Construction of a Ticket Selling System Database with Consideration for Transportation and Fan Camp Allocation for the Qatar World Cup 2022

Beatrice Sophia Pamittan
47558865

BUSA8090 - Assignment 1

Session 2 2023

Construction of the Database

I. Database Entities and Attributes

A. Customer

Attribute	**Data Type**
Cust_ID (PK)	varchar(10)
FirstName	char(50)
LastName	char(50)
Email	varchar(50)
Gender	char(50)
Age	int
CountryRes	char(50)

B. BooksFor (Unary Relationship)

Attribute	**Data Type**
BookingCust_ID (FK)	varchar(10)
TicketHolder_ID (FK)	varchar(10)

C. Venue

Attribute	**Data Type**
VenueID (PK)	varchar(50)
VenueName	char(50)
Capacity	int
VenueCost	int
NoOfSecurity	int
NoOfCaterers	int

NoOfEventStaff	int
PayPerHour	int

D. Game

Attribute	**Data Type**
GameID (PK)	varchar(10)
Team1ID	varchar(20)
Team2ID	varchar(20)
VenueID (FK)	varchar(50)
GameDate	date
GameTime	time

E. Train Station

Attribute	**Data Type**
TrainStationID (PK)	varchar(20)
StationName	varchar(50)
VenueID (FK)	varchar(50)
TrainOpsCost	int
StationStaffCost	int

F. Seat Category

Attribute	**Data Type**
SeatCatID (PK)	varchar(50)
SeatCategory	char(5)
Price	int

Quantity	int
VenueID (FK)	varchar(50)

G. Seat

Attribute	**Data Type**
SeatID (PK)	varchar(50)
SeatCatID (FK)	varchar(50)

H. Ticket

Attribute	**Data Type**
Ticket_ID (PK)	varchar(50)
GameID (FK)	varchar(50)
SeatID (FK)	varchar(50)

I. Booking

Attribute	**Data Type**
Booking_ID (PK)	varchar(50)
BookingDate	date
BookingCust_ID (FK)	varchar(50)

J. Ticket Booking

Attribute	**Data Type**
Booking_ID (PK)	varchar(20)

TicketHolder_ID (FK)	varchar(50)
Ticket_ID (FK)	varchar(50)
PaymentType (FK)	char(10)

K. Reservation

Attribute	**Data Type**
Reservation_ID (PK)	varchar(20)
ReservationDate	date
Booking_ID (FK)	varchar(20)
CheckInDate	date
CheckOutDate	date
StayDuration	int

L. Room

Attribute	**Data Type**
Room_No (PK)	varchar(10)
RmCapacity	int
Price	int
HouseKeepingCost	int
UtilitiesCost	int

M. Room Reservation

Attribute	**Data Type**
Reservation_ID (FK)	varchar(20)

Room_No (FK)	varchar(10)
PaymentType	char(10)

II. Relationships

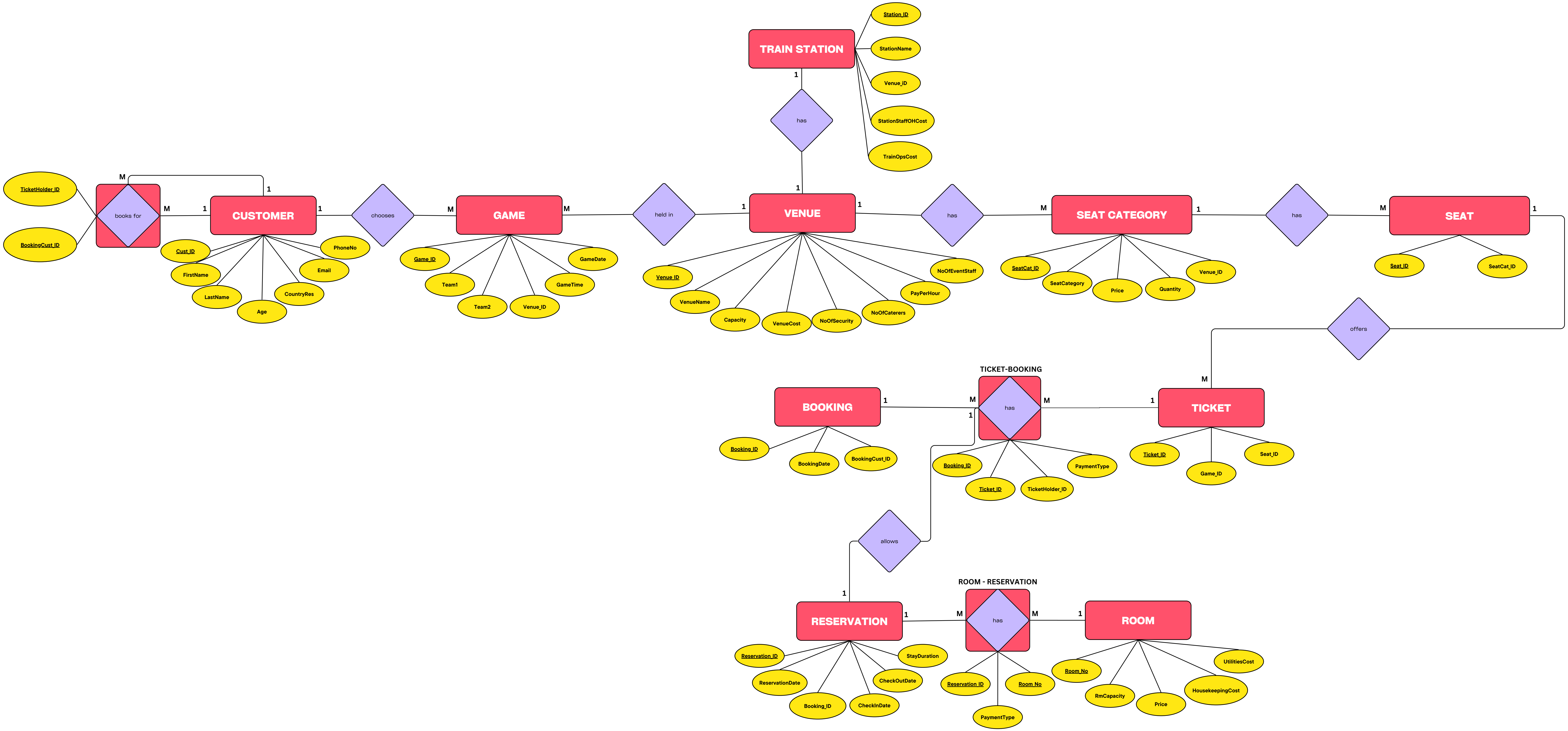
A. One to One
   1. 1 Venue has 1 Train Station

B. One to Many
   1. 1 Customer can choose Many Games
   2. 1 Venue can host Many Games
   3. 1 Venue has Many Seat Categories
   4. 1 Seat Category has Many Seats
   5. 1 Seat offers Many Tickets (across different games)

C. Many to Many
   1. Many customers can book for many customers
      a) Unary Relationship represented by TicketHolder_ID and BookingCust_ID
   2. Many Tickets in Many Bookings
      a) Ticket Booking is the associative entity
   3. Many Reservations has Many Rooms
      a) Room Reservations is the associative entity

III. Entity Relationship Diagram

A PDF copy of the ERD is also available in the Assignment folder submitted.



TRAIN STATION
Station_ID
StationName
Venue_ID
StationStaffOHCost
TrainOpsCost
has
TicketHolder_ID
BookingCust_ID
books for
CUSTOMER
Cust_ID
FirstName
LastName
Age
CountryRes
Email
PhoneNo
chooses
GAME
Game_ID
Team1
Team2
Venue_ID
GameTime
GameDate
held in
VENUE
Venue_ID
VenueName
Capacity
VenueCost
NoOfSecurity
NoOfCaterers
PayPerHour
NoOfEventStaff
has
SEAT CATEGORY
SeatCat_ID
SeatCategory
Price
Quantity
Venue_ID
has
SEAT
Seat_ID
SeatCat_ID
offers
TICKET-BOOKING
BOOKING
Booking_ID
BookingDate
BookingCust_ID
has
Booking_ID
Ticket_ID
TicketHolder_ID
PaymentType
TICKET
Ticket_ID
Game_ID
Seat_ID
allows
ROOM - RESERVATION
RESERVATION
Reservation_ID
ReservationDate
Booking_ID
CheckInDate
CheckOutDate
StayDuration
has
Reservation_ID
Room_No
PaymentType
ROOM
Room_No
RmCapacity
Price
HousekeepingCost
UtilitiesCost
1
M

Data Analytics Report

1. Introduction

1.1. Background

The ticket selling system for the Qatar World Cup 2022 has been constructed using a structured database where audience data is leveraged. Factors and concerns such as transportation, accommodations such as fan camp, and other various costs incurred by the event will be discussed in this report.

1.2. Objectives

This report focuses on extracting insights from the database which will aim to minimize the overhead costs of transportation and the fan camp. Concurrently, this will also aim to maximize the profit for the World Cup 2022 in Qatar and analyze areas for improvement in terms of cost and expenses to make the event possible.

1.2.1. Specific Objectives

- To identify the number of trains to arrange before and after each game
- To identify the cost of resources and overhead cost for the trains for each game
- To identify the utilization of the venues based on the attendance per game
- To identify the variable cost of resources at the venue
- To identify the peak booking dates that customers purchase tickets
- To identify the overhead cost percentage of each game against revenue
- To identify the number of rooms to give priority to ticket holders
- To identify the overhead cost of the fan camp
- To identify the fan camp utilization

1.3. Scope, Limitations, and Assumptions

1.3.1. Scope

This data analytics report will focus on identifying organizational issues and identify aspects of the event where overhead costs can be minimized, especially in transportation and the fan camp.

- Qatar World Cup 2022 is held from November 20 to November 25
- Customers can start booking tickets from June 1, 2022 to November 4, 2022

1.3.2. Limitations and Assumptions

- Only ticket holders are allowed to book a room at the fan camp
- Each train has a capacity of 30 people
- Venue costs per game are assumed to be if the venue was booked at full capacity, regardless of performance of ticket sales
- The data has been scaled down from the actual figures that occurred in 2022
- Customers can book multiple tickets
- There are no limits on the number of tickets a customer can buy

2. <u>Descriptive Analytics</u>

In order to have a better understanding of the demographics of the attendees and the scale of the World Cup games, a few descriptive statistics of the audience has been extracted from the database for better understanding and perspective moving forward.

2.1. Ticket Holder Demographics

Total Number of Ticket Holders

```sql
# Total Number of Ticket Holders
SELECT COUNT(DISTINCT TicketHolder_ID) AS NoOfTicketHolders
FROM TICKETBOOKING;
```

NoOfTicketHolders
880

The total number of people who have successfully purchased at least 1 ticket is 880.

Gender Distribution

```
# Number of Male and Female TicketHolders
SELECT CUSTOMER.Gender, COUNT(DISTINCT TICKETBOOKING.TicketHolder_ID) as Count
FROM CUSTOMER
RIGHT JOIN TICKETBOOKING ON CUSTOMER.Cust_ID = TICKETBOOKING.TicketHolder_ID
GROUP BY CUSTOMER.Gender;
```

Gender	Count
Female	461
Male	419

There are 461 female and 419 male ticket holders for the 2022 Qatar World Cup.

Domestic vs International Attendees

```
#Number of Domestic vs International Attendees
SELECT AttendeeType, COUNT(*) as Count
FROM (
SELECT DISTINCT(TICKETBOOKING.TicketHolder_ID),
    CASE WHEN CountryRes = 'Qatar' THEN 'Domestic' ELSE 'International'
    END AS AttendeeType
FROM CUSTOMER
RIGHT JOIN TICKETBOOKING ON CUSTOMER.Cust_ID = TICKETBOOKING.TicketHolder_ID
) as subquery_0
GROUP BY AttendeeType;
```

AttendeeType	Count
International	569
Domestic	311

There are 311 ticket holders who already reside in Qatar, while 569 are international travelers coming for the World Cup.

3. Potential Organizational Issues and Recommendations

Number of Trains to Arrange Per Game

SQL Query

```
# 1 Number of Trains to Arrange
SELECT GameID, Attendees, ROUND(Attendees/30) as NoOfTrainsBefore,  ROUND(Attendees/30) as NoOfTrainsAfter
FROM(
SELECT TICKET.GameID, COUNT(DISTINCT TICKETBOOKING.Ticket_ID) AS Attendees
FROM TICKETBOOKING
RIGHT JOIN TICKET on TICKETBOOKING.Ticket_ID = TICKET.Ticket_ID
GROUP BY TICKET.GameID)
as subquery_1;
```

The code above joins the TICKETBOOKING and TICKET tables through right join on the Ticket_ID attribute. This results in the tables of all tickets that are booked for all games in the Qatar World Cup. The query then groups the GameID, and counts all the distinct Ticket_IDs per game. In order to determine the number of trains needed before and after each game, it is assumed that the capacity of 1 train is 30 people and that all attendees will travel via train to and from the games. By taking the number of Attendees and dividing it by 30, and rounding up the answer, we obtain the number of trains needed to be arranged before and after each game.

Return Results

GameID	Attendees	NoOfTrainsBefore	NoOfTrainsAfter
G1	231	8	8
G2	138	5	5
G3	192	6	6
G4	187	6	6
G5	188	6	6
G6	343	11	11

Based on the results, Game 2 needed the least number of trains to be arranged for the Games, while Game 6 needed the most number of trains to be arranged for the games. It is also worth noting that Games 3, 4, and 5 all needed 6 trains before and after each game. It can be said that the number of attendees per game is directly proportional to the number of trains needed to be arranged per game. Moreover, careful scheduling of trains would be needed before and after games such that attendees would be able to reach the game on time and that trains would be available for attendees immediately after the game for a smooth flow of travel in and out of the area of the Stadium.

Cost of Resources and Overhead for Trains Each Game

SQL Query

```sql
# 2 Cost of Resources / Overhead for Trains Each Game
SELECT GameID, GameDate, GameTime, VenueName, Attendees, TotalNoOfTrains, TrainOpsCost, StationStaffCost, CostofTrainOps,
(CostofTrainOps+StationStaffCost) as TotalOHTrainCost
FROM (
SELECT GameID, GameDate, GameTime, VenueName, Attendees, (NoOfTrainsBefore+NoOfTrainsAfter) as TotalNoOfTrains, TrainOpsCost,
StationStaffCost, (TrainOpsCost*(NoOfTrainsBefore+NoOfTrainsAfter)) as CostofTrainOps
FROM (
SELECT GameID, GameDate, GameTime, VenueName, Attendees, ROUND(Attendees/30) as NoOfTrainsBefore,  ROUND(Attendees/30) as NoOfTrainsAfter,
TrainOpsCost, StationStaffCost
FROM (
SELECT GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, COUNT(TICKETBOOKING.TicketHolder_ID) as Attendees, TRAINSTATION.TrainOpsCost,
TRAINSTATION.StationStaffCost
FROM TICKET
INNER JOIN TICKETBOOKING ON TICKET.Ticket_ID = TICKETBOOKING.Ticket_ID
RIGHT JOIN GAME  ON TICKET.GameID = GAME.GameID
RIGHT JOIN VENUE on GAME.VenueID = VENUE.VenueID
RIGHT JOIN TRAINSTATION on VENUE.VenueID = TRAINSTATION.VenueID
GROUP BY GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, TRAINSTATION.TrainOpsCost, TRAINSTATION.StationStaffCost
ORDER BY GameDate)
as subquery_2)
as subquery_3)
as subquery_4;
```

The query comprises four select statements. The outermost SELECT statement is used to select specific columns and determine the total operational cost for each game, which includes Train Operational Costs as well as Station staff costs. Columns for GameID, GameDate, GameTime, VenueName, Attendees, TotalNoOfTrains, TrainOpsCost, StationStaffCost, and TotalOHTrainCost (the total of TrainOpsCost and StationStaffCost) which appear in the output. The innermost subquery gathers data on games, venues, ticket reservations, and train stations. It connects multiple tables based on their foreign and primary keys (Ticket_ID, GameID, and VenueID) and groups the results by game information, venue information, number of Attendees, TrainOpsCost, and StationStaffCost. In the following Select statement, the number of trains needed before and after each game are calculated. After which, the Total Number of Trains needed and the cost of train operations are aggregated. In the final Select statement, the Total Overhead Costs (TotalOHCost) is aggregated by adding together CostofTrainOps and StationStaffCost.

Return Results

GameID	GameDate	GameTime	VenueName	Atte...	TotalNoOfTrains	TrainOpsCost	StationStaffC...	CostofTrainOps	TotalOHTrainCost
G2	2023-11-22	19:00:00	Khalifa International Stadium	138	10	75	350	750	1100
G4	2023-11-21	16:00:00	Stadium 974	187	12	75	350	900	1250
G5	2023-11-21	22:00:00	Stadium 974	188	12	75	350	900	1250
G3	2023-11-23	18:00:00	Ahmad bin Ali Stadium	192	12	75	350	900	1250
G1	2023-11-20	19:00:00	Al Bayt Stadium	231	16	75	350	1200	1550
G6	2023-11-25	13:00:00	Lusail Stadium	343	22	75	350	1650	2000

The table obtained provides information about games, including the number of attendees, the total number of trains required, and the operational costs associated with train operations and station staff. It can be seen that the overhead train cost is the highest for Game 6 as it is the game with the most attendees. On the other hand, the lowest overhead would be Game 2 with the least number of attendees. It is also worth noting that it is assumed in this instance that the cost of resources across all games are constant. However, it could potentially be less for Game 2 due to the number of trains arriving and going for the event. In addition to that, Games 4 and 5 fall on the same day, at the same venue, but just at different times of the day. The cost can definitely increase or decrease depending if it would be the same staff working for both games. Moreover, less trains may also be needed depending on consumer behavior and if there are customers who would be watching both games and will only need to ride the trains before Game 4 and after Game 5. Another instance that could potentially be looked at is if trains that would drop off attendees for Game 5 can be the same train that would pick up attendees from the station after Game 4.

Utilization of Venues based on Attendance per Game

SQL Query

```
# 3 Utilization of Venues based on Attendance per Game
SELECT GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, COUNT(TICKETBOOKING.TicketHolder_ID) as Attendees, VENUE.Capacity,
(COUNT(TICKETBOOKING.TicketHolder_ID)/Capacity) as Utilization
FROM TICKET
INNER JOIN TICKETBOOKING ON TICKET.Ticket_ID = TICKETBOOKING.Ticket_ID
RIGHT JOIN GAME  ON TICKET.GameID = GAME.GameID
RIGHT JOIN VENUE on GAME.VenueID = VENUE.VenueID
RIGHT JOIN TRAINSTATION on VENUE.VenueID = TRAINSTATION.VenueID
GROUP BY GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, VENUE.Capacity
ORDER BY GameDate;
```

The SELECT statement selects the GameID, GameDate, GameTime, VenueName, the number of Attendees, capacity of the venue, and the utilization of the venue. Utilization is calculated by dividing the number of attendees by the capacity of the venue. These are all obtained through multiple inner joins on tables such as TICKET, TICKETBOOKING, GAME, VENUE, and TRAINSTATION. Group by then groups the results based on specific columns to aggregate the data. In this case, the grouping is done by GameID, GameDate, GameTime, VenueName, and Venue Capacity. It is then arranged based on the GameDate in ascending order.

Return Results

GameID	GameDate	GameTime	VenueName	Attendees	Capacity	Utilization
G1	2023-11-20	19:00:00	Al Bayt Stadium	231	350	0.6600
G4	2023-11-21	16:00:00	Stadium 974	187	200	0.9350
G5	2023-11-21	22:00:00	Stadium 974	188	200	0.9400
G2	2023-11-22	19:00:00	Khalifa International Stadium	138	200	0.6900
G3	2023-11-23	18:00:00	Ahmad bin Ali Stadium	192	200	0.9600
G6	2023-11-25	13:00:00	Lusail Stadium	343	400	0.8575

Based on the previous 2 queries, it can be said that the game with the most number of attendees does not necessarily mean that it would be the most utilized venue. As seen in the results above, Game 6 with the most number of Attendees has only about 86% utilization of the venue. The Top 3 Games with the highest venue utilizations are Game 3, Game 5, and Game 4 at 96%, 94%, and 93.5% respectively. These games almost reached full capacity of the venue, which is a good sign that these games were properly allocated and assigned to the correct venues based on possible forecasted demand on the tickets sold per Game. In order to increase the utilization of venues of other games, in order to minimize the overhead costs incurred during each game for under utilized venues would be to shift focus on selling more tickets, perhaps at a discounted price to make ticket sales to increase revenue and profit. Another strategy that could be adopted is to create bundle prices for 1 game that is high-performing in terms of ticket sales with a game that is underperforming in ticket sales to boost revenue and lower overhead costs overall.

Variable Cost of Resources (Staff) based on Utilization of Venue

SQL Query

```sql
# 4 Variable Cost of Resources (Staff) based on Utilization of Venue & when the venue is at full capacity
SELECT GameID, TotalStaff, PayPerHour, TotalVarCost, Utilization, (TotalVarCost * Utilization) as ActualNeededVarCost,
(TotalVarCost-ActualNeededVarCost) as Difference
FROM(
SELECT GameID, TotalStaff, PayPerHour, TotalVarCost, Utilization, (TotalVarCost * Utilization) as ActualNeededVarCost
FROM (
SELECT GameID, TotalStaff, PayPerHour, (TotalStaff*PayPerHour) as TotalVarCost, Utilization
FROM (
SELECT GAME.GameID, VENUE.VenueName, COUNT(TICKETBOOKING.TicketHolder_ID) as Attendees, VENUE.Capacity,
(COUNT(TICKETBOOKING.TicketHolder_ID)/Capacity) as Utilization, VENUE.VenueCost,
(VENUE.NoOfSecurity + VENUE.NoOfCaterers +VENUE.NoOfEventStaff) as TotalStaff, VENUE.PayPerHour
FROM TICKET
INNER JOIN TICKETBOOKING ON TICKET.Ticket_ID = TICKETBOOKING.Ticket_ID
RIGHT JOIN GAME  ON TICKET.GameID = GAME.GameID
RIGHT JOIN VENUE on GAME.VenueID = VENUE.VenueID
RIGHT JOIN TRAINSTATION on VENUE.VenueID = TRAINSTATION.VenueID
GROUP BY GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, VENUE.Capacity, VENUE.VenueCost, VENUE.NoOfSecurity,
VENUE.NoOfCaterers, VENUE.NoOfEventStaff, VENUE.PayPerHour
ORDER BY GameID
) as subquery_5
) as subquery_6
) as subquery_7;
```

For each game, the SQL code provided computes several cost-related metrics. It computes the necessary staff cost, variable cost, and venue utilization for a given game. The final result is formed by deriving ActualNeededVarCost and the Difference between TotalVarCost and ActualNeededVarCost from the outermost SELECT expression. These values are calculated via a sequence of subqueries, the first of which computes ActualNeededVarCost, the second of which computes TotalVarCost, and the last of which gathers information on games, venues, ticket bookings, and train stations. The code calculates the required staff, variable expenses, and utilization for each game.. For ease of analysis, the results are organized by GameID.

Return Results

GameID	TotalStaff	PayPerHour	TotalVarCost	Utilization	ActualNeededVarCost	Difference
G1	31	25	775	0.6600	511.5000	263.5000
G2	25	25	625	0.6900	431.2500	193.7500
G3	25	25	625	0.9600	600.0000	25.0000
G4	25	25	625	0.9350	584.3750	40.6250
G5	25	25	625	0.9400	587.5000	37.5000
G6	49	25	1225	0.8575	1050.4375	174.5625

Based on the results, it has provided significant insights into cost management and venue utilization. The number of staff working at a game should be proportionate to the number of Attendees at the game in order to avoid any unnecessary overhead costs. By allocating the optimal number of staff in a venue based on the number of attendees, it would greatly decrease the variable and overhead costs needed. It would be irresponsible to assume and staff each game as if the venue is at full capacity as this would in turn increase costs and prevent maximizing profits for the games and the World Cup overall.

Customer Booking Behavior and Peak Booking Periods

SQL Query

```
# 5 Identify Peak Booking Dates / Months #ok
SELECT MONTH(BookingDate) AS Month, COUNT(MONTH(BookingDate)) AS NoOfBookings
FROM BOOKING
GROUP BY MONTH(BookingDate)
ORDER BY MONTH(BookingDate);
```

This SQL code aggregates data from the BOOKING table to generate a monthly count of bookings. The MONTH function is used to extract the month from the BookingDate and rename it "Month." The COUNT function counts the amount of reservations for each month. The GROUP BY clause puts the results in ascending order of the month, which then displays a summary of monthly bookings.

Return Results

Month	NoOfBookings
6	41
7	41
8	37
9	42
10	43
11	4

The results above show that the number of bookings per month do not vary too much from each other, which can be interpreted as the demand for tickets are almost equal for the months leading up to the World Cup. However, there were slightly more ticket bookings 2 months prior to the World Cup. In order to minimize overhead costs for the World Cup as a whole, it could be suggested to have "Early Bird" Tickets to be sold at a slightly discounted price for a limited time in June to boost revenue on the first month of ticket selling. By extracting this information, organizers can better understand customer behavior, preferences and booking patterns to optimize marketing strategies to increase ticket sales.

Overhead Cost Percentage of Each Game against Revenue

SQL Query

```
# 6 Identify overhead cost percentage of game against revenue
SELECT GameID, TotalStaff, PayPerHour, TotalVarCost, Revenue, (TotalVarCost/Revenue) as OHPercentage
FROM (
SELECT GameID, TotalStaff, PayPerHour, (TotalStaff*PayPerHour) as TotalVarCost, Revenue
FROM (
SELECT GAME.GameID, (VENUE.NoOfSecurity + VENUE.NoOfCaterers +VENUE.NoOfEventStaff) as TotalStaff, VENUE.PayPerHour,
SUM(SEATCATEGORY.Price) as Revenue
FROM TICKET
INNER JOIN TICKETBOOKING ON TICKET.Ticket_ID = TICKETBOOKING.Ticket_ID
RIGHT JOIN SEAT ON TICKET.SeatID = SEAT.SeatID
RIGHT JOIN SEATCATEGORY ON SEAT.SeatCatID = SEATCATEGORY.SeatCatID
RIGHT JOIN GAME  ON TICKET.GameID = GAME.GameID
RIGHT JOIN VENUE on GAME.VenueID = VENUE.VenueID
RIGHT JOIN TRAINSTATION on VENUE.VenueID = TRAINSTATION.VenueID
GROUP BY GAME.GameID, GAME.GameDate, GAME.GameTime, VENUE.VenueName, VENUE.Capacity, VENUE.VenueCost, VENUE.NoOfSecurity,
VENUE.NoOfCaterers, VENUE.NoOfEventStaff, VENUE.PayPerHour
ORDER BY GameID
) as subquery_5
) as subquery_6;
```

This SQL query computes the overhead cost percentage for each game by taking into account staff costs and revenue. The innermost subquery collects information regarding games, venues, ticket sales, and seat categories. It computes the overall staff cost by adding the number of security, caterers, and event employees multiplied by their hourly wage. It also computes the revenue for each game by adding the prices of various seat classifications. The middle subquery computes the total variable cost (TotalVarCost) by multiplying the total number of employees by the hourly wage. The outermost query retrieves the GameID, TotalStaff, PayPerHour, TotalVarCost, and Revenue, then calculates the overhead percentage (OHPercentage) by dividing TotalVarCost by Revenue.

Return Results

GameID	TotalStaff	PayPerHour	TotalVarCost	Revenue	OHPercentage
G1	31	25	775	30075	0.0258
G2	25	25	625	16350	0.0382
G3	25	25	625	22800	0.0274
G4	25	25	625	22375	0.0279
G5	25	25	625	22550	0.0277
G6	49	25	1225	40525	0.0302

Overall, the results aid in the analysis of overhead expenditures in relation to revenue made by each game. It can be seen that OH costs account for less than 4% of the revenue for each game. This is good but it should also be worth noting that fixed costs are not yet taken into account so the figures above would not give the actual profit of each game just yet. The Total Staff and Variable Cost per Game are under the assumption that 100% of the venue was utilized. As seen in previous queries, none of the Games were sold out and hence, it would be wiser to optimize the allocation of resources per game based on the attendees. Revenue per Game is significantly higher than overhead costs at the venue which strongly suggest that each game is greatly operating at a profit.

Number of Rooms to give Priority to Ticket Holders

SQL Query

```
# 7 Number of Rooms to give Priority to Ticket Holders
SELECT RmCapacity, COUNT(Room_No) as NoOfRooms, (RmCapacity*COUNT(Room_No)) as MaxGuests
FROM ROOM
GROUP BY RmCapacity;
```

This SQL code calculates the maximum number of guests the fan camp can accommodate for each room capacity (RmCapacity). It then counts the number of rooms (NoOfRooms) for each room capacity and calculate the maximum number of guests that each RmCapacity can accommodate (MaxGuests) by multiplying the room capacity by the number of rooms. The GROUP BY clause organizes the results based on RmCapacity.

Return Results

RmCapacity	NoOfRooms	MaxGuests
2	50	100
4	50	200

The results above provide insight into the fan camp's maximum number of guests for different room types. The fan camp has a total of 100 room. It can be seen that the total number of guests that the fan camp can accommodate is 300, which is much less than the number of distinct ticket holders, 880. In the succeeding queries, it will be identified whether or not the number of rooms available are enough or not.

Overhead Cost of Fan Camp

SQL Query

```
# 8 Identify overhead cost of fan camp
SELECT RmCapacity, NoOfReservations, TotalHouseKeepingCost, TotalUtilitiesCost, TotalRevenue,
((TotalHouseKeepingCost+TotalUtilitiesCost)/(TotalRevenue)) as OverheadCostPercentage
FROM(
SELECT COUNT(ROOMRESERVATION.Reservation_ID) as NoOfReservations, ROOM.RmCapacity, SUM(ROOM.HouseKeepingCost) as TotalHouseKeepingCost,
SUM(ROOM.UtilitiesCost) as TotalUtilitiesCost, SUM(ROOM.Price) as TotalRevenue
FROM ROOMRESERVATION
INNER JOIN ROOM on ROOMRESERVATION.Room_No = ROOM.Room_No
GROUP BY ROOM.RmCapacity)
as subquery_7;
```

This SQL code calculates the overhead cost percentage for each room capacity in a lodging scenario. The inner subquery aggregates data related to room reservations and rooms, calculating the number of reservations (NoOfReservations), total housekeeping cost (TotalHouseKeepingCost), total utilities cost (TotalUtilitiesCost), and total revenue (TotalRevenue) based on room capacity (RmCapacity). The outer query selects the room capacity (RmCapacity) and computes the OverheadCostPercentage by dividing the sum of housekeeping and utilities costs by the total revenue.

Return Results

RmCapacity	NoOfReservations	TotalHouseKeepingCost	TotalUtilitiesCost	TotalRevenue	OverheadCostPercentage
2	44	1012	660	5280	0.3167
4	36	1260	1080	6900	0.3391

The results provide an overview to understand and consider carefully the proportion of operational costs to the total revenue for each room capacity. This further aids in cost analysis and decision-making as it would be easier to identify areas for improvement. It can be seen that the overhead cost for rooms for 4 people has a slightly higher percentage (~34%) as compared to a room for 2 people (31.6%). The overhead cost of the fan camp is noticeably higher than the overhead cost for the trains and the venue. This can be minimized by increasing the rates of each room type and also implementing dynamic pricing where the rates of each room would increase as the World Cup nears. This would also encourage Attendees to book their fan camp rooms earlier rather than later and this could lead to higher utilization rate of the fan camp.

Fan Camp Utilization
SQL Query

```
# 9 Identify fan camp utilization
SELECT TotalReservations, RmCapacity, TotalRoomsAvail, (TotalReservations/TotalRoomsAvail) as FanCampUtilization
FROM(
SELECT COUNT(Reservation_ID) as TotalReservations, ROOM.RmCapacity, COUNT(ROOM.Room_No) as TotalRoomsAvail
FROM ROOMRESERVATION
RIGHT JOIN ROOM ON ROOMRESERVATION.Room_No = ROOM.Room_No
GROUP BY ROOM.RmCapacity)
as subquery_8;
```

This SQL query calculates fan camp occupancy based on room reservations and availability. The inner subquery aggregates data from room reservations and rooms and calculates the total reservations and total available rooms for each kind of room capacity (RmCapacity). The outermost query selects TotalRegistrations, RmCapacity, and TotalRoomsAvail and calculates FanCampUtilization by dividing TotalResponses by TotalRoomsAvail.

Return Results

TotalReservations	RmCapacity	TotalRoomsAvail	FanCampUtilization
44	2	50	0.8800
36	4	50	0.7200

Based on the results, the fan camp was not fully utilized with bookings at 88% and 72% for the rooms that can accommodate 2 and 4 people respectively. This provides insight into space usage in the fan camp. Similar to ticket sales, in order to minimize overhead costs, we can increase revenue as well by encouraging Attendees especially those from overseas to book at the fan camp instead of other accommodations. This can be through genuine reviews, social media presence, and attractive pricing and offers.

Circling back to the number of rooms that should be prioritized for ticket holders, currently, the number of rooms available is more than enough given that the fan camp utilization did not exceed its capacity. The 100 rooms are enough for all ticket holders who booked at the fancamp. However, if it would be under the assumption that all International attendees would book a room at the fan camp, which is 569 people, at least 90 more rooms would be needed. The 90 rooms would be 45 rooms that can accommodate 2 people, and another 45 rooms that can accommodate 4 people.

4. Conclusion

With this data analysis report, we want to achieve several important goals aimed at optimizing resources and simplifying operations related to the Qatar World Cup while considering transportation and fan camp accommodation. One of our main goals was to determine the optimal number of trains needed before and after each game. By analyzing participant data and using connections with train traffic, we gained insight into the most efficient allocation of transport resources, ensuring a smooth journey for attendees of each game.

In addition, we were introduced to event-related cost accounting, focusing on both resource and overhead costs related to train traffic. This included a detailed estimate of the cost of coordinating train services and station staff. By calculating overheads and estimating their share of each game's revenue, we gained a comprehensive understanding of the financial impact and efficiency of these aspects.

Another important aspect of our analysis was to understand seat occupancy, variable venue costs and the number of preferred seats for ticket holders. By assessing attendance and booking patterns, we were able to pinpoint peak booking months and respective customer behaviors, enabling more targeted marketing strategies. Additionally, an analysis of fan usage and overheads would shed light on the effectiveness and cost-effectiveness of fan accommodation arrangements during these events. Overall, our analysis provides valuable information for strategic decision making, ensuring optimal resource allocation, minimizing overhead costs and a better customer experience.