

Rerservaciones hotel

Procesamiento de datos con Phytton
Perla Conchita

Notebook: <https://bit.ly/3KndG2i>



Contenido

El data set contiene información de dos hoteles de Portugal, resort (H1) y de ciudad (H2). Cuenta con 31 variables que describen 40,060 observaciones para el de resort y 79,330 observaciones para el de ciudad (H2). Comprende información del 01 julio del 2015 al 31 de agosto del 2017

[.https://www.sciencedirect.com/science/article/pii/S2352340918315191](https://www.sciencedirect.com/science/article/pii/S2352340918315191)

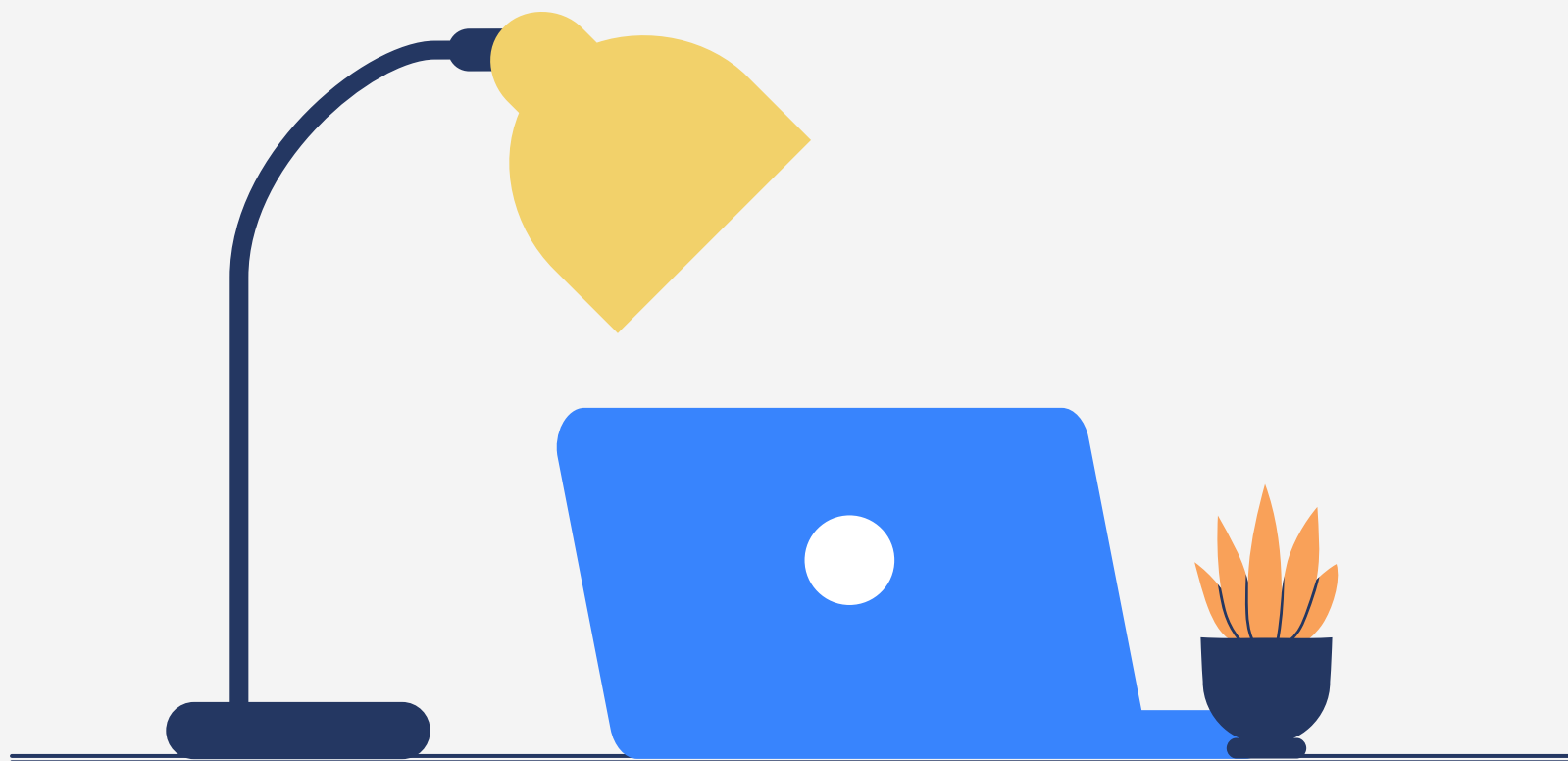


✓ Importación y
limpieza del dataset

✓ Casting y nuevas
variables

✓ Preguntas de
investigación

Importación y limpieza del dataset



01

Información

1. Cuenta con 20 variables numéricas y 12 categóricas
2. Las variables children, country, agent, company presentan valores nulos o faltantes

02

Variable agent

Trámite fue hecho de manera personal, por lo que esos casos carecen de código. Por lo que para fines del proyecto se sustituirán por valor 0.

03

Variable country, company

Country se cambió por la palabra unknown
company es eliminada ya que contiene mas del 50% de datos nulos.

04

Variable children

los valores nulos se reemplazan con el promedio de la columna

Variable agent Al analizar el documento adjunto a la base de datos se detecta que los valores nulos de las variables corresponde a que el trámite fue hecho de manera personal, por lo que esos casos carecen de código. Por lo que se sustituirán por valor 0.

```
[4] #para modificaciones trabajaremos con el siguiente data frame
df_demo = df
```

```
[5] df_agent_null = df[df.agent.isnull()] #identificar los valores nulos de la variable agent del dataset
```

```
[6] #rellenar con valor cero
df_demo["agent"] = df["agent"].fillna(0)
df_demo["agent"].isna().sum()
```

0

Variable country y company En el caso de country se cambió por la palabra *unknown* ya que desconoce que contiene mas del 50% de datos nulos.

```
✓ 0s df_demo["country"] = df_demo["country"].fillna("UNKNOWN")
df_demo["country"].isna().sum()
```

0

```
✓ 0s [9] df_demo.drop("company", inplace= True, axis=1)
```

```
✓ 0s [10] df_demo.columns.values == "company"

array([False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False])
```

Variable children los valores nulos se reemplazan con el promedio de la columna

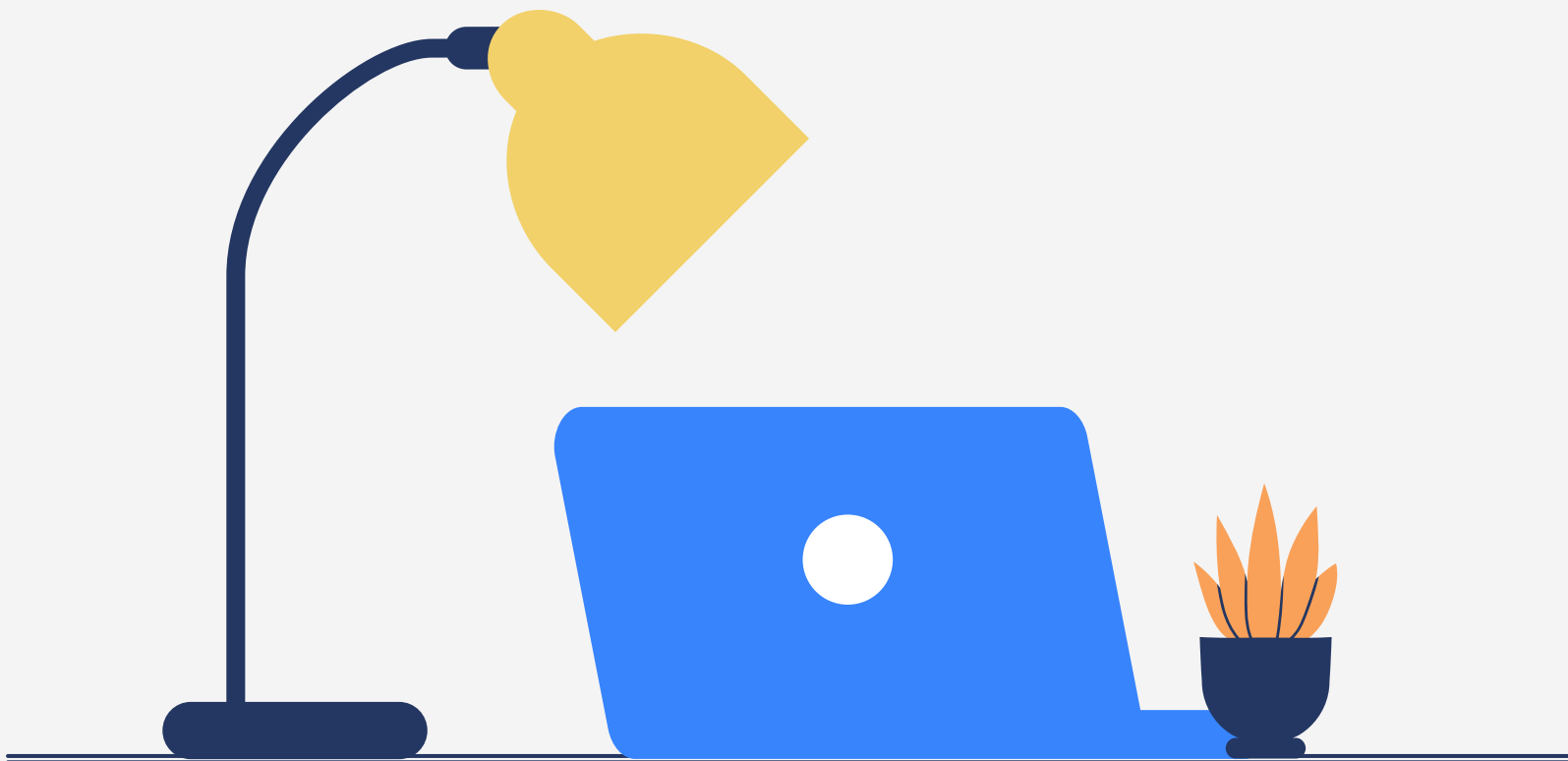
```
✓ 0s df_demo["children"] = df_demo["children"].fillna(df_demo["children"].median())
df_demo["children"].isna().sum()
```

0

```
✓ 0s [12] df_demo["children"].value_counts()
```

```
0.0    110800
1.0     4861
2.0     3652
3.0       76
10.0      1
Name: children, dtype: int64
```

Casting y nuevas variables



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	119390	non-null	float64
11	babies	119390	non-null	int64
12	meal	119390	non-null	object
13	country	119390	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	119390	non-null	float64
24	days_in_waiting_list	119390	non-null	int64
25	customer_type	119390	non-null	object
26	adr	119390	non-null	float64
27	required_car_parking_spaces	119390	non-null	int64
28	total_of_special_requests	119390	non-null	int64
29	reservation_status	119390	non-null	object
30	reservation_status_date	119390	non-null	object

Variables modificadas y creadas

01 reservation_status_date

Casting a formato fecha

30	reservation_status_date	119390	non-null	object
----	-------------------------	--------	----------	--------

02 Nueva variable: arrival_date

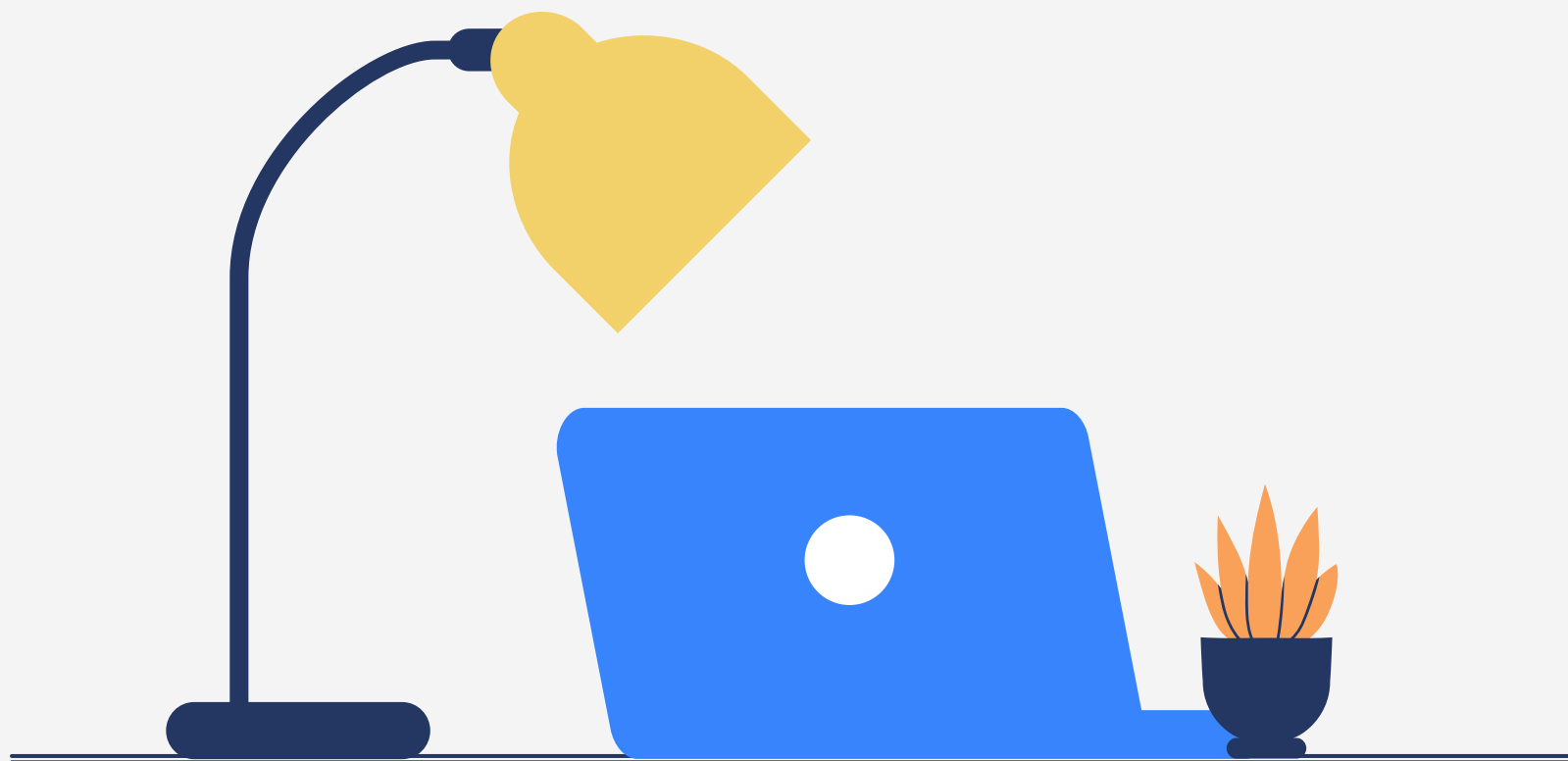
3	arrival_date_year	119390	non-null	int64
4	arrival_date_month	119390	non-null	object
5	arrival_date_week_number	119390	non-null	int64
6	arrival_date_day_of_month	119390	non-null	int64

03 Nueva variable: kids

10	children	119390	non-null	float64
11	babies	119390	non-null	int64



Preguntas de investigación



¿Cuántos menores llegan a cada tipo de hotel?

```
hotel
City Hotel      4946
Resort Hotel    3527
Name: kids, dtype: int64
```

	mean	sum
arrival_date_month		
August	1.380503	878
July	1.420345	740
June	1.451977	257
February	1.395349	240
May	1.396341	229
April	1.417219	214
December	1.327044	211
March	1.375000	187
October	1.338462	174
September	1.403361	167
January	1.402299	122
November	1.350000	108

¿Cuántos menores recibieron los hoteles resort mensualmente?



¿Cuál es el país de procedencia del que se realiza el mayor número de llegadas con menores de edad?
¿Cuántos menores comúnmente se registran por reservación de dicho país?



	mean	sum
country		
PRT	1.370436	2327
ESP	1.373984	1014
FRA	1.383041	946
GBR	1.446458	878
DEU	1.326007	362
ITA	1.336449	286
USA	1.558442	240
BRA	1.338983	237
BEL	1.473684	224
IRL	1.389262	207

Conclusiones

