Lorem Ipsum Hotel Group: Hotel Cancellation Prediction Challenge

Welcome, Data-Driven Consultants!

As the CEO of the **LOREM IPSUM HOTEL GROUP**, I am reaching out to you with a critical challenge that demands our immediate attention. Over recent months, we have witnessed a troubling trend of escalating cancellations within our hotel properties, posing significant implications for our business operations and bottom line. Previous attempts to address this issue have so far not yielded the desired results.

The decision to pursue aggressive overselling tactics initially seemed promising. However, this approach has proven to be unsustainable in the long term, as it has resulted in instances where we were unable to accommodate all booked guests, leading to guest dissatisfaction, negative reputational damage, and financial costs from having to pay to relocate guests to other hotels.

On the other hand, the implementation of strict cancellation policies was intended to mitigate the impact of cancellations on our revenue streams. While these policies were introduced with the aim of incentivizing guests to commit to their reservations, guests perceived our cancellation policies as inflexible and punitive, leading them to seek alternative accommodation options that offer more lenient terms. As a result, we have experienced a decrease in booking conversions.

It is evident that our current approach to managing cancellations is no longer sustainable and requires a fundamental reassessment. This project represents a critical opportunity for us to pivot towards a more strategic and data-driven approach to revenue management.

In your capacity as Data-Driven Consultants, we are trusting you with the responsibility of analyzing our historical booking data. Note that there are some peculiarities related to hotel booking data; you can read more about that in António et al. (2019).

We would like you to use KNIME Analytics software to develop a predictive model for hotel cancellation prediction. To help us understand how to achieve our vision, we want you to write a report consisting of **at most** 5-6 pages (excluding cover page and annexes), discussing the points below, making sure the report as a whole is accessible to individuals with varying levels of technical proficiency.

1. Executive Summary

   Provide an overview of your analysis approach, methodologies, and findings.

2. Business and Data Understanding; Data Engineering

   Discuss insights derived from the data exploration process. What features did you decide to use or exclude, and why? What kind of preprocessing did you perform?

3. Model Engineering and Evaluation

   Discuss insights derived from the model development process. Why did you select this model? Which evaluation metric did you decide to use and why?

4. Deployment

   Suggest how you would deploy the model. Offer actionable recommendations for optimizing revenue based on your predictions. If the company decides to deploy your model, how would this impact the business? Are there any risks to deploying the model?

5. Monitoring and Maintenance

   How would you make sure you are able to maintain the model's quality? Are there features that the hotel does not collect that you think might be useful? Are there other models that might be helpful to test?

6. Conclusion

   Describe any additional insights or suggestions you may have. What challenges did you encounter? Given more time or resources, what would you do differently?

7. Annexes

   Optional. Supplementary materials such as figures or tables to support your report. This does not count towards the page limit.

Project deliverables:

1. KNIME Workflow with trained classification model. Make sure that every KNIME workflow's nodes run with the dataset DDDM24_Hotel_Data.csv

   This file should be named **Hotel_99.knwf** where "99" is your group number.

2. Report with the aforementioned structure (maximum 5 pages excluding title page and annexes).

   This file should be named **Hotel_99.pdf** where "99" is your group number.

3. Predictions for the test set provided (DDDM24_Hotel_Test.csv).

   This file should be named **Hotel_99.csv** where "99" is your group number.

Project materials provided:

1. DDDM24_Hotel_Data.csv

   File containing booking information data to be used for training your model

2. DDDM24_Hotel_Test.csv

   File containing data to use for prediction

3. Dataset Description

   Next page

References:

António, N., Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. Data in Brief, 22, 41–49. https://doi.org/10.1016/j.dib.2018.11.126

Dataset Description (Based on António et al., 2019) :

Variable	Description
BookingID	Booking ID
ADR	Average Daily Rate
Adults	Number of adults
Agent	ID of the travel agency that made the booking
ArrivalDate	Date of arrival
Babies	Number of babies
BookingChanges	Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
Children	Number of children
Company	ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons
Country	Country of origin. Categories are represented in the ISO 3155-3:2013 format
CustomerType	Type of booking, assuming one of four possible categories: Contract – when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking
DaysInWaitingList	Number of days the booking was in the waiting list before it was confirmed to the customer
DepositType	Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made;

	Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay. In case no payments were found the value is "No Deposit". If the payment was equal or exceeded the total cost of stay, the value is set as "Non Refund". Otherwise the value is set as "Refundable"
DistributionChannel	Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
IsCanceled	Value indicating if the booking was canceled (1) or not (0)
IsRepeatedGuest	Value indicating if the booking name was from a repeated guest (1) or not (0)
LeadTime	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
MarketSegment	Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
Meal	Type of meal booked. Categories are presented in standard hospitality meal packages Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)
PreviousBookingsNotCanceled	Number of previous bookings not cancelled by the customer prior to the current booking
PreviousCancellations	Number of previous bookings that were cancelled by the customer prior to the current booking
RequiredCarParkingSpaces	Number of car parking spaces required by the customer
ReservedRoomType	Code of room type reserved. Code is presented instead of designation for anonymity reasons

StaysInWeekendNights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
StaysInWeekNights	Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
TotalOfSpecialRequests	Number of special requests made by the customer (e.g. twin bed or high floor)