

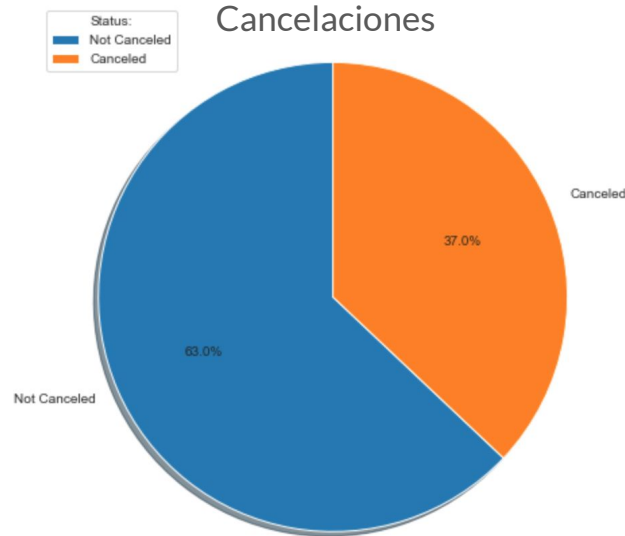


Resultados Modelo Predictivo de Cancelaciones

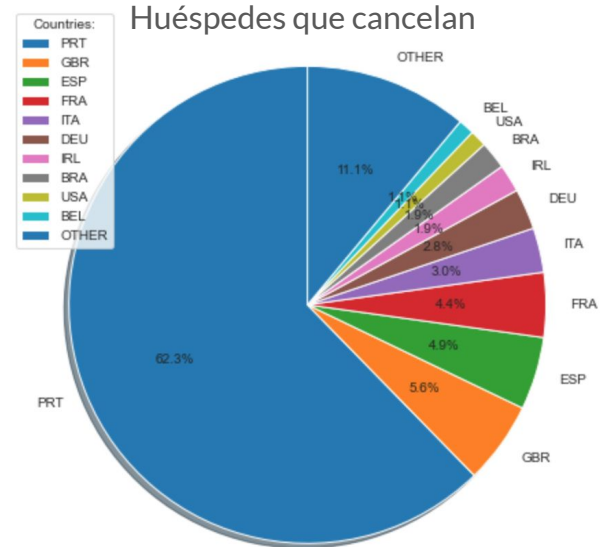
Julio 2021

Resumen de los datos encontrados

El **37%** de las reservaciones que se tienen registradas en los 3 años de datos fueron **canceladas**.



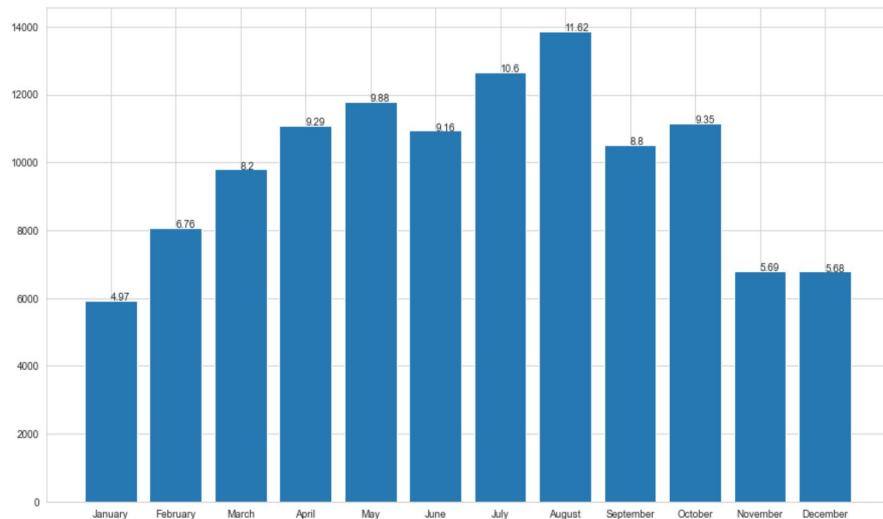
El **62%** de los huéspedes que cancelan su reservación son de **Portugal**, para los demás países sus porcentajes de cancelación son menores al **10%**.



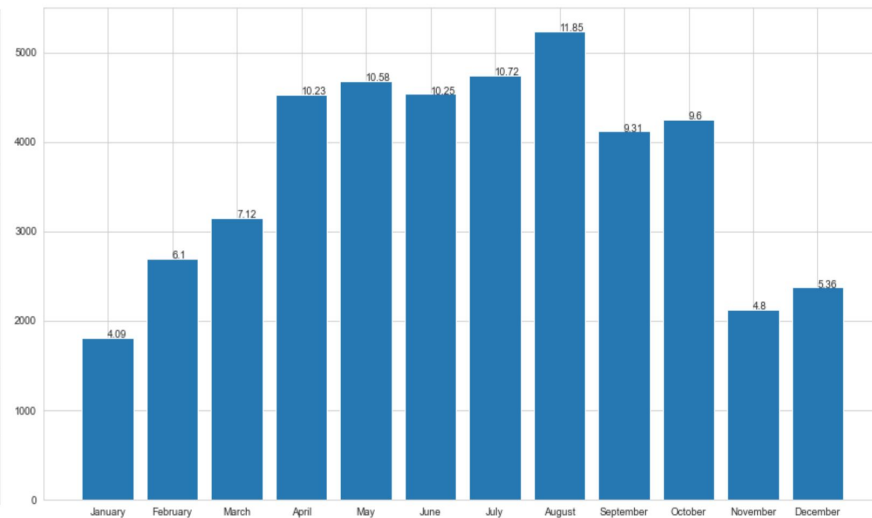
Resumen de los datos encontrados.

Los meses en los que se hacen más reservaciones son los meses de **primavera-verano**, desde Abril hasta Agosto, y estos mismos meses son en los que se encuentran el mayor porcentaje de cancelaciones.

Meses más ocupados



Meses con más cancelaciones



Importancia de variables

Con todos los datos y variables que se proporcionaron se hizo un análisis de importancia de variables para identificar cuales aportan más a la predicción de reservaciones canceladas.

Las 3 variables con mayor importancia son: *'lead_time'* que corresponde al tiempo entre la fecha de la reservación y cuando se hace, *'adr'* que es el precio promedio por noche por reservación y *'deposit_type'*.

Algunas de las variables que se eliminaron porque su importancia es mínima fueron *'babies'*, *'is_repeated_guest'*, *'days_in_waiting_list'* y *'previous_bookings_not_canceled'*.

	Feature	importance
0	lead_time	0.136052
13	adr	0.089061
73	deposit_type_Non Refund	0.069379
72	deposit_type_No Deposit	0.064497
2	arrival_date_day_of_month	0.062787
15	total_of_special_requests	0.057886
1	arrival_date_week_number	0.051109
4	stays_in_week_nights	0.042004
36	country_PRT	0.039197
35	country_OTRO	0.038717
3	stays_in_weekend_nights	0.026908
43	market_segment_Online TA	0.026583
9	previous_cancellations	0.025544
14	required_car_parking_spaces	0.021293
11	booking_changes	0.019194

Resultados del modelo



Se probaron 3 modelos diferentes para predecir cuando las reservaciones se cancelarán.

Modelo	Score Accuracy
Regresión logística	80.3%
SGD Classifier	80.7%
Random Forest	88.5%

El mejor resultado se obtiene con un **Random Forest** simple, sin ningún ajuste de hiper parámetros. Se puede continuar trabajando con este modelo para intentar mejorar este performance.