TFG Reservas Hotel Regresión logística

May 28, 2024

Modelos de aprendizaje automático para las reservas de un hotel

Importar los datos y librerias necesarios

[2]:
```
# Importar librerias de calculo

import numpy as np
import pandas as pd

# Importar librerias de visualizacion

import matplotlib.pyplot as plt
import seaborn as sns

# Importar librerias para el aprendizaje automatico

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
```

[3]:
```
# Importar los datos

df = pd.read_csv('./hotel_bookings.csv')
df.head()
```

[3]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month
0	Resort Hotel	0	342	2015	July	27	1
1	Resort Hotel	0	737	2015	July	27	1
2	Resort Hotel	0	7	2015	July	27	1
3	Resort Hotel	0	13	2015	July	27	1
4	Resort Hotel	0	14	2015	July	27	1

```
   stays_in_weekend_nights  stays_in_week_nights  adults  …  deposit_type  \
0                        0                     0       2  …    No Deposit
1                        0                     0       2  …    No Deposit
2                        0                     1       1  …    No Deposit
3                        0                     1       1  …    No Deposit
4                        0                     2       2  …    No Deposit

   agent company days_in_waiting_list customer_type   adr  \
0    NaN     NaN                    0     Transient   0.0
1    NaN     NaN                    0     Transient   0.0
2    NaN     NaN                    0     Transient  75.0
3  304.0     NaN                    0     Transient  75.0
4  240.0     NaN                    0     Transient  98.0

   required_car_parking_spaces  total_of_special_requests  reservation_status  \
0                            0                          0           Check-Out
1                            0                          0           Check-Out
2                            0                          0           Check-Out
3                            0                          0           Check-Out
4                            0                          1           Check-Out

  reservation_status_date
0              2015-07-01
1              2015-07-01
2              2015-07-02
3              2015-07-02
4              2015-07-03

[5 rows x 32 columns]
```

Preparar los datos

[4]:
```
# Metadatos

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
```

```
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  object
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

[5]:
```
#Limpieza y preparación de datos
## Modificar tipo de dato

df = df.astype(
    {'arrival_date_year': 'object', 'arrival_date_week_number': 'object',␣
 ↪'arrival_date_day_of_month' : 'object'}
)
```

[6]:
```
df.dtypes
```

[6]:
```
hotel                           object
is_canceled                      int64
lead_time                        int64
arrival_date_year               object
arrival_date_month              object
arrival_date_week_number        object
arrival_date_day_of_month       object
stays_in_weekend_nights          int64
stays_in_week_nights             int64
```

```
adults                              int64
children                          float64
babies                              int64
meal                               object
country                            object
market_segment                     object
distribution_channel               object
is_repeated_guest                   int64
previous_cancellations              int64
previous_bookings_not_canceled      int64
reserved_room_type                 object
assigned_room_type                 object
booking_changes                     int64
deposit_type                       object
agent                             float64
company                           float64
days_in_waiting_list                int64
customer_type                      object
adr                               float64
required_car_parking_spaces         int64
total_of_special_requests           int64
reservation_status                 object
reservation_status_date            object
dtype: object
```

[7]:
```
# Mapeo de los nombres de los meses a números
month_to_number = {
    'January': '01', 'February': '02', 'March': '03', 'April': '04',
    'May': '05', 'June': '06', 'July': '07', 'August': '08',
    'September': '09', 'October': '10', 'November': '11', 'December': '12'
}

# Aplicar el mapeo para convertir el mes de texto a número
df['month_number'] = df['arrival_date_month'].apply(lambda x:␣
 ↪month_to_number[x])

# Concatenar las columnas para formar la fecha en el formato DD/MM/YYYY
df['arrival_date'] = df['arrival_date_day_of_month'].astype(str) + '/' +␣
 ↪df['month_number'].astype(str) + '/' + df['arrival_date_year'].astype(str)

# Muestra las primeras filas para verificar el resultado
print(df[['arrival_date']].head())
```

```
  arrival_date
0    1/07/2015
1    1/07/2015
2    1/07/2015
3    1/07/2015
```

```
4    1/07/2015
```

[8]:
```
# Convertir la columna 'arrival_date' a tipo datetime
df['arrival_date'] = pd.to_datetime(df['arrival_date'], format='%d/%m/%Y')

print(df[['arrival_date']].head())
```

```
  arrival_date
0   2015-07-01
1   2015-07-01
2   2015-07-01
3   2015-07-01
4   2015-07-01
```

[9]:
```
df.dtypes
```

[9]:
```
hotel                            object
is_canceled                       int64
lead_time                         int64
arrival_date_year                object
arrival_date_month               object
arrival_date_week_number         object
arrival_date_day_of_month        object
stays_in_weekend_nights           int64
stays_in_week_nights              int64
adults                            int64
children                        float64
babies                            int64
meal                             object
country                          object
market_segment                   object
distribution_channel             object
is_repeated_guest                 int64
previous_cancellations            int64
previous_bookings_not_canceled    int64
reserved_room_type               object
assigned_room_type               object
booking_changes                   int64
deposit_type                     object
agent                           float64
company                         float64
days_in_waiting_list              int64
customer_type                    object
adr                             float64
required_car_parking_spaces       int64
total_of_special_requests         int64
reservation_status               object
reservation_status_date          object
```

```
month_number                            object
arrival_date                    datetime64[ns]
dtype: object
```

[10]:
```
# Buscar valores nulos
df.isnull().sum()
```

[10]:
```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        4
babies                          0
meal                            0
country                       488
market_segment                  0
distribution_channel            0
is_repeated_guest               0
previous_cancellations          0
previous_bookings_not_canceled  0
reserved_room_type              0
assigned_room_type              0
booking_changes                 0
deposit_type                    0
agent                       16340
company                    112593
days_in_waiting_list            0
customer_type                   0
adr                             0
required_car_parking_spaces     0
total_of_special_requests       0
reservation_status              0
reservation_status_date         0
month_number                    0
arrival_date                    0
dtype: int64
```

[11]:
```
df['children'].fillna(df['children'].median(), inplace=True)  ###Se reemplazan␣
↪valores por mediana
df['country'].fillna(df['country'].mode()[0], inplace=True) ###Se reemplazan␣
↪valores por moda
```

```
df['agent'].fillna(0, inplace=True) ###Se reemplazan valores por 0

country_column = df.pop('company') ###Se borra la columna country y se almacena␣
 ↪en otra variable

nulls = (df.children == 0) & (df.adults == 0) & (df.babies == 0)
```

[12]:
```
df.isnull().sum()
```

[12]:
```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        0
babies                          0
meal                            0
country                         0
market_segment                  0
distribution_channel            0
is_repeated_guest               0
previous_cancellations          0
previous_bookings_not_canceled  0
reserved_room_type              0
assigned_room_type              0
booking_changes                 0
deposit_type                    0
agent                           0
days_in_waiting_list            0
customer_type                   0
adr                             0
required_car_parking_spaces     0
total_of_special_requests       0
reservation_status              0
reservation_status_date         0
month_number                    0
arrival_date                    0
dtype: int64
```

[13]:
```
# Almacenar valores nulos en otro dataframe
df[nulls]
df = df[~nulls]
```

[14]:
```
# Resumen estadistico
stats = df.describe()
stats
```

[14]:

	is_canceled	lead_time	stays_in_weekend_nights
count	119210.000000	119210.000000	119210.000000
mean	0.370766	104.109227	0.927053
min	0.000000	0.000000	0.000000
25%	0.000000	18.000000	0.000000
50%	0.000000	69.000000	1.000000
75%	1.000000	161.000000	2.000000
max	1.000000	737.000000	19.000000
std	0.483012	106.875450	0.995117

	stays_in_week_nights	adults	children	babies
count	119210.000000	119210.000000	119210.000000	119210.000000
mean	2.499195	1.859206	0.104043	0.007961
min	0.000000	0.000000	0.000000	0.000000
25%	1.000000	2.000000	0.000000	0.000000
50%	2.000000	2.000000	0.000000	0.000000
75%	3.000000	2.000000	0.000000	0.000000
max	50.000000	55.000000	10.000000	10.000000
std	1.897106	0.575186	0.398836	0.097509

	is_repeated_guest	previous_cancellations
count	119210.000000	119210.000000
mean	0.031499	0.087191
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	1.000000	26.000000
std	0.174663	0.844918

	previous_bookings_not_canceled	booking_changes	agent
count	119210.000000	119210.000000	119210.000000
mean	0.137094	0.218799	74.889078
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	7.000000
50%	0.000000	0.000000	9.000000
75%	0.000000	0.000000	152.000000
max	72.000000	18.000000	535.000000
std	1.498137	0.638504	107.168884

	days_in_waiting_list	adr	required_car_parking_spaces
count	119210.000000	119210.000000	119210.000000
mean	2.321215	101.969092	0.062553

```
min                 0.000000      -6.380000                      0.000000
25%                 0.000000      69.500000                      0.000000
50%                 0.000000      94.950000                      0.000000
75%                 0.000000     126.000000                      0.000000
max               391.000000    5400.000000                      8.000000
std                17.598002      50.434007                      0.245360

       total_of_special_requests                   arrival_date
count                119210.000000                         119210
mean                      0.571504  2016-08-28 15:37:59.224897024
min                       0.000000            2015-07-01 00:00:00
25%                       0.000000            2016-03-13 00:00:00
50%                       0.000000            2016-09-06 00:00:00
75%                       1.000000            2017-03-18 00:00:00
max                       5.000000            2017-08-31 00:00:00
std                       0.792876                            NaN
```

[15]:
```
# Seleccionar solo columnas numéricas
numeric_df = df.select_dtypes(include=[float, int])
correlation = numeric_df.corr()
correlation
```

[15]:
```
                                is_canceled  lead_time  \
is_canceled                        1.000000   0.292876
lead_time                          0.292876   1.000000
stays_in_weekend_nights           -0.001323   0.085985
stays_in_week_nights               0.025542   0.166892
adults                             0.058182   0.117575
children                           0.004851  -0.037878
babies                            -0.032569  -0.021003
is_repeated_guest                 -0.083745  -0.123209
previous_cancellations             0.110139   0.086025
previous_bookings_not_canceled    -0.057365  -0.073599
booking_changes                   -0.144832   0.002230
agent                             -0.046770  -0.013114
days_in_waiting_list               0.054301   0.170008
adr                                0.046492  -0.065018
required_car_parking_spaces       -0.195701  -0.116624
total_of_special_requests         -0.234877  -0.095949

                                stays_in_weekend_nights  stays_in_week_nights  \
is_canceled                                   -0.001323              0.025542
lead_time                                      0.085985              0.166892
stays_in_weekend_nights                        1.000000              0.494175
stays_in_week_nights                           0.494175              1.000000
adults                                         0.094759              0.096214
children                                       0.046135              0.044652
```

```
babies                                0.018607             0.020373
is_repeated_guest                    -0.086009            -0.095302
previous_cancellations               -0.012769            -0.013976
previous_bookings_not_canceled       -0.042859            -0.048873
booking_changes                       0.050191             0.080018
agent                                 0.162411             0.196777
days_in_waiting_list                 -0.054399            -0.002026
adr                                   0.050670             0.066847
required_car_parking_spaces          -0.018520            -0.024933
total_of_special_requests             0.073124             0.068738

                                  adults  children    babies  \
is_canceled                     0.058182  0.004851 -0.032569
lead_time                       0.117575 -0.037878 -0.021003
stays_in_weekend_nights         0.094759  0.046135  0.018607
stays_in_week_nights            0.096214  0.044652  0.020373
adults                          1.000000  0.029409  0.017890
children                        0.029409  1.000000  0.023999
babies                          0.017890  0.023999  1.000000
is_repeated_guest              -0.140973 -0.032475 -0.008813
previous_cancellations         -0.007070 -0.024755 -0.007509
previous_bookings_not_canceled -0.108856 -0.021078 -0.006552
booking_changes                -0.041472  0.051000  0.085605
agent                           0.023370  0.050461  0.030235
days_in_waiting_list           -0.008365 -0.033293 -0.010627
adr                             0.224253  0.325057  0.029043
required_car_parking_spaces     0.014438  0.056247  0.037389
total_of_special_requests       0.123353  0.081747  0.097939

                                is_repeated_guest  previous_cancellations  \
is_canceled                             -0.083745               0.110139
lead_time                               -0.123209               0.086025
stays_in_weekend_nights                 -0.086009              -0.012769
stays_in_week_nights                    -0.095302              -0.013976
adults                                  -0.140973              -0.007070
children                                -0.032475              -0.024755
babies                                  -0.008813              -0.007509
is_repeated_guest                        1.000000               0.082740
previous_cancellations                   0.082740               1.000000
previous_bookings_not_canceled           0.420642               0.152570
booking_changes                          0.013044              -0.027261
agent                                   -0.051584              -0.018251
days_in_waiting_list                    -0.022057               0.005941
adr                                     -0.130807              -0.065974
required_car_parking_spaces              0.077928              -0.018540
total_of_special_requests                0.012963              -0.048488
```

	previous_bookings_not_canceled
is_canceled	-0.057365
lead_time	-0.073599
stays_in_weekend_nights	-0.042859
stays_in_week_nights	-0.048873
adults	-0.108856
children	-0.021078
babies	-0.006552
is_repeated_guest	0.420642
previous_cancellations	0.152570
previous_bookings_not_canceled	1.000000
booking_changes	0.011963
agent	-0.046348
days_in_waiting_list	-0.009416
adr	-0.072335
required_car_parking_spaces	0.047506
total_of_special_requests	0.037775

	booking_changes	agent
is_canceled	-0.144832	-0.046770
lead_time	0.002230	-0.013114
stays_in_weekend_nights	0.050191	0.162411
stays_in_week_nights	0.080018	0.196777
adults	-0.041472	0.023370
children	0.051000	0.050461
babies	0.085605	0.030235
is_repeated_guest	0.013044	-0.051584
previous_cancellations	-0.027261	-0.018251
previous_bookings_not_canceled	0.011963	-0.046348
booking_changes	1.000000	0.038555
agent	0.038555	1.000000
days_in_waiting_list	-0.011916	-0.041182
adr	0.026601	0.015711
required_car_parking_spaces	0.067490	0.119282
total_of_special_requests	0.055003	0.060783

	days_in_waiting_list	adr
is_canceled	0.054301	0.046492
lead_time	0.170008	-0.065018
stays_in_weekend_nights	-0.054399	0.050670
stays_in_week_nights	-0.002026	0.066847
adults	-0.008365	0.224253
children	-0.033293	0.325057
babies	-0.010627	0.029043
is_repeated_guest	-0.022057	-0.130807
previous_cancellations	0.005941	-0.065974
previous_bookings_not_canceled	-0.009416	-0.072335

```
booking_changes                             -0.011916  0.026601
agent                                       -0.041182  0.015711
days_in_waiting_list                         1.000000 -0.040859
adr                                         -0.040859  1.000000
required_car_parking_spaces                 -0.030601  0.056510
total_of_special_requests                   -0.082755  0.172308

                                 required_car_parking_spaces  \
is_canceled                                        -0.195701
lead_time                                          -0.116624
stays_in_weekend_nights                            -0.018520
stays_in_week_nights                               -0.024933
adults                                              0.014438
children                                            0.056247
babies                                              0.037389
is_repeated_guest                                   0.077928
previous_cancellations                             -0.018540
previous_bookings_not_canceled                      0.047506
booking_changes                                     0.067490
agent                                               0.119282
days_in_waiting_list                               -0.030601
adr                                                 0.056510
required_car_parking_spaces                         1.000000
total_of_special_requests                           0.082718

                                 total_of_special_requests
is_canceled                                      -0.234877
lead_time                                        -0.095949
stays_in_weekend_nights                           0.073124
stays_in_week_nights                              0.068738
adults                                            0.123353
children                                          0.081747
babies                                            0.097939
is_repeated_guest                                 0.012963
previous_cancellations                           -0.048488
previous_bookings_not_canceled                    0.037775
booking_changes                                   0.055003
agent                                             0.060783
days_in_waiting_list                             -0.082755
adr                                               0.172308
required_car_parking_spaces                       0.082718
total_of_special_requests                         1.000000
```

Modelos de Machine Learning

Modelo de regresión logística (predecir cancelaciones)

[16]:
```
# Crear una instancia del modelo de regresión logística
model = LogisticRegression()

#'is_canceled' es la variable a predecir
target = df['is_canceled']

# Selección de algunas características consideradas relevantes
features = df[['lead_time',␣
 ↪'total_of_special_requests','previous_cancellations', 'adults', 'children',␣
 ↪'babies', 'booking_changes']]

# Dividir los datos en conjunto de entrenamiento y de prueba
X_train, X_test, y_train, y_test = train_test_split(features, target,␣
 ↪test_size=0.2, random_state=42)
```

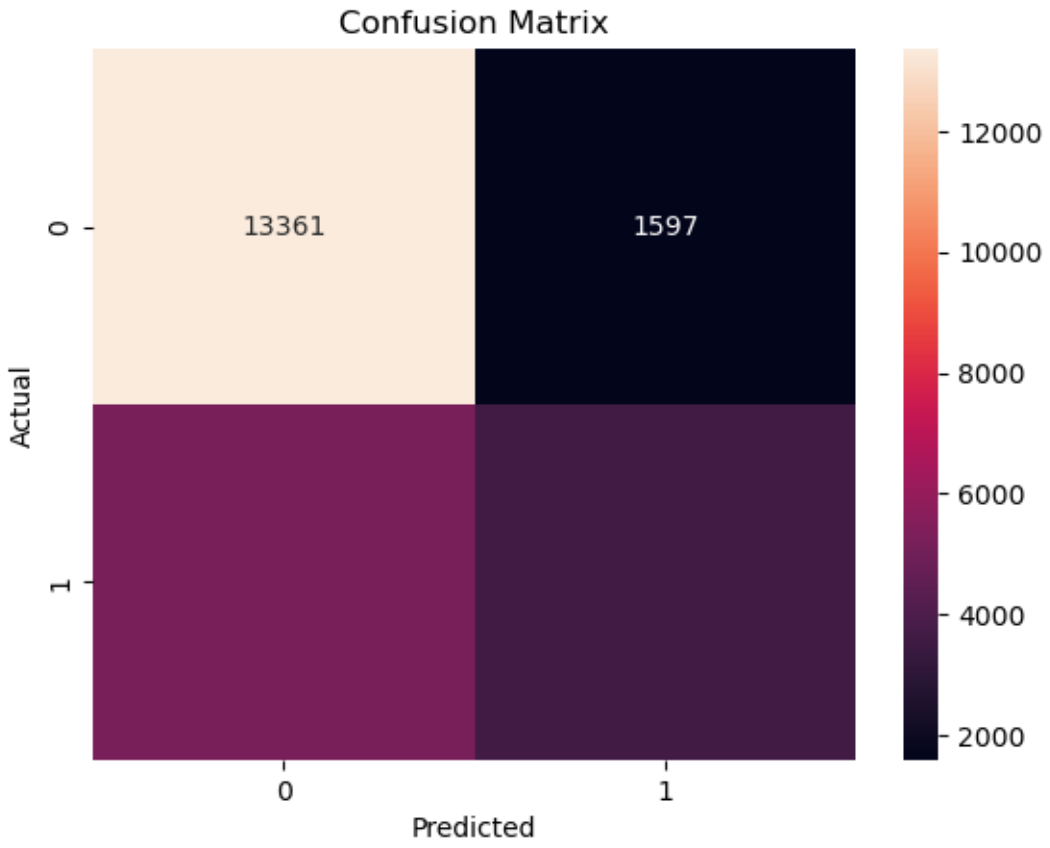
[17]:
```
## Entrenar el modelo
model.fit(X_train, y_train)
```

[17]:
```
LogisticRegression()
```

[18]:
```
# Predecir las etiquetas para el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular la precisión y generar un informe
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
Accuracy: 0.7140760003355423
              precision    recall  f1-score   support

           0       0.72      0.89      0.80     14958
           1       0.70      0.41      0.52      8884

    accuracy                           0.71     23842
   macro avg       0.71      0.65      0.66     23842
weighted avg       0.71      0.71      0.69     23842
```

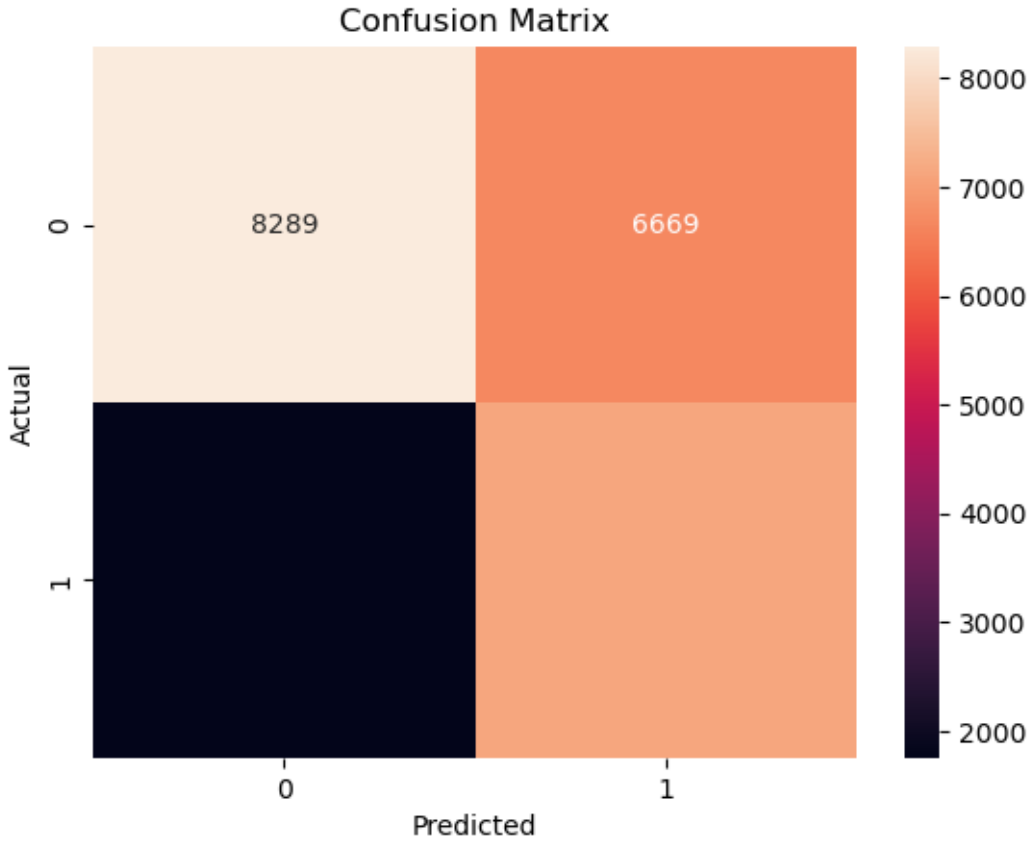
[19]:
```
# Matriz de confusión
from sklearn.metrics import confusion_matrix

conf_mat = confusion_matrix(y_test, y_pred)
sns.heatmap(conf_mat, annot=True, fmt='d')
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.title('Confusion Matrix')
plt.show()
```



[20]:
```
# Aplicar un nuevo umbral
probabilidades = model.predict_proba(X_test)[:, 1]
y_pred_umbral = np.where(probabilidades > 0.3, 1, 0)  # Umbral ajustado a 0,3

# Evaluar el nuevo umbral
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred_umbral))
```

```
Accuracy: 0.7140760003355423
              precision    recall  f1-score   support

           0       0.83      0.55      0.66     14958
           1       0.52      0.80      0.63      8884

    accuracy                           0.65     23842
   macro avg       0.67      0.68      0.65     23842
weighted avg       0.71      0.65      0.65     23842
```

[21]:
```
conf_mat = confusion_matrix(y_test, y_pred_umbral)
sns.heatmap(conf_mat, annot=True, fmt='d')
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.title('Confusion Matrix')
plt.show()
```



[24]:
```
#Exportar datos
df.to_csv('./hotel_bookings_regression_log.csv', index=False)
```

TFG Reservas Hotel Regresión lineal

May 28, 2024

0.3.2. Modelo de regresión lineal (estimar ADR -Average Daily Rate- o tarifa promedio)

[1]:
```
# Importar librerias de calculo
import numpy as np
import pandas as pd

# Importar librerias de visualizacion
import matplotlib.pyplot as plt
import seaborn as sns

# Importar librerias para el aprendizaje automatico
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error
```

[2]:
```
df = pd.read_csv('./hotel_bookings_regression_log.csv')
```

[3]:
```
# Lista de columnas para one-hot encoding
columns_to_encode = ['meal', 'arrival_date_month', 'hotel',␣
 ↪'reserved_room_type']

# Aplicar one-hot encoding
df = pd.get_dummies(df, columns=columns_to_encode)

# Imprimir las primeras filas para verificar los cambios
print(df.head())
```

```
   is_canceled  lead_time  arrival_date_year  arrival_date_week_number  \
0            0        342               2015                        27
1            0        737               2015                        27
2            0          7               2015                        27
3            0         13               2015                        27
4            0         14               2015                        27

   arrival_date_day_of_month  stays_in_weekend_nights  stays_in_week_nights  \
0                          1                        0                     0
1                          1                        0                     0
2                          1                        0                     1
3                          1                        0                     1
```

```
4                          1                          0                          2

   adults  children  babies  … hotel_Resort Hotel reserved_room_type_A  \
0       2       0.0       0  …               True                False
1       2       0.0       0  …               True                False
2       1       0.0       0  …               True                 True
3       1       0.0       0  …               True                 True
4       2       0.0       0  …               True                 True

  reserved_room_type_B  reserved_room_type_C  reserved_room_type_D  \
0                False                  True                 False
1                False                  True                 False
2                False                 False                 False
3                False                 False                 False
4                False                 False                 False

   reserved_room_type_E reserved_room_type_F  reserved_room_type_G  \
0                 False                False                 False
1                 False                False                 False
2                 False                False                 False
3                 False                False                 False
4                 False                False                 False

  reserved_room_type_H  reserved_room_type_L
0                False                 False
1                False                 False
2                False                 False
3                False                 False
4                False                 False

[5 rows x 57 columns]
```

[26]:
```
# Crear una instancia del modelo de regresión lineal
model = LinearRegression()

# Lista todas las columnas
features = ['adults', 'children', 'babies'] + [col for col in df.columns if col.
 ↪startswith(('meal_', 'arrival_date_month_', 'hotel_',␣
 ↪'reserved_room_type_'))]

# Selección de características
X = df[features]
y = df['adr']

# Dividir los datos en conjunto de entrenamiento y de prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

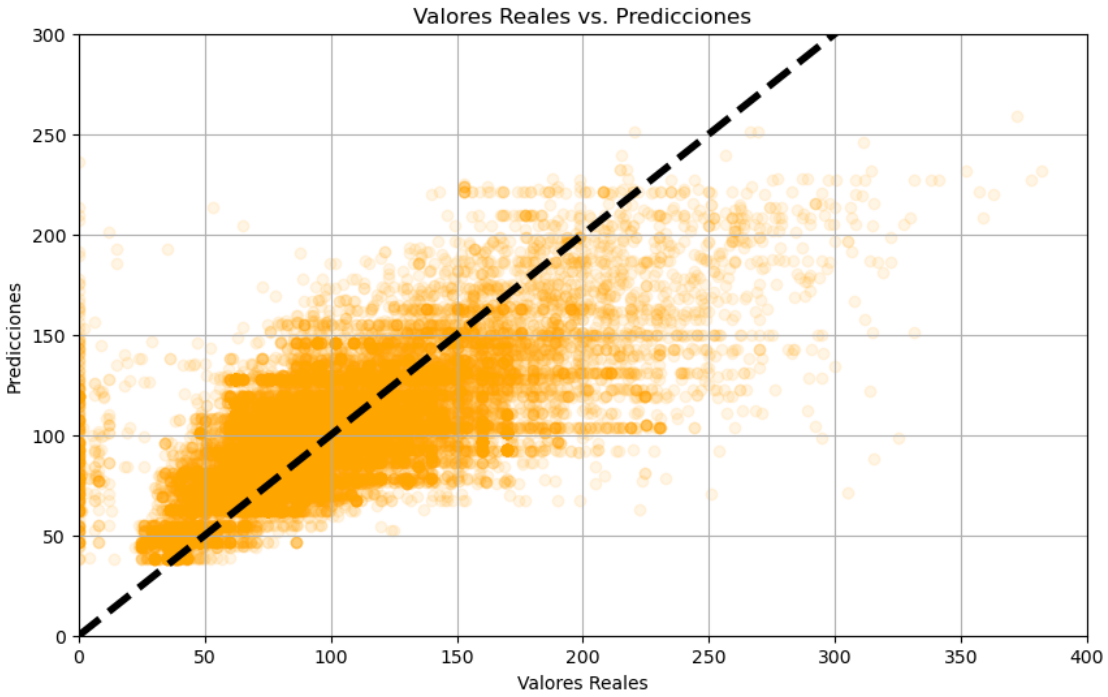
```
# Entrenar el modelo
model.fit(X_train, y_train)
```

[26]:
```
LinearRegression()
```

[27]:
```
# Predecir las tarifas para el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular RMSE
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print("RMSE:", rmse)
```

```
RMSE: 35.26137795550806
```

[28]:
```
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, alpha=0.1, color='orange')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--',␣
 ↪lw=4)  # Línea de predicción
plt.xlabel('Valores Reales')
plt.ylabel('Predicciones')
plt.title('Valores Reales vs. Predicciones')
plt.grid(True)
plt.xlim([0, 400])
plt.ylim([0, 300])
plt.show()
```



0.4. Exportar los datos

[14]:
```
#Exportar datos
df.to_csv('./hotel_bookings_modificado.csv', index=False)
```