**Technical document**

This EDA project subject to the analysis “Hotel Bookings” of dataset 1,19,390\*32 provided by the Alma Better for the capstone project in during the analysis we found 31,994 duplicate rows, after removing them we also removed the company column which we found no so useful for further analysis, net shape of the dataset in use is 87,396\*31.

Dataset Name – “Hotel Bookings”

Library used – Pandas as pd

Numpy as np

Matplotlib as plt

Seaborn as sns

**Points under consideration: -**

* Customer type preferred the most resort hotel.
* Hotel refunded the most
* Distribution channel is mostly used for booking City hotel
* Hotel preferred by customer having children.
* Hotel most preferred for weekend night.
* Highest booked and cancellation based on market segment (complementary).
* Preferred food by customer based on top 5 country.
* Most favorable month for booking.
* Top five agents of city hotel
* Hotel data based on month where repeated guest is 1

**Project Summery**

1. Transient types of customers visited to Resort Hotel most.
2. City hotel has non refunded the most 845 compared to Resort Hotel having 193.
3. TA/TO is most preferred for City Hotel 44549.
4. City hotel is more preferred by customers having children with age group as per dataset.
5. City hotel is most preferred for weekend nights.
6. The out of 2015, 2016, 2017 2016 has the highest booking and cancellation based on complimentary
7. BB is most preferred by the portugal
8. August month is most preferred for the booking
9. 9.0 is among the top five agents for the City Hotel
10. So, March has the most Repeated guest 418

**Bibliography**

**pandas documentation** <https://pandas.pydata.org/docs/>

**Geeksforgeeks**  https://www.geeksforgeeks.org