



# HOTEL BOOKING DEMAND

---

SAFA ALSAFARI

# OVERVIEW & OBJECTIVES



---

## PROBLEM STATEMENT

Building a machine learning model that classify booking statuses accurately can help hotels plan for:

- ❖ Refund policies
- ❖ Staffing schedules
- ❖ Targeting customers with offers and discounts



## DATASET

The dataset consists of 119,390 observations with 32 features.



	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_
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	

5 rows × 32 columns



# EXPLANATORY DATA ANALYSIS



# DATA CLEANING

## Checking features Types

- Features with Incorrect Types:4
- Handling techniques:
  - Change to object
  - Change to integer

- Features with Missing Values:4
- Handling techniques:
  - Column dropping
  - Rows dropping
  - Imputing with mean and mode

## Exploring & Handling the Missing values

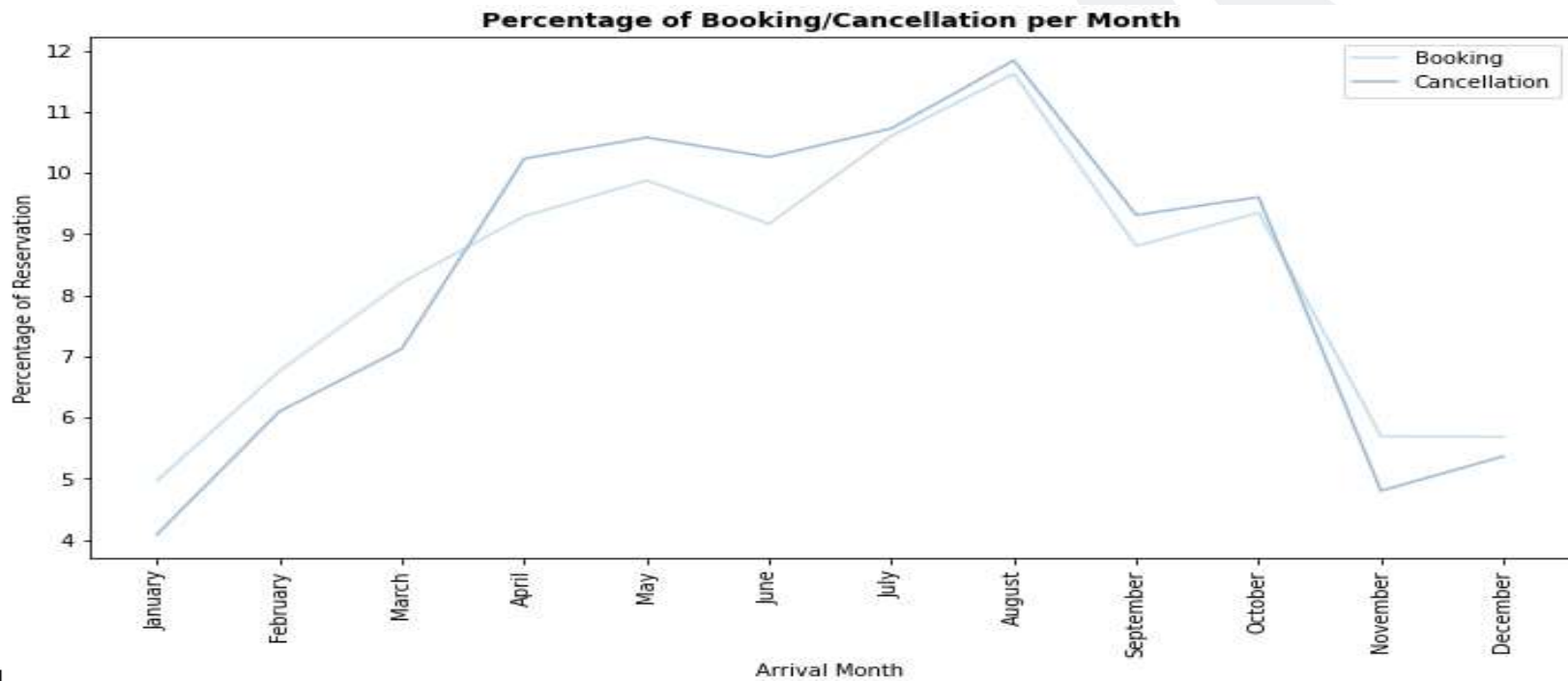
## HANDLING OUTLIERS

Features with outlier: 1

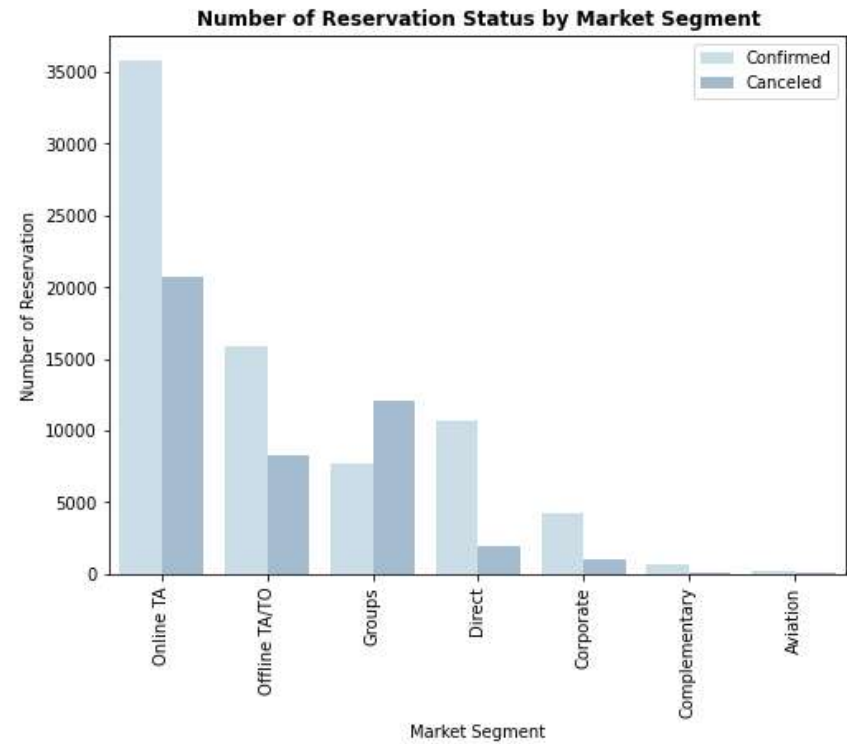
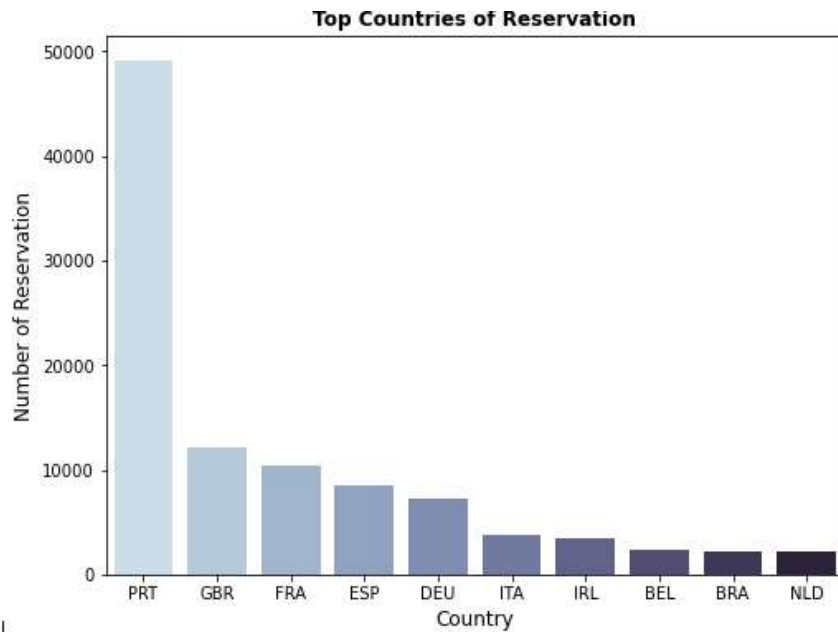
Handling techniques:

- Dropping of rows with outlier

# DATA EXPLORATION

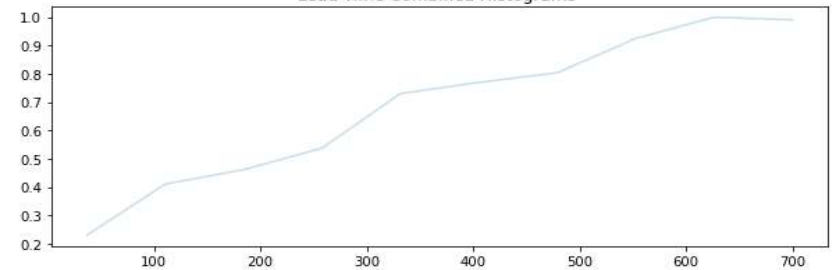
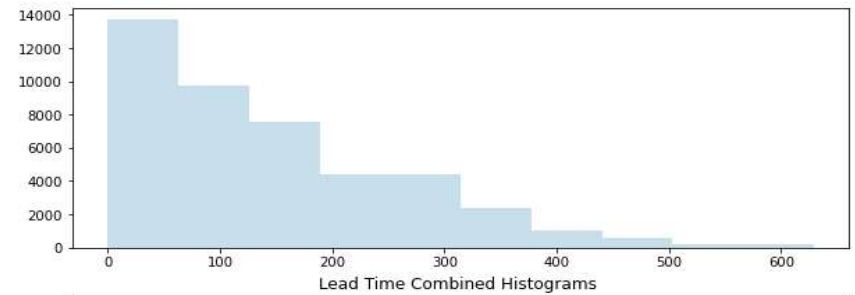
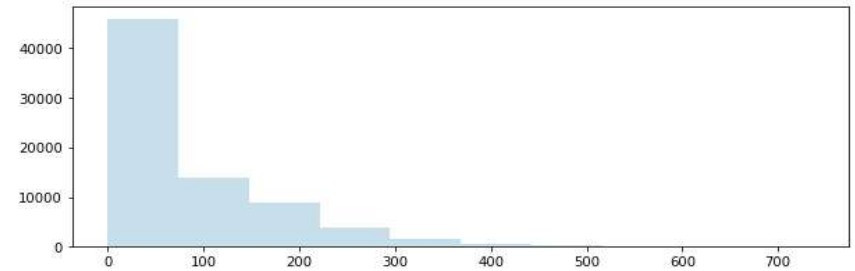
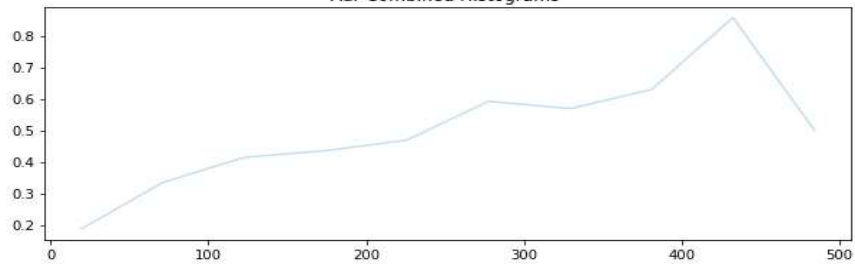
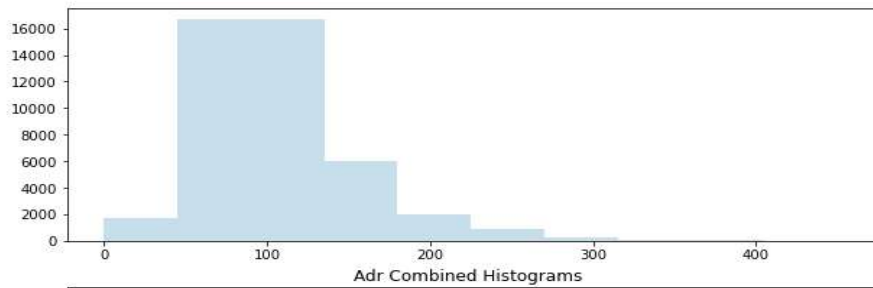
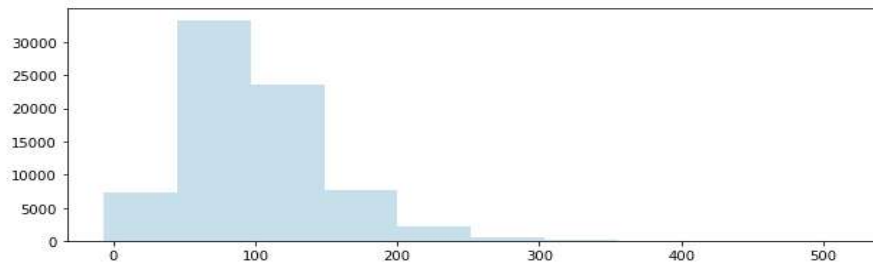


# DATA EXPLORATION





# DATA EXPLORATION

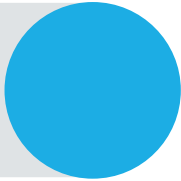


# MODEL BUILDING AND EVALUATION



# FEATURE ENGINEERING AND SELECTION

## FEATURES ENGINEERING



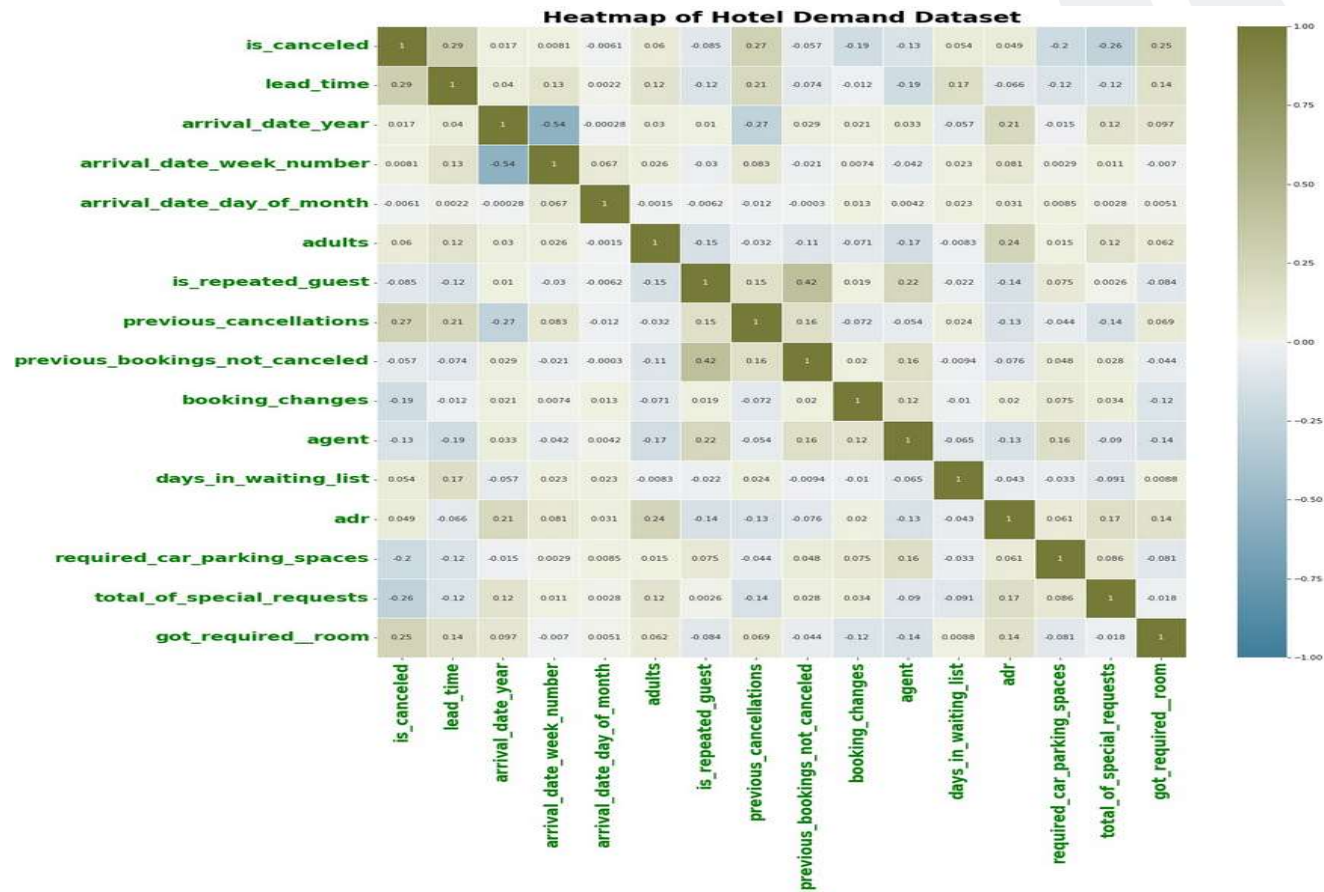
Adding feature  
Changing feature  
Numerical Features  
Scaling  
Encoding Categorical  
Features



Select feature based  
on their importance

## FEATURES SELECTION

# FEATURE CORRELATION MATRIX



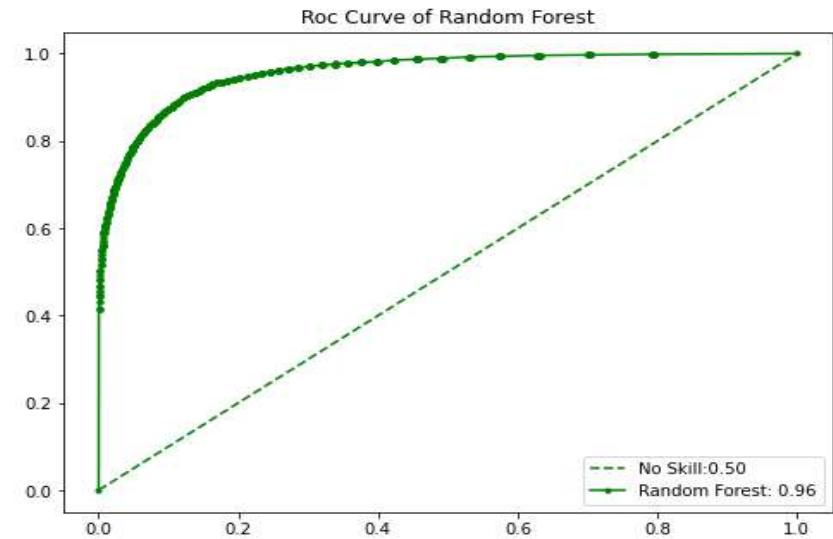
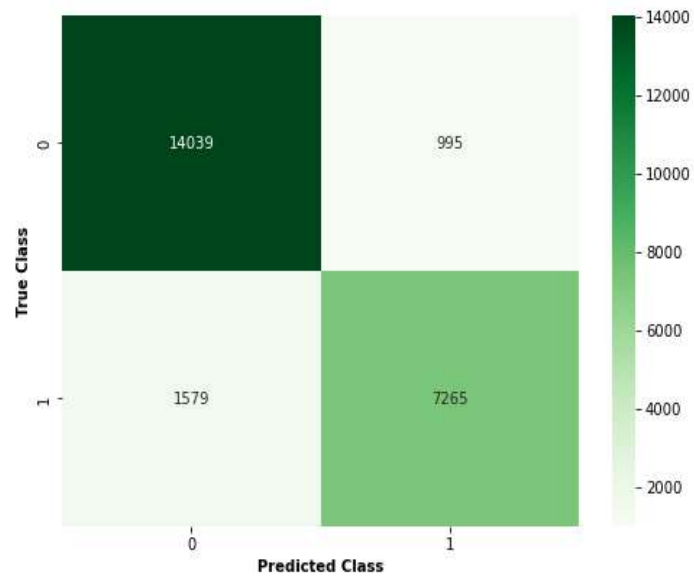
# MODEL SELECTION

Model	Precision	Recall	F-Macro (Cross Validation)	F-Macro (Holdout)
Logistic Regression (LR)	0.81	0.68	0.80	0.80
Naïve Base (NB)	0.80	0.54	0.74	0.75
K Nearest Neighbor (KNN)	0.87	0.68	0.82	0.83
Support Vector Machine (SVM)	0.82	0.67	0.80	0.80
Random Forest (RF)	0.88	0.81	0.88	0.89

Model Performance Using 5 Fold Cross Validation & Holdout



# RANDOM FOREST ANALYSIS



# CONCLUSION



# CONCLUSION

- Dataset requires cleaning and preparation
- The most important features are:
  - lead\_time
  - total\_of\_special\_requests
  - required\_car\_parking\_spaces
  - booking\_changes
- Best Model: Random Forest with F-Macro:0.89



# THANK YOU

---



[SBALSEFRI@GMAIL.COM](mailto:SBALSEFRI@GMAIL.COM)

