

Trabajo Práctico N°1
Reservas de Hotel

PRIMER CUATRIMESTRE DE 2023

Materia: Organización de Datos

Grupo N° 5

Integrantes:

Padrón	Apellido y Nombre
97529	BATALLAN, DAVID LEONARDO
98942	ROSELLÓ, JIMENA
104880	LOURENGO CARIDADE, LUCIA

Análisis Exploratorio y Preprocesamiento de Datos

Checkpoint 1:

A lo largo de este trabajo utilizaremos el dataset provisto por la cátedra, el archivo hotel_train.csv. El objetivo del trabajo será lograr predecir si una reserva será cancelada. Lo importamos mediante la librería pandas.

El dataset corresponde a un listado de las reservas de 2 hoteles: "City Hotel" y "Resort Hotel".

Analizamos cada una de las variables y las clasificamos en cualitativas, cuantitativas e irrelevantes.

Para las variables cuantitativas, buscamos su media, mediana y moda, y mediante histogramas analizamos sus distribuciones.

Para las variables cualitativas, estudiamos los valores que toman y cuán frecuentemente lo hacen mediante gráficos de barras y torta.

Por otro lado, analizamos la correlación existente entre las variables e hicimos un heatmap para visualizar la relación entre las variables y el target.

Por último, se detectaron los valores atípicos en los datos en forma univariada y multivariada. Para ello, estudiamos cómo se encuentran distribuidos los datos mediante gráficos boxplot y visualizamos los que se alejan de la distribución esperada.

Gráficos:

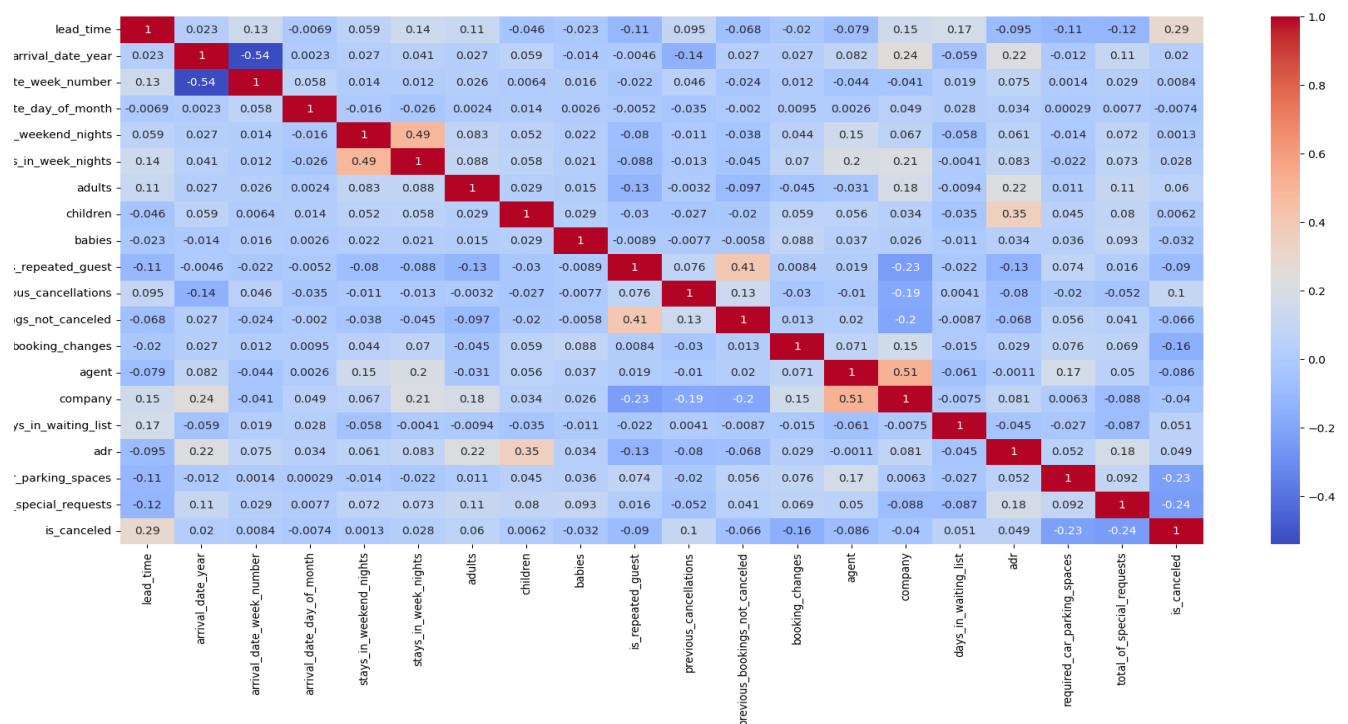


gráfico 1: de correlación entre las variables y el target

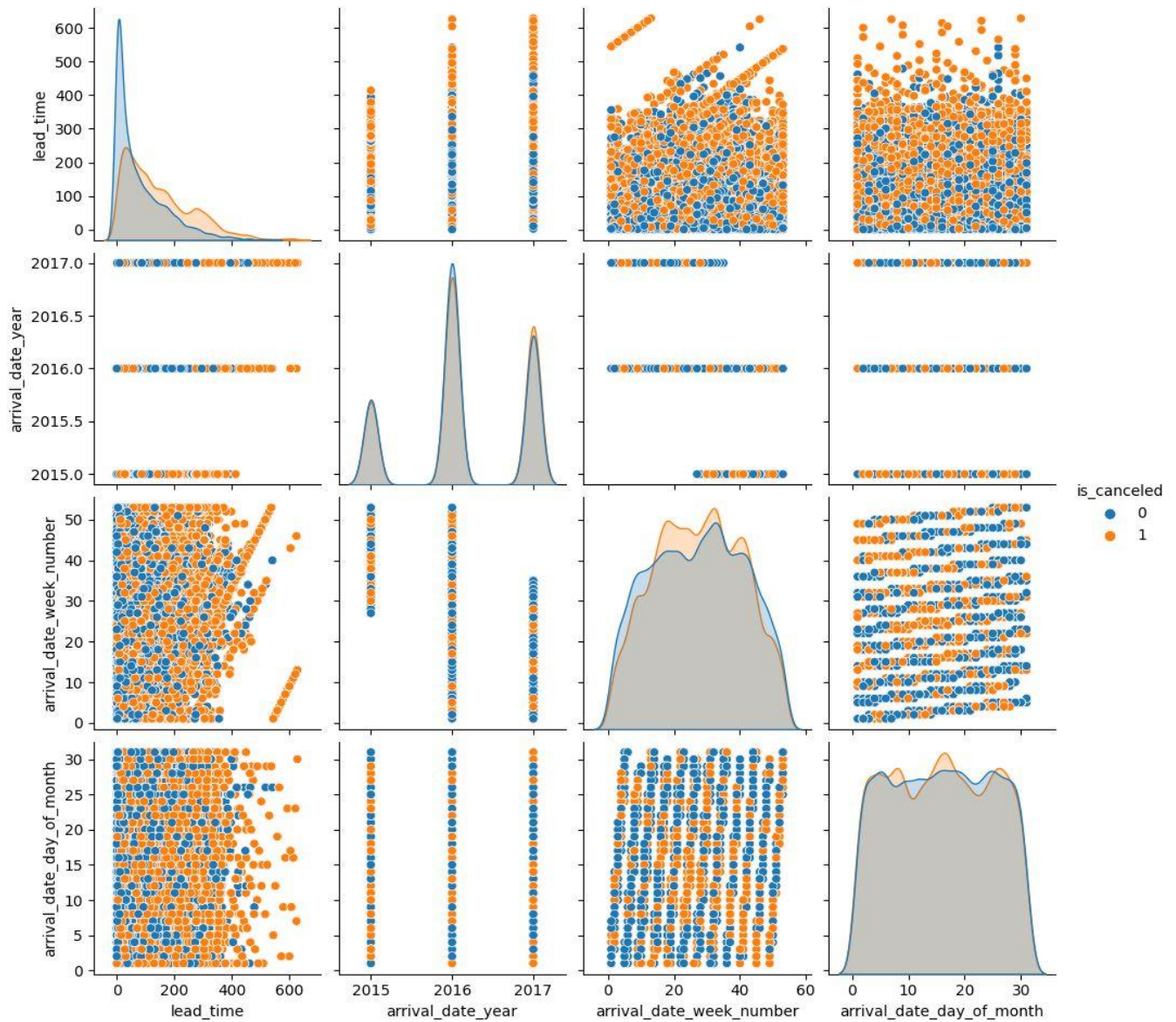


gráfico 2: Gráfico de a pares entre las variables lead_time y arrival year; week number y day_of_month

A partir del análisis encontramos que, entre las variables cuantitativas, las que mayor correlación mantienen con el target (is_canceled) son lead_time, total_of_special_requests y required_car_parking_spaces.