

# TP 1: Reservas de Hotel

## Grupo 35

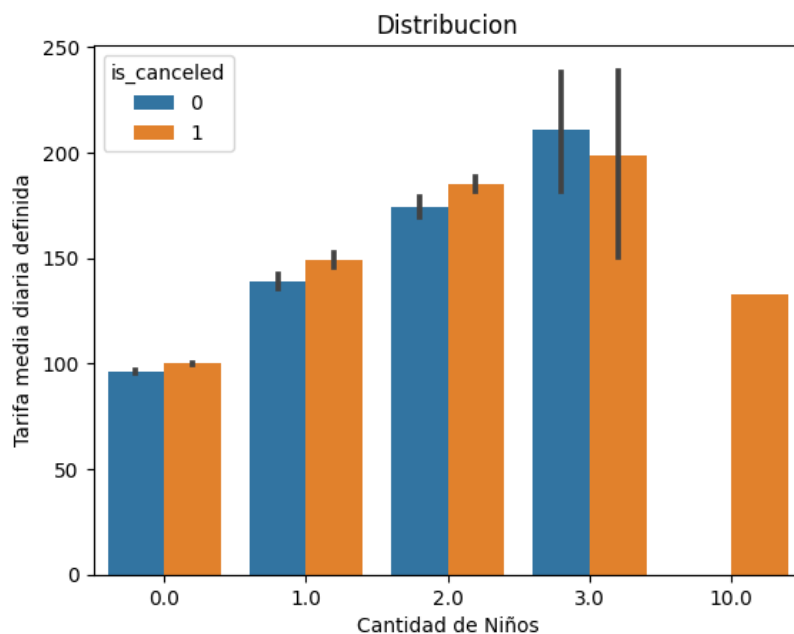
### Integrantes:

- Aylas, Brenda
- Cori, William
- Nazario, Ingrith

A lo largo de la notebook se fueron escribiendo los distintos descubrimientos del dataset.

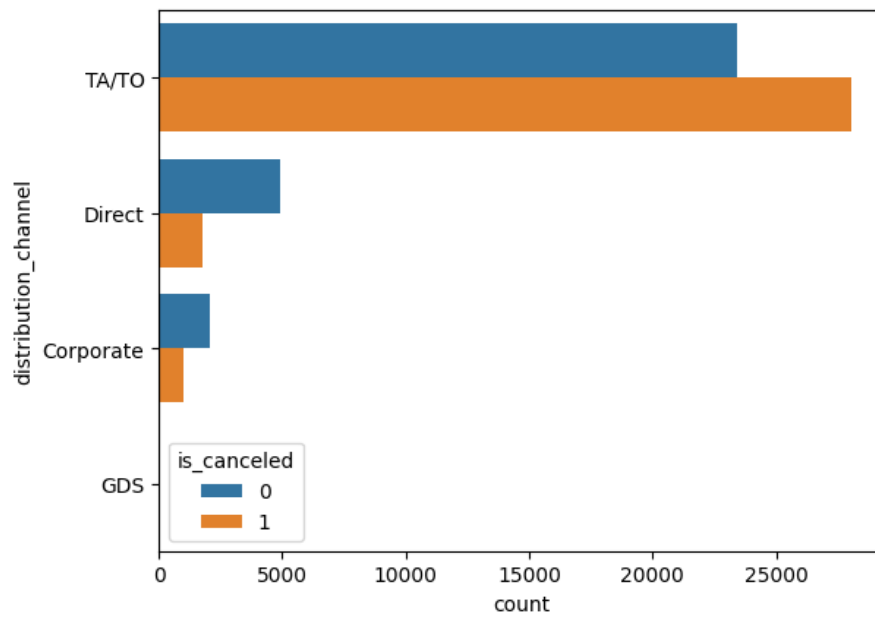
Se analizaron los tipos de datos, nulos y valores atípicos como así también las distribuciones de las variables tanto de forma individual como de a pares.

Entre ellas una de las más destacables fue la relación que pudimos visualizar a través de heatmap entre las variables ADR vs children (0.35) la que vemos que crecen de forma lineal ambos valores:



También nos llamó la atención el siguiente gráfico en el que a pesar de que la cantidad de registros informados por categoría son diferentes en cantidad se podría decir que si el canal que se usó para reservar el hotel fue de forma directa es mayor la probabilidad de que no se cancele una reserva:

```
[92]: <Axes: xlabel='count', ylabel='distribution_channel'>
```



Una de las columnas que decidimos eliminar fue la de compañía ya que había gran porcentaje de valores nulos y no eran nombres sino IDs lo que no nos daba información para utilizar para predecir el target. De forma contraria, para la variable de agente realizamos imputación para poder completar los datos faltantes con una de las técnicas aprendidas en clase.