HOTEL RESERVATION ANALYSIS

Mini Project 3 - March 2023

Ryan Stamp

CONTENTS

WHAT WE'LL DISCUSS

Industry Overview

Why are our guests cancelling?

Data pipeline

Facts & Figures

What did we discover?

Modelling

What does this tell us

Next steps

References

INDUSTRY OVERVIEW

-As of June 2022 there are 279 hotels and resorts in New Zealand

-The highest costs for Hotels and Resorts are: Wages (29.6%), Purchases (23.5%), and Rent & Utilities (8.9%)

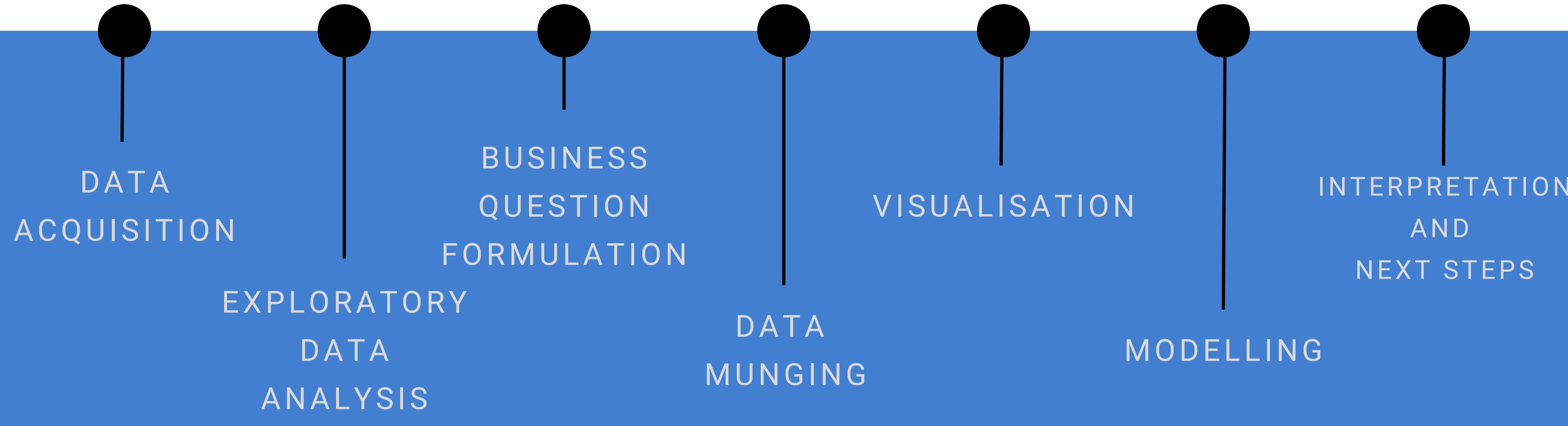
-The cost of cancellations: loss of income, impact of no-shows, lower revenue per average room when selling last minute, direct operating costs

WHY ARE OUR GUESTS CANCELLING?

Can we predict the booking status of future guests?

R Stamp 2023

DATA PIPELINE



R STAMP 2023

WHAT INFORMATION DO WE HAVE?

KAGGLE - HOTEL RESERVATON DATASET

Features used in the model:

- Number of people in booking
- Number of nights booked
- Type of meal plan booked
- Does the customer require a car parking space?
- Type of room reserved by the customer.
- Number of days between the date of booking and the arrival date
- Year of arrival date
- Market segment designation.
- Is the customer a repeated guest?
- Number of previous bookings that were canceled by the customer
- Number of previous bookings not canceled by the customer
- Average price per day of the reservation (prices of the rooms are dynamic)
- Total number of special requests made by the customer
- Flag indicating if the booking was canceled or not (also our target variable)

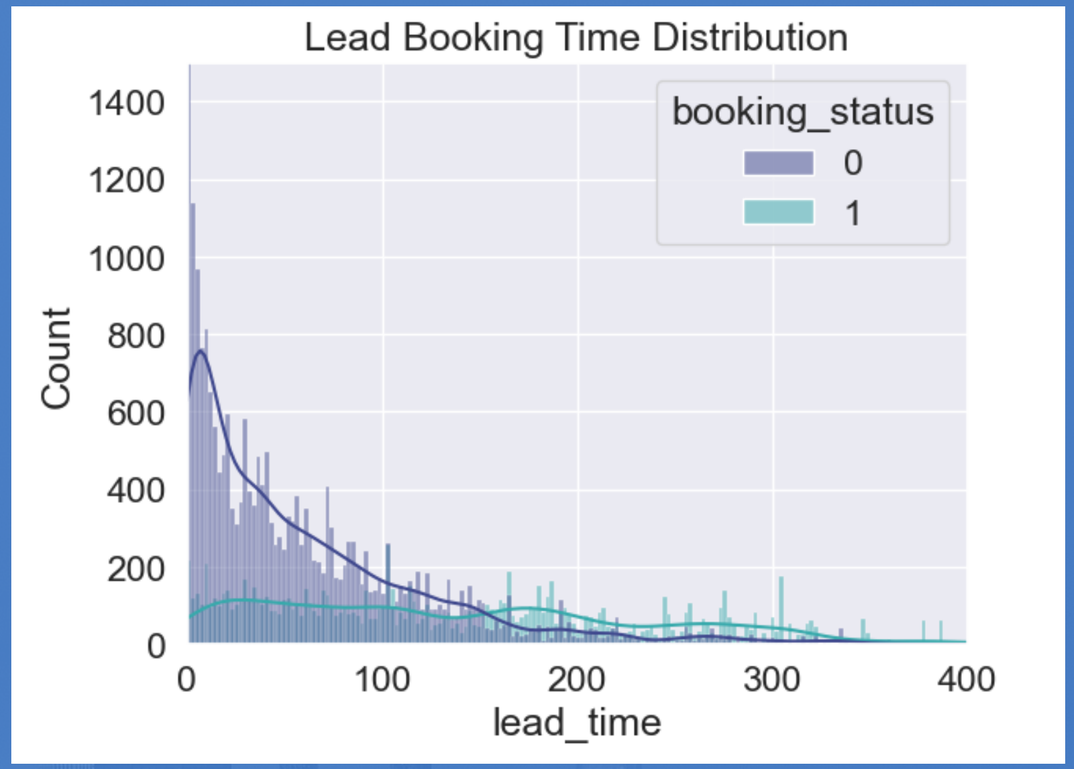
Features not used in the model:

- Booking ID
- Month of arrival date
- Date of the month

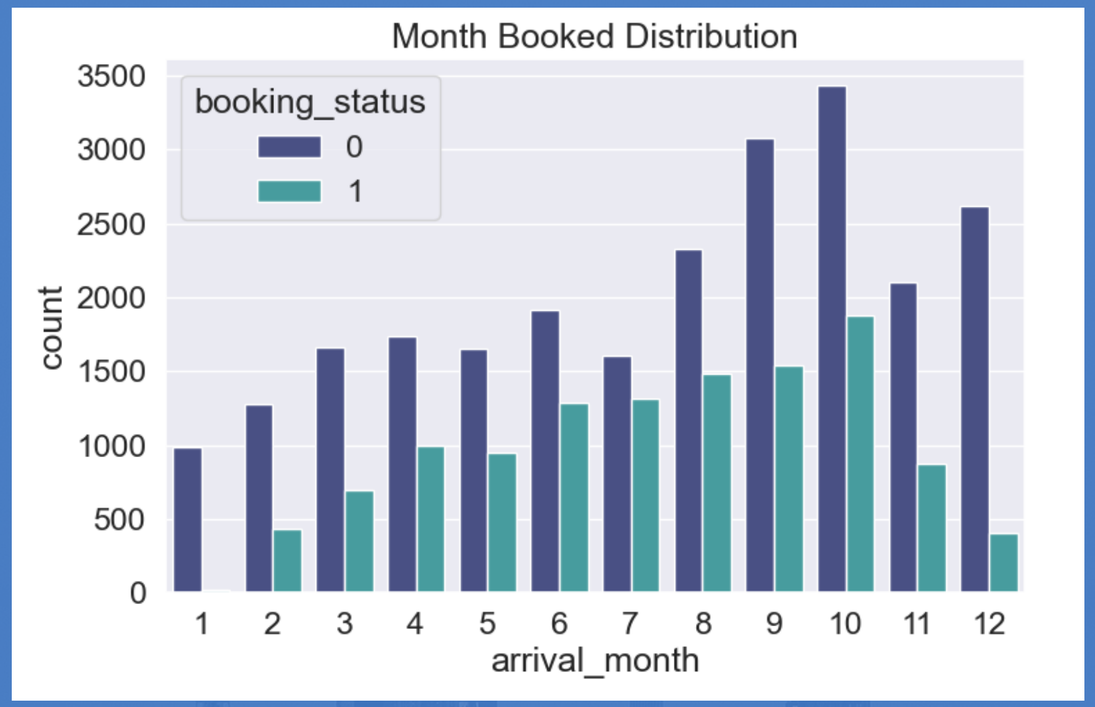
WHAT DID WE FIND?

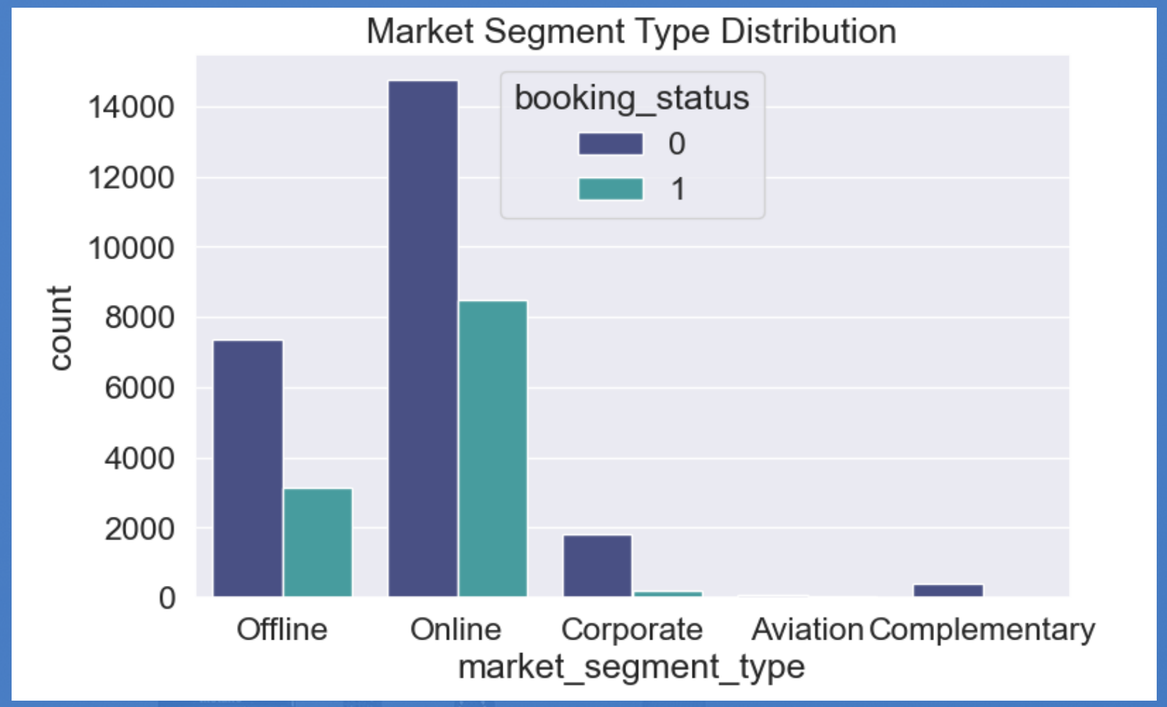
- Total number of bookings = 36,275
- Average lead time = 85 days
- Average price per room = 103€/Night
- Percentage of booking cancellations = 32%
- Strongest correlations with 'Booking Status' are Lead Time, Number of Special Requests and Average Price Per Room

WHAT DID WE FIND?



WHAT DID WE FIND?





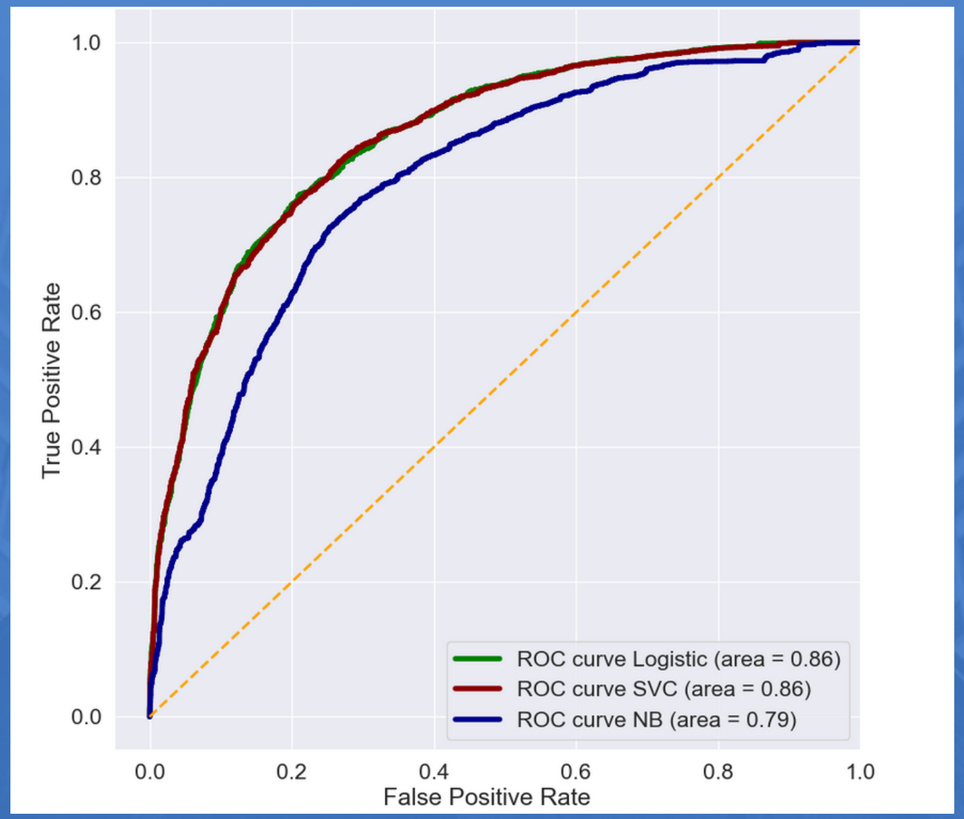
MODELLING & INTERPRETATION

Model	Precision	Recall	Accuracy
Logistic Regression	0.73	0.62	0.8
Support Vector Machine	0.74	0.6	0.8
Naive Bayes	0.35	0.97	0.4

Logistic Regression Model is recommended due to the highest recall score

MODELLING

Receiver Operator Characteristic Curve - Cancelled Reservations



SUMMARY AND NEXT STEPS

The strongest contributors to bookings being cancelled are:

-Longer Booking Lead time

-Higher Number of Special Requests

-Higher Average Price

>Find length of time between cancellations and booking dates

>Gather customer reasons for cancelling

>Further feature engineering to improve the model results

>Gather external data where possible which may be influencing cancellation rates

THANK YOU

ANY QUESTIONS?

R STAMP 2023

REFERENCES

HTTPS://WWW.KAGGLE.COM/DATASETS/AHSAN81/HOTEL-RESERVATIONS-CLASSIFICATION-DATASET

HTTPS://WWW.IBISWORLD.COM/NZ/NUMBER-OF-BUSINESSES/HOTELS-RESORTS/1811/

HTTPS://WWW.MIRAI.COM/BLOG/CANCELLATIONS-SHOOTING-UP-IMPLICATIONS-COSTS-AND-HOW-TO-REDUCE-THEM/