

# prueba-tecnica

November 25, 2025

```
[41]: pip install openpyxl
```

```
Requirement already satisfied: openpyxl in /usr/local/lib/python3.11/site-packages (3.1.5)
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.11/site-packages (from openpyxl) (2.0.0)

[notice] A new release of pip is
available: 25.2 -> 25.3
[notice] To update, run:
/usr/local/opt/python@3.11/bin/python3.11 -m pip install --upgrade
```

```
pip
```

Note: you may need to restart the kernel to use updated packages.

```
[1]: #cargar los datos
import pandas as pd
df = pd.read_excel('data/requests.xlsx')
pd.set_option('display.max_columns', None)
```

```
[2]: df.sample(5)
```

```
[2]:      Booking      Request date      Requested by \
109631  100/1109631  2024-04-22 00:00:00  user23@hotelbeds.com
164605  100/1164605  2024-05-30 00:00:00  user23@hotelbeds.com
213557  100/1213557  2024-07-08 00:00:00  user23@hotelbeds.com
128009  100/1128009  2024-05-08 00:00:00  user23@hotelbeds.com
199198  100/1199198  2024-07-05 00:00:00  user23@hotelbeds.com
```

```
      Authorized by Department Currency  Amount Reason \
109631  approver25@hotelbeds.com    Sales     USD   3.92 OTHERS
164605  approver25@hotelbeds.com    Sales     USD   1.24 OTHERS
213557  approver25@hotelbeds.com    Sales     USD   2.84 OTHERS
128009  approver25@hotelbeds.com    Sales     USD   3.00 OTHERS
199198  approver25@hotelbeds.com    Sales     USD   1.47 OTHERS
```

```
      Reason 2      Status CustomerShortname \
109631  SPECIFIC CUSTOMER AGREEMENTS  Applied           CLIENT3
164605  SPECIFIC CUSTOMER AGREEMENTS  Applied           CLIENT1
```

```

213557  SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT1
128009  SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT2
199198  SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT1

```

	CustomerRegion	Amount COMGES in EUR
109631	Region 3	3.659618
164605	Region 1	1.157634
213557	Region 1	2.651356
128009	Region 2	2.800728
199198	Region 1	1.372357

[44]: *#miramos los tipos de datos por columnas y los nulos.  
#Nos llama la atención Request date en tipo object  
df.info()*

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 239400 entries, 0 to 239399
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Booking           239400 non-null   object 
 1   Request date     239400 non-null   object 
 2   Requested by     239400 non-null   object 
 3   Authorized by    239398 non-null   object 
 4   Department        239400 non-null   object 
 5   Currency          239400 non-null   object 
 6   Amount             239400 non-null   float64
 7   Reason             239397 non-null   object 
 8   Reason 2          239316 non-null   object 
 9   Status             239400 non-null   object 
 10  CustomerShortname 239400 non-null   object 
 11  CustomerRegion    239400 non-null   object 
 12  Amount COMGES in EUR 239392 non-null   float64
dtypes: float64(2), object(11)
memory usage: 23.7+ MB

```

[45]: df.columns

```

[45]: Index(['Booking', 'Request date', 'Requested by', 'Authorized by',
       'Department', 'Currency', 'Amount', 'Reason', 'Reason 2', 'Status',
       'CustomerShortname', 'CustomerRegion', 'Amount COMGES in EUR'],
       dtype='object')

```

[46]: df.isnull().sum()

	0
Booking	0
Request date	0
Requested by	0

```

Authorized by      2
Department        0
Currency          0
Amount            0
Reason            3
Reason 2         84
Status            0
CustomerShortname 0
CustomerRegion    0
Amount COMGES in EUR 8
dtype: int64

```

[47]: df.describe().T

```

[47]:              count      mean       std      min    25%  \
Amount        239400.0  1408.656182  184094.647958 -219.9  1.89
Amount COMGES in EUR  239392.0   12.983835    99.618487 -219.9  1.77

                           50%      75%      max
Amount           2.730000  4.580000 38000000.00
Amount COMGES in EUR  2.557998  4.275778 18356.96

```

[48]: df.select\_dtypes(include=['object']).describe().T

```

[48]:      count  unique                      top  freq
Booking        239400  239393                100/1239393     3
Request date   239400    156  2024-05-30 00:00:00  44938
Requested by   239400    134  user23@hotelbeds.com  230627
Authorized by   239398     43  approver25@hotelbeds.com  230912
Department      239400     1                 Sales  239400
Currency        239400    26                 USD  228519
Reason          239397     7                 OTHERS 234622
Reason 2        239316    39  SPECIFIC CUSTOMER AGREEMENTS 234061
Status          239400     5                 Applied 221275
CustomerShortname  239400   687                 CLIENT1 84466
CustomerRegion   239400     5                 Region 1 86068

```

[49]: for column in df.columns:
 print(f'{column}: {df[column].unique()}'")

```

Booking: ['100/1000000' '100/1000001' '100/1000002' ... '100/1239397'
'100/1239398'
'100/1239399']
Request date: [datetime.datetime(2024, 1, 2, 0, 0) datetime.datetime(2024, 1, 3,
0, 0)
datetime.datetime(2024, 1, 4, 0, 0) datetime.datetime(2024, 1, 5, 0, 0)
datetime.datetime(2024, 1, 6, 0, 0) datetime.datetime(2024, 1, 8, 0, 0)
datetime.datetime(2024, 1, 9, 0, 0) datetime.datetime(2024, 1, 10, 0, 0)]

```



```
datetime.datetime(2024, 5, 14, 0, 0) datetime.datetime(2024, 5, 15, 0, 0)
datetime.datetime(2024, 5, 16, 0, 0) datetime.datetime(2024, 5, 17, 0, 0)
datetime.datetime(2024, 5, 18, 0, 0) datetime.datetime(2024, 5, 20, 0, 0)
datetime.datetime(2024, 5, 21, 0, 0) datetime.datetime(2024, 5, 22, 0, 0)
datetime.datetime(2024, 5, 23, 0, 0) datetime.datetime(2024, 5, 24, 0, 0)
datetime.datetime(2024, 5, 27, 0, 0) datetime.datetime(2024, 5, 28, 0, 0)
datetime.datetime(2024, 5, 29, 0, 0) datetime.datetime(2024, 5, 30, 0, 0)
datetime.datetime(2024, 5, 31, 0, 0) datetime.datetime(2024, 6, 3, 0, 0)
datetime.datetime(2024, 6, 4, 0, 0) datetime.datetime(2024, 6, 5, 0, 0)
datetime.datetime(2024, 6, 6, 0, 0) datetime.datetime(2024, 6, 7, 0, 0)
datetime.datetime(2024, 6, 10, 0, 0) datetime.datetime(2024, 6, 11, 0, 0)
datetime.datetime(2024, 6, 12, 0, 0) datetime.datetime(2024, 6, 13, 0, 0)
datetime.datetime(2024, 6, 14, 0, 0) datetime.datetime(2024, 6, 17, 0, 0)
datetime.datetime(2024, 6, 18, 0, 0) datetime.datetime(2024, 6, 19, 0, 0)
datetime.datetime(2024, 6, 20, 0, 0) datetime.datetime(2024, 6, 21, 0, 0)
datetime.datetime(2024, 6, 24, 0, 0) datetime.datetime(2024, 6, 25, 0, 0)
datetime.datetime(2024, 6, 26, 0, 0) datetime.datetime(2024, 6, 27, 0, 0)
datetime.datetime(2024, 6, 28, 0, 0) datetime.datetime(2024, 7, 1, 0, 0)
datetime.datetime(2024, 7, 2, 0, 0) datetime.datetime(2024, 7, 3, 0, 0)
datetime.datetime(2024, 7, 4, 0, 0) datetime.datetime(2024, 7, 5, 0, 0)
datetime.datetime(2024, 7, 8, 0, 0) datetime.datetime(2024, 7, 9, 0, 0)
datetime.datetime(2024, 7, 10, 0, 0) datetime.datetime(2024, 7, 11, 0, 0)
datetime.datetime(2024, 7, 12, 0, 0) datetime.datetime(2024, 7, 13, 0, 0)
datetime.datetime(2024, 7, 15, 0, 0) datetime.datetime(2024, 7, 16, 0, 0)
datetime.datetime(2024, 7, 17, 0, 0) datetime.datetime(2024, 7, 18, 0, 0)
datetime.datetime(2024, 7, 19, 0, 0) datetime.datetime(2024, 7, 22, 0, 0)]
```

Requested by: ['user94@hotelbeds.com' 'user46@hotelbeds.com'  
'user101@hotelbeds.com'  
'user117@hotelbeds.com' 'user39@hotelbeds.com' 'user88@hotelbeds.com'  
'user8@hotelbeds.com' 'user90@hotelbeds.com' 'user11@hotelbeds.com'  
'user9@hotelbeds.com' 'user48@hotelbeds.com' 'user81@hotelbeds.com'  
'user84@hotelbeds.com' 'user69@hotelbeds.com' 'user38@hotelbeds.com'  
'user59@hotelbeds.com' 'user13@hotelbeds.com' 'user118@hotelbeds.com'  
'user69@hotelbeds.com' 'user101@hotelbeds.com' 'user116@hotelbeds.com'  
'user54@hotelbeds.com' 'user104@hotelbeds.com' 'user102@hotelbeds.com'  
'user35@hotelbeds.com' 'user115@hotelbeds.com' 'user44@hotelbeds.com'  
'user56@hotelbeds.com' 'user63@hotelbeds.com' 'user20@hotelbeds.com'  
'user58@hotelbeds.com' 'user92@hotelbeds.com' 'user119@hotelbeds.com'  
'user10@hotelbeds.com' 'user124@hotelbeds.com' 'user29@hotelbeds.com'  
'user88@hotelbeds.com' 'user129@hotelbeds.com' 'user23@hotelbeds.com'  
'user127@hotelbeds.com' 'user125@hotelbeds.com' 'user68@hotelbeds.com'  
'user5@hotelbeds.com' 'user103@hotelbeds.com' 'user76@hotelbeds.com'  
'user66@hotelbeds.com' 'user126@hotelbeds.com' 'user77@hotelbeds.com'  
'user45@hotelbeds.com' 'user2@hotelbeds.com' 'user41@hotelbeds.com'  
'user6@hotelbeds.com' 'user40@hotelbeds.com' 'user82@hotelbeds.com'  
'user18@hotelbeds.com' 'user83@hotelbeds.com' 'user96@hotelbeds.com'  
'user51@hotelbeds.com' 'user80@hotelbeds.com' 'user131@hotelbeds.com'  
'user31@hotelbeds.com' 'user106@hotelbeds.com' 'user109@hotelbeds.com'

'user21@hotelbeds.com' 'user3@hotelbeds.com' 'user50@hotelbeds.com'  
'user89@hotelbeds.com' 'user33@hotelbeds.com' 'user114@hotelbeds.com'  
'user25@hotelbeds.com' 'user67@hotelbeds.com' 'user64@hotelbeds.com'  
'user105@hotelbeds.com' 'user85@hotelbeds.com' 'user108@hotelbeds.com'  
'user17@hotelbeds.com' 'user49@hotelbeds.com' 'user42@hotelbeds.com'  
'user1@hotelbeds.com' 'user14@hotelbeds.com' 'user99@hotelbeds.com'  
'user32@hotelbeds.com' 'user55@hotelbeds.com' 'user65@hotelbeds.com'  
'user24@hotelbeds.com' 'user38@hotelbeds.com' 'user30@hotelbeds.com'  
'user72@hotelbeds.com' 'user78@hotelbeds.com' 'user53@hotelbeds.com'  
'user75@hotelbeds.com' 'user61@hotelbeds.com' 'user97@hotelbeds.com'  
'user98@hotelbeds.com' 'user74@hotelbeds.com' 'user12@hotelbeds.com'  
'user111@hotelbeds.com' 'user26@hotelbeds.com' 'user107@hotelbeds.com'  
'user22@hotelbeds.com' 'user110@hotelbeds.com' 'user47@hotelbeds.com'  
'user43@hotelbeds.com' 'user112@hotelbeds.com' 'user86@hotelbeds.com'  
'user130@hotelbeds.com' 'user19@hotelbeds.com' 'user57@hotelbeds.com'  
'user28@hotelbeds.com' 'user122@hotelbeds.com' 'user100@hotelbeds.com'  
'user87@hotelbeds.com' 'user27@hotelbeds.com' 'user93@hotelbeds.com'  
'user7@hotelbeds.com' 'user95@hotelbeds.com' 'user73@hotelbeds.com'  
'user123@hotelbeds.com' 'user34@hotelbeds.com' 'user113@hotelbeds.com'  
'user62@hotelbeds.com' 'user79@hotelbeds.com' 'user91@hotelbeds.com'  
'user60@hotelbeds.com' 'user15@hotelbeds.com' 'user70@hotelbeds.com'  
'user71@hotelbeds.com' 'user36@hotelbeds.com' 'user37@hotelbeds.com'  
'user121@hotelbeds.com' 'user120@hotelbeds.com' 'user52@hotelbeds.com'  
'user4@hotelbeds.com' 'user16@hotelbeds.com']

Authorized by: ['approver21@hotelbeds.com' 'approver43@hotelbeds.com'  
'approver14@hotelbeds.com' 'approver37@hotelbeds.com'  
'approver17@hotelbeds.com' 'approver35@hotelbeds.com'  
'approver7@hotelbeds.com' 'approver2@hotelbeds.com'  
'approver42@hotelbeds.com' 'approver12@hotelbeds.com'  
'approver31@hotelbeds.com' 'approver9@hotelbeds.com'  
'approver8@hotelbeds.com' 'approver22@hotelbeds.com'  
'approver39@hotelbeds.com' 'approver10@hotelbeds.com'  
'approver28@hotelbeds.com' 'approver4@hotelbeds.com'  
'approver38@hotelbeds.com' 'approver16@hotelbeds.com' nan  
'approver25@hotelbeds.com' 'approver40@hotelbeds.com'  
'approver26@hotelbeds.com' 'approver15@hotelbeds.com'  
'approver29@hotelbeds.com' 'approver20@hotelbeds.com'  
'approver6@hotelbeds.com' 'approver24@hotelbeds.com'  
'approver30@hotelbeds.com' 'approver5@hotelbeds.com'  
'approver34@hotelbeds.com' 'approver19@hotelbeds.com'  
'approver1@hotelbeds.com' 'approver33@hotelbeds.com'  
'approver13@hotelbeds.com' 'approver27@hotelbeds.com'  
'approver32@hotelbeds.com' 'approver23@hotelbeds.com'  
'approver11@hotelbeds.com' 'approver41@hotelbeds.com'  
'approver3@hotelbeds.com' 'approver36@hotelbeds.com'  
'approver18@hotelbeds.com']

Department: ['Sales']

Currency: ['CNY' 'USD' 'AED' 'HKD' 'GBP' 'EUR' 'BRL' 'UDS' 'JPY' 'MXN' 'CYN']

'IDR'  
'AUD' 'KRW' 'NZD' 'ZAR' 'INR' 'KWR' 'THB' 'CAD' 'COP' 'SAR' 'CHF' 'SGD'  
'NOK' 'PHP']  
Amount: [7.482000e+01 4.243650e+02 1.040050e+02 ... 3.456300e+02 7.875000e+01  
7.003421e+06]  
Reason: ['BOOKING\_OPERATIONAL\_ISSUE' 'BOOKING\_TECHNICAL\_ISSUE' 'OTHERS'  
'RATE\_ERROR' 'CANCELLATIONS' 'Operational Issues' 'Others' nan]  
Reason 2: ['RESERVATION\_DISCREPANCIES' 'PRICE\_DISCREPANCY\_ACROSS\_BOOKING\_PROCESS'  
'BOOKOUT - HOTEL SOLD OUT BEFORE ARRIVAL' 'RESERVATION NOT FOUND ON SPOT'  
'PRICE\_MATCH' 'TARIFF\_ERROR\_PROTOCOL' nan 'SPECIFIC\_CUSTOMER AGREEMENTS'  
'TIME\_OUT/GHOST BOOKING' 'BOOKOUT - HOTEL SOLD OUT ON SPOT'  
'HOTEL\_MAPPING\_ERROR' 'CONTRACT\_ERROR' 'ROOM\_MAPPING\_ERROR'  
'BOOKOUT - HOTEL CLOSE BEFORE ARRIVAL' 'CANCELLATION\_WAIVE'  
'BOOKING\_OPERATIONAL\_ISSUE' 'RATE\_ERROR' 'BOOKING\_TECHNICAL\_ISSUE'  
'GIFTS' 'WEB\_ISSUE' 'RESORT\_FEE' 'NATIONALITY\_SUPPLEMENT' 'TARIFF\_ERROR'  
'FRAUD\_BOOKINGS' 'other' 'CANCELLATION\_WITH\_PENALTY'  
'BOOKOUT - HOTEL SOLD OUT ON THE SPOT'  
'BOOKOUT - UPGRADE\_TO\_AVOID\_REJECTION' 'CREDIT\_CARD\_PROCESS\_FAILURE'  
'Brazilian\_Law' 'OTHER' 'RESORT\_FEES' 'RE / TARIFF\_ERROR\_PROTOCOL'  
'Incentive \$150.00mxn x 3 bkngs.' 'OVERBOOKING' 'Relocation\_issue'  
'TE - WEB\_ISSUES' 'Wrong\_type\_description' 'NRF AGREEMENT'  
'GRACE\_PERIOD\_EXTENSION']  
Status: ['Applied' 'Removed' 'Rejected\_by\_Finance' 'On\_hold\_Finance'  
'Requested']  
CustomerShortname: ['CLIENT82' 'CLIENT141' 'CLIENT429' 'CLIENT176' 'CLIENT38'  
'CLIENT30'  
'CLIENT62' 'CLIENT625' 'CLIENT650' 'CLIENT112' 'CLIENT356' 'CLIENT45'  
'CLIENT101' 'CLIENT54' 'CLIENT44' 'CLIENT114' 'CLIENT180' 'CLIENT95'  
'CLIENT43' 'CLIENT17' 'CLIENT322' 'CLIENT166' 'CLIENT420' 'CLIENT21'  
'CLIENT107' 'CLIENT7' 'CLIENT47' 'CLIENT603' 'CLIENT450' 'CLIENT462'  
'CLIENT500' 'CLIENT207' 'CLIENT609' 'CLIENT232' 'CLIENT125' 'CLIENT246'  
'CLIENT142' 'CLIENT280' 'CLIENT678' 'CLIENT381' 'CLIENT402' 'CLIENT239'  
'CLIENT543' 'CLIENT642' 'CLIENT612' 'CLIENT106' 'CLIENT16' 'CLIENT193'  
'CLIENT65' 'CLIENT165' 'CLIENT333' 'CLIENT379' 'CLIENT216' 'CLIENT192'  
'CLIENT383' 'CLIENT463' 'CLIENT327' 'CLIENT15' 'CLIENT14' 'CLIENT13'  
'CLIENT12' 'CLIENT188' 'CLIENT515' 'CLIENT525' 'CLIENT139' 'CLIENT41'  
'CLIENT34' 'CLIENT59' 'CLIENT22' 'CLIENT76' 'CLIENT362' 'CLIENT31'  
'CLIENT49' 'CLIENT164' 'CLIENT251' 'CLIENT338' 'CLIENT479' 'CLIENT120'  
'CLIENT532' 'CLIENT153' 'CLIENT643' 'CLIENT135' 'CLIENT1' 'CLIENT2'  
'CLIENT3' 'CLIENT324' 'CLIENT100' 'CLIENT306' 'CLIENT273' 'CLIENT225'  
'CLIENT79' 'CLIENT417' 'CLIENT386' 'CLIENT219' 'CLIENT385' 'CLIENT623'  
'CLIENT683' 'CLIENT127' 'CLIENT508' 'CLIENT256' 'CLIENT140' 'CLIENT350'  
'CLIENT614' 'CLIENT320' 'CLIENT241' 'CLIENT270' 'CLIENT519' 'CLIENT167'  
'CLIENT640' 'CLIENT248' 'CLIENT178' 'CLIENT78' 'CLIENT48' 'CLIENT131'  
'CLIENT37' 'CLIENT200' 'CLIENT657' 'CLIENT52' 'CLIENT468' 'CLIENT367'  
'CLIENT305' 'CLIENT369' 'CLIENT428' 'CLIENT269' 'CLIENT75' 'CLIENT230'  
'CLIENT228' 'CLIENT304' 'CLIENT364' 'CLIENT19' 'CLIENT8' 'CLIENT91'

'CLIENT529' 'CLIENT499' 'CLIENT202' 'CLIENT119' 'CLIENT303' 'CLIENT93'  
'CLIENT414' 'CLIENT312' 'CLIENT572' 'CLIENT55' 'CLIENT328' 'CLIENT50'  
'CLIENT271' 'CLIENT88' 'CLIENT672' 'CLIENT144' 'CLIENT394' 'CLIENT80'  
'CLIENT64' 'CLIENT42' 'CLIENT108' 'CLIENT11' 'CLIENT210' 'CLIENT588'  
'CLIENT316' 'CLIENT396' 'CLIENT201' 'CLIENT355' 'CLIENT278' 'CLIENT476'  
'CLIENT231' 'CLIENT490' 'CLIENT194' 'CLIENT244' 'CLIENT33' 'CLIENT145'  
'CLIENT325' 'CLIENT400' 'CLIENT284' 'CLIENT289' 'CLIENT465' 'CLIENT35'  
'CLIENT536' 'CLIENT438' 'CLIENT686' 'CLIENT128' 'CLIENT51' 'CLIENT168'  
'CLIENT234' 'CLIENT681' 'CLIENT481' 'CLIENT590' 'CLIENT133' 'CLIENT371'  
'CLIENT453' 'CLIENT291' 'CLIENT434' 'CLIENT71' 'CLIENT477' 'CLIENT589'  
'CLIENT317' 'CLIENT412' 'CLIENT565' 'CLIENT18' 'CLIENT118' 'CLIENT659'  
'CLIENT551' 'CLIENT503' 'CLIENT342' 'CLIENT653' 'CLIENT478' 'CLIENT87'  
'CLIENT375' 'CLIENT86' 'CLIENT105' 'CLIENT656' 'CLIENT687' 'CLIENT639'  
'CLIENT575' 'CLIENT223' 'CLIENT635' 'CLIENT595' 'CLIENT58' 'CLIENT23'  
'CLIENT607' 'CLIENT569' 'CLIENT334' 'CLIENT426' 'CLIENT32' 'CLIENT634'  
'CLIENT387' 'CLIENT208' 'CLIENT518' 'CLIENT357' 'CLIENT242' 'CLIENT598'  
'CLIENT229' 'CLIENT677' 'CLIENT336' 'CLIENT527' 'CLIENT526' 'CLIENT184'  
'CLIENT501' 'CLIENT662' 'CLIENT156' 'CLIENT507' 'CLIENT224' 'CLIENT586'  
'CLIENT213' 'CLIENT172' 'CLIENT583' 'CLIENT68' 'CLIENT83' 'CLIENT123'  
'CLIENT645' 'CLIENT98' 'CLIENT528' 'CLIENT509' 'CLIENT368' 'CLIENT129'  
'CLIENT85' 'CLIENT235' 'CLIENT196' 'CLIENT254' 'CLIENT74' 'CLIENT576'  
'CLIENT685' 'CLIENT456' 'CLIENT443' 'CLIENT437' 'CLIENT363' 'CLIENT148'  
'CLIENT377' 'CLIENT191' 'CLIENT502' 'CLIENT171' 'CLIENT143' 'CLIENT593'  
'CLIENT5' 'CLIENT257' 'CLIENT567' 'CLIENT132' 'CLIENT554' 'CLIENT115'  
'CLIENT424' 'CLIENT70' 'CLIENT81' 'CLIENT204' 'CLIENT498' 'CLIENT622'  
'CLIENT562' 'CLIENT174' 'CLIENT432' 'CLIENT307' 'CLIENT160' 'CLIENT136'  
'CLIENT282' 'CLIENT154' 'CLIENT641' 'CLIENT181' 'CLIENT258' 'CLIENT440'  
'CLIENT268' 'CLIENT117' 'CLIENT337' 'CLIENT584' 'CLIENT116' 'CLIENT299'  
'CLIENT217' 'CLIENT198' 'CLIENT510' 'CLIENT329' 'CLIENT179' 'CLIENT147'  
'CLIENT272' 'CLIENT376' 'CLIENT89' 'CLIENT392' 'CLIENT416' 'CLIENT492'  
'CLIENT505' 'CLIENT599' 'CLIENT473' 'CLIENT485' 'CLIENT28' 'CLIENT157'  
'CLIENT9' 'CLIENT25' 'CLIENT252' 'CLIENT516' 'CLIENT667' 'CLIENT281'  
'CLIENT321' 'CLIENT197' 'CLIENT469' 'CLIENT658' 'CLIENT669' 'CLIENT209'  
'CLIENT348' 'CLIENT137' 'CLIENT496' 'CLIENT540' 'CLIENT452' 'CLIENT162'  
'CLIENT557' 'CLIENT511' 'CLIENT384' 'CLIENT444' 'CLIENT548' 'CLIENT349'  
'CLIENT77' 'CLIENT46' 'CLIENT27' 'CLIENT591' 'CLIENT393' 'CLIENT352'  
'CLIENT676' 'CLIENT522' 'CLIENT247' 'CLIENT29' 'CLIENT568' 'CLIENT286'  
'CLIENT343' 'CLIENT177' 'CLIENT109' 'CLIENT353' 'CLIENT466' 'CLIENT134'  
'CLIENT290' 'CLIENT313' 'CLIENT222' 'CLIENT539' 'CLIENT470' 'CLIENT413'  
'CLIENT110' 'CLIENT514' 'CLIENT582' 'CLIENT103' 'CLIENT195' 'CLIENT573'  
'CLIENT262' 'CLIENT533' 'CLIENT267' 'CLIENT602' 'CLIENT311' 'CLIENT149'  
'CLIENT264' 'CLIENT189' 'CLIENT421' 'CLIENT563' 'CLIENT122' 'CLIENT113'  
'CLIENT561' 'CLIENT183' 'CLIENT626' 'CLIENT552' 'CLIENT441' 'CLIENT493'  
'CLIENT664' 'CLIENT577' 'CLIENT315' 'CLIENT341' 'CLIENT405' 'CLIENT96'  
'CLIENT397' 'CLIENT220' 'CLIENT276' 'CLIENT253' 'CLIENT339' 'CLIENT606'  
'CLIENT435' 'CLIENT285' 'CLIENT629' 'CLIENT408' 'CLIENT419' 'CLIENT390'  
'CLIENT187' 'CLIENT491' 'CLIENT604' 'CLIENT407' 'CLIENT218' 'CLIENT553'  
'CLIENT454' 'CLIENT512' 'CLIENT538' 'CLIENT279' 'CLIENT537' 'CLIENT214'

'CLIENT295' 'CLIENT155' 'CLIENT495' 'CLIENT546' 'CLIENT366' 'CLIENT594'  
'CLIENT297' 'CLIENT480' 'CLIENT648' 'CLIENT126' 'CLIENT600' 'CLIENT158'  
'CLIENT467' 'CLIENT621' 'CLIENT459' 'CLIENT425' 'CLIENT559' 'CLIENT403'  
'CLIENT655' 'CLIENT675' 'CLIENT20' 'CLIENT351' 'CLIENT102' 'CLIENT169'  
'CLIENT130' 'CLIENT67' 'CLIENT332' 'CLIENT326' 'CLIENT517' 'CLIENT447'  
'CLIENT651' 'CLIENT361' 'CLIENT545' 'CLIENT458' 'CLIENT647' 'CLIENT359'  
'CLIENT274' 'CLIENT255' 'CLIENT578' 'CLIENT613' 'CLIENT631' 'CLIENT601'  
'CLIENT84' 'CLIENT475' 'CLIENT238' 'CLIENT345' 'CLIENT121' 'CLIENT523'  
'CLIENT682' 'CLIENT636' 'CLIENT671' 'CLIENT580' 'CLIENT488' 'CLIENT610'  
'CLIENT579' 'CLIENT175' 'CLIENT260' 'CLIENT170' 'CLIENT199' 'CLIENT206'  
'CLIENT550' 'CLIENT464' 'CLIENT138' 'CLIENT53' 'CLIENT185' 'CLIENT233'  
'CLIENT159' 'CLIENT69' 'CLIENT190' 'CLIENT484' 'CLIENT57' 'CLIENT36'  
'CLIENT422' 'CLIENT150' 'CLIENT263' 'CLIENT486' 'CLIENT616' 'CLIENT298'  
'CLIENT354' 'CLIENT451' 'CLIENT555' 'CLIENT124' 'CLIENT265' 'CLIENT679'  
'CLIENT415' 'CLIENT649' 'CLIENT287' 'CLIENT146' 'CLIENT374' 'CLIENT436'  
'CLIENT236' 'CLIENT406' 'CLIENT427' 'CLIENT654' 'CLIENT566' 'CLIENT644'  
'CLIENT300' 'CLIENT237' 'CLIENT592' 'CLIENT627' 'CLIENT524' 'CLIENT161'  
'CLIENT221' 'CLIENT186' 'CLIENT40' 'CLIENT574' 'CLIENT449' 'CLIENT430'  
'CLIENT442' 'CLIENT494' 'CLIENT243' 'CLIENT558' 'CLIENT638' 'CLIENT212'  
'CLIENT240' 'CLIENT556' 'CLIENT530' 'CLIENT398' 'CLIENT391' 'CLIENT596'  
'CLIENT445' 'CLIENT670' 'CLIENT92' 'CLIENT534' 'CLIENT308' 'CLIENT433'  
'CLIENT474' 'CLIENT611' 'CLIENT63' 'CLIENT471' 'CLIENT73' 'CLIENT275'  
'CLIENT205' 'CLIENT482' 'CLIENT104' 'CLIENT294' 'CLIENT292' 'CLIENT277'  
'CLIENT521' 'CLIENT668' 'CLIENT211' 'CLIENT628' 'CLIENT182' 'CLIENT360'  
'CLIENT370' 'CLIENT630' 'CLIENT448' 'CLIENT39' 'CLIENT296' 'CLIENT97'  
'CLIENT302' 'CLIENT373' 'CLIENT506' 'CLIENT163' 'CLIENT585' 'CLIENT457'  
'CLIENT261' 'CLIENT460' 'CLIENT330' 'CLIENT203' 'CLIENT152' 'CLIENT423'  
'CLIENT684' 'CLIENT531' 'CLIENT60' 'CLIENT618' 'CLIENT619' 'CLIENT461'  
'CLIENT665' 'CLIENT399' 'CLIENT660' 'CLIENT358' 'CLIENT227' 'CLIENT504'  
'CLIENT226' 'CLIENT266' 'CLIENT24' 'CLIENT250' 'CLIENT365' 'CLIENT674'  
'CLIENT411' 'CLIENT340' 'CLIENT615' 'CLIENT318' 'CLIENT472' 'CLIENT581'  
'CLIENT646' 'CLIENT410' 'CLIENT293' 'CLIENT446' 'CLIENT388' 'CLIENT513'  
'CLIENT215' 'CLIENT560' 'CLIENT404' 'CLIENT310' 'CLIENT431' 'CLIENT94'  
'CLIENT542' 'CLIENT382' 'CLIENT314' 'CLIENT323' 'CLIENT564' 'CLIENT389'  
'CLIENT633' 'CLIENT663' 'CLIENT347' 'CLIENT680' 'CLIENT346' 'CLIENT547'  
'CLIENT61' 'CLIENT66' 'CLIENT520' 'CLIENT249' 'CLIENT173' 'CLIENT72'  
'CLIENT10' 'CLIENT497' 'CLIENT632' 'CLIENT605' 'CLIENT418' 'CLIENT570'  
'CLIENT544' 'CLIENT111' 'CLIENT608' 'CLIENT335' 'CLIENT99' 'CLIENT380'  
'CLIENT378' 'CLIENT4' 'CLIENT288' 'CLIENT637' 'CLIENT652' 'CLIENT541'  
'CLIENT549' 'CLIENT395' 'CLIENT571' 'CLIENT489' 'CLIENT319' 'CLIENT372'  
'CLIENT26' 'CLIENT483' 'CLIENT401' 'CLIENT439' 'CLIENT409' 'CLIENT673'  
'CLIENT666' 'CLIENT245' 'CLIENT661' 'CLIENT90' 'CLIENT259' 'CLIENT151'  
'CLIENT6' 'CLIENT487' 'CLIENT624' 'CLIENT535' 'CLIENT344' 'CLIENT617'  
'CLIENT331' 'CLIENT301' 'CLIENT455' 'CLIENT587' 'CLIENT597' 'CLIENT56'  
'CLIENT620' 'CLIENT309' 'CLIENT283']

CustomerRegion: ['Region 4' 'Region 1' 'Region 3' 'Region 5' 'Region 2']

Amount COMGES in EUR: [ 9.61313615 54.52390436 13.36292737 ... 235.73182376

53.71027145

399.2810181 ]

[50]: *#Comprobamos si hay reservas con el mismo id y misma cantidad  
#Encontramos dos repetidas con el id 100/1200589, misma cantidad y misma fecha  
#y tres con el id 100/1239393, misma cantidad y misma fecha*

```
duplicates = df[df.duplicated(subset=['Booking', 'Amount'], keep=False)]
print(duplicates.T)
```

Booking	200589	\
Request date	100/1200589	
Requested by	2024-07-05 00:00:00	
Authorized by	user23@hotelbeds.com	
Department	approver25@hotelbeds.com	
Currency	Sales	
Amount	USD	
Reason	9.69	
Reason 2	OTHERS	
Status	SPECIFIC CUSTOMER AGREEMENTS	
CustomerShortname	Applied	
CustomerRegion	CLIENT2	
Amount COMGES in EUR	Region 2	
	9.046352	
Booking	200597	239392 \
Request date	100/1200589	100/1239393
Requested by	2024-07-05 00:00:00	2024-07-19 00:00:00
Authorized by	user23@hotelbeds.com	user72@hotelbeds.com
Department	approver25@hotelbeds.com	approver39@hotelbeds.com
Currency	Sales	Sales
Amount	USD	USD
Reason	9.69	144.54
Reason 2	OTHERS	OTHERS
Status	SPECIFIC CUSTOMER AGREEMENTS	NaN
CustomerShortname	Applied	Requested
CustomerRegion	CLIENT1	CLIENT56
Amount COMGES in EUR	Region 1	Region 4
	4.3598	134.939084
Booking	239393	239394
Request date	100/1239393	100/1239393
Requested by	2024-07-19 00:00:00	2024-07-19 00:00:00
Authorized by	user72@hotelbeds.com	user72@hotelbeds.com
Department	approver39@hotelbeds.com	approver39@hotelbeds.com
Currency	Sales	Sales
Amount	USD	USD
Reason	144.54	144.54
	OTHERS	OTHERS

```

Reason 2           NaN           NaN
Status            Requested      Requested
CustomerShortname CLIENT56      CLIENT56
CustomerRegion    Region 4      Region 4
Amount COMGES in EUR 134.939084 134.939084

```

[51]: df = df.drop\_duplicates(subset=['Booking', 'Amount'], keep='first')

[52]: #comprobamos los duplicados y los borramos  
df.duplicated().sum()

[52]: np.int64(0)

[53]: #renombramos las columnas para que no tengan espacios  
name\_columns = {col: col.replace(" ", "\_") for col in df.columns}  
df = df.rename(columns= name\_columns)

[54]: df.sample(2)

[54]: Booking Request\_date Requested\_by \
216954 100/1216954 2024-07-08 00:00:00 user23@hotelbeds.com
206062 100/1206062 2024-07-08 00:00:00 user23@hotelbeds.com

	Authorized_by	Department	Currency	Amount	Reason	\
216954	approver25@hotelbeds.com	Sales	USD	5.19	OTHERS	
206062	approver25@hotelbeds.com	Sales	USD	1.58	OTHERS	

	Reason_2	Status	CustomerShortname	\
216954	SPECIFIC CUSTOMER AGREEMENTS	Applied	CLIENT3	
206062	SPECIFIC CUSTOMER AGREEMENTS	Applied	CLIENT2	

	CustomerRegion	Amount_COMGES_in_EUR
216954	Region 3	4.84526
206062	Region 2	1.47505

[55]: #comprobamos los nulos en Authorized\_by  
authorized\_null = df[df['Authorized\_by'].isna()]

[56]: print(authorized\_null)

	Booking	Request_date	Requested_by	Authorized_by	\
471	100/1000471	2024-01-16 00:00:00	user88@hotelbeds.com	NaN	
239395	100/1239395	2024-07-19 00:00:00	user16@hotelbeds.com	NaN	

	Department	Currency	Amount	Reason	\
471	Sales	EUR	211.00	BOOKING_OPERATIONAL_ISSUE	
239395	Sales	BRL	1105.44	NaN	

	Reason_2	Status	CustomerShortname	CustomerRegion	\
--	----------	--------	-------------------	----------------	---

471		NaN	Applied	CLIENT532	Region 2
239395	RESERVATION DISCREPANCIES	Requested		CLIENT620	Region 2

	Amount_COMGES_in_EUR
471	211.000000
239395	184.763497

[57]: *#vemos que los dos son de la región 2 y comprobamos si los correos de aprobación se repiten.*

```
authorized_by_region2 = df[df['CustomerRegion'] == "Region 2"]['Authorized_by']
```

[58]: *#No hay un patrón, así que no podemos averiguar cuáles son los faltantes*

```
print(authorized_by_region2.unique())
```

```
['approver14@hotelbeds.com' 'approver21@hotelbeds.com'
 'approver37@hotelbeds.com' 'approver31@hotelbeds.com'
 'approver43@hotelbeds.com' 'approver39@hotelbeds.com'
 'approver10@hotelbeds.com' 'approver4@hotelbeds.com'
 'approver16@hotelbeds.com' 'approver2@hotelbeds.com' nan
 'approver25@hotelbeds.com' 'approver40@hotelbeds.com'
 'approver26@hotelbeds.com' 'approver17@hotelbeds.com'
 'approver15@hotelbeds.com' 'approver7@hotelbeds.com'
 'approver29@hotelbeds.com' 'approver35@hotelbeds.com'
 'approver38@hotelbeds.com' 'approver20@hotelbeds.com'
 'approver42@hotelbeds.com' 'approver30@hotelbeds.com'
 'approver34@hotelbeds.com' 'approver32@hotelbeds.com'
 'approver5@hotelbeds.com' 'approver22@hotelbeds.com'
 'approver23@hotelbeds.com' 'approver28@hotelbeds.com'
 'approver1@hotelbeds.com' 'approver33@hotelbeds.com'
 'approver24@hotelbeds.com' 'approver36@hotelbeds.com'
 'approver19@hotelbeds.com' 'approver18@hotelbeds.com']
```

[59]: *#por danto, decidimos eliminar los nulos*

```
df = df.dropna(subset=['Authorized_by'])
```

[60]: *#Vemos que los correos deberían tener el @ antes de "hotelbeds"*

```
df_requested = df['Requested_by']
print(df_requested)
```

0	user94@hotelbeds.com
1	user94@hotelbeds.com
2	user94@hotelbeds.com
3	user94@hotelbeds.com
4	user94@hotelbeds.com
	...
239392	user72@hotelbeds.com
239396	user97@hotelbeds.com
239397	user7@hotelbeds.com
239398	user7@hotelbeds.com

```
239399    user44@hotelbeds.com
Name: Requested_by, Length: 239395, dtype: object
```

```
[61]: #comprobamos qué correos, tanto de requested_by como de authorized_by no tienen ↴ la estructura deseada
```

```
import re
email_pattern = r'^[a-zA-Z0-9_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+$'
invalid_emails = df[~df['Requested_by'].str.contains(email_pattern, na=False) | ~df['Authorized_by'].str.contains(email_pattern, na=False)]
print(invalid_emails)
```

	Booking	Request_date	Requested_by	\
173	100/1000173	2024-01-06 00:00:00	user69@hotelbeds.com	
174	100/1000174	2024-01-08 00:00:00	user38@hotelbeds.com	
187	100/1000187	2024-01-09 00:00:00	user101@hotelbeds.com	

	Authorized_by	Department	Currency	Amount	\
173	approver43@hotelbeds.com	Sales	CNY	2487.00	
174	approver14@hotelbeds.com	Sales	USD	58.00	
187	approver14@hotelbeds.com	Sales	USD	790.99	

	Reason	Reason_2	\
173	OTHERS	PRICE MATCH	
174	BOOKING_TECHNICAL_ISSUE	PRICE DISCREPANCY ACROSS BOOKING PROCESS	
187	BOOKING_OPERATIONAL_ISSUE	RESERVATION NOT FOUND ON SPOT	

	Status	CustomerShortname	CustomerRegion	Amount_COMGES_in_EUR	
173	Applied	CLIENT47	Region 1	319.538487	
174	Applied	CLIENT603	Region 5	54.147412	
187	Applied	CLIENT280	Region 5	738.449330	

```
[62]: #Creamos una función para sustituir los correos erróneos y añadirles el @
```

```
import re
def change_email(row):
    email_pattern = r'^[a-zA-Z0-9_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+$'
    if not re.match(email_pattern, row):
        if "hotelbeds" in row and "@" not in row:
            return row.replace("hotelbeds", "@hotelbeds")
    return row
```

```
[63]: df['Requested_by'] = df['Requested_by'].apply(change_email)
```

```
[64]: df.columns
```

```
[64]: Index(['Booking', 'Request_date', 'Requested_by', 'Authorized_by',
       'Department', 'Currency', 'Amount', 'Reason', 'Reason_2', 'Status',
       'CustomerShortname', 'CustomerRegion', 'Amount_COMGES_in_EUR'],
```

```

        dtype='object')

[65]: #vemos las divisas y comprobamos que hay errores. De todas formas, vamos a
      ↪sustituirlo todo por EUR y borrar la otra columna de divisa.

df['Currency'].unique()

[65]: array(['CNY', 'USD', 'AED', 'HKD', 'GBP', 'EUR', 'BRL', 'UDS', 'JPY',
       'MXN', 'CYN', 'IDR', 'AUD', 'KRW', 'NZD', 'ZAR', 'INR', 'KWR',
       'THB', 'CAD', 'COP', 'SAR', 'CHF', 'SGD', 'NOK', 'PHP'],
       dtype=object)

[66]: #vamos a cambiar esa columna a euro con un diccionario y una función.

exchange_rates = {
    'CNY': 0.13, 'USD': 0.85, 'AED': 0.23, 'HKD': 0.11, 'GBP': 1.15, 'EUR': 1.0,
    'BRL': 0.17, 'JPY': 0.0064, 'MXN': 0.048, 'IDR': 0.000056, 'AUD': 0.63,
    'KRW': 0.00066, 'NZD': 0.59, 'ZAR': 0.049, 'INR': 0.012, 'THB': 0.025,
    'CAD': 0.63, 'COP': 0.00022, 'SAR': 0.23, 'CHF': 0.92, 'SGD': 0.63,
    'NOK': 0.085, 'PHP': 0.015
}

[67]: def convert_to_eur(row):
    currency = row['Currency']
    amount = row['Amount']
    if currency in exchange_rates:
        return amount * exchange_rates[currency]
    return amount

[68]: df['Amount'] = df.apply(convert_to_eur, axis=1)

[69]: df['Currency'] = "EUR"

[70]: df.sample(5)

[70]: Booking Request_date Requested_by \
146063 100/1146063 2024-05-30 00:00:00 user23@hotelbeds.com
41431 100/1041431 2024-02-23 00:00:00 user23@hotelbeds.com
91126 100/1091126 2024-03-21 00:00:00 user23@hotelbeds.com
1650 100/1001650 2024-01-16 00:00:00 user23@hotelbeds.com
24366 100/1024366 2024-01-16 00:00:00 user23@hotelbeds.com

          Authorized_by Department Currency Amount Reason \
146063 approver25@hotelbeds.com Sales EUR 5.5420 OTHERS
41431 approver25@hotelbeds.com Sales EUR 3.1195 OTHERS
91126 approver25@hotelbeds.com Sales EUR 3.1960 OTHERS
1650 approver25@hotelbeds.com Sales EUR 2.3205 OTHERS
24366 approver25@hotelbeds.com Sales EUR 8.5900 OTHERS

```

```

                    Reason_2    Status CustomerShortname \
146063  SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT2
41431   SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT3
91126   SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT3
1650    SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT1
24366   SPECIFIC CUSTOMER AGREEMENTS  Applied      CLIENT3

CustomerRegion  Amount_COMGES_in_EUR
146063        Region 2          6.086916
41431         Region 3          3.426224
91126        Region 3          3.510246
1650         Region 1          2.548663
24366        Region 3          8.590000

```

[71]: *#vemos cuántos nulos hay en Reason*  
`df['Reason'].isnull().sum()`

[71]: `np.int64(2)`

[72]: `df['Reason'].unique()`

[72]: `array(['BOOKING_OPERATIONAL_ISSUE', 'BOOKING_TECHNICAL_ISSUE', 'OTHERS',
 'RATE_ERROR', 'CANCELLATIONS', 'Operational Issues', 'Others', nan],
 dtype=object)`

[73]: *#comprobamos cuántas razones hay de cada tipo*  
`cancel_count = df['Reason'].value_counts()`  
`print(cancel_count)`

```

Reason
OTHERS                  234619
BOOKING_OPERATIONAL_ISSUE    2538
BOOKING_TECHNICAL_ISSUE     1408
CANCELLATIONS              429
RATE_ERROR                 397
Operational Issues           1
Others                      1
Name: count, dtype: int64

```

[74]: *#agrupamos por departamento, region y cliente, por si encontramos algún patrón*  
`grouped_by_department = df.groupby(['Department', 'Reason']).size()`.  
 `↪reset_index(name='Counts')`  
`grouped_by_region = df.groupby(['CustomerRegion', 'Reason']).size()`.  
 `↪reset_index(name='Counts')`  
`grouped_by_shortname = df.groupby(['CustomerShortname', 'Reason']).size()`.  
 `↪reset_index(name='Counts')`

*#.size cuenta el total de filas creadas por la agrupación*

```
[75]: print("Cancelaciones agrupadas por Department y Reason:")
print(grouped_by_department)

print("\nCancelaciones agrupadas por CustomerRegion y Reason:")
print(grouped_by_region)

print("\nCancelaciones agrupadas por CustomerShortname y Reason:")
print(grouped_by_shortname)
```

Cancelaciones agrupadas por Department y Reason:

	Department	Reason	Counts
0	Sales	BOOKING_OPERATIONAL_ISSUE	2538
1	Sales	BOOKING_TECHNICAL_ISSUE	1408
2	Sales	CANCELLATIONS	429
3	Sales	OTHERS	234619
4	Sales	Operational Issues	1
5	Sales	Others	1
6	Sales	RATE_ERROR	397

Cancelaciones agrupadas por CustomerRegion y Reason:

	CustomerRegion	Reason	Counts
0	Region 1	BOOKING_OPERATIONAL_ISSUE	331
1	Region 1	BOOKING_TECHNICAL_ISSUE	728
2	Region 1	CANCELLATIONS	333
3	Region 1	OTHERS	84592
4	Region 1	RATE_ERROR	83
5	Region 2	BOOKING_OPERATIONAL_ISSUE	894
6	Region 2	BOOKING_TECHNICAL_ISSUE	148
7	Region 2	CANCELLATIONS	42
8	Region 2	OTHERS	68734
9	Region 2	Others	1
10	Region 2	RATE_ERROR	76
11	Region 3	BOOKING_OPERATIONAL_ISSUE	381
12	Region 3	BOOKING_TECHNICAL_ISSUE	68
13	Region 3	CANCELLATIONS	23
14	Region 3	OTHERS	64254
15	Region 3	RATE_ERROR	51
16	Region 4	BOOKING_OPERATIONAL_ISSUE	644
17	Region 4	BOOKING_TECHNICAL_ISSUE	77
18	Region 4	CANCELLATIONS	17
19	Region 4	OTHERS	14281
20	Region 4	Operational Issues	1
21	Region 4	RATE_ERROR	101
22	Region 5	BOOKING_OPERATIONAL_ISSUE	288
23	Region 5	BOOKING_TECHNICAL_ISSUE	387
24	Region 5	CANCELLATIONS	14

```

25      Region 5          OTHERS    2758
26      Region 5          RATE_ERROR    86

```

Cancelaciones agrupadas por CustomerShortname y Reason:

	CustomerShortname	Reason	Counts
0	CLIENT1	OTHERS	84465
1	CLIENT10	BOOKING_TECHNICAL_ISSUE	335
2	CLIENT100	RATE_ERROR	6
3	CLIENT101	BOOKING_OPERATIONAL_ISSUE	3
4	CLIENT101	BOOKING_TECHNICAL_ISSUE	1
..	..	..	..
810	CLIENT97	BOOKING_TECHNICAL_ISSUE	1
811	CLIENT98	BOOKING_OPERATIONAL_ISSUE	4
812	CLIENT98	BOOKING_TECHNICAL_ISSUE	1
813	CLIENT98	OTHERS	1
814	CLIENT99	CANCELLATIONS	6

[815 rows x 3 columns]

```
[76]: #mostramos solo los nulos y, efectivamente, en ambos casos el cliente y el ↴authorized_by es el mismo
nuls = df[df['Reason'].isna()]
```

```
[77]: print(nuls)
```

	Booking	Request_date	Requested_by			
96323	100/1096323	2024-04-11 00:00:00	user109@hotelbeds.com			
96324	100/1096324	2024-04-11 00:00:00	user109@hotelbeds.com			
	Authorized_by	Department	Currency	Amount	Reason	
96323	approver30@hotelbeds.com	Sales	EUR	121.70	Nan	
96324	approver30@hotelbeds.com	Sales	EUR	217.36	Nan	
	Reason_2	Status	CustomerShortname	CustomerRegion		
96323	CANCELLATION	WAIVE	Applied	CLIENT113	Region 4	
96324	CANCELLATION	WAIVE	Applied	CLIENT113	Region 4	
	Amount_COMGES_in_EUR					
96323		121.70				
96324		217.36				

```
[78]: client113 = df[df['CustomerShortname'] == "CLIENT113"]
```

```
[79]:
```

```
#Revisamos que todas las entradas con CLIENT113 que tienen la razon ↴
↳ "cancellation waive", tienen como reason "CANCELLATIONS" y la única que no ↴
↳ es así, tiene reason "Booking_operational_issue". Podríamos sustituir los ↴
↳ nulos por "CANCELLATIONS", que es la opción más probable, o eliminar esas ↴
↳ filas o consultar con el cliente. Decidimos sustituir los nulos por ↴
↳ "CANCELLATIONS"
```

```
print(client113)
```

	Booking	Request_date	Requested_by	\
96322	100/1096322	2024-04-11 00:00:00	user109@hotelbeds.com	
96323	100/1096323	2024-04-11 00:00:00	user109@hotelbeds.com	
96324	100/1096324	2024-04-11 00:00:00	user109@hotelbeds.com	
96325	100/1096325	2024-04-11 00:00:00	user109@hotelbeds.com	
96326	100/1096326	2024-04-11 00:00:00	user109@hotelbeds.com	

	Authorized_by	Department	Currency	Amount	\
96322	approver30@hotelbeds.com	Sales	EUR	146.76	
96323	approver30@hotelbeds.com	Sales	EUR	121.70	
96324	approver30@hotelbeds.com	Sales	EUR	217.36	
96325	approver30@hotelbeds.com	Sales	EUR	343.22	
96326	approver30@hotelbeds.com	Sales	EUR	149.93	

	Reason	Reason_2	Status	\
96322	BOOKING_OPERATIONAL_ISSUE	RESERVATION DISCREPANCIES	Applied	
96323	NaN	CANCELLATION WAIVE	Applied	
96324	NaN	CANCELLATION WAIVE	Applied	
96325	CANCELLATIONS	CANCELLATION WAIVE	Applied	
96326	CANCELLATIONS	CANCELLATION WAIVE	Applied	

	CustomerShortname	CustomerRegion	Amount_COMGES_in_EUR	
96322	CLIENT113	Region 4	146.76	
96323	CLIENT113	Region 4	121.70	
96324	CLIENT113	Region 4	217.36	
96325	CLIENT113	Region 4	343.22	
96326	CLIENT113	Region 4	149.93	

```
[80]: df['Reason'] = df['Reason'].fillna("CANCELLATIONS")
```

```
[81]: df['Reason'].isnull().sum()
```

```
[81]: np.int64(0)
```

```
[82]: #De la primera razón ya no hay nulos, en la siguiente hay 82, pero son ↴
↳ optativos.
#Aun así, los sustituimos por UNKNOWN para evitar errores.
df['Reason_2'].isnull().sum()
```

```
[82]: np.int64(81)

[83]: df['Reason_2'] = df['Reason_2'].fillna("Unknown")

[84]: df['Reason_2'].unique()

[84]: array(['RESERVATION DISCREPANCIES',
       'PRICE DISCREPANCY ACROSS BOOKING PROCESS',
       'BOOKOUT - HOTEL SOLD OUT BEFORE ARRIVAL',
       'RESERVATION NOT FOUND ON SPOT', 'PRICE MATCH',
       'TARIFF ERROR PROTOCOL', 'Unknown', 'SPECIFIC CUSTOMER AGREEMENTS',
       'TIME OUT/GHOST BOOKING', 'BOOKOUT - HOTEL SOLD OUT ON SPOT',
       'HOTEL MAPPING ERROR', 'CONTRACT ERROR', 'ROOM MAPPING ERROR',
       'BOOKOUT - HOTEL CLOSE BEFORE ARRIVAL', 'CANCELLATION WAIVE',
       'BOOKING OPERATIONAL ISSUE', 'RATE ERROR',
       'BOOKING TECHNICAL ISSUE', 'GIFTS', 'WEB ISSUE', 'RESORT FEE',
       'NATIONALITY SUPPLEMENT', 'TARIFF ERROR', 'FRAUD BOOKINGS',
       'other', 'CANCELLATION WITH PENALTY ',
       'BOOKOUT - HOTEL SOLD OUT ON THE SPOT',
       'BOOKOUT - UPGRADE TO AVOID REJECTION',
       'CREDIT CARD PROCESS FAILURE', 'Brazilian Law', 'OTHER',
       'RESORT FEES', 'RE / TARIFF ERROR PROTOCOL',
       'Incentive $150.00mxn x 3 bkngs.', 'OVERBOOKING ',
       'Relocation issue', 'TE - WEB ISSUES', 'Wrong type description',
       'NRF AGREEMENT', 'GRACE PERIOD EXTENSION'], dtype=object)
```

```
[85]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 239395 entries, 0 to 239399
Data columns (total 13 columns):
 #  Column           Non-Null Count  Dtype  
--- 
 0  Booking          239395 non-null   object 
 1  Request_date     239395 non-null   object 
 2  Requested_by    239395 non-null   object 
 3  Authorized_by   239395 non-null   object 
 4  Department       239395 non-null   object 
 5  Currency         239395 non-null   object 
 6  Amount           239395 non-null   float64
 7  Reason           239395 non-null   object 
 8  Reason_2         239395 non-null   object 
 9  Status            239395 non-null   object 
 10 CustomerShortname 239395 non-null   object 
 11 CustomerRegion    239395 non-null   object 
 12 Amount_COMGES_in_EUR 239387 non-null   float64
dtypes: float64(2), object(11)
memory usage: 25.6+ MB
```

```
[86]: df.isnull().sum()
```

```
[86]: Booking          0  
Request_date       0  
Requested_by        0  
Authorized_by       0  
Department         0  
Currency           0  
Amount              0  
Reason              0  
Reason_2            0  
Status              0  
CustomerShortname   0  
CustomerRegion      0  
Amount_COMGES_in_EUR 8  
dtype: int64
```

```
[87]: #pasamos la columna de date a formato fecha  
df['Request_date'] = pd.to_datetime(df['Request_date'])
```

```
[88]: nuls = df[df['Amount_COMGES_in_EUR'].isnull()]  
print(nuls)
```

	Booking	Request_date	Requested_by	\\
34	100/1000034	2024-01-04	user11@hotelbeds.com	
200	100/1000200	2024-01-10	user94@hotelbeds.com	
31812	100/1031812	2024-01-31	user119@hotelbeds.com	
82259	100/1082259	2024-03-21	user23@hotelbeds.com	
82260	100/1082260	2024-03-21	user23@hotelbeds.com	
82261	100/1082261	2024-03-21	user23@hotelbeds.com	
82262	100/1082262	2024-03-21	user23@hotelbeds.com	
185340	100/1185340	2024-05-30	user23@hotelbeds.com	

	Authorized_by	Department	Currency	Amount	\\
34	approver21@hotelbeds.com	Sales	EUR	9.44	
200	approver21@hotelbeds.com	Sales	EUR	84.29	
31812	approver38@hotelbeds.com	Sales	EUR	302609.00	
82259	approver25@hotelbeds.com	Sales	EUR	1.51	
82260	approver25@hotelbeds.com	Sales	EUR	4.03	
82261	approver25@hotelbeds.com	Sales	EUR	2.42	
82262	approver25@hotelbeds.com	Sales	EUR	13.47	
185340	approver25@hotelbeds.com	Sales	EUR	5.17	

	Reason	Reason_2	\\
34	BOOKING_OPERATIONAL_ISSUE	RESERVATION NOT FOUND ON SPOT	
200	BOOKING_OPERATIONAL_ISSUE	RESERVATION DISCREPANCIES	
31812	BOOKING_OPERATIONAL_ISSUE	BOOKOUT - HOTEL SOLD OUT BEFORE ARRIVAL	
82259	OTHERS	SPECIFIC CUSTOMER AGREEMENTS	

```

82260          OTHERS      SPECIFIC CUSTOMER AGREEMENTS
82261          OTHERS      SPECIFIC CUSTOMER AGREEMENTS
82262          OTHERS      SPECIFIC CUSTOMER AGREEMENTS
185340         OTHERS      SPECIFIC CUSTOMER AGREEMENTS

```

	Status	CustomerShortname	CustomerRegion	Amount_COMGES_in_EUR
34	Applied	CLIENT44	Region 3	NaN
200	Applied	CLIENT106	Region 2	NaN
31812	Applied	CLIENT355	Region 2	NaN
82259	Applied	CLIENT1	Region 1	NaN
82260	Applied	CLIENT1	Region 1	NaN
82261	Applied	CLIENT3	Region 3	NaN
82262	Applied	CLIENT2	Region 2	NaN
185340	Applied	CLIENT2	Region 2	NaN

```
[89]: df = df.drop('Amount_COMGES_in_EUR', axis=1)
```

```
[90]: df.head()
```

```

[90]:      Booking Request_date      Requested_by      Authorized_by \
0  100/1000000  2024-01-02  user94@hotelbeds.com  approver21@hotelbeds.com
1  100/1000001  2024-01-02  user94@hotelbeds.com  approver21@hotelbeds.com
2  100/1000002  2024-01-02  user94@hotelbeds.com  approver21@hotelbeds.com
3  100/1000003  2024-01-02  user94@hotelbeds.com  approver21@hotelbeds.com
4  100/1000004  2024-01-02  user94@hotelbeds.com  approver21@hotelbeds.com

```

	Department	Currency	Amount	Reason	\
0	Sales	EUR	9.72660	BOOKING_OPERATIONAL_ISSUE	
1	Sales	EUR	55.16745	BOOKING_OPERATIONAL_ISSUE	
2	Sales	EUR	13.52065	BOOKING_TECHNICAL_ISSUE	
3	Sales	EUR	5.07650	BOOKING_OPERATIONAL_ISSUE	
4	Sales	EUR	127.58980	BOOKING_OPERATIONAL_ISSUE	

	Reason_2	Status	CustomerShortname	\
0	RESERVATION DISCREPANCIES	Applied	CLIENT82	
1	RESERVATION DISCREPANCIES	Applied	CLIENT141	
2	PRICE DISCREPANCY ACROSS BOOKING PROCESS	Applied	CLIENT141	
3	BOOKOUT - HOTEL SOLD OUT BEFORE ARRIVAL	Applied	CLIENT141	
4	RESERVATION NOT FOUND ON SPOT	Applied	CLIENT82	

	CustomerRegion
0	Region 4
1	Region 1
2	Region 1
3	Region 1
4	Region 4

```
[91]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 239395 entries, 0 to 239399
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Booking          239395 non-null   object  
 1   Request_date     239395 non-null   datetime64[ns]
 2   Requested_by     239395 non-null   object  
 3   Authorized_by    239395 non-null   object  
 4   Department        239395 non-null   object  
 5   Currency          239395 non-null   object  
 6   Amount            239395 non-null   float64 
 7   Reason            239395 non-null   object  
 8   Reason_2          239395 non-null   object  
 9   Status             239395 non-null   object  
 10  CustomerShortname 239395 non-null   object  
 11  CustomerRegion    239395 non-null   object  
dtypes: datetime64[ns](1), float64(1), object(10)
memory usage: 23.7+ MB

```

[92]: `#Hacemos limpieza de columnas`

[93]: `df['Reason'].unique()`

[93]: `array(['BOOKING_OPERATIONAL_ISSUE', 'BOOKING_TECHNICAL_ISSUE', 'OTHERS',
 'RATE_ERROR', 'CANCELLATIONS', 'Operational Issues', 'Others'],
 dtype=object)`

[94]: `df['Reason'] = df['Reason'].str.title()`

[95]: `df['Reason'].unique()`

[95]: `array(['Booking_Operational_Issue', 'Booking_Technical_Issue', 'Others',
 'Rate_Error', 'Cancellations', 'Operational_Issues'], dtype=object)`

[96]: `for column in df.columns:
 print(df[column].unique())`

```

['100/1000000' '100/1000001' '100/1000002' ... '100/1239397' '100/1239398'
 '100/1239399']
<DatetimeArray>
['2024-01-02 00:00:00', '2024-01-03 00:00:00', '2024-01-04 00:00:00',
 '2024-01-05 00:00:00', '2024-01-06 00:00:00', '2024-01-08 00:00:00',
 '2024-01-09 00:00:00', '2024-01-10 00:00:00', '2024-01-11 00:00:00',
 '2024-01-12 00:00:00',
 ...
 '2024-07-10 00:00:00', '2024-07-11 00:00:00', '2024-07-12 00:00:00',
 '2024-07-13 00:00:00', '2024-07-15 00:00:00', '2024-07-16 00:00:00',
 '2024-07-17 00:00:00', '2024-07-18 00:00:00', '2024-07-19 00:00:00'],

```

```
'2024-07-22 00:00:00']  
Length: 156, dtype: datetime64[ns]  
['user94@hotelbeds.com' 'user46@hotelbeds.com' 'user101@hotelbeds.com'  
'user117@hotelbeds.com' 'user39@hotelbeds.com' 'user88@hotelbeds.com'  
'user8@hotelbeds.com' 'user90@hotelbeds.com' 'user11@hotelbeds.com'  
'user9@hotelbeds.com' 'user48@hotelbeds.com' 'user81@hotelbeds.com'  
'user84@hotelbeds.com' 'user69@hotelbeds.com' 'user38@hotelbeds.com'  
'user59@hotelbeds.com' 'user13@hotelbeds.com' 'user118@hotelbeds.com'  
'user116@hotelbeds.com' 'user54@hotelbeds.com' 'user104@hotelbeds.com'  
'user102@hotelbeds.com' 'user35@hotelbeds.com' 'user115@hotelbeds.com'  
'user44@hotelbeds.com' 'user56@hotelbeds.com' 'user63@hotelbeds.com'  
'user20@hotelbeds.com' 'user58@hotelbeds.com' 'user92@hotelbeds.com'  
'user119@hotelbeds.com' 'user10@hotelbeds.com' 'user124@hotelbeds.com'  
'user29@hotelbeds.com' 'user129@hotelbeds.com' 'user23@hotelbeds.com'  
'user127@hotelbeds.com' 'user125@hotelbeds.com' 'user68@hotelbeds.com'  
'user5@hotelbeds.com' 'user103@hotelbeds.com' 'user76@hotelbeds.com'  
'user66@hotelbeds.com' 'user126@hotelbeds.com' 'user77@hotelbeds.com'  
'user45@hotelbeds.com' 'user2@hotelbeds.com' 'user41@hotelbeds.com'  
'user6@hotelbeds.com' 'user40@hotelbeds.com' 'user82@hotelbeds.com'  
'user18@hotelbeds.com' 'user83@hotelbeds.com' 'user96@hotelbeds.com'  
'user51@hotelbeds.com' 'user80@hotelbeds.com' 'user131@hotelbeds.com'  
'user31@hotelbeds.com' 'user106@hotelbeds.com' 'user109@hotelbeds.com'  
'user21@hotelbeds.com' 'user3@hotelbeds.com' 'user50@hotelbeds.com'  
'user89@hotelbeds.com' 'user33@hotelbeds.com' 'user114@hotelbeds.com'  
'user25@hotelbeds.com' 'user67@hotelbeds.com' 'user64@hotelbeds.com'  
'user105@hotelbeds.com' 'user85@hotelbeds.com' 'user108@hotelbeds.com'  
'user17@hotelbeds.com' 'user49@hotelbeds.com' 'user42@hotelbeds.com'  
'user1@hotelbeds.com' 'user14@hotelbeds.com' 'user99@hotelbeds.com'  
'user32@hotelbeds.com' 'user55@hotelbeds.com' 'user65@hotelbeds.com'  
'user24@hotelbeds.com' 'user30@hotelbeds.com' 'user72@hotelbeds.com'  
'user78@hotelbeds.com' 'user53@hotelbeds.com' 'user75@hotelbeds.com'  
'user61@hotelbeds.com' 'user97@hotelbeds.com' 'user98@hotelbeds.com'  
'user74@hotelbeds.com' 'user12@hotelbeds.com' 'user111@hotelbeds.com'  
'user26@hotelbeds.com' 'user107@hotelbeds.com' 'user22@hotelbeds.com'  
'user110@hotelbeds.com' 'user47@hotelbeds.com' 'user43@hotelbeds.com'  
'user112@hotelbeds.com' 'user86@hotelbeds.com' 'user130@hotelbeds.com'  
'user19@hotelbeds.com' 'user57@hotelbeds.com' 'user28@hotelbeds.com'  
'user122@hotelbeds.com' 'user100@hotelbeds.com' 'user87@hotelbeds.com'  
'user27@hotelbeds.com' 'user93@hotelbeds.com' 'user7@hotelbeds.com'  
'user95@hotelbeds.com' 'user73@hotelbeds.com' 'user123@hotelbeds.com'  
'user34@hotelbeds.com' 'user113@hotelbeds.com' 'user62@hotelbeds.com'  
'user79@hotelbeds.com' 'user91@hotelbeds.com' 'user60@hotelbeds.com'  
'user15@hotelbeds.com' 'user70@hotelbeds.com' 'user71@hotelbeds.com'  
'user36@hotelbeds.com' 'user37@hotelbeds.com' 'user121@hotelbeds.com'  
'user120@hotelbeds.com' 'user52@hotelbeds.com' 'user4@hotelbeds.com'  
'user16@hotelbeds.com']  
['approver21@hotelbeds.com' 'approver43@hotelbeds.com'  
'approver14@hotelbeds.com' 'approver37@hotelbeds.com']
```

```

'approver17@hotelbeds.com' 'approver35@hotelbeds.com'
'approver7@hotelbeds.com' 'approver2@hotelbeds.com'
'approver42@hotelbeds.com' 'approver12@hotelbeds.com'
'approver31@hotelbeds.com' 'approver9@hotelbeds.com'
'approver8@hotelbeds.com' 'approver22@hotelbeds.com'
'approver39@hotelbeds.com' 'approver10@hotelbeds.com'
'approver28@hotelbeds.com' 'approver4@hotelbeds.com'
'approver38@hotelbeds.com' 'approver16@hotelbeds.com'
'approver25@hotelbeds.com' 'approver40@hotelbeds.com'
'approver26@hotelbeds.com' 'approver15@hotelbeds.com'
'approver29@hotelbeds.com' 'approver20@hotelbeds.com'
'approver6@hotelbeds.com' 'approver24@hotelbeds.com'
'approver30@hotelbeds.com' 'approver5@hotelbeds.com'
'approver34@hotelbeds.com' 'approver19@hotelbeds.com'
'approver1@hotelbeds.com' 'approver33@hotelbeds.com'
'approver13@hotelbeds.com' 'approver27@hotelbeds.com'
'approver32@hotelbeds.com' 'approver23@hotelbeds.com'
'approver11@hotelbeds.com' 'approver41@hotelbeds.com'
'approver3@hotelbeds.com' 'approver36@hotelbeds.com'
'approver18@hotelbeds.com']

['Sales']
['EUR']
[ 9.7266 55.16745 13.52065 ... 217.7469 49.6125 392.191576]
['Booking_Operational_Issue' 'Booking_Technical_Issue' 'Others'
 'Rate_Error' 'Cancellations' 'Operational Issues']
['RESERVATION DISCREPANCIES' 'PRICE DISCREPANCY ACROSS BOOKING PROCESS'
 'BOOKOUT - HOTEL SOLD OUT BEFORE ARRIVAL' 'RESERVATION NOT FOUND ON SPOT'
 'PRICE MATCH' 'TARIFF ERROR PROTOCOL' 'Unknown'
 'SPECIFIC CUSTOMER AGREEMENTS' 'TIME OUT/GHOST BOOKING'
 'BOOKOUT - HOTEL SOLD OUT ON SPOT' 'HOTEL MAPPING ERROR' 'CONTRACT ERROR'
 'ROOM MAPPING ERROR' 'BOOKOUT - HOTEL CLOSE BEFORE ARRIVAL'
 'CANCELLATION WAIVE' 'BOOKING OPERATIONAL ISSUE' 'RATE ERROR'
 'BOOKING TECHNICAL ISSUE' 'GIFTS' 'WEB ISSUE' 'RESORT FEE'
 'NATIONALITY SUPPLEMENT' 'TARIFF ERROR' 'FRAUD BOOKINGS' 'other'
 'CANCELLATION WITH PENALTY' 'BOOKOUT - HOTEL SOLD OUT ON THE SPOT'
 'BOOKOUT - UPGRADE TO AVOID REJECTION' 'CREDIT CARD PROCESS FAILURE'
 'Brazilian Law' 'OTHER' 'RESORT FEES' 'RE / TARIFF ERROR PROTOCOL'
 'Incentive $150.00mxn x 3 bkngs.' 'OVERBOOKING' 'Relocation issue'
 'TE - WEB ISSUES' 'Wrong type description' 'NRF AGREEMENT'
 'GRACE PERIOD EXTENSION']

['Applied' 'Removed' 'Rejected by Finance' 'On hold Finance' 'Requested']
['CLIENT82' 'CLIENT141' 'CLIENT429' 'CLIENT176' 'CLIENT38' 'CLIENT30'
 'CLIENT62' 'CLIENT625' 'CLIENT650' 'CLIENT112' 'CLIENT356' 'CLIENT45'
 'CLIENT101' 'CLIENT54' 'CLIENT44' 'CLIENT114' 'CLIENT180' 'CLIENT95'
 'CLIENT43' 'CLIENT17' 'CLIENT322' 'CLIENT166' 'CLIENT420' 'CLIENT21'
 'CLIENT107' 'CLIENT7' 'CLIENT47' 'CLIENT603' 'CLIENT450' 'CLIENT462'
 'CLIENT500' 'CLIENT207' 'CLIENT609' 'CLIENT232' 'CLIENT125' 'CLIENT246'
 'CLIENT142' 'CLIENT280' 'CLIENT678' 'CLIENT381' 'CLIENT402' 'CLIENT239'

```

'CLIENT543' 'CLIENT642' 'CLIENT612' 'CLIENT106' 'CLIENT16' 'CLIENT193'  
'CLIENT65' 'CLIENT165' 'CLIENT333' 'CLIENT379' 'CLIENT216' 'CLIENT192'  
'CLIENT383' 'CLIENT463' 'CLIENT327' 'CLIENT15' 'CLIENT14' 'CLIENT13'  
'CLIENT12' 'CLIENT188' 'CLIENT515' 'CLIENT525' 'CLIENT139' 'CLIENT41'  
'CLIENT34' 'CLIENT59' 'CLIENT22' 'CLIENT76' 'CLIENT362' 'CLIENT31'  
'CLIENT49' 'CLIENT164' 'CLIENT251' 'CLIENT338' 'CLIENT479' 'CLIENT120'  
'CLIENT153' 'CLIENT643' 'CLIENT135' 'CLIENT1' 'CLIENT2' 'CLIENT3'  
'CLIENT324' 'CLIENT100' 'CLIENT306' 'CLIENT273' 'CLIENT225' 'CLIENT79'  
'CLIENT417' 'CLIENT386' 'CLIENT219' 'CLIENT385' 'CLIENT623' 'CLIENT683'  
'CLIENT127' 'CLIENT508' 'CLIENT256' 'CLIENT140' 'CLIENT350' 'CLIENT614'  
'CLIENT320' 'CLIENT241' 'CLIENT270' 'CLIENT519' 'CLIENT167' 'CLIENT640'  
'CLIENT248' 'CLIENT178' 'CLIENT78' 'CLIENT48' 'CLIENT131' 'CLIENT37'  
'CLIENT200' 'CLIENT657' 'CLIENT52' 'CLIENT468' 'CLIENT367' 'CLIENT305'  
'CLIENT369' 'CLIENT428' 'CLIENT269' 'CLIENT75' 'CLIENT230' 'CLIENT228'  
'CLIENT304' 'CLIENT364' 'CLIENT19' 'CLIENT8' 'CLIENT91' 'CLIENT529'  
'CLIENT499' 'CLIENT202' 'CLIENT119' 'CLIENT303' 'CLIENT93' 'CLIENT414'  
'CLIENT312' 'CLIENT572' 'CLIENT55' 'CLIENT328' 'CLIENT50' 'CLIENT271'  
'CLIENT88' 'CLIENT672' 'CLIENT144' 'CLIENT394' 'CLIENT80' 'CLIENT64'  
'CLIENT42' 'CLIENT108' 'CLIENT11' 'CLIENT210' 'CLIENT588' 'CLIENT316'  
'CLIENT396' 'CLIENT201' 'CLIENT355' 'CLIENT278' 'CLIENT476' 'CLIENT231'  
'CLIENT490' 'CLIENT194' 'CLIENT244' 'CLIENT33' 'CLIENT145' 'CLIENT325'  
'CLIENT400' 'CLIENT284' 'CLIENT289' 'CLIENT465' 'CLIENT35' 'CLIENT536'  
'CLIENT438' 'CLIENT686' 'CLIENT128' 'CLIENT51' 'CLIENT168' 'CLIENT234'  
'CLIENT681' 'CLIENT481' 'CLIENT590' 'CLIENT133' 'CLIENT371' 'CLIENT453'  
'CLIENT291' 'CLIENT434' 'CLIENT71' 'CLIENT477' 'CLIENT589' 'CLIENT317'  
'CLIENT412' 'CLIENT565' 'CLIENT18' 'CLIENT118' 'CLIENT659' 'CLIENT551'  
'CLIENT503' 'CLIENT342' 'CLIENT653' 'CLIENT478' 'CLIENT87' 'CLIENT375'  
'CLIENT86' 'CLIENT105' 'CLIENT656' 'CLIENT687' 'CLIENT639' 'CLIENT575'  
'CLIENT223' 'CLIENT635' 'CLIENT595' 'CLIENT58' 'CLIENT23' 'CLIENT607'  
'CLIENT569' 'CLIENT334' 'CLIENT426' 'CLIENT32' 'CLIENT634' 'CLIENT387'  
'CLIENT208' 'CLIENT518' 'CLIENT357' 'CLIENT242' 'CLIENT598' 'CLIENT229'  
'CLIENT677' 'CLIENT336' 'CLIENT527' 'CLIENT526' 'CLIENT184' 'CLIENT501'  
'CLIENT662' 'CLIENT156' 'CLIENT507' 'CLIENT224' 'CLIENT586' 'CLIENT213'  
'CLIENT172' 'CLIENT583' 'CLIENT68' 'CLIENT83' 'CLIENT123' 'CLIENT645'  
'CLIENT98' 'CLIENT528' 'CLIENT509' 'CLIENT368' 'CLIENT129' 'CLIENT85'  
'CLIENT235' 'CLIENT196' 'CLIENT254' 'CLIENT74' 'CLIENT576' 'CLIENT685'  
'CLIENT456' 'CLIENT443' 'CLIENT437' 'CLIENT363' 'CLIENT148' 'CLIENT377'  
'CLIENT191' 'CLIENT502' 'CLIENT171' 'CLIENT143' 'CLIENT593' 'CLIENT5'  
'CLIENT257' 'CLIENT567' 'CLIENT132' 'CLIENT554' 'CLIENT115' 'CLIENT424'  
'CLIENT70' 'CLIENT81' 'CLIENT204' 'CLIENT498' 'CLIENT622' 'CLIENT562'  
'CLIENT174' 'CLIENT432' 'CLIENT307' 'CLIENT160' 'CLIENT136' 'CLIENT282'  
'CLIENT154' 'CLIENT641' 'CLIENT181' 'CLIENT258' 'CLIENT440' 'CLIENT268'  
'CLIENT117' 'CLIENT337' 'CLIENT584' 'CLIENT116' 'CLIENT299' 'CLIENT217'  
'CLIENT198' 'CLIENT510' 'CLIENT329' 'CLIENT179' 'CLIENT147' 'CLIENT272'  
'CLIENT376' 'CLIENT89' 'CLIENT392' 'CLIENT416' 'CLIENT492' 'CLIENT505'  
'CLIENT599' 'CLIENT473' 'CLIENT485' 'CLIENT28' 'CLIENT157' 'CLIENT9'  
'CLIENT25' 'CLIENT252' 'CLIENT516' 'CLIENT667' 'CLIENT281' 'CLIENT321'  
'CLIENT197' 'CLIENT469' 'CLIENT658' 'CLIENT669' 'CLIENT209' 'CLIENT348'

'CLIENT137' 'CLIENT496' 'CLIENT540' 'CLIENT452' 'CLIENT162' 'CLIENT557'  
'CLIENT511' 'CLIENT384' 'CLIENT444' 'CLIENT548' 'CLIENT349' 'CLIENT77'  
'CLIENT46' 'CLIENT27' 'CLIENT591' 'CLIENT393' 'CLIENT352' 'CLIENT676'  
'CLIENT522' 'CLIENT247' 'CLIENT29' 'CLIENT568' 'CLIENT286' 'CLIENT343'  
'CLIENT177' 'CLIENT109' 'CLIENT353' 'CLIENT466' 'CLIENT134' 'CLIENT290'  
'CLIENT313' 'CLIENT222' 'CLIENT539' 'CLIENT470' 'CLIENT413' 'CLIENT110'  
'CLIENT514' 'CLIENT582' 'CLIENT103' 'CLIENT195' 'CLIENT573' 'CLIENT262'  
'CLIENT533' 'CLIENT267' 'CLIENT602' 'CLIENT311' 'CLIENT149' 'CLIENT264'  
'CLIENT189' 'CLIENT421' 'CLIENT563' 'CLIENT122' 'CLIENT113' 'CLIENT561'  
'CLIENT183' 'CLIENT626' 'CLIENT552' 'CLIENT441' 'CLIENT493' 'CLIENT664'  
'CLIENT577' 'CLIENT315' 'CLIENT341' 'CLIENT405' 'CLIENT96' 'CLIENT397'  
'CLIENT220' 'CLIENT276' 'CLIENT253' 'CLIENT339' 'CLIENT606' 'CLIENT435'  
'CLIENT285' 'CLIENT629' 'CLIENT408' 'CLIENT419' 'CLIENT390' 'CLIENT187'  
'CLIENT491' 'CLIENT604' 'CLIENT407' 'CLIENT218' 'CLIENT553' 'CLIENT454'  
'CLIENT512' 'CLIENT538' 'CLIENT279' 'CLIENT537' 'CLIENT214' 'CLIENT295'  
'CLIENT155' 'CLIENT495' 'CLIENT546' 'CLIENT366' 'CLIENT594' 'CLIENT297'  
'CLIENT480' 'CLIENT648' 'CLIENT126' 'CLIENT600' 'CLIENT158' 'CLIENT467'  
'CLIENT621' 'CLIENT459' 'CLIENT425' 'CLIENT559' 'CLIENT403' 'CLIENT655'  
'CLIENT675' 'CLIENT20' 'CLIENT351' 'CLIENT102' 'CLIENT169' 'CLIENT130'  
'CLIENT67' 'CLIENT332' 'CLIENT326' 'CLIENT517' 'CLIENT447' 'CLIENT651'  
'CLIENT361' 'CLIENT545' 'CLIENT458' 'CLIENT647' 'CLIENT359' 'CLIENT274'  
'CLIENT255' 'CLIENT578' 'CLIENT613' 'CLIENT631' 'CLIENT601' 'CLIENT84'  
'CLIENT475' 'CLIENT238' 'CLIENT345' 'CLIENT121' 'CLIENT523' 'CLIENT682'  
'CLIENT636' 'CLIENT671' 'CLIENT580' 'CLIENT488' 'CLIENT610' 'CLIENT579'  
'CLIENT175' 'CLIENT260' 'CLIENT170' 'CLIENT199' 'CLIENT206' 'CLIENT550'  
'CLIENT464' 'CLIENT138' 'CLIENT53' 'CLIENT185' 'CLIENT233' 'CLIENT159'  
'CLIENT69' 'CLIENT190' 'CLIENT484' 'CLIENT57' 'CLIENT36' 'CLIENT422'  
'CLIENT150' 'CLIENT263' 'CLIENT486' 'CLIENT616' 'CLIENT298' 'CLIENT354'  
'CLIENT451' 'CLIENT555' 'CLIENT124' 'CLIENT265' 'CLIENT679' 'CLIENT415'  
'CLIENT649' 'CLIENT287' 'CLIENT146' 'CLIENT374' 'CLIENT436' 'CLIENT236'  
'CLIENT406' 'CLIENT427' 'CLIENT654' 'CLIENT566' 'CLIENT644' 'CLIENT300'  
'CLIENT237' 'CLIENT592' 'CLIENT627' 'CLIENT524' 'CLIENT161' 'CLIENT221'  
'CLIENT186' 'CLIENT40' 'CLIENT574' 'CLIENT449' 'CLIENT430' 'CLIENT442'  
'CLIENT494' 'CLIENT243' 'CLIENT558' 'CLIENT638' 'CLIENT212' 'CLIENT240'  
'CLIENT556' 'CLIENT530' 'CLIENT398' 'CLIENT391' 'CLIENT596' 'CLIENT445'  
'CLIENT670' 'CLIENT92' 'CLIENT534' 'CLIENT308' 'CLIENT433' 'CLIENT474'  
'CLIENT611' 'CLIENT63' 'CLIENT471' 'CLIENT73' 'CLIENT275' 'CLIENT205'  
'CLIENT482' 'CLIENT104' 'CLIENT294' 'CLIENT292' 'CLIENT277' 'CLIENT521'  
'CLIENT668' 'CLIENT211' 'CLIENT628' 'CLIENT182' 'CLIENT360' 'CLIENT370'  
'CLIENT630' 'CLIENT448' 'CLIENT39' 'CLIENT296' 'CLIENT97' 'CLIENT302'  
'CLIENT373' 'CLIENT506' 'CLIENT163' 'CLIENT585' 'CLIENT457' 'CLIENT261'  
'CLIENT460' 'CLIENT330' 'CLIENT203' 'CLIENT152' 'CLIENT423' 'CLIENT684'  
'CLIENT531' 'CLIENT60' 'CLIENT618' 'CLIENT619' 'CLIENT461' 'CLIENT665'  
'CLIENT399' 'CLIENT660' 'CLIENT358' 'CLIENT227' 'CLIENT504' 'CLIENT226'  
'CLIENT266' 'CLIENT24' 'CLIENT250' 'CLIENT365' 'CLIENT674' 'CLIENT411'  
'CLIENT340' 'CLIENT615' 'CLIENT318' 'CLIENT472' 'CLIENT581' 'CLIENT646'  
'CLIENT410' 'CLIENT293' 'CLIENT446' 'CLIENT388' 'CLIENT513' 'CLIENT215'  
'CLIENT560' 'CLIENT404' 'CLIENT310' 'CLIENT431' 'CLIENT94' 'CLIENT542'

```
'CLIENT382' 'CLIENT314' 'CLIENT323' 'CLIENT564' 'CLIENT389' 'CLIENT633'  
'CLIENT663' 'CLIENT347' 'CLIENT680' 'CLIENT346' 'CLIENT547' 'CLIENT61'  
'CLIENT66' 'CLIENT520' 'CLIENT249' 'CLIENT173' 'CLIENT72' 'CLIENT10'  
'CLIENT497' 'CLIENT632' 'CLIENT605' 'CLIENT418' 'CLIENT570' 'CLIENT544'  
'CLIENT111' 'CLIENT608' 'CLIENT335' 'CLIENT99' 'CLIENT380' 'CLIENT378'  
'CLIENT4' 'CLIENT288' 'CLIENT637' 'CLIENT652' 'CLIENT541' 'CLIENT549'  
'CLIENT395' 'CLIENT571' 'CLIENT489' 'CLIENT319' 'CLIENT372' 'CLIENT26'  
'CLIENT483' 'CLIENT401' 'CLIENT439' 'CLIENT409' 'CLIENT673' 'CLIENT666'  
'CLIENT245' 'CLIENT661' 'CLIENT90' 'CLIENT259' 'CLIENT151' 'CLIENT6'  
'CLIENT487' 'CLIENT624' 'CLIENT535' 'CLIENT344' 'CLIENT617' 'CLIENT331'  
'CLIENT301' 'CLIENT455' 'CLIENT587' 'CLIENT597' 'CLIENT56' 'CLIENT309'  
'CLIENT283']  
['Region 4' 'Region 1' 'Region 3' 'Region 5' 'Region 2']
```

```
[97]: df['Reason_2'] = df['Reason_2'].str.title()
```

```
[98]: df['Reason_2'].unique()
```

```
[98]: array(['Reservation Discrepancies',  
           'Price Discrepancy Across Booking Process',  
           'Bookout - Hotel Sold Out Before Arrival',  
           'Reservation Not Found On Spot', 'Price Match',  
           'Tariff Error Protocol', 'Unknown', 'Specific Customer Agreements',  
           'Time Out/Ghost Booking', 'Bookout - Hotel Sold Out On Spot',  
           'Hotel Mapping Error', 'Contract Error', 'Room Mapping Error',  
           'Bookout - Hotel Close Before Arrival', 'Cancellation Waive',  
           'Booking Operational Issue', 'Rate Error',  
           'Booking Technical Issue', 'Gifts', 'Web Issue', 'Resort Fee',  
           'Nationality Supplement', 'Tariff Error', 'Fraud Bookings',  
           'Other', 'Cancellation With Penalty ',  
           'Bookout - Hotel Sold Out On The Spot',  
           'Bookout - Upgrade To Avoid Rejection',  
           'Credit Card Process Failure', 'Brazilian Law', 'Resort Fees',  
           'Re / Tariff Error Protocol', 'Incentive $150.00Mxn X 3 Bkngs.',  
           'Overbooking ', 'Relocation Issue', 'Te - Web Issues',  
           'Wrong Type Description', 'Nrf Agreement',  
           'Grace Period Extension'], dtype=object)
```

```
[99]: #guardamos el dataframe  
df.to_excel('archivo.xlsx')
```

```
[100]: #A continuación, pasamos a visualización
```

```
[ ]: pip install seaborn
```

```
[105]: import matplotlib.pyplot as plt  
import seaborn as sns
```

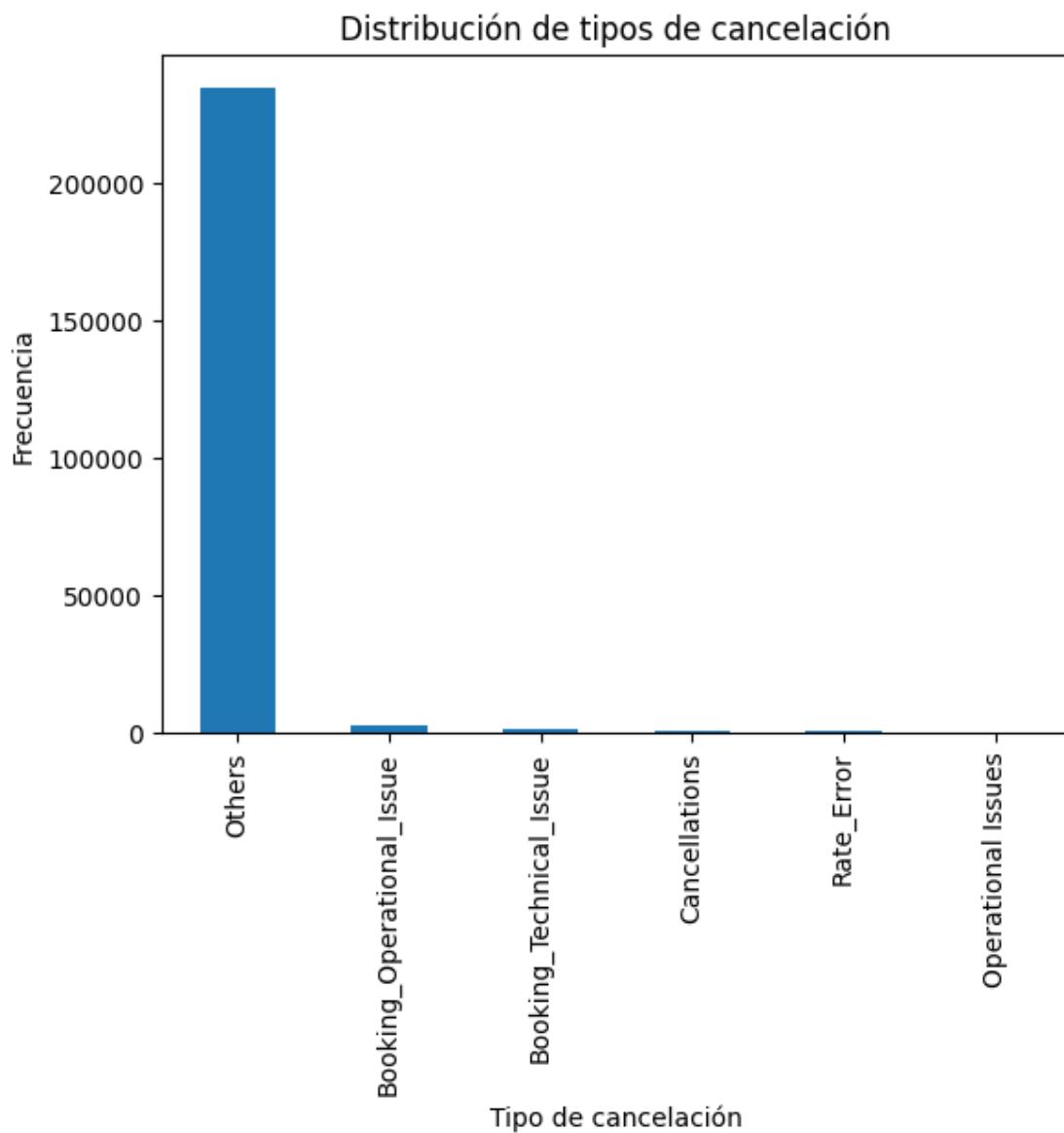
```
[107]: df.columns
```

```
[107]: Index(['Booking', 'Request_date', 'Requested_by', 'Authorized_by',
       'Department', 'Currency', 'Amount', 'Reason', 'Reason_2', 'Status',
       'CustomerShortname', 'CustomerRegion'],
      dtype='object')
```

```
[ ]: #Distribución por tipo de cancelación
```

```
[108]: import matplotlib.pyplot as plt
```

```
reason_counts = df['Reason'].value_counts()
reason_counts.plot(kind='bar', title='Distribución de tipos de cancelación')
plt.xlabel('Tipo de cancelación')
plt.ylabel('Frecuencia')
plt.show()
```

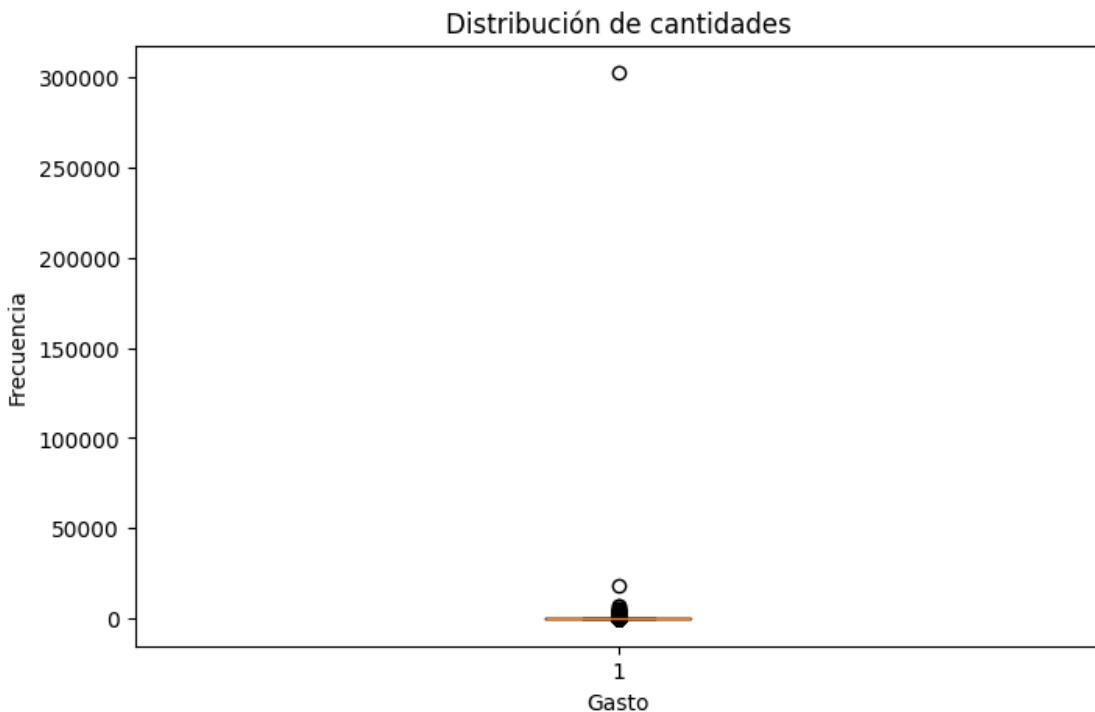


```
[ ]: #Distribución gasto
```

```
[109]: import matplotlib.pyplot as plt

plt.figure(figsize=(8,5))
plt.boxplot(df['Amount'])

plt.xlabel('Gasto')
plt.ylabel('Frecuencia')
plt.title('Distribución de cantidades')
plt.show()
```



```
[ ]: #Euros por tipo de cancelación
```

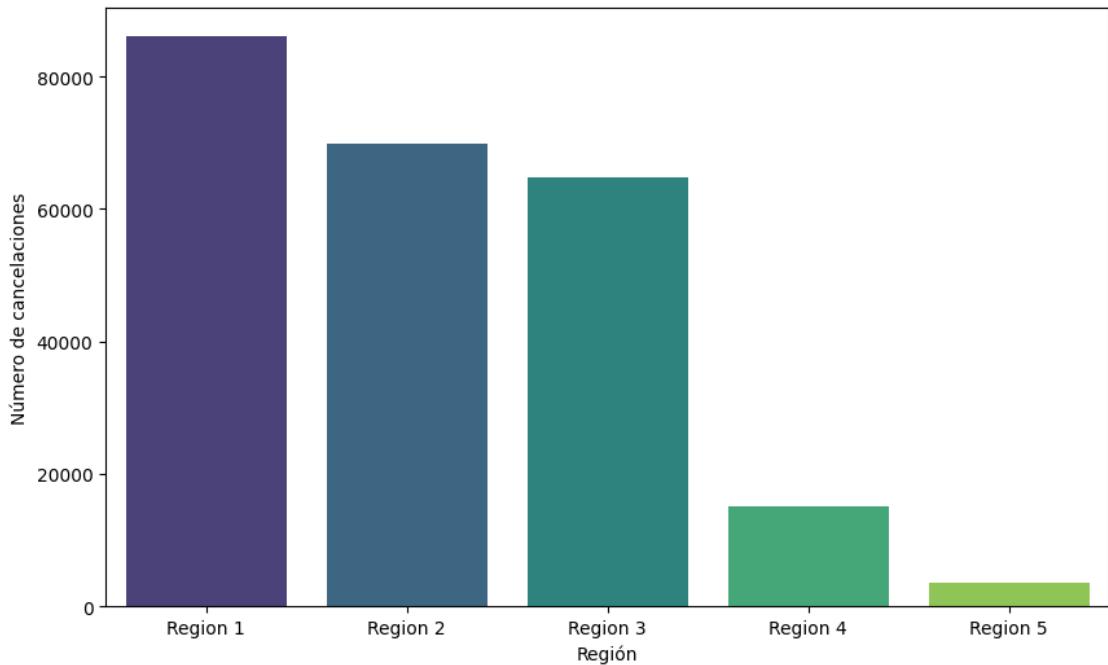
```
[ ]: #cancelaciones por región
```

```
[110]: plt.figure(figsize=(10, 6))
region_counts = df['CustomerRegion'].value_counts().reset_index()
region_counts.columns = ['CustomerRegion', 'Count']
sns.barplot(data=region_counts, x='CustomerRegion', y='Count',
             palette='viridis')
plt.xlabel('Región')
plt.ylabel('Número de cancelaciones')
plt.show()
```

```
/var/folders/cr/f3dqyy756z9ftxxsrmtgjjh0000gn/T/ipykernel_14136/2017763889.py:4
: FutureWarning:
```

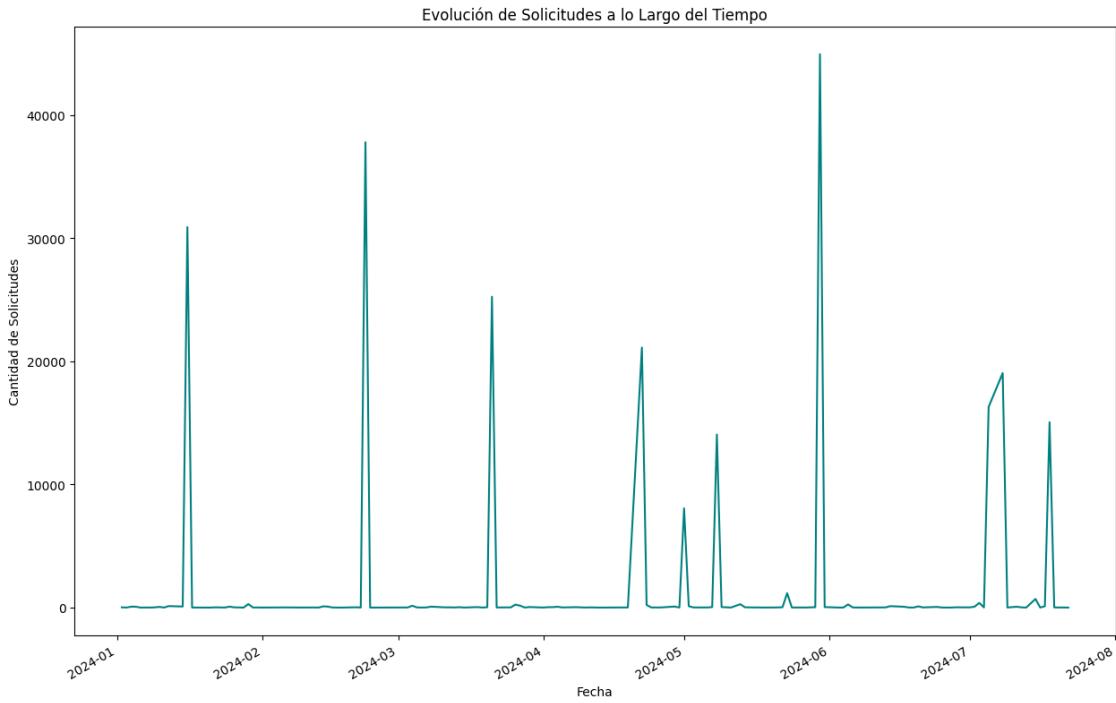
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=region_counts, x='CustomerRegion', y='Count',
             palette='viridis')
```



```
[ ]: #Solicitudes en el tiempo
```

```
[111]: df.groupby('Request_date').size().plot(figsize=(15,10), linestyle='--',  
      color='teal')  
  
plt.title("Evolución de Solicitudes a lo Largo del Tiempo")  
plt.xlabel("Fecha")  
plt.ylabel("Cantidad de Solicitudes")  
plt.show()
```



```
[ ]: #correos con más peticiones
```

```
[112]: correos_frecuencia = df['Requested_by'].value_counts().reset_index()
correos_frecuencia.columns = ['Requested_by', 'Count']

top_correos = correos_frecuencia.head(10)

plt.figure(figsize=(10, 6))
sns.barplot(data=top_correos, x='Requested_by', y='Count', palette='pastel')
plt.title('Correos que Más Aparecen en Requested_by')
plt.xlabel('Correo')
plt.ylabel('Número de Apariciones')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

```
/var/folders/cr/f3dqyy756z9ftxxsrmtgjjh0000gn/T/ipykernel_14136/128422515.py:9:
FutureWarning:
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.
```

```
sns.barplot(data=top_correos, x='Requested_by', y='Count', palette='pastel')
```

