**FINAL PROJECT REPORT**

**Nitish Kumar Reddy Vinta (82702859)**

**Introduction:**

The ability to get and investigate data is main part in the business. Data and information represent a lot of information related to data and makes it more useful by outlines, charts and other visible tools. Similarly, hotel management can detect similar patterns, make hypothesis, and build a better independent way. There are few separate ways that they can use this information to get perfect advantage of this project.

In Travel and Tourism market, Hotel market is one of the most popular. The projected revenue growth for hotel segment is 312.30bn in 2022. The analysts are expecting 11.34% growth in revenue incoming four years (2022-2026). An average individual is spending about $0.34k. These are some of the statistics published in “Statista”.

I also found some remarkably interesting statistics in this hotel market. So, I decided to investigate this segment to do visualizations. But later I got to know is getting related data for this market is difficult and we need to pay to the third parties to access their data. Finding direct data related to my project has been difficult. There are few datasets in this segment which are floating online, but they are not matching my research questions. At last, I found only one dataset in Kaggle which have many unwanted variables. So, I decided to use that dataset with only the variables I need.

My main idea behind this project is to predict and understand the trend the hotel booking system for the past few years. Hotel management collects the data for room availability and the total number of bookings to increase the cost of a room. To analyze the business logic behind all this I thought this would be good project to go with, unfortunately there is not much data available to public. But I found one “Hotel booking demand” dataset in Kaggle. The second data set I used is “Popular Indicators” from the World Bank.

My goal is to find some of the research questions like:

* What is the peak season for hotel bookings? How long are they staying in the hotel?
* Do the GDP of a country will change the hotel booking system?
* Which market segment plays majority role in hotel bookings?

**Methodology:**

The Hotel Demand dataset holds reservation details of two types of Hotels: City Hotel and Resort Hotel. It is vast data set which has 119,390 records of hotels from different countries. For this project I didn’t do any data cleanup as our main goal is to focus on visualizations from combining the two datasets available.

Popular Indicators dataset has Average GDP, GDP per capita and population for all the countries for the years 2015, 2016 and 2017. I have taken these three years only because I have the booking details for only these years.

Variables which are used for this project which are present in “Hotel booking demand” dataset.

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| Hotel | Type of the Hotel:  Resort Hotel  City Hotel |
| is\_canceled | Indicating booking is cancelled or not. |
| lead\_time | Number of days before the arrival data. |
| arrival\_date\_year | Arrival year |
| arrival\_date\_month | Arrival month |
| arrival\_date\_week\_number | Arrival week number |
| arrival\_date\_day\_of\_month | Date of arrival. |
| stays\_in\_weekend\_nights | Number of nights stays in weekends |
| stays\_in\_week\_nights | Number of nights stays in weekdays. |
| adults | Number of adults |
| children | Number of children |
| babies | Number of babies |
| country | Country codes of different countries |
| market\_segment | Market segment designations.  (Offline TA/TO – offline travel agents and tour operators.  Online TA- online travel agents |
| is\_repeated\_guest | Value showing is the guest came earlier or not |
| previous\_cancellations | Information about earlier cancellations. |
| reserved\_room\_type | Code for room type reserved.  \*Information about room types is not mentioned properly. |
| deposit\_type | Type of deposit made:  No Deposit  Deposit  Nonrefundable deposit |

Variables which are used for visualization which are present in “Popular Indicators”

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| GDP per capita | GDP per person  (Value is in U.S $) |

There are many variables present in the “Hotel booking demand” and “Popular Indicators” dataset which I didn’t want to use. So, I mentioned only few variables in those datasets which I used for the visualizations. But full dataset has been loaded into the tableau to create required visualizations. I have used link method to connect the two tables. I joined them with “country” variable in hotel demand booking dataset to “country code” in popular indicators.

Main dataset “Hotel booking demand” is sourced from Kaggle, but it is originally from the article “Hotel Booking Demand Datasets” (Antonio, 2019).

**Analysis:**

**Visualization 1:**

**Total Time spent in two hotels and their reserved room types**

**Graphical user interface, chart, scatter chart

Description automatically generated**

In this visualization, I did analysis for the total time spent by guests in two different hotels based on their room types. As we can see city hotels has the highest value other than the resort value. At the bottom of the visualization, we can also see the different room types, where room type “A” is the most preferred room other than all other types in both city and resort hotels.

In city hotels, many of the guests are not interested in “P “& “C”. The time spent on these two room types are very less compared to other room types.

In resort hotels, less preferred room types are “L” & “B”. The time spent on these two rooms are less.

**Visualization 2:**

**Cancellations vs Booking Changes for 2015, 2016 & 2017**

**Chart, bar chart, histogram

Description automatically generated**

From the above visualization, line graph shows the cancellation status, and the bar graph shows booking changes taken place in every month in 2015,2016 till September of 2017.

The highest booking changes was obtained in October of 2016 and lowest booking changes were seen July 2015. As we see in this there is some missing data in the left side of the graph so there is a huge drop in the graph. Let’s not consider this for our analysis. We should consider only in the middle pattern where there are vast amount of data. I can say that there is some pattern in the middle also.

January month of both 2016 and 2017 has the highest number of room cancellations. I can expect it might be similar in 2015 also as there is not much difference in the values of 2016 & 2017. The cancellations pattern is kind of simar for mid months of the year.

**Visualization 3:**

**Countries which have City Hotels & Resort Hotels**

In the below world map, we can see it is divided into two parts based on the type of hotel. This visualization shows how many countries have resort hotels and city hotels.

Almost all the countries have city hotels, where few of the countries don’t have resort hotels. In South America, Paraguay and Bolivia didn’t have any resort hotels. We can observe this difference if we see Africa clearly, few of the African countries don’t have resort hotels as majority of them have city hotels.

**Graphical user interface, application

Description automatically generated**

**Visualization 4:**

**Visualization between Average GDPs per capita and deposit types**

**Chart, bar chart

Description automatically generated**

According to the bar chart observed in the visualization 4 between average gdp per capita and the people deposit types, the most percentage of the people consider the option of refundable deposit while booking their hotel reservations. The average GDP per capita on “no deposit” is 18628. I believe the people booking this type of hotels do not require a security deposit to book the room. The average GDP per capita on nonrefundable deposit is 30,853. The people booking this hotel reservation are more compared to the no deposit. The most important reason most hotels hold non-refundable is to avoid any cancellations that might happen in the last minute. This could be profit to the hotels.

Most individuals chose to the refundable deposit which is 38,386 higher than the no deposit and non-refundable. The price may be higher but there is an insurance for the individuals to cancel at any time and choose the other one. Sometimes hotels use the policy of refundable when the customers use the credit card instead of debit and cash. So, the people using the credit cards are higher compared to the other payment methods.

**Visualization 5:**

**Total Reservations made by different market segments**

**Chart

Description automatically generated**

According to the bar diagram shown in the visualization 5 between total reservations made by the different sectors, the most persons chose to reserve a hotel room using the online ticket agents which is 56,477. After that people choose to book the room by offline ticket agents which they book by phone call or meeting in the office. The total reservations in this sector are 24,219. The offline TA includes the corporate ticket agents and people who work in the physical offices rather than websites. On the third row is groups having reservations of 19811. This group module is used to make reservations for five or more guests who have same features in common.

They are all from same company and are arriving at same time. The least number of people chooses to make reservations in the aviation sector, which implies that most people don’t book a room while traveling. The hotel industry relies on the airlines to carry its demand. The corporate sector is taking up a lot of the room space in hotel bookings compared to complementary.

**Visualization 6:**

**Total count of adults, children & babies for three years**

**Graphical user interface, application, PowerPoint

Description automatically generated**

In reference to the column chart between total adults, children, and babies with respect to the three consecutive years of 2015, 2016, 2017, most of the children booked the city hotel compared to the resort hotel in all three years. This implies that children prefer to stay in the cities near the amenities, schools. Most of the babies chooses to the stay in the resort hotels compared to the city hotel.

This implies that most babies are self-occupied to having the entertainment, food etc. most of the adults chooses to stay in the city hotels compared to the resort hotels because the city hotels are in the vicinity of metropolitan areas, as they are willing to stay near the corporate offices, hotels, restaurants, airports etc. according to the data in the visualization 6, there is minimum amount of date found in the year 2015.

Resort hotels mostly focuses on the travelers. This means a smaller number of people are travelled in those years. There is less amount of data found in the year 2015, so we cannot mostly conclude there are less individuals chooses difference of resort and city hotels.

**Visualization 7:**

**Repeated guest statistics**

**Bar chart

Description automatically generated**

In this visualization, it shows If the customer is a repeated guest or a previous cancellation upon their arrival date of month.

**Dashboard 1:**

**Chart

Description automatically generated with medium confidence**

**Chart, bar chart, waterfall chart

Description automatically generated**

In the dashboard, I had used filter method for all the countries. This dashboard contains five visualizations related to hotel bookings.

**Dashboard 2:**

**Graphical user interface, chart

Description automatically generated**

In this dashboard, we have two visualizations which has a map plot and animated plot.

**Conclusion:**

After doing visualizations with the data I have, I tried to answer all the questions which I mentioned. Countries GDP per capita is also responsible for deposits the guests are making. As my research shows Online travel agents (online TA) plays key role in hotel bookings. We can expect this pattern will continue to dominate in the market.

We can also see a clear pattern of cancellations and changes in the bookings for few months. January month has the highest spike in both of them

**Other research question:**

**Visualization 8:**

**Total bookings and cancelations done over different market segments**

**Application

Description automatically generated with low confidence**

When I am surfing in web to find relevant information and datasets for my project, I found another interesting prediction that is the total hotel revenue will be generated through online sales by 2026. To analyze this trend, I tried a see if this trend is continuing over the decade.

As we can see from the above visualization Online TA (online travel agents) are making more hotel bookings than many other departments in the market segment. For bookings changes, booking the hotel or for the cancellation they are preferring Online TA over others.

Very less bookings were happening in Aviation and Complimentary segments. Groups and offline TA/TO share common market.