# Predicting Hotel Bookings Cancellation

By: Eman Alwusaybie

# Table of Contents

BACKGROUND

Data Overview

RESULT SUMMARY

# BACKGROUND

Hotel Booking Demands It is an industry in itself So, it is important to manage reservations very efficiently, based on modern technologies such as machine learning

In this project, I tried to predict the possibility of a booking for a hotel based on different factors and tried to predict if they need special requests based on different features

## Data Overview

# ABOUT DATA

To provide a solution to this problem, I did some analysis using the hotel_booking dataset from kaggle with 119.000 rows and 31 columns

# The features contained in this dataset:

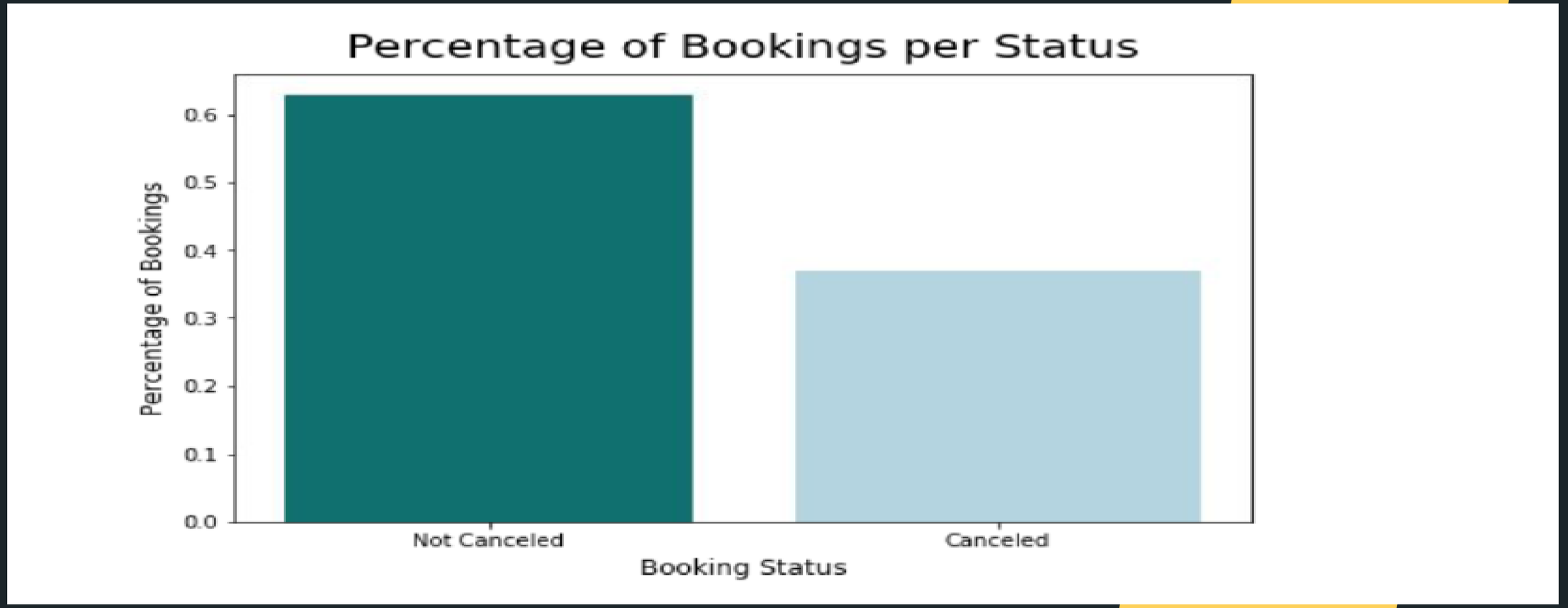
- HOTEL
- IS CANCELED
- LEAD TIME
- ARRIVAL DATE YEAR
- ARRIVAL DATE MONTH
- ARRIVAL DATE WEEK NUMBER
- ARRIVAL DATE DAY OF MONTH
- STAYS IN WEEKEND NIGHTS

- STAYS IN WEEK NIGHTS
- ADULTS
- CHILDRE
- N BABIES
- MEAL
- COUNTRY
- MARKET SEGMENT
- DISTRIBUTIO
  N CHANNEL

- IS REPEATED GUEST
- PREVIOUS CANCELLATIONS
- PREVIOUS BOOKINGS NOT CANCELED
- RESERVED ROOM TYPE
- ASSIGNED ROOM TYPE
- BOOKING CHANGES
- DEPOSIT TYPE
- AGENT

- COMPANY
- DAYS IN WAITING LIST
- CUSTOMER TYPE
- ADR
- REQUIRED CAR PARKING SPACES
- RESERVATION STATUS
- RESERVATIO
  N STATUS
  DATE

# Distribution of target variable



# Exploratory Data Analysis

## Key Findings

- Columns “reservation status” and “reservation status date” are updated after the booking has been cancelled

- Columns “company” and “agent” have a lot of missing values

Solution: Drop these columns

(handled by pipeline)

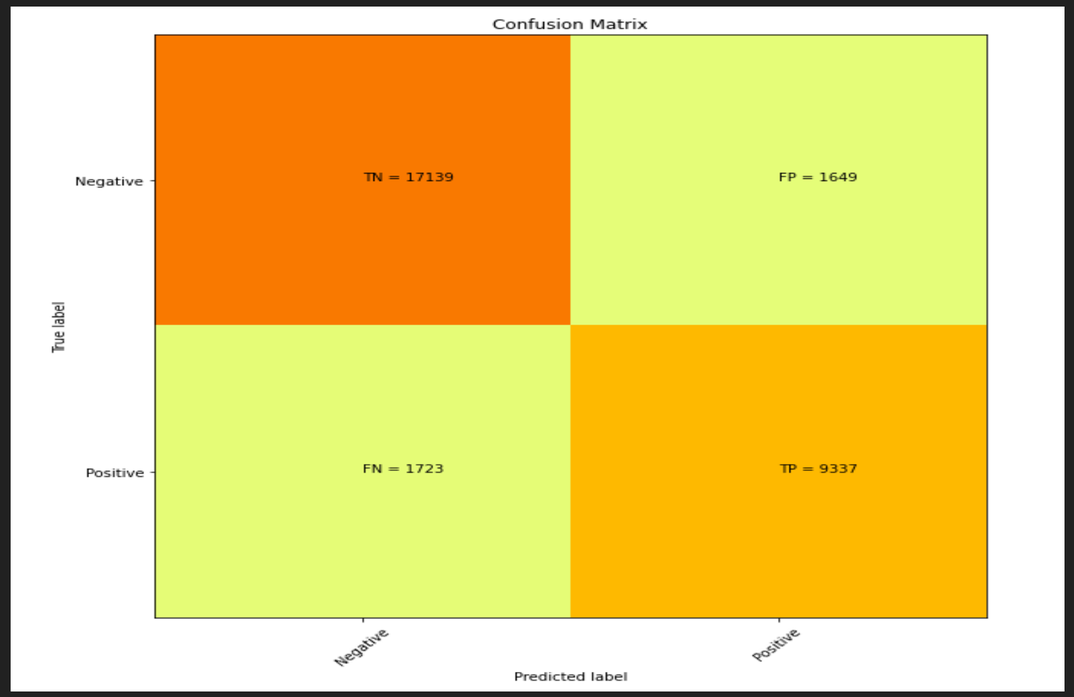
# RESULTSUMMARY

Logistic Regression Model: 0.82

Random Forest Model : 0.84:

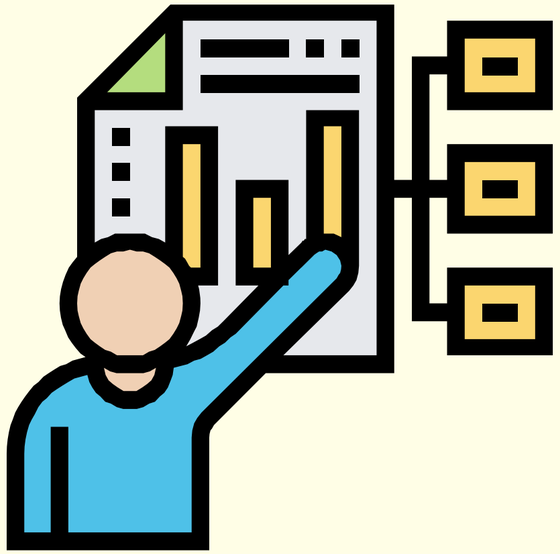
Xgboost Model: 0.89

# Xgboost
# Confusion Matrix




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.91 | 0.91 | 0.91 | 18788 |
| 1 | 0.85 | 0.84 | 0.85 | 11060 |
| accuracy | | | 0.89 | 29848 |
| macro avg | 0.88 | 0.88 | 0.88 | 29848 |
| weighted avg | 0.89 | 0.89 | 0.89 | 29848 |

# RESULT SUMMARY



## CONCLUSION

The XgBoost model proved to be a best choice to predict whether a hotel booking would be cancelled given the current and prior booking information about the type of hotel, room, customer, stay, payment status, etc..

# Thank You!

-END-