



# Airbnb Data Analysis

a guide for future hosts





Noelia  
Álvarez



Elena  
Candela



Miriam  
Sánchez



Sandra  
Rodríguez



Carol  
Vílchez

# Data Sensei Team





target  
audience



## ● Future hosts

key information



decision making



where



how



# selected KPIs



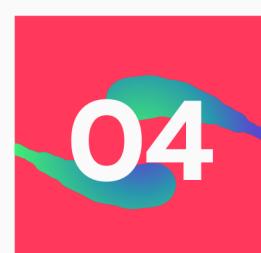
Estimated earnings according to neighborhood and room type



"Gold Area"



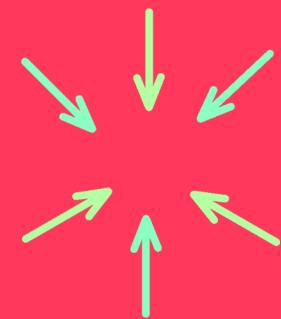
Regulatory compliance of tourist apartments



Factors that affect the rental price

# Gold area

"Neighborhood or location that is particularly popular and in high demand among guests"



**central location**



**high profits**



**tourist attractions**



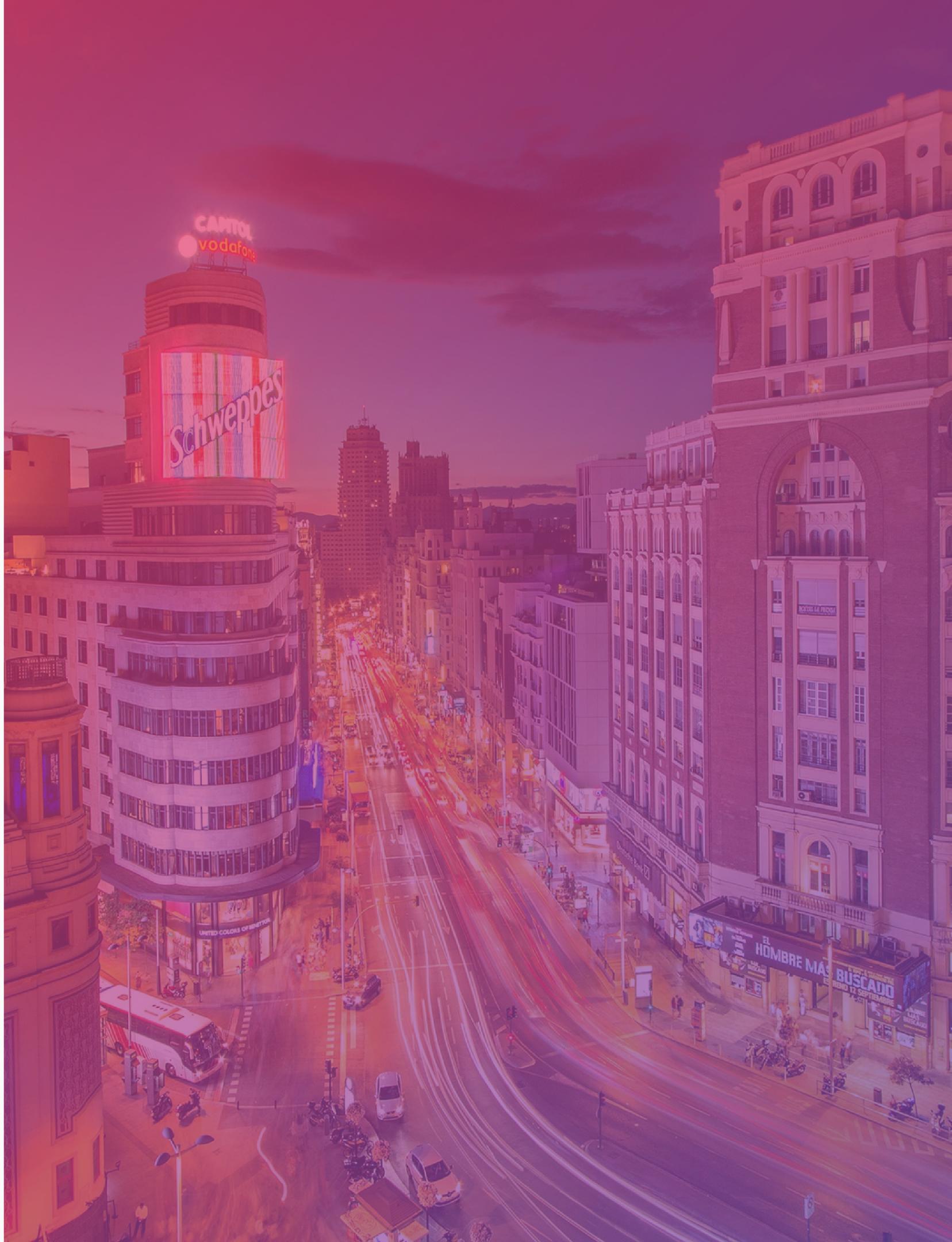
**accesibility**

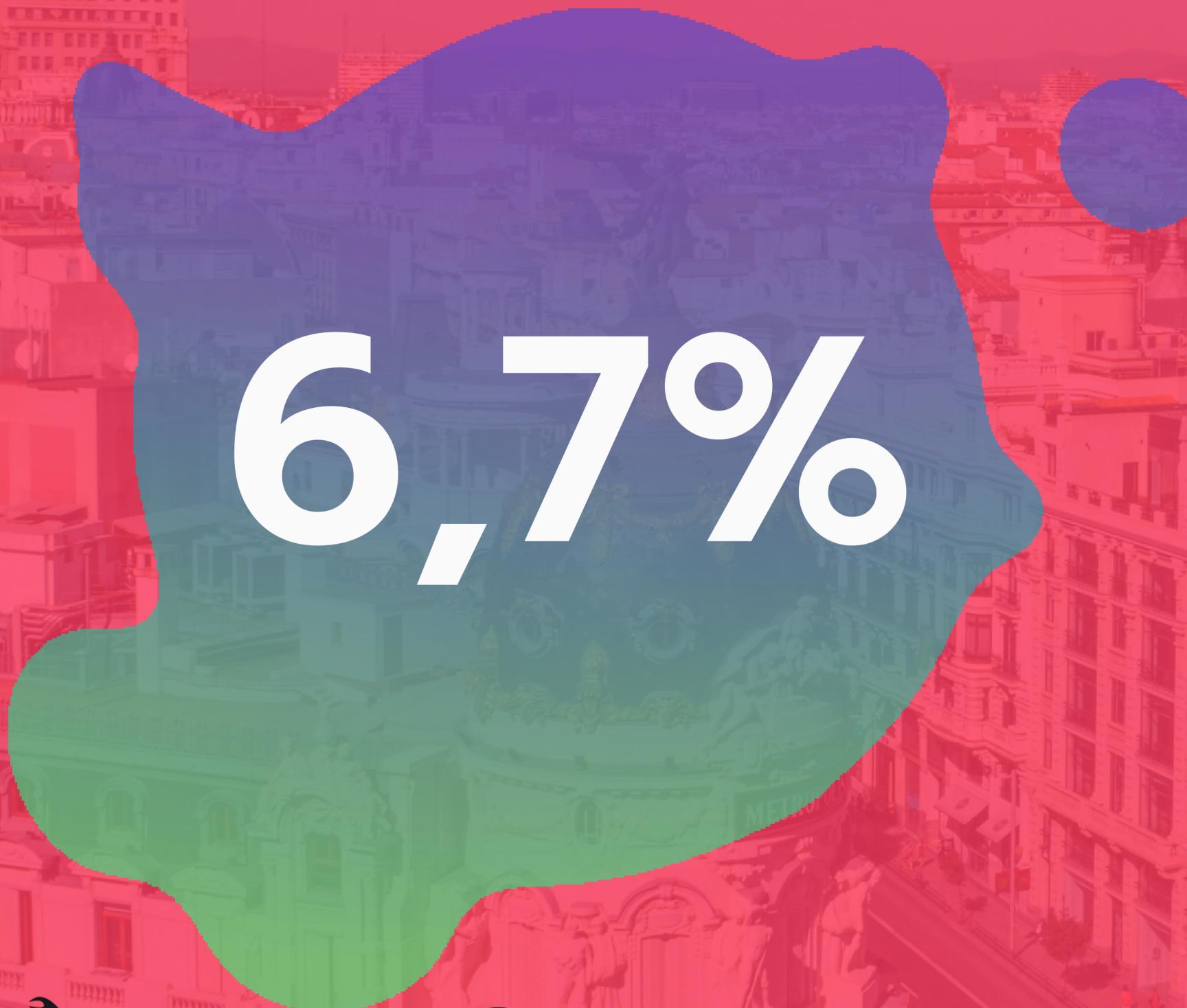
# Airbnb Requirements in Madrid

- License and registration with the Comunidad de Madrid
- Rented for a maximum of **90 nights**



Glovo?

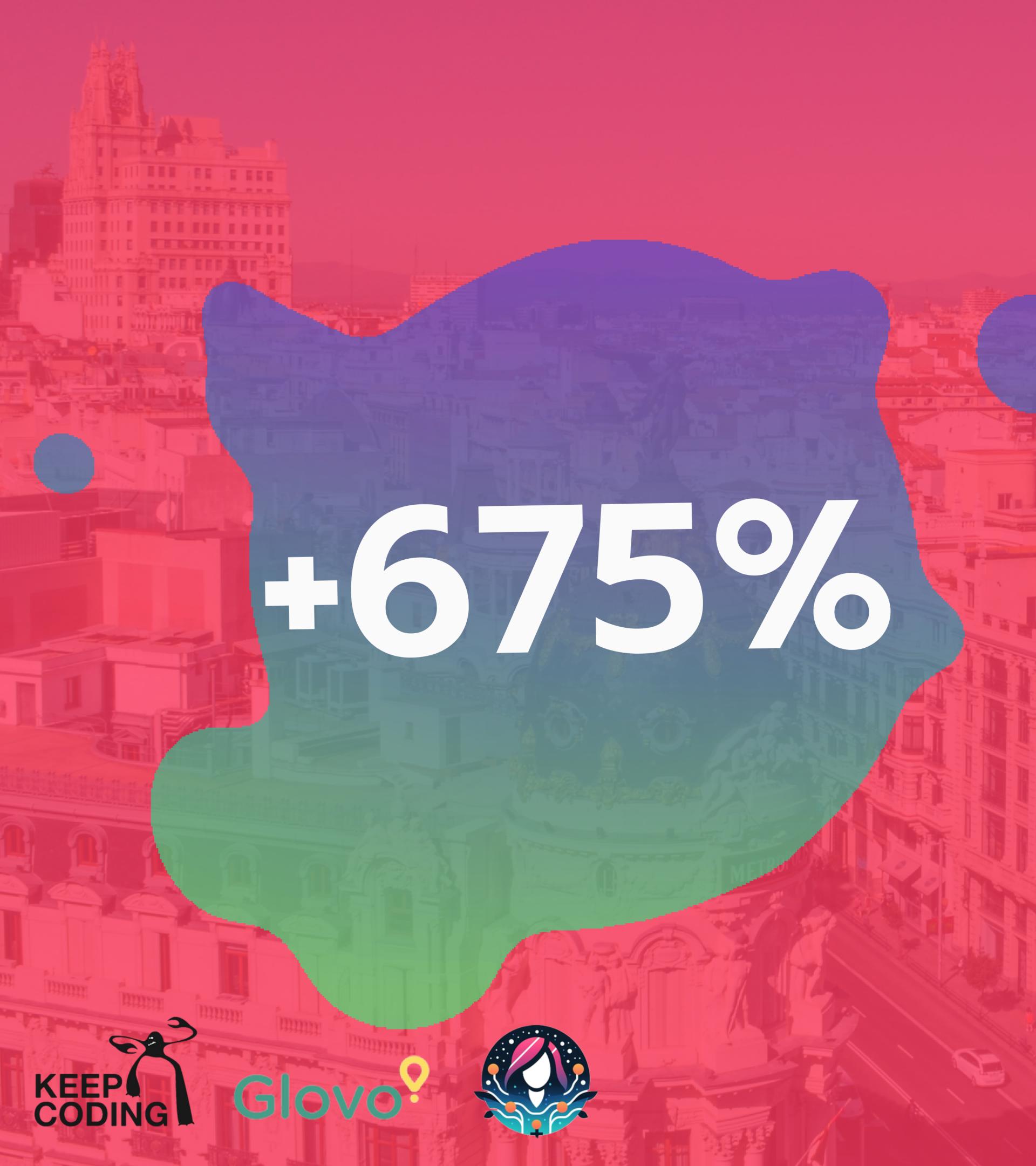




6,7%

---

Of the hosts from the dataset do  
not comply with the regulations



+675%

---

Increment of New Hosts from 2018  
to 2020

# Learning path



Analysis



User stories



KPIs

# Tools

- Agile Method

git

taiga.io

discord



# Tools

- SCRUM board

taiga.io

The screenshot shows a Scrum board for the project "airbnb-sprint1" from May 6 to May 14, 2024. The board has five columns: NEW, IN PROGRESS, READY FOR TEST, CLOSED, and NEEDS INFO. The NEW column contains three user stories. The IN PROGRESS column contains four user stories. The READY FOR TEST column contains one user story. The CLOSED column contains four user stories. The NEEDS INFO column contains one user story. The top bar shows a progress bar at 17% completion, 18 total points, 3 completed points, 8 open tasks, 4 closed tasks, and 0 cocaine doses.

USER STORY	NEW	IN PROGRESS	READY FOR TEST	CLOSED	NEEDS INFO
#1 As a member of the team I want to analyze the dataset in order to familiarize with the data 3 pts	NEW				
#2 As a sensei I want to create a db model architecture in order to define the best model for the data 5 pts	NEW				
#3 As a sensei I want to load the csv file data into the db 5 pts	NEW				
#29 normalize ERD model					
#30 unify the design of the display					
#28 create ERD Entity-relationship diagram					
#31 import csv intodbeaver					
#36 create meeting schedule					
#24 write conclusions tonotebook					
#25 create a logo team					
#23 create calculations for graphical representations					
#18 download dataset and evaluate					



Glovo?

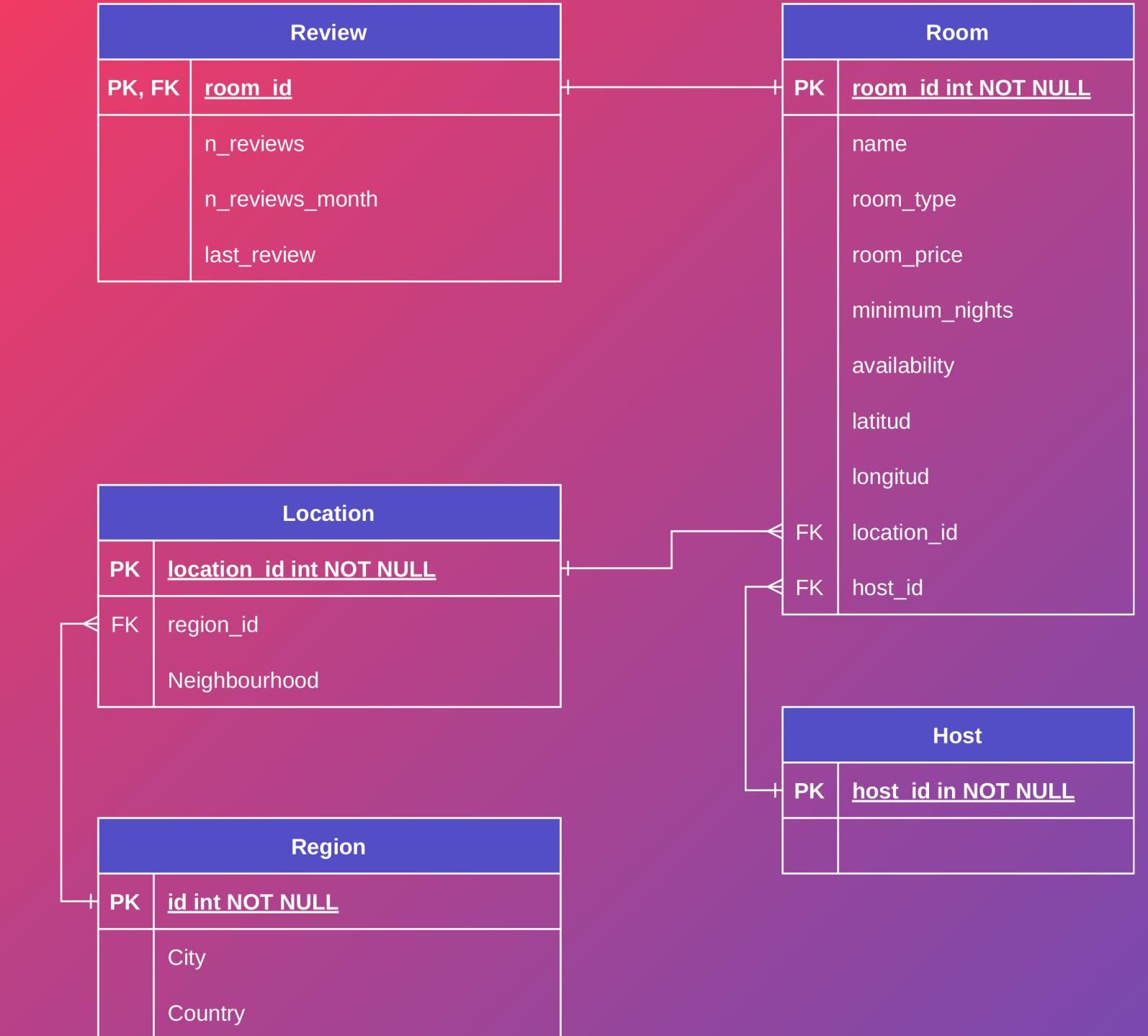


# Tools

- Implementing Data Warehouse
- ETL

draw.io

DBeaver



# Tools

- Sampling and initial exploration of the data
- Exploratory analysis
- Pre-processing and Modeling prediction

R



## Linear regression model

**independent variable:**

room\_price

**dependent variables:**

neighbourhood

room\_type

availability

minimum\_nights

# data visualization



## ● Future hosts

key information



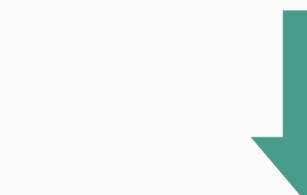
decision making



where



how



tableau

# Conclusions

room price



hotel room

centric neighbourhood

low availability



higher minimum nights

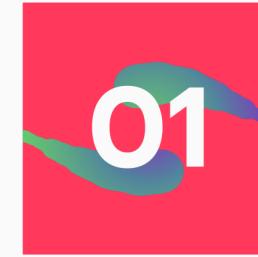


KEEP  
CODING

Glovo!



# What we would do again?



## Detailed Data Exploration

Understanding the variables, its limitations and their relationships



## Accurate Model Evaluation

Using RMSE and  $R^2$  as a metric to assess the model performance



## Documentation process and effective communication within the team

# Lessons learned

01

**Importance of data cleaning as the first step of analysis**

02

**Source data quality limitation**

03

**Linear model is not the most appropriate in this case**

due to the complex non-linear relationships between variables

# Thank you!



DataSensei Team