

## **Abstract**

The goal of the project is to analyze cancellations in hotels, whether resorts or city hotels, to help devise new methods to reduce the consequent losses.

Cancellations limit the production of accurate forecasts, a critical tool in terms of revenue management performance. To circumvent the problems caused by booking cancellations, hotels implement rigid cancellation policies and overbooking strategies, which can also have a negative influence on revenue and reputation

Using data sets from cancellation of hotel booking demand as a problem in the scope of data science, the results allow hotel managers to accurately predict net demand and build better forecasts, improve cancellation policies, define better overbooking tactics and thus use more assertive pricing and inventory allocation strategies, the used data in this project are provided by Kaggle.

## **Design**

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project.

## **Data**

The dataset is provided in .CSV format , its contains a single file which compares various booking information, this data set contains booking information for a city hotel and a resort hotel. and includes 119390 rows and 32 columns such as cancellation, deposit type, lead time, and the total of special requests, among other things.

## **Algorithms**

Feature Engineering

- Cleaning the dataset by dropping feature that won't be relevant in the analysis, which is the company column, and dropping Null rows.
- Cleaning the average\_daily\_rate column by dropping the negative values
- Cleaning by dropping the duplicates
- Rename the column ADR to average\_daily\_rate to be more readable

## **Tools**

- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting

## **Communication**

The slides are provided [here](#) .