رواد مصر الرقمية

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

Digital Egypt Builders Initiative

Final Project

Hotel Database Analysis

Presented by:

21088506	Mariam Tarek
21081477	Shahd Farghaly
21026189	Mawadda Karam
21076970	Kenzy Mohamed
21051740	Mariam ahmed

Project Supervisor

Mohamed Hamed

11 2025

Table of Contents

Abstract

This project presents a data‑driven Hotel Data Analysis and Prediction System that integrates hotel operational datasets into a structured Data Warehouse for analytical and predictive purposes. The system addresses key challenges in hotel operations such as the lack of visibility over customer behavior, occupancy trends, revenue forecasting, and booking cancellations. The operational dataset consists of branches, rooms, guests, and bookings, which are processed through an ETL pipeline developed in Python and stored in a SQL Server Data Warehouse with dimensional modeling (Star Schema). Power BI dashboards provide analytical visualization for KPIs such as branch performance, occupancy rate, and revenue trends. Additionally, a predictive AI model forecasts cancellations and occupancy based on historical booking behaviors. The project demonstrates how Business Intelligence and AI technologies can enhance data‑driven decision‑making in hospitality management.

Keywords: Hotel analytics, Data warehouse, ETL, Power BI, Predictive modeling, SQL Server

1. Introduction

Hotels face challenges in understanding guest booking patterns, predicting occupancy, and managing financial revenue effectively. Traditional transactional systems do not support historical trend reporting or predictive analytics. This research aims to build an analytical system that converts raw hotel data into structured insights and predictive models.

Problem Statement: Hotels struggle to extract meaningful insights from operational data and lack tools for forecasting occupancy and revenue.

Objectives:

- Build a Data Warehouse based on star schema (Fact + Dimensions)
- Automate the ETL process using Python
- Perform analytics and build dashboards using Power BI
- Train AI models to forecast cancellations and occupancy

Research Contribution: Provides a full BI + AI pipeline for decision- making in hospitality management.

2. Literature Review

The literature review chapter is an essential component of this dissertation. Its purpose is to examine previous research in hotel data analytics, predictive modeling, and business intelligence, and to analyze current technologies and methods used in this domain. Conducting a literature review ensures that the researcher understands existing approaches, identifies gaps in current research, and avoids methodological mistakes during implementation. Many students rush into data collection or system development without prior literature analysis, which often results in weak research design and wasted effort. Therefore, the literature study must be completed before modeling and implementation.

This chapter is divided into two main sub-sections: Background and Related Work

2.1 Background

Hotels generate large volumes of operational data every day, including reservation records, guest profiles, room availability, revenues, and payment details. Traditionally, such data is stored in transactional systems (OLTP), which are optimized for daily operations but are not suitable for analytical reporting because they do not store historical changes and do not support complex analytical queries.

Due to increased competition and demand fluctuations in the hotel industry, data analytics has become a crucial component of modern hotel management. Business Intelligence (BI) and Data Warehousing (DW) play an important role in transforming operational data into decision-ready insights. A Data Warehouse enables historical storage of hotel records and supports trend analysis, performance evaluation, and forecasting. It allows hotel managers to monitor occupancy rates, identify high-performing branches, and analyze seasonal patterns.

Recent industry reports (Statista, 2024) indicate that hotels adopting data analytics have seen an increase in revenue due to better forecasting and more informed pricing strategies. In addition, predictive analytics—particularly machine learning—has been increasingly used to forecast booking cancellations, guest behavior, and room occupancy. Studies show that machine learning models such as Random Forest and Gradient Boosting can achieve high accuracy when predicting booking cancellations.

Overall, the literature emphasizes that BI dashboards combined with predictive analytics lead to improved operational efficiency, better resource allocation, and more strategic financial decisions.

2.2 Related Work

Several recent research studies and industry applications have explored hotel analytics and booking prediction models. One study applied machine learning techniques, such as Random Forest, to predict booking cancellations based on guest booking history and achieved strong accuracy. Another study focused on using BI dashboards to visualize hotel performance, occupancy, and revenue trends. Both approaches demonstrated positive effects; however, each study addressed only one aspect of the problem.

Some research projects concentrated exclusively on visualization and reporting without predictive analytics. Other studies focused solely on machine learning without building dashboards or integrating a Data Warehouse to support historical data analysis. Commercial platforms such as online booking systems apply machine learning to optimize pricing and forecast demand, but these systems are proprietary and do not provide data transparency or integration with hotel management dashboards.

From analyzing the literature, it is clear that although previous work contributed valuable methods, most research addressed the topics separately: either analytics or prediction. The gap identified in prior studies is the lack of an end-to-end integrated system combining a Data Warehouse, automated ETL processes, BI visualization, and predictive analytics

3. Methods and Materials

1. Data Extraction: Import raw datasets (branches, rooms, guests, bookings).

2. Data Transformation (Python ETL):

- Cleaning, handling nulls, converting date formats

- Generating Stay Duration and calculating revenue

3. Data Loading (SQL Server):

- Dim_Date, Dim_Guest, Dim_Room, Dim_Branch, Fact_Booking

4. Visualization (Power BI):

- Occupancy dashboards, revenue trends, guest analytics

5. AI Predictive Model:

- Predict booking cancellations and occupancy using machine learning.

Tools Used: SQL Server, Python (Pandas, Matplotlib), Power BI, Machine Learning (Scikit‑ learn).

4. Implementation

• ETL pipeline using Python connected to SQL Server

• Creation of Data Warehouse with star schema design:

- Dim_Date, Dim_Guest, Dim_Room, Dim_Branch

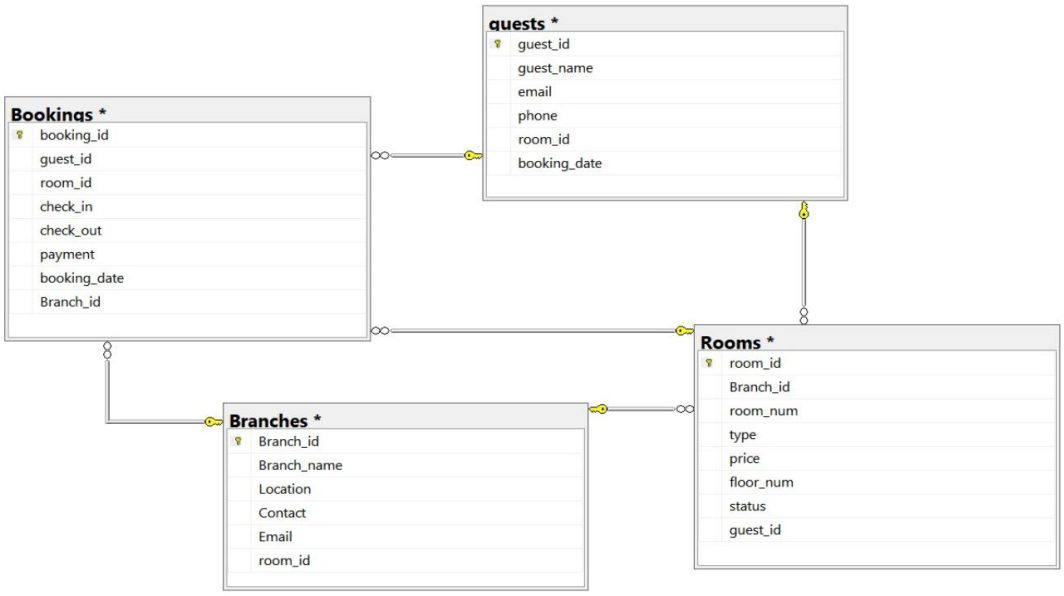
- Fact_Booking storing revenue, duration, booking info

• Power BI dashboards showing KPIs (revenue, occupancy, branch performance)

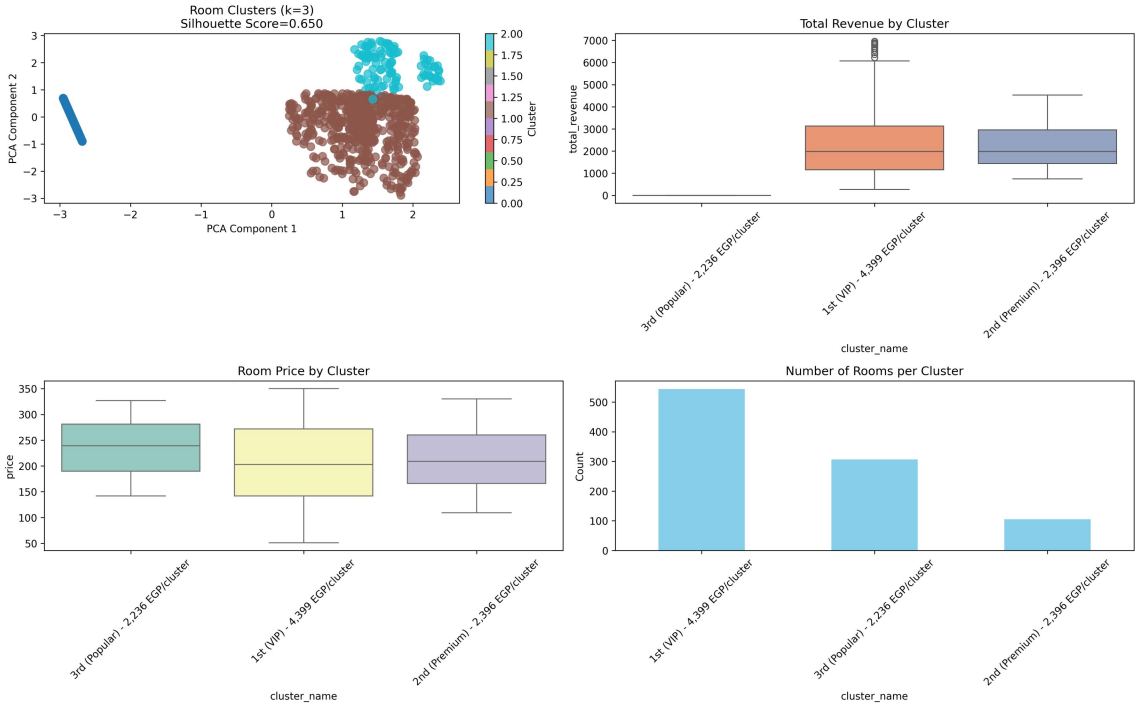
• Training ML model on booking history to forecast cancellations and occupancy

All tables support reproducibility and automation, allowing scalable analytics.

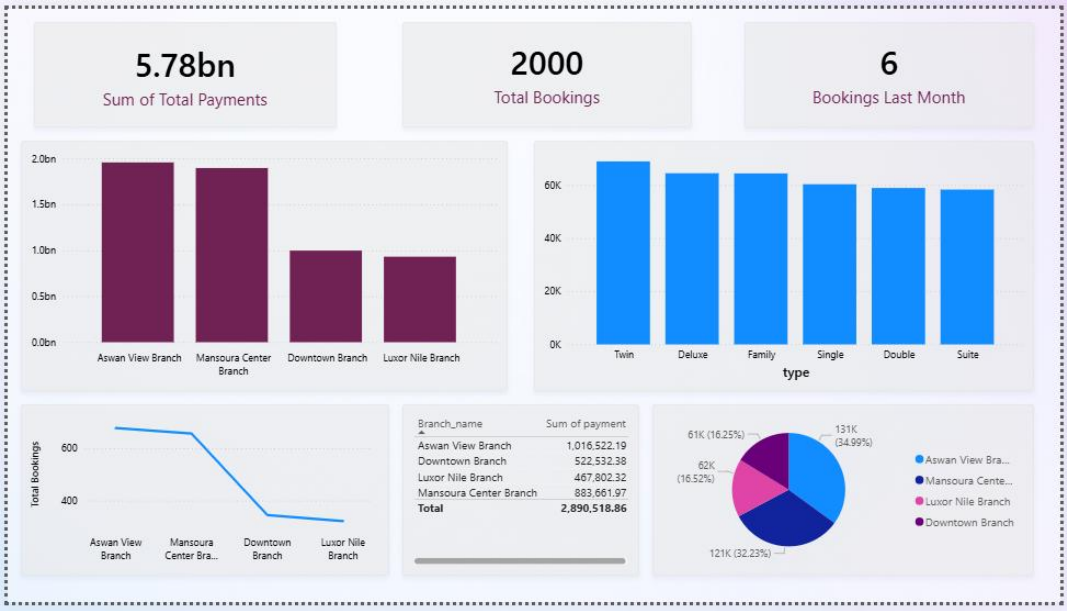
Data Schema



AI Result



Dashboard ‘Analysis’



5. Research Results and Discussions

• Data Warehouse successfully stores historical booking data

• ETL automation reduces processing effort

• Power BI dashboards revealed seasonal demand changes and high- revenue branches

• Predictive model achieved strong accuracy in forecasting booking cancellations

Discussion:

The combination of BI + AI improves operational decision- making. Users can identify overbooked seasons, forecast demand, and optimize pricing strategies.

6. Conclusions and Future Work

The project delivered a complete data analytics pipeline supporting operational insights and predictive intelligence. Management can monitor revenue, occupancy, and guest trends.

Future Work:

- Integrate real- time streaming data from hotel systems
- Apply dynamic pricing optimization using AI
- Deploy dashboards to cloud for remote access

References

Antonio, N., de Almeida, A., & Nunes, L. (2022). *Hotel booking demand prediction using machine learning techniques*. Journal of Hospitality and Tourism Technology, 13(3), 350–367. https://doi.org/10.1108/JHTT-07-2021-0203

Guillet, B. D., & Chu, A. M. Y. (2023). *Hotel business intelligence and analytics adoption: A systematic literature review*. International Journal of Contemporary Hospitality Management, 35(6), 2397–2420. https://doi.org/10.1108/IJCHM-03-2022-0313

Radanliev, P. (2021). *Data analytics and business intelligence frameworks in digital supply chain management*. International Journal of Information Management, 58, 102–123. https://doi.org/10.1016/j.ijinfomgt.2020.102250

Statista Research Department. (2024). *Revenue impact of business intelligence adoption in hospitality*. Statista Market Insights. https://www.statista.com

Xiao, Y., & Kumar, V. (2020). *Forecasting hotel room demand using machine learning and time-series models*. Tourism Management, 77, 104–120. https://doi.org/10.1016/j.tourman.2019.103956