A logo for college computing

Description automatically generated

# HotelRank: Elevating Revenue Performance Through Machine Learning and Deep Learning Techniques.

|  |  |
| --- | --- |
| *Student Name:* ***Giorgio Zoppi*** |  |
| *Student Number:* ***sba2301*** |  |
| *Module Name:* ***Strategic Thinking (HDip in Data Analytics - Feb 2024 - HCI cohort)*** |  |
| *Assignment Title:* ***CA 1 – Capstone Project Proposal*** |  |
| *Assessment Due Date:* ***29th March 2024*** |  |
| *Date of Submission* |  |

**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.

Contents

[HotelRank: Elevating Revenue Performance Through Machine Learning and Deep Learning Techniques. 1](#_Toc162131219)

[*Module Name:* ***Strategic Thinking (HDip in Data Analytics - Feb 2024 - HCI cohort)*** 1](#_Toc162131220)

[Introduction 1](#_Toc162131221)

[Forecasting of hotel room demand. 1](#_Toc162131222)

[Prediction of booking cancellations. 1](#_Toc162131223)

[Online hotel reputation. 1](#_Toc162131224)

[Problem Domain and Objectives 2](#_Toc162131225)

[Demand forecasting. 2](#_Toc162131226)

[HotelRank Score. 2](#_Toc162131227)

[Scope. 2](#_Toc162131228)

[**Phase** 3](#_Toc162131229)

[**Objectives** 3](#_Toc162131230)

[**Milestone Deadline** 3](#_Toc162131231)

[**21st May 2024** 3](#_Toc162131232)

[*Explore common proprieties in the datasets.* 3](#_Toc162131233)

[*Visualization of the datasets. Understanding data patterns.* 3](#_Toc162131234)

[**4th June 2024** 3](#_Toc162131235)

[*Check the quality of data.* 3](#_Toc162131236)

[**15th June 2024** 3](#_Toc162131237)

[*Create a merged datasets to include weather and events.* 3](#_Toc162131238)

[*Cross reference hotel and user reviews. Select/Add/Remove features.* 3](#_Toc162131239)

[**15th July 2024** 3](#_Toc162131240)

[*Handling missing values.* 3](#_Toc162131241)

[*Handling Duplicates.* 3](#_Toc162131242)

[*Assure Data Consistency.* 3](#_Toc162131243)

[**1st August 2024** 3](#_Toc162131244)

[*Once the data is clean format in a query able data source to facilitate training.* 3](#_Toc162131245)

[**1st August 2024** 3](#_Toc162131246)

[*Creating models, training and evaluating their performance using an iterative approach. Detect overfitting and underfitting. K-Fold Cross validation to determine which model performs better.* 4](#_Toc162131247)

[**1st September 2024** 4](#_Toc162131248)

[*Creating models, training and evaluating their performance using an iterative approach. Detect overfitting and underfitting. K-Fold Cross validation to determine which model performs better.* 4](#_Toc162131249)

[**1st September 2024.** 4](#_Toc162131250)

[*Tune ML and deep learning models hyperparameters, visualization and cross validation.* 4](#_Toc162131251)

[**15th October 2024** 4](#_Toc162131252)

[*Top reviews analysis and score computation using classification algorithms.* 4](#_Toc162131253)

[**15th November 2024** 4](#_Toc162131254)

[*Select the models and weight for HotelRank and perform the computation on the dataset. Perform revenue predictions and visualize them.* 4](#_Toc162131255)

[**15th December 2024** 4](#_Toc162131256)

[**20th January 2025** 4](#_Toc162131257)

[**1st February 2025.** 4](#_Toc162131258)

[DataSources 4](#_Toc162131259)

[Ethical Considerations 4](#_Toc162131260)

[References 5](#_Toc162131261)

# Introduction

Revenue management involves employing information systems and pricing strategies to effectively assign appropriate capacity to customers, ensuring it aligns with the optimal price and timing (Ivanov, 2014).

Hotel revenue management is a critical to maintain profitability. Among the various factors contributing to an hotel revenue management, some stand out prominently:

* Forecasting of hotel room demand.
* Prediction of booking cancellations.
* Online hotel reputation.

### Forecasting of hotel room demand.

Forecasting hotel room demand involves predicting the anticipated demand for rooms on any given day throughout the year. Accurate forecasts enable hotels and revenue managers to adjust prices dynamically, thus maximizing revenue potential. However, forecasting hotel room demand is a complex task, influenced by a multitude of factors that vary widely across different regions and contexts. Variables such as location, cultural events, weather, seasonal patterns, and more significantly impact optimal hotel revenue management strategies (Apostolos, 2021).

### Prediction of booking cancellations.

Booking cancellations (Nuno Antonio, 2019) are an issue for the hotel revenue manager because they make harder to predict the number of booked rooms. A common pattern to address this issue is trying to achieve the overbooking by lowering the price but when overbooking is real, it is a problem because it damages hotel reputation and revenue. People might find themselves without a room and complain online. So, from revenue manager perspective having a way to predict cancellation in advance is crucial in his decision-making process.

### Online hotel reputation.

Using the framework developed in (Diana-Jens & Rodríguez Ruibal, 2015) , we define what means online hotel reputation. Online reputation is *“the result of what clients, former clients, future clients, employees, etc. say, write and communicate to another anywhere in the internet social media based on their perceptions and experience in any moment of their relationship, direct or indirect, with the brand”* The same study performed in the four main Spanish cities states that room price increases whenever the hotel is positioned in the top positions in TripAdvisor Index. As demonstrated in a Chinese study on ten luxury hotels (Wang, et al., 2023), negative online reviews can lead to a low booking rate that takes months to recover. So online reputation is an important factor to consider for revenue management.

Our focus is therefore to create a hotel ranking algorithm called **HotelRank** that considers all three factors (demand forecasting, cancellations and online reputation) in a weighted manner using machine learning and deep learning techniques.This research project will be executed in strict collaboration with **Blastness Spa,** the #1 provider in Italy for luxury hotels with a portfolio of over nine hundred hotels, that it will provide the consent of anonymized data.

# Problem Domain and Objectives

Our main goal is to increase the profit for our customers that are mainly hotel revenue managers providing a way to compare with competitors. To achieve our main goal, we want to build put our focus mainly on:

* model demand forecasting.
* model booking cancellation.
* analyse online hotels reputation.

## Demand forecasting.

Demand forecasting can depend on internal or external factors. Internal factors are the location, the historical data from the hotel and the segmentation of the customers. External factors are season, events, reputation. Our challenge is to compare our dataset with datasets about weather and events obtained using public API (i.e. weather.com and predicthq.com). Once we’ve completed our challenge, we will have enough information to define a score. Our hypothesis to validate is that we can define a score that it is able to indicate the evolution of the demand in time.

Booking Cancellation.

As per demand forecasting, we use the same approach for booking cancellation, using our booking dataset to determine a score factor that represents the probability of cancellation in time.

Analyse Online Reputation.

Here our objective is to collect TripAdvisor data for the hotels in our dataset and see how the reviews and comments, providing our own reputation score. The hypothesis to validate here is that higher is the rank, higher is the revenue.

## HotelRank Score.

Once we’ve created the model, we can define an iterative process to craft **HotelRank** ranking score.

**HotelRank** might be a weighted combination linear between those factors:

During the data exploration and normalization, we want to refine the above formula for ranking using an iterative approach and set the correct weight **w1, w2, w3, w4.**

Once **HotelRank** is defined our goal is to productionize solution for future booking using a data engineering approach in the cloud.

# Scope.

We’re going to follow on the CRISP-DM methodology except for the infrastructure and deployment of the model we’re working on with Scrumban (Alliance, 2017). An

important point in CRISP-DM is that these phases are not necessarily sequential; the

construction of a model usually requires several cycles. Over the arc of two semesters the scope of the project will try to answer the following questions:

* Which is most accurate model for demand forecasting?

Can the weather impact or external events impact the demand?

* Can cancellations affect revenue performance?
* How does our online reputation affect revenue performance?
* Can we determine how good we are respect our competitors?

We can summarize key milestones defined according to the project description and its deliverables. The deadlines might have some minor deviations due to the project complexity.

|  |  |  |
| --- | --- | --- |
| **Phase** | **Objectives** | **Milestone Deadline** |
| **Hotel Domain Knowledge Research** | *Domain Analysis.*  *Understand how ML techniques are used (Regression, Classification)*  *Understand how Deep Learning are used in the domain (RNN, LSTM, Prophet)*  *Report about domain knowledge.* | **20th April 2024** |
| **Data Collection** | **Hotel Bookings:** *The dataset consists of the bookings of ten Italian luxury hotels in a two-year period.*  **Weather.com API:** *A subscription to Weather.com API and an HTTP based client will allow us to collect and store in Azure Blob Storage the weather condition in the temporal to match.*  **PredictHP.com dataset:** *A subscription to the event API will allow to infer events in proximity to the hotels in our dataset.*  ***TripAdvisor*** *will provide access to user reviews for our hotel set.* | **21st May 2024** |
| **Data Exploration.** | *Explore common proprieties in the datasets.* *Visualization of the datasets. Understanding data patterns.* | **4th June 2024** |
| **Data Quality Checks.** | *Check the quality of data.* | **15th June 2024** |
| **Feature Engineering** | *Create a merged datasets to include weather and events.* *Cross reference hotel and user reviews. Select/Add/Remove features.* | **15th July 2024** |
| **Cleaning Data** | *Handling missing values.**Handling Duplicates.**Assure Data Consistency.* | **1st August 2024** |
| **Integrate Data in Iceberg Tables** | *Once the data is clean format in a query able data source to facilitate training.* | **1st August 2024** |
| **Modelling: Demand and cancellation forecast models** | *Creating models, training and evaluating their performance using an iterative approach. Detect overfitting and underfitting. K-Fold Cross validation to determine which model performs better.* | **1st September 2024** |
| **Modelling: Demand and cancellation forecast models.** | *Creating models, training and evaluating their performance using an iterative approach. Detect overfitting and underfitting. K-Fold Cross validation to determine which model performs better.* | **1st September 2024.** |
| **Hyperparameters tuning.** | *Tune ML and deep learning models hyperparameters, visualization and cross validation.* | **15th October 2024** |
| **Compute Reputation Score** | *Top reviews analysis and score computation using classification algorithms.* | **15th November 2024** |
| **Compute HotelRank.** | *Select the models and weight for HotelRank and perform the computation on the dataset. Perform revenue predictions and visualize them.* | **15th December 2024** |
| **Data Engineering Automation to adapt the model and the score to new data and customers.** | *Deploy the selected models in the cloud.*  *Automate the data collection in the cloud.*  *Automate the data cleaning in the cloud.*  *Automate chosen models training.*  *Automate the HotelRank computing.*  *REST API and Documentation for hotel revenue managers that give them access to the models for prediction and ranking.* | **20th January 2025** |
| **Reporting and Project Close.** | *Project report with summary of the results indicating the chosen models and the process.*  *Project review: Retrospective document to indicate what went well and which are the areas of improvement.* | **1st February 2025.** |

# DataSources

# Ethical Considerations

Our data come from **Blasteness** that owns the explicit consent to anonymize customer names and hotel names. The data is placed in a public available Azure Blob Storage at s3://

Other sources of information come from API subscription provided by **Blastness** to TripAdvisor, Weather.com and PredictHQ. In this case we

# References