**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Strategic Thinking/ Higher Diploma Data Analytics for Business |
| **Assessment Title:** | Individual / Practical |
| **Lecturer Name:** | James Garza |
| **Student Full Name:** | Miroslava Slavikova |
| **Student Number:** | sba22382 |
| **Assessment Due Date:** | 17 Dec 2023 |
| **Date of Submission:** | Dec 2023 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**Hospitality Management and the importance of Data Analysis**

**Subject area: Hospitality Management with focus on Revenue Management**

Table of Contents

**Introduction3**

**Business Analysis and Project Plan3**

**Machine Learning Models1**

**Libraries 2**

**Dataset and Data Understanding3**

**Data Preparation1**

**Data Visualization2**

**Hotel Revenue Management3**

**Challenges1**

**Milestones1**

**Analysis of Results1**

**Conclusion1**

**Bibliography & References1**

Introduction

The hotel industry generates tens of thousands of data from numerous data points every single day. Online reservations made by guests constitute data including personal information about each and every traveller. Hotel collects data from guest surveys, digital invoicing, special request, Wi-Fi sign ups, and even on-property purchase behaviours.

Hotel Revenue Management requires a solid understanding of business data, be it managing room rates, negotiating agreements about commissions through a variety of distribution channels, or scouting the competition’s room rates in order to remain competitive in the market. In today’s data driven world, it is increasingly important to track, collect and analyse hotel’s data to transform it into actionable results in order to target the audience effectively, to streamline operations and to retain competitive advantage.

Business Analysis and Project Plan

Historical data help us to predict future guests’ behaviour, predict demand and identify new trends. They also demonstrate cycles, seasonality, patterns and any anomalies that can impact the sales and future demand. Hotel data are typically gathered from a diverse source, including industry reports, market research, sales records, online analytics, and guests’ surveys.

We’re going to look at the historical data of the two different properties in Portugal, one being hotel is located in the city of Lisbon and the second property is a holiday resort in the region of Algarve and identify key areas of revenue enhancement.

Machine Learning Models

I have used three different supervised machine learning models:

* Decision Tree with accuracy score of 0.810
* KNN with accuracy score of 0.800
* Logistic Regression with accuracy score of 0.775

The results provided reasonable values, albeit not outstanding. Therefore, I have tried to perform PCA that not only compresses data, but aims to improve the accuracy score.

Running the KNN model thereafter improved the accuracy score to 0.842.

Libraries

I have downloaded all the necessary libraries, panda for data manipulation and analysis library, matplotlib.pyplot and seaborn for data visualisation, numpy for numerical computing, sklearn libraries for future scaling for ML models, PCA, encoding, training and testing sets, linear regression model for predictive modelling. We also included a Folium, very useful Python Library used for visualizing geospatial data. Libraries have been imported and assigned the abbreviated formats. The abbreviated format makes recalling and use of these libraries more efficient. Lastly, we have uploaded a csv file that we renamed as a hotel.df for easy reference.

Dataset and Data Understanding

Data was acquired from Property Management Systems SQL databases. The dataset contains actual bookings due to arrive from 01st July 2015 to 31st Aug 2017 and includes bookings that have materialized and bookings that were cancelled. The data set has 119390 observations and 32 features, where each observation represents a hotel booking.

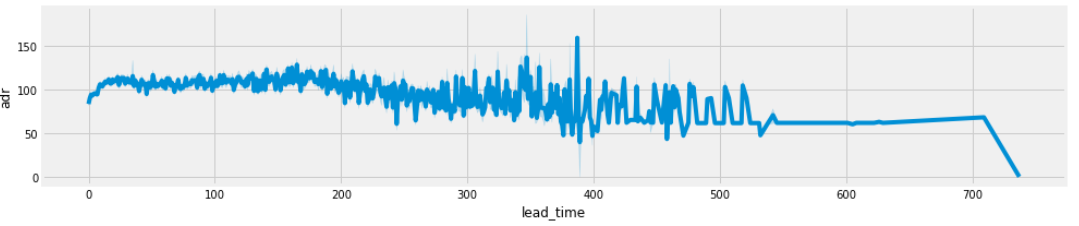
Data Preparation

* Counting the sum of missing values, the dataset shows country, agent and company
* Filling the missing values with zero
* Deleting rows with zero values for adults, children and babies in the same row, total of 180
* Identifying the target value ADR, vs is-cancelled

Data Visualisation

I did several visualizations to review the data and have better understanding of the dataset.

1. pair plot of the first ten variables that shows us the relationships between pairs of variables (ref. to Jupyter file, code block: )
2. line plot shows lead time for all bookings and there’re some outliers booking as far as 2 years in advance



1. Heatmap of correlations shows us relationships between variables and how they are correlated. We can see the following correlations: that repeated guests to bookings not cancelled, lead time and is cancelled, stays weekend and weeknights, agent and company, ADR and children, agent and company.

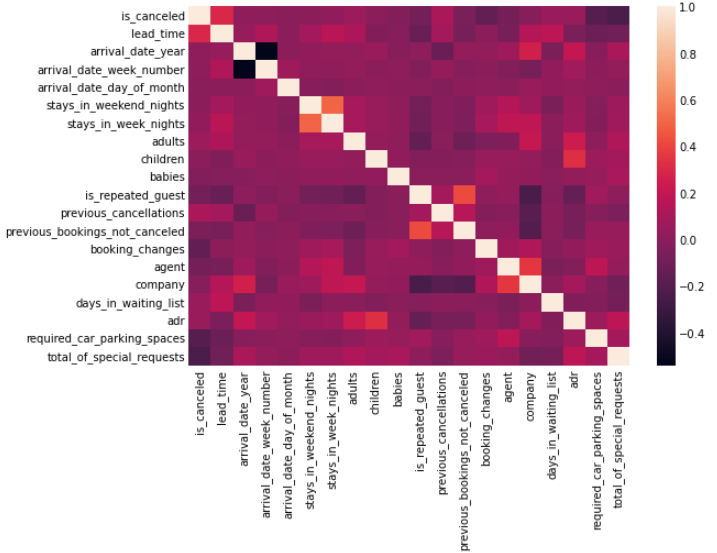


Figure x Heatmap of correlations

Hotel Revenue Management

Revenue management professionals suggest that total hotel Revenue Management (RM) is the wave of the future and that, going forward, technology and data analytics will help enhance RM decisions in the hospitality industry. These are among the findings in a new study from the Cornell Center for Hospitality Research (CHR) based on surveys designed to determine how hotel RM practices have evolved over the past six years and where they are headed. (Kimes, 2017)

When it comes to hotel revenue management, the following are essential business knowledge areas:

* Evaluate historical data to identify Average Daily Rate (ADR)
* Revenue per Available Room (RevPAR)
* % occupancy rates
* Direct bookings and distribution channels
* Meeting, Incentive, Conferences and Events (MICE) - average group sizes and even activity

Effective hotel Revenue Management is crucial and can improve numerous aspects of the hotel operations, such as:

* Selling rooms at the highest possible rate
* Managing staff roster effectively by having the right number of staff on duty
* Managing occupancy by minimising cancellations or unsold rooms
* Managing resources more effectively by reducing the waste products such as food and beverages (F&B)

Let’s have a look at few categories:

1. Market Segment (online, groups)

Hotel market segmentation will help to identify different categories of guest based on the travel habits and reservation patterns. It helps the hotel to detect new business in certain areas and the hoteliers can target travellers with tailored marketing materials while focusing on the specific pricing plans. Market segmentation can benefit in many ways to better understand their guest’s needs like buying power, booking patters and preferred choice of hotels. By addressing those needs, hoteliers can get better insights, better forecasting and improve the budgeting decision making process. In return, they’ll be able to increase profitability and operational efficiency.

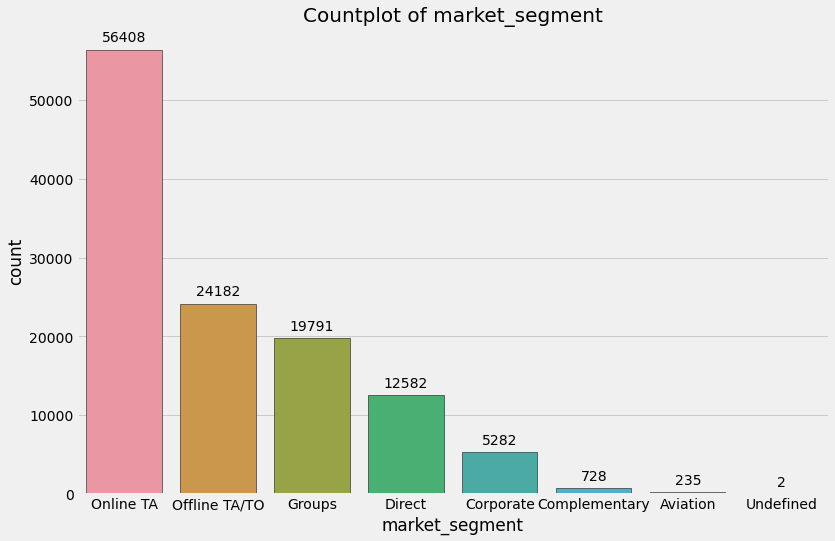


Figure x: Countplot of market segment

1. Arrival month

To predict data with the goal of maximising revenue, it is crucial to take few factors into consideration. Understanding the world’s holiday seasons, global conditions and special events in the region will help to understand when is an excellent opportunity to boost hotel rates because there’ll be a higher demand for the rooms due to the large number of reservations. for the off-peak season, when there’s a lower demand, hotel can create special deals to attract more traffic and foster consumer’s loyalty. creating a hotel package will additionally diversify the revenue streams by upselling a fine dining experience, spa treatments or variety upgrades, be it a better room or an executive floor usually reserved for premium guests.

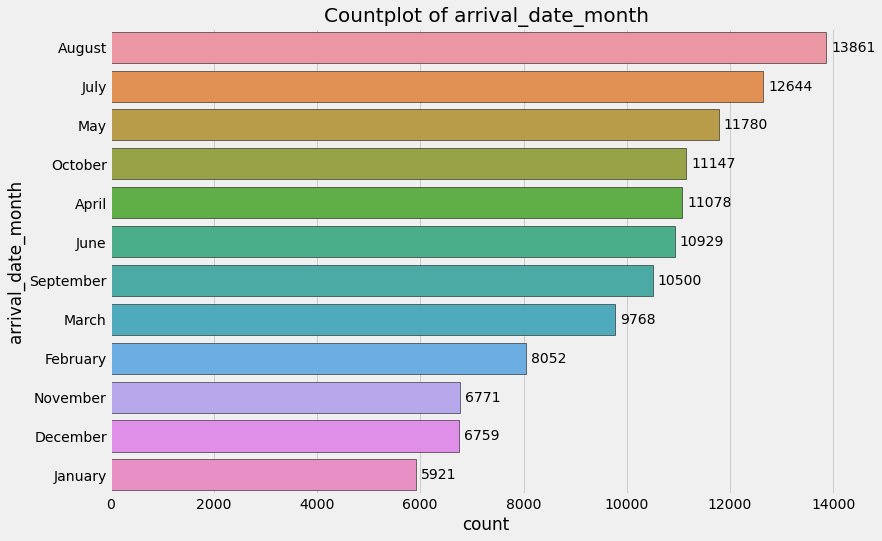


Figure x: Countplot of guest arrivals per calendar month

1. Cancellations

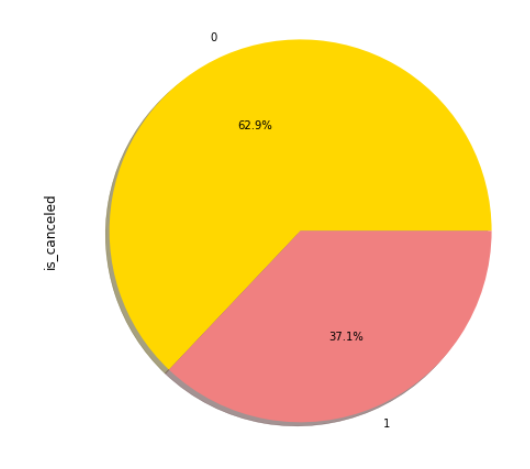
Cancellations directly impact revenue, particularly last-minute cancellations lead to loss of revenue. Even, if the hotel is able to sell the room again, it’ll likely be offered at a lower room rate or it’ll be a loss-making entity for the day. The data analytics can help and understand the reasons for cancellations and how to avoid it, since it leads to extra work and preparations for guests’ arrival. There could be a genuine reason for cancellations, such as event or conference is re-rescheduled, family holidays not materialised due to some personal emergency but it could also be fraudulent bookings that guest or agency make in order to obtain visa for the customer coming from a country with stringent visa policy into the respective destination. It is crucial for hotel to have a clear refund & cancellation policy that is clearly communicated to guests and groups alike.

Figure % of room cancellations

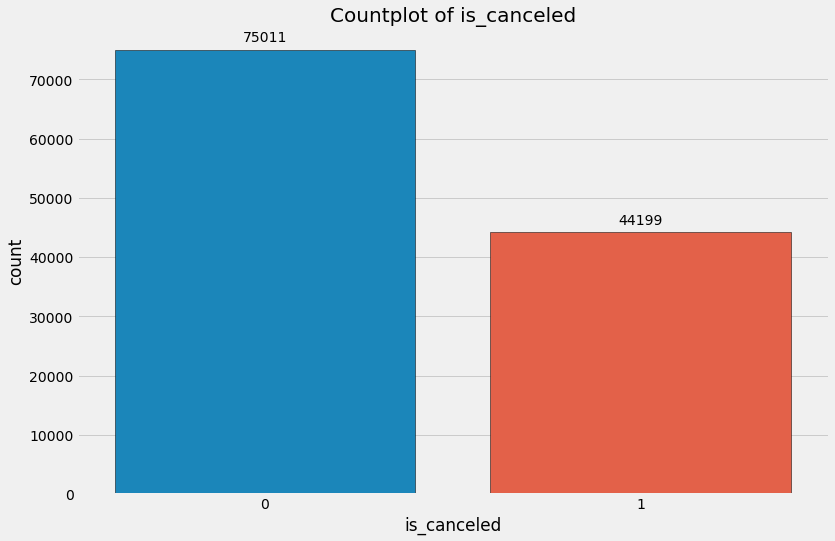
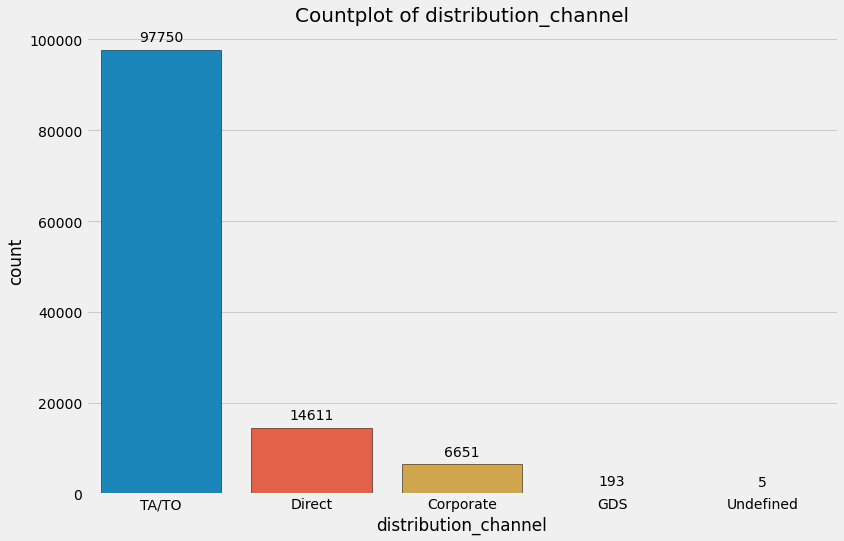


Figure Countplot of total cancellations

1. Distribution channels

Hotels have a wide range of distribution channels at their disposal, some examples are travel agents, Global Distribution System (GDS like Amadeus, or Galileo Airlines reservation systems), variety of booking platforms like [Booking.com](http://booking.com/), [hotels.com](http://hotels.com/) and many more. While these wholesalers are important to fill that last-minute rooms, to book a block of rooms for group travel, or to have online booking available to guest 247/ with instant confirmation, they should not be the primary distributors and the profit margin might be lower for this type of distribution due to commission being changed by the online portals and special room rates negotiated by the travel agents.



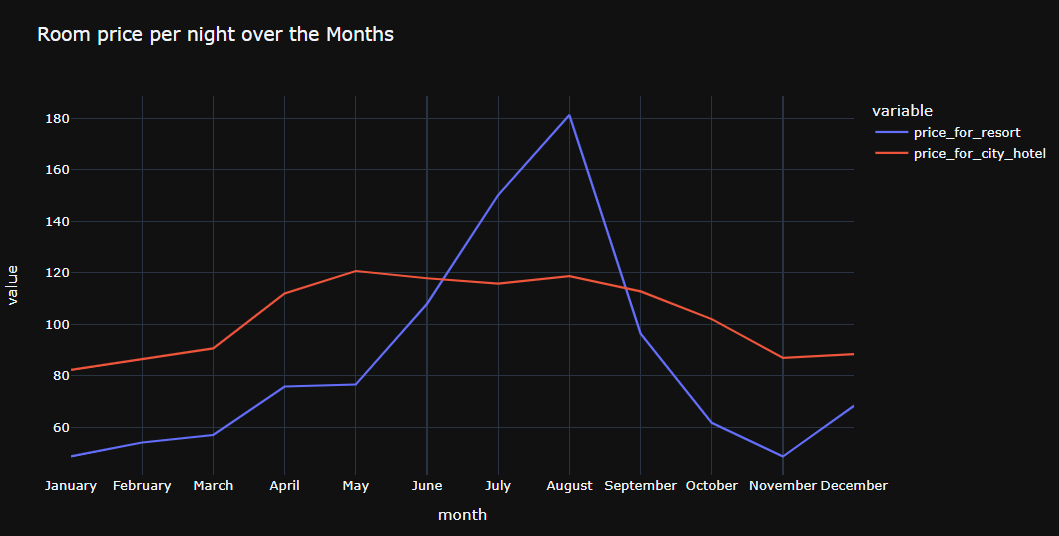
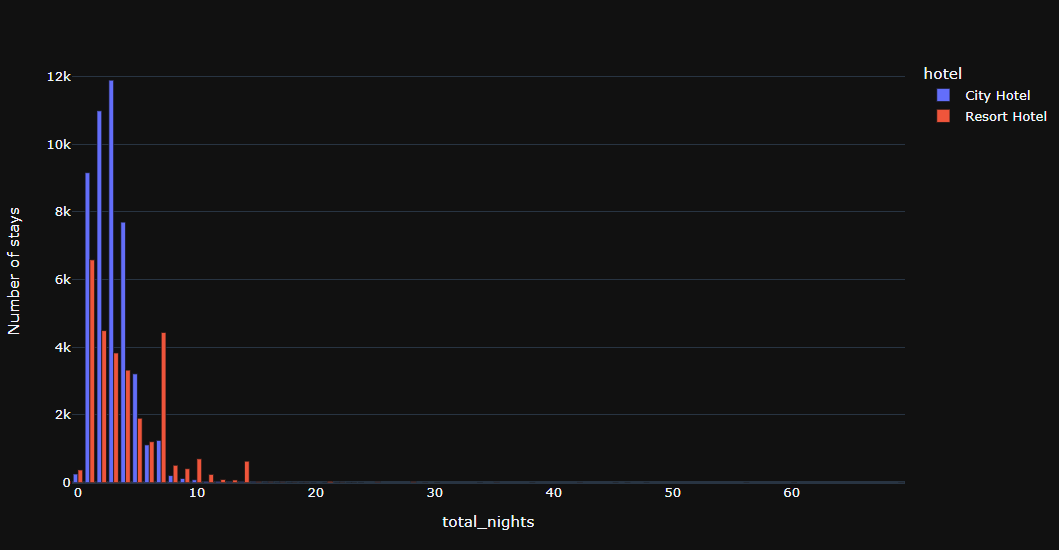
Figure

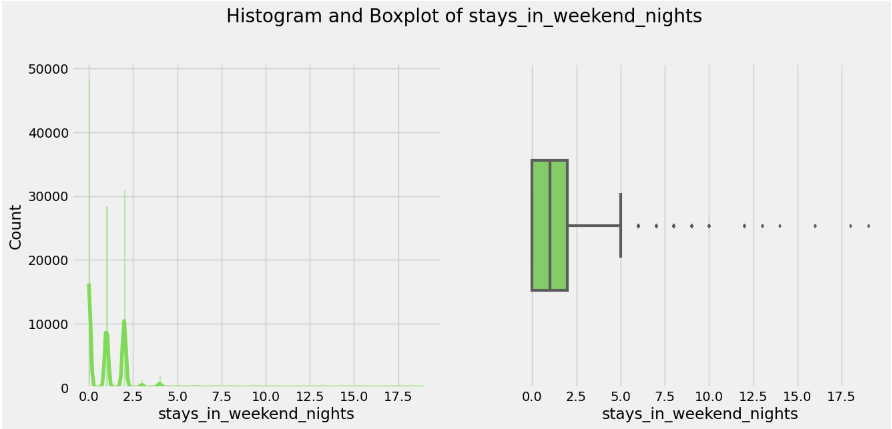
1. Prices of room per night (room rate)

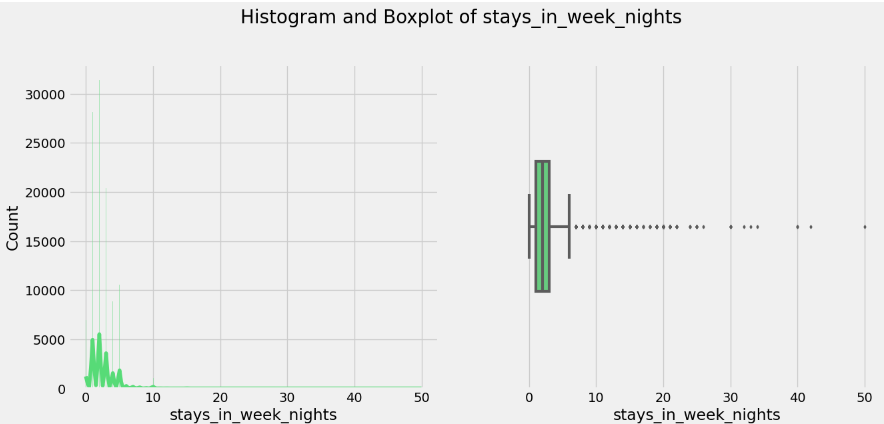
Hotel pricing is typically determined by the supply and demand, during the high seasons like holidays, regional events or seasons depended destinations determined by weather, be it a summer for beach holidays or skiing in winter.

During the high demand or peak-seasons prices tend to climb up whereas in low seasons also called off-peak time, the room rates are lower and travellers can get a good deal on the hotel stay as long as they don’t mind the off-peak travel. While this could be a primary factor in determining the pricing strategy, there’re other factors to take into consideration as well. Hoteliers need to compare the competition rates, calculate the necessary operational cost and profit margins to keep the hotel afloat while delivering an optimal occupancy and good revenue. There’s no one size fits all model and every property has to do their due diligence and decide what works for them.

Number of stays - city hotel more more nights.. resorts loaners







**Analysis of Results**

1. City hotels generated more bookings than resorts
2. Pie  charts show us the demand is greater for city hotels than a resort hotels

 3.

1. Pie chart show us that 37.1 % of bookings are cancelled
2. Direct bookings with hotel are   , 79.1% is made though Travel Agents or Tour Operators
3. Top 3 countries with highest nubmer of guest arrviing: Portugal 20977, Great Britain 9668, adn France 8464…
4. Average ADR for city hotel is higher than for resort hotels.

## Challenges encountered

I have encountered numerous challenges while working on this dataset. These may include looking for the suitable dataset, they were either too large or too small, I went with the bigger size of dataset. Larger volume of data provides more information for the training and might lead to better performance.

Additionally, I’ve identified a lot of missing vales, had a challenge with identifying the target value - (ADR, vs is-cancelled) and identifying the best performing model with better accuracy score.

Admittedly, the project was a challenging task , but generally speaking, I really enjoyed learning and attempting to understand such a complex subject of study as Data Analysis and Machine Learning.

**Conclusion and future recommendations**

GitHub link:

Timeline:

[Project Timeline\_Miroslava Slavikova\_CA2.odp](Project%20Timeline_Miroslava%20Slavikova_CA2.odp)

**References & Bibliography:**

# Hotel Asset Management: Principles and Practices, 1 Mar. 2016, by [Richard E. Musgrove](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_1?ie=UTF8&field-author=RIchard+E.+Musgrove&text=RIchard+E.+Musgrove&sort=relevancerank&search-alias=books-uk) (Author), [CHAM](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_2?ie=UTF8&field-author=CHAM&text=CHAM&sort=relevancerank&search-alias=books-uk) (Author), [CHA](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_3?ie=UTF8&field-author=CHA&text=CHA&sort=relevancerank&search-alias=books-uk) (Author), [Lori E. Raleigh](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_4?ie=UTF8&field-author=Lori+E.+Raleigh&text=Lori+E.+Raleigh&sort=relevancerank&search-alias=books-uk) (Author), [ISHC](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_5?ie=UTF8&field-author=ISHC&text=ISHC&sort=relevancerank&search-alias=books-uk) (Author), [A.J. Singh](https://www.amazon.co.uk/s/ref=dp_byline_sr_book_6?ie=UTF8&field-author=A.J.+Singh&text=A.J.+Singh&sort=relevancerank&search-alias=books-uk) (Author)

<https://revenue-hub.com/future-hotel-revenue-management/>

<https://www.sciencedirect.com/science/article/pii/S2352340918315191>

<https://www.kaggle.com/>

<https://towardsdatascience.com/how-to-clean-your-data-in-python-8f178638b98d>

<https://www.analyticsvidhya.com/blog/2020/06/guide-geospatial-analysis-folium-python/#:~:text=Folium%20is%20a%20Python%20library,library%20for%20plotting%20interactive%20maps>.

<https://realpython.com/>

<https://www.geeksforgeeks.org/>