



DEPI – Data Analytics Graduation Project



Project Planning

Project Overview

ProHost is a data analytics project that simulates a hotel management reporting system similar to the Opera Property Management System (PMS) developed by Oracle Hospitality.

The project aims to analyze hotel data from different operational areas, including human resources, customer relations, finance, reservations, and services, to uncover meaningful insights that can support data-driven decisions and improve hotel performance.

Project Objectives

- To collect, clean, and integrate hotel data from multiple departments.
- To analyze key operational indicators such as occupancy rate, revenue, customer satisfaction, and staff performance.
- To develop interactive dashboards and visual reports using Power BI and Excel.
- To provide clear, actionable insights for improving hotel operations, profitability, and guest experience.

Scope of Work

The project covers five main areas of hotel analytics:

- Customer Analysis: Demographics, satisfaction, and feedback trends.
- Profitability & Cost Analysis: Revenue, expenses, and profit margins.
- Reservation Analysis: Booking patterns, occupancy rates, and cancellation trends.
- Hotel Services Analysis: Usage of services (e.g., restaurant, spa) and guest preferences.

Tools & Technologies

- SQL Server: For database storage and querying.
- Excel / Power Query: For data cleaning and transformation.
- Power BI: For dashboard design and visualization.
- Python: For advanced data analysis and visualization.

Expected Deliverables

- Clean and well-structured hotel database.
- Analytical reports and dashboards for each focus area.
- A summary report highlighting insights and recommendations.
- Presentation of findings during the DEPI graduation showcase.

Timeline (Tentative)

<i>Phase</i>	<i>Task</i>	<i>Duration</i>
1	Project Planning	Week 1
2	Stakeholder Analysis	Week 2
3	Database Design	Week 3
4	Data Analysis & Dashboard Development	Weeks 4 - 5
5	UI/UX Design	Week 6
6	Documentation & Presentation	Week 7

Project Goal

To create a data-driven hotel analytics report that delivers actionable insights and supports strategic decision-making across all departments.

Stakeholder Analysis

Purpose

The goal of the stakeholder analysis is to identify the main individuals and departments who will benefit from or contribute to the **ProHost** analytics system. Understanding their needs helps ensure that the project delivers relevant insights and supports effective decision-making within the hotel.

Stakeholder Table

<i>Stakeholder</i>	<i>Role / Responsibility</i>	<i>Expected Benefits</i>
Hotel Management	Oversees all hotel operations and strategic planning.	Access to summarized dashboards for overall performance, financial results, and customer satisfaction.
Finance Manager	Handles budgeting, cost control, and profitability analysis.	Detailed financial reports showing revenue, expenses, and profitability by department.
Marketing Team	Promotes hotel services and manages guest engagement campaigns.	Data on customer demographics, preferences, and booking trends to plan targeted campaigns.
Public Relations Manager	Manages guest feedback, complaints, and brand image.	Access to guest satisfaction analysis and sentiment trends to improve customer experience.

Stakeholder Communication Plan

<i>Stakeholder</i>	<i>Communication Method</i>	<i>Frequency</i>	<i>Purpose</i>
Hotel Management	Dashboard presentations and summary reports	Monthly	Review of hotel KPIs and strategic performance.
Finance Manager	Financial dashboards and reports	Monthly	Track income, expenses, and profit margins.
Marketing Team	Marketing performance reports	Monthly	Evaluate customer segments.
Public Relations Manager	Feedback and sentiment analysis report	Monthly	Improve guest satisfaction and reputation.

Summary

The ProHost system ensures that every department receives data-driven insights tailored to its responsibilities.

By combining information across Finance, Marketing, and Customer Relations, hotel leaders can make more informed, coordinated decisions to enhance performance and guest satisfaction.

Database Design

Introduction

The Database Design phase focuses on building a structured, scalable, and realistic database that simulates a hotel management environment like the Opera Property Management System (PMS) used in real hotels. This database serves as the foundation for all subsequent analysis related to guests, reservations, hotel operations, financial performance, and services.

During this phase, we aimed to design a relational database that reflects real hotel workflows while ensuring data integrity, consistency, and optimal performance for analytical queries.

Data Source Explanation

At the beginning of the project, the team spent significant time searching for high-quality, real-world hotel datasets. We explored various platforms such as Kaggle, GitHub, open data portals, and academic repositories. However, most available datasets were either incomplete, outdated, or did not contain the operational details required for a full PMS simulation.

Because of this challenge, we decided to generate **synthetic data** tailored to the hotel domain.

What is Synthetic Data?

Synthetic data is artificially generated information created using scripts, algorithms, or statistical models to replicate real-world data patterns without using actual sensitive or private records.

It is commonly used when:

- Real data is unavailable or confidential
- Projects require large datasets for testing
- Privacy and compliance regulations prevent sharing real data
- A customized data structure is needed for simulations

Synthetic data allowed us to:

- Build a realistic hotel management system
- Control data quality, relationships, and size
- Avoid privacy concerns
- Create a complete dataset aligned with our project goals

Note on HR Schema

Due to time limitations in the graduation project timeline, we were not able to include the **HR analytical processes** in this phase.

However, the **HR schema design remains fully implemented**, and the data model has been documented so it can be extended and analyzed in future iterations of the project.

Database Structure Overview

The database is named **ProHost** and contains multiple schemas representing different hotel operations:

- **Guests** → Guest information, preferences, feedback, reservations
- **Operations** → Rooms, room types, pricing, services, billing
- **HR** → Employees, positions, departments, shifts (not analyzed yet)

This structure helps separate business domains, improve organization, and maintain clarity during analysis.

Core Master Tables

Guests.Guests

It contains personal information about each guest including contact information, nationality, gender, and registration date.

Guests.GuestPreferences

Stores any special preferences linked to each guest (e.g., quiet room, no smoking, high floor).

Guests.GuestsFeedback

Captures feedback and ratings provided by guests on various hotel services.

Guests.Guests

It contains personal information about each guest including contact information, nationality, gender, and registration date.

Guests.GuestPreferences

Stores any special preferences linked to each guest (e.g., quiet room, no smoking, high floor).

Guests.GuestsFeedback

Captures feedback and ratings provided by guests on various hotel services.

Guests.Reservations

Captures every booking made by guests, including room details, stay duration, number of guests, meal plans, payment status, and booking source.

Operations.RoomTypes

Defines available room categories such as Single, Double, Suite, etc.

Operations.Rooms

Stores individual room details including room number, floor, hotel, room type, and availability status.

Operations.RoomPricing

Contains time-based pricing rules for each room type, enabling seasonal or dynamic pricing analysis.

Operations.Billings

Represents the master bill for each reservation, including issue dates, totals, discounts, and payment details.

Operations.BillingItems

Stores detailed charges associated with each bill, such as restaurant charges, spa services, or other extra fees.

Operations.Services

Contains a list of services provided by the hotel (e.g., spa, laundry, room service, gym).

Operations.ServiceUsage

Tracks how and when guests use various services, linking usage data to reservations and billing.

Operations.Hotels

Stores general information about the hotel such as name, address, rating, number of rooms, and chain affiliation.

HR.Departments

Defines the different departments within the hotel (e.g., Front Office, Housekeeping). Includes minimum staffing requirements and links to the hotel.

HR.Positions

Contains job positions within each department, salary ranges, job levels, and descriptions.

HR.Employees

Stores detailed employee information such as name, position, department, employment status, salary, demographics, education, and manager relationships.

HR.Shifts

Tracks employee scheduling includes shift dates, hours worked, overtime, department, and status.

These tables fully support HR analytics and will be used in the future version of the project.

ERD — Entity Relationship Diagram

The full ERD visual representation of all tables and relationships is available here:

<https://dbdiagram.io/d/ProHostDiagram-68f178d02e68d21b41e1b23d>

SQL Script

The full SQL script is available publicly:

https://github.com/mariam-tahaa/DEPI_Project/blob/main/Database/ProHost-DatabaseCreation.sql

Data

Analysis

Data Analysis & Dashboard Development

The Data Analysis & Dashboard Development phase represents the core of the **ProHost** project, where the cleaned and structured hotel data is transformed into meaningful insights. After completing the data preparation and ensuring the accuracy of all datasets, this phase focuses on exploring the data, identifying trends, and answering the key business questions defined during project planning.

The analysis covers multiple perspectives including reservations, finance, guests, and hotel operations to understand performance, behavior, and opportunities for improvement. Different analytical techniques were used based on the nature of each dataset, including Exploratory Data Analysis (EDA) in Python, data transformation in Power Query, SQL queries, and advanced modeling using DAX in Power BI.

Reservations Dataset

Data Cleaning & Transformation Steps

This section documents all transformation steps applied to the Reservations dataset using Power Query. The goal was to prepare clean, accurate, and analysis-ready data for the Reservation Dashboard.

1. Data Source Connection

The dataset was loaded directly from SQL Server using a custom SQL query that joins:

- Guests.Reservations (main table)
- Operations.RoomPricing (to retrieve the nightly room price)

Purpose:

To extract only the data needed for analysis and merge price information in a single query.

2. Filter Years (2020–2025)

The SQL query filters Check-In and Check-Out dates to only include reservations between 2020 and 2025.

Purpose:

To limit the analysis to recent years and remove old irrelevant records.

3. Change Data Types

Converted fields to appropriate types:

- Dates → date
- Number of Adults / Children → whole number
- PricePerNight → currency

Purpose:

Ensures accurate calculations and avoids type conversion errors.

4. Remove Unneeded Columns

Removed the column:

- PricingID

Purpose:

The field is only needed for SQL joining and not required for analysis.

5. Trim & Clean Text Fields

Eliminated duplicate records based on all columns.

6. Remove Duplicates

Ensure each reservation appears only once.

7. Remove Blank Rows

Filtered out rows where all fields were null or empty.

8. Calculate Leading Time

$$\text{LeadingTime} = \text{CheckInDate} - \text{BookingDate}$$

Purpose:

Shows the number of days customers book in advance.

Used for demand forecasting and customer behavior analysis.

9. Calculate Nights

$$\text{Nights} = \text{CheckOutDate} - \text{CheckInDate}$$

Purpose:

To calculate length of stay (LOS), a key hotel performance metric.

10. Standardize Booking Source

Replaced shorthand codes:

Original	Updated
OTA	Online Travel Agency

Purpose:

Improves clarity and consistency in dashboard visuals.

11. Remove Checked-In Status

Filtered rows where:

Status ≠ “CheckedIn”

Purpose:

Some Checked-In records may represent operational movement rather than new reservations. We focus on actual booked reservations.

12. Expand Meal Plan Description

Replaced meal codes with full names:

Code	Meaning
BB	Bed & Breakfast
FB	Full Board
HB	Half Board
BO	Bed Only

Purpose:

Better readability in dashboards.

13. Guest Type Classification

Created a new field:

Type = “Family” if NumberOfChildren ≠ 0

Else → “Individuals”

Purpose:

To analyze family vs individual travelers.

Reservations – Analytical Questions

The reservation analysis focuses on understanding booking behavior, customer patterns, pricing trends, and seasonality.

The following analytical questions were used to guide the exploration and development of the Power BI dashboards:

1. Reservation Volume & Growth

- What is the total number of reservations for each year?
- What is the Year-Over-Year (YoY) growth in reservations?

2. Booking Behavior – Leading Time

- What is the average Leading Time (days between booking and check-in)?
- How is Leading Time changing Year-Over-Year?

3. Guest Stay Duration

- What is the Average Length of Stay (LOS)?
- How does the LOS change Year-Over-Year?

4. Pricing Trends

- What is the Average Room Price per night?
- How does the Average Room Price Change Year-Over-Year?

5. Reservation Status Distribution

- What is the contribution of each reservation status (Checked Out, Cancelled, No-Show)?

6. Guest Segmentation

- How many reservations were made by Families vs. Individuals?

7. Room & Meal Preferences

- Which room types are most preferred by guests?
- Which meal plans are most selected?

8. Booking Channel Analysis

- What is the distribution of booking channels?
(Direct, Online Travel Agency, Walk-in, etc.)

9. Monthly Reservation Trends

- What is the total number of reservations per month compared to the same month last year?
- What is the monthly growth rate?

10. Monthly Pricing Trends

- How does the Average Room Price change month over month compared to last year?
- What is the monthly growth rate in pricing?

11. Monthly Length of Stay Trends

- How does the Average Length of Stay change month over-month compared to last year?
- What is the monthly growth rate?

12. Monthly Leading Time Trends

- How does the Average Leading Time change month-over-month compared to last year?
- What is the monthly growth rate?

Guests Data

GuestPreferences Table

1. Load Data

- Imported the Guests.GuestPreferences table from SQL Server.

2. Remove Extra Quotes

- Cleaned the Preferences column by removing unnecessary double quotes (").

3. Split Preferences into Separate Columns

- Split the Preferences column by comma (,)

4. Set Correct Data Types

- Converted all preference columns to text.

5. Unpivot Preference Columns

- Converted the split preference columns into row-level values.
- This produced a cleaner structure: one row = one guest preference.

6. Remove Helper Column

- Removed the "Attribute" column created during unpivoting.

7. Rename Columns

- Renamed Value → Preference.

8. Trim & Clean Text

- Removed leading/trailing spaces.
- Removed invisible/unwanted characters using Text.Clean.

9. Remove Duplicate Records

- Ensured unique combinations of GuestID + Preference.

10. Remove Blank Records

- Deleted rows where Preference or GuestID was blank or null.

Final Output: A clean, normalized table showing each guest and their individual preferences.

Guests Table

1. Load Data

- Imported the Guests.Guests table from SQL Server.

2. Merge First & Last Name

- Combined FirstName and LastName into a single field: FullName.

3. Convert Date Columns

- Change:
 - DateOfBirth → Date
 - RegistrationDate → Date

4. Trim & Clean Text Fields

5. Remove Duplicates

- Removed repeated guest records.

6. Remove Blank Rows

- Filtered out rows with missing essential fields.

7. Calculate Guest Age

- Created a new calculated column Age:
 - Age = (Today – DateOfBirth) ÷ 365
 - Converted age to whole number.

8. Create Age Groups

Added a new AgeGroup column:

- < 20 → “-20”
- 20–29 → “20s”
- 30–39 → “30s”
- 40–49 → “40s”
- 50–59 → “50s”
- 60+ → “60s”

Final Output: A cleaned Guests table with full names, age, and age groups for demographic insights.

GuestsFeedback Table

1. Load Data With Join

- Loaded GuestsFeedback table and joined it with Operations.Services (to get Service Type and Service Name).

2. Remove Unnecessary Columns

- Removed ServiceID since it was replaced by ServiceName & ServiceType.

3. Convert FeedbackDate

- Ensured FeedbackDate is in date format.

4. Standardize Service Type Values

- Replaced “F&B” → “Food & Beverages” for consistency.

5. Trim & Clean Text

6. Remove Duplicates

- Ensured each feedback entry was unique.

7. Remove Blank Records

- Removed any rows where important fields were blank or null.

Final Output: A clean table containing guest feedback with service details and standardized text.

TotalRespondents Table

1. Load Data With Join

- Loaded the same joined dataset as GuestsFeedback.

2. Convert Data Types

- Converted:
 - FeedbackDate → Date
 - GuestID → Text

3. Keep Unique Responses

- Deduplicated by FeedbackDate + GuestID
(to count each guest once per day).

4. Group by Date

- Counted the number of respondents per day.
- Created a new column: TotalRespondents.

Final Output: A table showing how many guests provided feedback each day.

DAX Columns for NPS Analysis (Guests Table)

```
NPSScore =  
IF(  
    ISBLANK(AVERAGEX(RELATEDTABLE(GuestsFeedback),  
    GuestsFeedback[Rating])),  
    1,  
    AVERAGEX(RELATEDTABLE(GuestsFeedback), GuestsFeedback[Rating])  
)
```

Purpose:

- Calculates the average feedback rating per guest.
 - If a guest has no feedback → set score to 1 "Lowest value".
-

```
NPS =  
SWITCH(  
    TRUE(),  
    Guests[NPSScore] >= 4, "Promoters",  
    Guests[NPSScore] >= 3, "Passives",  
    Guests[NPSScore] < 3, "Detractors"  
)
```

Purpose:

- Categorizes each guest according to NPS standards:
 - Promoters → Rating ≥ 4
 - Passives → Rating = 3
 - Detractors → Rating < 3

These columns power the NPS metrics in your dashboard.

Customer Analysis – Analytical Questions

This part of the analysis focuses on understanding guest demographics, satisfaction levels, preferences, and overall behavior. These questions guide the development of insights and visualizations within the dashboard.

1. What is the total number of respondents and the Year-over-Year (YoY) growth rate?

This question helps measure the overall engagement in customer feedback surveys and identifies whether guest participation is increasing or decreasing over time.

2. What is the average rating?

The average rating provides a general evaluation of guest satisfaction with their stay and services offered by the hotel.

3. What is the gender contribution?

This identifies the distribution of respondents by gender, helping the hotel understand which groups are more actively sharing feedback.

4. What is the NPS contribution and the YoY growth rate (Promoters – Passives – Detractors)?

Net Promoter Score (NPS) evaluates overall customer loyalty.

- **Promoters** → Highly satisfied guests
- **Passives** → Neutral
- **Detractors** → Unsatisfied guests

Tracking YoY growth helps monitor changes in customer loyalty.

5. What is the new customer acquisition rate over months compared to last year?

This measures how many guests are first-time visitors and whether new customer acquisition is improving.

6. What is the average rating over months compared to last year?

This tracks satisfaction trends over time and helps determine if customer experience is improving monthly.

7. What is the average rating contribution across hotel services and service types?

This identifies which services receive high or low ratings, helping management improve weaker areas (e.g., housekeeping, dining, facilities, etc.).

8. What is the age distribution of respondents?

Understanding the age profile of guests helps identify key demographic segments (young adults, families, seniors, etc.).

9. What are the most frequent preferences requested by guests?

This question reveals what guests commonly request such as room views, late check-outs, bed types, or special arrangements allowing the hotel to enhance personalization and service quality.

Financial Data

The Finance module consists of three key tables imported from SQL Server: Billings, Expenses, and BillingItems.

Each table undergoes a series of cleaning and transformation steps to ensure consistency, accuracy, and readiness for analysis and dashboard development.

Billings Table

1. Source Import

- The table is loaded from SQL Server using a direct query from Operations.Billings.
- Records are filtered to include only billing entries between 2020 and 2025.

2. Change Column Types

- Date fields converted to date type:
IssueDate, DueDate, PaymentDate
- Financial fields converted to Currency type:
AmountPaid, DiscountAmount
- Key fields converted to text type:
BillingID, ReservationID

3. Trim Text Fields

- Leading/trailing spaces removed from:
BillingID, ReservationID, Status, PaymentMethod

4. Clean Text Fields

- Non-printable characters removed from the same key text fields to prevent data inconsistencies.

5. Remove Duplicates

- Ensures unique billing records based on all columns.

6. Remove Blank Rows

- Deletes any row where all fields are empty or null.

Final Output: A fully cleaned Billings table containing accurate, formatted, and deduplicated financial billing data.

Expenses Table

1. Source Import

- Data imported from Operations.Expenses joined with Operations.ExpenseCategories.
- Records filtered for the period 2020 to 2025.

2. Remove Unnecessary Columns

- Removed fields not needed for analysis:
HotelID, ExpenseCategoryID2, ExpenseCategoryID.

3. Change Column Types

- *ExpenseDate* converted to date type.
- *Amount* converted to Currency type.
- *ExpenseID* converted to text type.

4. Trim Text Fields

- Trimmed leading/trailing spaces from:
CategoryName, Status, PaymentMethod, VendorName, Description, ExpenseID

5. Clean Text Fields

- Removed invisible characters to prevent inconsistent categories or values.

6. Remove Duplicates

- Ensures each expense entry appears once.

7. Remove Blank Rows

- Deletes rows containing entirely empty or null values.

Final Output: A clean Expenses dataset with reliable categories, dates, vendors, and financial amounts.

BillingItems Table

1. Source Import

- Data loaded from Operations.BillingItems.
- Filtered to include service records from 2020 to 2025.

2. Change Column Types

- Converted to proper data types:
UnitPrice → Currency
ServiceDate → Date
BillingID, *BillingItemID* → Text

3. Remove Duplicates

- Ensures each billing item appears only once.

4. Remove Blank Rows

- Removes rows with all fields empty or null.

5. Replace Values

- Standardized item category naming:
"F&B" → "Food & Beverages"

Final Output: A cleaned BillingItems table ready for service-based financial analysis.

Finance Analysis – Analytical Questions

The Finance module focuses on understanding the hotel's financial performance by analyzing revenue, expenses, profitability, discounts, payment behaviors, and service-level contributions. The following analytical questions guide the financial insights generated in the project:

1. What are the total Revenue, Expenses, Discounts, Profit, and Profit Margin, and what is their Year-over-Year (YoY) growth rate?

This question evaluates the hotel's overall financial performance and compares current results to the previous year to identify financial improvement or decline.

2. What are the monthly Revenue, Expenses, Discounts, Profit, and Profit Margin, and what is their Month-over-Month (MoM) growth rate compared to the same month last year?

This helps measure seasonality, monthly performance trends, and whether financial results are improving or declining compared to the previous year.

3. What is the Revenue distribution by service category?

Breaks down revenue by service types such as:

- Room services
- Food & Beverages
- Spa & Wellness
- Transport
- Other hotel services

This identifies the most profitable services and areas needing improvement.

4. What is the Expense distribution by expense category?

Shows where the hotel spends most of its budget (e.g., maintenance, utilities, housekeeping, marketing, F&B supplies).

It helps measure cost efficiency and control opportunities.

5. What is the discount rate for each payment method, and what is the total revenue associated with each?

Analyzes how payment preferences (cash, credit card, online, transfer) affect:

- Discount levels
- Total revenue collected
- Customer purchasing behavior

This can help optimize payment policies.

6. What is the payment status distribution (Paid – Overdue – Pending), and what is the total outstanding amount?

Evaluates the efficiency of billing and payment collection by identifying:

- How many invoices are fully paid
- How many are still pending
- How many are overdue
- Total unpaid amounts affecting cash flow

This is important for financial planning and credit risk management.

Dashboard Design

UI/UX Design

In this phase, the goal was to create a clean, intuitive, and visually appealing dashboard interface that enhances the user experience while ensuring smooth navigation and clear data storytelling. To achieve this, the dashboard layout and background were fully designed outside Power BI using **PowerPoint**, allowing complete creative control over structure, spacing, and visual hierarchy.

The steps followed during this phase were:

1. Designing the Dashboard Background in PowerPoint

PowerPoint was used to design the dashboard layout because it provides flexible tools for:

- Aligning shapes and grid elements
- Creating clean sections for KPIs, charts, and filters
- Applying brand colors, icons, and consistent styling
- Building smooth visual flow between dashboard elements

Custom placeholders were created for:

- Header and date controls
- KPI cards
- Visual sections (charts, tables, trends)
- Navigation elements (if needed)

2. Exporting the Layout as an SVG File

After finalizing the design in PowerPoint, the background was exported as an **SVG** file. SVG format was chosen because it ensures:

- High quality with no pixelation
- Sharp visuals at any screen size
- Better performance inside Power BI

3. Importing the SVG Design into Power BI

Inside Power BI, the exported SVG file was added as a **page background**, giving the dashboard a polished and professional look.

This approach separates **UI/UX design** from **data visualization**, ensuring:

- Consistent branding
- Clean and modern visual structure
- Better focus on insights rather than raw visuals
- A premium design similar to professional BI studios

Final Deliverables

Conclusion & Final Deliverables

This documentation highlights the full process followed throughout the project—from data extraction, cleaning, and transformation to analysis, modeling, and dashboard design. The detailed steps, calculations, and analytical frameworks reflect the collaborative effort of our team in transforming raw operational data into meaningful insights that support decision-making.

All final results, insights, KPIs, and visual dashboards can be explored in the project **Power BI report** and the **presentation slides**. These deliverables include the complete analytical findings for Reservations, Guests, and Finance, along with interactive dashboards designed for a seamless user experience.

Power BI report:

<https://app.powerbi.com/view?r=eyJrIjoiMTc5MDk1OGItNTFIZS00NmRhLWFmNGQtNWFjMWM1ZTUXMWY4IiwidCI6IjBiMGU4YWY3LTlwY2QtNDIzM05YzgzLWM0NWM2OGU4OTkyYyJ9&pageName=321c72300d01340ee3aa>

Presentation slides:

https://github.com/mariam-tahaa/DEPI_Project/tree/main/ProjectDocumentation

Team Members

This project was completed through the combined effort of our team members, each contributing through their unique skills during the EDA, data modeling, analysis, and design phases:

- **Mariam Taha**
- **Shahd Mohammed**
- **Yasmin Omar**
- **Shrouk Abdelwence**
- **Hazem Mustafa**
- **Kareem Shaaban**