Cloud for AI

Task 1: Project Proposal

*Jordy Corthals, Raymond Tréville, Yrja Van Kersschaver*

## Business Problem

A hotel business notices a lot of cancellations in their past bookings. This leads to a loss of revenue and an imbalance of personnel. They would like to have a system provide the ideal daily rate in order to maximize room occupation while focusing on customer retention. Having guests stay over a longer period of time increases the personal experience they can provide hoping that guests return after a satisfactory stay. For the hotel itself, longer stays are easier to manage than short ones. It will help simplify inventory management as well as with scheduling the staff. Less cancellations will save time on last minute rescheduling.

Based on the hotel’s request, and after a quick look in the data provided, we have some potential solutions in mind. Considering the focus on the financial aspect, linear machine learning algorithms emerge as potential solutions. Particularly, Linear Regression and K-means. The data that has been provided works well for supervised models and will aid in making informed decisions for the future based on the historical data.

## Feasibility

Feasibility of a solution regarding this problem is supported by the amount of data available over the course of two years in combination with a broad spectrum of labels within this data pertaining to the issue. The existence of businesses already applying these type of solutions, like Booking.com, suggest that this topic is not unique and that solutions might already be available.

## Benchmarks

We will be comparing our model to the performance of the hotel’s previous way of working. We aim to measure the improvement in the measures we describe in the metrics part below. These are only the measurably features. We do think that next to this there will be quite some qualitative improvements as well. Less stress to have to handle last minute schedule adjustments as well as providing more time to focus on other tasks. We are unable to measure these ourselves, but some time after deployment of our model we could gather some feedback from the hotel for these indirect influences.

## Metrics

#### Machine Learning Metric

We will be using Root Mean Squared Error (RMSE) as our machine learning metric. It is suitable for machine learning models that focus on numerical values and is commonly used as a standard metric.

Our solution will not be needing any heavy computational force or expensive infrastructure. We will be scripting from our own laptops and deployment will be done by using open-source software.

#### Software/Business Metric

The business metrics include higher revenue, improved customer retention, and an increase in returning customers. All of this in comparison to the previous values of these metrics before deployment of our solution.

## Societal value and ethical considerations

In our analysis we will be excluding unnecessary personal data such as credit card details, telephone numbers and e-mails. These are not relevant to the purpose of our project. We will not be making any distinctions in our analysis based on gender and age. While we will be including country data, we will not be using it to distinguish daily rates in order to ensure fairness in our system. We do think there might be quite some relevant insights to be gained from the country data, however we would like to focus on a base of non-discrimination. We will leave people’s names in our analysis so we can measure customer retention and return rates. Other than for this purpose, we will refrain from using names for any other reason and will limit the use within our scope of the project.