2022-09-01 07:00:00
4	4	2022-09-01 07:00:00
...	...	...
6362615	6362615	2022-10-02 05:00:00
6362616	6362616	2022-10-02 05:00:00
6362617	6362617	2022-10-02 05:00:00
6362618	6362618	2022-10-02 05:00:00
6362619	6362619	2022-10-02 05:00:00

6362620 rows × 2 columns

```
dfinal.dtypes

t_id            int64
fecha y hora    datetime64[ns]
dtype: object
```

Balance

	cliente	t_id	balance_prev	balance_post
0	C1231006815	0	170136.00	160296.36
1	M1979787155	0	0.00	0.00
2	C1666544295	1	21249.00	19384.72
3	M2044202225	1	0.00	0.00
4	C1305486145	2	181.00	0.00
5	C553264065	2	0.00	0.00
6	C553264065	6107	0.00	109905.65
7	C553264065	7903	109905.65	99729.44
8	C553264065	16381	99729.44	5254.44
9	C553264065	20003	211351.84	5254.44

```
tipob=df_balance.dtypes
tipob

cliente         object
t_id            int64
balance_prev    float64
balance_post    float64
dtype: object
```

```
TIPOS DE DATOS
t_id: int32
origen: string | varchar(20)
destino: string/object? | varchar(20)
cuantia: float
fraude: bool (lo cambiamos a dummy?)
tiempo: int32
fecha: string | varchar
fecha: date (SQL)
hora: time (SQL)
tipo: string/object | varchar(10)
balance_prev: float
balance_post: float
```

Transacción

PK	t_id
	origen
	destino
	tipo
	tiempo
	Cuantía

Balances

	Cliente
FK	t_id
	balances_prev
	balances_post

Fraude	tipo de dato (P/SQY)
alarma	dummies
es fraude	bool, dummies
t_id	int

Información de los datos

Cuantía

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6362620 entries, 0 to 6362619
Data columns (total 2 columns):
 #   Column   Dtype
---  ------   -----
 0   t_id     int32
 1   cuantia  float64
dtypes: float64(1), int32(1)
memory usage: 72.8 MB
```

```
#missing values
df.isna().sum()

t_id       0
cuantia    0
dtype: int64
```

```
df.isnull().sum()

t_id       0
cuantia    0
dtype: int64
```

```
#Valores únicos
df.t_id.value_counts()

0           1
4241732     1
4241754     1
4241753     1
4241752     1
            ..
2120309     1
2120368     1
2120367     1
2120366     1
6362619     1
Name: t_id, Length: 6362620, dtype: int64
```

```
#Valores únicos
dfc.monto.value_counts()

12000000.00    1297
12000.00         88
5000.00          79
15000.00         68
520.00           65
                ...
8513.48           1
189137.90         1
1615.78           1
128801.19         1
12257.89          1
Name: monto, Length: 6316000, dtype: int64
```

es fraude

```
dfef.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6362520 entries, 0 to 6362519
Data columns (total 3 columns):
 #   Column        Dtype
---  ------        -----
 0   t_id          int32
 1   Resultado     string
 2   fraude_dummy  int32
dtypes: int32(2), string(1)
memory usage: 97.1 MB
```

```
#missing values
dfef.isna().sum()

t_id            0
Resultado       0
fraude_dummy    0
dtype: int64
```

```
dfef.isnull().sum()

t_id            0
Resultado       0
fraude_dummy    0
dtype: int64
```

```
dfef['Resultado'].unique()

<StringArray>
['False', 'True']
Length: 2, dtype: string
```

alarma fraude

```
dfaf.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   t_id            16 non-null     int32
 1   mensaje_alarma  16 non-null     object
dtypes: int32(1), object(1)
memory usage: 320.0+ bytes
```

```
x=dfaf['mensaje_alarma'].unique()
x

array(['Detectado_fraude'], dtype=object)
```

```
dfaf['t_id']=dfaf['t_id'].astype('int64')
```

```
dfaf.isnull().sum()

t_id              0
mensaje_alarma    0
dtype: int64
```

```
dfaf.isna().sum()

t_id              0
mensaje_alarma    0
dtype: int64
```

```
dfaf.mensaje_alarma.value_counts()

Detectado_fraude    16
Name: mensaje_alarma, dtype: int64
```

Tipo

```
dft.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6362620 entries, 0 to 6362619
Data columns (total 2 columns):
 #   Column  Dtype
---  ------  -----
 0   t_id    int32
 1   tipo    object
dtypes: int32(1), object(1)
memory usage: 72.8+ MB
```

```
df.isna().sum()

t_id    0
tipo    0
dtype: int64
```

```
df.isnull().sum()

t_id    0
tipo    0
dtype: int64
```

```
df.tipo.value_counts()

CASH_OUT    2237500
PAYMENT     2151495
CASH_IN     1399284
TRANSFER     532909
DEBIT         41432
Name: tipo, dtype: int64
```

Balances

```
df_balance.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12725240 entries, 0 to 12725239
Data columns (total 4 columns):
 #   Column        Dtype
---  ------        -----
 0   cliente       object
 1   t_id          int32
 2   balance_prev  float64
 3   balance_post  float64
dtypes: float64(2), int32(1), object(1)
memory usage: 339.8+ MB
```

```
#missing values
df_balance.isna().sum()

cliente         0
t_id            0
balance_prev    0
balance_post    0
dtype: int64
```

```
df_balance.isnull().sum()

cliente         0
t_id            0
balance_prev    0
balance_post    0
dtype: int64
```

```
df_balance.balance_prev.value_counts()

0.00          4806837
184.00            934
133.00            925
164.00            925
109.00            922
               ...
53670.82            1
92985.71            1
65591.57            1
374580.44           1
1153042.12          1
Name: balance_prev, Length: 5303616, dtype: int64
```

```
df_balance.balance_post.value_counts()

0.00          6048999
10000000.00        53
971418.91          32
19169204.93        29
1254956.07         25
               ...
345176.34           1
110979.08           1
216364.13           1
444884.39           1
1929841.08          1
Name: balance_post, Length: 6167874, dtype: int64
```

```
display(df_balance.isna().sum(axis = 0))

cliente         0
t_id            0
balance_prev    0
balance_post    0
dtype: int64
```

Clientes

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6362620 entries, 0 to 6362619
Data columns (total 3 columns):
 #   Column   Dtype
---  ------   -----
 0   t_id     int32
 1   origen   object
 2   destino  object
dtypes: int32(1), object(2)
memory usage: 121.4+ MB
```

```
#missing values
df.isna().sum()

t_id       0
origen     0
destino    0
dtype: int64
```

```
df.isnull().sum()

t_id       0
origen     0
destino    0
dtype: int64
```

```
#Valores únicos
df.t_id.value_counts()

0          1
4241732    1
4241754    1
4241753    1
4241752    1
          ..
2120869    1
2120868    1
2120867    1
2120866    1
6362619    1
Name: t_id, Length: 6362620, dtype: int64
```

```
df.origen.value_counts()

C1902386530    3
C363736674     3
C545315117     3
C724452879     3
C1784010646    3
              ..
C98968405      1
C720209255     1
C1567523029    1
C644777639     1
C1280323807    1
Name: origen, Length: 6353307, dtype: int64
```

```
df.destino.value_counts()

C1286084959    113
C985934102     109
C665576141     105
C2083562754    102
C248609774     101
              ...
M1470027725      1
M1330329251      1
M1784358659      1
M2081431099      1
C2080388513      1
Name: destino, Length: 2722362, dtype: int64
```

En cuantia me sal