

Team Project

Data Visualisation (3860 - 070)

Submitted to:

Rupa Manabala

Aman Gupta

300348626

guptaa18@student.douglascollege.ca

**Details of Dataset**

The present dataset is completely distinct from the previous task because the prior dataset had only a limited number of fields and a small amount of information, which made it challenging to generate visualizations.

Previous Dataset: Suicide Rate

**New Dataset:** Hotel Reservation

**Data Description:** The hotel reservation dataset is a collection of data related to bookings made in hotels, including information on the date of booking, check-in and check-out dates, room type, arrival date etc. The dataset can be used to analyze booking patterns, occupancy rates, revenue management, and customer segmentation in the hotel industry.

**Source**: Kaggle.com

**Number of fields and rows:** 22 fields and 36275 rows

**Group Member Details-**

1. **Simran Saini (300347691)**
2. **Aman Gupta (300348626)**

**Group Member Contribution towards final term project-**

Part A-Tableau visuals were prepared by Simran Saini and the documentation was done by Aman Gupta.

Part B-Power BI visuals were prepared by Aman Gupta and the documentation was done by Simran Saini.

**Part A-TABLEAU**

**Analysis Questions:**

**Graphical user interface, application

Description automatically generated**

**Description about visuals**

1. **Story:** The story page includes an introduction section which features the title "Customer Booking Insights", as well as a booking analysis, insights on total nights distribution, and a price forecasting model.

**Graphical user interface, text, application

Description automatically generated**

1. **Dashboard:** This dashboard provides a comprehensive analysis of customer booking behavior and preferences, allowing us to gain insights by analyzing arrival month, leading time, and room type. It enables us to understand customer patterns and make data-driven decisions.

A picture containing text

Description automatically generated

1. **Booking Analysis Insights (Sorting and Filtering):** A bar chart has been created to display the average lead time for various types of room reservations based on meal plans and booking status. The chart is sorted in descending order based on the lead time, with the longest lead time displayed first. The visual also includes filters for room type and meal plan selection. This allows for viewing of the chart for a specific meal plan.

**A picture containing bar chart

Description automatically generated**

1. **Average Lead Time by Room Type and Booking Status (Calculation Field):** A bar chart has been created based on a new calculated field called the average lead time. This field is computed by dividing the sum of lead time by the count of booking IDs. The chart has columns showing the average lead time and rows indicating the room type reserved and booking status. The visualization displays the average lead time for various room types based on their booking status, sorted in descending order.

**Chart, bar chart

Description automatically generated**

1. **Total Nights Distribution (Table Calculation):** A bar chart has been created based on table calculations. The columns show the booking ID, and the rows show the total nights calculated field, which is created by adding the number of weekend nights to the number of weeknights. The visualization shows the percentile of total nights for various booking IDs. The table calculation is performed by selecting the percentile option from the quick table calculation menu on the total nights calculated field.

Chart, bar chart

Description automatically generated

1. **Lead Time Trends (Trend Lines):** A line chart has been created based on trend lines. The columns show the arrival month, which is a continuous attribute. The rows display the lead time. The visualization shows the lead time for each month of the year and includes linear trend lines. The minimum lead time was observed in the first month, with a lead time of 35,387, while the highest lead time was observed in the tenth month, with a lead time of 551,523.

**Chart, line chart

Description automatically generated**

1. **Price Forecasting (Forecasting):** A line chart has been created based on forecasting. The columns show the arrival date, and the rows display the average price per room. The visualization displays the average price for different dates and was created by selecting the forecast option from the analytics. The lowest average price rate observed was 104,143, while the highest average price rate was 139,999.

Chart, line chart

Description automatically generated

1. **Booking Distribution by month (Distribution):** A line chart has been created based on the distribution band. The columns show the arrival month, and the rows display the number of previous bookings not cancelled data. The visualization is created by adding the distribution band from the analytics option to the table of the number of bookings not cancelled data. The distribution band helps to segregate the data by adding lines for the different ranges in the number of bookings not cancelled data and the arrival month. The lowest point was observed in the eighth month, and the highest was observed in the twelfth month.

**Chart, line chart

Description automatically generated**

1. **Cancellation per Booking ID:** A treemap visualization has been generated to depict the count of cancellations made for different booking IDs. The treemap shows that INN12098 has the highest number of cancellations, with a count of 13, while INN24541 has the lowest number of cancellations, with a count of only 1.

**Chart, treemap chart

Description automatically generated**

1. **Room Type per Adult:** A pie chart has been generated to represent the distribution of preferred room types based on the number of adults. According to the chart, Room Type 1 was the most popular among adults, with a count of 49,799, while Room Type 7 was the least popular.

**Chart, pie chart

Description automatically generated**

1. **Arrival Analysis:** The line graph and bar graph combination provide an analysis of the relationship between the number of children, arrival month, and special requests. The data reveals that in arrival month 8, there is the highest number of children (759) compared to other months, while arrival month 1 has the lowest number of children (73). On the other hand, arrival month 10 has the highest number of special requests (3428) of all the months.

**Chart, histogram

Description automatically generated**

1. **Meal Type Preferences:** The following analysis investigate Food Choices Bases on Number of children. The maximum preference was for meal plan 1 and the least amount of children preferred no meal plan.

**Chart, bubble chart

Description automatically generated**

**PART-B – POWER BI**

**Details of dataset**

The dataset used previously and the current one is identical.

**Description of transformation**

1. **Creating a new column:** A transformation was performed by creating a new column named "Date\_of\_arrival." This was accomplished by concatenating three existing columns, namely "arrival\_date," "arrival\_month," and "arrival\_year."

**Query used:** Text.From([arrival\_year]) & "-" & Text.PadStart(Text.From([arrival\_month]), 2, "0") & "-" & Text.PadStart(Text.From([arrival\_date]), 2, "0")

1. **Change data type of repeated\_guest and replaced value:** The data type of the "repeated\_guest" column was changed from whole number to text and then values where replaced i.e., 0 with False and 1 with True where False shows guest is not repeated, and True shows guest is repeated.
2. **Change data type of avg\_price\_per\_room:** The data type of “avg\_price\_per\_room” column was changed from whole number to currency.

A screenshot of a computer

Description automatically generated with medium confidence

**Description about visuals**

1. **Introduction:**

**Graphical user interface, application, Teams

Description automatically generated**

1. **Analysis Questions:**

**Text

Description automatically generated**

1. **Sum of arrival\_year by booking\_status:** The analysis involves computing the sum of the arrival year by the booking status, with a total of 49 million bookings not cancelled and 24 million bookings cancelled. This indicates that the number of bookings not cancelled is more than twice the number of bookings that were cancelled.

Chart, pie chart

Description automatically generated

1. **Count of repeated\_guest by arrival\_month:** The analysis is focused on the count of repeated guests by arrival month. It helps to understand the pattern of guests who have visited more than once and their distribution across different months.

sChart, line chart

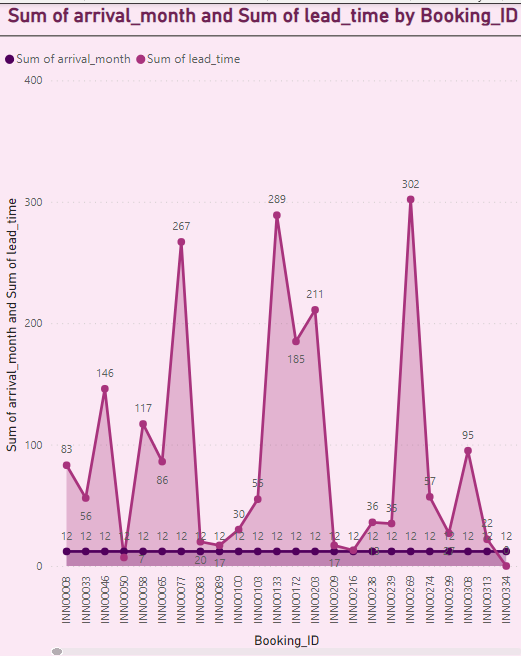
Description automatically generated

1. **Sum of no\_of\_adults by room\_type\_reserved:** The analysis involves computing the sum of the number of adults by the room type reserved, with room type 1 being at the highest with a total of 50 thousand adults.

**Graphical user interface, chart, application

Description automatically generated**

1. **Sum of arrival\_month and Sum of lead\_time by Booking\_ID:** The analysis is a line chart that shows the sum of the arrival month and the sum of the lead time for the different booking Id. The highest lead time is 302.

****

1. **Sum of lead\_time and Sum of no\_of\_adults:** The analysis is based on the gauge chart that displays the values for the sum of lead time and number of adults.

**Chart

Description automatically generated**

1. **Sum of no\_of\_children by room\_type\_reserved:** The analysis is a bar chart that displays the sum of number of children based on the room type reserved, with room type 6 having the most number of children with the total of 1631 and room type 3 having the least number of children.

**Chart, bar chart

Description automatically generated**

1. **Sum of lead\_time and Sum of no\_of adults by Date\_of\_arrival:** The analysis is a funnel chart that displays the sum of lead time and the number of adults by the arrival date.

**Chart, funnel chart

Description automatically generated**

1. **Sum of lead time Sum of no of adults and Sum of no of previous bookings not canceled by market segment type and booking status:** The analysis is a scatter chart that displays the booking status based on the sum of number of adults and sum of lead time.

**Chart, scatter chart, bubble chart

Description automatically generated**

1. **Conclusion:**

**Graphical user interface, text

Description automatically generated**

**Model View**

This shows the relational model of the data, consisting of no relation as it requires two tables to form a relation.

**Graphical user interface, application

Description automatically generated**

**Conclusion**

The following can be concluded-

1.The 10th month has the highest frequency of repeated guests.

2.The sum of arrival years relate to the booking status are 24M for cancelled status and 49M for the not cancelled status.

3. Room type 1 has the highest number of reservations made by adults.

4.INN00269 is the booking Id that has the longest lead time.

5. Room type 6 has the highest number of preferences for the children.