**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  |  | | --- | --- | --- | --- | | **Sl. No** | **Name** | **Email** | **Contribution** | |  | Rohith V | rohithdevang13@gmail.com | 1. Data Clean up: 2. Created the function to find out the percentage of Null Values and visualized using bar plot. 3. Tackled all the Null values and handling of the missing data’s were done. 4. EDA performing: 5. EDA Based on Cancelation Bookings:  * Performed the bookings cancelled based on months. * Performed hotel which has high cancelation rate.  1. EDA based on Meal category:  * Performed EDA on which meal category was booked in each of the hotel.  1. EDA based on Country:  * Performed EDA on which country the highest booking was made.  1. EDA Based on agents:  * Performed EDA, whether bookings involved the help of agents or not? | |  | Abhishek V L | abhisheklakshmana7@gmail.com | 1. EDA Performing:  * Created the Correlation Heatmap for the numerical data. * Checked for the outliers present in the data and removed the outlier.  1. EDA based on the arrival period:  * Performed EDA for the customer arrival period based on years, months and days in months.  1. EDA based on Market Segment and Distribution Channel:  * Performed EDA on mode of bookings done majorly. | |  | Nikita Hubballi | hubballinikita111@gmail.com | 1. Data processing:  * Created 2 columns for Total stay in hotel by concocting the duration of stays made in weekend and weekday nights and for total people by concocting the adults, children and babies.  1. EDA performing: 2. EDA based on the hotels:  * Performed EDA on which among hotel the highest booking was made. * Performed EDA on the duration of stays made in each of the hotel.  1. EDA based on Weekday and Weekend day nights.  * Performed EDA on number of stays made in weekday nights. * Performed EDA on number of stays made in weekend day nights.  1. EDA based on repeated guest.  * Performed EDA on whether the guests were repeated or not? And how many of them are repeated guests | | 4. | Aditi Sharma | aditisharma98014@gmail.com | 1. Data Processing:  * Removed all the duplicate data’s which were present in the dataset.  1. EDA Performing: 2. EDA based on type of Visitors:  * Analyzed the performance of types of visitors visited the most.  1. EDA based on Deposit types:  * Analyzed the type of deposit was made more.  1. EDA on Room types:  * Analyzed which type of room was mostly preferred by single adult. * Analyzed which type of room was mostly preferred by couple travelers. * Analyzed which type of room was mostly preferred by group travelers. | |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/Abhisheklakshmana/Hotel-Booking-Aanalysis.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  **Summary :**  **Firstly we took this project because of the mutual interest in travelling, so at least we had a basic knowledge what all could have in the data set. And we got some of the data’s which I was expected.**  **After downloading the data set, we just divided according to the columns and then data cleaning and handling of missing values was done, then we performed few of the EDA’s. Here we analyzed the EDA’s with the help of graphs with the help pf matplotlib.pyplot and seaborn.**  **Problem Statement:**  **The dataset represents the data of City hotels and Resort Hotels. The dataset have 32 variables with 119,390 observations. Each observation represents a hotel booking. The datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were canceled. Since this is hotel real data, all data elements pertaining hotel or customer identification were removed.**   1. **In which of the hotel the highest booking was made** 2. **In which type of hotel people prefer stay in terms of days** 3. **Cancelation Done Based on months** 4. **High Cancelation rate in hotels** 5. **Arrival Seasonality based on years, months and days** 6. **Check whether the guest is repeated or not** 7. **Mode of bookings through market segment and distribution channel** 8. **When was the majority stays made?** 9. **What type of meal preferred by the customer in each hotel** 10. **In which country the highest booking was made** 11. **What kind of visitors travelled more?** 12. **Which was the most preferred deposit type made** 13. **Which kind of rooms were assigned for different adult size**     **Conclusion:**  **From the given dataset, after performing EDA based upon all of the above mentioned, we conclude the inferences which we found were:**   1. **Majority of the hotels booked are city hotel. Definitely need to spend the most targeting fund on those hotel.** 2. **People also prefer to stay for longer duration in Resort Hotels and prefer City Hotels for Shorter duration.** 3. **We also realize that the high rate of cancellations can be due high no deposit policies.** 4. **We should also target months between May to Aug. Those are peak months due to the summer period.** 5. **We see there is less amount of repeated guest.** 6. **Majority of the bookings are done through online travel agents.** 7. **Most the peoples prefer to stay in week day nights compared to weekend’s night.** 8. **Majority of the peoples prefer BB (Bed & Breakfast) category in the meal section.** 9. **Generally couples travel most from the given data.** 10. **From the given set of data Portugal is the country where majority of the bookings were done.** 11. **Majority of the hotels does not require deposit type, so this may also be the reason for the high cancelation rates.** 12. **Lastly coming to room type, solo travelers prefer type A rooms, couples prefer type A rooms and traveling in a group or families prefer G type of rooms.**  * **There was a lot of duplicate data.** * **Choosing appropriate visualization techniques to use was difficult.** * **A lot of null values were there in the dataset.** |