



INFRINGEMENT DETECTION ON AIRBNB

Capstone Project
04.05.2023





TEAM

**SUPERPOWERS:
STRAIGHT THINKING
AGILE BY NATURE**

Johanna Krüger



Dr. Krisztina Nagy



Uta Viegener





INTRO



AIRBNB



INTRO

WHAT IS AIRBNB?

worldwide-used online marketplace for homestays with an increasing share of the accommodation market

AIRBNB



INTRO

WHAT IS AIRBNB?

worldwide-used online marketplace for
homestays with an increasing share of the
accommodation market

900.000.000

AIRBNB





INTRO

WHAT IS AIRBNB?

worldwide-used online marketplace for homestays with an increasing share of the accommodation market

900.000.000

overnight stays were booked through Airbnb between 2008 and 2020

AIRBNB



INTRO

WHAT IS AIRBNB?

worldwide-used online marketplace for homestays with an increasing share of the accommodation market

900.000.000

overnight stays were booked through Airbnb between 2008 and 2020

CONCERNS

short-term rental puts pressure on housing market

non-declared incomes, lack of taxes being collected

overtourism and public danger

AIRBNB





INTRO

DATA



INTRO

WHAT IS INSIDE AIRBNB?

mission driven activist project

provide scraped data of Airbnb listings

aim to understand the impact of short-term rentals
on housing and residential communities

DATA



INTRO

DATA

WHAT IS INSIDE AIRBNB?

mission driven activist project

provide scraped data of Airbnb listings

aim to understand the impact of short-term rentals on housing and residential communities

"CONSIDER THE MISSION OF THE PROJECT AND HOW YOUR USE OF THE DATA CAN CONTRIBUTE TO IT. DO GOOD WITH THE DATA."



INTRO



MISSION

CHECK-IN



INTRO

EXPLORE



types of accommodations offered on Airbnb

understand who are the hosts and what do they offer

investigate the length of stay, specifically the minimum nights of stay



INTRO

EXPLORE



type of accommodations offered on Airbnb

understand who are the hosts and what do they offer

investigate the length of stay, specifically the minimal nights stays

MODEL

create a model that detects potentially fraudulent listings in one of these cities



MISSION

CHECK-IN

type of accommodations offered on Airbnb

understand who are the hosts and what do they offer

investigate the length of stay, specifically the minimal nights stays

MODEL

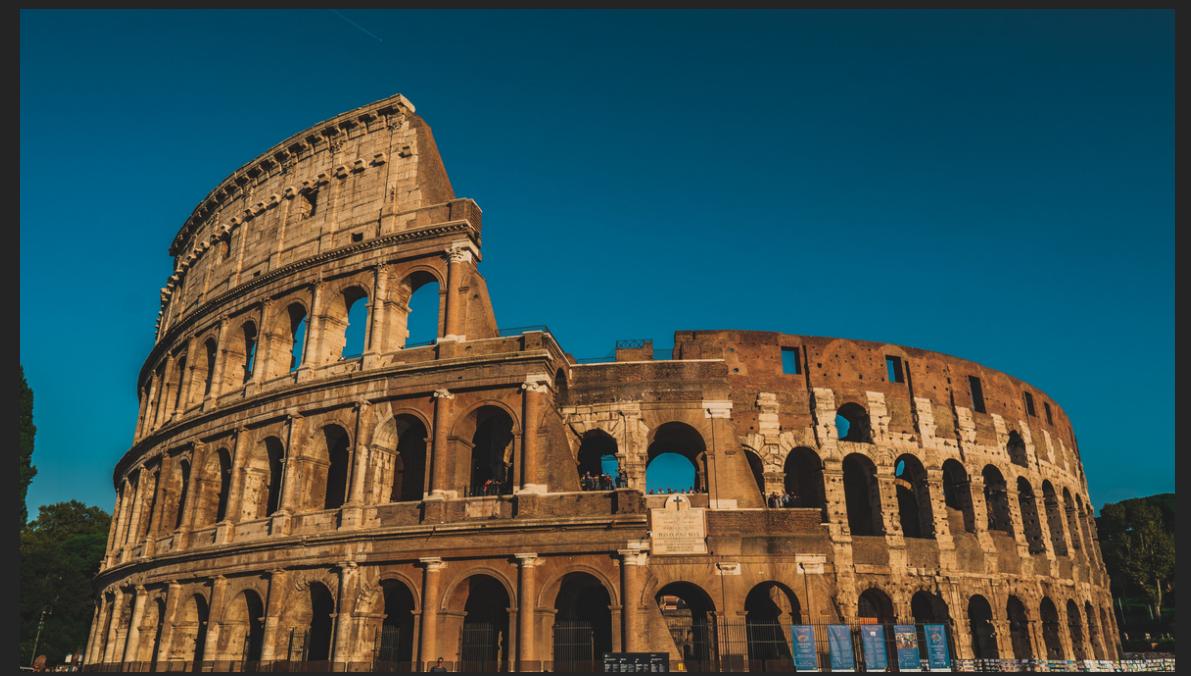
create a model that detects potentially fraudulent listings in one of these cities

STAKEHOLDER

city council of Barcelona



ANALYSIS





ANALYSIS

STOCKHOLM



ROME



BARCELONA





ANALYSIS

COMPARATIVE ANALYSIS

focus on European cities with similar Covid related touristic regulations

well-known touristic targets

reported problems due to short-term rentals





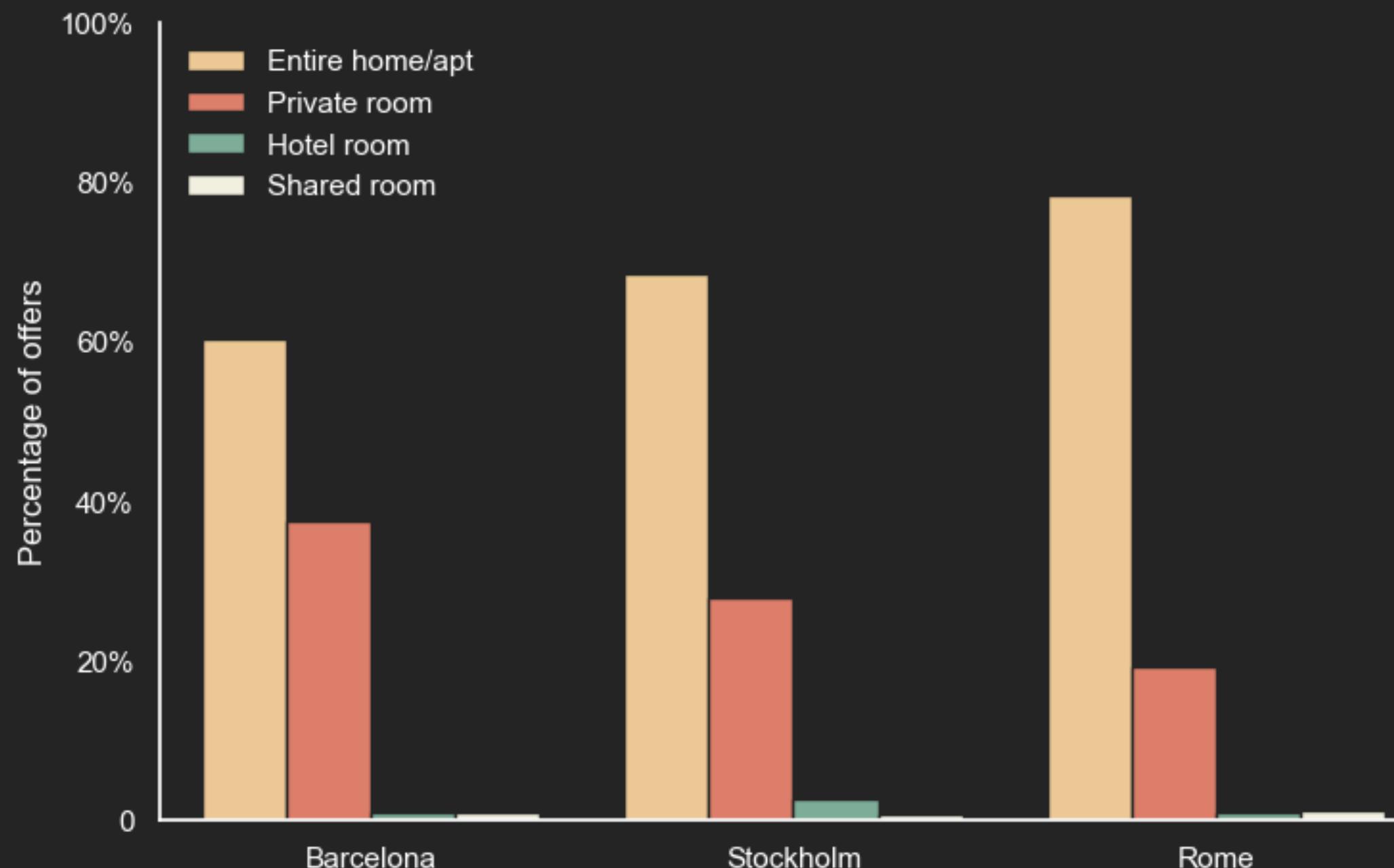
ANALYSIS

ACCOMMODATION TYPES



ANALYSIS

ACCOMMODATION TYPES

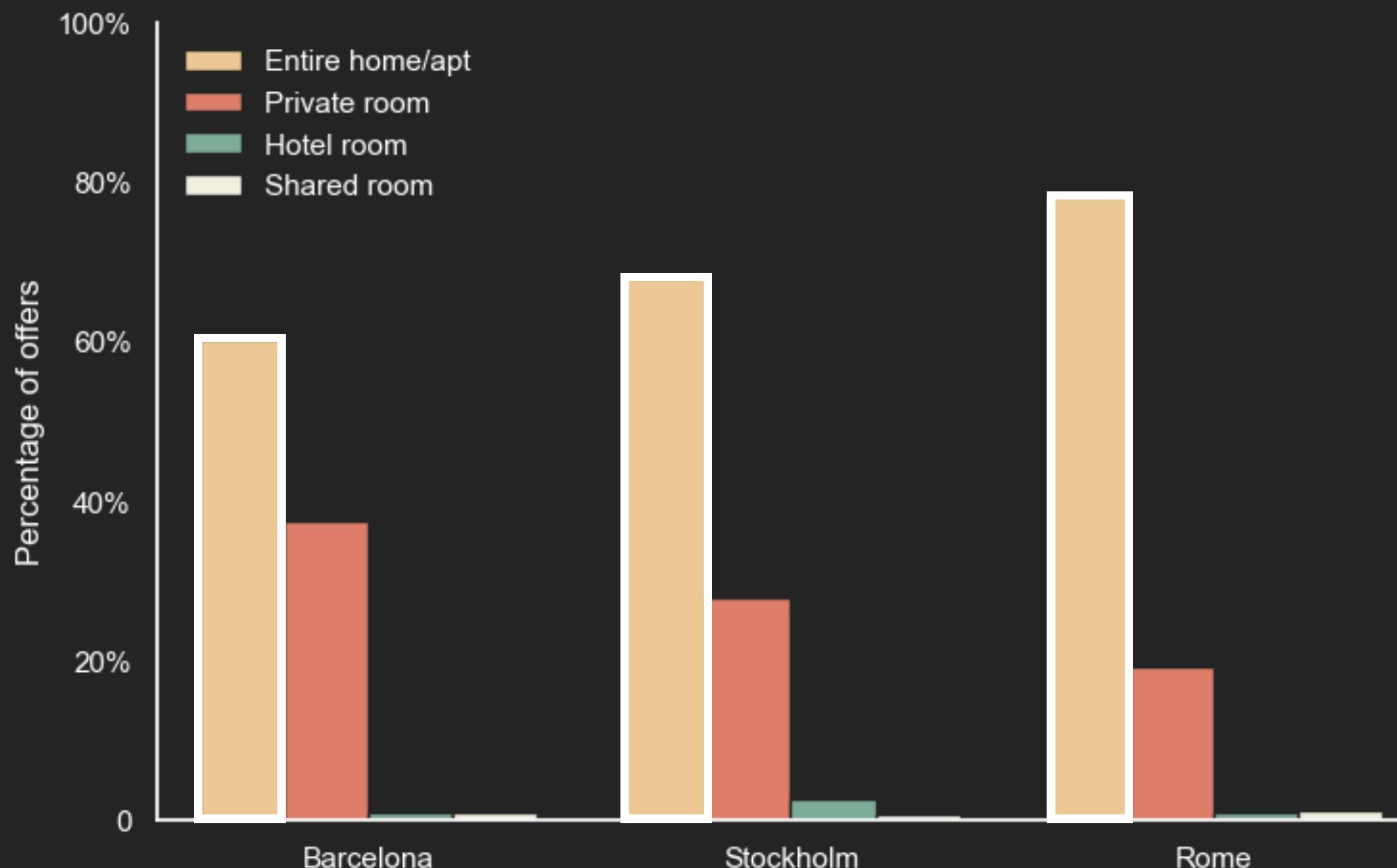




ANALYSIS

ACCOMMODATION TYPES

In all cities the highest proportion of the accommodation types is entire apartments





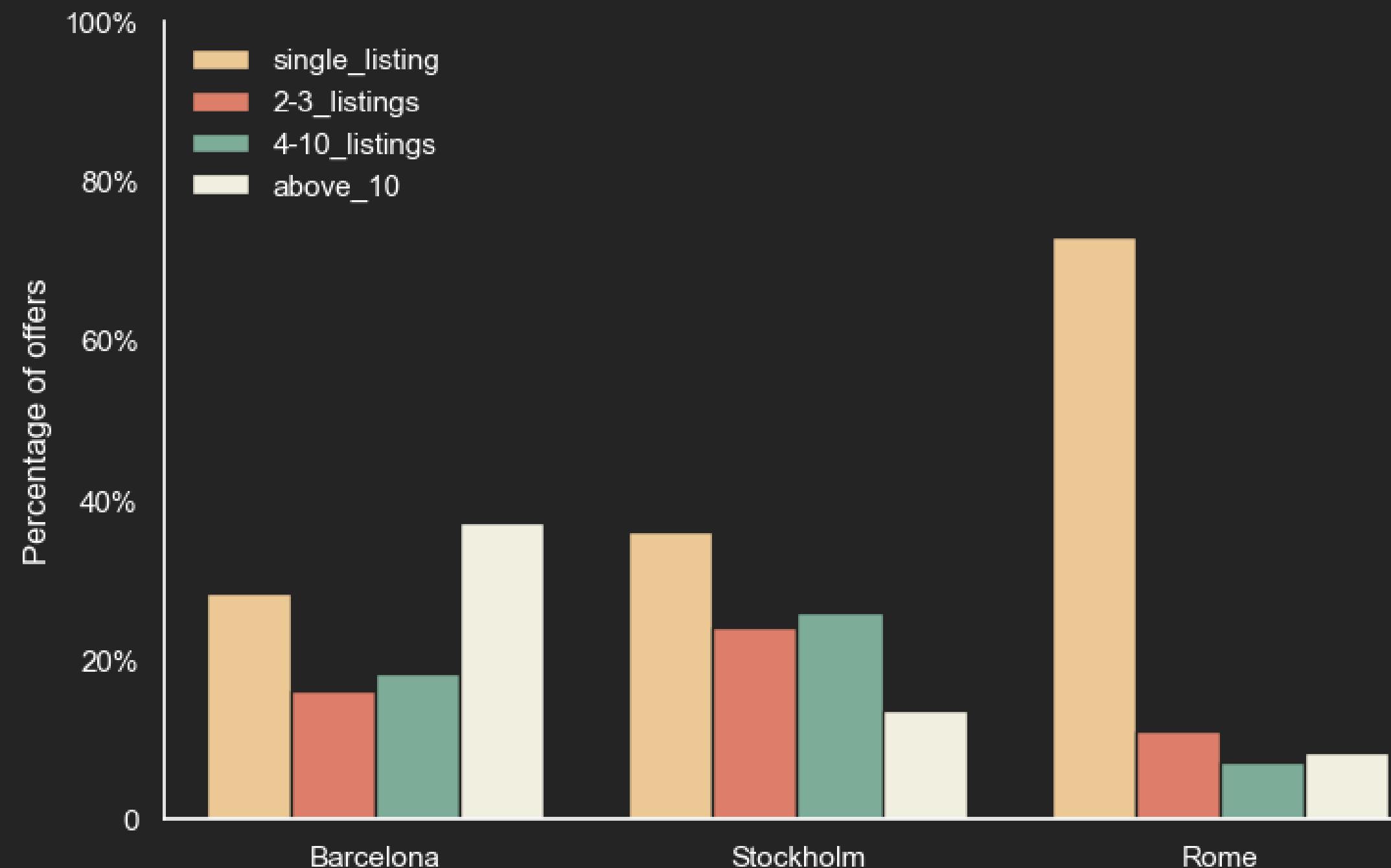
ANALYSIS

HOST TYPES



ANALYSIS

HOST TYPES

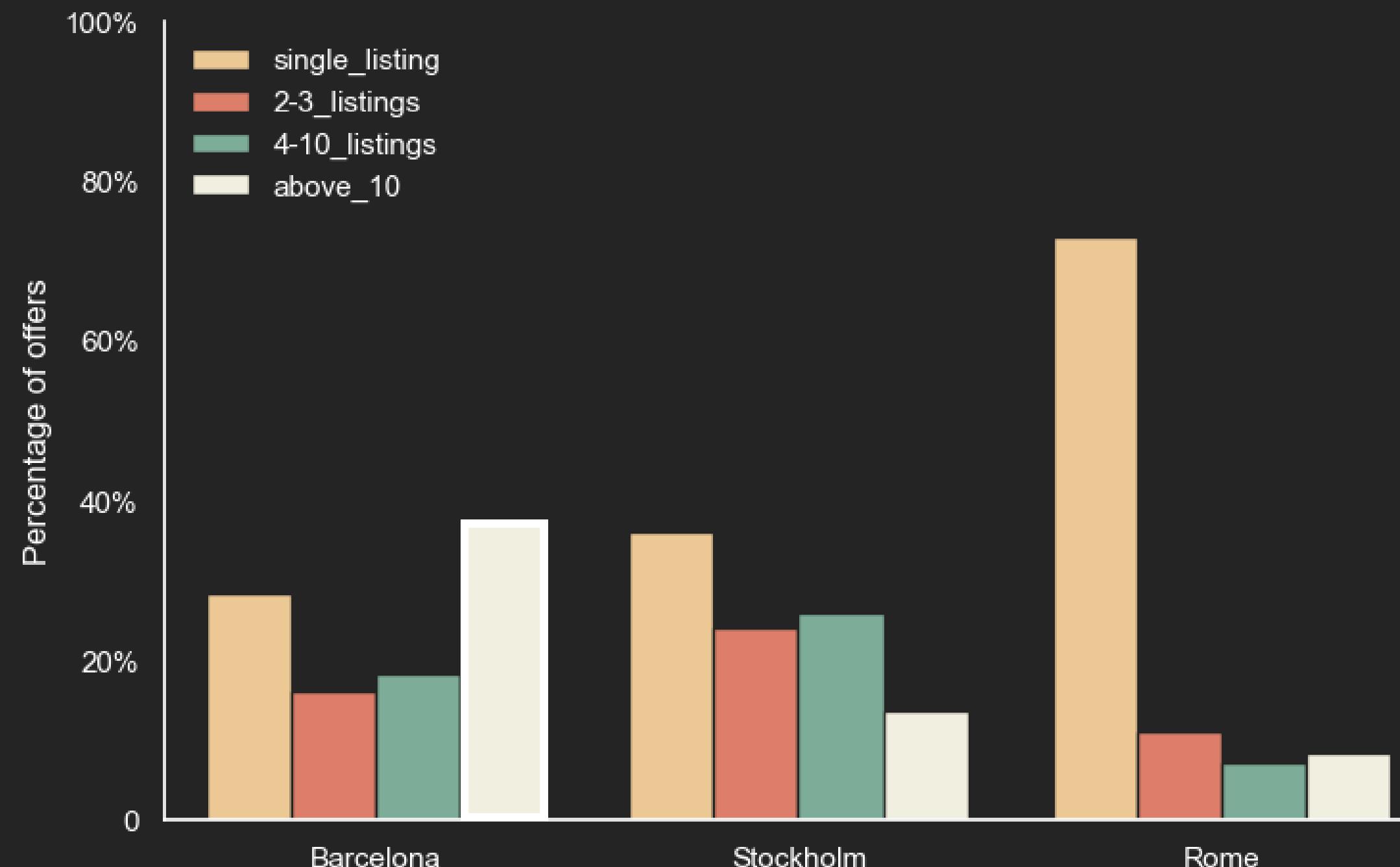




ANALYSIS

HOST TYPES

In all cities there are business providers with more than 10 offers, the proportion of them is the highest in Barcelona





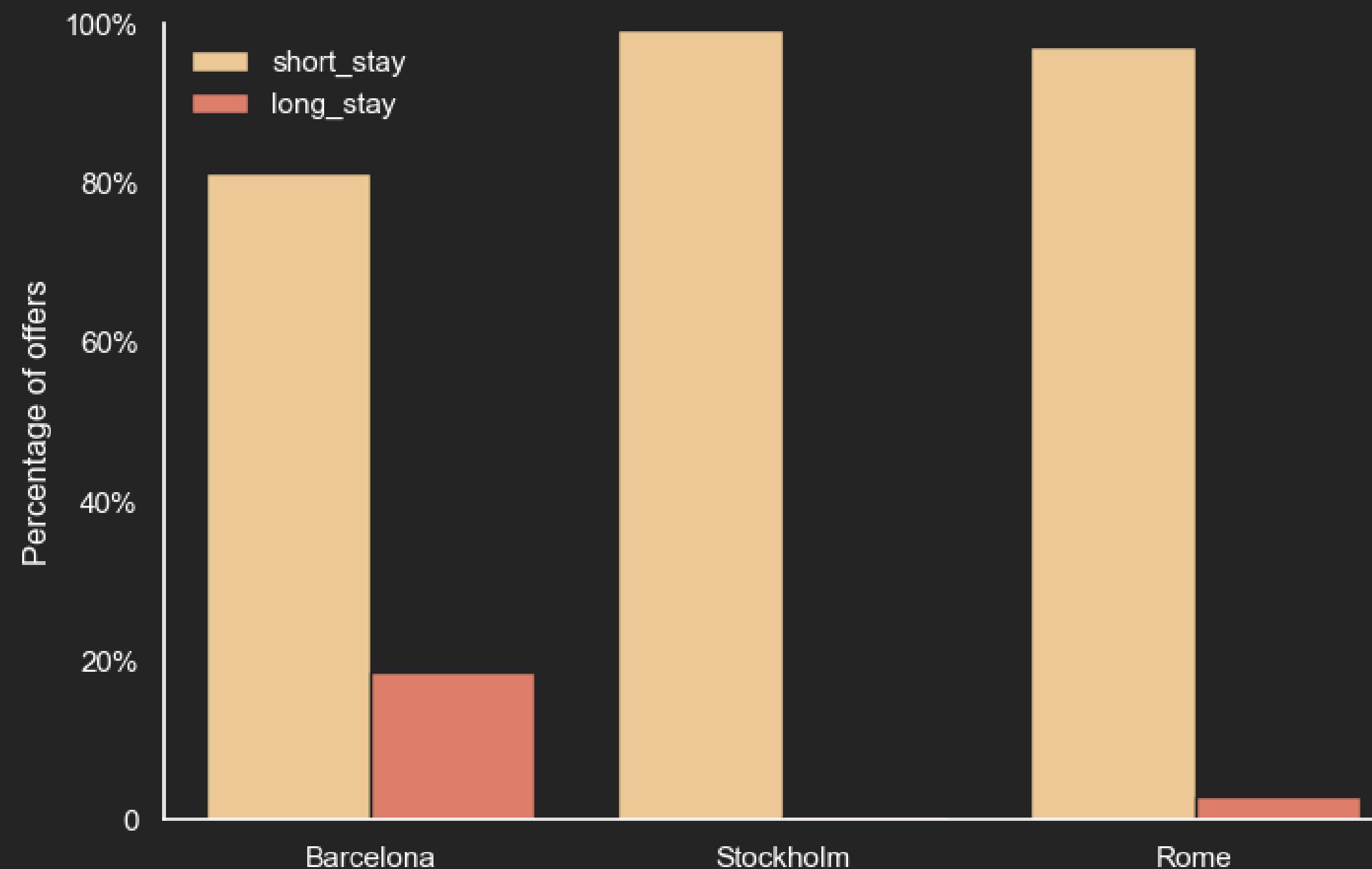
ANALYSIS

MINIMUM LENGTH OF STAY



ANALYSIS

MINIMUM LENGTH OF STAY

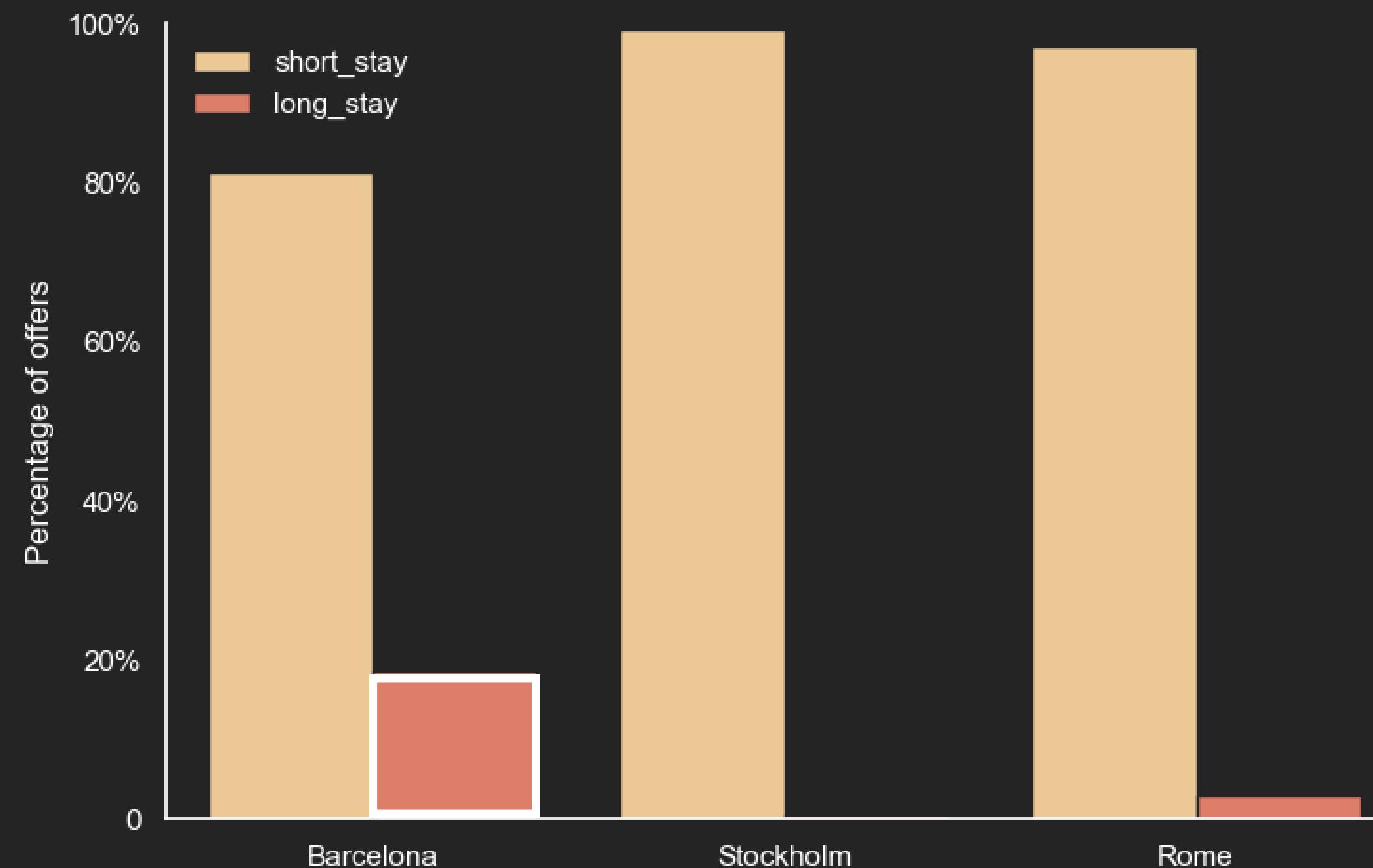




ANALYSIS

MINIMUM LENGTH OF STAY

Most listings allow short-term stay, while due to regulation the proportion of long-term offers in Barcelona is highest





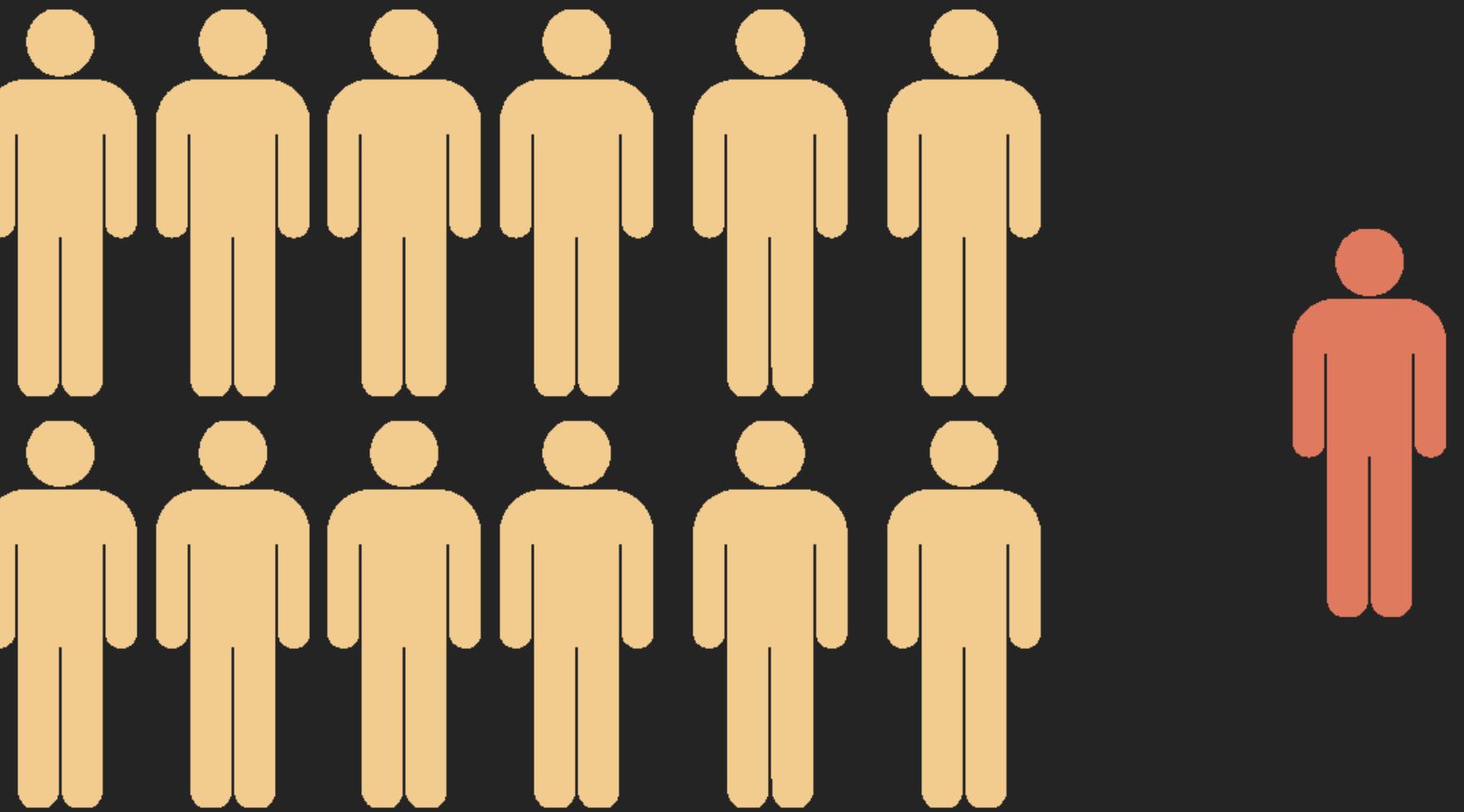
BARCELONA

AIRBNB IN BARCELONA



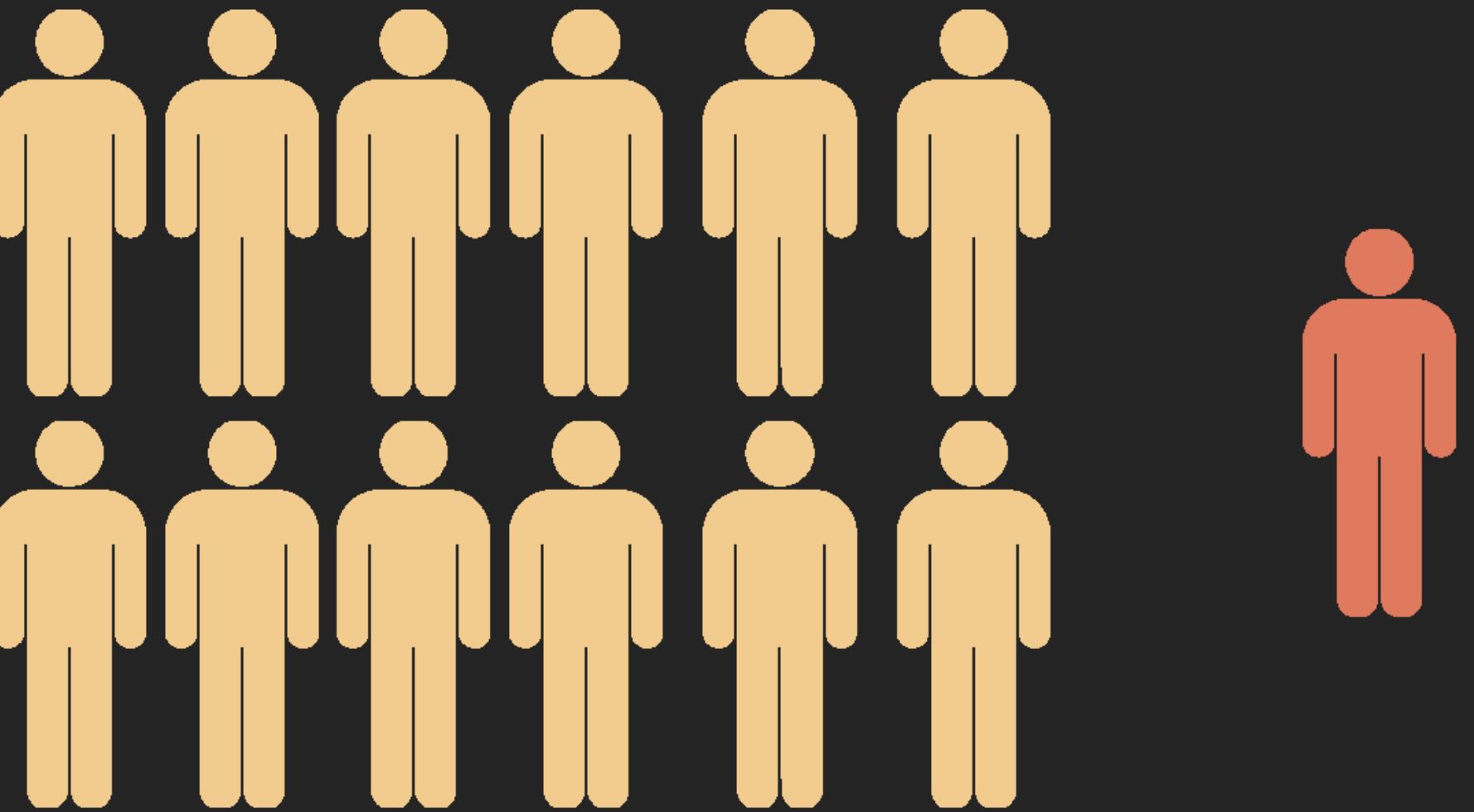


BARCELONA





BARCELONA



1,6 MIO INHABITANTS

19 MIO TOURISTS PER YEAR



BARCELONA

SHORT-TERM STAYS

Hosts have to register their short-term offers to obtain a license

License number has to be displayed on the listing

REGULATIONS





BARCELONA

SHORT-TERM STAYS

Hosts have to register their short-term offeres to obtain a license

License number has to be displayed on the listing

COMMERCIAL HOSTS

Hosts who offer accommodation on a permanent basis and for profit or it's their main source of income

REGULATIONS





Hosts have to register their short-term offers to obtain a license

License number has to be displayed on the listing

COMMERCIAL HOSTS

Hosts who offer accommodation on a permanent basis and for profit or it's their main source of income

AIRBNB'S ROLE

Airbnb does not assume legal responsibilities

REGULATIONS





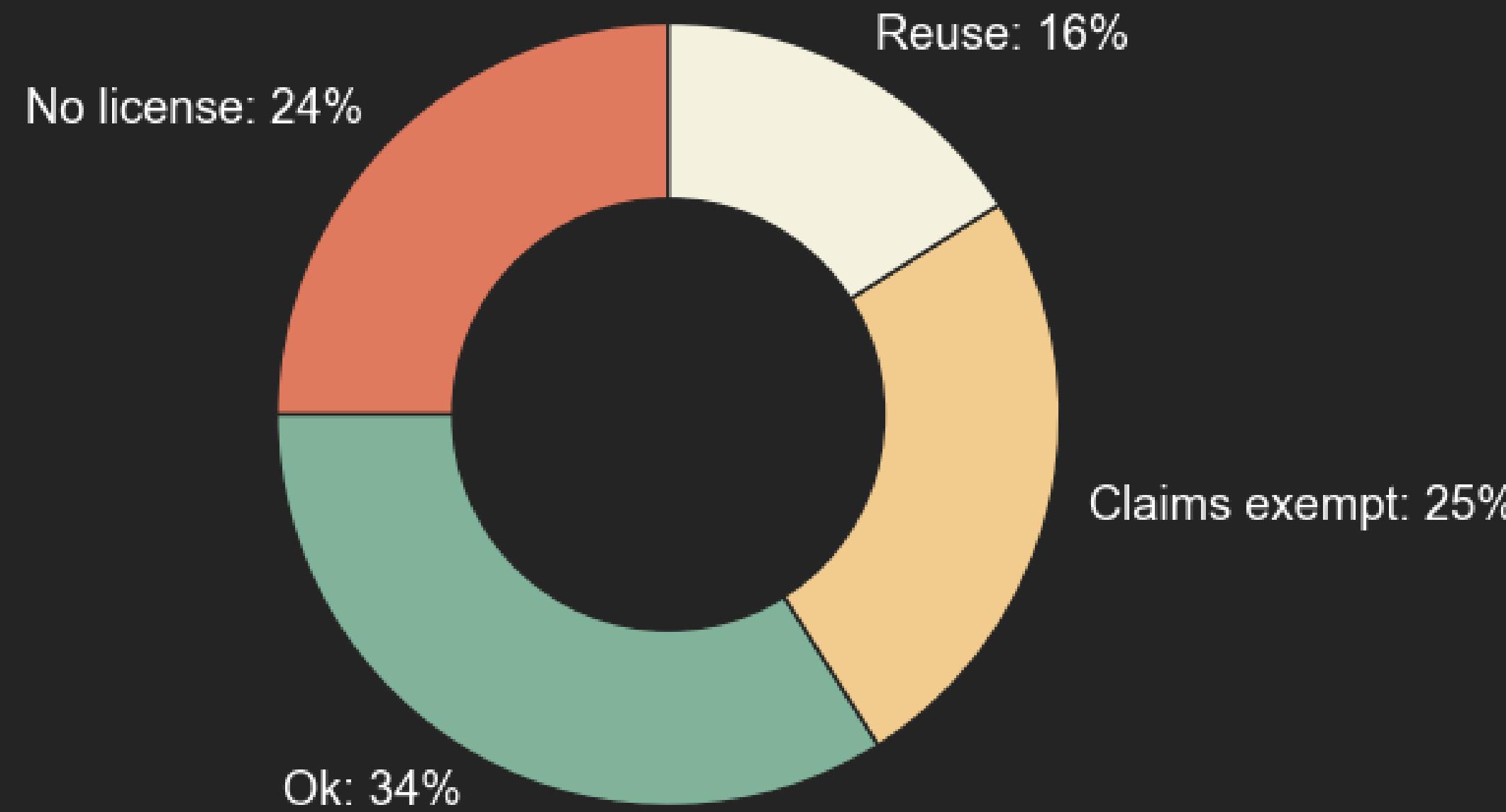
BARCELONA

LICENSES IN SHORT-TERM STAY OFFERS



BARCELONA

LICENSES IN SHORT-TERM STAY OFFERS

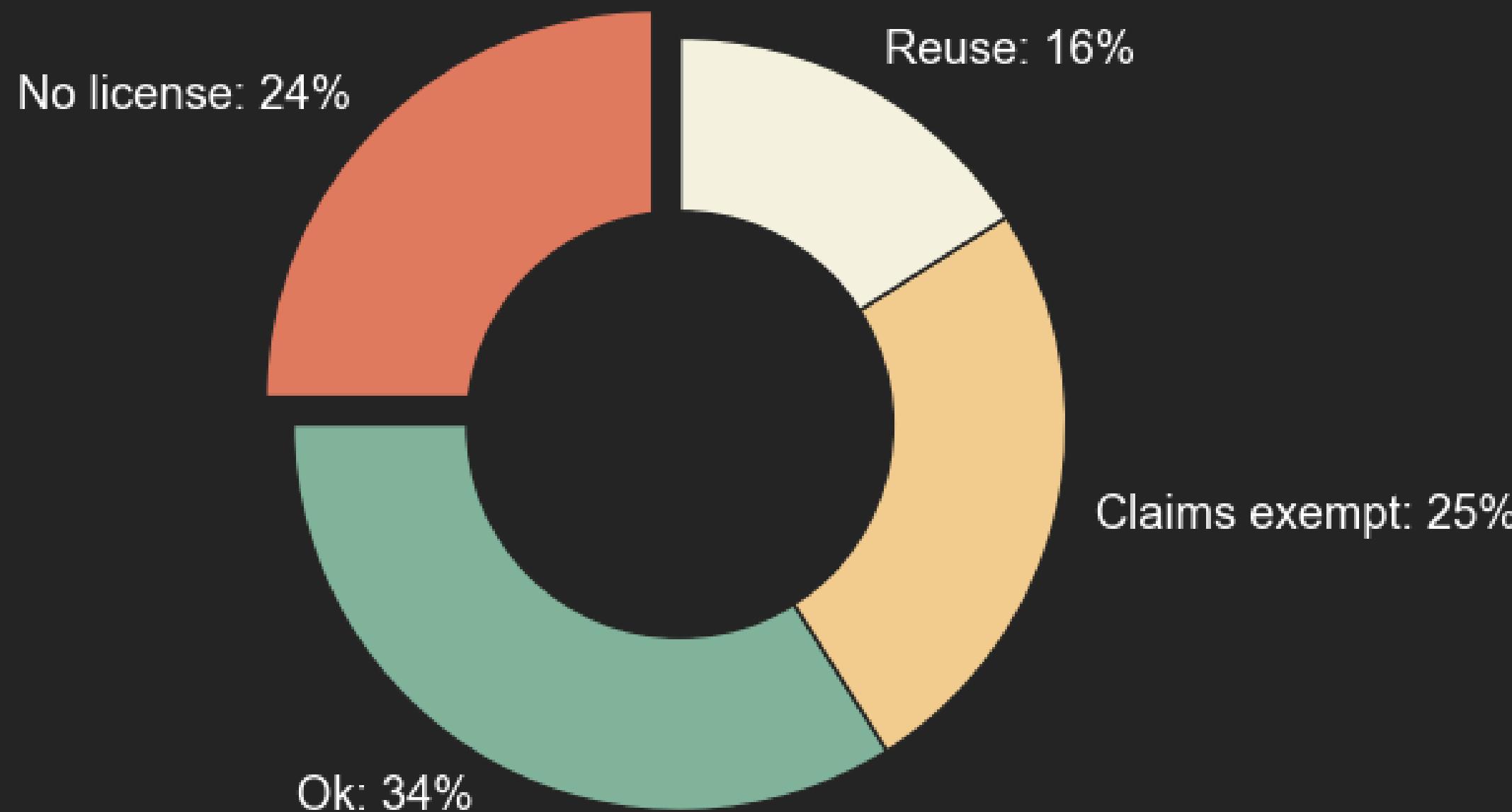




BARCELONA

LICENSES IN SHORT-TERM STAY OFFERS

Many listings offer short-term stays without valid license

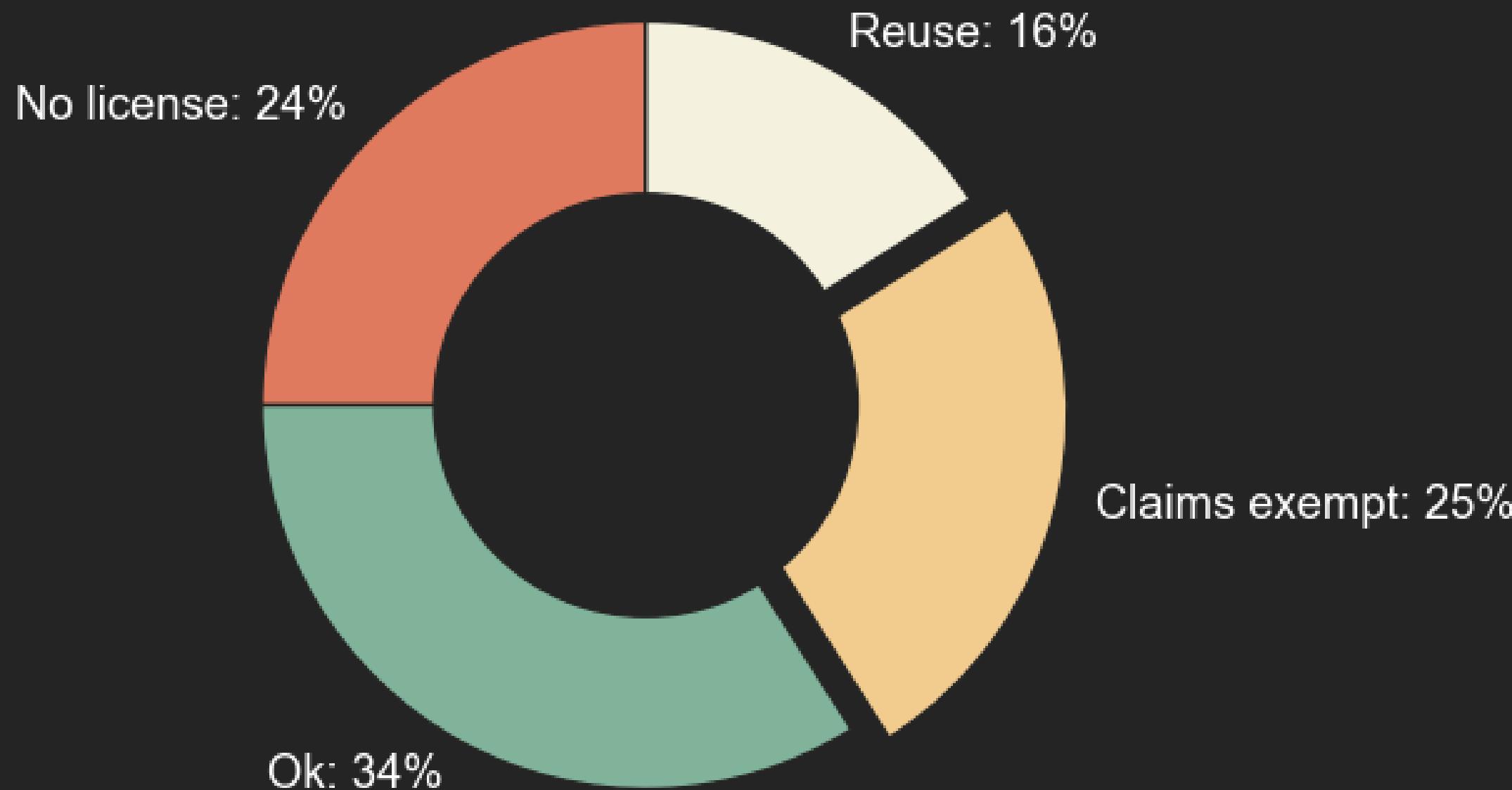




BARCELONA

LICENSES IN CONNECTION WITH SHORT STAY OFFERS

Many listings offer short-term stays without valid license

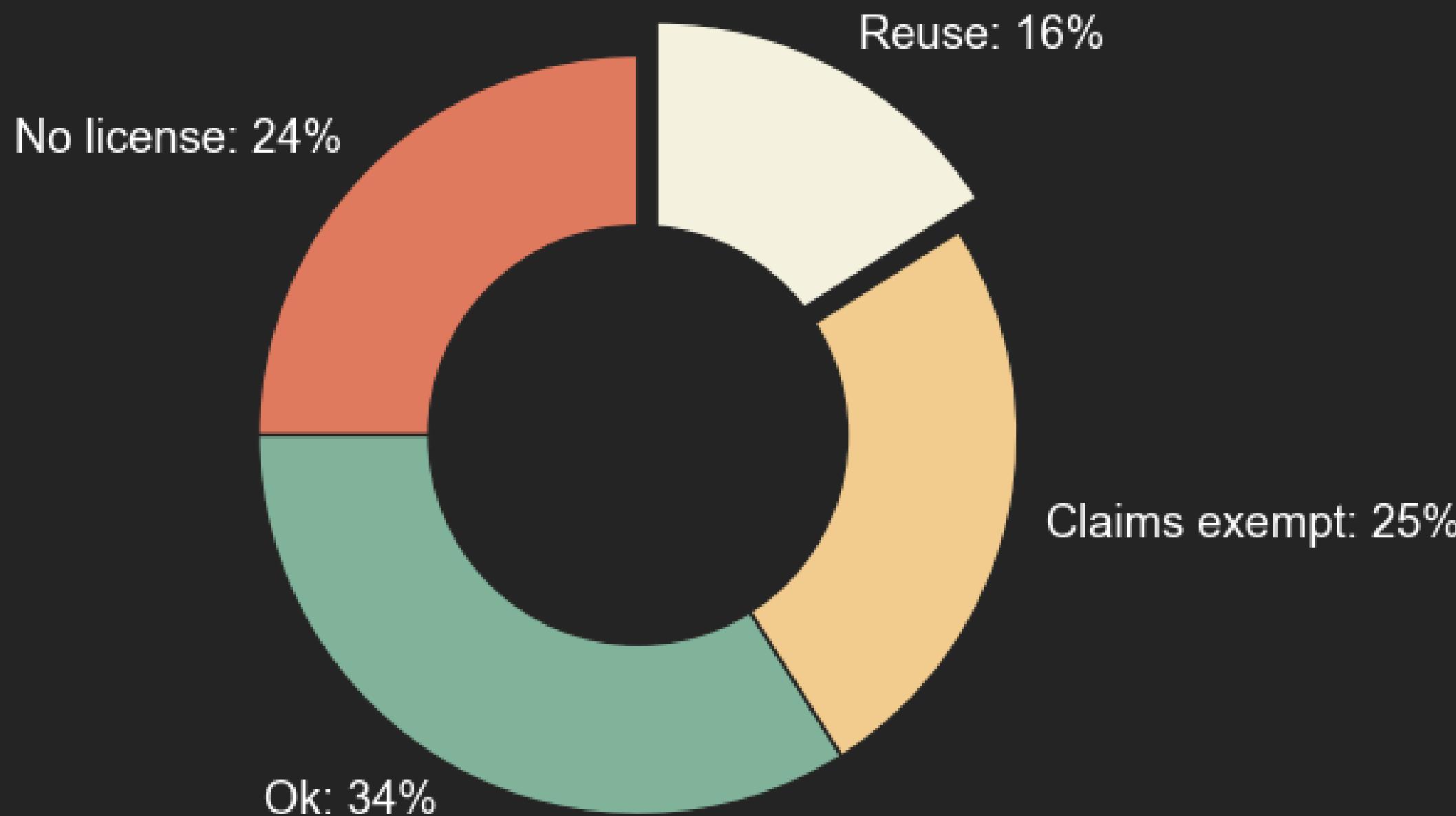




BARCELONA

LICENSES IN CONNECTION WITH SHORT STAY OFFERS

Many listings offer short-term stays without valid license





BARCELONA

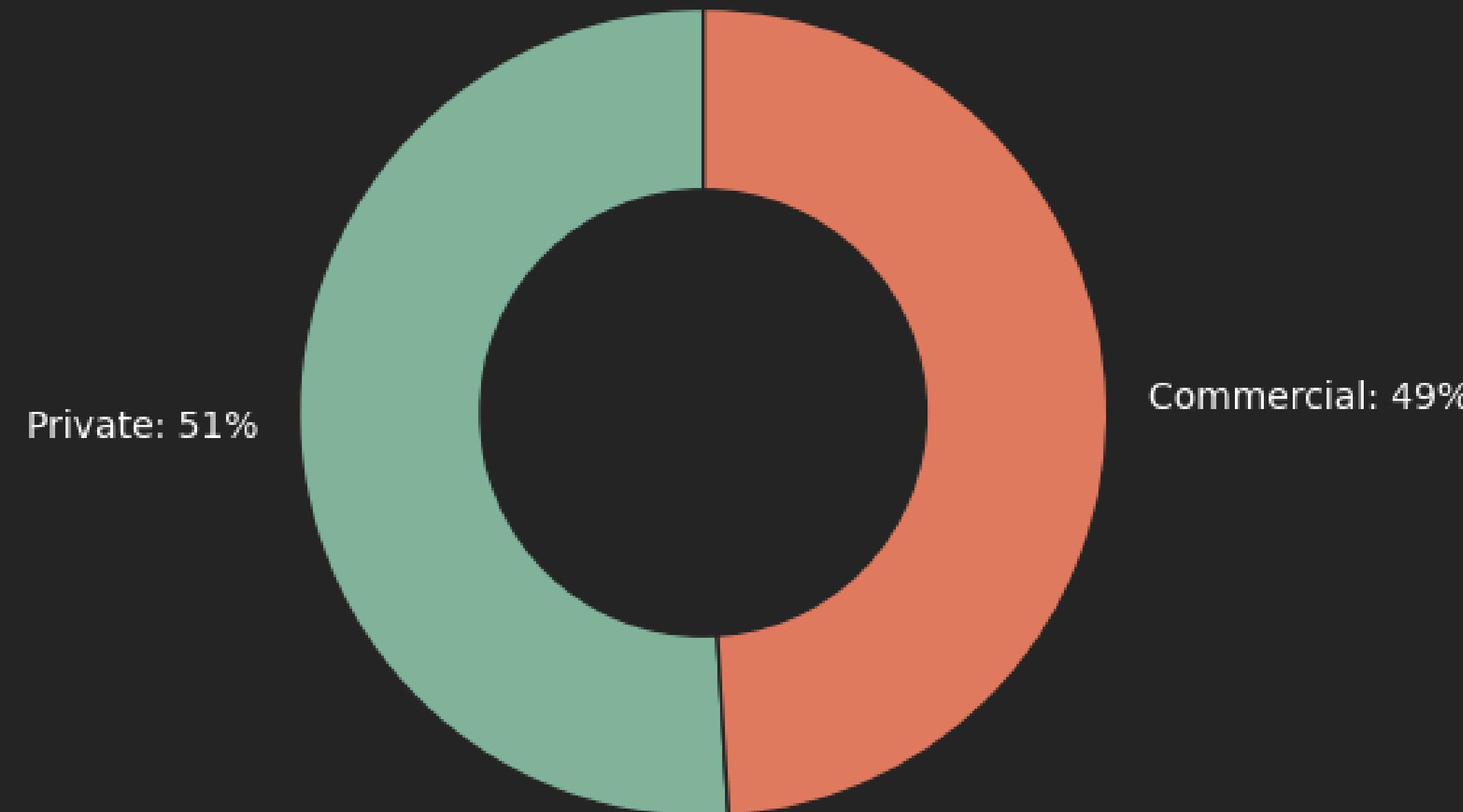
HOST TYPES



BARCELONA

HOST TYPES

Private and commercial listings are almost balanced





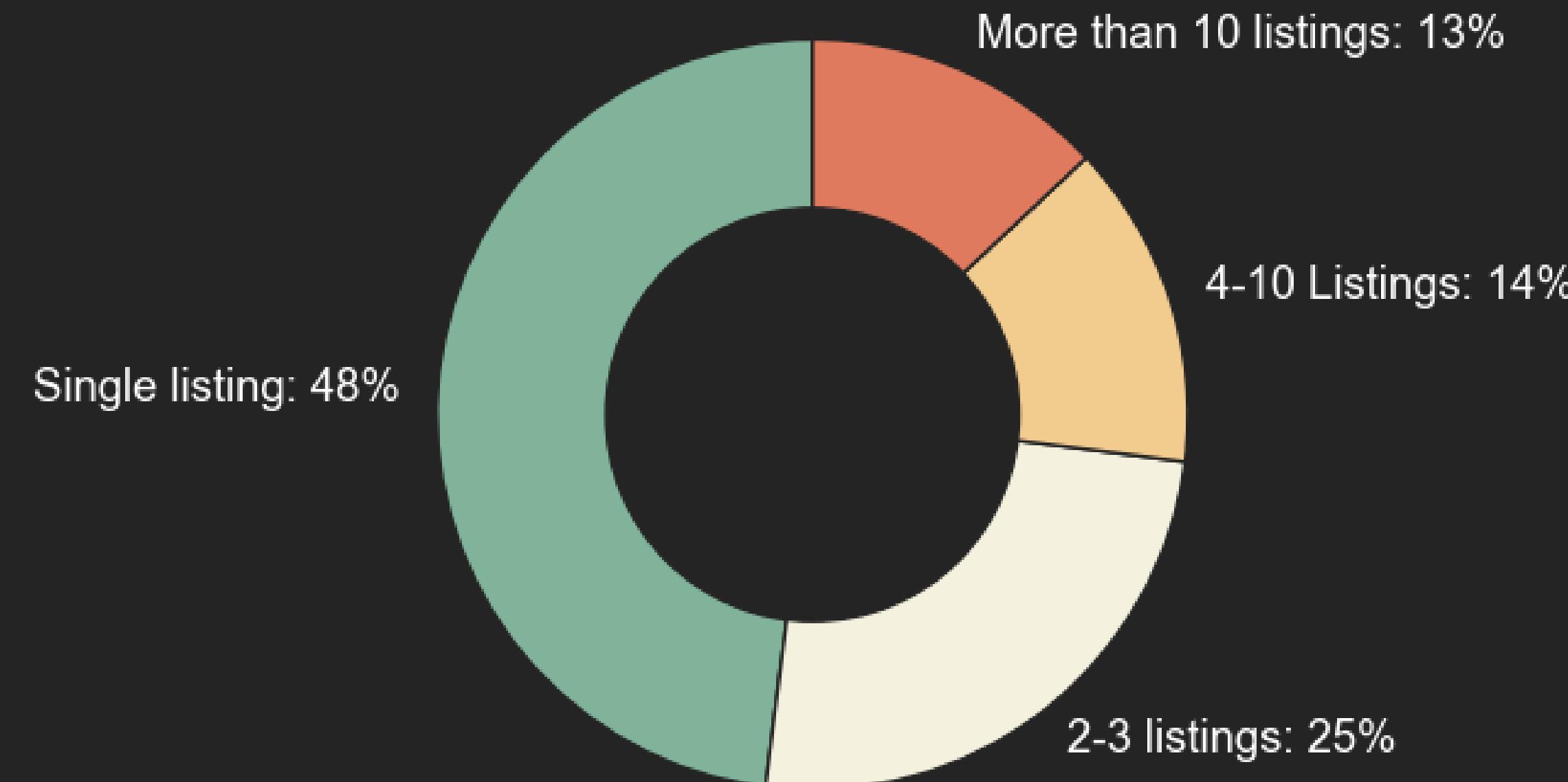
BARCELONA

PRIVATE HOSTS WITH MANY LISTINGS



BARCELONA

PRIVATE HOSTS WITH MANY LISTINGS

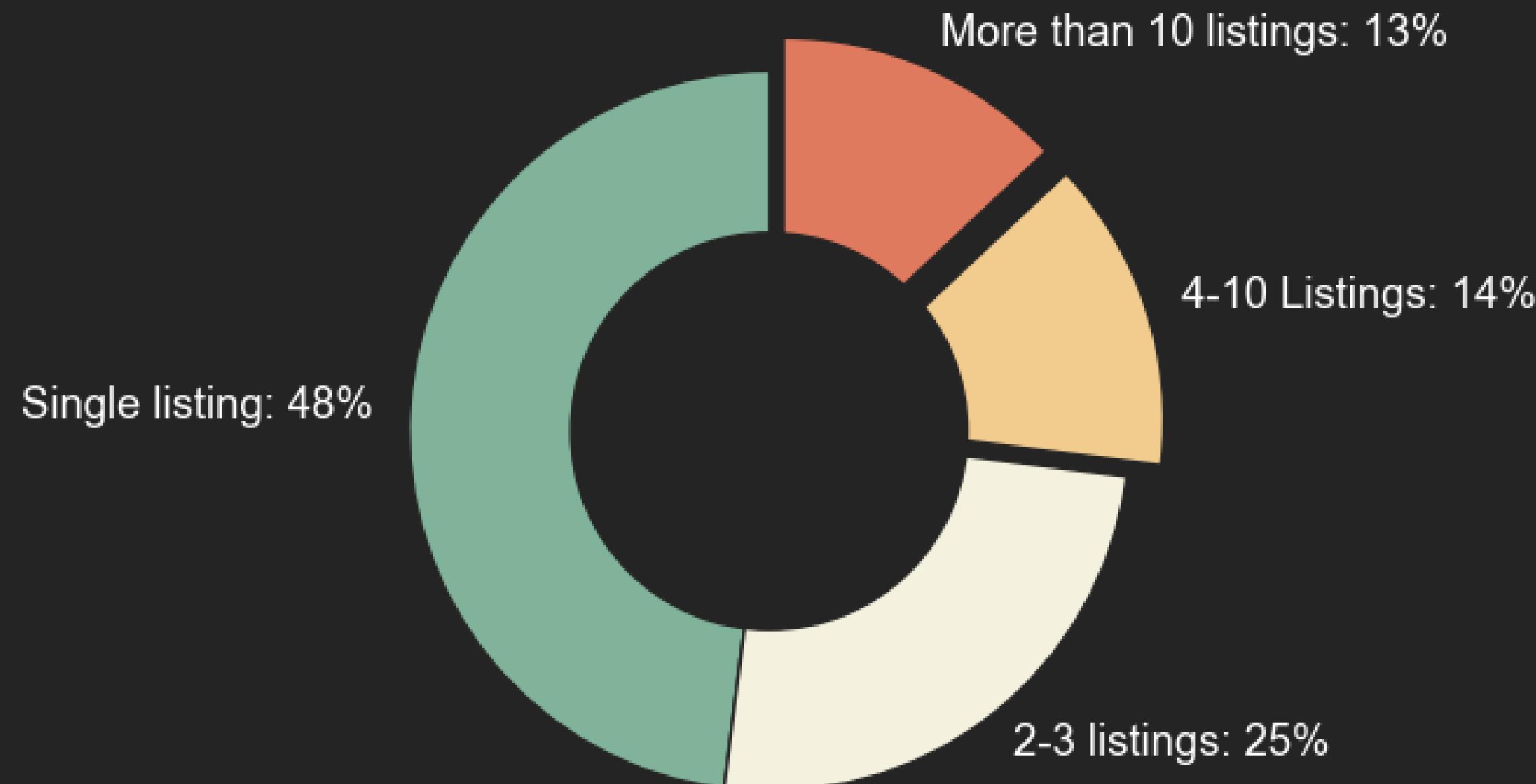




BARCELONA

PRIVATE HOSTS WITH MANY LISTINGS

Many private hosts offer more than 3 accommodations





BARCELONA



SUSPICIOUS LISTINGS



BARCELONA

SUSPICIOUS

SUSPICIOUS LISTINGS





BARCELONA

SUSPICIOUS

1. Listings that offer short-term stay but do not provide a valid license

SUSPICIOUS LISTINGS





BARCELONA

SUSPICIOUS

SUSPICIOUS LISTINGS

1. Listings that offer short-term stay but do not provide a valid license
2. Listings of hosts who claim they're operating privately but offer more than 3 accomodations



BARCELONA

SUSPICIOUS

SUSPICIOUS LISTINGS

1. Listings that offer short-term stay but do not provide a valid license
2. Listings of hosts who claim they're operating privately but offer more than 3 accomodations

Compliant: 54%

Suspicious: 46%





MODEL





MODEL



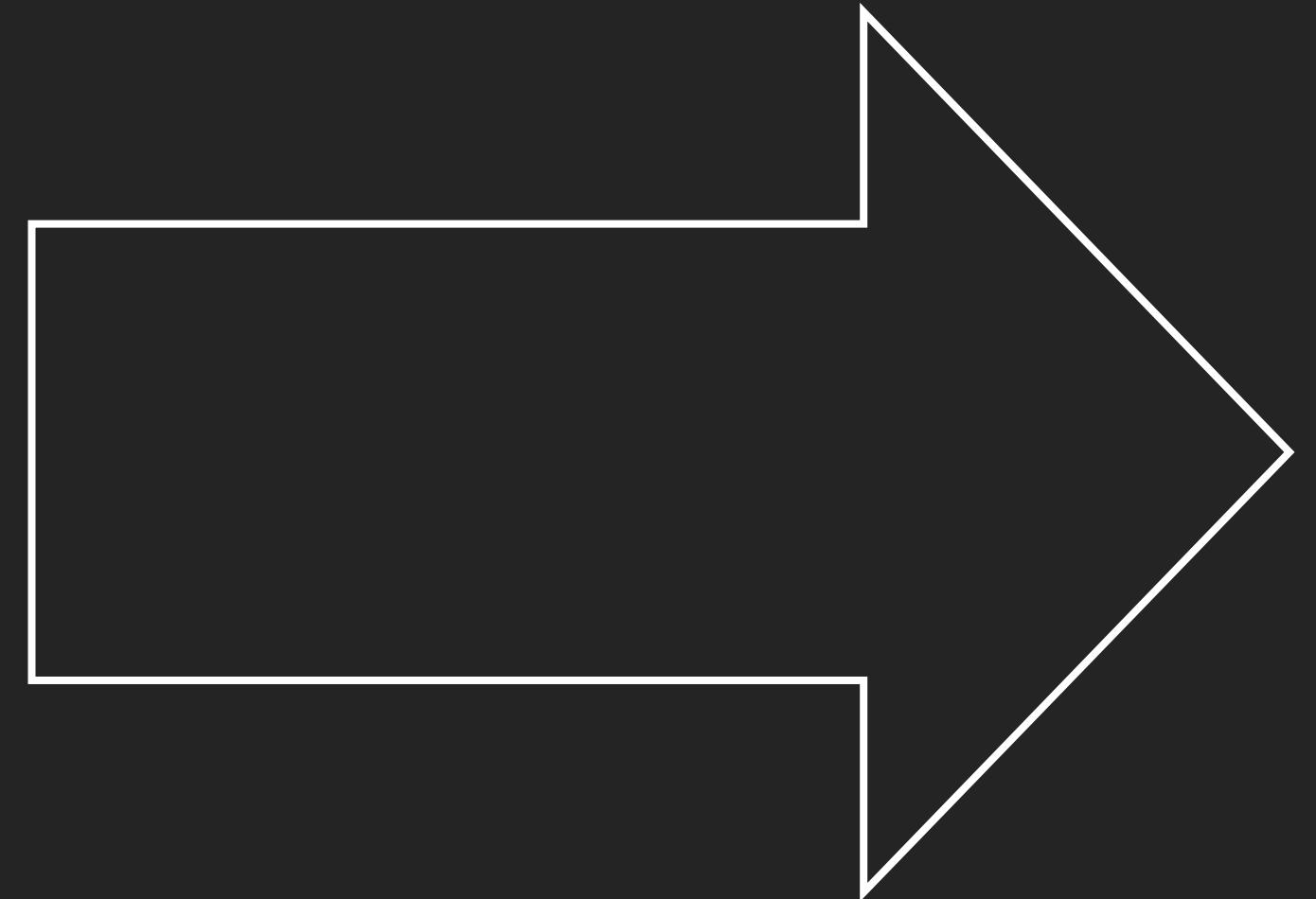
**BASELINE MODEL
LOGISTIC REGRESSION
ACCURACY: 79%**



MODEL



BASELINE MODEL
LOGISTIC REGRESSION
ACCURACY: 79%





MODEL

BASELINE MODEL
LOGISTIC REGRESSION
ACCURACY: 79%



imputation of missing values
outlier handling
feature selection
feature scaling
hyperparameter tuning



MODEL



BASELINE MODEL

LOGISTIC REGRESSION

ACCURACY: 79%

imputation of missing values
outlier handling
feature selection
feature scaling
hyperparameter tuning

FINAL MODEL

RANDOM FOREST

ACCURACY: 99%



MODEL



ACCURACY: 99%



MODEL

FINAL MODEL

CONFUSION MATRIX

		Predicted Value	
		Compliant	Suspicious
Actual Value	Compliant	True Neg 1210 53.35%	False Pos 22 0.97%
	Suspicious	False Neg 8 0.35%	True Pos 1028 45.33%

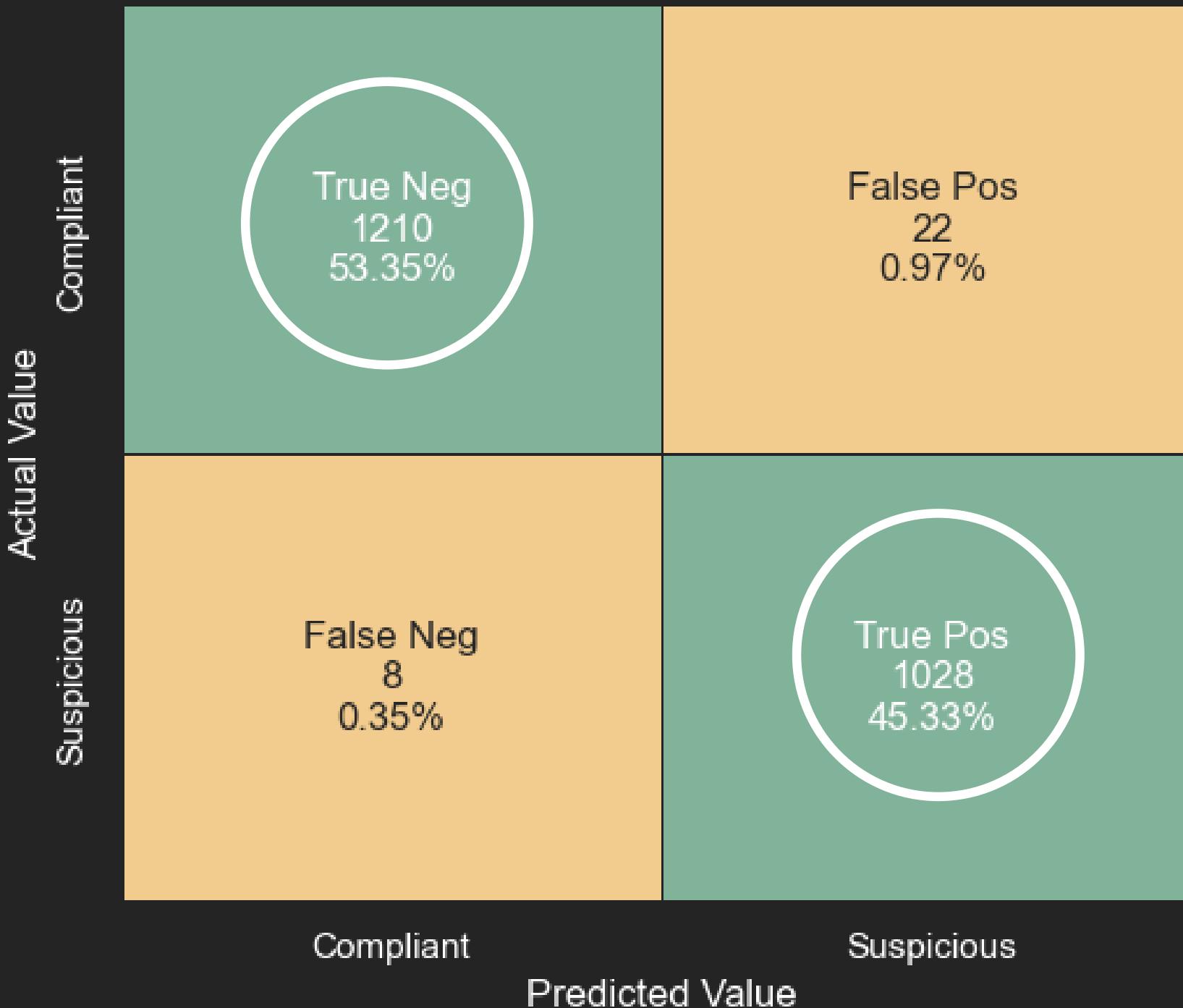
ACCURACY: 99%



MODEL

FINAL MODEL

CONFUSION MATRIX



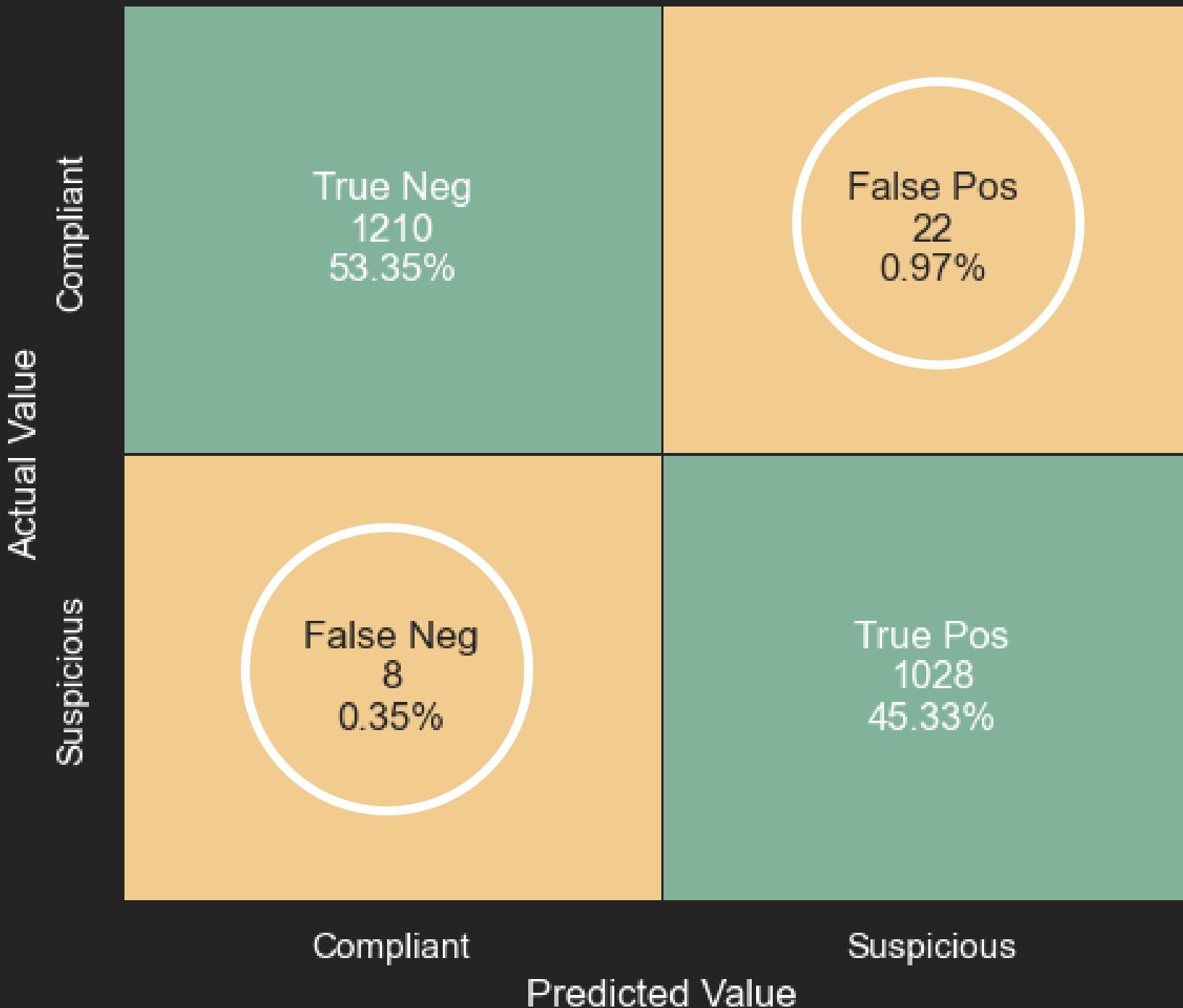
ACCURACY: 99%



MODEL

FINAL MODEL

CONFUSION MATRIX



ACCURACY: 99%



MODEL

USE AND LIMITATIONS



MODEL

USE AND LIMITATIONS

- The model is able to repeat our analysis briefly in an automated way which saves manpower and money
- It can be used by the city council of Barcelona to check on Airbnb listings instantly for irregularities
- Hosts of suspicious listings can be contacted and reminded of the regulations in place
- The city regains control over the activities on Airbnb and ensures that taxes are paid



MODEL

USE AND LIMITATIONS

- The model can be used by our stakeholder to check on Airbnb listings instantly for irregularities
- It is able to repeat the analysis briefly in an automated way which saves manpower and money
- Hosts of suspicious listings can be contacted and reminded of the regulations in place
- The city regains control over the activities on Airbnb and ensures that taxes are paid

DO'S

- apply to Barcelona only
- use the model as a tourist to only book legal accommodation
- question our assumptions!



MODEL

USE AND LIMITATIONS

- The model can be used by our stakeholder to check on Airbnb listings instantly for irregularities
- It is able to repeat the analysis briefly in an automated way which saves manpower and money
- Hosts of suspicious listings can be contacted and reminded of the regulations in place
- The city regains control over the activities on Airbnb and ensures that taxes are paid

DO'S

- apply to Barcelona only
- use the model as a tourist to only book legal accommodation
- question our assumptions!

DONT'S

- apply to other cities
- consider suspicious activity immediately as criminal



OUTRO

TECHNOLOGIES/TOOLS

Python
Unsupervised Machine Learning (Clustering)
Supervised Machine Learning
Webscraping
Tableau
Jupyter
Github
Kanban





OUTRO

TECHNOLOGIES/TOOLS

Python
Unsupervised Machine Learning (Clustering)
Supervised Machine Learning
Webscraping
Tableau
Jupyter
Github
Kanban

CREDITS/THANKS

Inside Airbnb
neuefische
Google and the world wide web





OUTRO

**THANK YOU
FOR YOUR
ATTENTION!**

