Hotel Booking Analysis

Robby Sutiawan

Contents

Hotel Overview

The dataset comes from hotel booking data in **Lisbon**, **Portugal**, from **July 1, 2015, to August 31, 2017**.

€17.5 M
Total Revenue

79k bookings
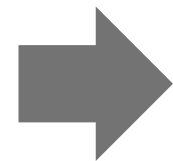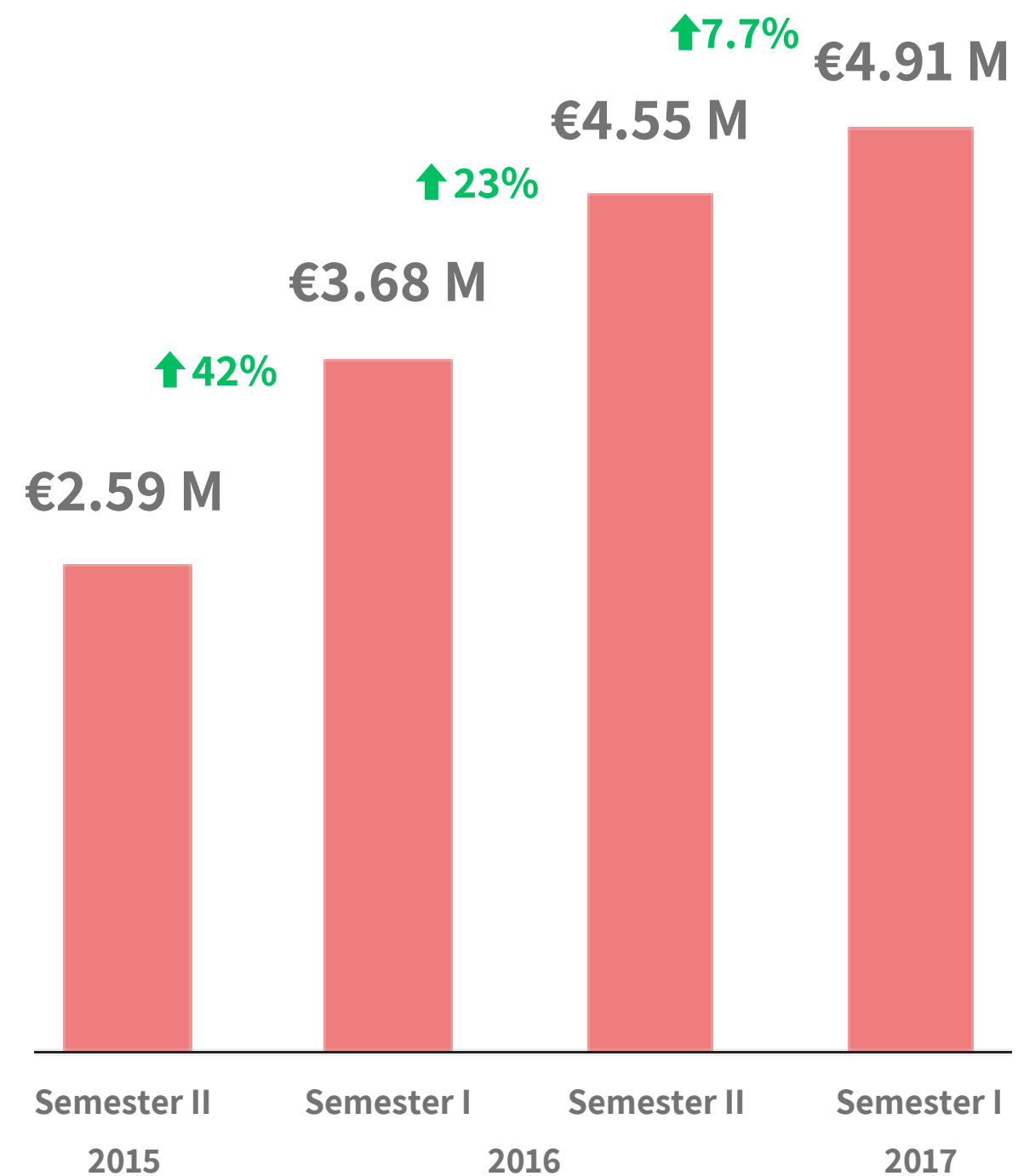
2.97 avg total night

€105.5 avg daily rate

154k total guests

167 countries

Problem Statement

Revenue Trend **Slowed Down** in the First Half of 2017

↑7.7%

€4.91 M

↑23%

€4.55 M

€3.68 M

↑42%

€2.59 M

Semester II	Semester I	Semester II	Semester I
2015	2016		2017

The number of bookings **declined** in the first half of 2017

S2 2015	**7.6K**
S1 2016	**10.7K**
S2 2016	**11.9K**
S1 2017	**11.5K**

↓ 3.2%

Out of all bookings, **41.8% were canceled**

No Canceled

58.2%

41.8%

Canceled

63% deposit
no refund
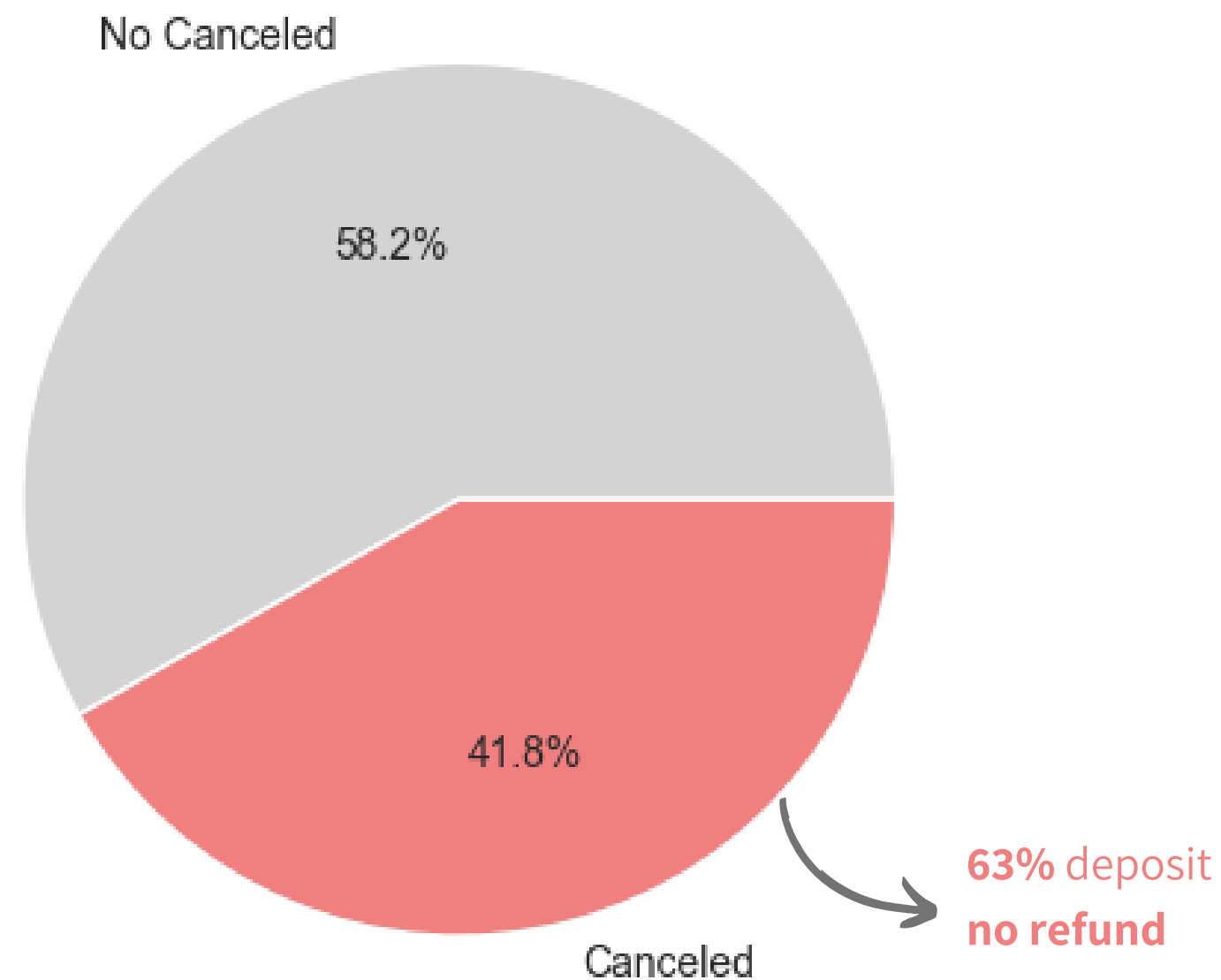
From these cancellations, the hotel **lost potential revenue** of approximately **€7.7 million** in the last 2 years!

Note: Guests who cancel their bookings with a **non-refundable** deposit type will still **generate revenue**.
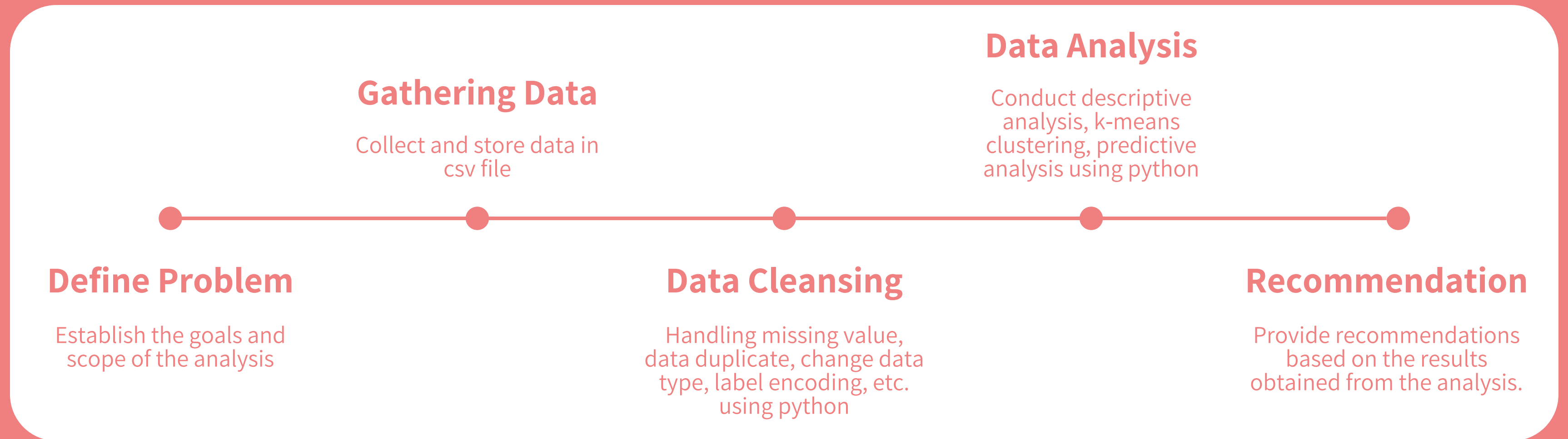
Bussiness Objective

How can we **optimize revenue?**

- **Pricing strategy**
- **Reduce the cancellation rate**

Methodology

**Gathering Data**

Collect and store data in csv file

Data Analysis

Conduct descriptive analysis, k-means clustering, predictive analysis using python

Define Problem

Establish the goals and scope of the analysis

Data Cleansing

Handling missing value, data duplicate, change data type, label encoding, etc. using python

Recommendation

Provide recommendations based on the results obtained from the analysis.

Data Analysis

The average daily rate tends to increase during peak months

Booking Number vs ADR

Booking Number & **ADR** increased

Spring, Summer, Autumn
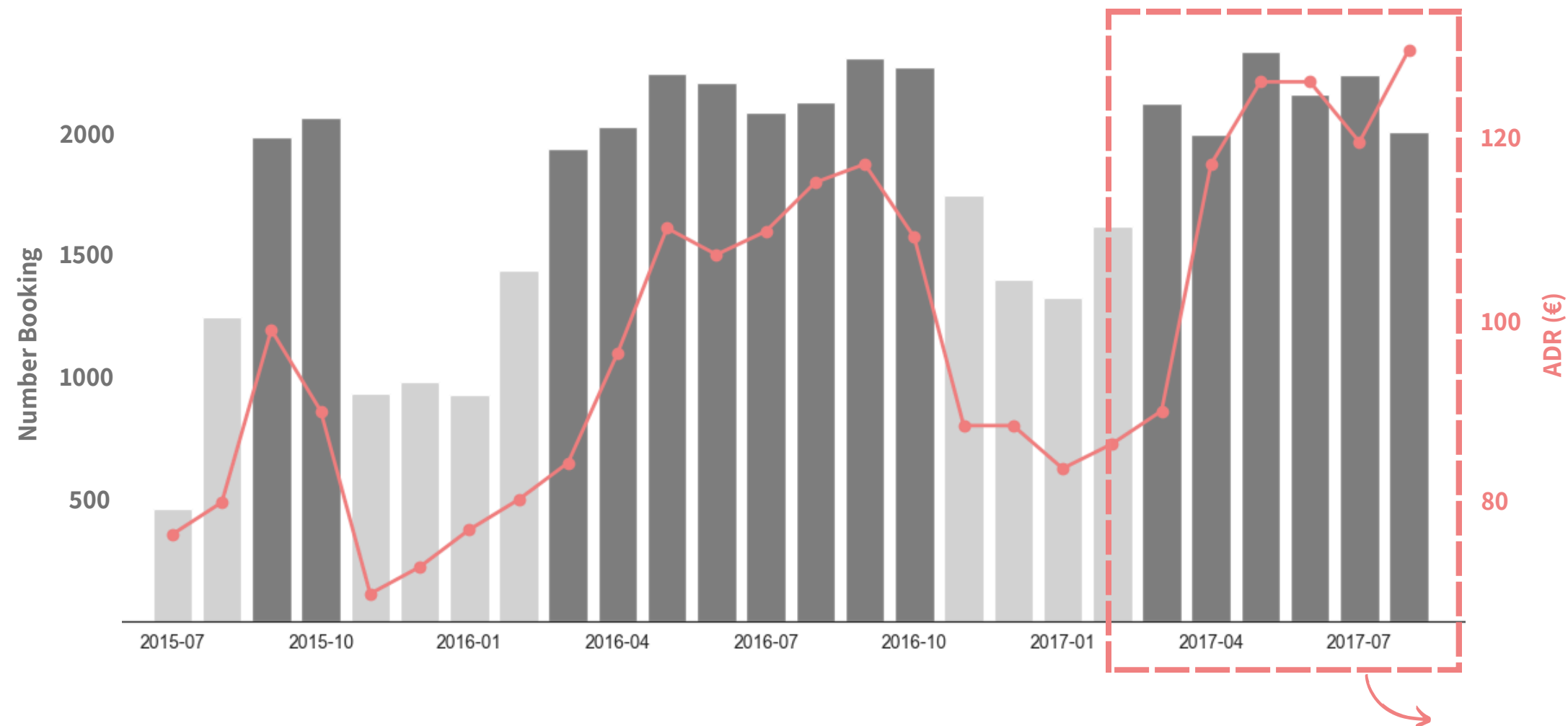
(Mar - Oct)

Rates have **sharply increased**, but this has **not been followed by booking growth** from the previous period.

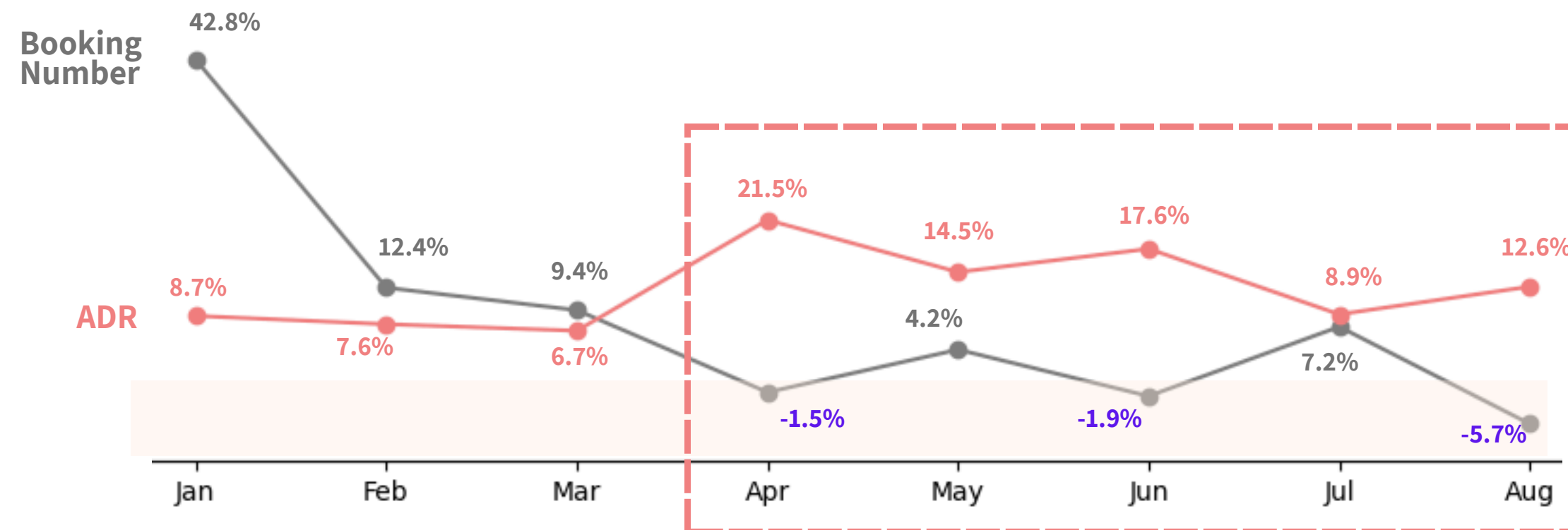# A price increase of over 10% leads to a decline in booking growth

Booking Number vs ADR (% Diff 2017 & 2016)

An increase in rates of **more than 10%** results in a **decrease in booking growth**, which in turn slows down revenue growth.
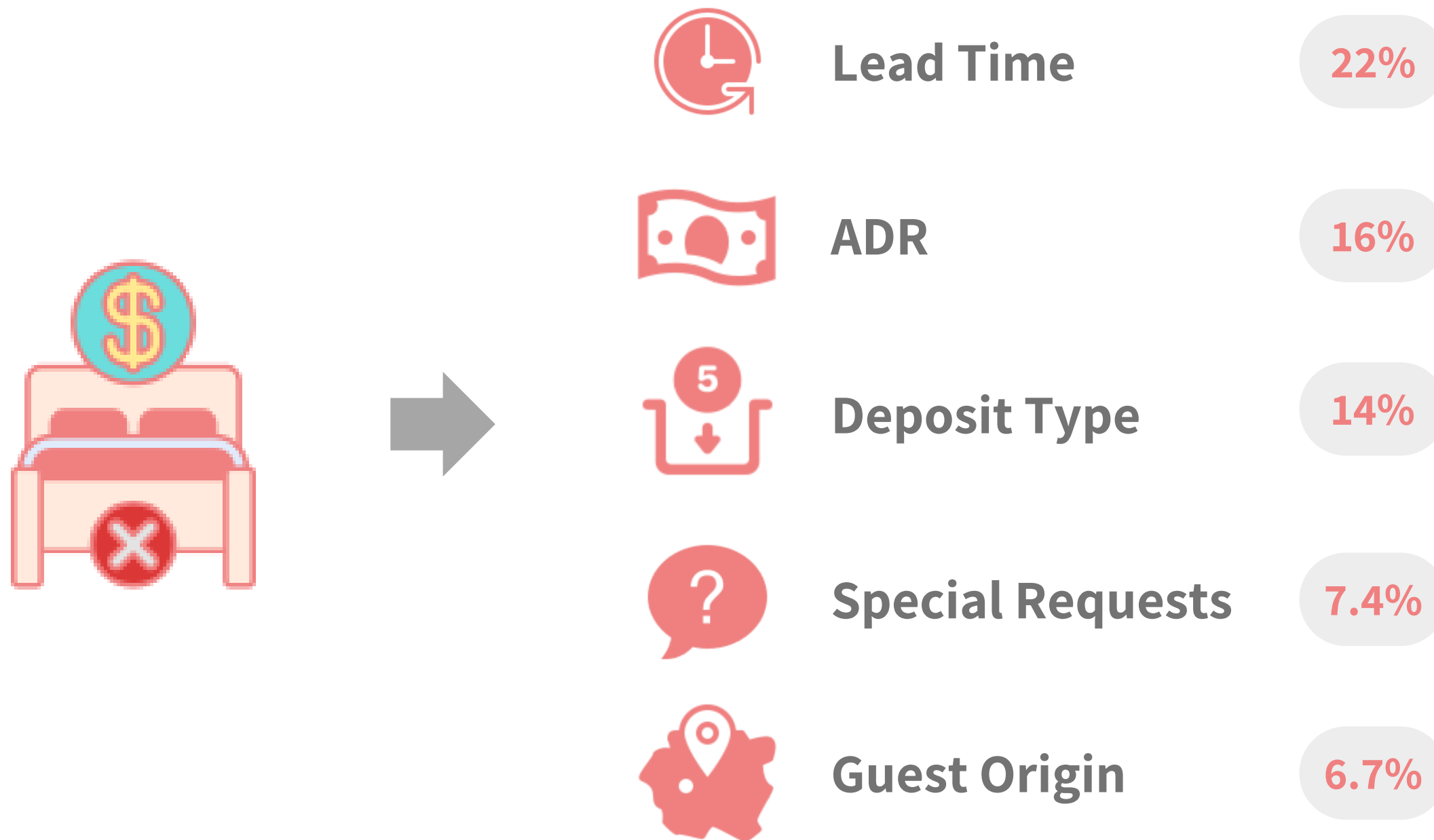
To **optimize revenue**, we can **increase rates YoY during peak months**, but **not by more than 10%.**

Implement treatments according to customer segments

Make clustering using K-Means

|	Proportion	ADR	Total Night	Favorite Month	Market	Treatment
Budget Seekers	**60.4%**	€ 94	2 Night	Mar - Oct	Domestic	Design friendly budget package
Long-Term Stayers	24.3%	€ 99	**5 Night**	Mar-Apr, Jul-Aug	Domestic & International (FRA, DEU)	Design attractive entertainment & discount package
Luxury Seekers	15.3%	**€ 165**	3 Night	Jul-Aug	Domestic & International (FRA, ESP)	Design exclusive and premium experience package

Top 5 factors* contributing to booking cancellations

	Factor	%
🕐	**Lead Time**	**22%**
💵	**ADR**	**16%**
	Deposit Type	**14%**
❓	**Special Requests**	**7.4%**
📍	**Guest Origin**	**6.7%**

*based on random forest model

Lead time has **high correlation** to cancelation rate

Lead time is **the time gap** between **booking** and **the arrival date**.
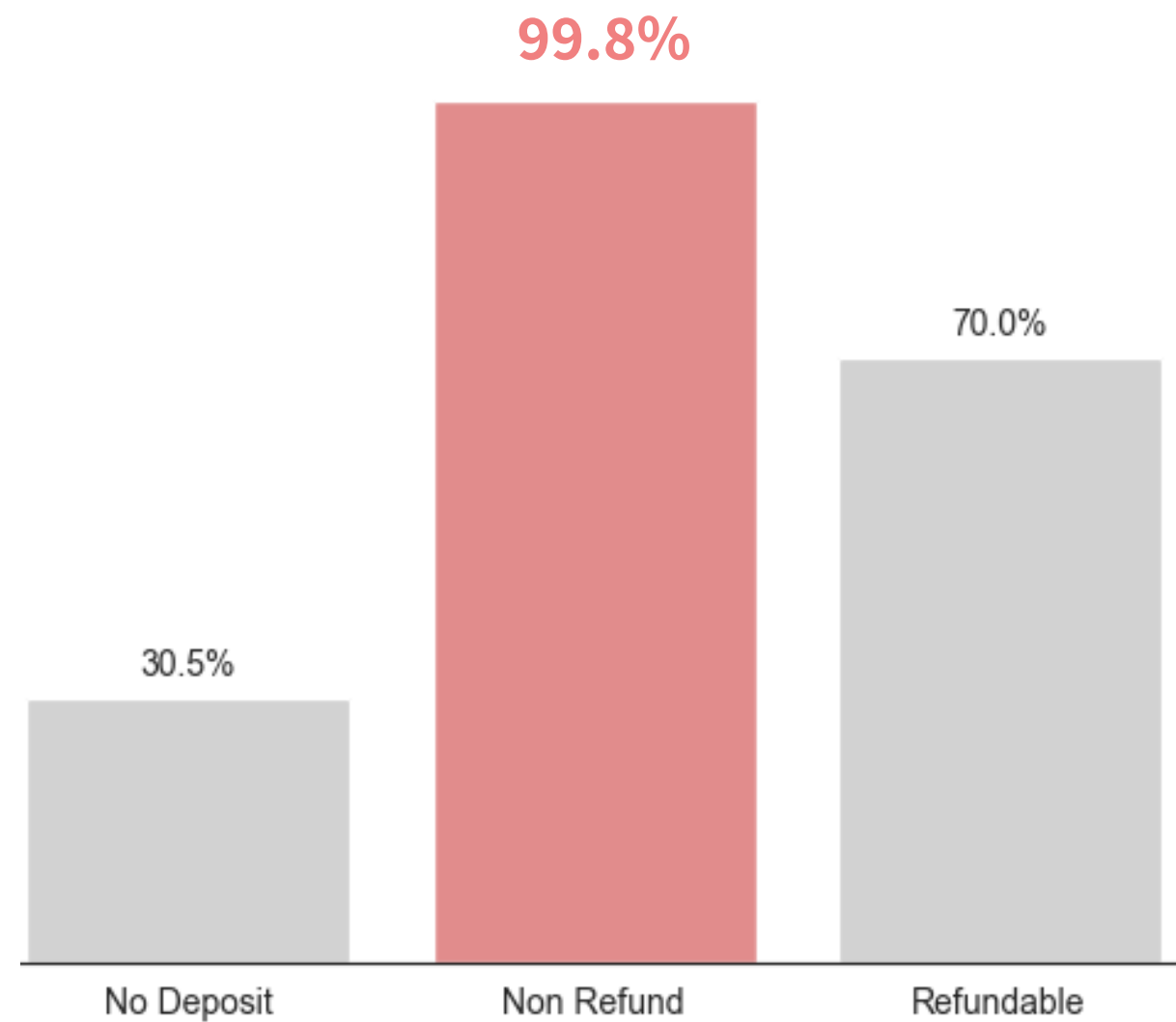
Lead time & cancellation rate
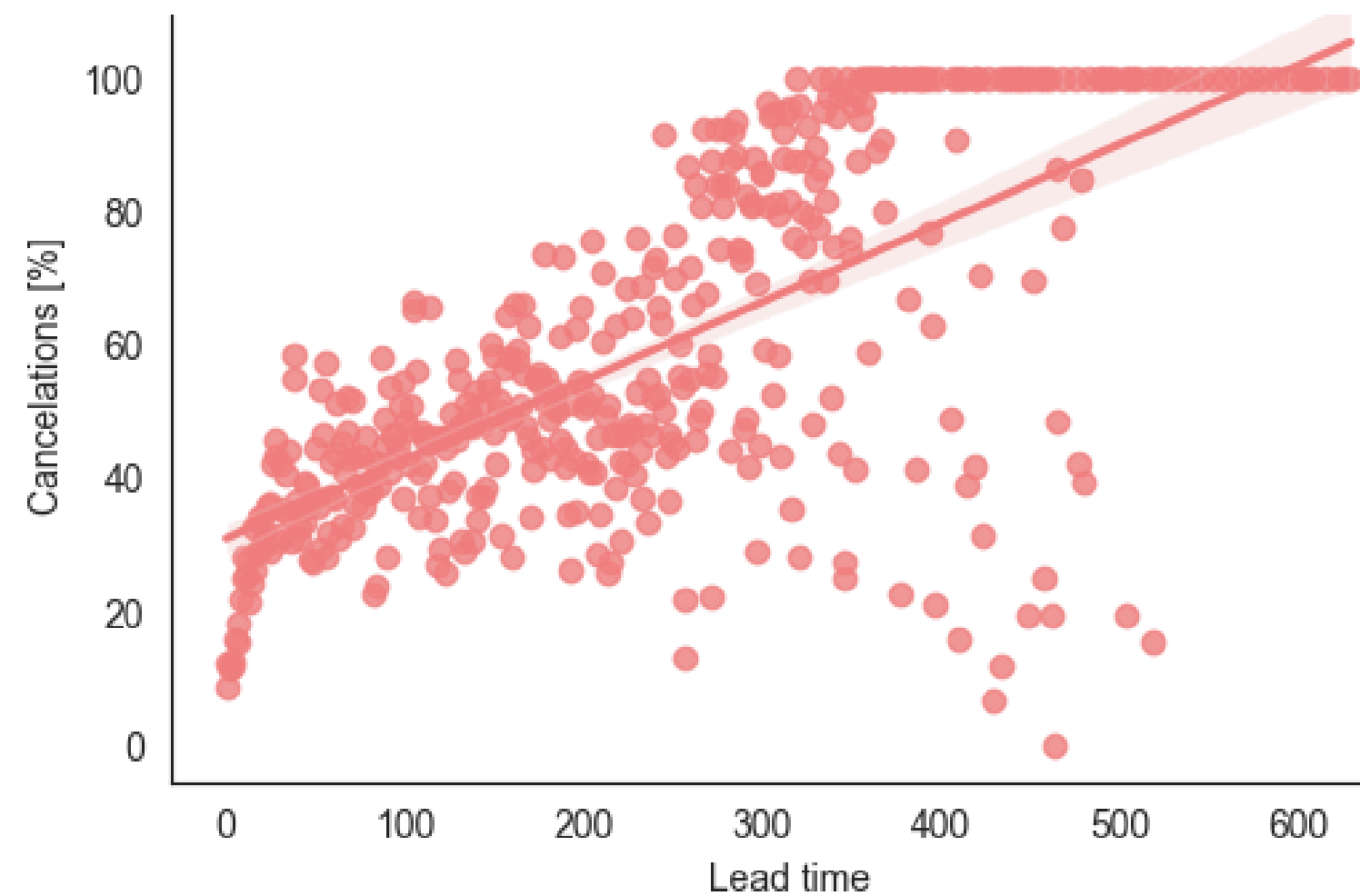
The longer the **lead time**, the higher the **cancellation rate**.

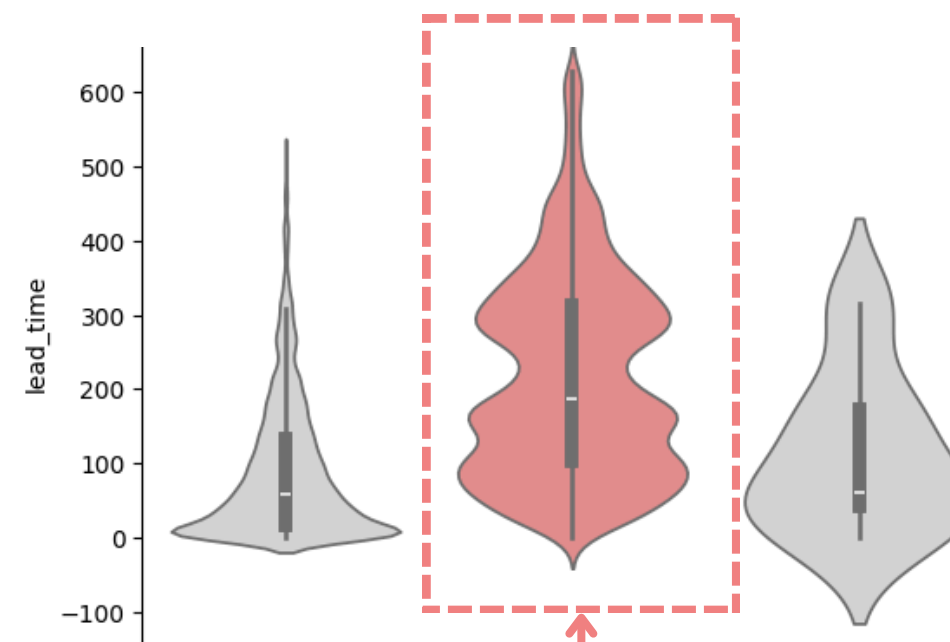| Lead Time (days)	Cancelation Rate
0 - 60	29 %
60 - 120	43%
120 - 180	48%
180 - 240	50%
> 240	71%

Lead time **> 2 months** has a cancellation rate of **over 40%**.

The No Refund deposit has a cancellation rate of 99.8%
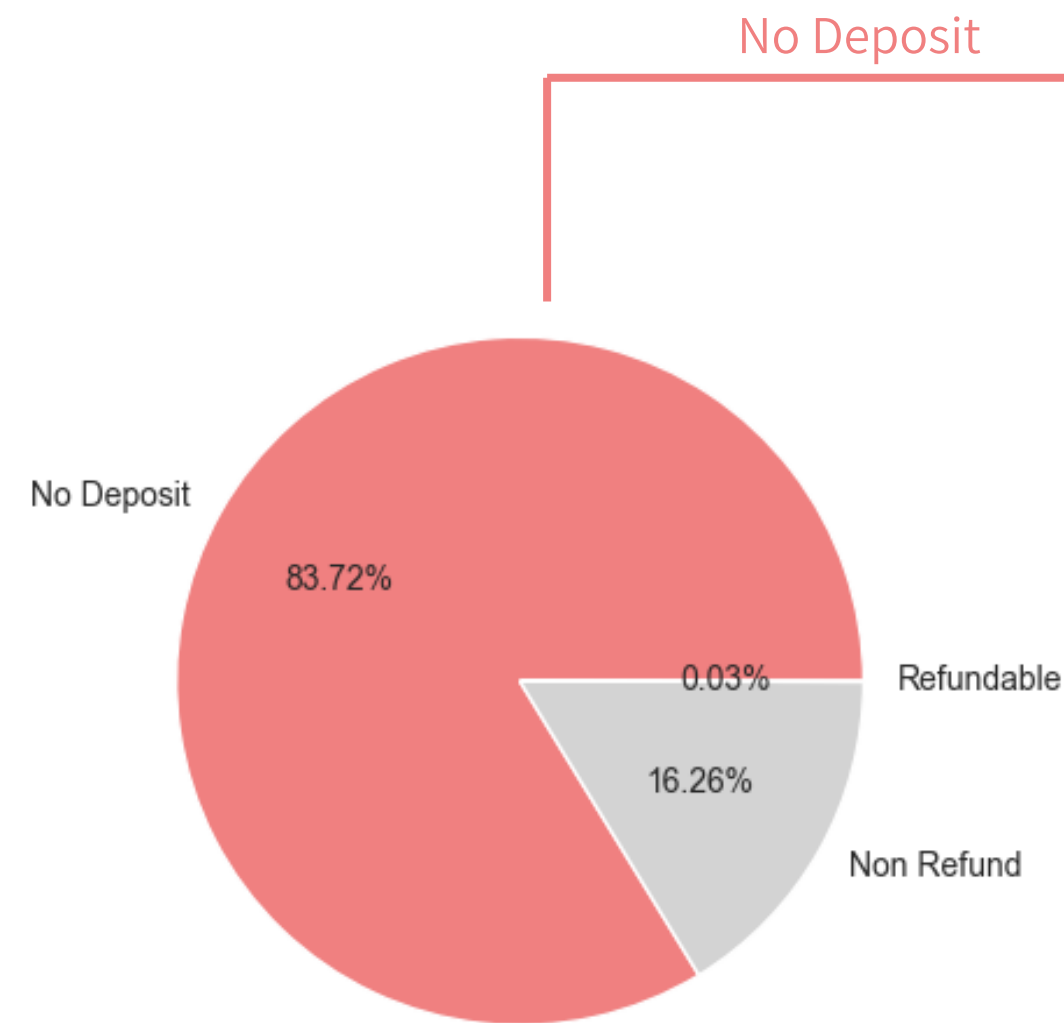
Cancelation Rate based on Deposit Type

Lead Time Non Refund is higher than others

Although the cancellation rate for Non Refund is very high, it **does not affect revenue**. To optimize revenue, we can implement a policy by **limiting the lead time** for Non Deposit and Refundable deposit types, and **redirecting high lead times** to the Non Refund type.

Redirect high lead time from the No Deposit type to the No Refund deposit

No Deposit

No Deposit is the **most favorite** option for booking

| Lead Time (days)	Cancelation Rate	Proportion
0 - 60	26 %	48%
60 - 120	34%	22%
120 - 180	36%	14%
180 - 240	38%	8%
> 240	43%	8%

70% of the No Deposit type has a lead time of **less than 120 days (4 months)**, with a low cancellation rate **below 35%**. To minimize the cancellation rate, we can set a **Maximum lead time of 120 days for the No Deposit type**.

Therefore, if a guest books more than 120 days before the arrival date, they are **required to use the Non Refund deposit type**.

Cancel guests tend to make fewer special requests

Total Special Requests besed on Cancelation
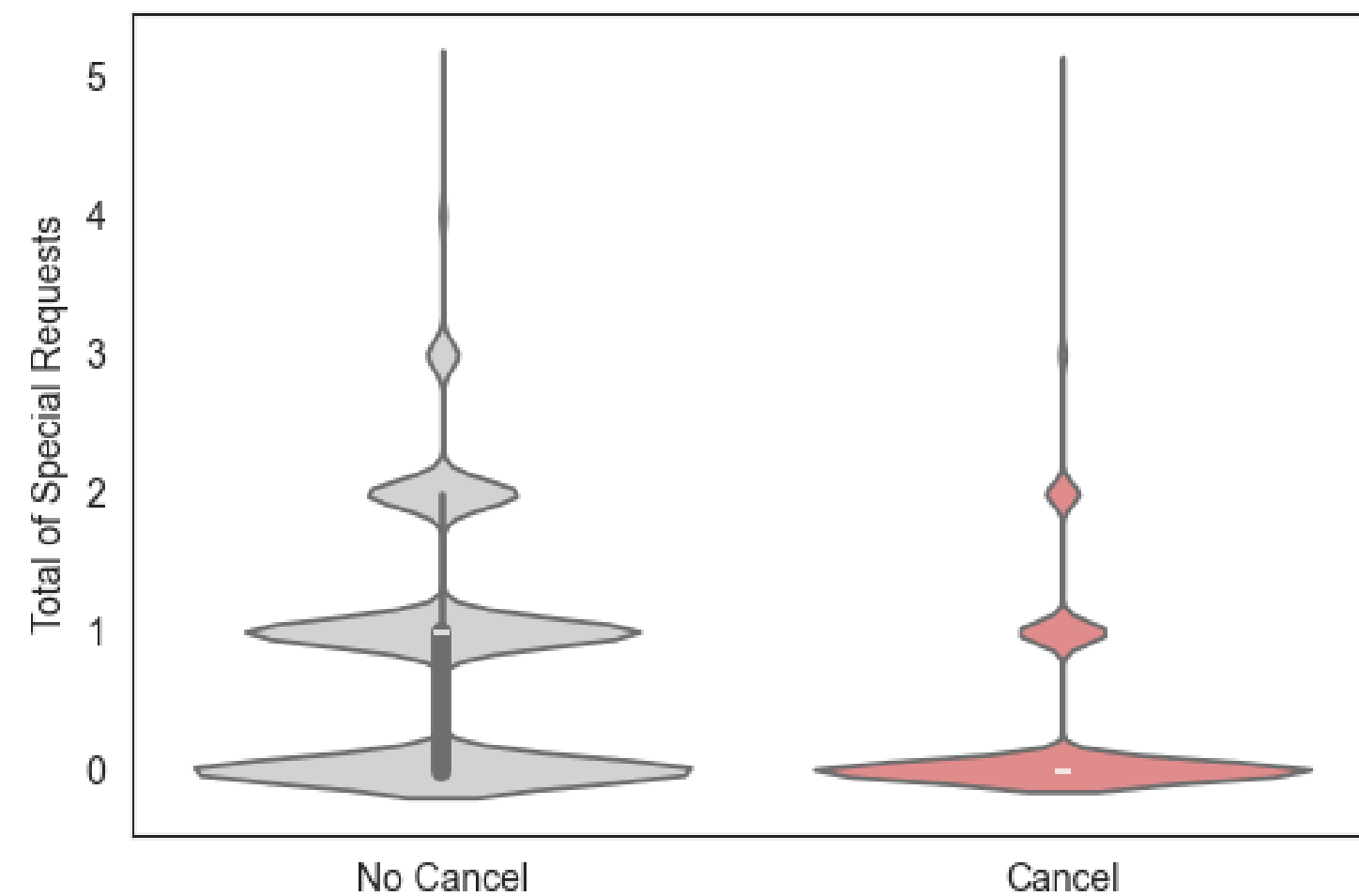
Guests who cancel their reservations tend to **make fewer special requests** compared to those who do not cancel.

ADR & Cancelation Rate based on total of special requests

| Total Special Requests	Cancelation Rate	ADR
0	55%	€99
1	22%	€111
2	21%	€120
3	17%	€123
4	10%	€121
5	4%	€120

With a **low ADR cause no special request**, guests tend to be **less hesitant** to cancel their reservations.

Domestic guests have highest cancelation rate

Cancelation Rate based on Guests Origin
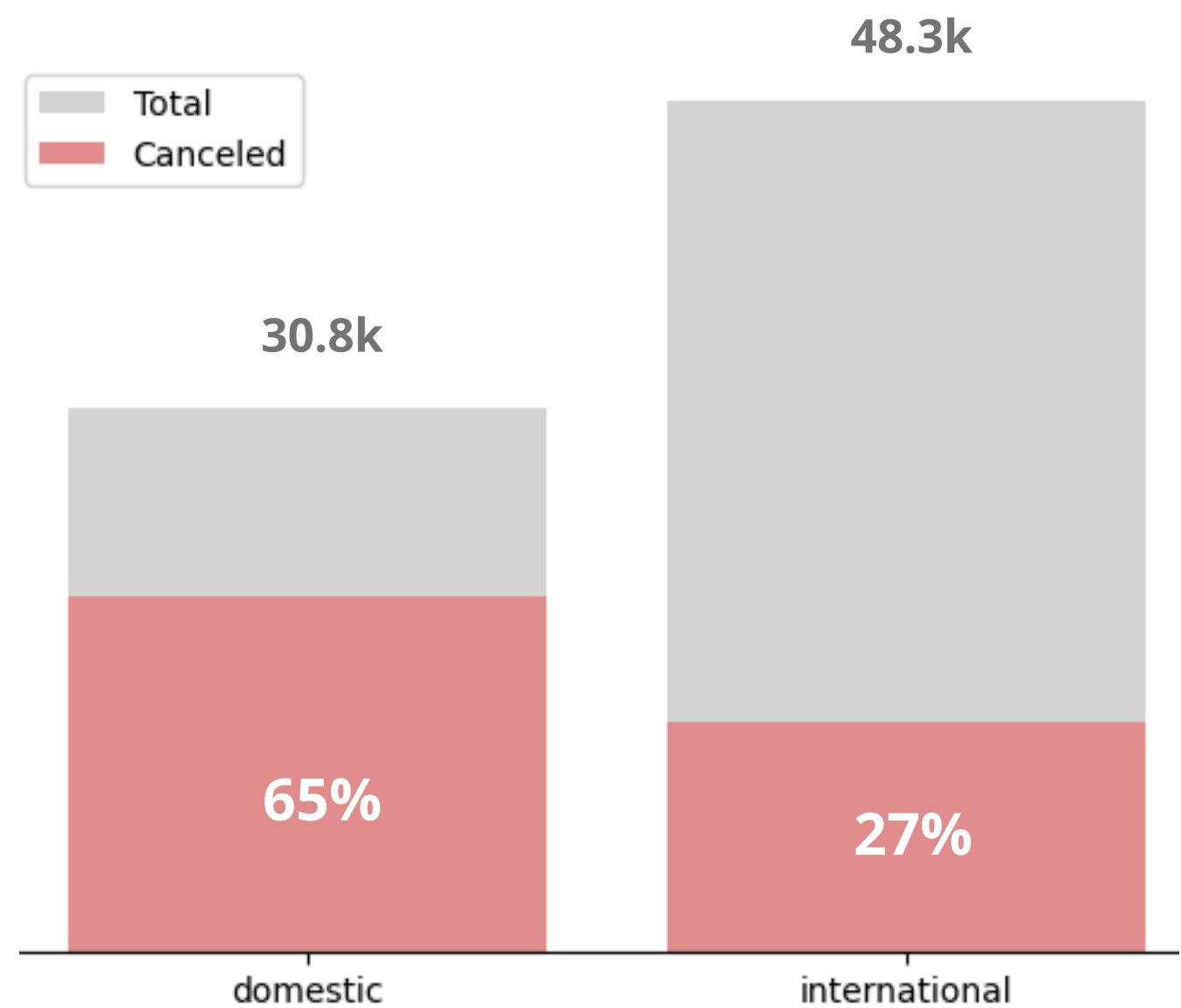
Deposit type proportion from domestic guests who cancel booking

No Deposit

36.7%

0.0%   Refundable

63.3%
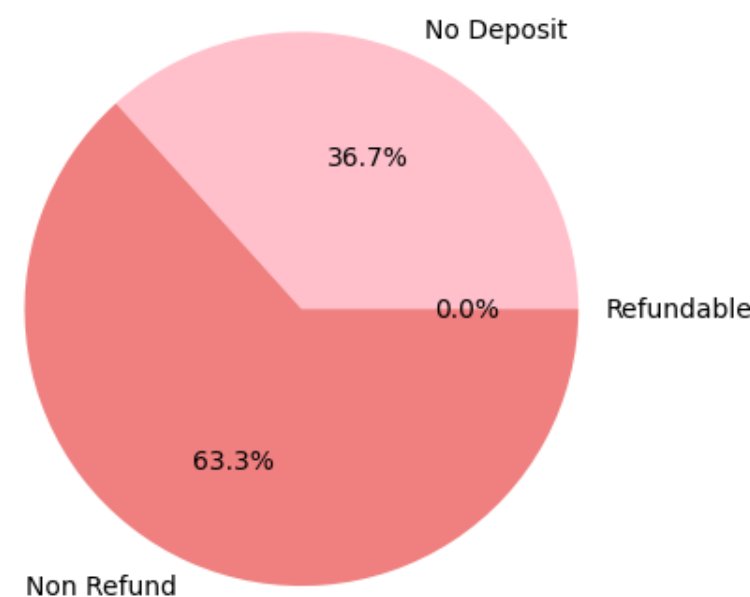
Non Refund

Top market segments

Groups	9.5k
Offline TA/TO	7k
Online TA	2.3k

- Domestic guests have a **cancellation rate of 65%**.
- Most domestic guests who cancel use the **non-refund deposit type** and come from the **market segment "groups"**.

Recommendations

Recommendation for Pricing Strategy

Factor	Objective	Recommendation

Pricing Issues

How to optimize revenue through pricing strategy?

Increase ADR (YoY) during **peak months (Mar-Oct)** by **no more than 10%**

Monitor competitor rates to ensure pricing remains competitive

Offer **packages tailored to segments** (Budget, Luxury, Long-Term)

Require bookings with **a lead time > 4 months** to use **a non-refundable deposit**

Recommendation to reduce cancelation rate

Factor	Objective	Recommendation

High Cancelation Rate

How to reduce cancelation rate?

Limit the lead time for **no deposit** bookings to **a maximum of 4 months**

Provide flexibility in changing the arrival date for **domestic guests**, especially for the **"group"** market segment
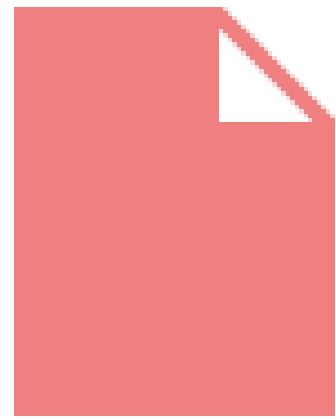
Engage in **frequent communication** and **offer attractive special requests** for **non-refundable** deposit bookings

Conduct **surveys** to understand the **reasons** behind guest cancellations
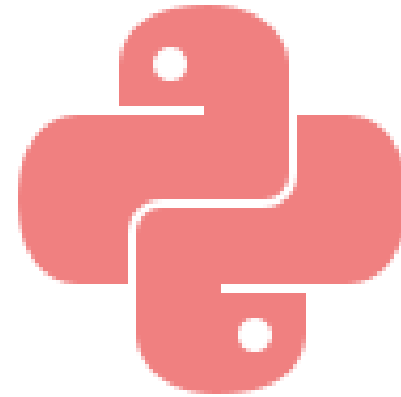
Thank You !
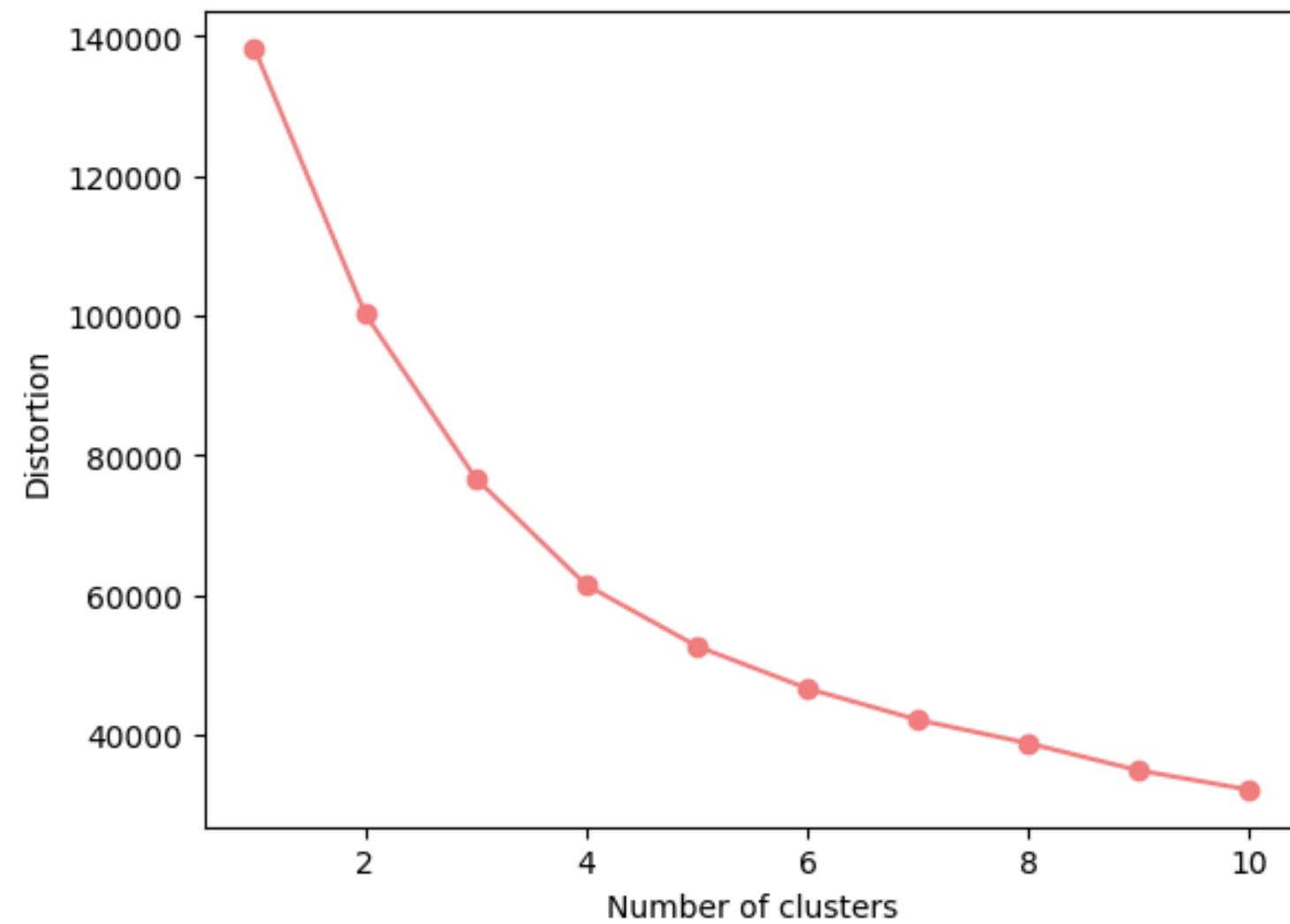
Appendix
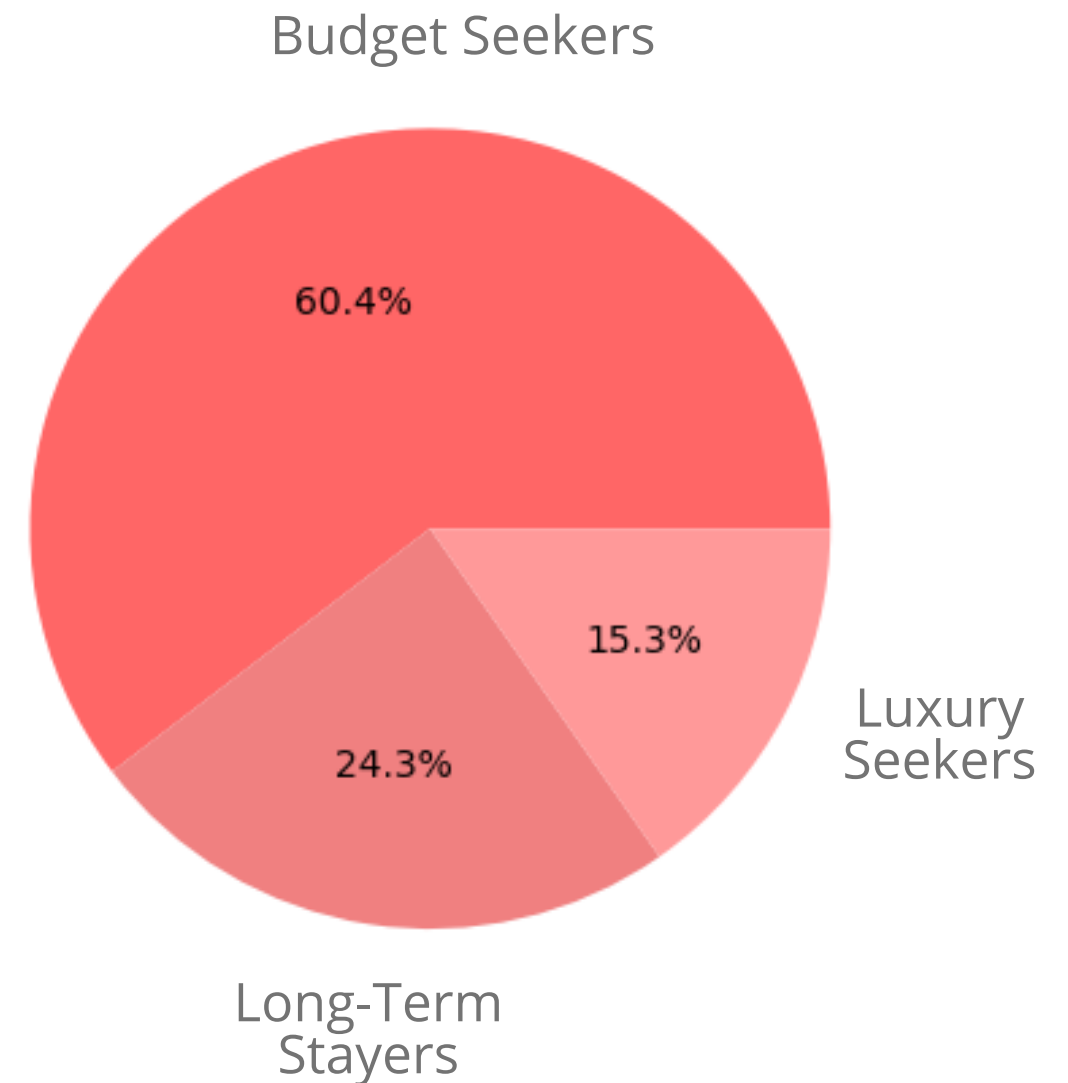
Visit the link for see more

Data Source

Python Code

There are 3 customer clusters that can be created using K-Means

Elbow Plot

Cluster Proportion

Random Forest became the best model for predicting booking cancellations

Evaluation Model

	Accuracy	Precision	Recall	F1
Logistis Regression	79%	82%	63%	71%
Decision Tree	83%	79%	80%	80%
Random Forest	86%	85%	81%	83%

The **Random Forest** model performed the best in terms of **accuracy, precision, recall, and F1 score**.

Confusion Matrix Random Forest

	0	1
0	12546	1328
1	1876	7999