Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. Swapnil Wankhede
swapnilwankhede8600@gmail.com

Contribution:

Data Exploration and Data Cleaning
- Loading dataset and exploration of data
- Removing duplicates rows
- Handling Null Values
- Converting Columns to appropriate data types
- Creating new columns

Analysis on problems.

2. Sahil Jagtap
jagtapsahil098@gmail.com

Contribution:

Data Exploration and Data Cleaning
- Loading dataset and exploration of data
- Removing duplicates rows
- Handling Null Values
- Converting Columns to appropriate data types
- Creating new columns

Analysis on problems.

3. Shweta Jagtap
shwetajagtap500@gmail.com

Contribution:

Data Exploration and Data Cleaning
- Loading dataset and exploration of data
- Removing duplicates rows
- Handling Null Values
- Converting Columns to appropriate data types
- Creating new columns

Analysis on problems.

Please paste the GitHub Repo link.

Github Link: https://github.com/wankhede10swapnil/Hotel_Bookings_Analysis

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Exploratory Data Analysis in the hotel industry is key to marketing strategy, building customer trustworthiness and enhancing productivity. By analyzing the patterns available in the given dataset which has 119390 observations with 32 features including some missing and duplicates values. It contains the information for a city hotel and a resort hotel, and includes several variables such as when the bookings were made, length of stay, the number of adults, children, and etc. All personally identifying information has been removed from the data. To Explore and analyze the data to discover when is the best time to book hotel? what would be the optimal length of stay in order to get the best daily rate? Comparison of hotel's ADR week wise and many more important factors that have been analyzed.

It is observed that data contain lots of duplicate's data, lots of null values and some features are not in correct datatype format after loading the dataset. That was the tough challenge. Removing duplicates rows by using drop function, handling null values, converting columns to appropriate data types and creating new columns then updated dataset has nearly 81228 observations with 33 variables. some actions which have been taken to overcome the problem and to make the better analysis using NumPy, Pandas, Matplotlib, seaborn libraries.

Some problems are taken into consideration and are analysed. It is concluded that August is the best time of year to book a hotel room. City hotels are mostly booked in weekend nights and week nights and maximum number of cancellations of booking as compare to resort hotel. Most of the rooms of city hotels are sold by TA/TO distribution channel as compared to resort hotel. The city hotel got the higher number of special requests. More booking is in august month followed by July month. ADR for the week 28 to 35 are more for the resort Hotel than city hotel. In 2016, maximum number of reservation status is checked out. For every customer, BB (Bed and Breakfast) is most preferable meal type. Portugal, United Kingdom, France and Spain are the top four countries where the bookings are maximum. The best optimal length of stay is 1 to 10 week and weekends nights in order to get the best daily rate. Average daily rate (ADR) is positively correlated to children, adults, and babies with 33%, 25% and 2% respectively. It means that maximum and minimum revenue is generated by occupied rooms with children and babies respectively. It is observed that arrival date week number and arrival date year are 51% negatively co-related. In short, this project gave real insights of hospitality sector.

Drive link: https://drive.google.com/drive/folders/1w2bsJj3wtlKfkWUn1zCe-SQ4U3VFYSkF?usp=sharing