Contents

[Introduction 1](#_Toc150954973)

[Datasets summary 1](#_Toc150954974)

[References 2](#_Toc150954975)

# Introduction

The objective of this project is to perform Time Series Analysis and Text Analytics for Machine Learning for Business and create an interactive Dashboard for Data Visualization Techniques. Through both analyses, we aim to gain valuable insights into two different real-world datasets: Political Social Media Posts and Hotel Booking.

On one hand, we will use the dataset called Political Social Media Posts to apply Text Analytics where we will analyse 5000 messages from politicians’ social media accounts, along with human judgments about the purpose, partisanship, and audience of the messages posted on Twitter and Facebook.

On the other hand, we will work on the dataset called Hotel Booking to apply Time Series Analysis and use the same dataset to create an interactive Dashboard to effectively communicate the key insights derived from the exploratory data analysis. In this particular dataset, we can find information of two different hotels: the City Hotel and Resort Hotel, offering a comprehensive look into one key factor influencing the hospitality industry: seasonality, where there is a repeated behaviour. This behaviour is related to peak seasons (high demand) and off-peak seasons (low demand).

# Datasets summary

The Hotel Booking dataset contains over 100,000 observations and 32 attributes, with the target variable ‘is\_canceled’ column indicating whether the bookings were cancelled or not. Specifically, there are 75,166 bookings not canceled, while there are 44,224 bookings canceled.

There are key features to use in the interactive Dashboard.

* Type of Hotel
* Country: Represents where the guests’ country origin
* Average Daily Rate
* Arrival Date Month: Represents the number of arrivals per month.

For the Time Series Analysis, the key features include:

* Reservation Status Date: Represents the date where the guests booked a room.
* Average Daily Rate: Total revenue generated by all the occupied rooms.

The other dataset, Political Social Media Posts, provides 5000 messages, where the sources of the messages are equally distributed: 2500 are from Twitter and 2500 from Facebook posts.

The key features for Text Analytics are:

* Unique ID: Identifies each unique message.
* Judgement timestamp: Contains the date and time when the message was posted.
* Text: is the message itself sent by the people

# References

All, Moez. “Time Series Forecasting Tutorial.” Www.datacamp.com, 1 Feb. 2020, www.datacamp.com/tutorial/tutorial-time-series-forecasting.

Jason Brownlee. “A Gentle Introduction to the Box-Jenkins Method for Time Series Forecasting.” *Machine Learning Mastery*, 21 Feb. 2017, machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/.

“Using Machine Learning for Time Series Forecasting Project.” CodeIT, codeit.us/blog/machine-learning-time-series-forecasting.

“ARIMA and SARIMAX Models with Python.” Cienciadedatos.net, cienciadedatos.net/documentos/py51-arima-sarimax-models-python.html.

<https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf>

Kumar, Vijay. “Statistical Tests to Check Stationarity in Time Series.” Analytics Vidhya, 16 June 2021, [www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/](http://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/).

Rategain. “Impact of Seasonality on Revenue: Strategies to Navigate through the Peak and Off-Peak Seasons.” *RateGain*, 8 June 2023, rategain.com/blog/the-impact-of-seasonality-on-hotel-revenue/#:~:text=Seasonality%20is%20a%20crucial%20factor. Accessed 12 Nov. 2023.

“Sort a Pandas Dataframe Series by Month Name.” Stack Overflow, stackoverflow.com/questions/48042915/sort-a-pandas-dataframe-series-by-month-name. Accessed 12 Nov. 2023.

“Pie Charts.” Plotly.com, plotly.com/python/pie-charts/.

“Plotly.subplots.make\_subplots — 5.10.0 Documentation.” Plotly.com, plotly.com/python-api-reference/generated/plotly.subplots.make\_subplots.html.

plotlygraphs. “Subplots.” Plotly.com, 3 July 2019, plotly.com/python/subplots/.