Data Science Challenge

Showcase your data science expertise using a hands-on example

Overview

Context

You work for the **data science department** of a **hotel chain** operating hotels in several European countries.

At the moment, the chain is building a **new data-science-heavy tool to increase revenues**.

The tool leverages **prediction** and **optimization** techniques to find the revenue-optimizing price for each room across all its hotels

The tool is **jointly developed** by the **data science** department and the **revenue management team**.

Your Task

In preparation for an upcoming meeting with the revenue management team, your boss asks you to **explore potential inputs** of the future pricing tool and the **prospects of predicting hotel bookings** using data science techniques.

Summarize your answers to their questions on **5 to 10 slides** and be prepared to present the slides and the corresponding code.

In what follows, you can find a list of all questions to be addressed in your presentation.

Data and Tech-Stack

The hotel_bookings.csv dataset contains bookings for a selection of hotels where the revenue tool will be rolled out first.

Python should be the technical backbone of your tech-stack.

You may use additional tools where you feel they are necessary. In case you do, please plan additional time to explain why you used these tools and how they work.

General instructions

Input

Data

You can find the only dataset required to successfully complete the data science challenge ("hotel_bookings.csv") attached to your recruiter's email.

Please contact david.nielsen@tuicruises.com if you face difficulties in opening the dataset.

Submitting Results

Code

Please submit your code and all supplementary files (virtual environments, etc.) we may need to execute your code as a .zip file to your recruiter.

Slides

Please send your slides as a .pdf file to your recruiter. Please include an estimation of how much time you spend in total on the data science challenge. Sharing this information will not be part your assessment but will help us in further developing the data science challenge

Task 1 | Explorative Analysis of Potential Inputs to the Pricing Tool

TUI Cruises
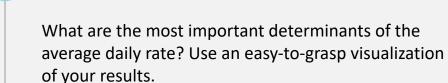
Customers

Questions

From which countries are the most guests from? Pick a suited visualization to display your findings.

Along which dimensions do guests from different source countries differ the most?

What implications do your findings from the customer analysis have for the hotel chain's pricing strategy?

Prices

Questions

How do prices measured by the average daily rate (adr) differ across hotels and time? Display your results using a suitable visualization.

What are the most important determinants of the average daily rate? Use an easy-to-grasp visualization of your results.

What implications do your findings have for the hotel chain's pricing strategy?

Task 2 | Predicting Bookings using Machine Learning

TUICruises

Task

Goal

Build a machine learning model that allows to predict weekly bookings for each hotel over time.

Algorithms

Your machine learning pipeline should train at least two algorithms suited to solve the problem at hand and evaluate their performance on *unseen* data using appropriate metrics. Be prepared to explain why you choose these algorithms and metrics.

Further Instructions

Packages

Feel free to use any packages you want. Please be prepared to give a quick introduction to packages less familiar even to data science practitioners.

Delivery

Summarize the main insights from your analysis using suitable visualizations or tables.

Please don't forget to submit your solutions and slides in time.

TUI Cruises