



BORCELLE HOTEL

# HOTEL BOOKING ANALYSIS

## *Strategies to Reduce Cancellations*

Prepared by: Ade Dwi Aryani



# Table of *Content*

**01.** Executive Summary

**02.** Borcelle Hotel Overview

**03.** Problem Statement  
and Objectives

**04.** Methodology

**05.** Data Analysis

**06.** Recommendation



# Executive Summary

## Project Description:

Borcelle hotel faces a high booking cancellation rate of 36.83% in August 2017, which are causing revenue instability and operational challenges.

"How can Borcelle Hotel reduce its booking cancellation rate from 36.83% to 20% by 2018?"

## Objective:

To find and implement strategies to reduce the booking cancellation rate from 36.83% to 20% within the next 12 months (by 2018)

## Tools:



For data analysis, including logistic regression and K-Means clustering



For data visualization and presentation

## Metrics:

### Cancellation Rate

The percentage of bookings canceled out of the total bookings

### Logistic Regression Accuracy

**81% accuracy** with an F1-score of 78%

### K-Means Clustering

Used to segment customers into 3 categories  
**(Budget-Conscious Travelers, Leisure Travelers, Frequent Business Travelers)**

## Analysis Result & Recommendation

Factors to focus on:	Insight	Recommendation
Required Car Parking Spaces	High required car parking spaces will reduce the cancellation rate	Optimize car parking availability
Deposit Type	Non-refundable deposits lead to lower cancellation rates	Highlight the non-refundable deposit policy
Previous Cancellations	High previous cancellations likely lead to a high cancellation rate	Highlight clear booking and cancellation policies
Budget-Conscious Travelers	Highest cancellation rate	Offer discounts for non-refundable bookings
Leisure Travelers	Moderate cancellations rate	Offer no deposit with flexible cancellations to attract customers
Frequent Business Travelers	Lowest cancellation rate	Enhance loyalty programs



# Borcelle Hotel Overview

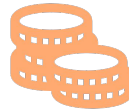
Borcelle is a hotel in Portugal with two properties: a **City Hotel** in **Lisbon** and a **Resort Hotel** in the **Algarve**



Borcelle Hotels offers **premium stays** in exotic destinations, attracting **guests from around the globe**



**€29,45M total revenue**



**€102,67 average daily rate**



**3-4 nights average stays**



**118,408 bookings**



**44,014 bookings cancelled, or 37.17% of the total.**  
The cancellation rate is **relatively high**



**145,1K guests**



**176 countries of origin**

\* Period: July 2015 - August 2017

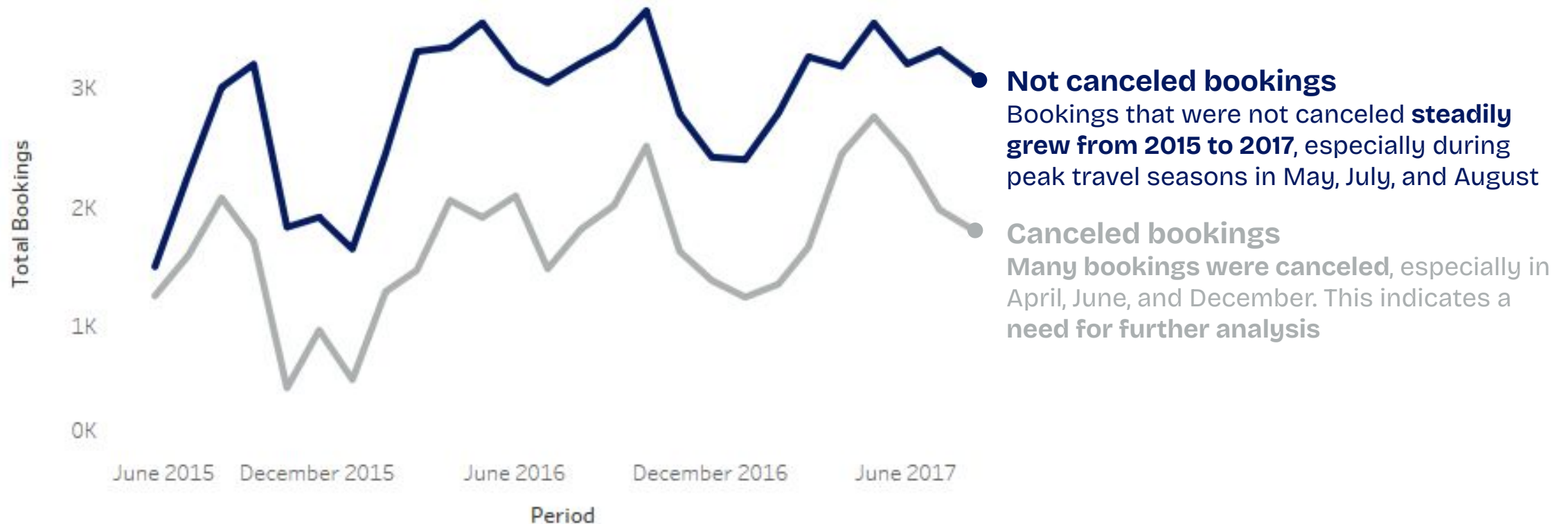


# Borcelle Hotel Overview

Period: July 2015 - August 2017

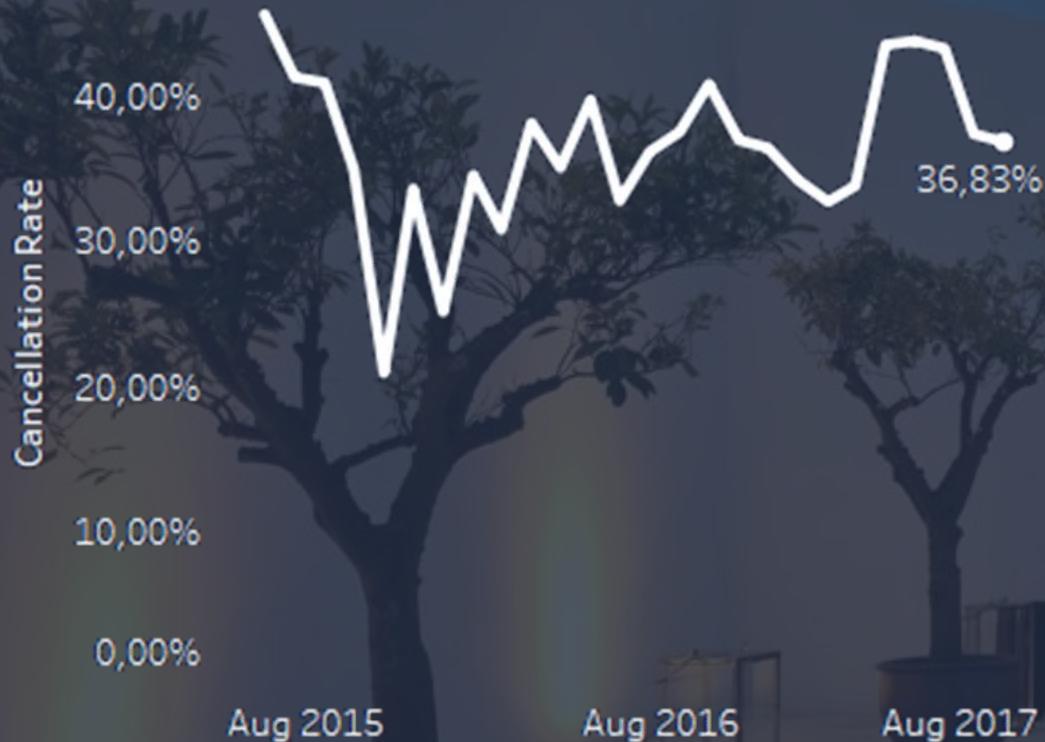


## Trends in Canceled and Not Canceled Bookings



# Problem *Statement*

Trends in Cancellation Rate



Head of Marketing ask to Data Team

“How can we reduce the booking cancellation rate from 36.83% to 20% within the next 12 months (by 2018)?”

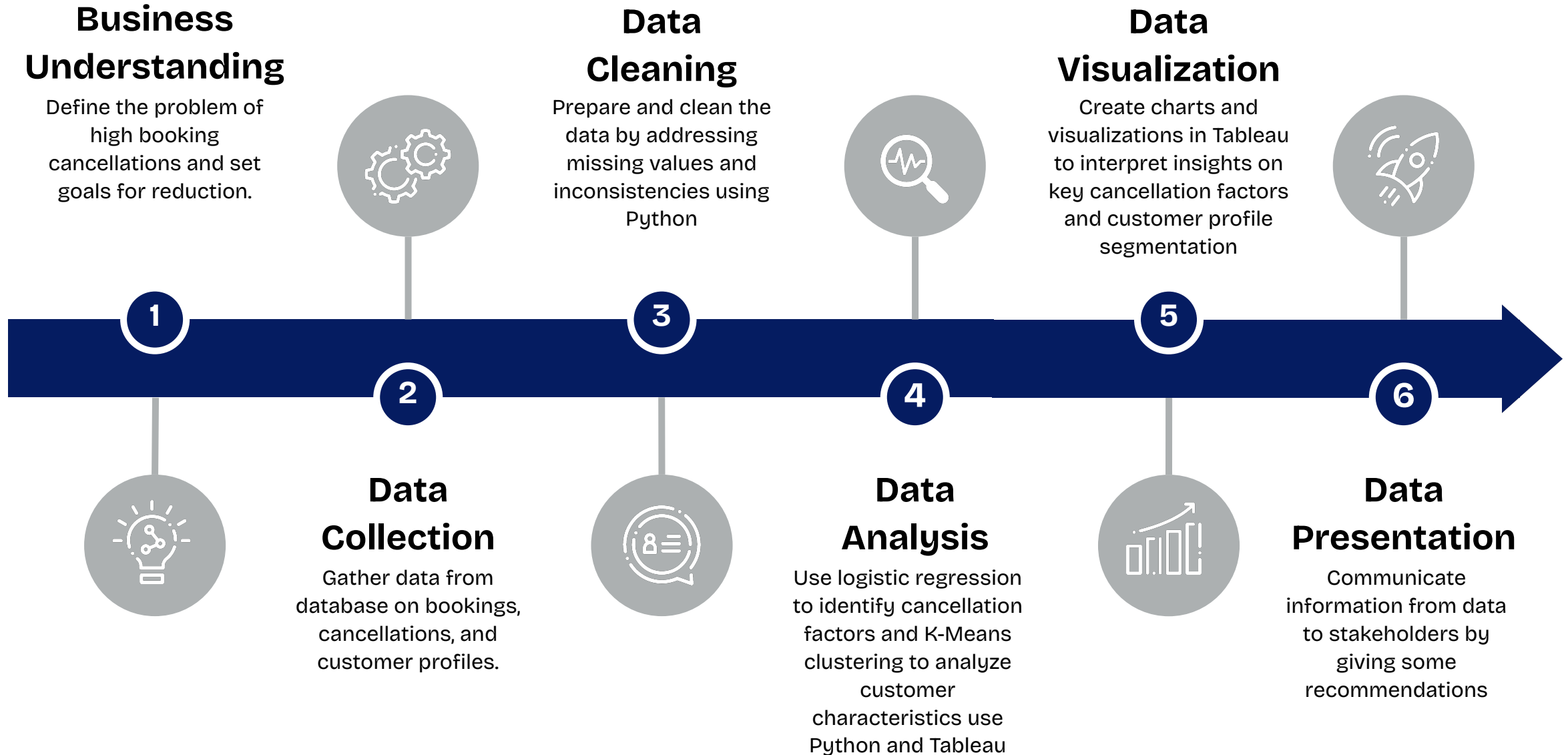


# *Objective*

**To find and implement strategies to reduce the booking cancellation rate from 36.83% to 20% within the next 12 months (by 2018)**



# Research Methodology







# Hotel Booking Analysis

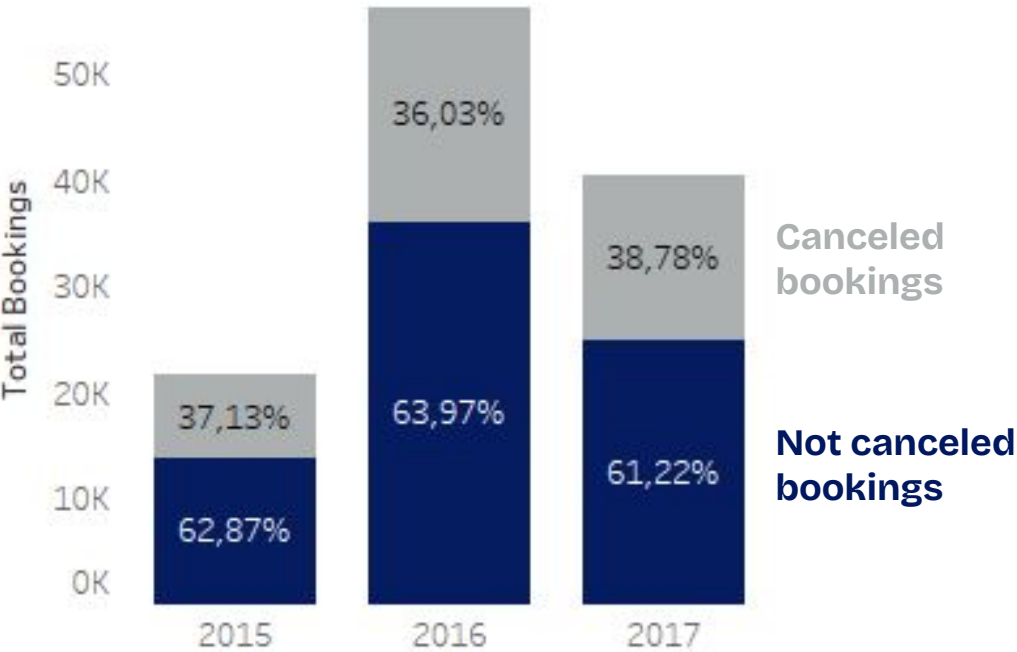
## *Strategies to Reduce Cancellations*



# Rising cancellation rates are causing revenue instability and planning challenges

## Increase in Cancellation Rate

▲ 1,65% (Jul 2015 – Aug 2017)



There was a **1.65% increase** in the cancellation rate from July 2015 to August 2017, indicating **potential losses for the company** due to revenue impact from canceled bookings.

High booking cancellation rate **can lead to several negative consequences for hotels**, including:

01.	Financial Losses	Lost revenue from unfilled rooms, especially with last-minute cancellations
02.	Inefficient Resource Allocation	Wasted resources as prepared rooms and services go unused
03.	Operational Disruption	Challenges in staff scheduling and service preparation
04.	Reputation Damage	Customers may view the hotel as unreliable, affecting future bookings.

Source : Lin et al. (2023)

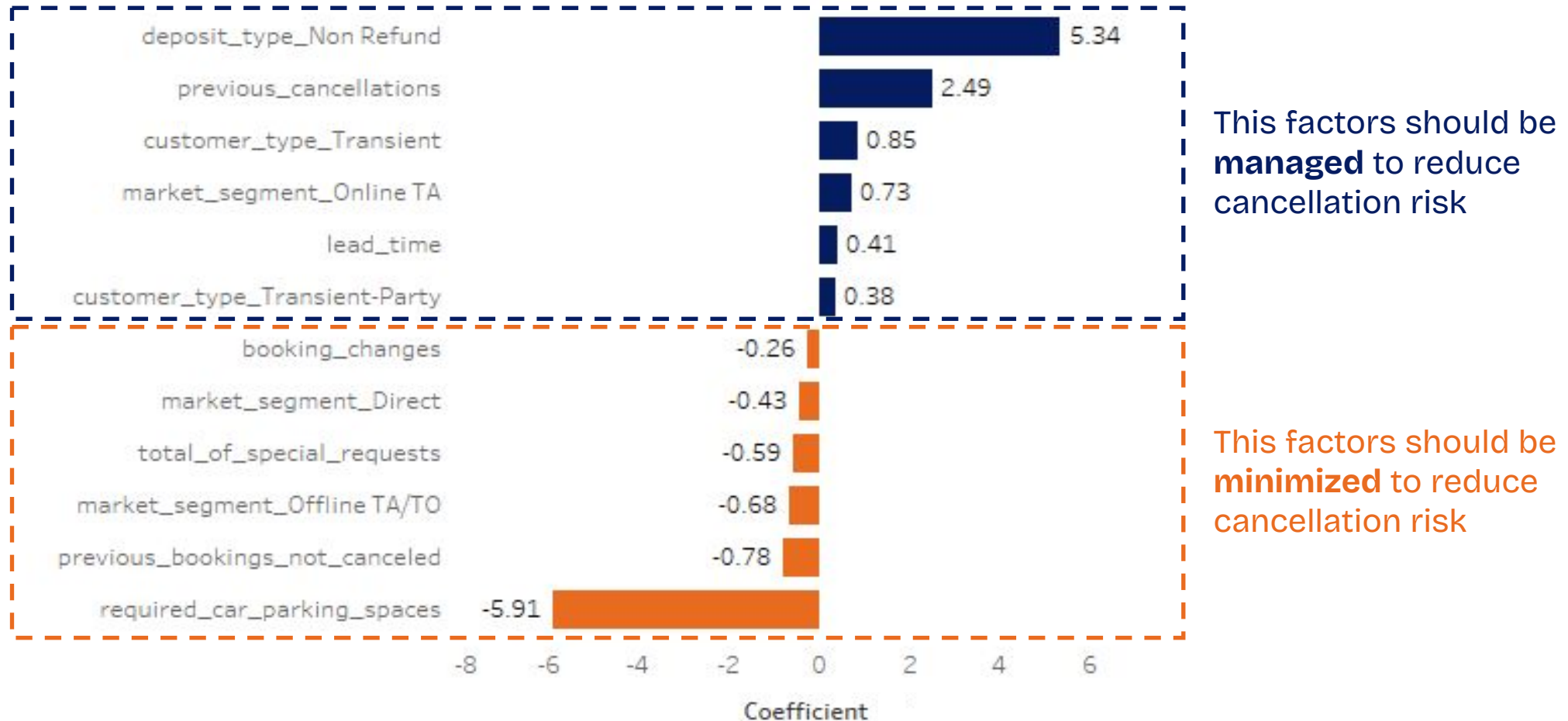
# Non-canceled bookings are more profitable, while canceled ones result in significant potential revenue losses

	Not Canceled Booking	Canceled Booking
Revenue	€25.85M	€3.60M
Avg. Lead Time	2-3 Months	4-5 Months
Avg. Days in Waiting List	1-2 Days	3-4 Days
Avg. Booking Changes	0.28 times (More changes)	0.09 times (Lower changes)
Total Bookings	74,394 Bookings	44,014 Bookings

- Non-canceled bookings yield **higher revenue** with **fewer changes** and **shorter wait times**,
- Canceled bookings, generating **revenue only from non-refundable** reservations, lead to **more changes, longer wait times, and potential revenue losses**.

# Manage factors with positive coefficients and minimize factors with negative coefficients to reduce cancellation risk

## Impact of Key Factors on Booking Cancellation

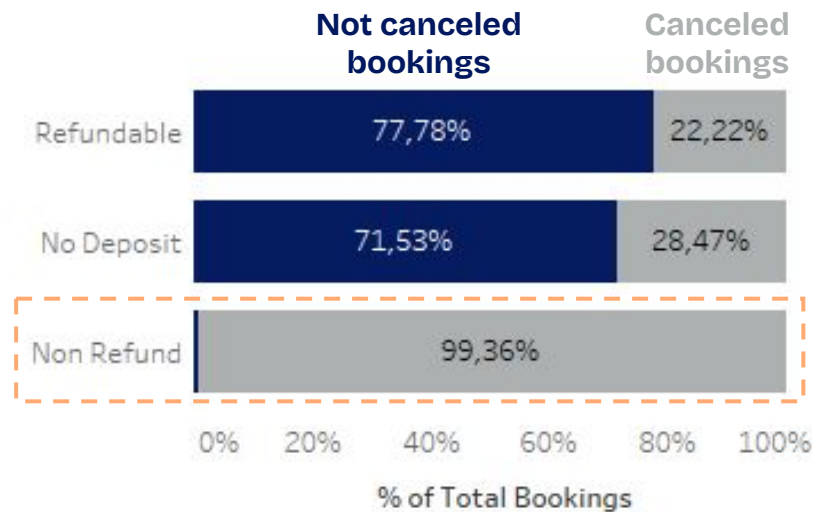


Key factors for predicting booking cancellations were identified using **logistic regression with 21 features**. The model achieved an **accuracy of 81%** and an **F1-score of 78%** on both training and test data.



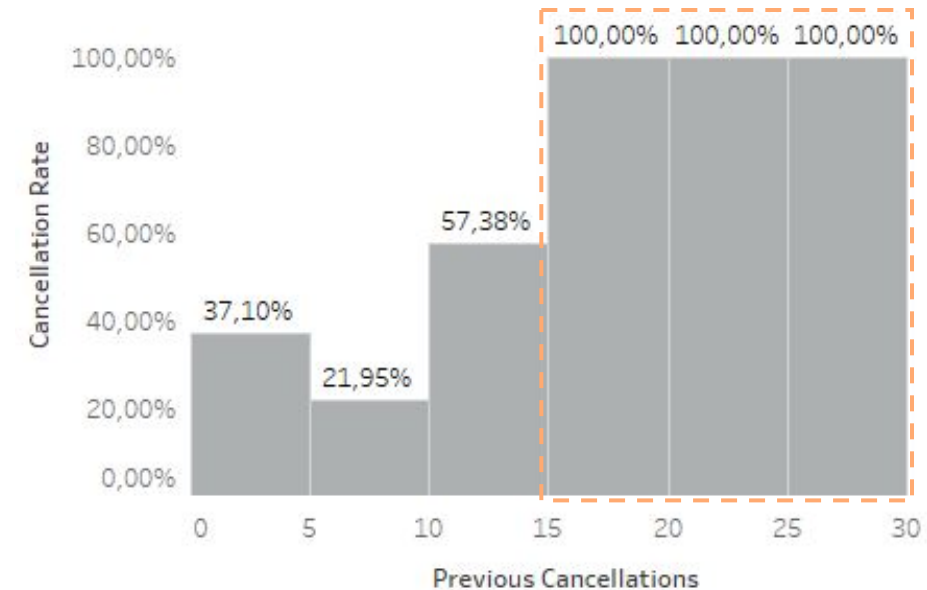
# Non-refundable bookings and entities with more past cancellations have higher cancellation rates

## Higher Cancellation Rates for Non-Refundable Deposits



- **Most canceled bookings** are from the **Non Refund**, indicating a higher tendency to cancel when deposits are non-refundable.
- People may tend to cancel non-refundable bookings due to **urgent situations or changes in plans**, even if it means losing the deposit.

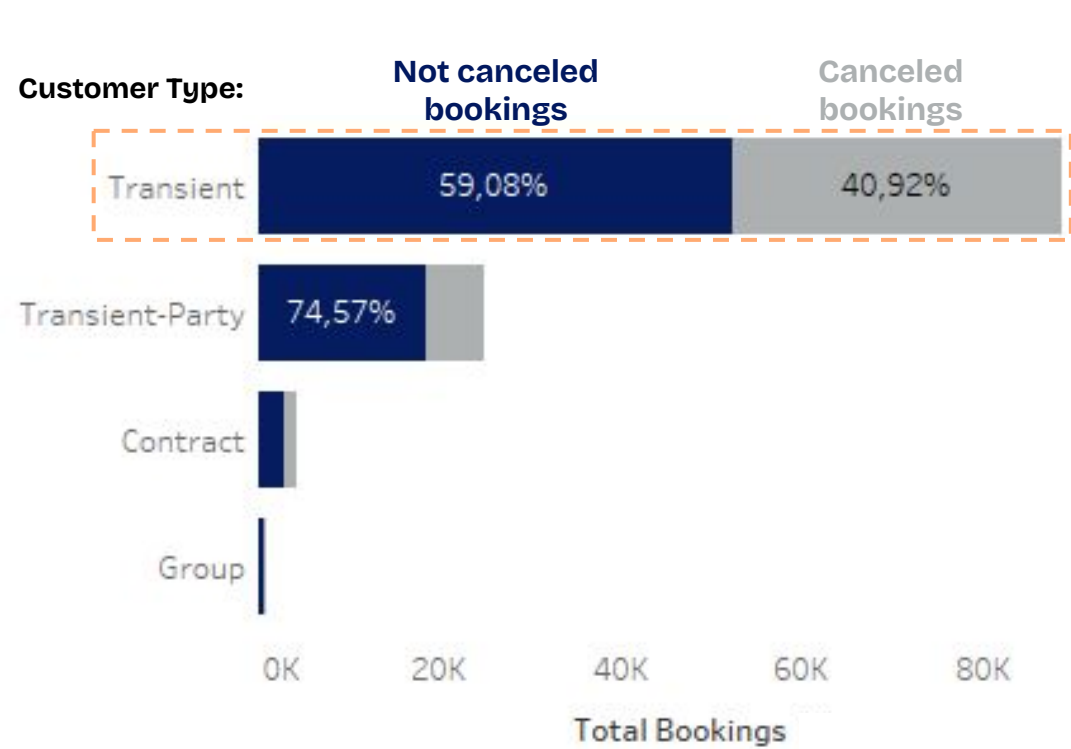
## More Previous Cancellations Mean Higher Risk



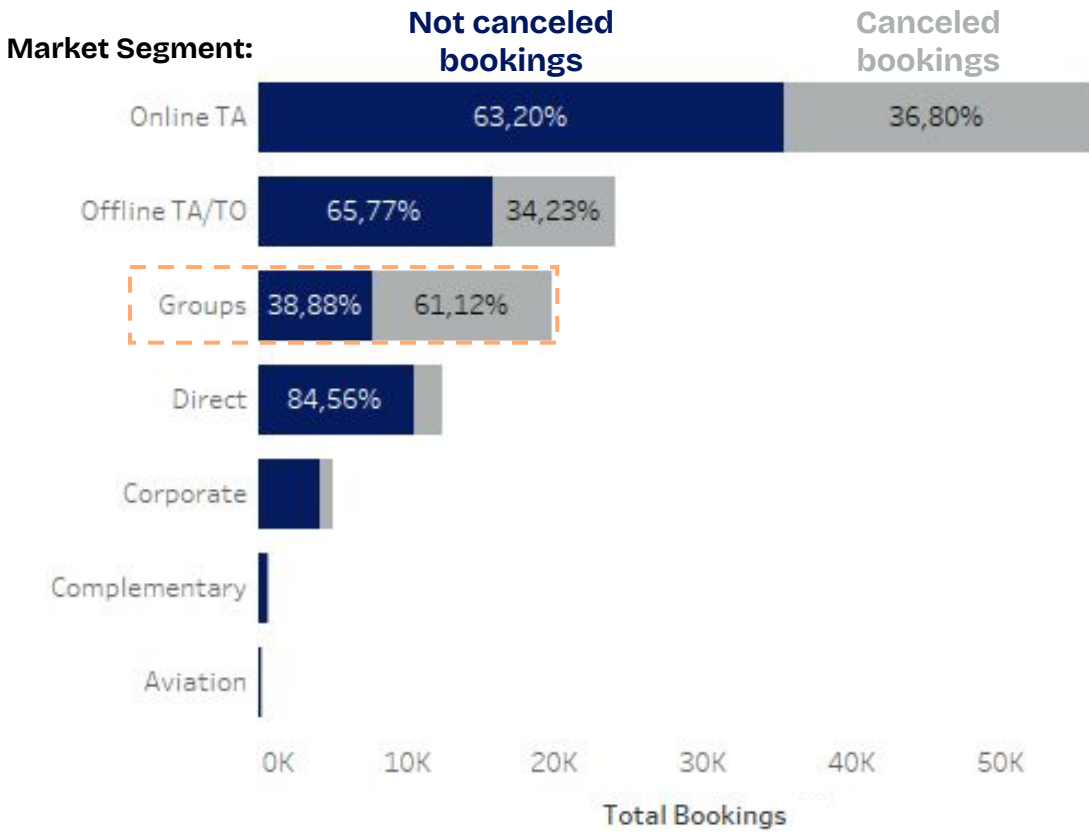
- The **more previous cancellations** an entity has, the **higher** the likelihood of **future cancellations**.
- Conversely, with fewer or no previous cancellations, the cancellation rate tends to be lower but still significant.

# Transient customers and groups both show high cancellation rates

Transient customers cancel the most



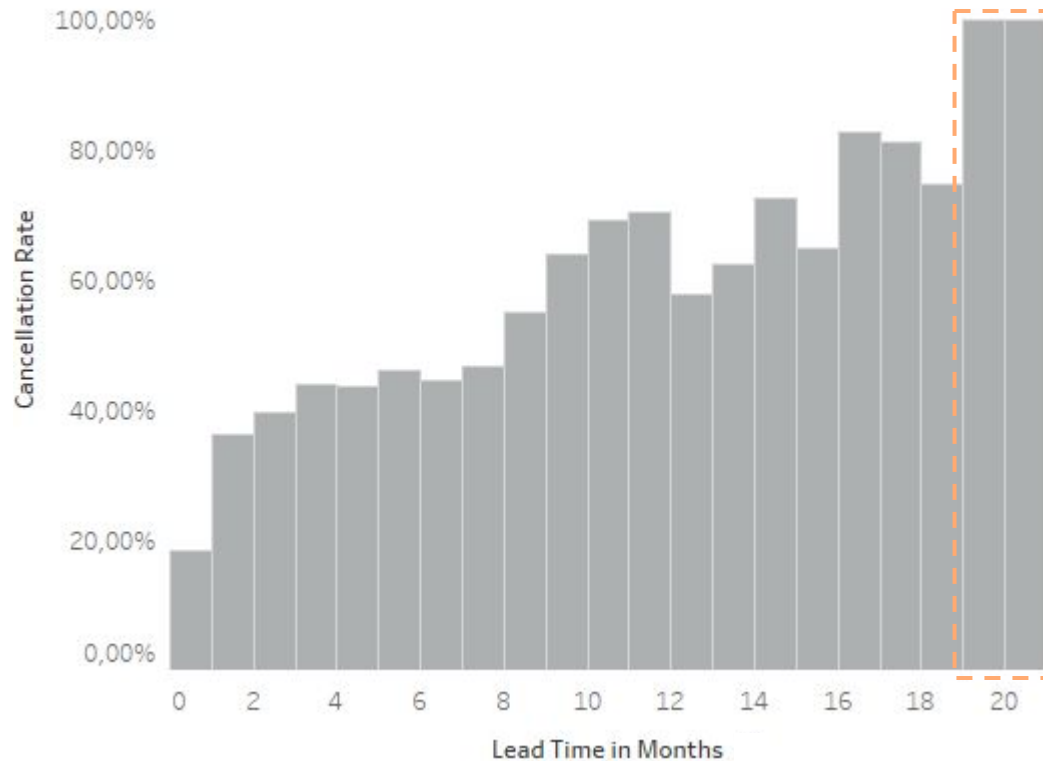
Groups have the highest cancellation rate



- **Group customers have the lowest cancellation rate**, indicating high commitment. While **transient customers** show the **highest cancellation rate**, making them less stable.
  - **Market segment: Groups** have the **highest cancellation rate**, indicating they are the least stable.
- Notes:** "Groups" in market segment refers to the market category serving groups, while "Group" in customer type refers to the category of customers who come in groups.

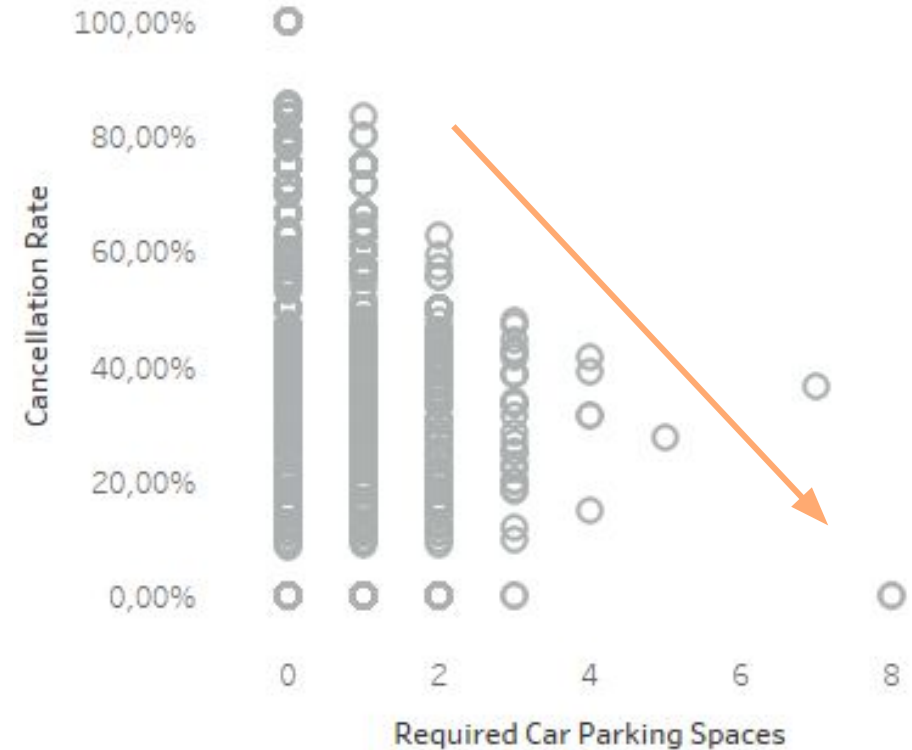
# Cancellation rates drop as the execution date nears and with more car parking spaces

Cancellation rates peak early and drop sharply near the stay date



- Cancellation rates are **highest with longer lead times**, particularly at 20-21 months.
- As the execution date approaches, the cancellation rates gradually **decrease**, with a **significant drop by the 0-month** mark.

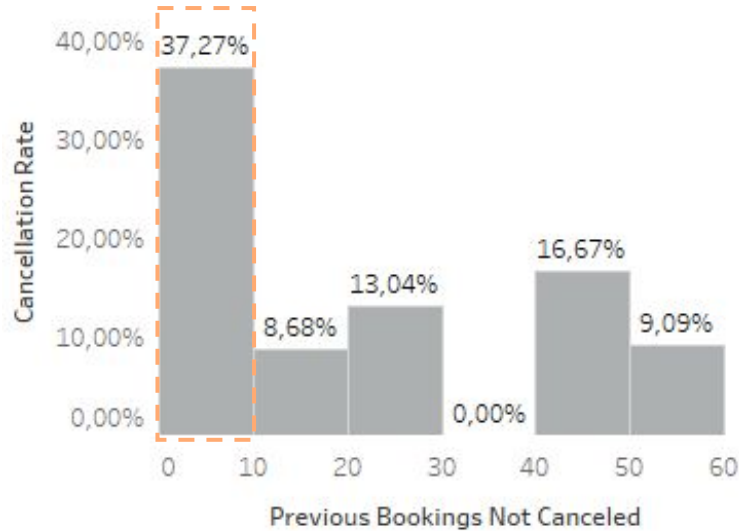
More car parking spaces reduces cancellations



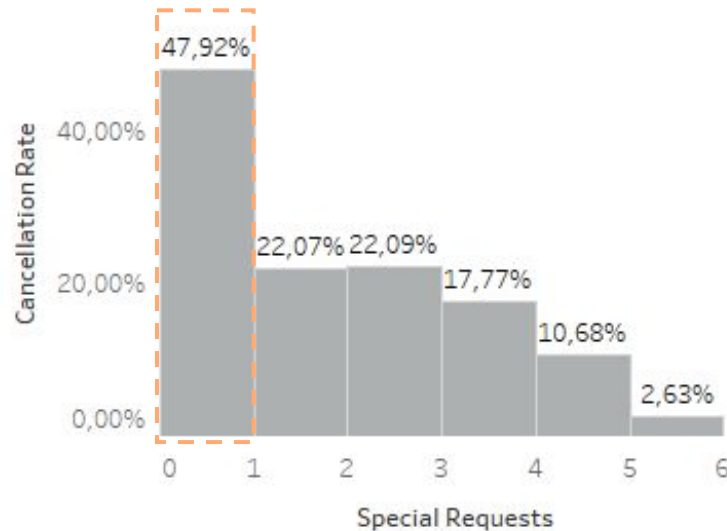
- As the number of **required car parking spaces increases**, the **cancellation rate tends to decrease**

# Customers with more previous bookings not canceled, special requests, or changes tend to cancel less

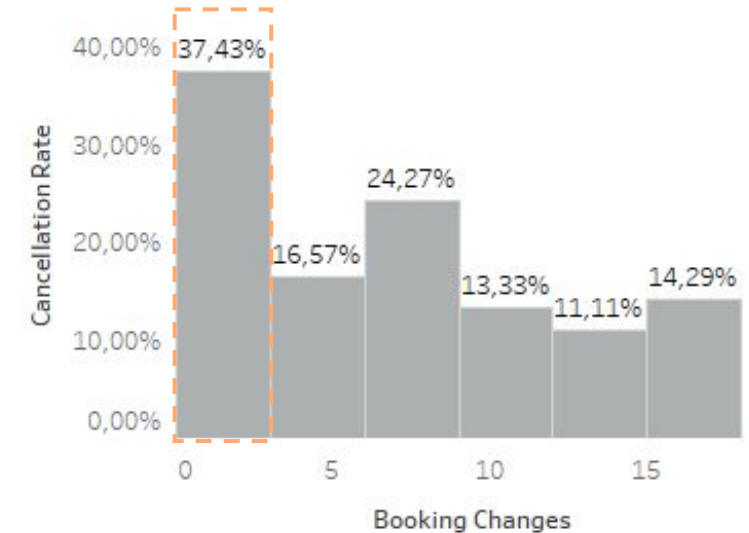
**Previous bookings not canceled reduces cancellation rates**



**More special requests reduce cancellation rates**



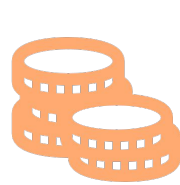
**More changes reduces cancellations**



- The **more previous bookings, special requests, or changes to a booking**, the **lower the cancellation rate**. Customers who are more engaged in their bookings are less likely to cancel.
- Conversely, new or less active customers tend to have a higher cancellation rate.



# After applying KMeans Clustering, we identified three segments:



Budget-Conscious  
Travelers



Leisure  
Travelers



Frequent Business  
Travelers

Cancellation Rate	47.9% (High)	25.3% (Moderate)	0.04% (Very Low)
Lead Time	123.9 days (Long)	79.6 days (Moderate)	55.2 days (Short)
Average Daily Rate	€94.93 (Lower)	€112.70 (Higher)	€113.58 (Highest)
Special Requests	0.05 (Few)	1.42 (Many)	0.83 (Moderate)
Booking Changes	38.7% (More)	24.4% (Moderate)	18.6% (Fewer)

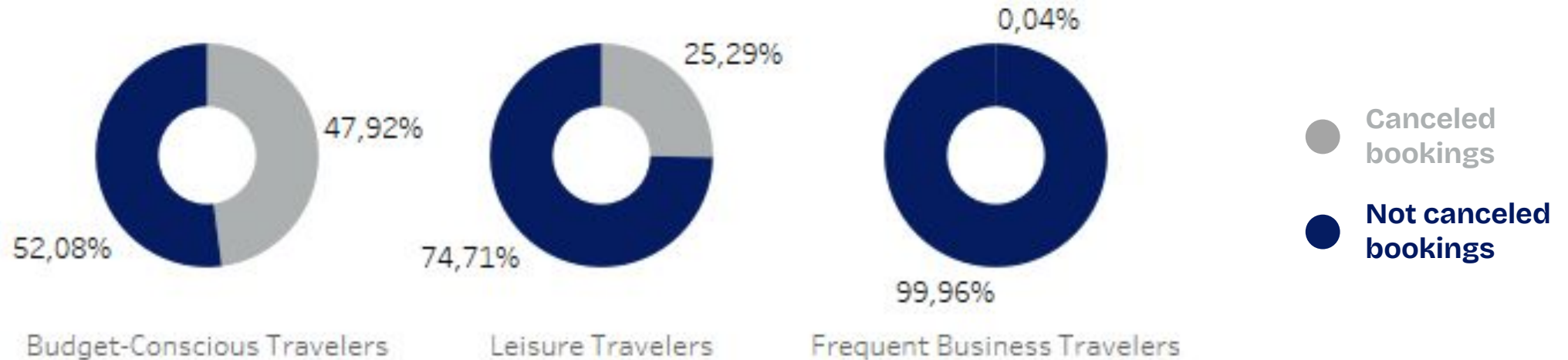
**Budget-conscious travelers** prefer affordability and flexibility, prioritizing cost over luxury and showing less commitment to their plans.

**Leisure travelers** desire comfort and personalization, are willing to pay more for higher quality and plan well in advance.

**Frequent business travelers** require convenience and high service levels, often booking at the last minute and expecting specific amenities for work.

# Frequent Businesses rarely cancel, Leisure moderately cancel, and Budget-Conscious travelers cancel most

Cancellation patterns and booking sustainability based on customers segments



- **Frequent Business Travelers rarely cancel bookings**, showing strong commitment.
- **Leisure Travelers have moderate cancellation rates** but most still proceed with bookings.
- **Budget-conscious travelers have the highest cancellation rates**, reflecting flexibility in their travel plans.

# Insights *Summary*

**Factor-Based:**  
Prioritize factors  
with a high  
coefficient

	Insights	Recommendation
Required Car Parking Spaces	High required car parking spaces will reduce the cancellation rate	Optimize car parking availability
Deposit Type	Non-refundable deposit types tend to have a lower cancellation rate	Highlight the non-refundable deposit policy
Previous Cancellations	High previous cancellations likely lead to a high cancellation rate	Highlight clear booking and cancellation policies, including consequences of cancellations
Customer Type	Transient customer type tends to increase the cancellation rate	Offer incentives to transient customers for longer stays
Previous Booking Not Canceled	High previous bookings not canceled will reduce the cancellation rate	Create a loyalty program for customers who rarely cancel
Market Segment	Online TA market segment is associated with a higher cancellation rate, Offline TA/TO and direct market segment tends to reduce the cancellation rate	Improve online booking experience, showcase deals for offline travel agency bookings on the website, and promote direct bookings with exclusive offers and perks

# Insights *Summary*

	Insights	Recommendation
Special Requests	Total of special requests reduces the cancellation rate	Improve the special requests section on the booking page
Lead Time	Longer lead time generally increases the cancellation rate	Offer incentives for early booking
Budget-Conscious Travelers	Highest cancellation rate	Offer discounts for non-refundable bookings to encourage customers to commit to their reservations
Leisure Travelers	Moderate cancellations	Offer no deposit with flexible cancellations to attract customers
Frequent Business Travelers	Lowest cancellation rate	Enhance loyalty programs

**Segment-Based:**  
Focus on budget-conscious travelers to reduce cancellations effectively





# Recommendation

*Strategies to Reduce Cancellations*



# Recommendations to Reduce Cancellations Based on Factor Importance

	Strategy	Action Plan	
Required Car Parking Spaces	Add a section on the website highlighting car parking options (e.g., valet, on-site)	Create a car parking page with visuals, booking info, and FAQs	Prioritize factors with a high coefficient
Deposit Type - Non Refund	Communicate benefits of non-refundable rates clearly on the booking page	Highlight non-refundable options with a "Best Value" badge and a comparison chart	
Previous Cancellations	Target customers with high cancellation rates with personalized email offers	Show special offers for high-cancellation users upon login using personalization	
Customer Type Transient	Offer incentives to transient customers for longer stays	Create special offers or discounts for longer bookings	
Previous Bookings Not Canceled	Create a loyalty program for customers who rarely cancel	Add a loyalty dashboard to track rewards for non-cancellations	
Market Segment - Online TA	Improve online booking experience	Enhance website UX and streamline booking process for online TA users	
Market Segment - Offline TA/TO	Showcase deals for offline travel agency bookings on the website	Include a "Travel Agents" section with exclusive offers and incentives	
Total Special Requests	Improve the special requests section on the booking page	Add fields for special requests and highlight accommodations on the homepage	
Market Segment - Direct	Promote direct bookings with exclusive offers and perks	Highlight direct booking benefits on the homepage and during the booking process	
Lead Time	Offer incentives for early bookings	Introduce early booking discounts or rewards	

# Recommendations to Reduce Cancellations Based on Customer Segments

	Cancellation Rate	Strategy	Reasoning	
Budget-Conscious Travelers	47.90%	<ul style="list-style-type: none"><li>Offer more significant discounts for non-refundable bookings</li></ul>	<ul style="list-style-type: none"><li>Budget-conscious travelers have the highest cancellation rate. Offering more significant non-refundable discounts and budget packages will reduce cancellations</li></ul>	We should focus on this segment to reduce cancellations most effectively
		<ul style="list-style-type: none"><li>Provide budget-friendly packages with limited services</li></ul>		
Leisure Travelers	25.30%	<ul style="list-style-type: none"><li>Increase flexibility in booking changes without penalty</li></ul>	<ul style="list-style-type: none"><li>Leisure Travelers often request booking changes. By providing more flexibility and fulfilling special requests, we can improve satisfaction and reduce cancellations</li></ul>	
		<ul style="list-style-type: none"><li>Enhance the customers experience by fulfilling more special requests</li></ul>		
Frequent Business Travelers	0.04%	<ul style="list-style-type: none"><li>Optimize the loyalty program</li></ul>	<ul style="list-style-type: none"><li>Frequent Business Travelers rarely cancel, so maintaining this low rate through loyalty programs and booking flexibility will be effective</li></ul>	
		<ul style="list-style-type: none"><li>Offer additional flexibility for business trip booking changes</li></ul>		

# BORCELLE BOOKING DASHBOARD

■ Not Canceled ■ Canceled ☐

## Cancellation Rate

Rate of booking cancellations

37.17%

## Canceled Bookings

Total canceled bookings

44,014

## Total Bookings

Total number of bookings

118,408

## Total Revenue

Total booking revenue

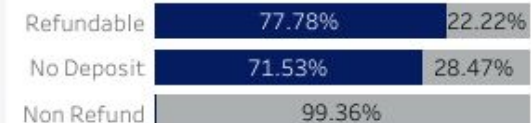
€29.45M

## Total Guests

Number of guest staying

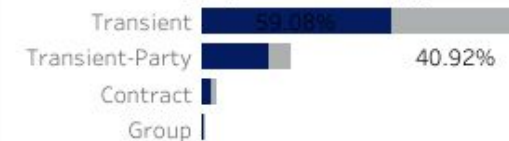
145,041

## Bookings by Deposit Type



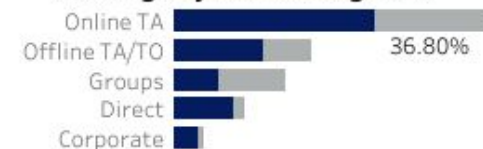
0% 20% 40% 60% 80% 100%  
% of Total Bookings

## Bookings by Customer Type



0K 20K 40K 60K 80K  
Total Bookings

## Bookings by Market Segment

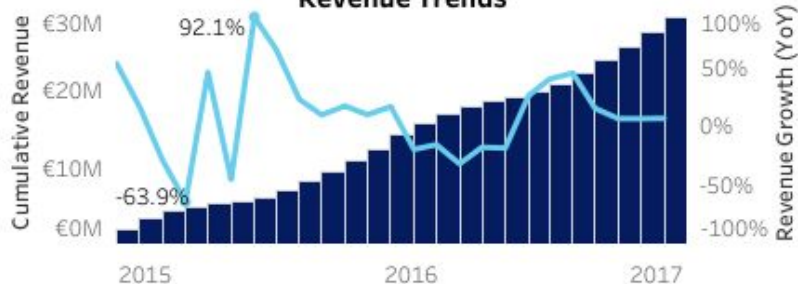


0K 20K 40K  
Total Bookings

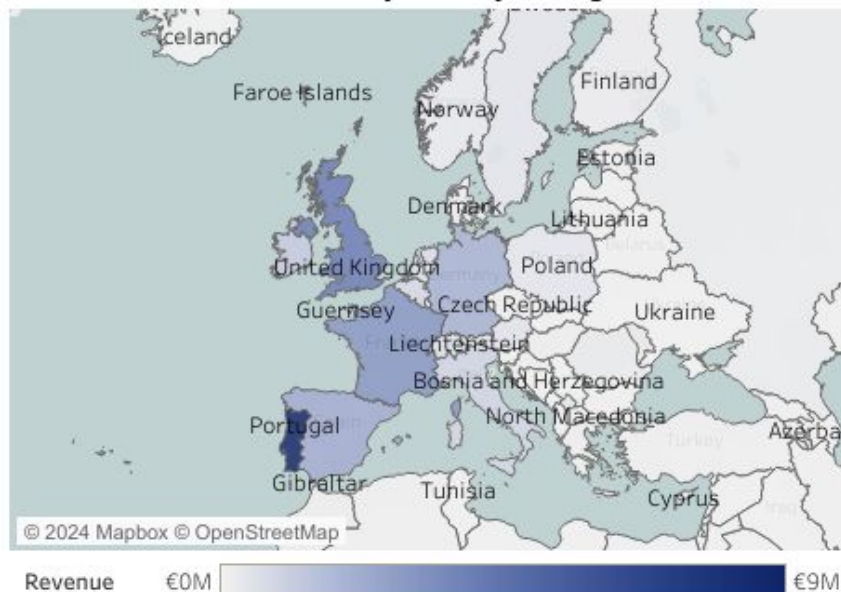
## Booking Trends



## Revenue Trends



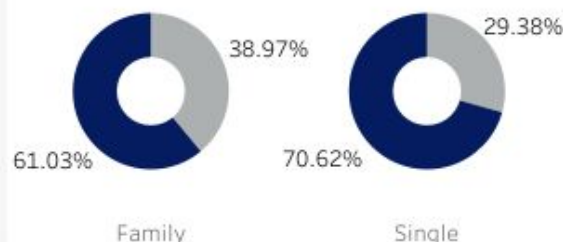
## Revenues by Country of Origin



## Guest Trends



## Bookings by Family Status



## Bookings by Hotel Type



## Cancellation Rate vs Parking Spaces





# Appendix

**Dataset**

<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>

**Google Colab**

[https://colab.research.google.com/drive/12D2pawcONq0EeTztYx6wcaPv\\_795sm9y](https://colab.research.google.com/drive/12D2pawcONq0EeTztYx6wcaPv_795sm9y)

**Linkedin**

<https://www.linkedin.com/in/ade-dwi-aryani/>

**Email**

[adedwiaryan@gmail.com](mailto:adedwiaryan@gmail.com)

**Tableau**

<https://public.tableau.com/app/profile/ade.dwi.aryani/vizzes>



*Thank You*



# Root Cause *Analysis*

**How can we reduce the booking cancellation rate from 36.83% to 20% within the next 12 months (by 2018)?**

## **Factors Affecting Booking Cancellation**

- Insufficient parking availability may lead to cancellations
- Ambiguous deposit requirements and additional fees can cause cancellations
- Customers with a history of cancellations are more likely to cancel again
- Customers who have not canceled previous bookings are less likely to cancel in the future
- Cancellation rates vary across different market segments
- Numerous special requests and frequent booking changes increase the risk of cancellations
- Both excessively long and very short lead times can impact cancellation rates

## **Customer Behavior and Preferences**

- Price-sensitive customers often cancel to seek better deals
- Frequent cancellers are likely to cancel again
- Lead time and multiple booking changes impact cancellation risk

# Borcelle Hotel Overview

Period: July 2015 - August 2017



## Revenue Trends and Key Fluctuations



## Average Daily Rate Trends



- The **hotel's revenue grew steadily** from April 2016 to August 2017, peaking at **€29 million**, despite significant fluctuations in October 2015 and November 2016.
- This growth aligns with the **rising average daily rates (ADR)**, which are highest in peak months like July and August, while lower ADR in off-peak months causes revenue fluctuations.

# Borcelle Hotel Overview

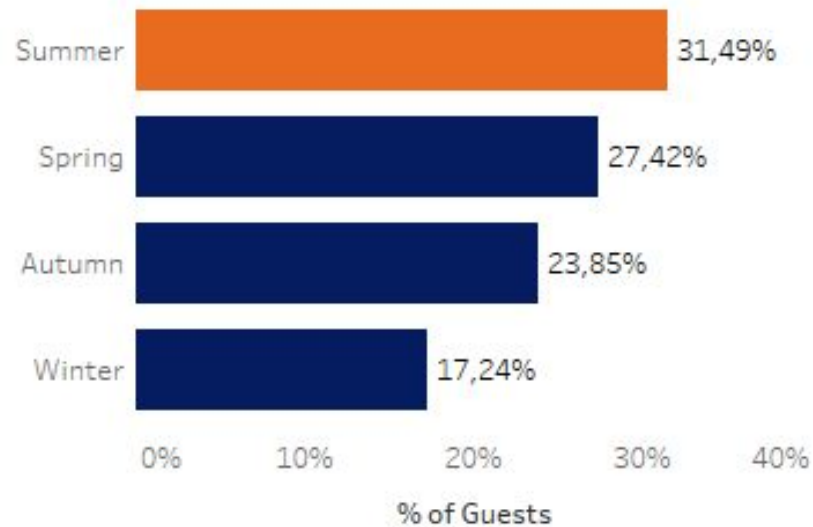
Period: July 2015 - August 2017



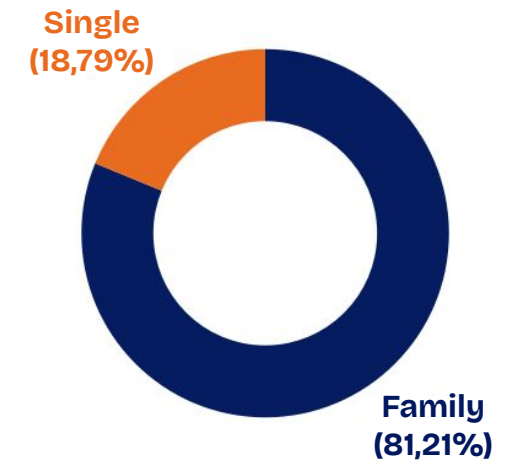
## Guests Trends



## % Distribution of Guests by Season



## Guests by Family Status



- The hotel consistently experienced **an increase in guest numbers**, peaking during the **summer months**.
- This seasonal pattern aligned with guest preferences, as **summer was the most popular season**, while winter had the fewest guests.
- Additionally, the **majority of guests were families**, comprising approximately 81% of the total, while solo travelers accounted for around 19%.



# Data Cleaning

## Initial Data Assessment

**Raw Dataset:**  
Rows: 119,390  
Columns: 36

## Remove Irrelevant Columns

**Dropped Columns:**  
agent, company,  
phone-number,  
credit\_card

## Handle Missing Values

**Affected Columns:**  
children, country  
Approach: removal of  
rows with missing values

## Convert Data Types

- Children: From float to integer
- Reservation Status Date: Converted to datetime format

## Final Dataset Preparation

Rows: 118,408  
Columns: 33

## Verify and Clean

**Duplicates:**  
No duplicates found  
Ensure data consistency

## Handle Noisy Data

- Removed negative ADR values
- Removed zero adults value
- Removed rows with excessively high numbers of children/babies (>10)

**Ready for Analysis**

**The features used for analysis are:**

Hotel, lead time, stays in weekend nights, stays in week nights, adults, children, babies, market segment, is a repeated guest, previous cancellations, previous bookings not canceled, reserved room type, booking changes, deposit type, days in waiting list, customer type, average daily rate (ADR), required car parking spaces, total number of special requests, and whether the booking is canceled.

# Logistic Regression

## Classification Report

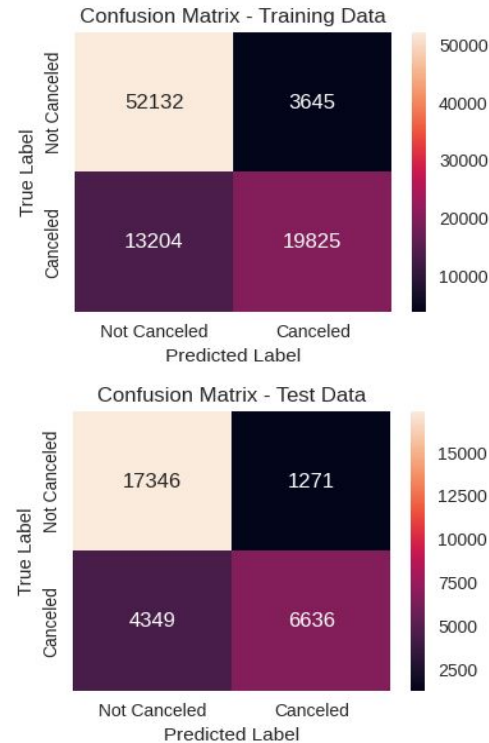
Classification Report - Training Data (Logistic Regression):				
	precision	recall	f1-score	support
0	0.80	0.93	0.86	55777
1	0.84	0.60	0.70	33029
accuracy			0.81	88806
macro avg	0.82	0.77	0.78	88806
weighted avg	0.82	0.81	0.80	88806

Classification Report - Test Data (Logistic Regression):				
	precision	recall	f1-score	support
0	0.80	0.93	0.86	18617
1	0.84	0.60	0.70	10985
accuracy			0.81	29602
macro avg	0.82	0.77	0.78	29602
weighted avg	0.81	0.81	0.80	29602

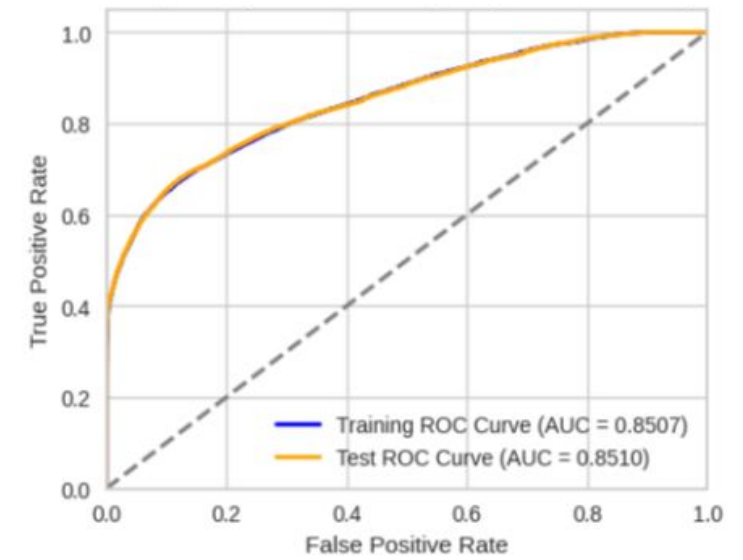
- The model has an overall accuracy of 81% on both training and testing data.
- Precision for class 0 (Not Canceled) is better on testing data, while precision for class 1 (Canceled) is better on training data.
- Overall, the model performs well with minor differences between training and testing.

## Confusion Matrix



- Model performs well in predicting "Not canceled but struggles with "Canceled" (cancellation).
- Despite high accuracy, it could improve in handling cancellations, especially on testing data

## Receiver Operating Characteristic (ROC) Curve



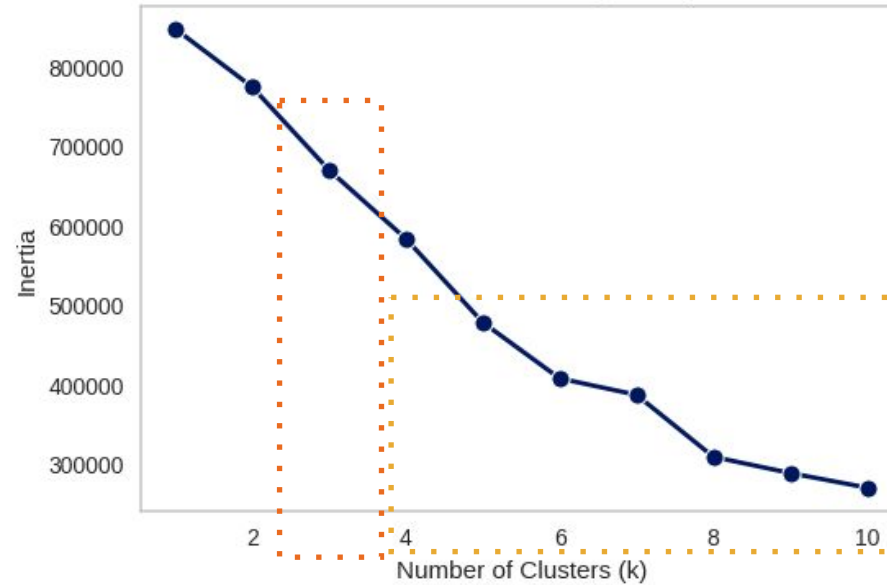
- ROC curve approaching the upper left corner indicates good performance.
- AUC close to 0.85 shows that the model has good ability to distinguish between bookings that will be canceled and those that won't.
- ROC curve better than the diagonal line (AUC = 0.5) indicates that the model performs better than random guessing.

# K-Means Clustering

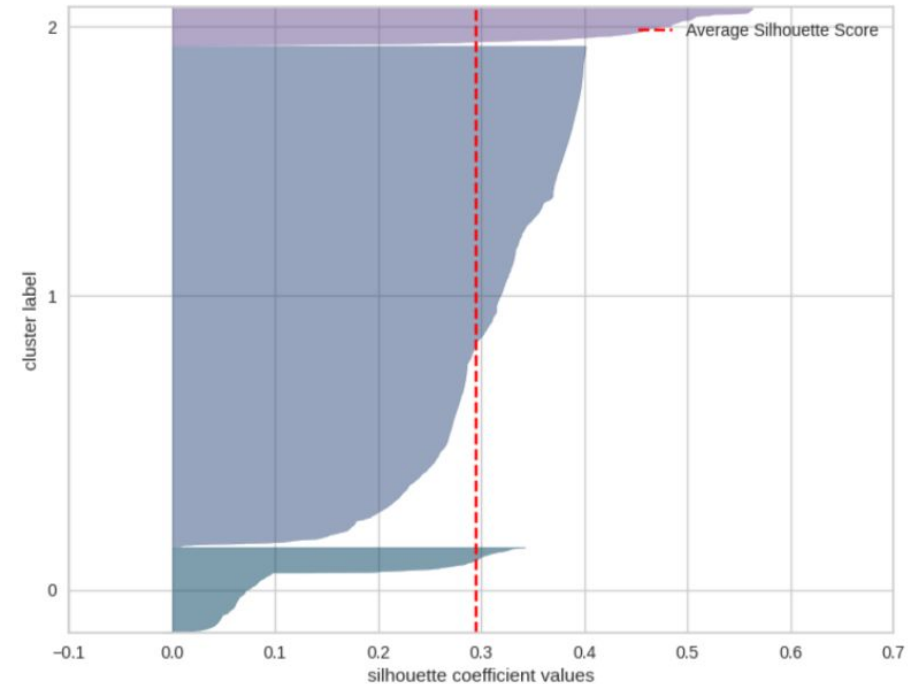
Inertia and  
Silhouette Scores

k	Inertia	Silhouette Score
2	774,582.14	0.2737
3	669,951.42	0.2947
4	583,440.54	0.2461
5	478,103.03	0.2549
6	407,597.42	0.2593
7	386,852.39	0.2772
8	308,863.88	0.2288
9	288,415.59	0.2473

The Elbow Method  
Showing the Optimal k



Silhouette Plot of  
KMeans Clustering in 3 Centers



- **Choosing  $k = 3$**

Although inertia decreases significantly up to  $k = 3$  and continues to decrease at a slower rate with additional clusters, **the highest Silhouette score of 0.2947 for  $k = 3$**  strongly supports this choice. This score indicates the best balance of cluster separation and cohesion, reflecting that  $k = 3$  results in **well-defined, distinct clusters with optimal internal consistency**.