

**Materia:** 75.06/95.58 - Organización de Datos

**Cuatrimestre:** 1° cuatrimestre 2023

| Padron | Apellido y Nombre |
| --- | --- |
| 87943 | Gomez, Juan Angel |

* respecto del punto a y c
  + se ven bastantes nulos en la variable company, se decide borrarla ya que involucra casi el 95% de valores nulos, es irrelevante para el analisis
  + se ven pocos Nan en children, con solo 4 y se los preemplaza por el valor mas cerca de la media.
  + respecto de company, se interpreta que si no tiene valor, implica que el viaje lo hicieron por si mismos, se reemplaza por ceros
  + Conclusion de vinculacion con la variable target: la que mas rekacionada esta es con lead time
* visualizacion de los datos
  + la mayor parte de las reservas viene de portugal (la moda)
* respecto de valores atipicos:
  + la deteccion lo haria con la funcion que calcula los posibles, que pone en la mira los valores que se desvian mas de 1.5 veces el rango intercuartilico
  + podrian ser que haya niños sin adultos supervisandolos (adults, babies en 0)

en este caso se eliminarian

* + un grupo de varias personas que son estan en un grupo, ejemplo el registro con mas de 9 adultos sin que pertenezca a un grupo, es para tenerlo en la mira.
  + se ve del filtrado y del dataset que la gran mayoria de los registros estan con menos de 5 adultos