HOTEL RESERVATIONS

Scarlett Valentin

08/01/24

TABLE OF CONTENTS

TABLE OF CONTENTS

BUSINESS PROBLEM

Goal: Combat hotel cancelations and no-shows.

How:

- Create a model to predict the number of hotel booking cancelations.
- Evaluate the model to assess the efficacy in using it to predict cancelations.
- Determine the 4 most important factors that contribute to a cancelation.

TABLE OF CONTENTS

DATA UNDERSTANDING

Hotel Reservations Dataset:

- 36,275 entries of unique bookings ranging from 2017 to 2018

- Includes 17 attributes of customer reservation details

- Target column: Booking Status

TABLE OF CONTENTS

DATA PREPARATION

1

Explore data

2

Convert values to numeric

3

Train-test split

5

Standard Scaler

4
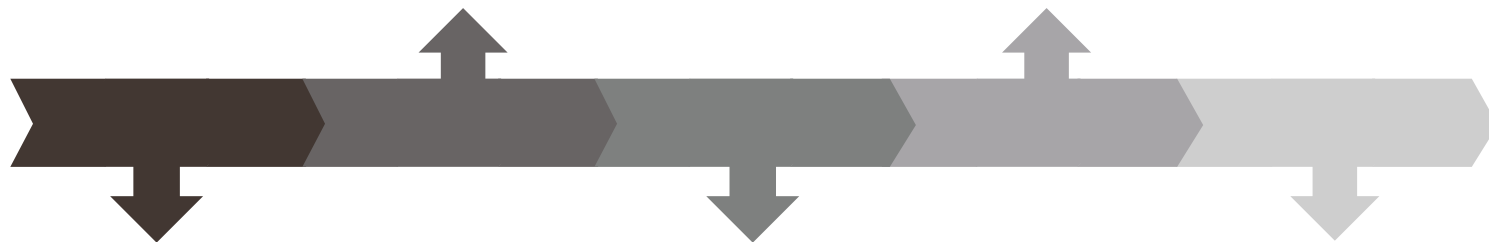
One Hot Encoder
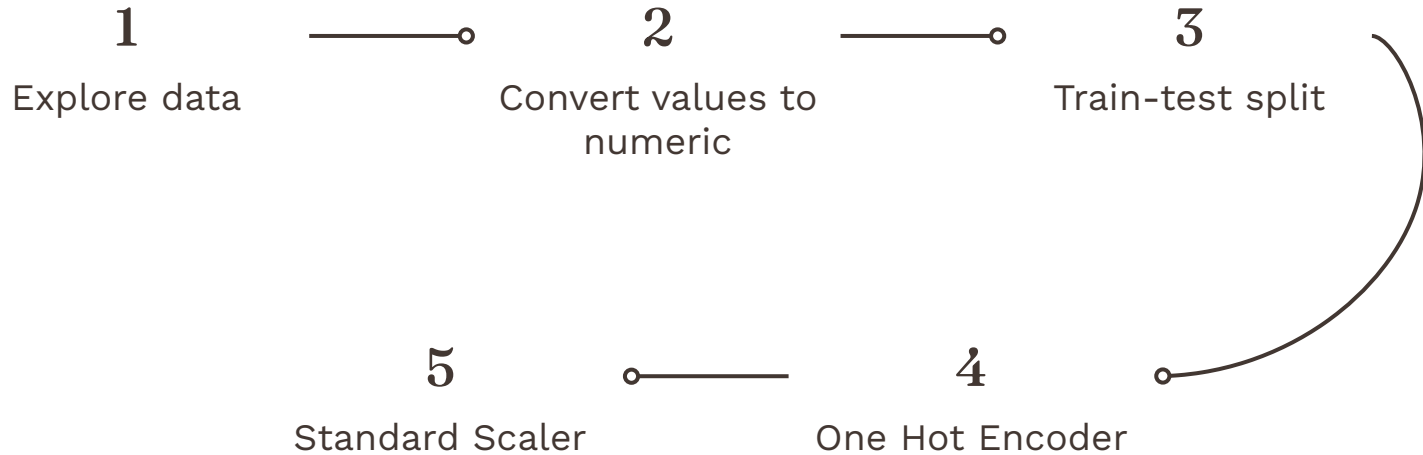
TABLE OF CONTENTS

MODELING

Significant Features

Removed features
with p-value > 0.05

Final Model

max_depth = 7

Logistic Regression

Used for binary
outcomes

Decision Tree Classifier

Hyperparameter Tuning & Pruning:
- Maximum depth
- Minimum samples leaf
- Minimum leaf sample size
- Maximum features

Feature Importance

1. Lead time
2. Online booking
3. Average price per room
4. Number of special requests

TABLE OF CONTENTS

EVALUATION

85% ACCURACY

86% PRECISION

0.8 AUC

CONFUSION MATRIX
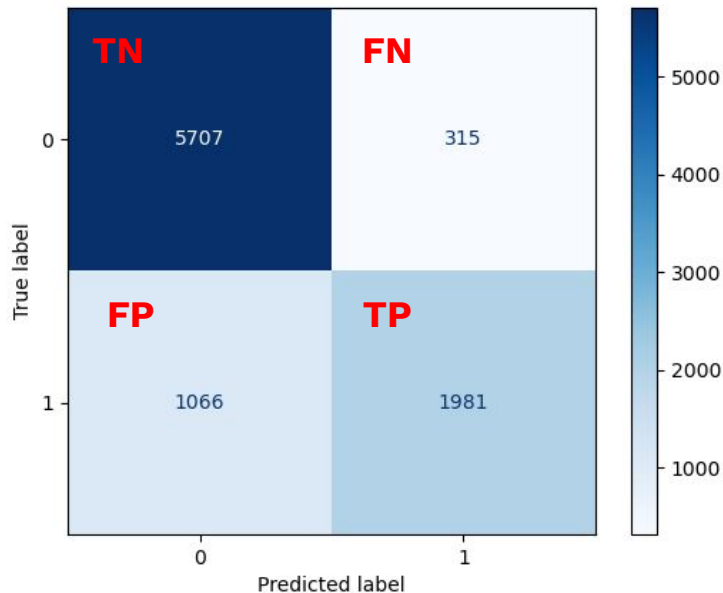
315 retained bookings that were predicted to cancel

TABLE OF CONTENTS
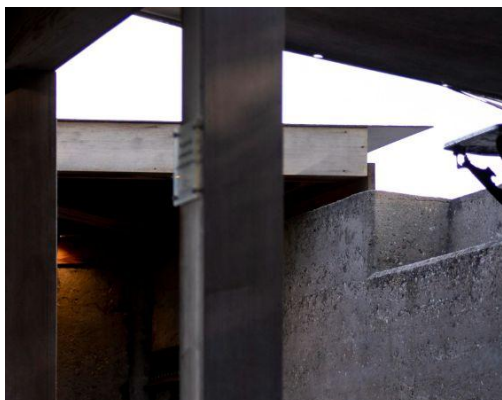
NEXT STEPS

1. Experiment with different classifier models to **increase the precision score** of the predictions.

2. **Gather recent data** from our guest reservations to continue tuning and testing the model.

3. Further **investigate the top 4 factors** of hotel cancelations.

THANK YOU!

Do you have any questions?

Contact Information:
Email: scarlettmvalentin@gmail.com
GitHub: @scarlettvalentin
LinkedIn: linkedin.com/in/scarlett-valentin/