

Checkpoint 1 - Grupo 28 - The kingdom

Hotel Booking Demand

Es un dataframe es un conjunto de reservas de hoteles que consta de 61,913 filas (registros) y 31 columnas (variables), donde cada registro es una reserva de hotel y las variables son distintas que nos brindan información de los registros.

En nuestra notebook trabajamos de la siguiente forma:

Análisis y procesamiento de Datos

En esta etapa importamos lo necesario para trabajar con el dataframe. Analizamos las cantidades de fila. los tipos de dichas columnas/variables.

Describimos las variables según su significado y su tipo cuantitativo / cualitativo..

Análisis Univariado

Estudiamos las distintas variables cuantitativas / cualitativas.

Cuantitativas

En cada una de estas variables:

- Análisis Estadístico
- Analisis de Valores faltantes
- Análisis Gráfico: usamos countplot o boxplot para detectar valores atípicos.
- Análisis de Outliers: eliminamos o reemplazamos a partir de detectar esos valores atípicos a partir del análisis del gráfico.

Después de evaluar las variables cuantitativas detectamos:

- Encontramos valores nulos/faltantes como en **adr**, en **children**: eliminamos esos registros al ser porcentajes mínimos.
- Encontramos valores atípicos en **previous_booking_not_canceled** : eliminamos los outliers pero al no considerarse relevante esta información para el target , eliminamos esta variable.

Cualitativas

En cada una de las variables analizamos :

- Análisis Estadístico
- Analisis de Valores faltantes
- Analisis de Grafico

Después de evaluar encontramos que:

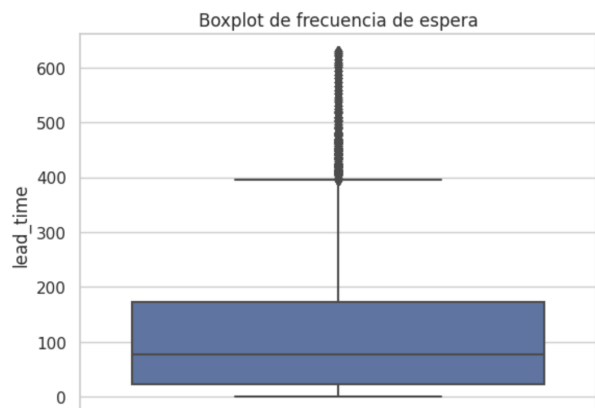
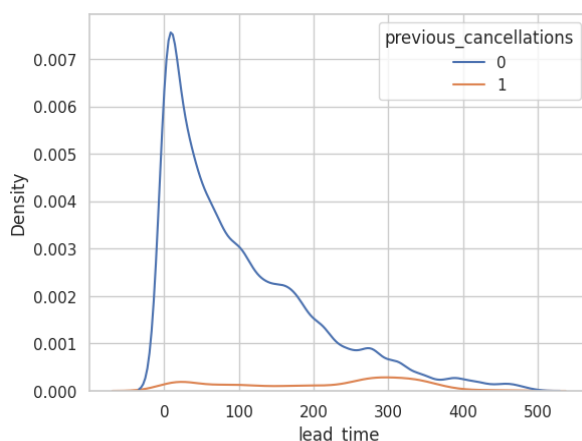
- Los de los valores faltantes eliminamos tanto a la variable **company** por tener un porcentaje alto de falta de datos como la variable **agent** por tener muchos valores faltantes. La variable country se reemplazaron los valores por el más frecuente.
- Los de valores atípicos como: **lead_time**, **children**, **adults**, **days_in_waiting_list** se decidieron eliminar las reservas o reemplazar los valores.

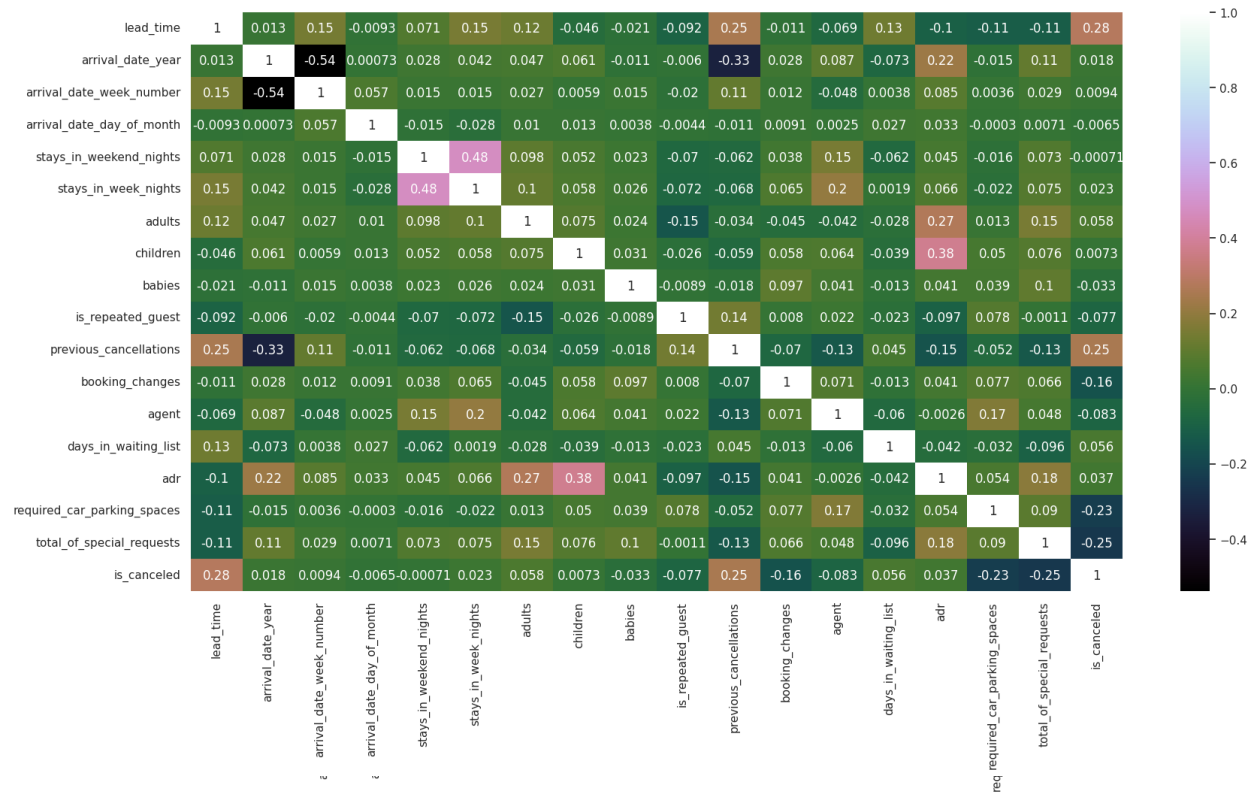
Análisis Multivariado

Analizamos las parejas de variables mediante gráficos y sacando conclusiones. Estas son las correlaciones que surgieron:

- **stays_in_week_nights** y **stays_in_week_nights**
- **children** y **adr**
- **previous_cancellation** y **lead_time**
- **lead_time** y **is_canceled**

En el caso de la correlación **previous_cancellation** y **lead_time** podemos ver en su gráfico que al ser tener menos días del día de reserva al día de llegada mayor habrá mayor posibilidad de ser cancelada.





Tareas Realizadas

Integrante	Tarea
Jurgens Cecilia Ines	Análisis de variables/outliers Análisis de Correlaciones
Soto Nicolo Marilyn	Detección de Outliers Análisis de Variables/ valores faltantes
Schipani Martin Sebastian	Armado de Reporte Análisis de Correlaciones/outliers

