

Informe TP1 : Reservas de Hotel

Integrantes:

Agama Avila, Arely - 105829
Martinez, Selene Anahi - 100439
Meichtri, Melany - 102330

Checkpoint 1: Análisis Exploratorio y Preprocesamiento de Datos

El objetivo es realizar un análisis con el dataset brindado 'hoteles_train.csv', analizando de qué variables dependerá que las reservas sean canceladas. Mostrando gráficos de cada atributo, detectando los datos faltantes y valores atípicos y con ellos tomar una decisión ya sea eliminando o modificando.

a) Exploración Inicial

Básicamente, clasificamos los atributos en variables cualitativas y cuantitativas. Mostrando los valores que toman y la frecuencia en que lo hacen. Decidimos cambiar el nombre del atributo 'adr' dado que no daba mucha información, siendo cambiado por 'average_daily_rate'. También cambiamos 'arrival date month', 'arrival_date_year', 'arrival_date_day', esto para poder juntar estos atributos en uno solo y castearlo a datetime más adelante. Así mismo, mostramos gráficos de las variables, las correlaciones correspondientes y la relación de las variables con el target(is_canceled). Cada uno está explicado detalladamente en la notebook.

b) Visualización de los datos

En esta sección mostramos algunas predicciones que tuvimos, sobre qué variables pueden estar más relacionadas con el target, mostrándolo mediante gráficos y justificando su utilización.

c) Datos faltantes

Analizamos el porcentaje de datos faltantes encontrados en nuestro dataframe. Tomando decisiones eliminando atributos o modificandolos. Hacemos un análisis previo para ver la cantidad de datos faltantes por variable, tomamos la decisión y mostramos que la distribución del antes y después se conserve.

d) Valores atípicos

Analizamos datos atípicos para las variables teniendo en cuenta valores posibles razonables dado el contexto del dataset y donde se concentran la cantidad de valores de las variables.

Dando un ejemplo, definimos como valores atípicos reservas sin integrantes o cuyos integrantes sean solo bebés.