PREDICTING AND REDUCING THE IMPACT OF HOTEL RESERVATIONS CANCELLATION RATE

January
2023

IRON HACK

AGENDA

1. INTRODUCTION
2. HOW TO SOLVE OUR PROBLEM
3. MATERIALS AND METHOD
4. RESULTS
5. CONCLUSIONS

INTRODUCTION

WE WANT TO MINIMIZE THE NEGATIVE IMPACT OF CANCELLATIONS IN OUR HOTELS

- MORE THAN 10 MILLION CANCELLATIONS PER YEAR IN PORTUGAL.
- **HYPOTHETICAL SITUATION:** THE PROPERTY IS COMPLAINING ABOUT THE BIG AMOUNT OF CANCELLATIONS THEY ARE RECEIVING, THEY CLAIM THAT THE AMOUNT OF CANCELLATIONS IS GREATER THAN 25% FOR BOTH HOTELS COMBINED.

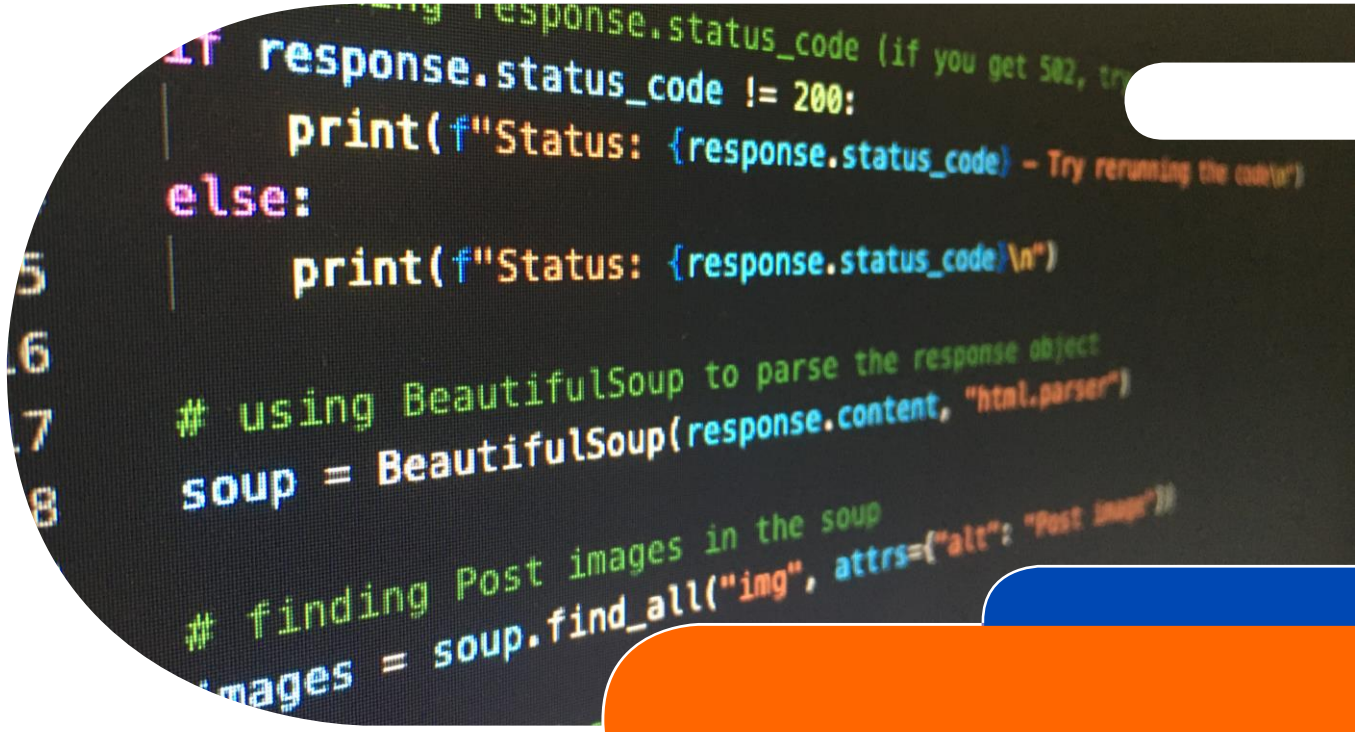
HOW TO SOLVE OUR PROBLEM

WE ARE GOING TO MODIFY OUR CANCELLATION POLICIES THANKS TO DATA ANALYSIS

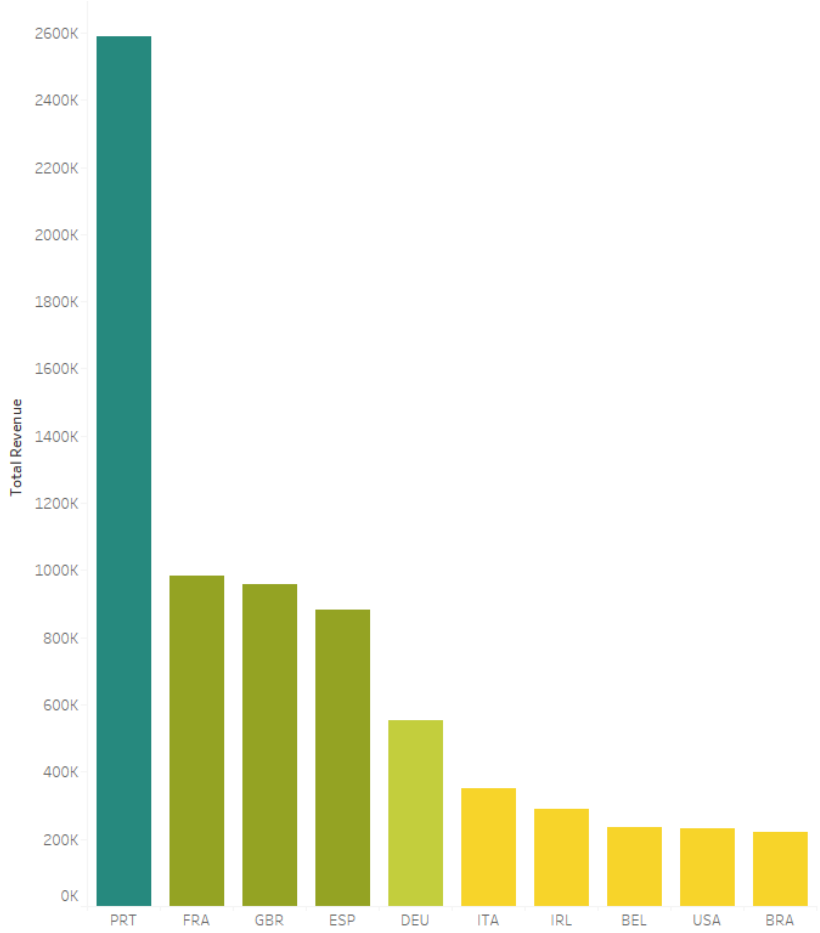
- ANALYZING DIFFERENT PARAMETERS TO IDENTIFY TRENDS
- BUILDING MACHINE LEARNING MODELS TO TRY TO PREDICT RESERVATIONS CANCELLATION RATE

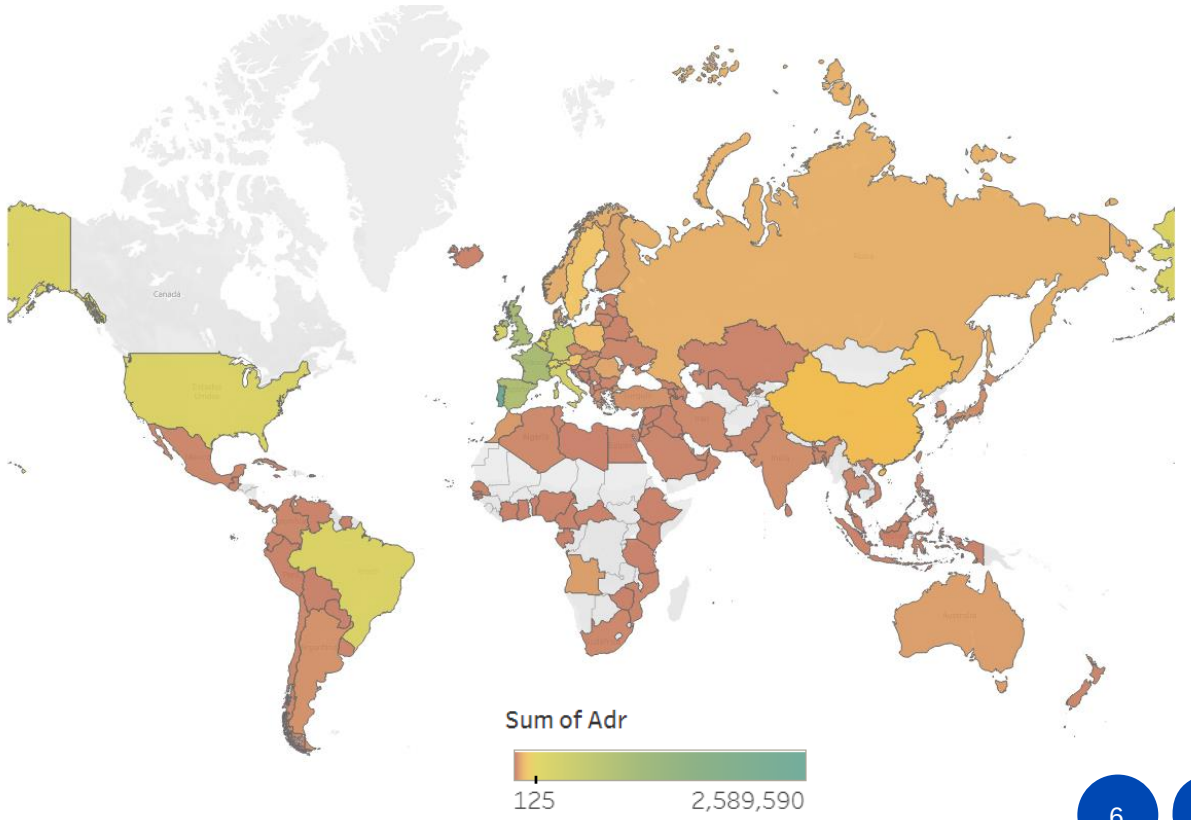
MATHERIALS & METHOD

- GET DATA
- EDA
- DATA CLEANING
- DATA VISUALIZATION
- HYPOTHESIS TESTING
- BUILDING MODELS



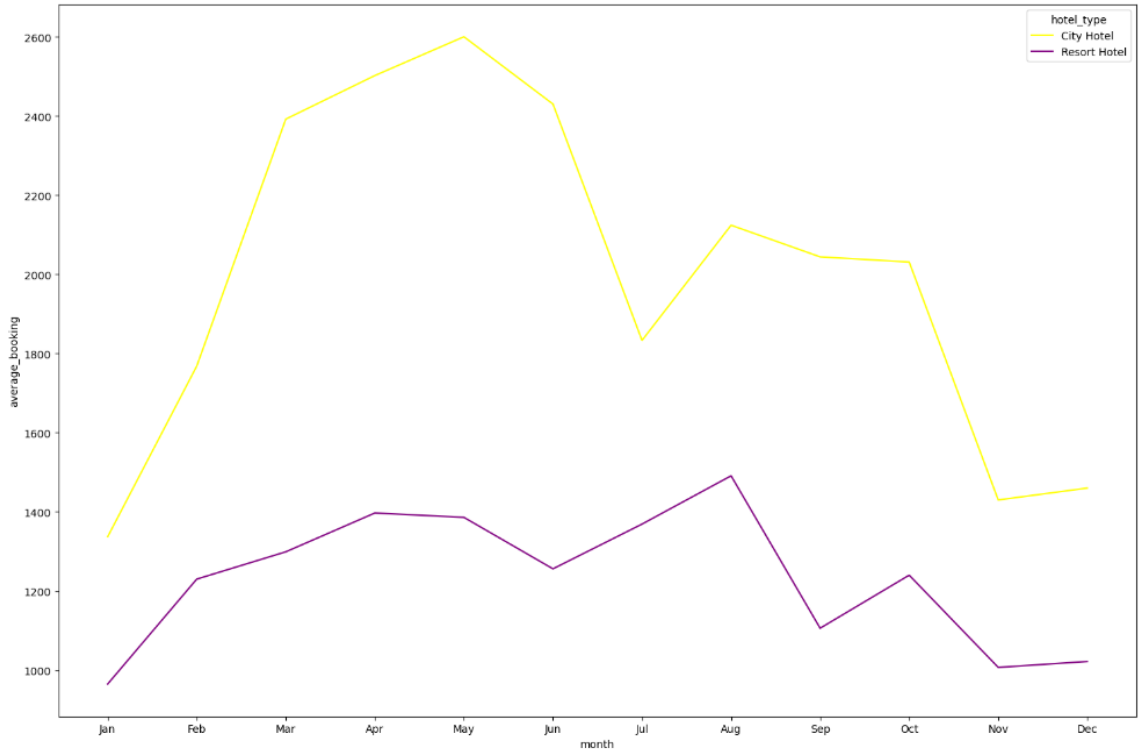
MAIN GUESTS' NATIONALITIES

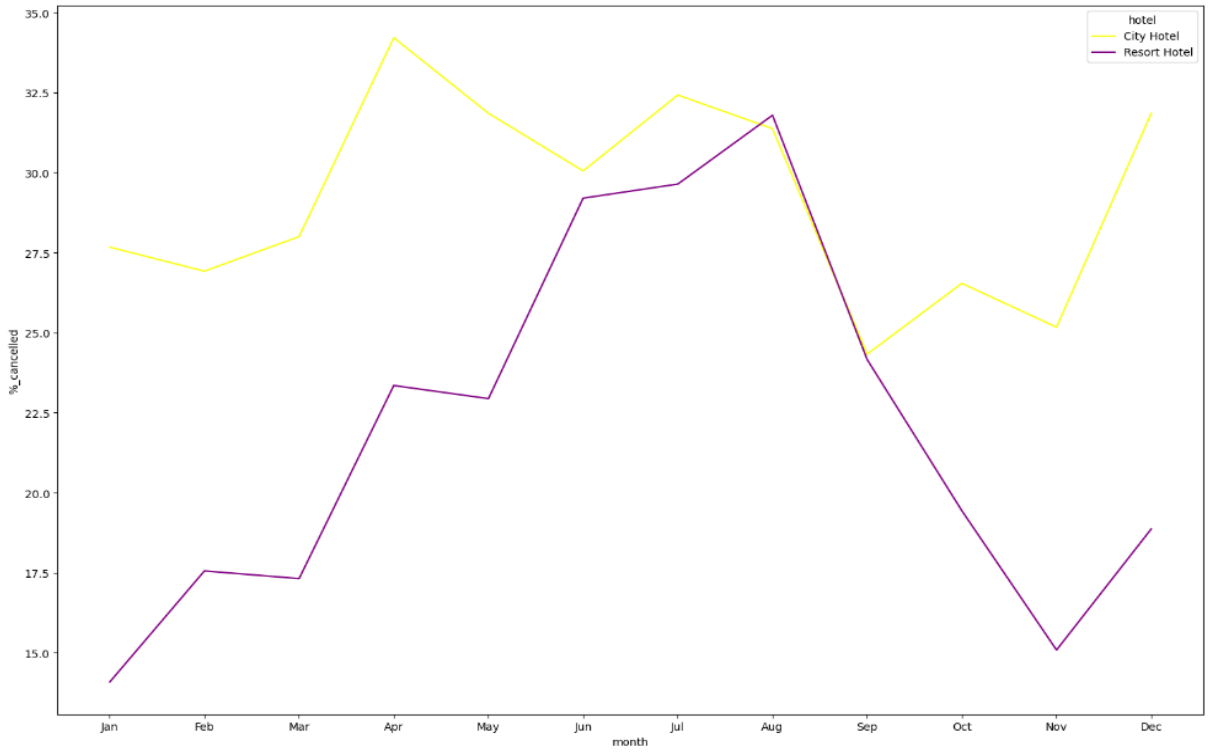




BOOKINGS

AVERAGE BOOKING AND CANCELLATION % PER HOTEL AND MONTH





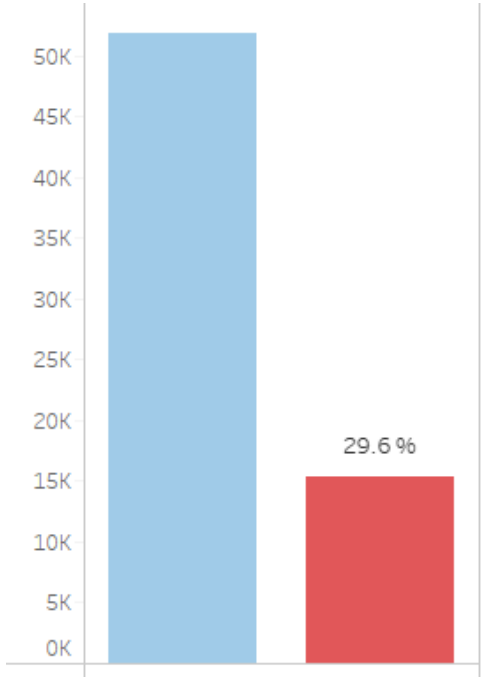
CITY HOTEL

RESORT HOTEL

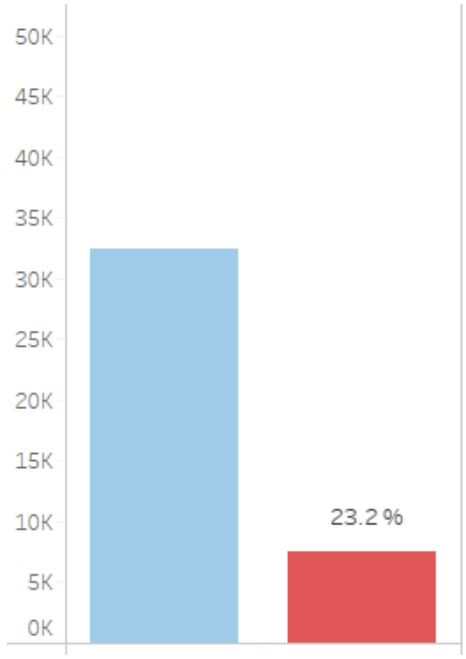
COMPARISON

PERCENTAGE OF CANCELLATIONS PER TYPE OF HOTEL

CITY HOTEL



RESORT HOTEL



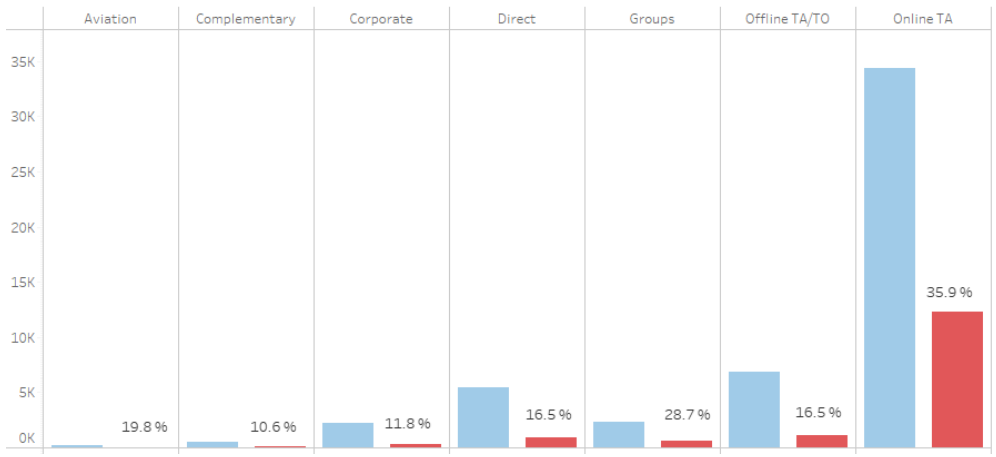
RESERVATIONS

CANCELLATIONS

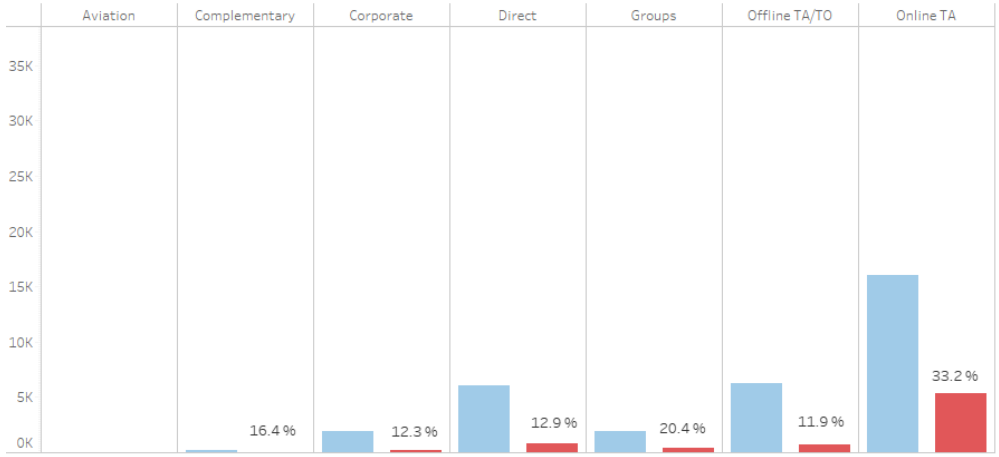
COMPARISON

PERCENTAGE OF CANCELLATIONS PER MARKET SEGMENT

CITY HOTEL



RESORT HOTEL

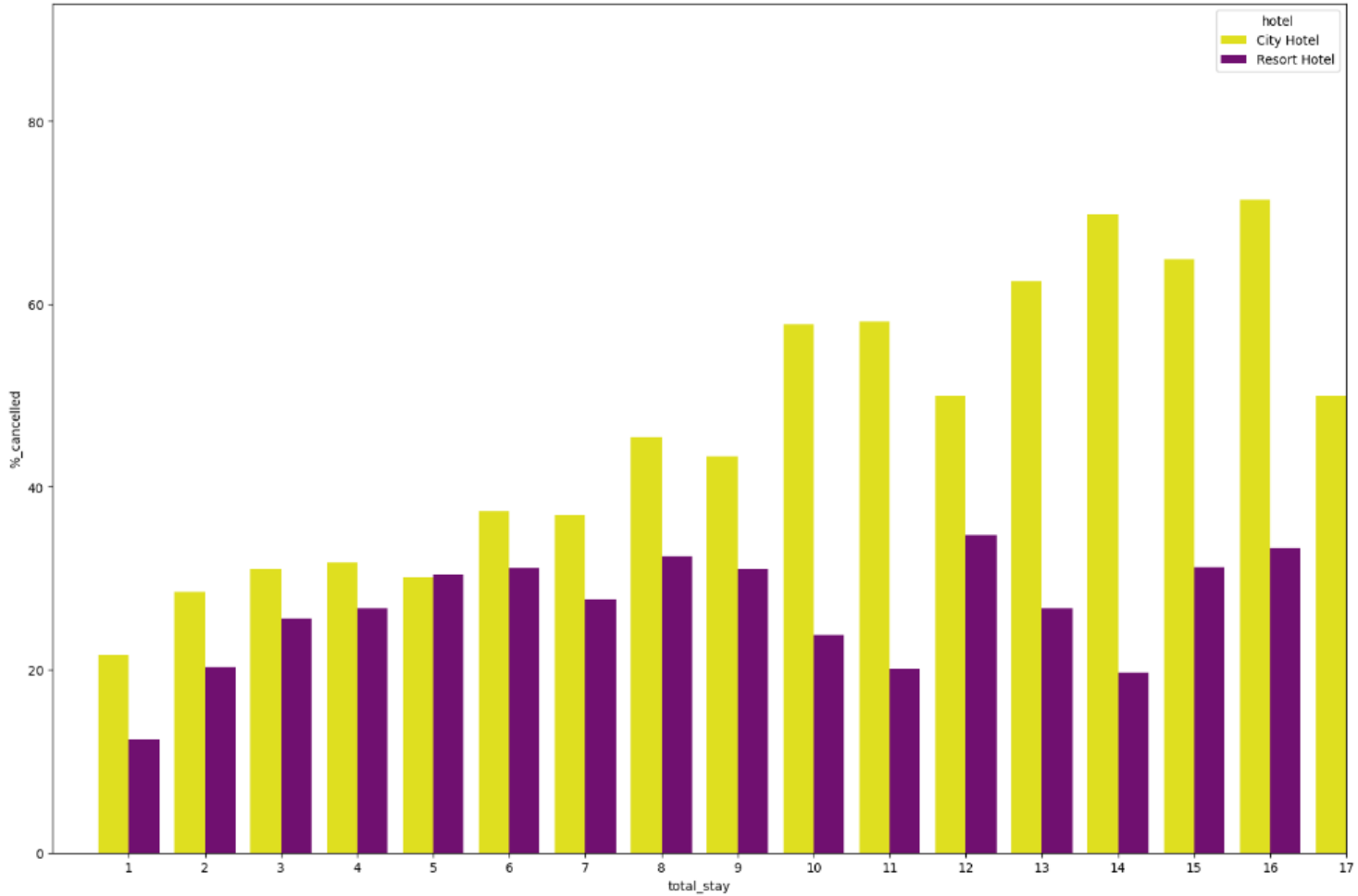


RESERVATIONS

CANCELLATIONS

CANCELLATIONS

% OF CANCELLATIONS VS LENGHT OF STAY

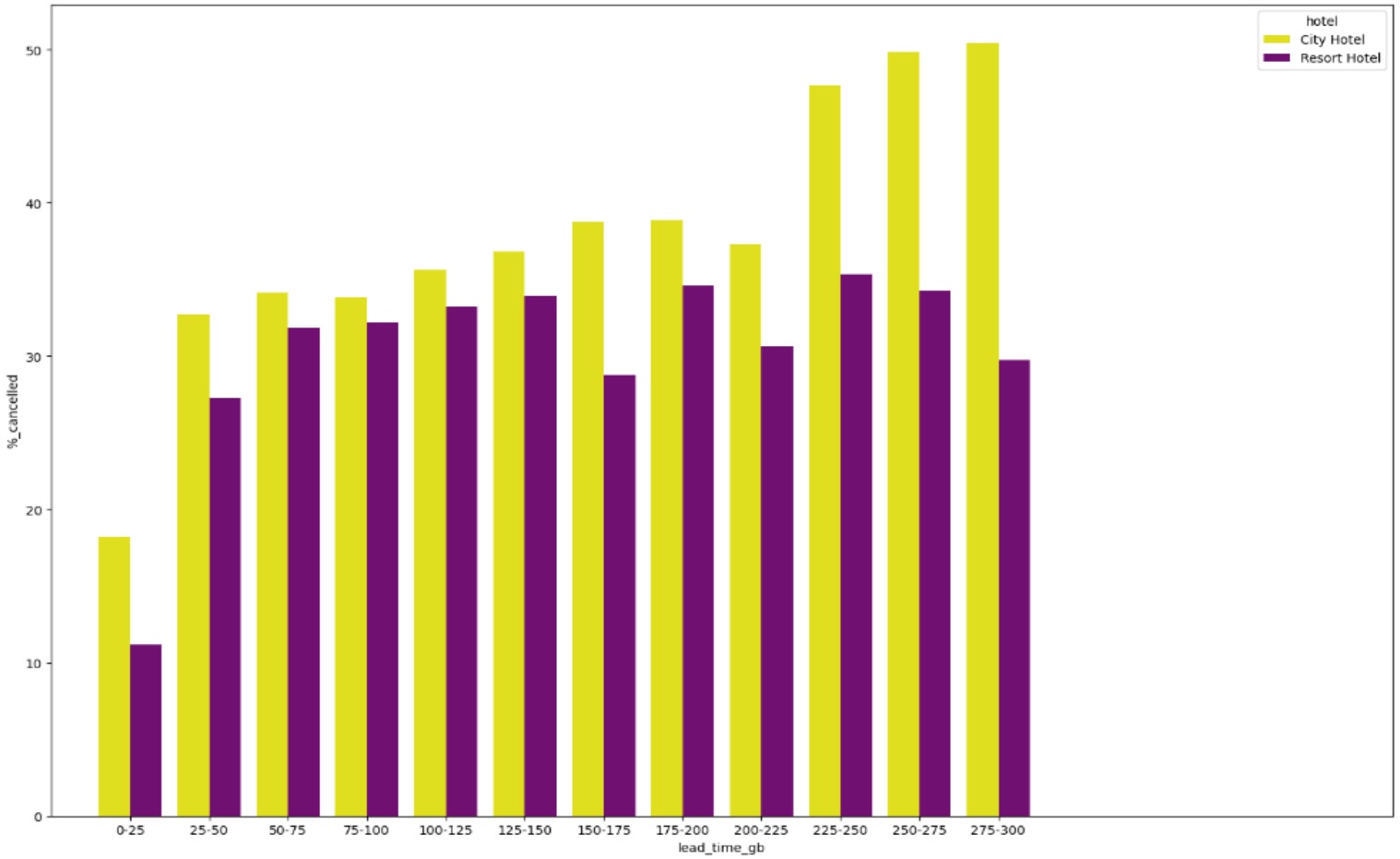


CITY HOTEL

RESORT HOTEL

CANCELLATIONS

% OF CANCELLATIONS VS LEAD TIME



CITY HOTEL

RESORT HOTEL

RESULTS

```
# Choosing between Logistic Regression and MLPClassifier:

# Logistic Regression scores: 0.7423891757165747

# precision-->  0.5199477731031481
# recall-->  0.7764298093587522
# f1-->  0.6228169258840908

# Logistic Regression confusion matrix

# [8926, 3309]
# [1032, 3584]

# MLPClassifier scores: 0.81087175835262

# precision-->  0.6278861643100053
# recall-->  0.7599653379549394
# f1-->  0.687640889934333

# MLPClassifier confusion matrix

# [10156,  2079]
# [ 1108,  3508]
```

CONCLUSIONS

1. Both hotels have more guests and cancellation % during high seasons. The peak season for the City Hotel goes from March to May and for the Resort Hotel from June to August. Both hotels lower season is Winter
2. The percentage of cancellations is not related to the country of origin of our guests
3. The percentage of cancellations is higher in the City Hotel than in the Resort
4. The company should focus on growing other market segments rather than Online TA
5. The longer the stay is, the higher is the probability of a cancellation. This trend is more clear in the City Hotel.
6. The longer the lead time is, the higher is the probability of a cancellation. Again the trend is steeper in the City Hotel.
7. Accurate Machine Learning Model available in order to perform predictions.

```
print('Thank you for your attention!')
```

```
Thank you for your attention!
```

Christian M.
García Delgado

IRON HACK