# Introduction

Revenue management (Ivanov, 2014) is a critical to maintain hotel profitability. Among the various factors contributing to a hotel revenue management strategy, three main factors 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, factors such as location, cultural events, weather, seasonal patterns and more, significantly impact on optimal hotel room forecasting (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 a revenue manager perspective it is important having cancellations predictions.

### 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”*. 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.

Our focus is therefore to create a hotel ranking algorithm called **HotelRank** that considers all these three factors in a weighted manner using AI.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 has provided the anonymized data and its consent to use.

# 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 put our focus on creating models for demand forecasting, booking cancellation and online 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, and reputation. Our challenge is to validate our dataset with datasets about weather and events obtained using a public API (i.e. weather.com and predicthq.com), cleaning the data, selecting the model, evaluate his performance and predict results. Our hypothesis to validate is that we can define a score that it is able to indicate the evolution of the demand in time using ML models.

## Booking Cancellation.

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

## 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:

After model training, the correct weights **w1, w2, w3, w4** are set.

Once **HotelRank** is defined our goal become to make models predictions and hotel ranking accessible to our customers in Azure.

# Scope.

Project management CRISP-DM methodology will be used. An important point in CRISP-DM is that it is meant to be an iterative process since the construction of a model 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?

In the table below we summarize key milestones. The deadlines might have some minor deviations due to the project complexity, so we’ve kept one month buffer at the end.

|  |  |  |
| --- | --- | --- |
| **Phase** | **Objectives** | **Milestone Deadline** |
| **Hotel Domain Knowledge Research** | *Domain Analysis. Understand how ML techniques are used. Understand how Deep Learning are used in the domain. Report about domain knowledge.* | **20th April 2024** |
| **Data Collection** | *The dataset consists of the bookings of eight Italian luxury hotels in a two-year period. Data Collection from Weather.com. Data Collection from PredictHP.com. Data Collection from TripAdvisor.* | **21st May 2024** |
| **Data Exploration.** | *Explore common proprieties in the datasets. First cleaning the data, remove all Italian references.* *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 using ML.** | *Creating models, training and evaluating their performance using an iterative approach using XGBoost, Regression, RNN, LSTM and Prophet. 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 and visualization.* | **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.** | *Deploy the selected models in the cloud.*  *Automate all data flow from ingestion to model training. Provide model access to hotel revenue managers via REST API.* | **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.** |

# Data Sources.

## Hotel Datasets.

The data has been provided by Blastness.com and released on Creative Common License in the GitHub repository <https://github.com/CCT-Dublin/capstone-project-feb-2024-pt-giorgiozoppi> in the folder hoteldataset. That folder contains data of booking about ten hotels but only eight are fully usable, during a period of two years at least. For some hotels we’ve more then two years and data cleaning should match the exact date frames.

Blasteness.com has also provided the location, latitude and longitude of each hotel to cross reference with TripAdvisor, PredictHQ.com and Weather.com API but this will not disclosed for keeping the hotel name private. We will disclose the city during the data cleaning and feature extraction.

## Other datasets.

# References