# 



# School of InfoComm Technology

**Data Visualisation**

Diploma in FI

Diploma in IT

April 2021 Semester

**ASSIGNMENT 1**

**(Individual Assignment)**

**Submission Deadline:**

**27th Jun 2021 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **P04** |
| **Student Name** | **:** | Chen Han |
| **Student Number** | **:** | S10202961 |

**Penalty for late submission:**

10% of the marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 4th Jul 2021 (Sunday), 11:59PM.

1. **Project Objectives**

**Primary Exploratory Questions**

Customer Analysis (customer behaviors and profile)

1. What are the preferred visiting months for customers?
2. What is the most famous meal package that customer order?
3. Which country of customers visited the most?
4. What is the most famous booking channel for customers? (Distribution channels)
5. What are the percentages of new customers and repeated customers?
6. What is the preferred room type for customers?
7. What is the preferred deposit type for customers?
8. What is the percentage of new customers and returned customers (repeated)?

Booking Cancellation (finding the reasons for cancellation of bookings)

1. How many bookings are cancelled by the customer?
2. Which distribution channel has the most customers cancelling the booking?
3. What type of customer has the most cancellation of the booking?
4. How is the average day in waiting list related to the cancellation of the booking?

Profitability (areas that the company can pay attention to increase the profit)

1. How is the length of staying? night related to the customer Average Daily Rate (ADR)?
2. How is the meal type that the customer chose related to the customer ADR?
3. How is the customer type related to the customer ADR?

**Why should NP Management Ltd care about the primary exploratory questions?**

The primary exploratory questions provide an insightful thought to break down and analyze the data. By categorizing the data into three parts, being the customer analysis, booking cancellation and profitability, the company is able to go in-depth of each category. For customer analysis, it provides a visualization of customer behaviors such as type of meal ordered, which country of people visited the hotel most, and how many new customers are there etc. From there, we can find out how to maintain new customers and retain returned customers based on their profiles. For booking cancellation, it provides a visualization of different possible factors that lead to the cancellation of booking by customer. Upon identifying possible reasons for cancellation of bookings, we can investigate that part and come up with solutions to avoid cancellation of bookings to happen. Finally, profitability, which is the area that the company can look into to increase the profit or revenue. In this category, we are investigating how is each factor affects the customer ADR (Average Daily Rate) and possible strategies to take to increase the profit.

1. **Data Preparation**

* **State of data**

The initial state of data is not cleaned yet, as some of the attributes such as is\_canceled and is\_repeated\_guest is represented in 0 and 1, which might cause confusion to Tableau to interpret it as measures. Thus, a substantial data cleanup is needed in order to make data visualization easier in Tableau.

* **Data cleansing**

1. Make a new calculated field “IsCanceled” with the following command:   
   **﻿[Is Canceled] = 1**, so as to convert the 0s and 1s in “Is Canceled” to Boolean values.
2. Make a new calculated field “IsRepeatedGuest” with the following command: ﻿ ﻿  
   **IF [Is Repeated Guest] = 1 THEN "Old Customer"**

**ELSE "New Customer"**

**END**

1. Make a new calculated field “Arrival Date” with the following command:  
   **﻿**﻿**MAKEDATE([Arrival Date Year],MONTH(DATEPARSE("MMM", [Arrival Date Month])), [Arrival Date Day Of Month])**
2. Make a new calculated field “Total Stay” with the following command:  
   **﻿[Stays In Week Nights] + [Stays In Weekend Nights]**
3. **Exploratory Data Analysis and Visualization**

* **Exploratory Questions**

**Customer Analysis (customer behaviors and profile)**

1. What are the preferred visiting months for customers?

Table

Description automatically generated

Answer: The preferred visiting month for customer is around the period of March to June, as well as August to October. So the company might need to take a glance at the months having less visitors (Nov to Feb) for better marketing strategies to attract more visitors.

Visualization used: Heat map, as the data involves of date parts from July 2015 to August 2017, using a heat map allows the user to visualize any seasonal patterns or hotspots (like peak and off-peak periods in this case), we can see that during winter periods, there are significant drop of visitors’ flow

1. What is the most famous meal package that customer order?

Chart

Description automatically generated

Answer: The most famous meal package ordered by customer is the Bed & Breakfast meal.

Visualization used: Pie chart, as it gives the user an overall view of proportions in each attribute, we can easily tell that Bed & Breakfast is the most preferred meal out of the rest.

1. Which country of customers visited the most?

Chart, treemap chart

Description automatically generated

Answer: Most of the customers visited the hotel came from Portugal.

Visualization used: Tree Map, as the data involves of countries with high cardinality, a pie chart will be insufficient to have so many proportions divided. Using a tree map allows the user to spot the difference in each attribute easily, as well as the maximum value from the data (it is sorted accordingly), which in this case, PRT has the highest percentage of total of 34.25% out of all visitors.

1. What is the most famous booking channel for customers? (Distribution channels)

Chart, pie chart

Description automatically generated

Answer: The most famous booking channel for customers is Travel Agent and Tour Operators.

Visualization used: Pie Chart

1. What are the percentages of new customers and repeated customers?

Chart

Description automatically generated

Answer: The percentage of new customer is 98.29% whereas the percentage of repeated customer is 1.71%.

Visualization used: Pie Chart

1. What is the preferred room type for customers?

Chart, pie chart

Description automatically generated

Answer: Most of the customers prefer Room A.

Visualization used: Pie Chart

1. What is the preferred deposit type for customers?

A picture containing graphical user interface

Description automatically generated

Answer: Most of the customers has no deposit.

Visualization used: Pie Chart

**Booking Cancellation (finding the reasons for cancellation of bookings)**

1. How many bookings are cancelled by the customer?

Chart, bar chart

Description automatically generated

Answer: There is a total of 44224 bookings cancelled by customers, occupying almost one third of total bookings, which is a relatively large number of booking cancellation.

Visualization used: Bar chart, which is suitable to compare two values and see the

differences.

1. Which distribution channel has the most customers cancelling the booking?

Chart, bubble chart

Description automatically generated

Answer: Most of the customers cancelled their booking are booked through Travel Agent or Tour Operator. The company should pay attention to this distribution channel to reduce the amount of booking cancellations.

Visualization used: Packed bubbles, which compares the size of the data with circles, in this case, TA/TO occupying the most amount of space, thus making it easier to spot the highest number.

1. What type of customer has the most cancellation of the booking?

Chart, bubble chart

Description automatically generated

Answer: Most of the booking cancellations are made by transient customers, might be simply because they are walk-in guest, last minute bookers, who ultimately changed their mind and cancelled the bookings. The company can come out with ways to avoid transient customers to cancel their bookings such as providing better customer services.

Visualization used: Packed bubbles.

1. How is the average day in waiting list related to the cancellation of the booking?

Chart, bar chart

Description automatically generated

Answer: The average day in waiting list is 3.564 for customers cancelled their booking, which is way higher than 1.590 days for customers did not cancel the booking. This shows that most customer did not satisfy with the long waiting time for the hotel to confirm their booking, thus, the hotel management should increase the efficiency of assigning rooms for customer reservation with an ideal response time of 1.59 days.

Visualisation used: Bar chart as it is easy to show the difference between the two values.

**Profitability (areas that the company can pay attention to increase the profit)**

1. How is the length of staying night related to the customer average ADR?

Chart, scatter chart

Description automatically generated

Answer: The correlation for average total stay night and average ADR for both hotels are showing decreasing trend with only a few outliers on top right-hand corners. This shows that the longer that customer stay in the hotel, the lower their average ADR. In order to increase the profit, the company can consider assigning more rooms for short-term visitors than long-term visitors of as they are likely to spend more on services and goods. Especially during peak periods as mentioned in Customer Analysis Part 1, there might be more demands for rooms and long-term visitors might not contribute as much ADR as short-term visitors.

Visualization used: Scatterplot graph as it is efficient to show the relationship/correlation between two measures (Avg. Total Stay and Avg. ADR), so that a trend can be spotted to better analysis.

1. How is the meal type that the customer chose related to the customer ADR?

Chart, waterfall chart

Description automatically generated

Answer: Customers who ordered half board meal has the highest average ADR.

Visualization used: A combination of bar chart and heat map, as it would be clearer to see the highest value by the color and the height of the square. Also, it is much more aesthetic and eye-appealing than bar chart alone.

1. How is the customer type related to the customer ADR?

Chart

Description automatically generated with medium confidence

Answer: Transient customer has the highest average ADR of $107.01 and visitors come in group has the lowest average ADR of $83.49. To increase the profit, the company could accept more transient customers than group visitors especially during peak period as they are contributing more ADR than each member in the group visitors.

Visualization used: A combination of bar chart and heat map, as it would be clearer to see the highest value and lowest value by the color and the height of the square, which in this case, Transient and Group customers.

1. **Dashboard**
2. **Customer Analysis**

Chart, treemap chart

Description automatically generated

1. **Booking Cancellation**

Chart, bar chart

Description automatically generated

1. **Profitability**

Chart, scatter chart

Description automatically generated

----The end of the report----