Module 01: Introduction to Python

Pandas Group Quiz Exercise

Dataset Description

This dataset provides a detailed profile of hotel bookings for two hotels

The columns are as follows:

1. Hotel: Hotel (H1 = Resort Hotel or H2 = City Hotel)
2. is_canceled: Value indicating if the booking was canceled (1) or not (0)
3. lead_time: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
4. arrival_date_year: Year of arrival date
5. arrival_date_month: Month of arrival date
6. arrival_date_week_number: Week number of year for arrival date
7. arrival_date_day_of_month: Day of arrival date
8. stays_in_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
9. stays_in_week_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
10. adults: Number of adults
11. children: Number of children
12. babies: Number of babies
13. meal: Type of meal booked.
14. Country: Country of origin.
15. market_segment: Market segment designation.
16. distribution_channel: Booking distribution channel.
17. is_repeated_guest: Value indicating if the booking name was from a repeated guest (1) or not (0)
18. previous_cancellations: Number of previous bookings that were cancelled by the customer prior to the current booking
19. previous_bookings_not_canceled: Number of previous bookings not cancelled by the customer prior to the current booking
20. reserved_room_type: Code of room type reserved.
21. assigned_room_type: Code for the type of room assigned to the booking..
22. 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
23. deposit_type: Indication on if the customer made a deposit to guarantee the booking.
24. Agent: ID of the travel agency that made the booking
25. 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
26. days_in_waiting_list: Number of days the booking was in the waiting list before it was confirmed to the customer
27. customer_type: Type of booking, assuming one of four categories:
28. Contract - when the booking has an allotment or other type of contract associated to it.
29. Adr: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

30. required_car_parking_spaces: Number of car parking spaces required by the customer
31. total_of_special_requests: Number of special requests made by the customer (e.g. twin bed or high floor)
32. reservation_status: Reservation last status
33. reservation_status_date: Date at which the last status was set.

1. Print the column names of the data frame.
2. Find the unique hotel names.

There are two hotels in the dataset. Make two separate data frames for both the hotels and do the following for both of them:

3. Use shape command on the datasets and find the number of samples in each dataset.
4. Find the percentage of cancelled bookings for each hotel.
5. Take all the bookings where is_canceled == 0 from each dataset. Basically you need to remove all the rows where is_canceled == 1.
6. Find the month which has the highest number of bookings using the **mode** command.
7. Find the most frequent room type reserved.
8. Find the average lead time for the bookings.
9. Find the average number of adults who stayed for each booking.
10. Find the unique distribution channels for the bookings made.
11. Using pivot table find the total number of adults who stayed in each hotel grouped by the distribution channels they used for the booking.
12. Identify the unique deposit types. Find the most frequent deposit type.
13. Find the unique country of origin of the people who made bookings to the hotel.
14. Using pivot table find the aggregate number of people who have made bookings from each unique country of origin.
15. At how many occasions reserved room type was not the same as the assigned room type.
16. Find the average length of the stay. For this you need to sum the weekend and weekday stay columns and the find the mean.

Submission: Upload your notebooks to the classroom using the name convention, i.e. YourName_M1_PandasQuiz