

# Parrot 9기 전처리 미니 프로젝트

## 1. Dataset

- ✓ hotel: Hotel (H1 = Resort Hotel or H2 = City Hotel)
- ✓ is\_canceled: Value indicating if the booking was canceled (1) or not (0)
- ✓ lead\_time: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
- ✓ arrival\_date\_year: Year of arrival date
- ✓ arrival\_date\_month: Month of arrival date
- ✓ arrival\_date\_week\_number: Week number of year for arrival date
- ✓ arrival\_date\_day\_of\_month: Day of arrival date
- ✓ stays\_in\_weekend\_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- ✓ stays\_in\_week\_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
- ✓ adults: Number of adults
- ✓ children: Number of children
- ✓ babies: Number of babies
- ✓ meal: Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)
- ✓ country: Country of origin. Categories are represented in the ISO 3155–3:2013 format
- ✓ market\_segment: Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- ✓ distribution\_channel: Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- ✓ is\_repeated\_guest: Value indicating if the booking name was from a repeated guest (1)

or not (0)

- ✓ previous\_cancellations: Number of previous bookings that were cancelled by the customer prior to the current booking
- ✓ previous\_bookings\_not\_canceled: Number of previous bookings not cancelled by the customer prior to the current booking
- ✓ reserved\_room\_type: Code of room type reserved. Code is presented instead of designation for anonymity reasons.
- ✓ assigned\_room\_type: Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons.
- ✓ booking\_changes: Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
- ✓ deposit\_type: Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.
- ✓ agent: ID of the travel agency that made the booking
- ✓ company: ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons
- ✓ days\_in\_waiting\_list: Number of days the booking was in the waiting list before it was confirmed to the customer
- ✓ customer\_type: Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking
- ✓ adr: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
- ✓ required\_car\_parking\_spaces: Number of car parking spaces required by the customer
- ✓ total\_of\_special\_requests: Number of special requests made by the customer (e.g. twin

bed or high floor)

- ✓ reservation\_status: Reservation last status, assuming one of three categories: Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – customer did not check-in and did inform the hotel of the reason why
- ✓ reservation\_status\_date: Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel

## 2. 과제 내용

- ✓ 본 과제는 EDA(데이터 탐색/정제/전처리)가 목표입니다!
- ✓ 데이터셋 변수 특성 파악
- ✓ 결측치, 이상치 처리
- ✓ 변수 정규화(표준화)와 리스케일링
- ✓ 변수 간 상관관계 분석
- ✓ 시각화
- ✓ 불필요한 변수 제거

## 3. 제출 기한과 방법

- ✓ 제출 기한은 10/8(일) 11:59입니다. 추석 연휴동안 조원들과 상의하여 프로젝트를 진행해 주세요! 제출 기한 엄수 바랍니다.
- ✓ 조마다 1명이 대표로 패럿 네이버 카페(<https://cafe.naver.com/sgparrot2019>) 에 제출해 주시면 됩니다.
- ✓ 파일명은 “\_\_조\_전처리\_프로젝트.ipynb” 로 해주세요.