

Investigate Business Hotel using Data Visualization

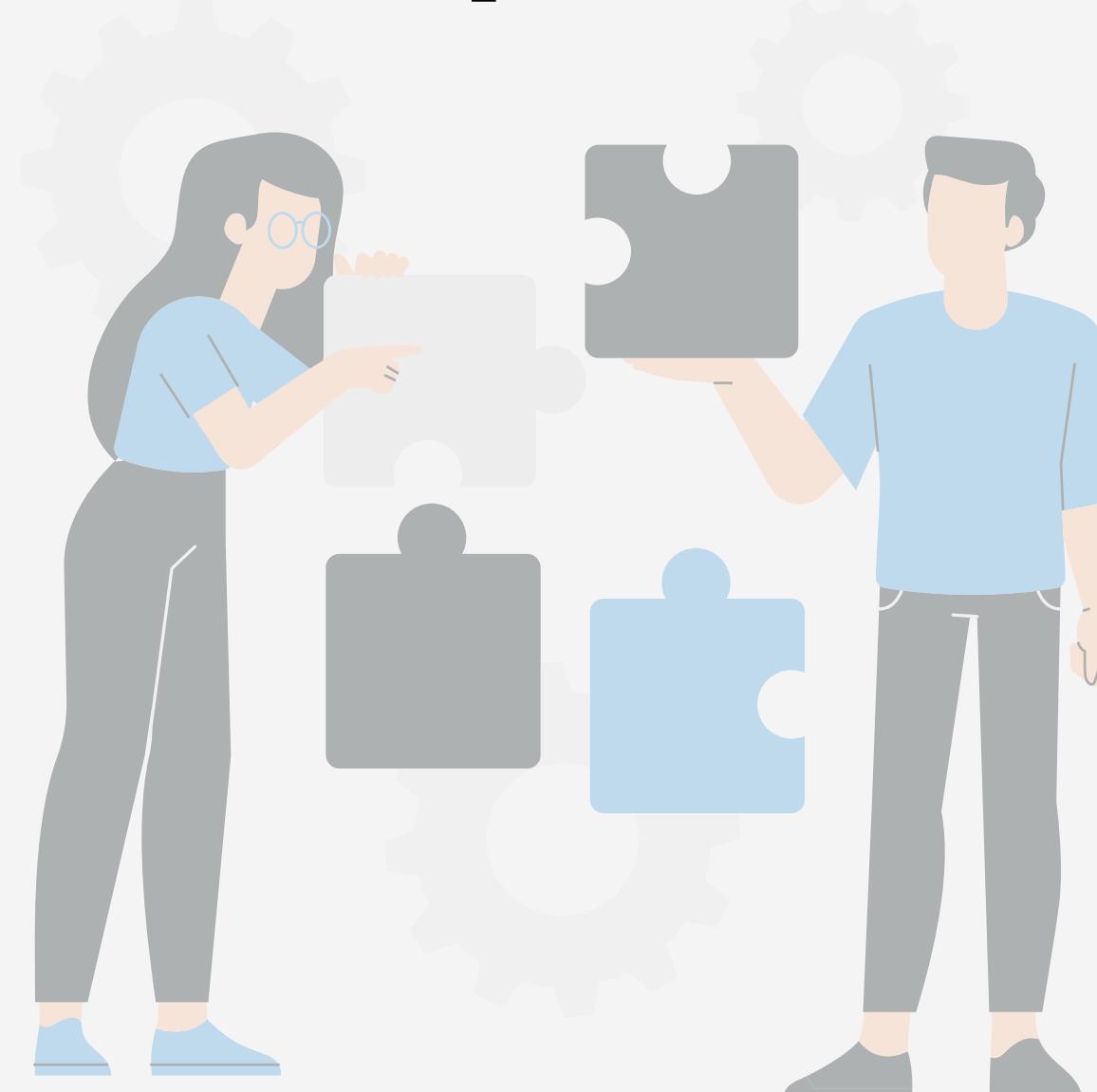
Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com



Created by:
Lutfia Husna Khoirunnisa
Your Email:
lutfiahusnakhoirunnisa@gmail.com
Your linkedIn Profile
linkedin.com/in/lutfiahk

“Lutfi is a junior data analyst experienced on data analysis, business analysis, and data science with a background in mathematics. Experienced in handling and interpreting diverse data sets, extracting valuable insights, and making datadriven recommendations ”

Step by Step



1

Project Introduction

2

Problem Statement

3

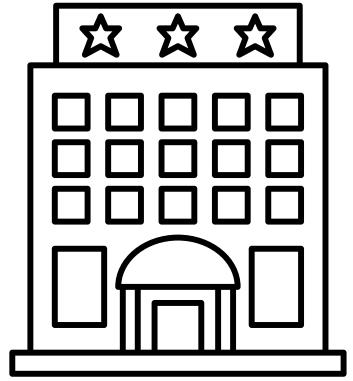
Data Overview and Cleansing

4

Data Analysis

5

Conclusion



Project Introduction

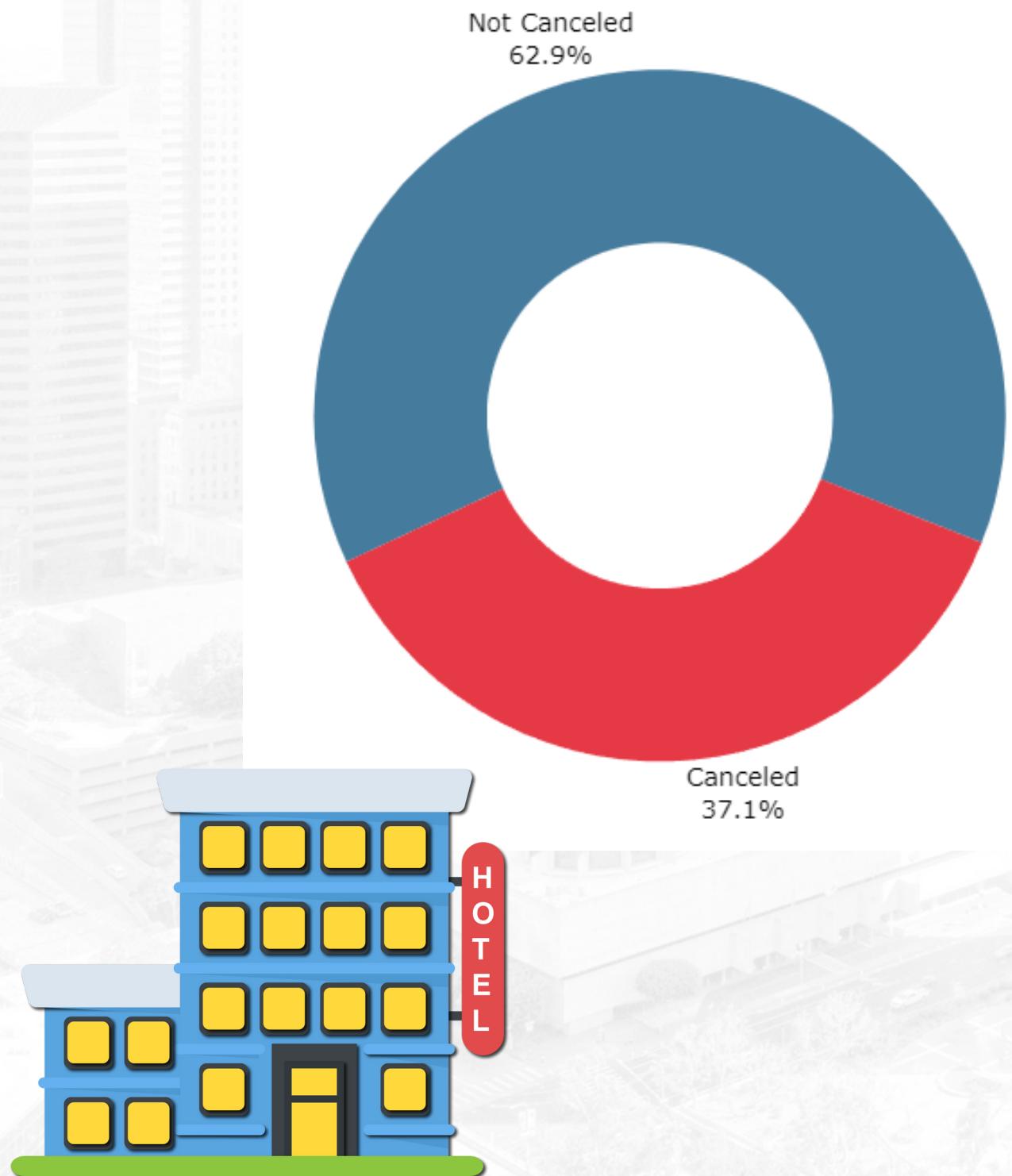
"It is crucial for a company to constantly analyze its business performance. On this project, we will analyze deeper into the business in the hotel industry"

Our focus is to understand the behavior of our customers when making hotel reservations and its relationship to the cancellation rate of hotel bookings. The insights we discover will be presented in the form of data visualization to make it easier to understand and more persuasive.



All data visualizations in this project are built with Python Plotly

Problem Statement



37.1%
of the bookings have
been canceled.

"If the cancellation rate of hotel bookings continues to increase, several consequences can occur, such as revenue loss that can lead to financial and operational problem."

"Therefore, the company needs to develop a strategy to reduce the cancellation rate."

Data Overview & Cleansing



This project will use a dataset that contains booking information from resorts and hotels from 2017 until 2019, including the number of guests, arrival time, reservation status, and more. The dataset consists of 119,390 records and 30 features.



The dataset is not yet clean as it contains missing values, incorrect values, and unused records that require handling

access the dataset and notebook [here](#)

column	missing_value
children	4
city	488
agent	16340
company	112593

Data Overview & Cleansing

column	missing_value
children	4
city	488
agent	16340
company	112593

There are several columns that have missing values, so we will handle the missing values as follows:

- For the "children" column, we will fill the missing values with 0.
- For the "city" column, we will fill the missing values with 'others'.
- For the "agent" column, we will fill the missing values with 0, indicating that no travel agent made the booking.
- For the "company" column, we will fill the missing values with 0, indicating that no company made the booking or is responsible for paying the booking.

```
set(df['meal'])
✓ 0.1s
{'Breakfast', 'Dinner', 'Full Board', 'No Meal', 'Undefined'}
```

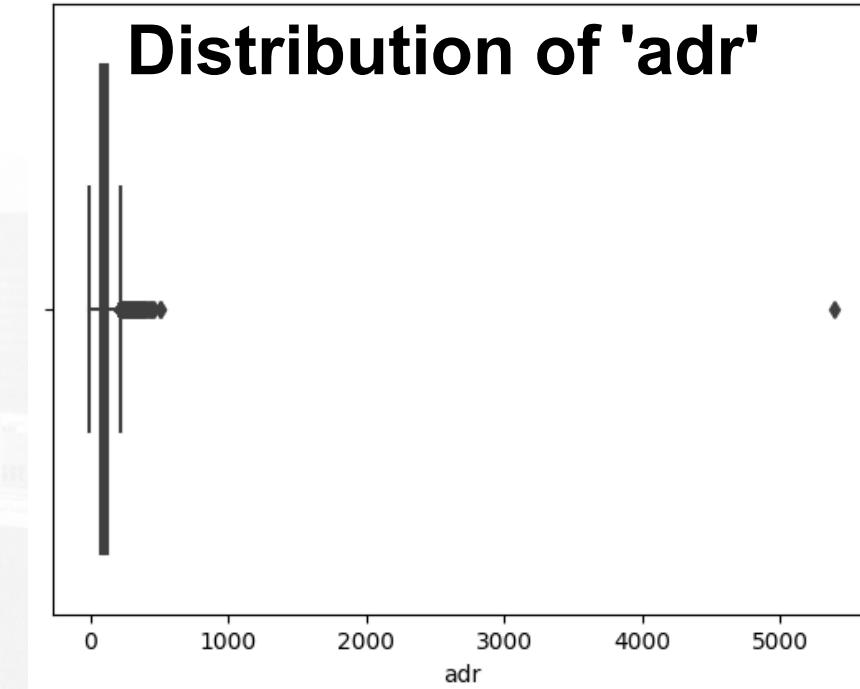
There is incorrect data input in the 'meal' column. We will replace the 'undefined' data with 'No meal'.

Data Overview & Cleansing

We will drop the data records that have extremely high and negative 'adr' value.

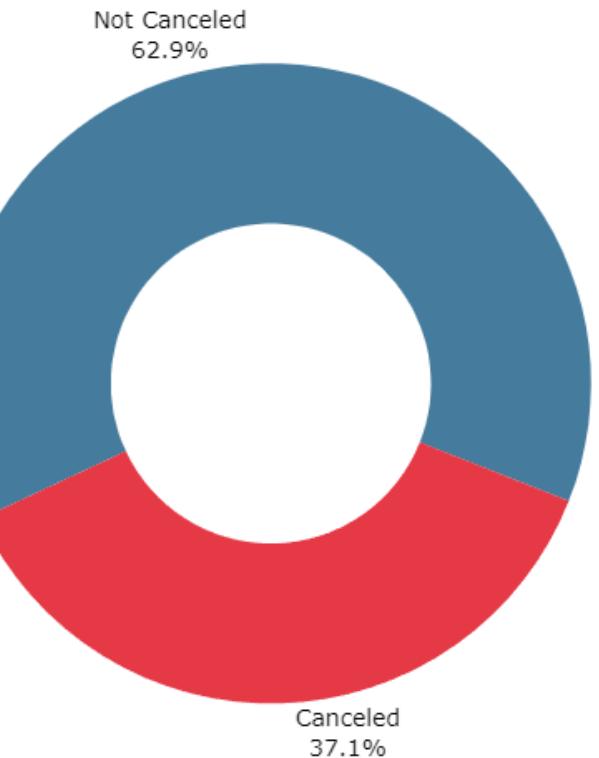
	hotel	is_canceled	total_guest
2224	Resort Hotel	0	0
2409	Resort Hotel	0	0
3181	Resort Hotel	0	0
3684	Resort Hotel	0	0
3708	Resort Hotel	0	0
...
115029	City Hotel	0	0
115091	City Hotel	0	0
116251	City Hotel	0	0
116534	City Hotel	0	0
117087	City Hotel	0	0

180 rows × 3 columns



180 bookings are not canceled but do not have any guests, so we will also drop these records

Data Analysis



37.1% of the 119,209 bookings have been canceled.

The majority of customers is a **new customer**, and the number of canceled booking from new customer is higher than from the repeated customer.

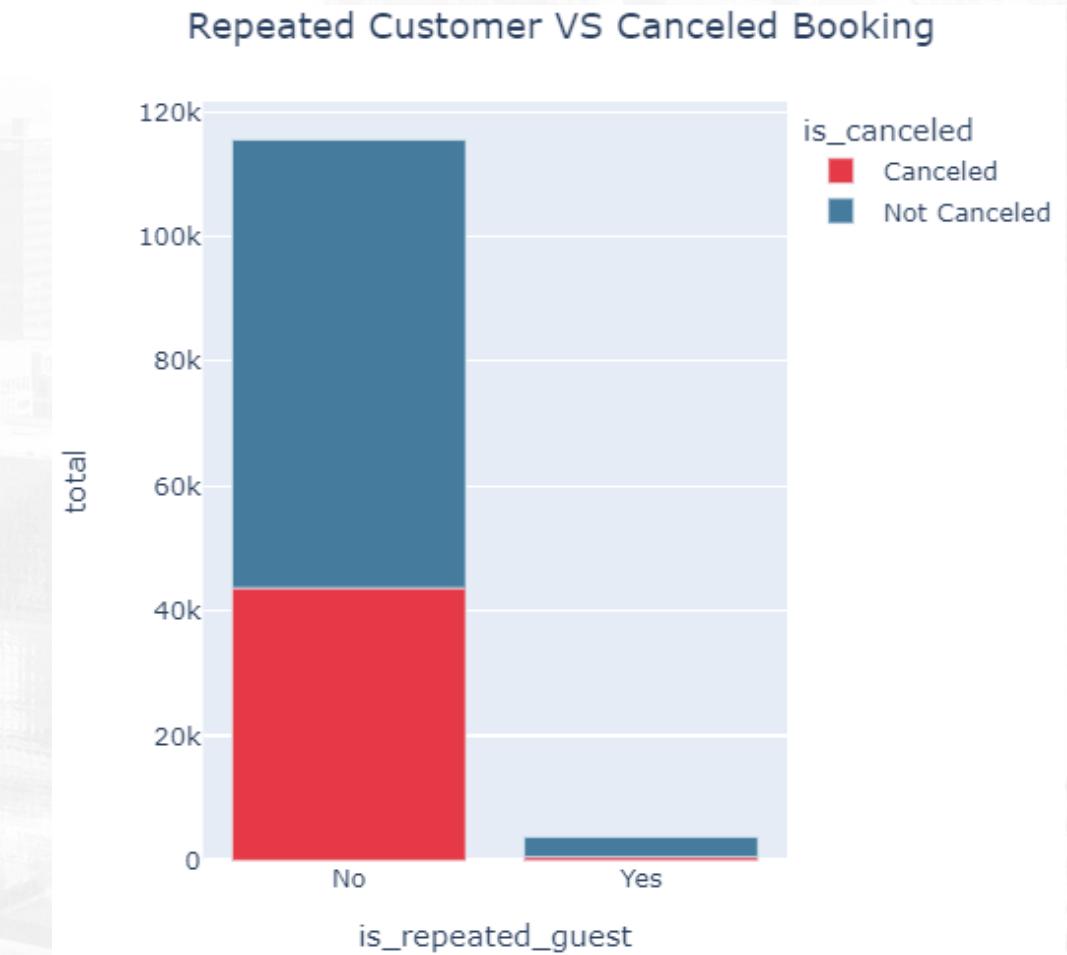
The z-proportion test revealed a significant difference between the new customer and repeated guest groups in terms of canceled bookings, with new customers having the highest proportion at 38% of total new customers.

Z-statistic: 28.91

p-value: 0.0

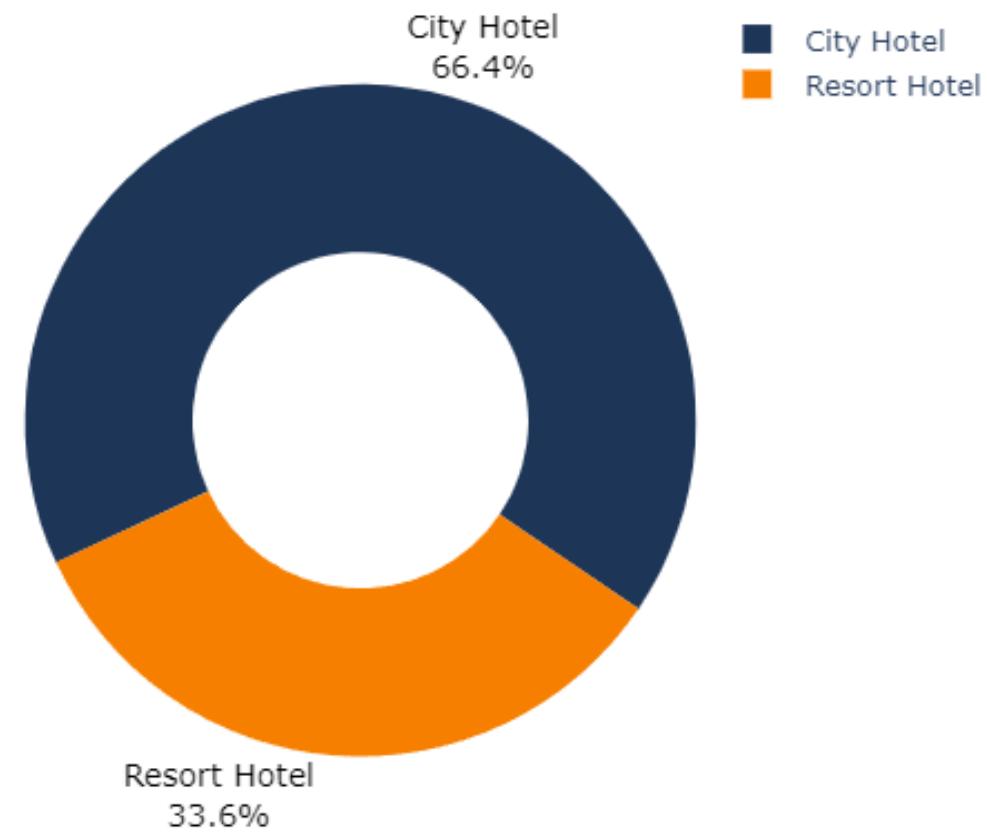
The difference in proportions between the two groups is statistically significant.

Canceled Booking of New Customer has the highest proportion: 0.38



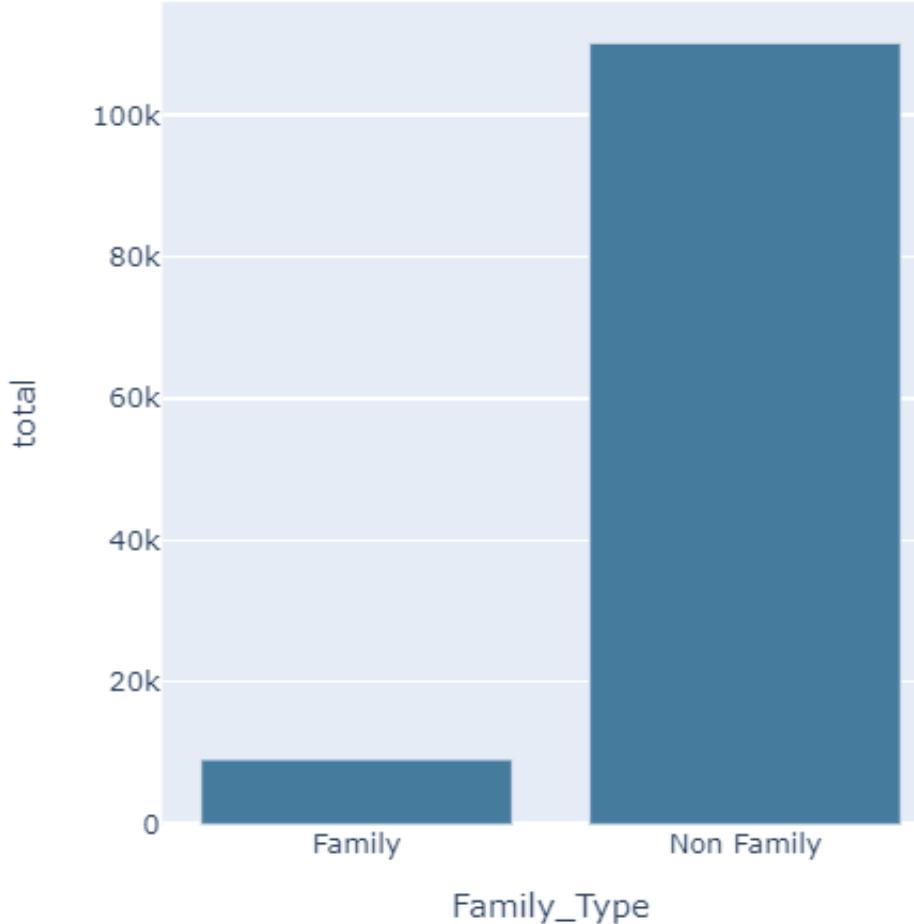
Data Analysis

Canceled Booking



66.4% of bookings are for City Hotels, which are usually found in urban areas, providing convenient access to business districts, shopping centers, and transportation hubs. Resort hotels, on the other hand, are situated in scenic destinations like beaches or mountains, offering natural surroundings and recreational activities.

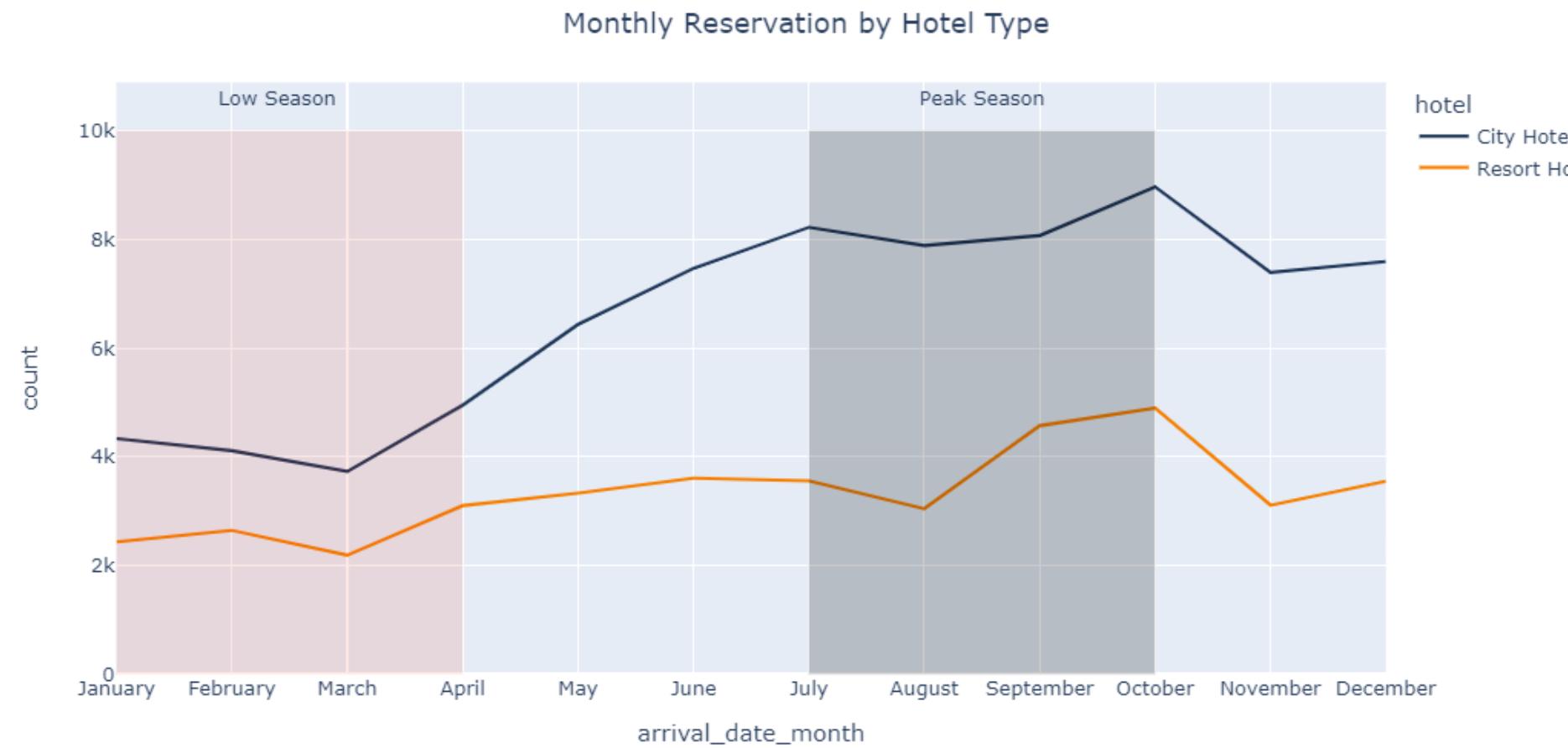
Distribution of Family Type



This makes sense because the majority of customers are not families, indicating that there may be more customers using hotels for business travel or reasons other than vacation purposes.

Data Analysis

Monthly Hotel Booking Analysis Based on Hotel Type

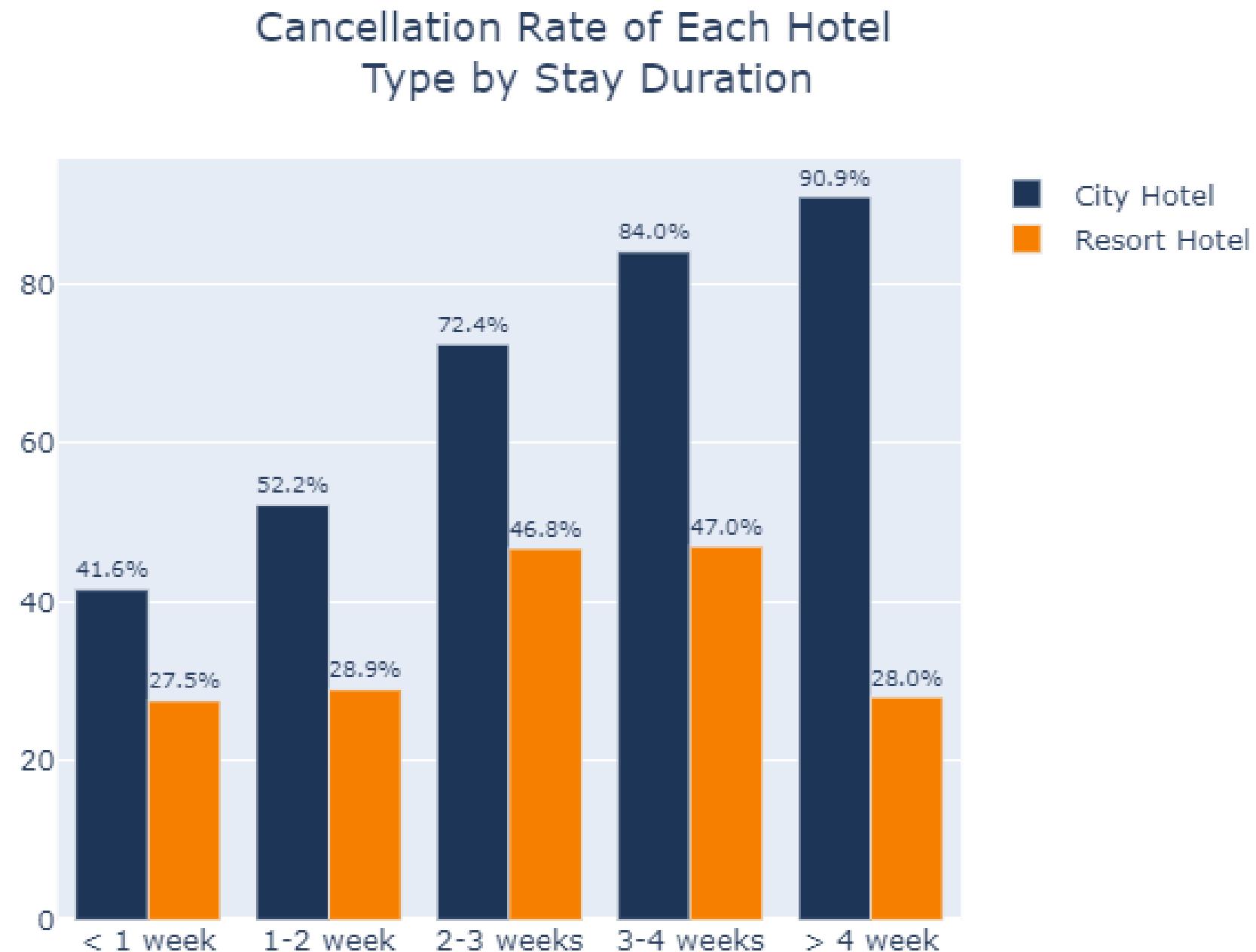


High booking numbers are around the months of July until October. During the year, city hotels are booked more frequently than resort hotels.

As we know, July to October is not the holiday season, and the majority of customers are not families. Therefore, it is likely that many customers are traveling for business purposes or reasons other than vacation. Moreover, city hotels are typically located in the center of the city where many offices and economic activities take place.

Data Analysis

Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates

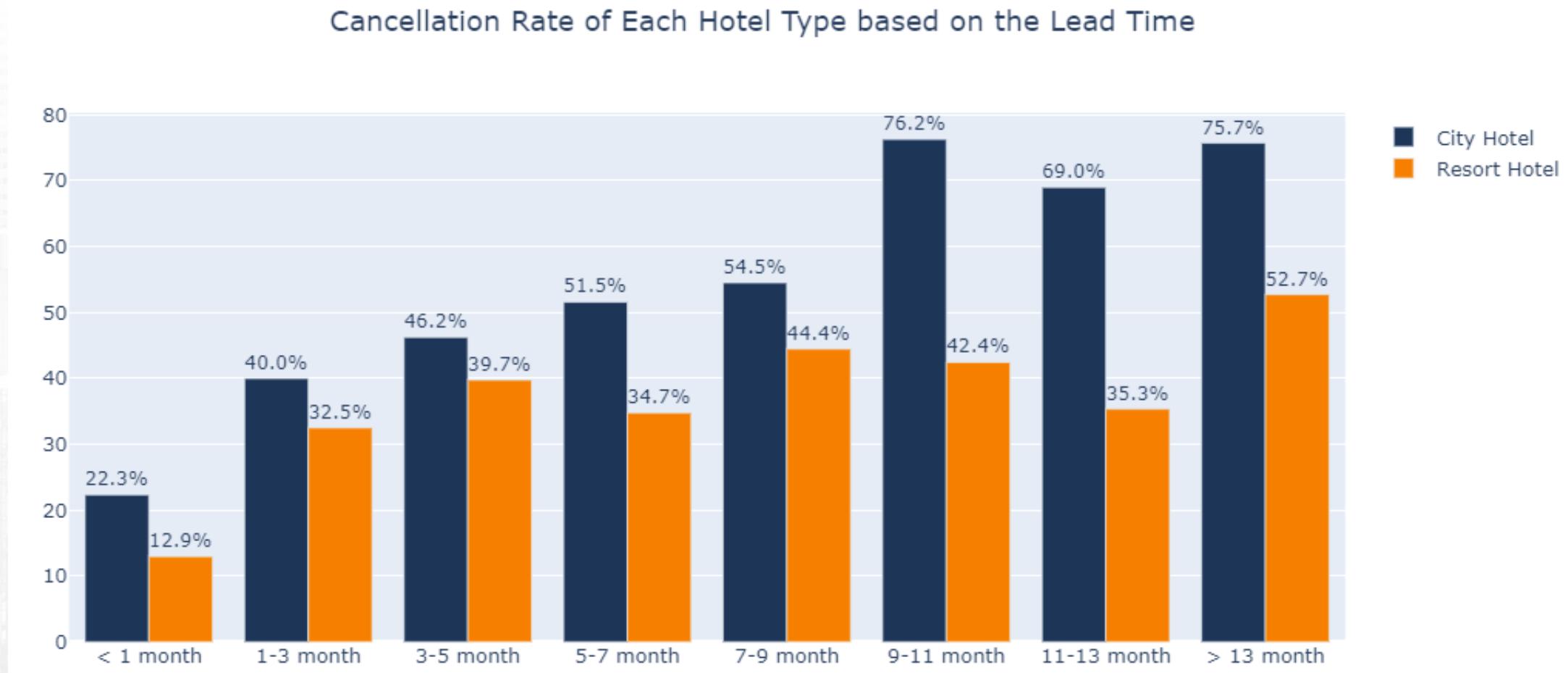


Basically, the cancellation rate of the city hotel is higher than the resort hotel. It is known that the longer the stay duration of customers, the higher the cancellation rate at the city hotel. The cancellation rate of the resort hotel tends to be consistent, ranging from 27% to 45%.

"We need to highlight the cancellation rate of the city hotel. The cancellation rate for bookings with a stay duration of more than 2 weeks is over 50%, which is a high number. "

Data Analysis

Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate



The cancellation rate of the city hotel for each lead time is higher than that of the resort hotels. As we know, the lead time of more than 9 months is particularly high, especially for the city hotel that reach above 70%.

Therefore, the company needs to develop a strategy to decrease the cancellation rate by considering offering special packages or deals specifically for bookings with high lead time. This can provide added value to customers and encourage them to proceed with their bookings.

Conclusion



The majority of customers are **non-family customers**, who book city hotels which are mainly located in the center of the city. The majority of the customer tends to be on non-vacancy purpose.

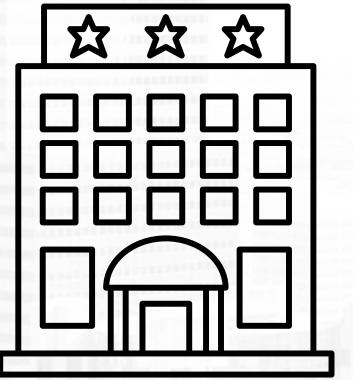
The company needs to make strategies or marketing that **refer to the needs and preferences** of these customers.



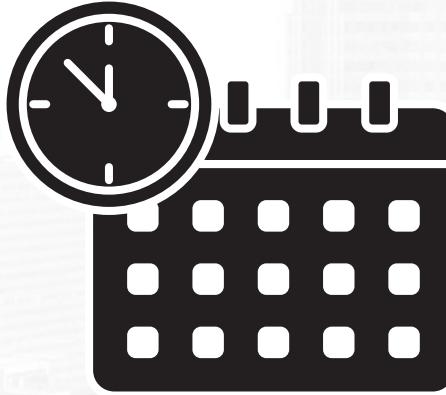
Above 90% of the customers are new customers. Having a high number of new customers is a positive indication of the company's ability to attract new business, but the company needs to focus on building customer loyalty.

Therefore, the company should implement strategies that prioritize **customer retention**, such as loyalty programs, personalized experiences, and excellent customer service, to ensure a strong foundation of loyal customers and sustained business growth.

Conclusion



For the city hotel, the longer the stay duration, the higher the cancellation rate tends to be. Therefore, the company needs to implement promotions to decrease the cancellation rate, such as offering discounts for customers who stay for more than 2 weeks.



As we know the longer the lead time, the cancellation rate tends to be higher, the company needs to develop a strategy to decrease the cancellation rate by considering offering special packages or deals specifically for bookings with high lead time. This can make customers change their minds and encourage them to proceed with their booking.

Mini Project

Completion Certification

has been presented to

LUTFIA HUSNA KHOIRUNNISA

For successfully completing data scientist
Mini Project Investigate Hotel Business using Data Visualization

CEO Rakamin Academy



Andika Deni Prasetya



Data Scientist,
Bukalapak



Figarri Keisha

Thank you

Check out my profile and other portfolios:



Linkedin



GitHub