

**MH6341 - Data Management And Business Intelligence**

**Group Project Report**

| **NAME** | **MATRICULATION NUMBER** |
| --- | --- |
| Khoo Chun Yun | G2302847F |
| Goel Ishita | G2302904C |
| Du Siyi | G2303045K |
| Ang Shu Wei | G2302794C |
| Zhang Danrui | G2302855K |

**Table of Content**

1. **Overview**

1.1 Introduction

1.2 Description of Business

1.3 Objectives

1. **Design of the OLTP Database**

2.1 ER Diagram

2.2 Table Structure

1. **OLTP Data Generation Process**
2. **Business Intelligence Questions**
3. **Design of the OLAP database**

5.1 Design Diagram

5.2 Tables Structure

1. **ETL Process**
2. **Business Intelligence Dashboard Design**
3. **Overview**

**1.1 Introduction**

Hotel reservation system is a type of transaction systems, which allows guests to make easy, secure, and direct reservations without the need to involve intermediaries.

It serves a two-way purpose, for the customers it serves as a platform for making reservations, as well as a source of information about the room fares, availability etc., and on the other hand it serves the purpose for the hotel management to keep track of the customer bookings, manage overbookings, schedules and much more.

Hence, for the management and smooth functioning of a hotel, a good reservation system is important. It involves various features such as room type selection, reservation management, payment, etc.

In this report we will explore the hotel reservation system for one specific hotel in Singapore, that is the Outpost Hotel Sentosa.

**1.2 Description of Business**

The Outpost Hotel Sentosa is an adult-exclusive 5-star hotel situated in prime location, within Sentosa Island Singapore. The hotel has 193 rooms and offers various activities and workshops for its customers. With an ever-increasing customer base, the need to develop and manage its operations is necessary for the hotel to maintain its prestigious reputation.

Hence, in the report, we will explore one aspect of this hotel’s operation,namely the Hotel Reservation System.

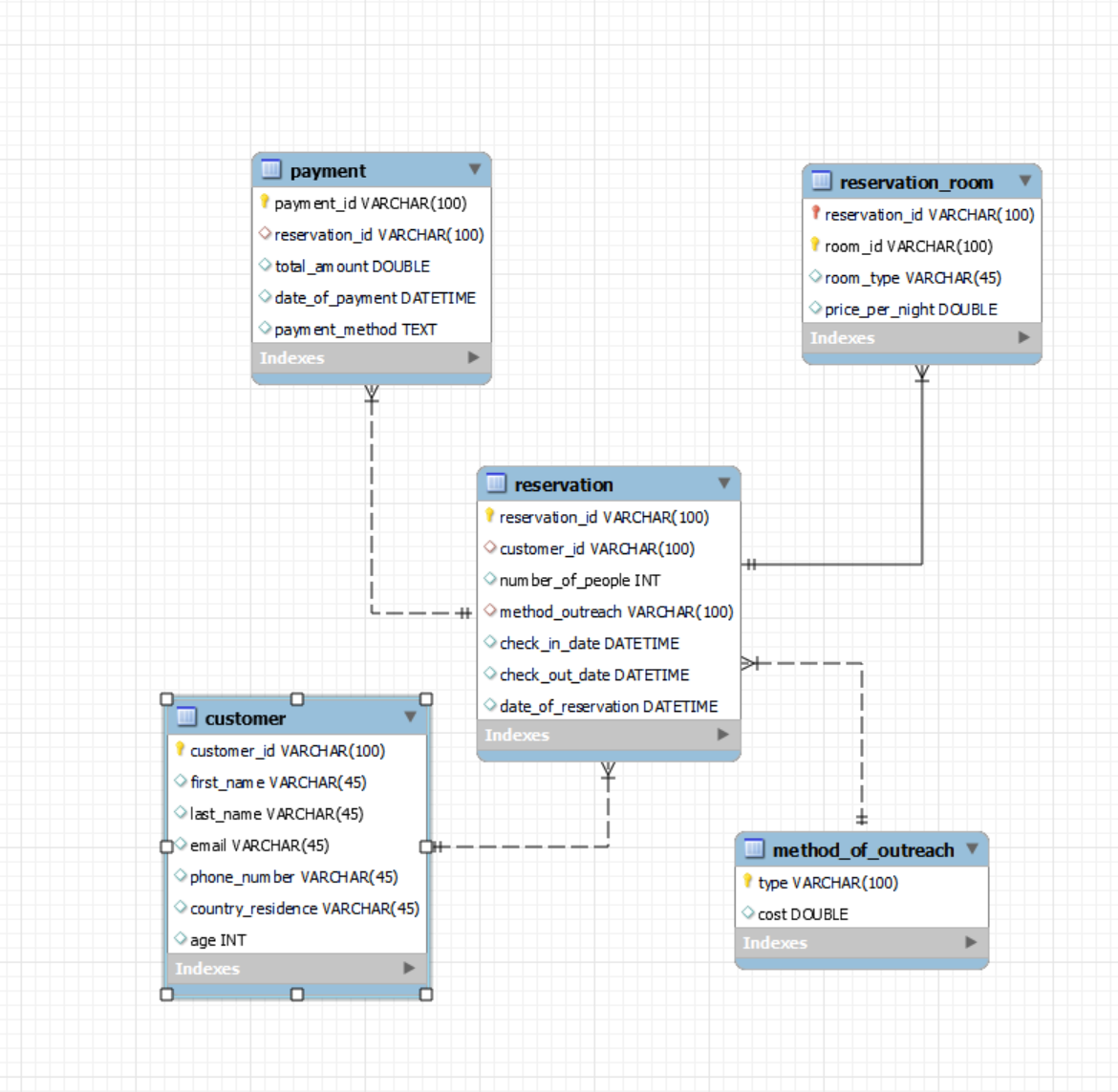
**1.3 Objectives**

The primary objective is to architect a data management and analytical framework for the Outpost Hotel Sentosa to enhance the operational efficiency of the Hotel Reservation System, as well as decision-making. It involves setting up an OLTP (Online Transaction Processing) database for managing daily reservations and an OLAP (Online Analytical Processesing) database for in-depth historical analysis. An ETL process is also designed for seamless data integration between these databases and for creating a BI dashboard. The dashboard will display key metrics and insights, directly supporting the hotel management in strategic planning and operational improvements.

Additionally, incorporating data spanning the COVID-19 pandemic allows for a comprehensive understanding of its impact, guiding the hotel to navigate challenges, adjust to market shifts, and leverage emerging opportunities for recovery and growth. To summarize, this initiative is designed to empower the hotel with data-driven insights to optimize business operations and strategies.

1. **Design of the OLTP Database**

**2.1 ER Diagram**

****

**A customer can make many reservations (One-To-Many Relationship)**

**Each reservation has one payment invoice (One-To-One Relationship)**

**Each reservation can be for more than one rooms (One-To-Many Relationship)**

**A reservation has a method of outreach (One-To-One Relationship)**

**2.2 Table Structure**

The Table Structure for the OLTP database is given as follows:

**Customer**

This table contains the information about customer demographics, such as age, country etc.

| **Column Name** | **Column Type** |
| --- | --- |
| Customer ID (PK) | VARCHAR |
| First Name | VARCHAR |
| Last Name | VARCHAR |
| Email | VARCHAR |
| Phone Number | VARCHAR |
| Country of Residence | VARCHAR |
| Age | INT |

**Reservation**

The reservation table contains information about the reservation details of the customer.

| **Column Name** | **Column Type** |
| --- | --- |
| Reservation ID (PK) | VARCHAR |
| Number of Rooms | INT |
| Number of People | INT |
| Method of Outreach (FK) | VARCHAR |
| CheckIn Date | DATETIME |
| CheckOut Date | DATETIME |
| Date of Reservation | DATETIME |

**Payment**

The table contains information about the payment detail of the reservation.

| **Column Name** | **Column Type** |
| --- | --- |
| PaymentID (PK) | VARCHAR |
| ReservationID (FK) | VARCHAR |
| Total Amount | DOUBLE |
| Date of Payment | DATETIME |
| Method | TEXT |

**Method Of Outreach**

The method and the cost associated with the method of outreach (that is, how the customers got to know about this hotel).

| **Column Name** | **Column Type** |
| --- | --- |
| Type (PK) | VARCHAR |
| Cost | DOUBLE |

**Reservation\_Room**

This is an associative entity of the reservation. That is, it is a weak entity where the room ID is assigned for the reservation.

| **Column Name** | **Column Type** |
| --- | --- |
| Reservation ID (FK) | VARCHAR |
| Room ID (PK) | VARCHAR |
| Room Type | VARCHAR |
| Price Per Night | DOUBLE |

Note that for *Reservation\_Room*, the primary key is **Reservation ID, Room ID.**

1. **OLTP Data Generation Process**

For generating the data for the OLTP database, we made use of the Python programming language. Specifically for the generation of the data we made use of the Faker library to generate information regarding the customer demographics etc.We generated the data in accordance with the four tables as described in *Section 2*. The constraints for the primary and foreign keys were kept in consideration while generating the data.

In addition, to represent variation in hotel booking with respect to peak periods (such as December-February) as well as the impact of COVID in our data, we generated the data monthly. Various factors were kept in mind for generating the data such as the circuit-breaker period, the occupancy of the hotels as well as the customer demographics in terms of the country of residence.

After generating the data, the data was incorporated into the csv files for further analysis. Their are 178003 reservations.

1. **Business Intelligence Questions**

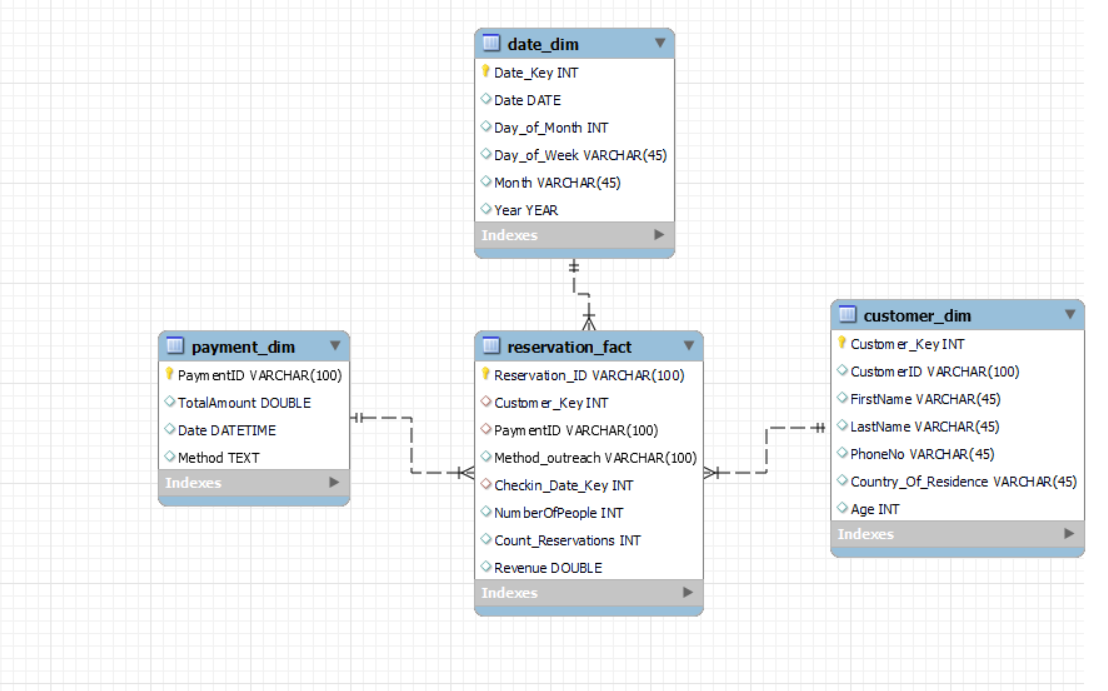
The following are some of the questions that the hotel management may want to discern from the design and implementation of our data warehouse design.

* Who are our current key customer demographic (to ascertain marketing strategies)? [Matches with tableau]
* How have consumption patterns changed over time(pre-covid, during covid, post-covid)? [Use checkout date]
* Which customer group generates the highest revenue?
* Which method of outreach generates the highest revenue for each customer group?
* Which method of payment is most/least used? Study the trend of payment methods over the years.

1. **Design of the OLAP database**

For the OLAP database design, we are implementing it according to the star schema design.

**5.1 Design Diagram**



**5.2 Tables Structure**

The table structure for the star schema is given as follows.

**Reservation\_Fact Table**

| **Column Name** | **Column Type** |
| --- | --- |
| Reservation\_ID (PK) | VARCHAR |
| Customer\_Key (FK) | INT |
| PaymentID (FK) | INT |
| Checkin\_Date\_Key (FK) | INT |
| NumberOfPeople | INT |
| Count\_Reservations | INT |
| Revenue | DOUBLE |
| Method\_of\_Outreach | VARCHAR |

**Customer\_Dim**

| **Column Name** | **Column Type** |
| --- | --- |
| Customer\_Key (PK) | INT |
| CustomerID | VARCHAR |
| FirstName | VARCHAR |
| LastName | VARCHAR |
| PhoneNo | VARCHAR |
| Country Of Residence | VARCHAR |
| Age | INT |

**Payment\_Dim Table**

| **Column Name** | **Column Type** |
| --- | --- |
| PaymentID | VARCHAR |
| TotalAmount | DOUBLE |
| Date | DATETIME |
| Method | TEXT |

**Date\_Dim**

| **Column Name** | **Column Type** |
| --- | --- |
| Date\_Key (PK) | INT |
| Date | DATE |
| Day\_of\_Month | INT |
| Day\_of\_Week | VARCHAR |
| Month | VARCHAR |
| Year | YEAR |

1. **ETL Process**

The data from the OLTP database was transferred to the OLAP data warehouse.The data specific for the analysis and to answer the questions is filtered out. First we define the focus of our data warehouse. In our case we focus on the individual reservations, that is, each row represents a single reservation made by the customer.

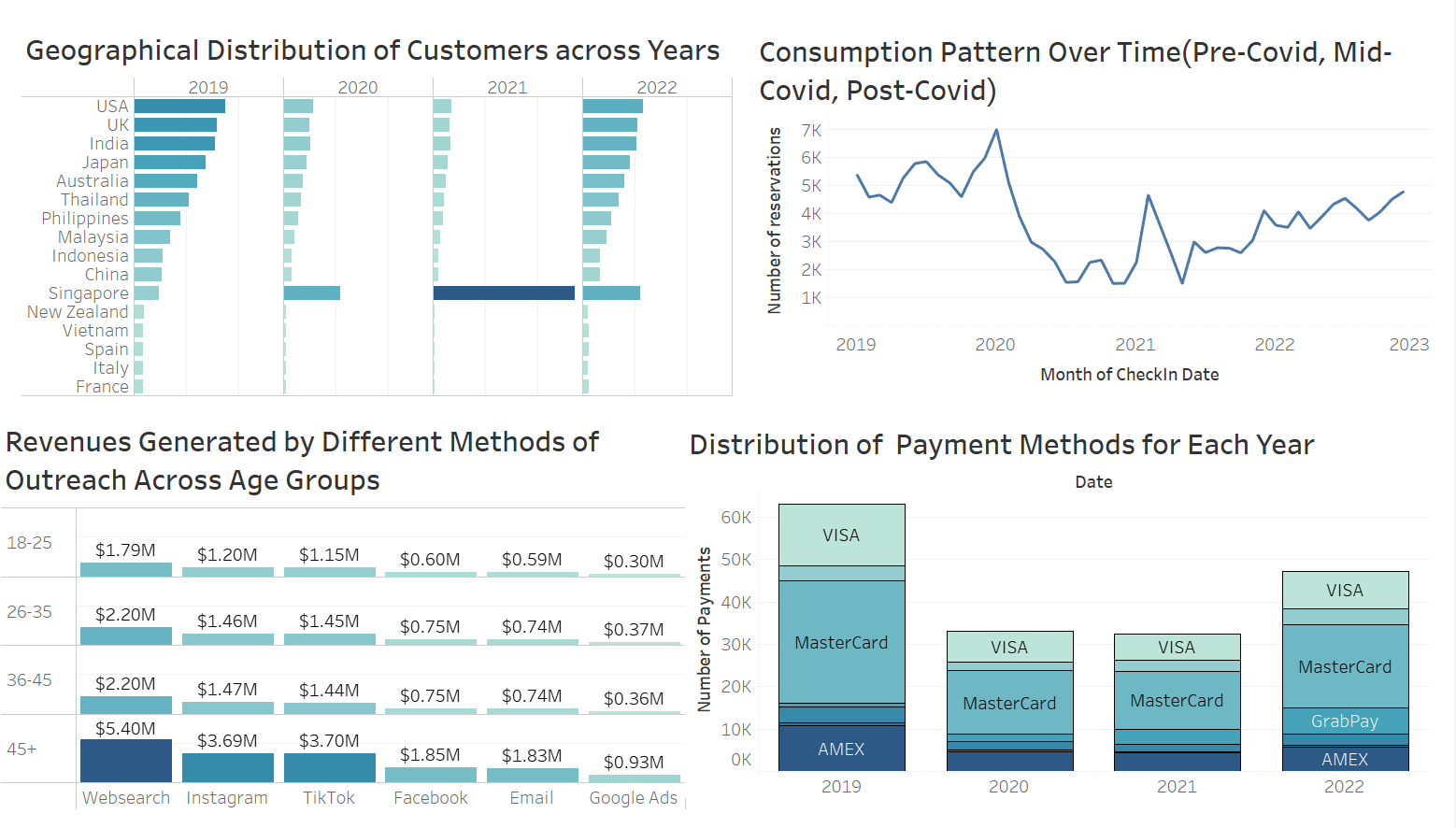
The dimensions are then described to give additional information about the customers, the date dimension as well as the payment dimension. These describe some attributes related to reservations that are analyzed such as customer demographics, outreach methods, etc.

For the dimension tables we add surrogate keys for each. The date dimension table is created using python to get the various dates. (1460 dates, 4 years from 2019 to 2022). The format of the date is adjusted to be the same as the one we have in our dataset.

The data is then loaded into the data warehouse. First the customer, payment and the date dimension are loaded from the respective OLTP tables. Then, the fact table of our star schema is created with the use of join functions.The joins are performed with the respective dimension tables to include the surrogate keys.The data is loaded in a year by year basis (due to the long processing time).

After the data warehouse generation, the analytical questions we listed are carried out. (Two examples are shown in the appendix).

1. **Business Intelligence Dashboard Design**



Appendix

