Análisis Reservas Hotel

Predicción CANCELACIONES

SUPERVISADO

1

Objetivo

Conseguir un modelo de predicción para saber si van a cancelar una reserva del hotel.

Dataset Kagle

- Datos del 1 julio 2015 al 31 agosto 2017
- 119390 reservas
- 32 features
- 2 hoteles

2 HOTELES

40060 RESERVAS



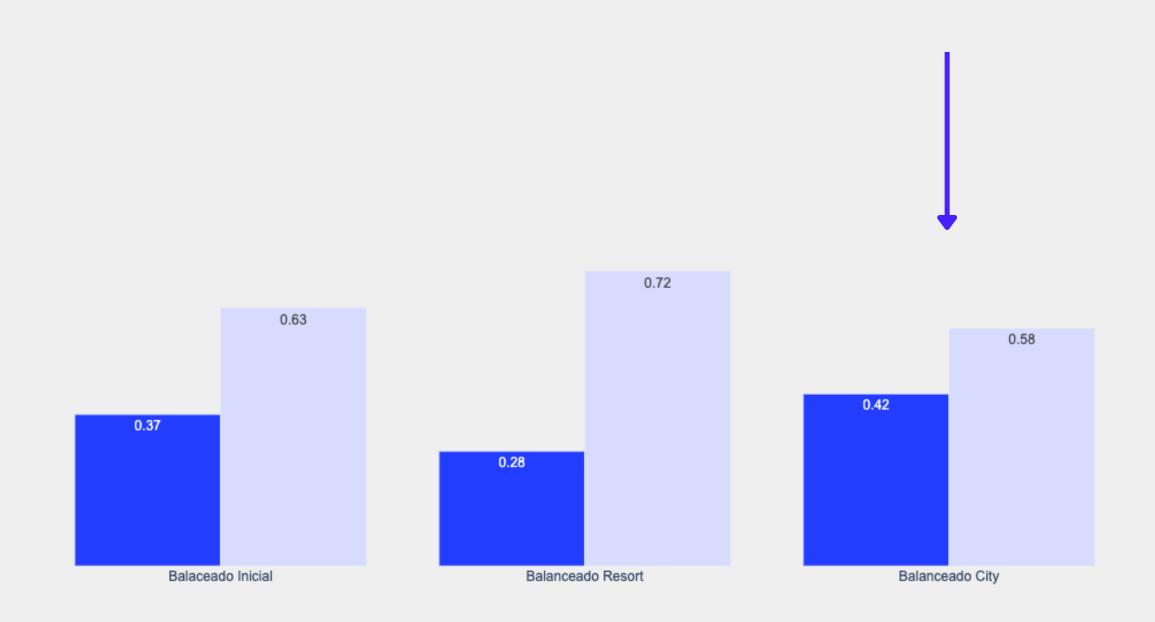
Resort
Hotel situado en la zona del Algarve

79330 RESERVAS



Ciudad Hotel situado en la ciudad de Lisboa

Target





Tarifa Media Diaria por habitación ocupada:

80-125€

Estancia Media

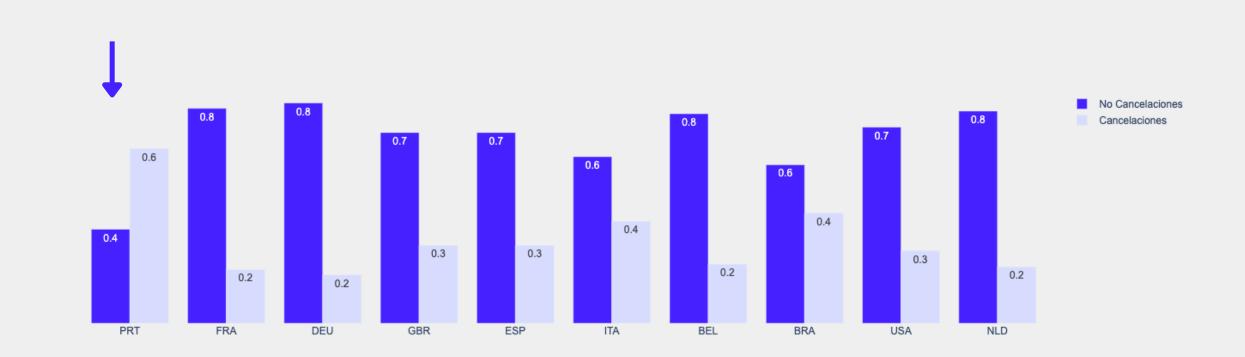
3 noches



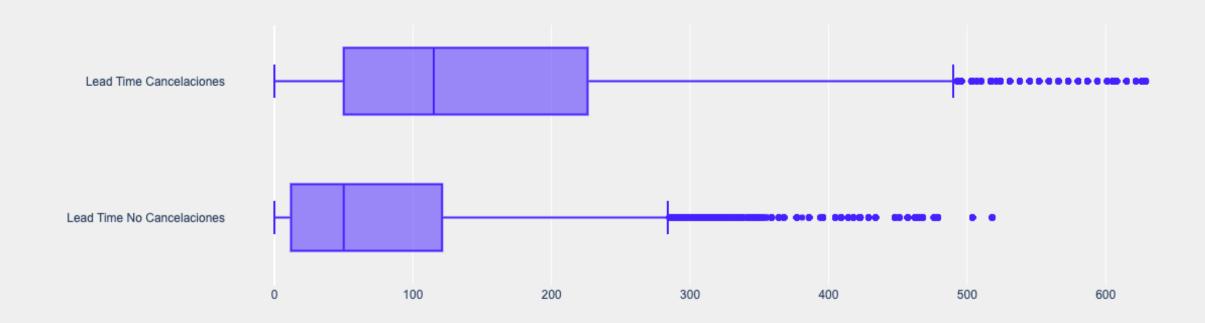


Perfil Huésped





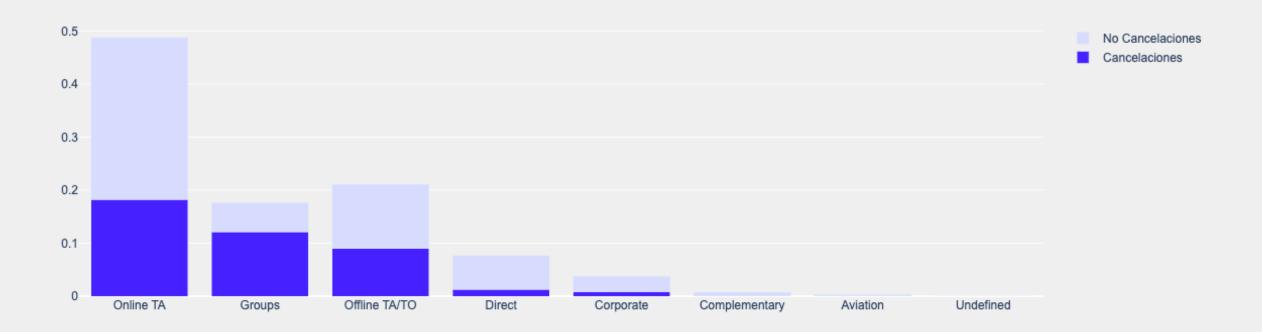
Reservas



Las cancelaciones suelen reservarse con más antelación:

12 - 121 días

50 - 226 días



VARIABLES

15

Variables numéricas

adr

lead_time

days_in_waiting_list

dias_modificacion_llegada

previous_bookings_canceled

previous_cancellations

booking_changes

total_of_special_requests

required_parking_spaces

stay_in_weekend_nights

stay_in_week_nights

total_n_noches

adults

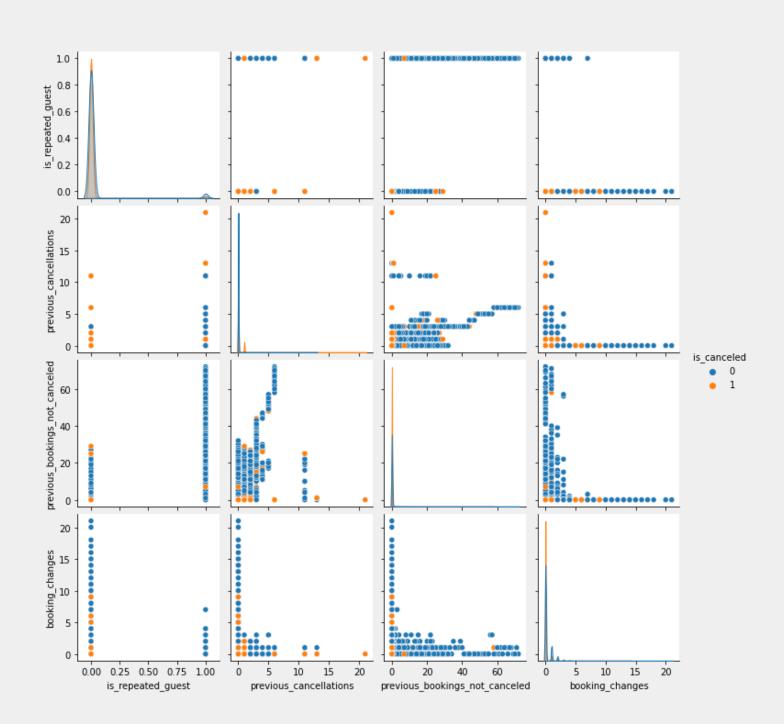
children

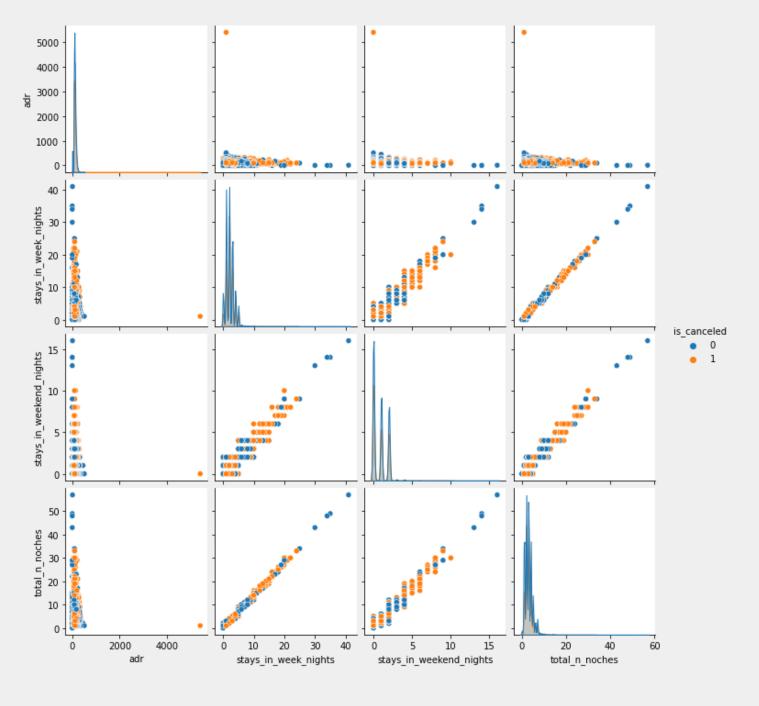
babies

is_canceled -	1	0.31	-0.0073	0.049	0.035	0.053	-0.027	-0.03	-0.066	0.17	-0.053	-0.15	0.061	-0.012	-0.13	-0.29	06
lead_time -	0.31	1	-0.024	0.058	0.034	0.12	-0.054	-0.03	-0.11	0.12	-0.065	-0.031	0.19	-0.14	-0.078	-0.13	0.6
stays_in_weekend_nights -	-0.0073	-0.024	1	0.22	0.65	0.082	0.038	0.0065	-0.08	-0.039	-0.027	0.046	-0.056	0.037	-0.024	0.069	-0.085
stays_in_week_nights -	0.049	0.058	0.22	1	0.89	0.088	0.037	0.0055	-0.094	-0.051	-0.041	0.08	0.016	0.057	-0.038	0.064	0.014
total_n_noches -	0.035	0.034	0.65	0.89	1	0.11	0.047	0.0073	-0.11	-0.058	-0.045	0.084	-0.014	0.062	-0.042	0.083	-0.029
adults -	0.053	0.12	0.082	0.088	0.11	1	-0.0058	0.012	-0.17	-0.038	-0.11	-0.088	-0.0077	0.26	0.011	0.16	0.06
children -	-0.027	-0.054	0.038	0.037	0.047	-0.0058	1	0.024	-0.029	-0.044	-0.017	0.047	-0.036	0.31	0.055	0.11	-0.043
babies -	-0.03	-0.03	0.0065	0.0055	0.0073	0.012	0.024	1	-0.0019	-0.0087	-0.0033	0.066	-0.0091	0.014	0.02	0.072	-0.024
is_repeated_guest -	-0.066	-0.11	-0.08	-0.094	-0.11	-0.17	-0.029	-0.0019	1	0.27	0.45	0.014	-0.024	-0.15	0.093	0.018	-0.017
previous_cancellations -	0.17	0.12	-0.039	-0.051	-0.058	-0.038	-0.044	-0.0087	0.27	1	0.39	-0.04	0.021	-0.13	-0.0034	-0.077	0.14
previous_bookings_not_canceled -	-0.053	0.065	-0.027	-0.041	-0.045	-0.11	-0.017	-0.0033	0.45	0.39	1	0.017	-0.0093	-0.075	0.068	0.046	-0.035
booking_changes -	-0.15	-0.031	0.046	0.08	0.084	-0.088	0.047	0.066	0.014	-0.04	0.017	1	-0.012	0.024	0.042	0.063	-0.12
days_in_waiting_list -	0.061	0.19	0.056	0.016	-0.014	-0.0077	-0.036	-0.0091	-0.024	0.021	-0.0093	-0.012	1	-0.067	0.024	-0.093	0.04
adr -	-0.012	-0.14	0.037	0.057	0.062	0.26	0.31	0.014	-0.15	-0.13	-0.075	0.024	-0.067	1	0.06	0.17	-0.092
required_car_parking_spaces -	-0.13	-0.078	0.024	-0.038	-0.042	0.011	0.055	0.02	0.093	-0.0034	0.068	0.042	-0.024	0.06	1	0.088	-0.079
total_of_special_requests -	-0.29	-0.13	0.069	0.064	0.083	0.16	0.11	0.072	0.018	-0.077	0.046	0.063	-0.093	0.17	0.088	1	0.25
das_modificacion_llegada -	0.6	0.6	-0.085	0.014	-0.029	0.06	-0.043	-0.024	-0.017	0.14	-0.035	-0.12	0.04	-0.092	0.079	0.25	1
	is_canceled -	lead_time	sekend_nights -	week_nights -	otal_n_noches -	adults	children .	babies	peated_guest_	cancellations	- pajaces - ou	king_changes ·	n_waiting_list -	- Jpe	arking_spaces -	ecial_requests -	acion_llegada -

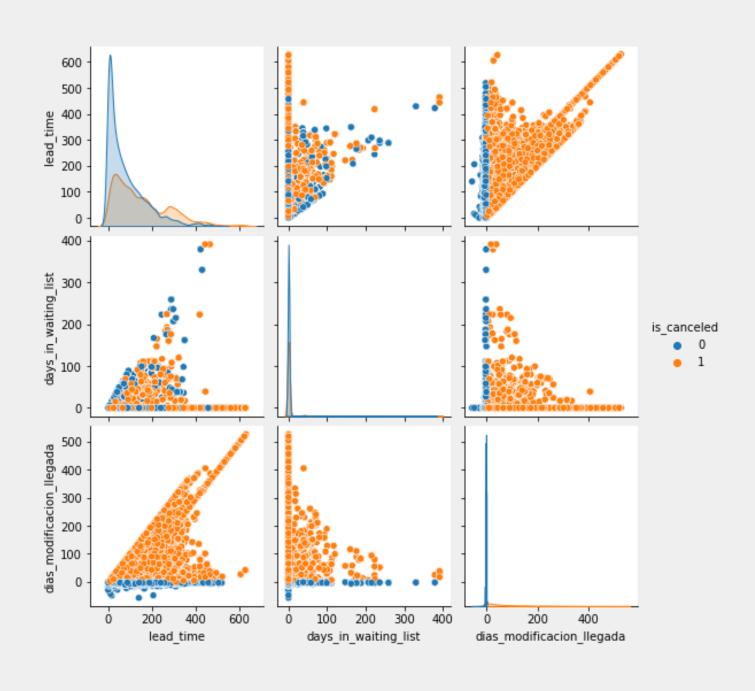
[•] No incluiremos en el modelo predictivo las variables con fondo azul claro por correlación con otras.

NO TENEMOS CORRELACIÓN LINEAL CON EL TARGET



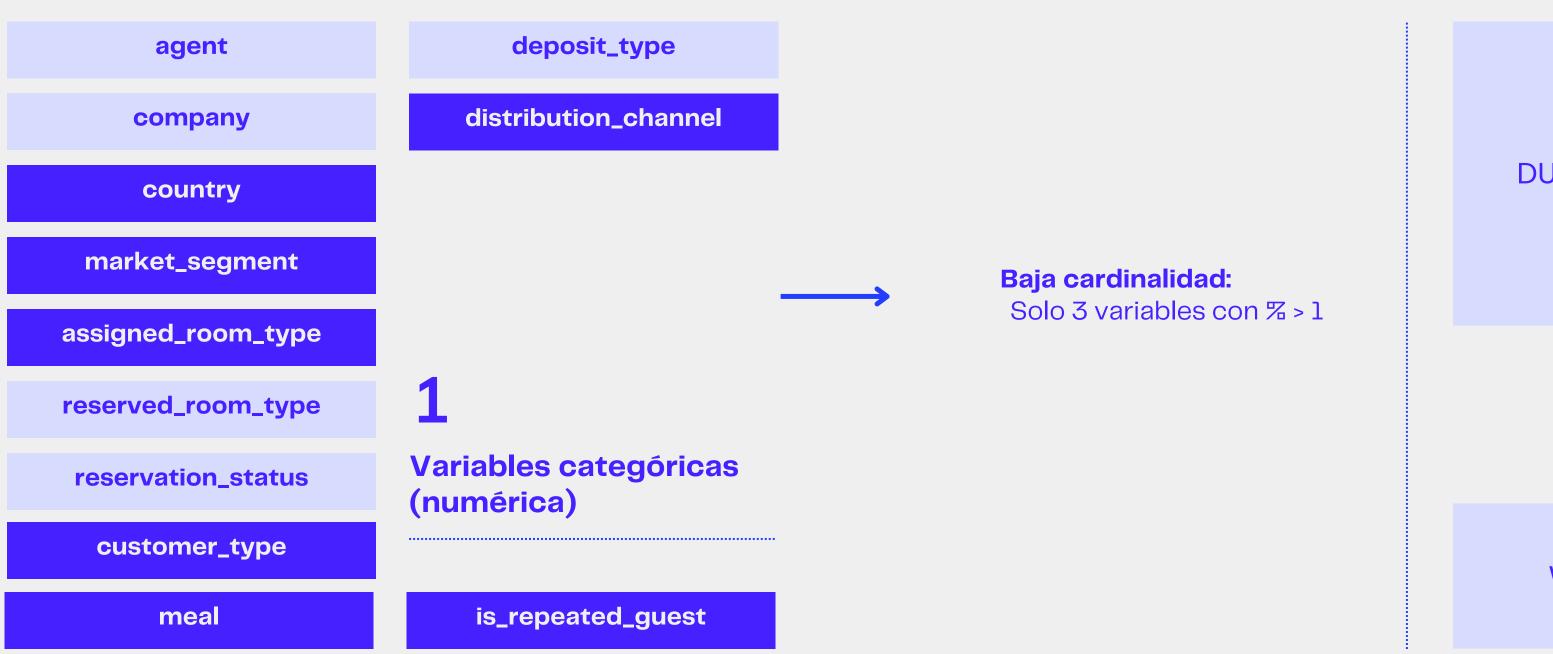


NO TENEMOS CORRELACIÓN LINEAL CON EL TARGET



12

Variables categóricas



DUMMIES

WOE

[•] No incluiremos en el modelo predictivo las variables con fondo azul claro por correlación con otras o por demasiada relación con el target (contaminan el modelo).

Variables fechas

reservation_status_date

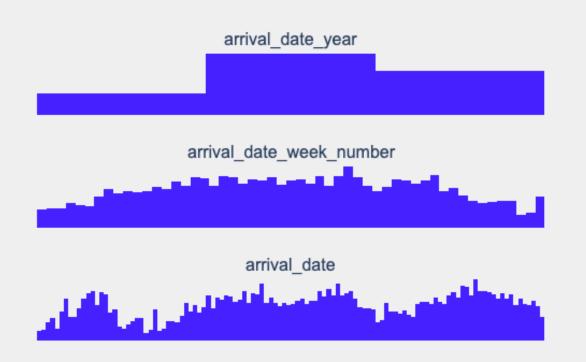
arrival_date

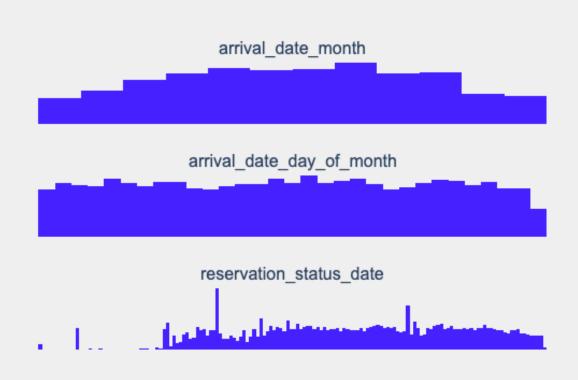
arrival_date_month

arrival_date_week_number

arrival_date_year

arrival_date_day_of_month





• No incluiremos en el modelo predictivo las variables con fondo azul claro por correlación con otras.

MATRIZ PHIK NO HAY CORRELACIÓN LINEAL

is_canceled -	1	0.39	0.032	0.069	0.065	0.051	0.038	0.051	0.07	0.054	0.027	0.069	0.5	0.39	0.14	0.1	0.032	0.062	0.1	0.18	0.059	0.32	0.088	0.093	0.076	0.23	0	0.2	0.46	1	0.66	0.039	0.4	1
lead_time -	0.39	1	0.17	0.45	0.44	0.15	0.085	0.1	0.23	0.069	0.035	0.2	0.34	0.36	0.29	0.19	0.046	0.067	0.14	0.16	0.03	0.45	0.37	0.14	0.31	0.24	0	0.076	0.14	0.35	0.83	0.075	0.84	0.9
arrival_date_year -	0.032	0.17	1	0.57	0.56	0.05	0.039	0.019	0.028	0.02	0	0.12	0.16	0.13	0.067	0.0068	0.02	0.03	0.1	0.084	0	0.015	0.16	0.11	0.13	0.15	0	0.057	0.089	0.022	0.99	0.012	1	0.15
arrival_date_month -	0.069	0.45	0.57	1	0.98	0.17	0.068	0.12	0.2	0.11	0.026	0.15	0.29	0.17	0.12	0.066	0.047	0.061	0.083	0.05	0.04	0.15	0.24	0.2	0.15	0.22	0.007	0.05	0.085	0.097	0.96	0.085	1	0.33
arrival_date_week_number -	0.065	0.44	0.56	0.98	1	0.3	0.072	0.13	0.2	0.11	0.029	0.16	0.29	0.17	0.12	0.066	0.052	0.065	0.083	0.051	0.029	0.16	0.27	0.2	0.14	0.23	0	0.046	0.089	0.096	0.96	0.092	1	0.34
arrival_date_day_of_month -	0.051	0.15	0.05	0.17	0.3	1	0.045	0.027	0.069	0.019	0.017	0.11	0.2	0.098	0.072	0.025	0.032	0	0.026	0.026	0.031	0.066	0.18	0.14	0.11	0.075	0	0.0063	0.035	0.057	0.87	0.027	1	0.12
stays_in_weekend_nights -	0.038	0.085	0.039	0.068	0.072	0.045	1	0.9	0.11	0.038	0	0.027	0.36	0.12	0.078	0.067	0	0.012	0.064	0.16	0.52	0.16	0.029	0.069	0.041	0.046	0	0.015	0.055	0.13	0.49	0.98	0.63	0.062
stays_in_week_nights -	0.051	0.1	0.019	0.12	0.13	0.027	0.9	1	0.13	0.043	0	0.018	0.36	0.11	0.061	0.04	0	0	0.079	0.12	0.71	0.12	0.028	0.098	0.025	0.067	0	0	0.049	0.1	0.36	0.98	0.38	0.04
adults -	0.07	0.23	0.028	0.2	0.2	0.069	0.11	0.13	1	0.18	0.046	0.093	0.26	0.3	0.39	0.15	0.042	0.14	0.45		0.14	0.11	0.15	0.29	0.04	0.11	0	0.017	0.14	0.1	0.43	0.093	0.46	0.17
children -	0.054	0.069	0.02	0.11	0.11	0.019	0.038	0.043	0.18	1	0.029	0.12	0.19	0.26	0.064	0.045	0	0	0.79	0.58	0.038	0.086	0.059	0.037	0.035	0.15	0	0.082	0.13	0.027	0.24	0.044	0.27	0.044
babies -	0.027	0.035	0	0.026	0.029	0.017	0	0	0.046	0.029	1	0	0	0.06	0.088	0	0	0	0.046	0.052	0.027	0.027	0	0	0	0.017	0	0.025	0.07	0.032	0.14	0	0.21	0.036
meal -	0.069	0.2	0.12	0.15	0.16	0.11	0.027	0.018	0.093	0.12	0	1	0.36	0.51	0.1	0.084	0.016	0.026	0.26	0.18	0	0.13	0.095	0.083	0.088	0.41	0	0.021	0.13	0.04	0.74	0.02	0.68	0.1
country -	0.5	0.34	0.16	0.29	0.29	0.2	0.36	0.36	0.26	0.19	0	0.36	1	0.55	0.21	0.19	0	0	0.53	0.52	0.059	0.69	0.38	0.25	0.25	0.24	0	0.075	0.36	0.54	0.85	0.37	0.84	0.38
market_segment -	0.39	0.36	0.13	0.17	0.17	0.098	0.12	0.11	0.3	0.26	0.06	0.51	0.55	1	0.85	0.55	0.14	0.24	0.43	0.25	0.037	0.54	0.32	0.59	0.18	0.55	0	0.19	0.37	0.32	0.72	0.11	0.7	0.31
distribution_channel -	0.14	0.29	0.067	0.12	0.12	0.072	0.078	0.061	0.39	0.064	0.088	0.1	0.21	0.85	1	0.3	0.092	0.3	0.17	0.16	0.042	0.12	0.42	0.63	0.055	0.092	0	0.085	0.068	0.17	0.45	0.047	0.46	0.25
is_repeated_guest -	0.1	0.19	0.0068	0.066	0.066	0.025	0.067	0.04	0.15	0.045	0	0.084	0.19	0.55	0.3	1	0.34	0.54	0.049	0.083	0	0.035	0.054	0.23	0.026	0.09	0	0.14	0.037	0.041	0.43	0.035	0.35	0.15
previous_cancellations -	0.032	0.046	0.02	0.047	0.052	0.032	0	0	0.042	0	0	0.016	0	0.14	0.092	0.34	1	0.69	0.031	0.03	0	0.023	0	0.012	0.011	0.021	0	0.02	0.037	0.038	0.36	0	0.36	0.05
ous_bookings_not_canceled -	0.062	0.067	0.03	0.061	0.065	0	0.012	0	0.14	0	0	0.026	0	0.24	0.3	0.54	0.69	1	0.027	0.032	0	0.027	0	0.1	0	0.023	0	0.067	0.068	0.057	0.11	0	0.18	0.048
reserved_room_type -	0.1	0.14	0.1	0.083	0.083	0.026	0.064	0.079	0.45	0.79	0.046	0.26	0.53	0.43	0.17	0.049	0.031	0.027	1	0.92	0.036	0.24	0.092	0.21	0.061	0.26	0	0.081	0.14	0.087	0.38	0.075	0.36	0.09
assigned_room_type -	0.18	0.16	0.084	0.05	0.051	0.026	0.16	0.12	0.46	0.58	0.052	0.18	0.52	0.25	0.16	0.083	0.03	0.032	0.92	1	0.12	0.39	0.091	0.2	0.076	0.15	0	0.063	0.14	0.29	0.36	0.16	0.32	0.16
booking_changes -	0.059	0.03	0	0.04	0.029	0.031	0.52	0.71	0.14	0.038	0.027	0	0.059	0.037	0.042	0	0	0	0.036	0.12	1	0.049	0.039	0.022	0	0.038	0	0	0.036	0.052	0.16	0.57	0.13	0.038
deposit_type -	0.32	0.45	0.015	0.15	0.16	0.066	0.16	0.12	0.11	0.086	0.027	0.13	0.69	0.54	0.12	0.035	0.023	0.027	0.24	0.39	0.049	1	0.16	0.074	0.28	0.11	0.0038	0.052	0.52	0.72	0.79	0.14	0.61	0.6
agent -	0.088	0.37	0.16	0.24	0.27	0.18	0.029	0.028	0.15	0.059	0	0.095	0.38	0.32		0.054	0	0	0.092	0.091	0.039	0.16	1	0.07	0.075	0.16	0	0.036	0.14	0.096	0.77	0.016	0.81	0.25
company -	0.093	0.14	0.11	0.2	0.2	0.14	0.069	0.098	0.29	0.037	0	0.083	0.25	0.59	0.63	0.23	0.012	0.1	0.21	0.2	0.022	0.074	0.07	1	0	0.092	0	0.045	0.075	0.089	0.69	0.048	0.75	0.11
days_in_waiting_list -	0.076	0.31	0.13	0.15	0.14	0.11	0.041	0.025	0.04	0.035	0	0.088	0.25	0.18	0.055	0.026	0.011	0	0.061	0.076	0	0.28	0.075	0	1	0.13	0	0.019	0.1	0.13	0.81	0.03	0.83	0.13
customer_type -	0.23	0.24	0.15	0.22	0.23	0.075	0.046	0.067	0.11	0.15	0.017	0.41	0.24	0.55	0.092	0.09	0.021	0.023	0.26	0.15	0.038	0.11	0.16	0.092	0.13	1	0	0.08	0.16	0.12	0.78	0.072	0.72	0.26
adr -	0	0	0	0.007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0038	0	0	0	0	1	0	0	0	0.071	0	0	0
equired_car_parking_spaces -	0.2	0.076	0.057	0.05	0.046	0.0063	0.015	0	0.017	0.082	0.025	0.021	0.075	0.19	0.085	0.14	0.02	0.067	0.081	0.063	0	0.052	0.036	0.045	0.019	0.08	0	1	0.082	0.1	0.17	0.005	0.14	0.12
total_of_special_requests -	0.46	0.14	0.089	0.085	0.089	0.035	0.055	0.049	0.14	0.13	0.07	0.13	0.36	0.37	0.068	0.037	0.037	0.068	0.14	0.14	0.036	0.52	0.14	0.075	0.1	0.16	0	0.082	1	0.5	0.48	0.05	0.42	0.28
reservation_status -	1	0.35	0.022	0.097	0.096	0.057	0.13	0.1	0.1	0.027	0.032	0.04	0.54	0.32	0.17	0.041	0.038	0.057	0.087	0.29	0.052	0.72	0.096	0.089	0.13	0.12	0	0.1	0.5	1	0.67	0.14	0.5	0.8
reservation_status_date -	0.66	0.83	0.99	0.96	0.96	0.87		0.36	0.43	0.24	0.14	0.74	0.85	0.72		0.43	0.36	0.11	0.38	0.36	0.16	0.79	0.77	0.69	0.81	0.78	0.071	0.17		0.67		0.37	1	0.84
total_n_noches	0.039	0.075	0.012	0.085	0.092	0.027	0.98	0.98	0.093	0.044	0	0.02	0.37	0.11	0.047	0.035	0	0	0.075	0.16	0.57	0.14	0.016	0.048	0.03	0.072	0	0.005	0.05	0.14	0.37	1	0.37	0.03
arrival_date -	0.4	0.84	1	1	1	1	0.63	0.38	0.46	0.27	0.21	0.68	0.84	0.7	0.46	0.35	0.36	0.18	0.36	0.32	0.13	0.61	0.81	0.75	0.83	0.72	0	0.14	0.42	0.5	1	0.37	1	0.81
dias_modificacion_llegada -	1	0.9	0.15	0.33	0.34	0.12	0.062	0.04	0.17	0.044	0.036	0.1	0.38	0.31	0.25	0.15	0.05	0.048	0.09	0.16	0.038	0.6	0.25	0.11	0.13	0.26	0	0.12	0.28	0.8	0.84	0.037	0.81	1
	is_canceled -	lead_time -	arrival_date_year -	arrival_date_month -	arrival date week number -	arrival_date_day_of_month	stays_in_weekend_nights -	stays_in_week_nights -	adults -	children -	babies -	- meal -	country -	market_segment -	distribution_channel -	is_repeated_guest -	previous_cancellations -	us_bookings_not_canceled -	reserved_room_type -	assigned room type -	booking_changes -	deposit_type -	agent -	- company -	days in waiting list -	customer_type -	- adr	required_car_parking_spaces -	total of special requests -	reservation_status -	reservation_status_date -	total_n_noches -	arrival_date -	dias_modificacion_llegada -

1.0

- 0.4

- 0.2

MODELO PREDICTIVO

PROBLEMA DE CLASIFICACIÓN BINARIA

- Random Forest
- SVM
- Bagging Random Forest
- Gradient BoostingClassifier
- XGBoost
- Red Neuronal

	Modelo	Precisió	n	F1 Score		Accurac	Y	AUC	AUC					
		Train	Test	Train	Test	Train	Test	Train	Test					
	Gradient Boosting Classif.	0.84	0.84	0.81	0.81	0.84	0.84	0.84	0.83					
	SVM (con feature selection)	0.85	0.84	0.78	0.78	0.81	0.81	0.82	0.82					
	Bagging Tree Classifier	0.83	0.83	0.74	0.74	0.80	0.80	0.78	0.79					
	MLC (Red Neuronal)	0.84	0.81	0.85	0.83	0.88	0.86	0.88	0.85					
POCO ERFITT	DE Random Forest	0.93	0.87	0.90	0.84	0.91	0.87	0.91	0.84					
	XGBoost	0.99	0.87	0.99	0.85	0.99	0.87	0.99	0.87					

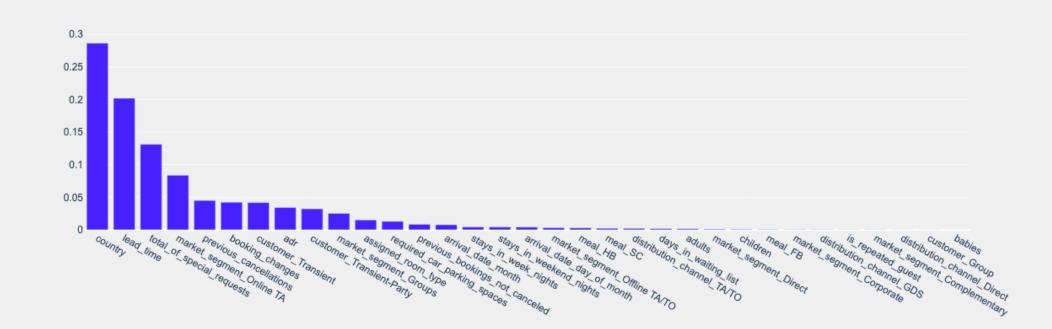
Pasos a Mejorar

Probar con otras opciones de parámetros para intentar aumentar el scoring

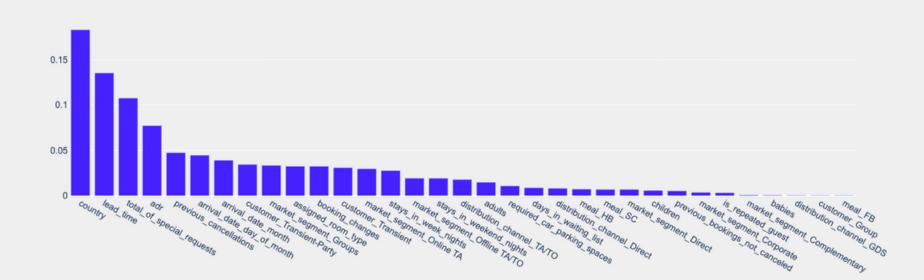
Encontrar posibles variables transformadas para mejorar la predicción

FEATURE IMPORTANCE

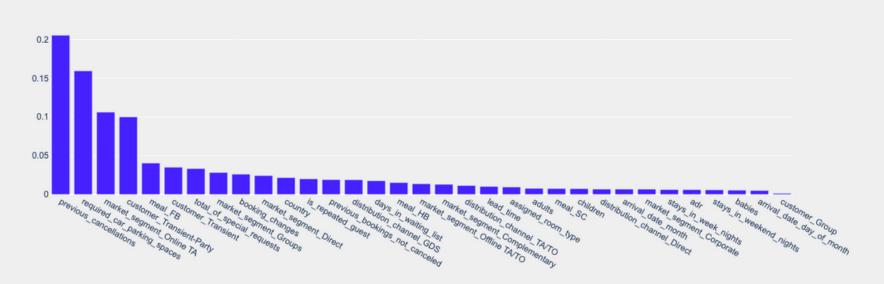
Gradient Boosting Classifier



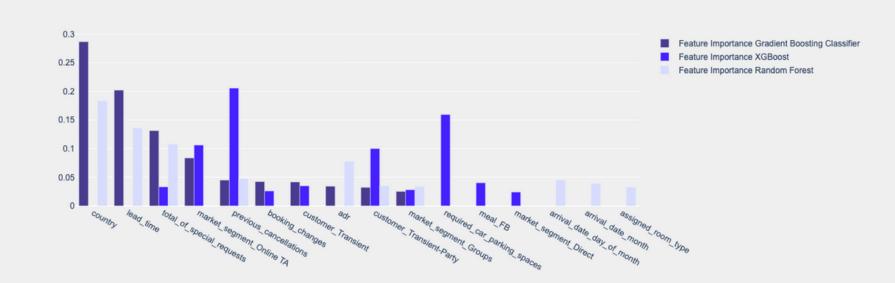




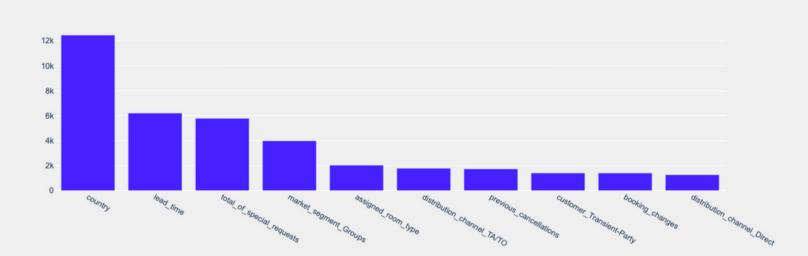
XGBoost



TOP 10 RF, GB y XGBoost



Select KBest



VARIABLES MÁS IMPORTANTES

country

lead_time

arrival date day of month

arrival date month

total of special requests

required car parking spaces

market segment

previous cancellations

booking changes

customer (Transient)

country

Los huéspedes procedentes de ciertos países tienen más probabilidad de cancelar:

PORTUGAL: 60% de las cancelaciones

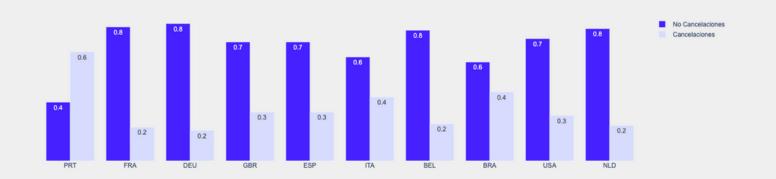
ITALIA: 40% BRASIL: 40%

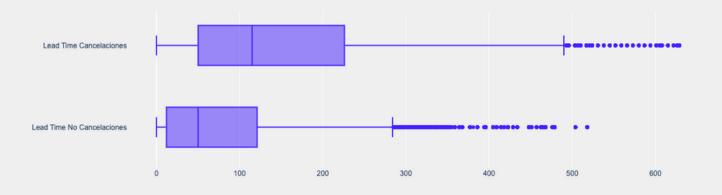
tienen más probabilidad de cancelar.

lead time

Tiempo entre la fecha de reserva y la fecha de llegada.

Como vimos al principio, cuánta mayor antelación, mayor posibilidad de cancelación.



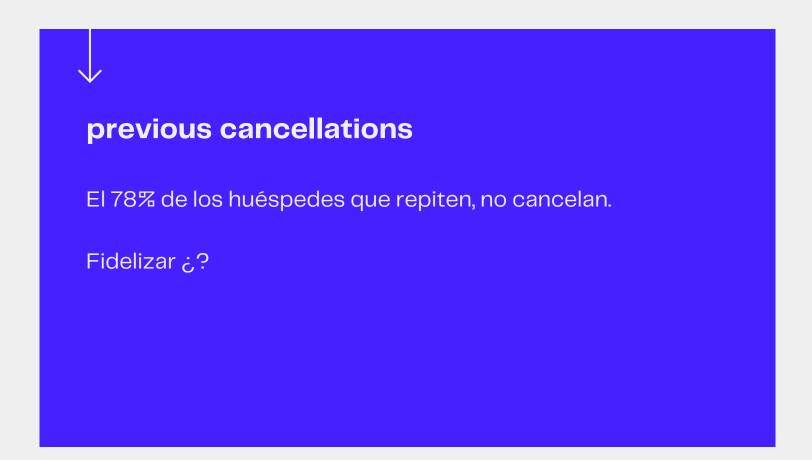


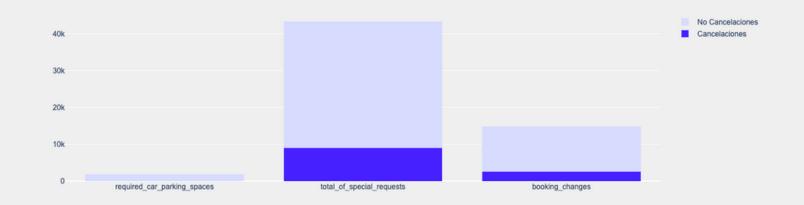
total of special request

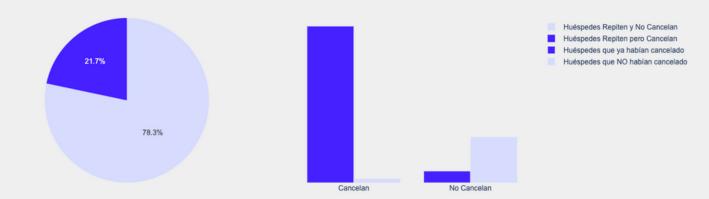
required car parking spaces

booking changes

Lo más común en estas columnas es 0. Que algún huésped haga cambios o tenga peticiones especiales indica que probablemente no cancele.







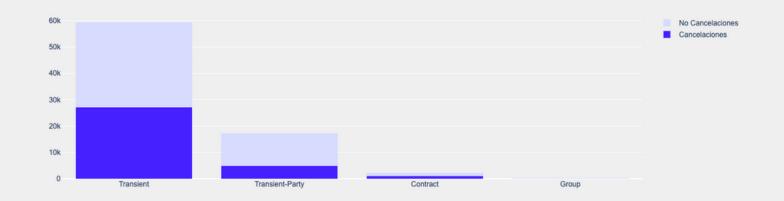
transient transient-party

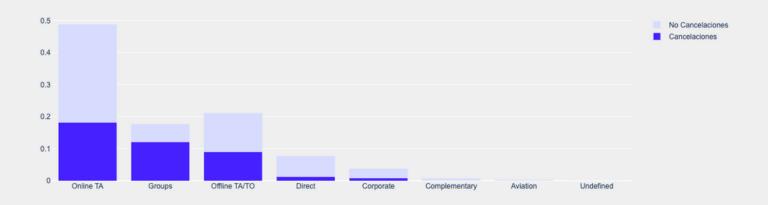
Reservas Individuales

Casi todas las reservas son de este tipo, no nos da mucha información esta columna.

market segment

Destaca la categoría 'Groups, con más cancelaciones que reservas sin cancelar.



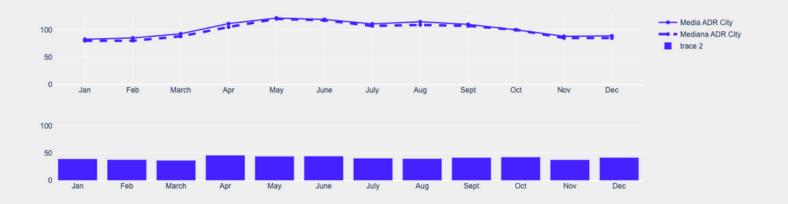


arrival date month

arrival date day of month

El día y el mes de la fecha de reserva.

Destacamos que al ser un hotel de ciudad, no hay picos de precio tan destacados como en los hoteles con estacionalidad.



GRACIAS

SUPERVISADO