**Hotel Bookings**

**Assignment M9.9**

May 6, 2022

Saksham Bansal

BIA-610 APPLIED ANALYTICS

Stevens Institute of Technology – Spring 2022

Table of Contents

[1 Volksvagen group 2](#_Toc102234874)

[1.1 what is the landscape of the global automobile industry? how did vw group’s position evolve over the year? 2](#_Toc102234875)

[1.2 What is vw group’s strategy for 2018? 2](#_Toc102234876)

[1.3 What is big data? What role could big data play in this strategy? 2](#_Toc102234877)

[1.4 If you were the head of a business unit within VW Group, how would you leverage big data to help the company achieve this strategy? You can select any unit, such as any brand or function. 3](#_Toc102234878)

[1.5 How would you assess and prioritize different big data-related initiatives so that you could select a winning proposal for investment? 3](#_Toc102234879)

Table of Figures

No table of figures entries found.

# Hotel Bookings

## Introduction

The problem that is being addressed is that there are too many hotel bookings being cancelled all over the globe and hoteliers would like to find out why is this happening to devise a solution for it. To find the solution I am going to find insights using the data. When I am done with my analysis, hoteliers would get to know the statistical reasoning behind the cancellations of bookings and solve the problem.

## Data preparation

The data used in this analysis is obtained from [Github](https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-11/readme.md). The original source of the data is an open hotel booking demand dataset from [Antonio, Almeida and Nunes, 2019](https://www.sciencedirect.com/science/article/pii/S2352340918315191#f0010). The dataset contains a total of 32 variables. To clean the data, I used data interpreter in Tableau.

The final form of data can be seen in the picture below.

Table

Description automatically generated

|  |  |  |
| --- | --- | --- |
| **variable** | **class** | **description** |
| hotel | character | Hotel (H1 = Resort Hotel or H2 = City Hotel) |
| is\_canceled | double | Value indicating if the booking was canceled (1) or not (0) |
| lead\_time | double | Number of days that elapsed between the entering date of the booking into the PMS and the arrival date |
| arrival\_date\_year | double | Year of arrival date |
| arrival\_date\_month | character | Month of arrival date |
| arrival\_date\_week\_number | double | Week number of year for arrival date |
| arrival\_date\_day\_of\_month | double | Day of arrival date |
| stays\_in\_weekend\_nights | double | Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel |
| stays\_in\_week\_nights | double | Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel |
| adults | double | Number of adults |
| children | double | Number of children |
| babies | double | Number of babies |
| meal | character | Type of meal booked |
| country | character | Country of origin. Categories are represented in the ISO 3155–3:2013 format |
| market\_segment | character | Market segment designation |
| distribution\_channel | character | Booking distribution channel |
| is\_repeated\_guest | double | Value indicating if the booking name was from a repeated guest (1) or not (0) |
| previous\_cancellations | double | Number of previous bookings that were cancelled by the customer prior to the current booking |
| previous\_bookings\_not\_canceled | double | Number of previous bookings not cancelled by the customer prior to the current booking |
| reserved\_room\_type | character | Code of room type reserved |
| assigned\_room\_type | character | Code for the type of room assigned to the booking. |
| booking\_changes | double | 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 |
| deposit\_type | character | Indication on if the customer made a deposit to guarantee the booking. |
| agent | character | ID of the travel agency that made the booking |
| company | character | ID of the company/entity that made the booking or responsible for paying the booking |
| days\_in\_waiting\_list | double | Number of days the booking was in the waiting list before it was confirmed to the customer |
| customer\_type | character | Type of booking |
| adr | double | Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights |
| required\_car\_parking\_spaces | double | Number of car parking spaces required by the customer |
| total\_of\_special\_requests | double | Number of special requests made by the customer (e.g. twin bed or high floor) |
| reservation\_status | character | Reservation last status |
| reservation\_status\_date | double | Date at which the last status was set. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## exploratory data analysis

The first thing that I wanted to find out was what kind of data is available. So, I prepared the below mentioned analysis. As it can be clearly seen, in this dataset, maximum number of hotels are in Portugal and so, the maximum number of cancellations are also in Portugal. Map

Description automatically generated

Then, I counted the total number of city hotel cancellations and resort hotel cancellations and checked how many of the cancellations were made by repeated guests. City hotels had a very high number of cancellations compared to resort hotel. Repeated guests are less likely to make cancellations.

Chart, treemap chart

Description automatically generated

Further, I compared all the market segments to find out which segment had the highest average of cancellation rate and different room types and their cancellations. Groups had the highest cancellations average and most A type rooms were assigned and they had the highest cancellation number too.

Graphical user interface

Description automatically generated with medium confidence

Lastly, I wanted to know the effect of days in waiting list on the cancellation rate and different groups and their required car parking space to compare whether booking was cancelled on the basis of this or not. The higher the days in waiting list, the higher is the chance of cancellation. Transients required the most car parking space.

Graphical user interface, application, Word

Description automatically generated

I used RapidMiner to find out the variables that play the most important role. And the results are mentioned below. Clearly deposit type played the most important role.

## Summary

I wanted to find out the reason behind cancellations. Firstly, most data is Portugal based. If the customer is repeated one, chances of cancellation is lower. Action needs to be taken to prevent groups from cancelling. D type rooms should be assigned more. Lastly, less waiting list time is preferred.

To efficiently figure out the reason behind cancellations, more data is required on hotels from all over the world, to reduce the chances of bias.

## References

* <https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-11/readme.md>
* https://www.sciencedirect.com/science/article/pii/S2352340918315191#f0010