

HANDLING MISSING VALUES USING POWER-BI

- Handling missing values is one of the most crucial parts of predicting the models using machine learning.
- There are different kinds of methods for handling missing values and different platforms can be used to handle the missing.
- Handling missing values in Power BI involves identifying and removing or replacing missing values to ensure that the data is accurate and meaningful.

STEPS USED TO HANDLE THE MISSING VALUES IN THE COLLECTED DATASET

1. The dataset that I collected from the open source was hotel booking.
2. There are a lot many missing values in this dataset.

```
hotel                0
is_canceled          0
lead_time            0
arrival_date_year     0
arrival_date_month    0
arrival_date_week_number 0
arrival_date_day_of_month 0
stays_in_weekend_nights 0
stays_in_week_nights  0
adults               0
children             4
babies               0
meal                 0
country              488
market_segment        0
distribution_channel  0
is_repeated_guest     0
previous_cancellations 0
previous_bookings_not_canceled 0
reserved_room_type    0
assigned_room_type    0
booking_changes       0
deposit_type          0
agent                16340
company              112593
days_in_waiting_list  0
customer_type         0
adr                  0
required_car_parking_spaces 0
total_of_special_requests 0
reservation_status     0
reservation_status_date 0
dtype: int64
```

3. So, by analyzing the count of missing values in python, the column “company” was removed.
4. The agent’s null values were replaced by 0.
5. The null values in the country were dropped by deleting that whole row.

In the above 5 steps, the null values are being replaced and removed, now we should take look at some domain knowledge.

The hotel booking will be done by either children or adults but not with babies, so as per that the filtration we should perform is to remove([adults] == 0 and [children] == 0 and [babies] == 0)

6. The condition was applied to the 3 columns and the filtration was done.
7. The 3 new columns are being added.
 - $\text{Total_members} = \text{adults} + \text{children} + \text{babies}$
 - $\text{Total stays} = \text{stays_in_week_nights} + \text{stays_in_week_end_nights}$
 - $\text{Revenue} = \text{adr} * \text{Total stays}$
8. The reason to add these columns is to help to visualize the information more detailed.