



SPRINT 1

TORONTO AIRBNB PRICE PREDICTION

By:

 Apurba Pandey



1. Problem Statement

2. Data Driven Insights

3. Impact of Data Driven Solution

- ## 4.
- Dataset
 - EDA
 - WHATS NEXT!!

Table of contents



Problem Statement

Challenges Faced by Hosts:

- Difficulty determining the best area for investment.
- Struggles with pricing and understanding occupancy rates.
- Newcomers often make suboptimal decisions.



Solution

- **Goal:**

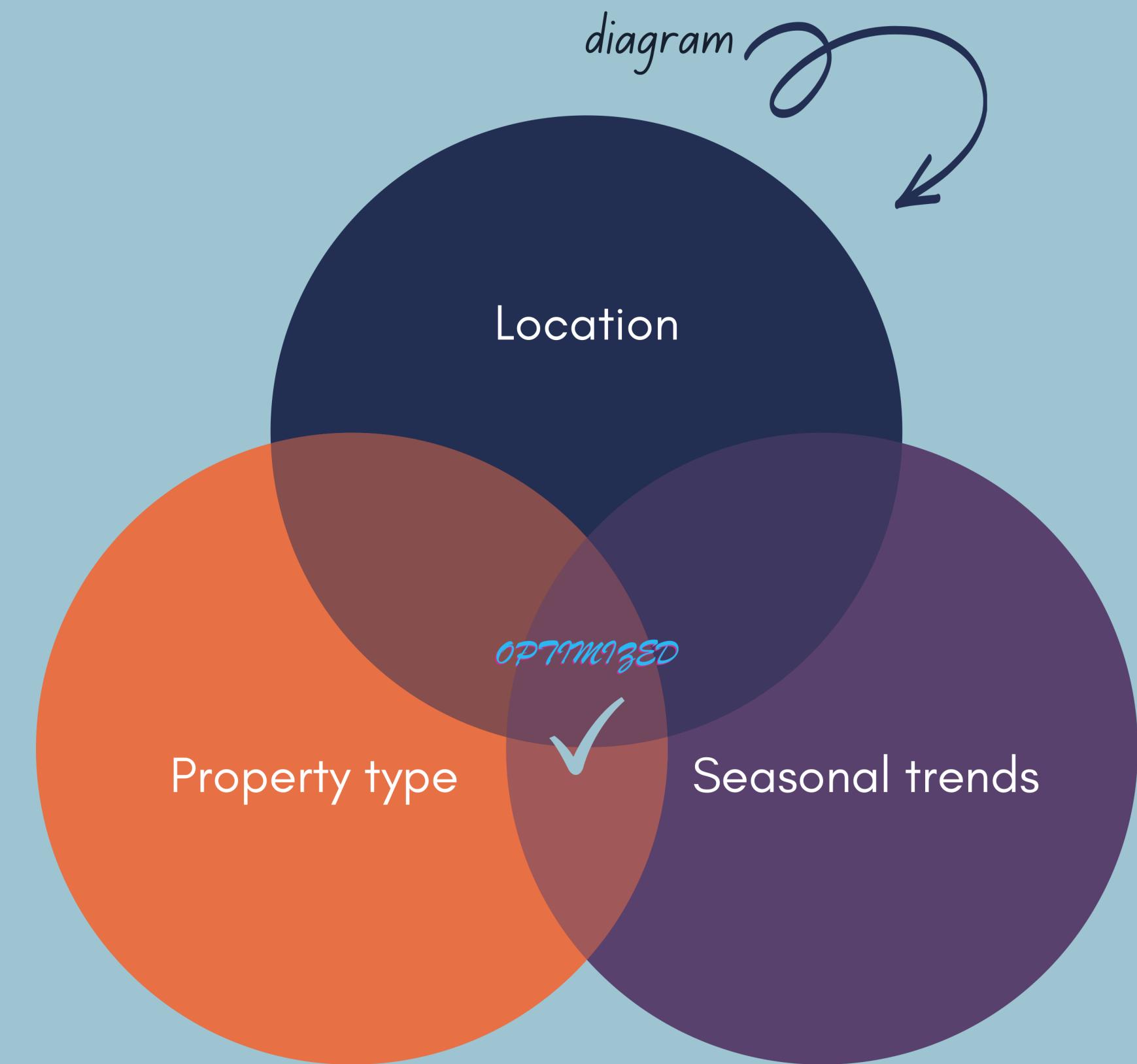
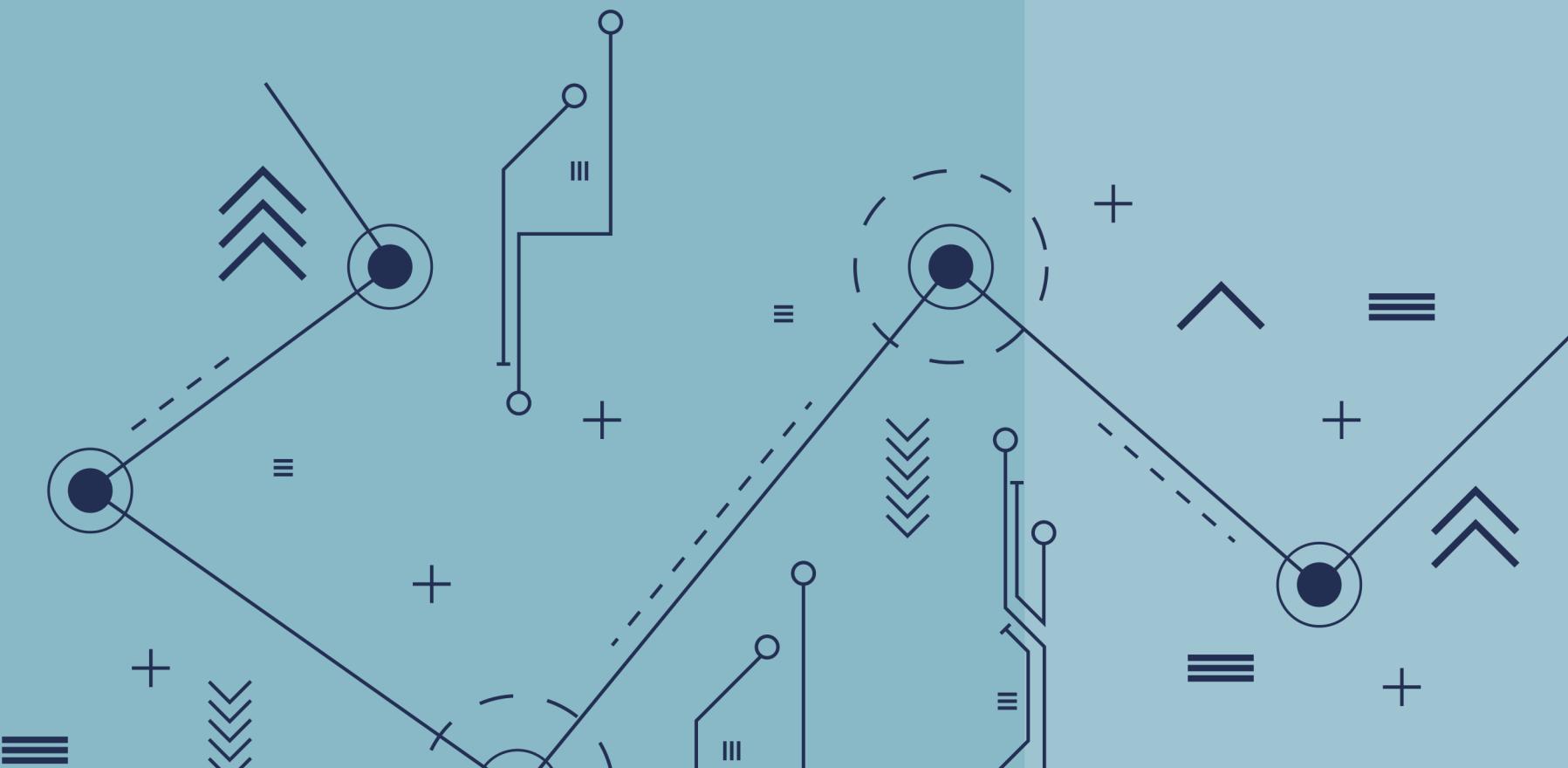
- Develop a machine-learning model for Airbnb hosts.

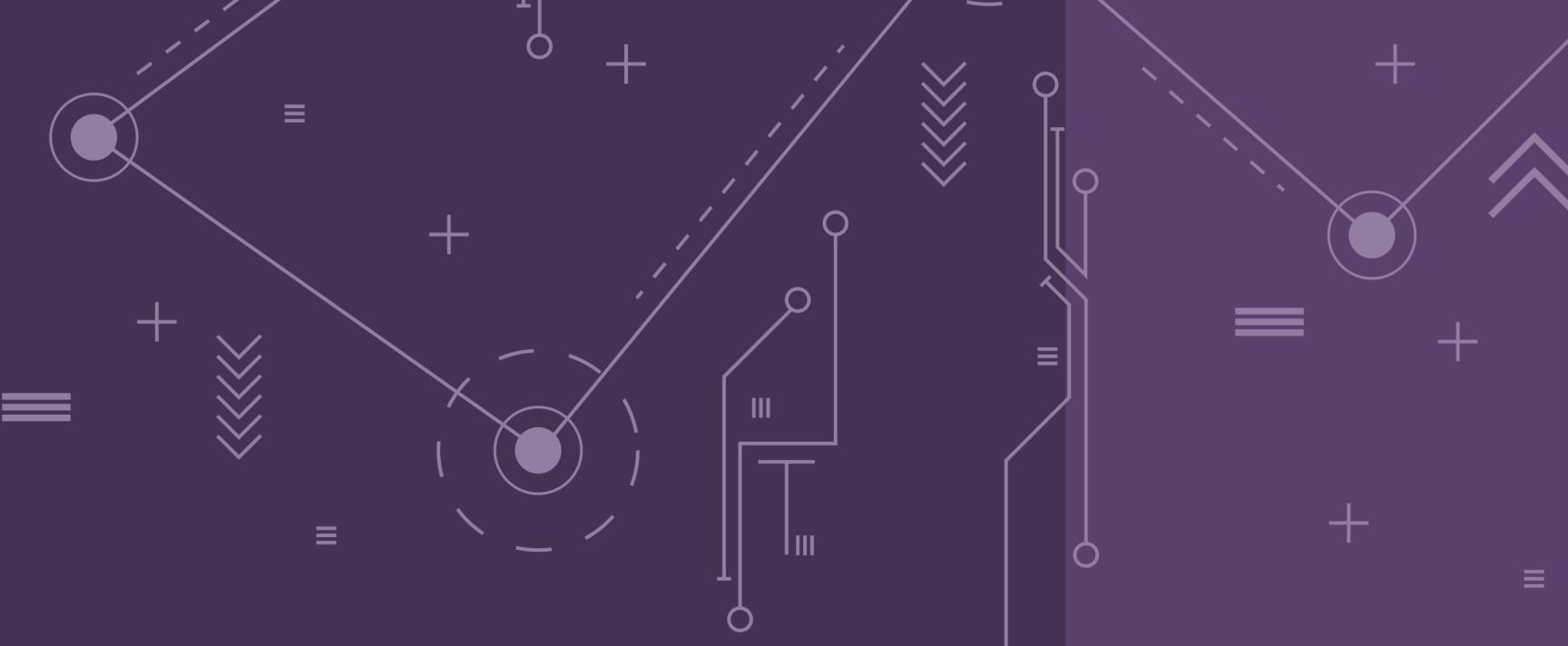
- **Data Science Techniques to be Used:**

- Extensive dataset analysis
- Supervised techniques: Linear regression, Neural networks, XGBoost.

- **Anticipated Outcomes:**

- Predictive models for occupancy rates and rental income.
- Data-driven insights into pricing, property type optimization, and more.



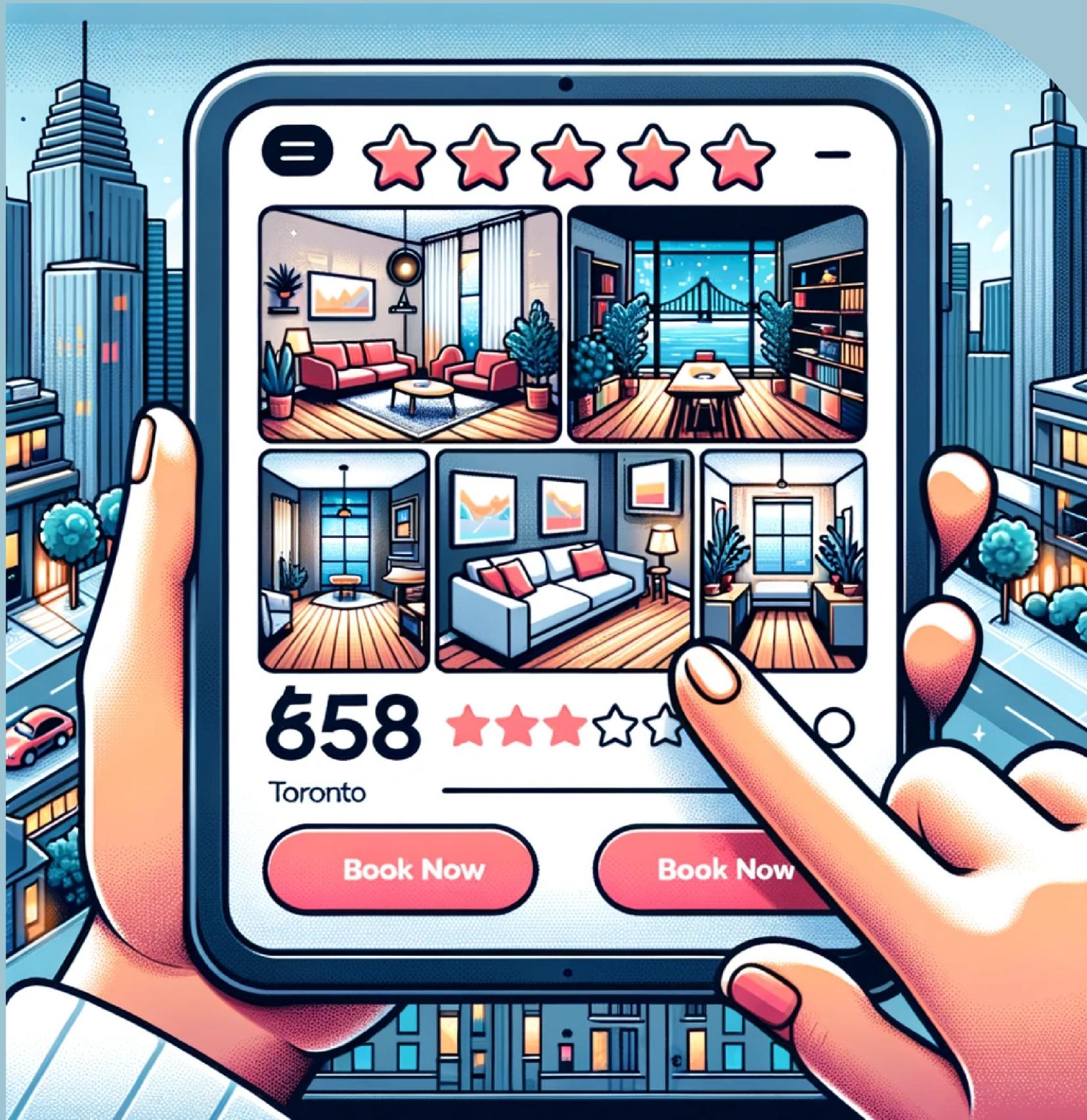


Estimated Potential Impact



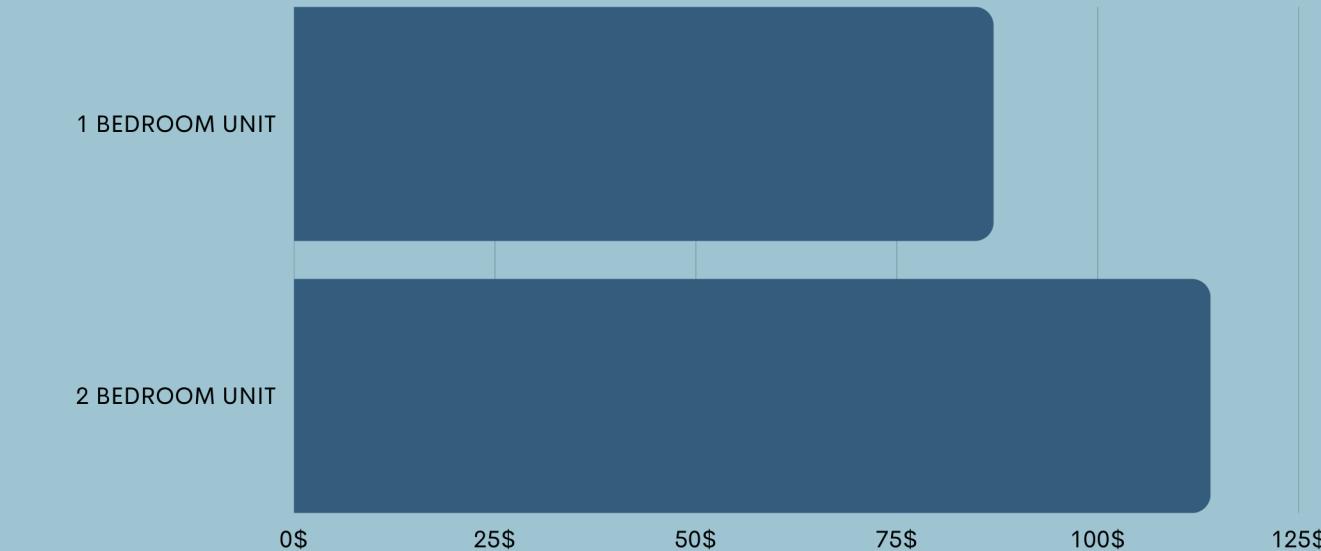
- **Boost in Profitability:**
 - Potential for thousands of dollars in additional earnings for hosts.
- **Time Savings:**
 - Collective savings of thousands of hours.
- **Enhanced Guest Experience:**
 - Aim for higher guest satisfaction.
 - Anticipation of more positive reviews, which further benefits hosts.

PRICE/COST Ratio



According to CP24.com

- Average one bedroom apartment in Toronto goes for 2600
- Average two bedroom apartment in Toronto goes for 3413



```
### Check the mean prices for different bedrooms  
df_combined.groupby("bedrooms")["price"].mean()
```

bedrooms

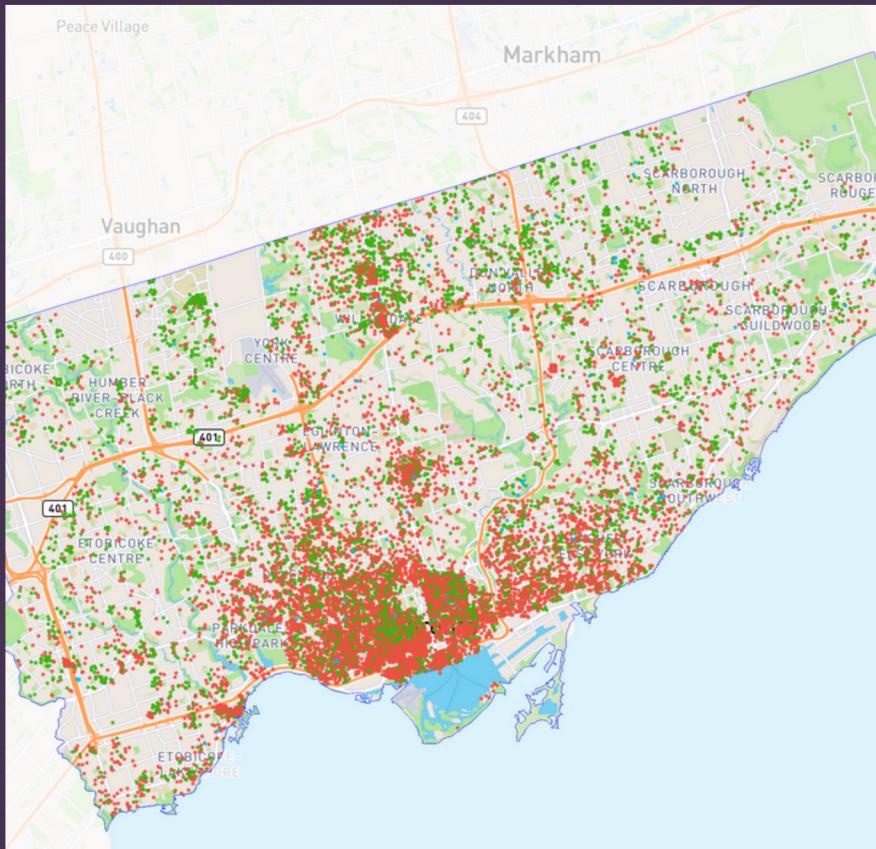
1.00	145.95
2.00	241.31



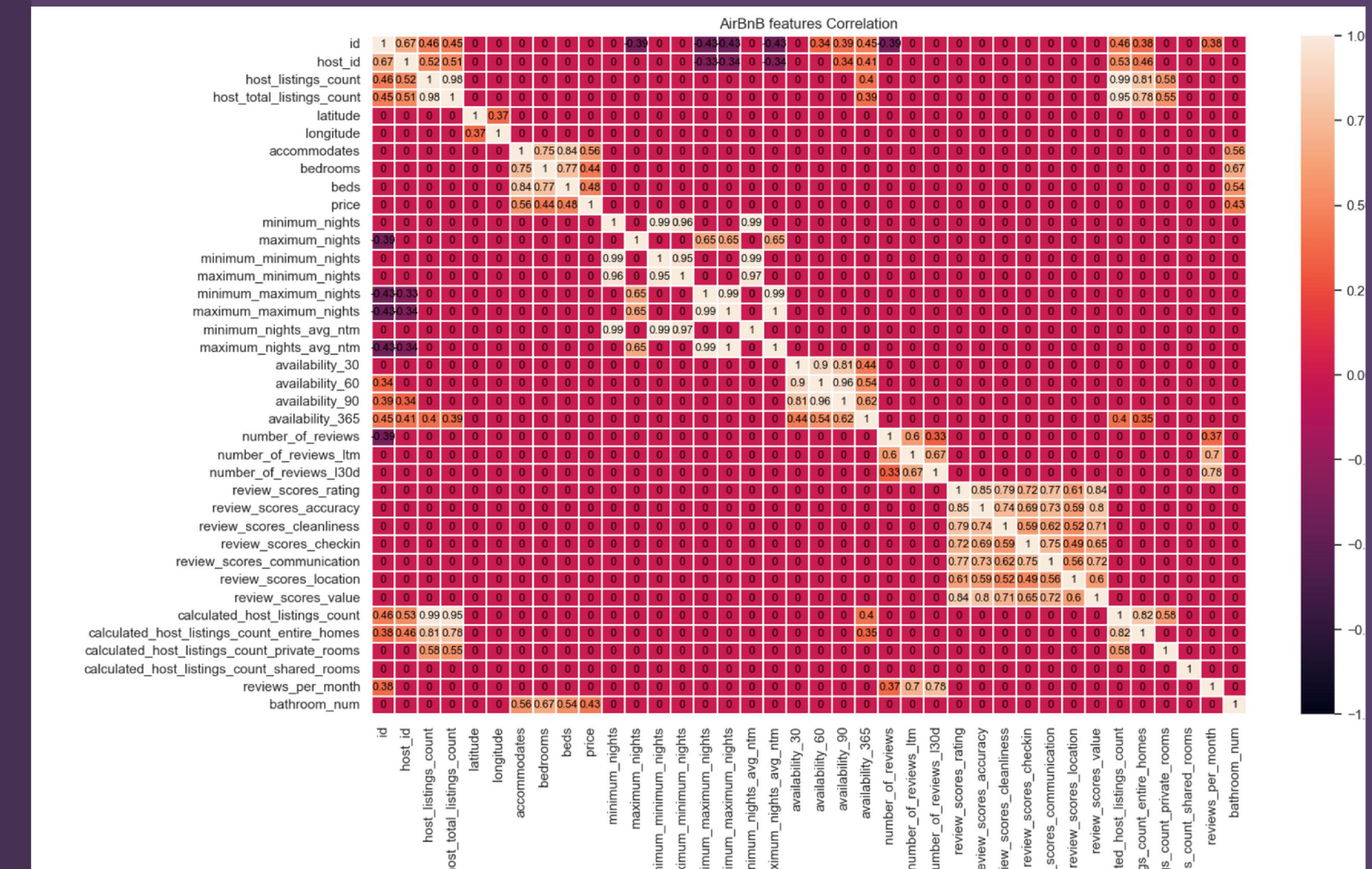
Dataset Source



- Open Data



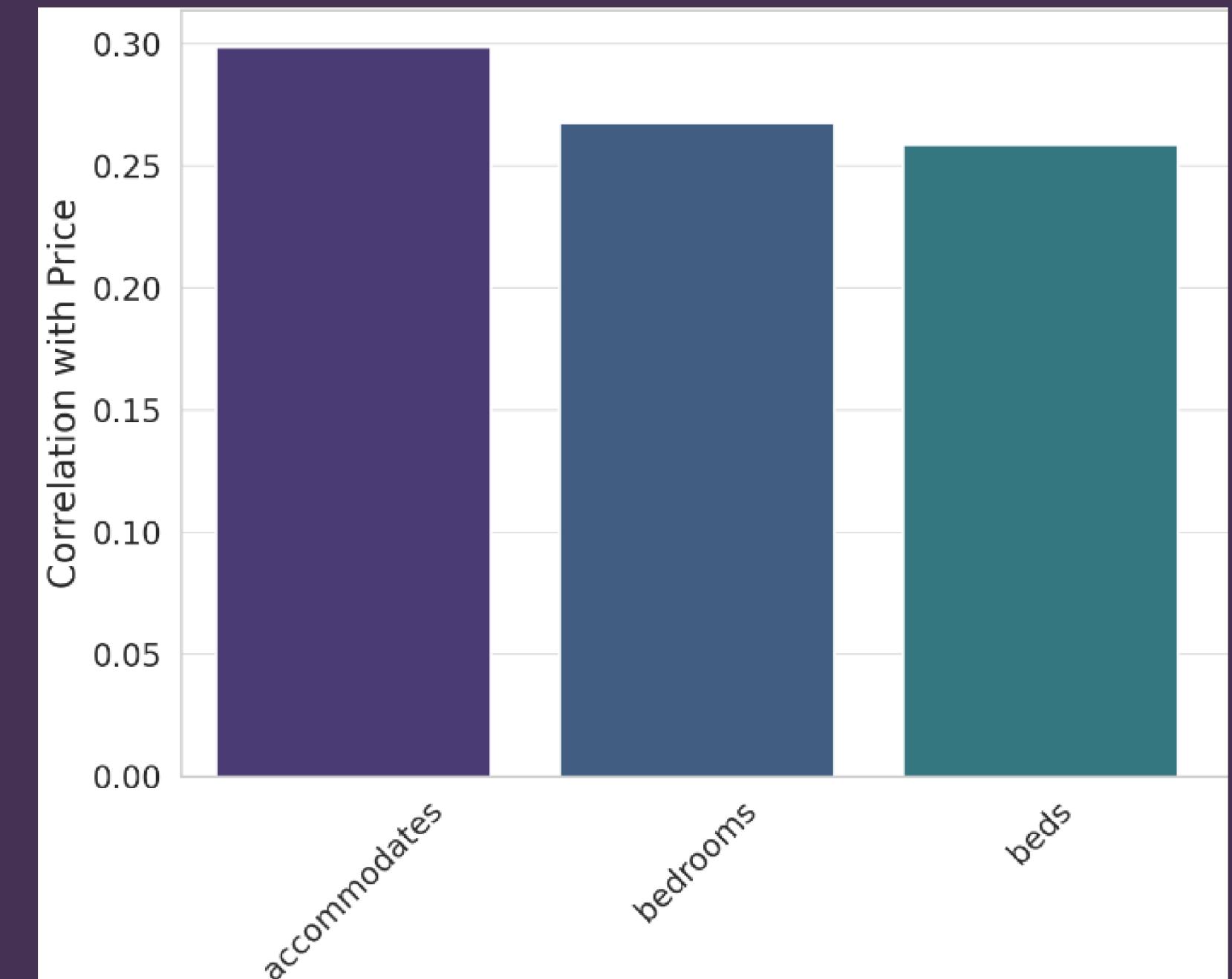
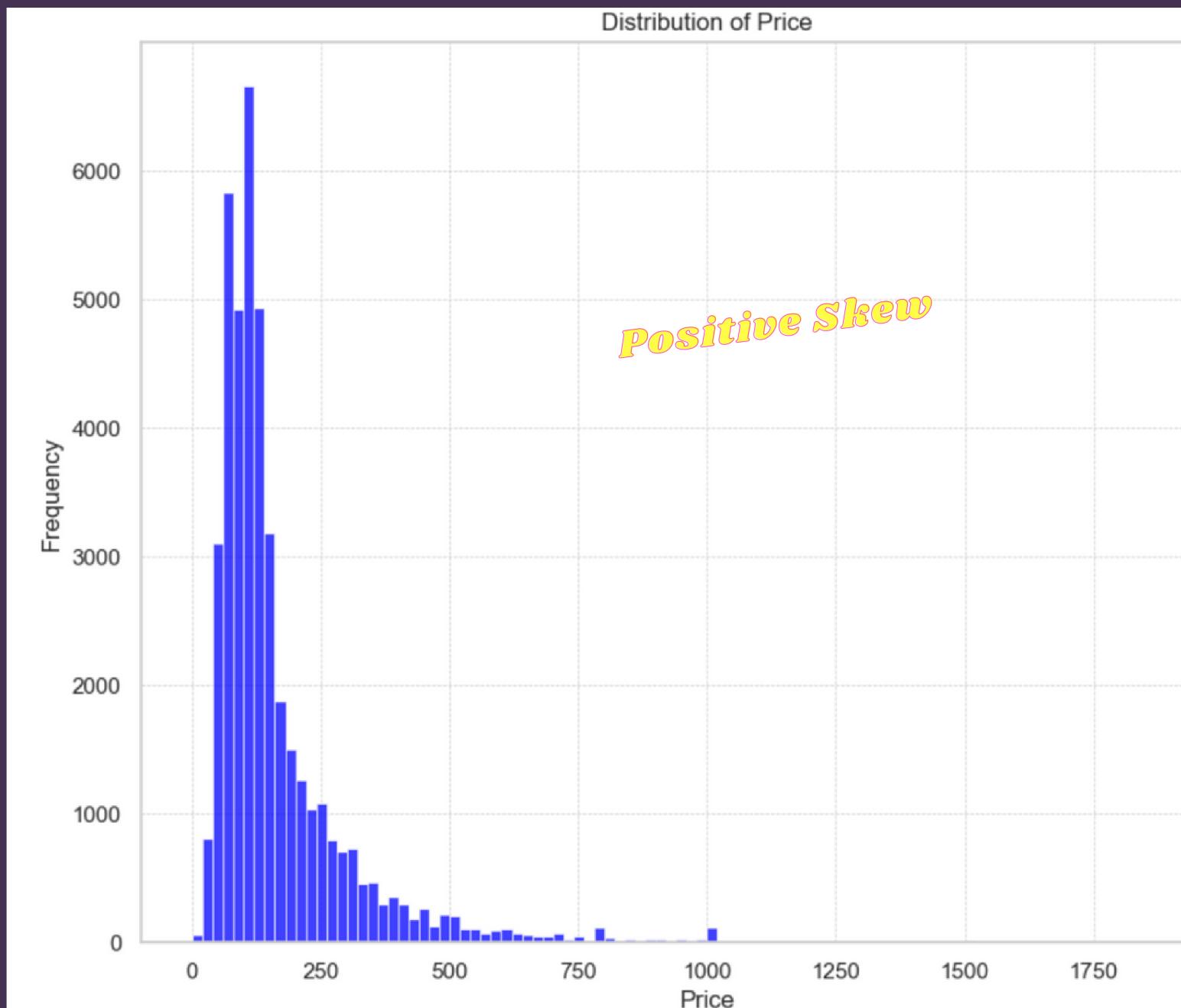
- Inside Airbnb



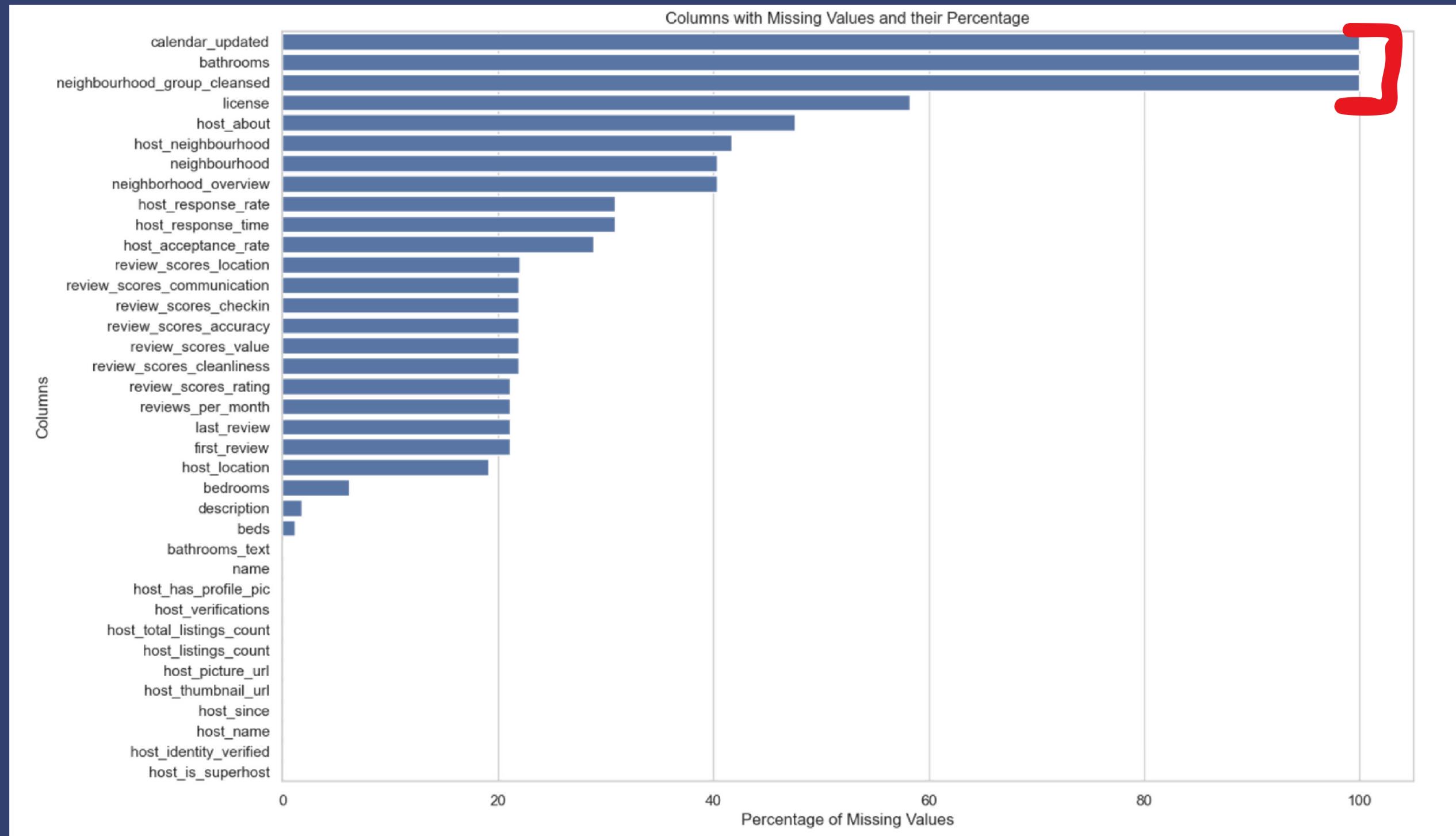
PRICE / DAY

OUTLIER
ALERT!

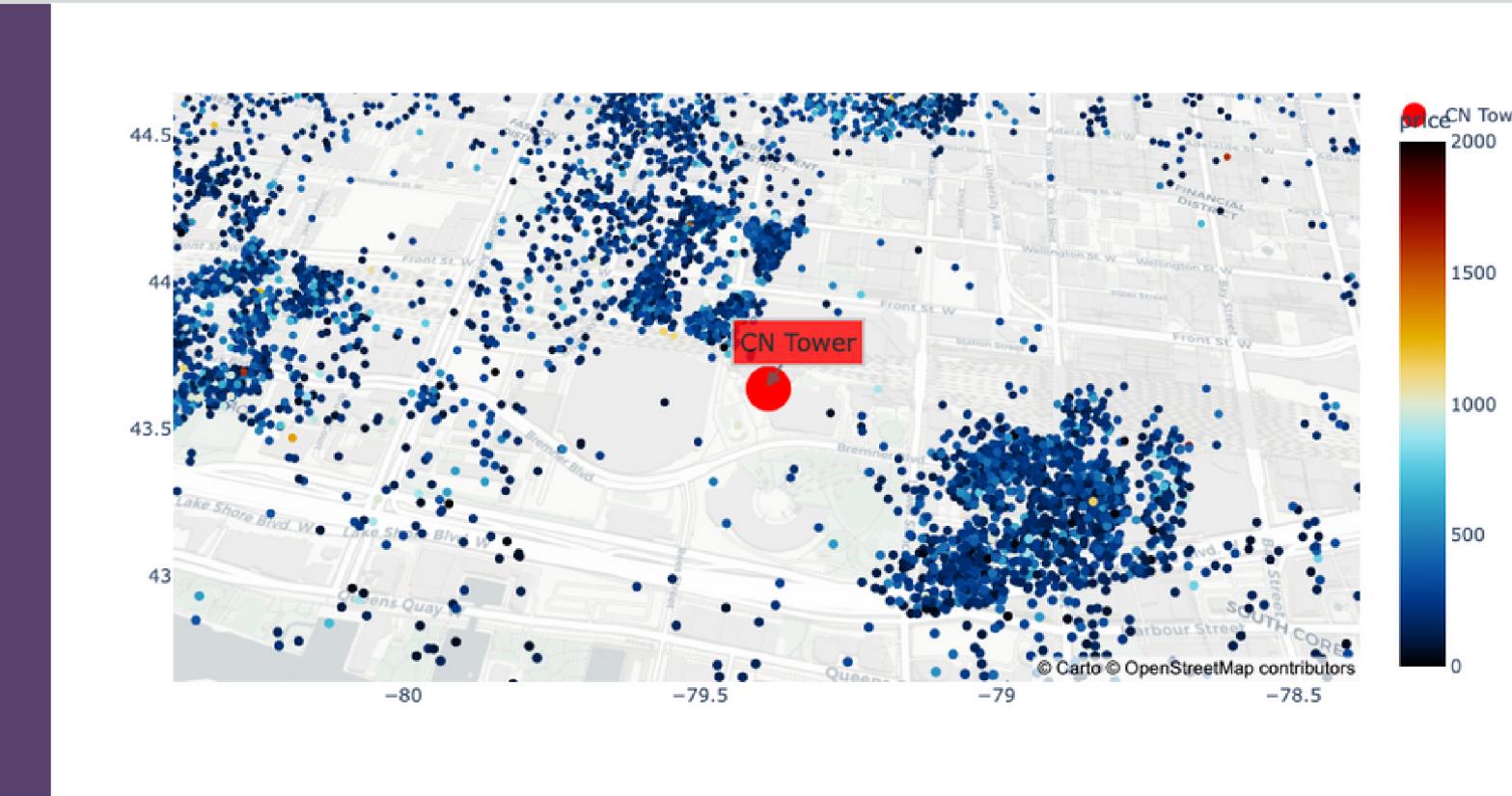
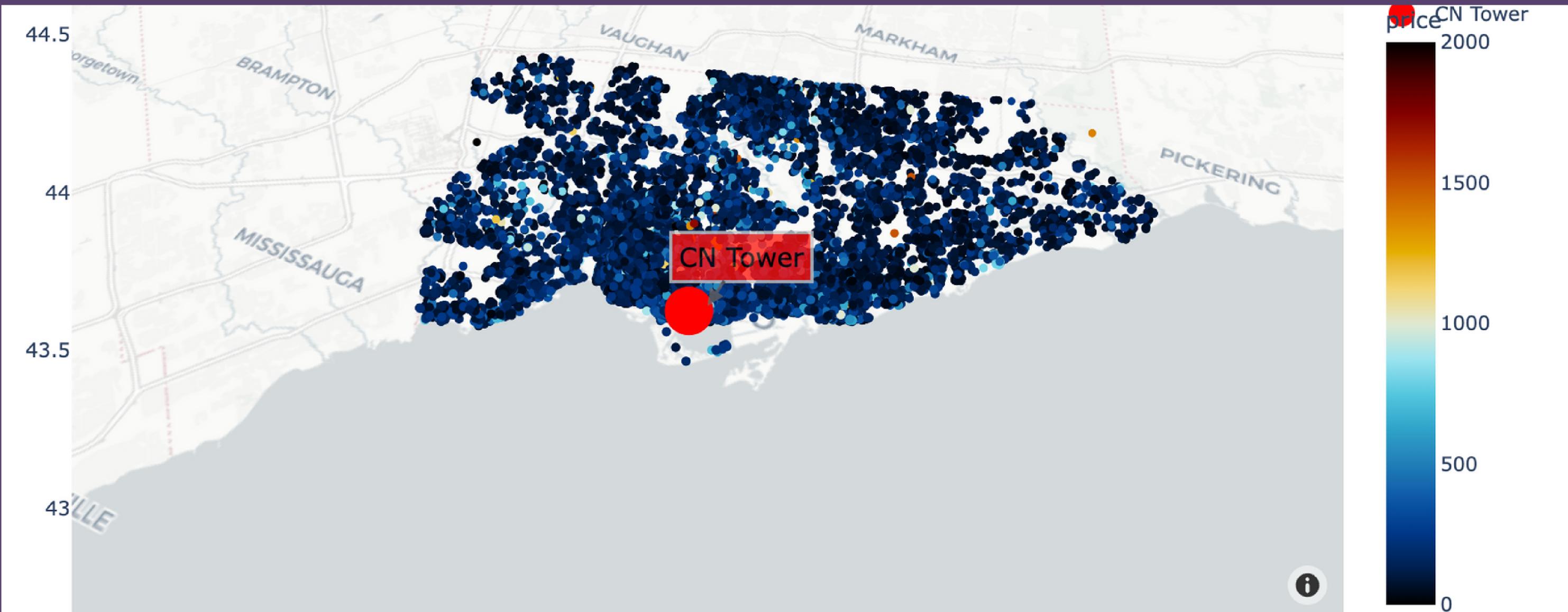
df_combined["price"].describe()	
count	42889.000000
mean	182.290955
std	768.800200
min	0.000000
25%	85.000000
50%	120.000000
75%	192.000000
max	110114.000000
Name:	price, dtype: float64

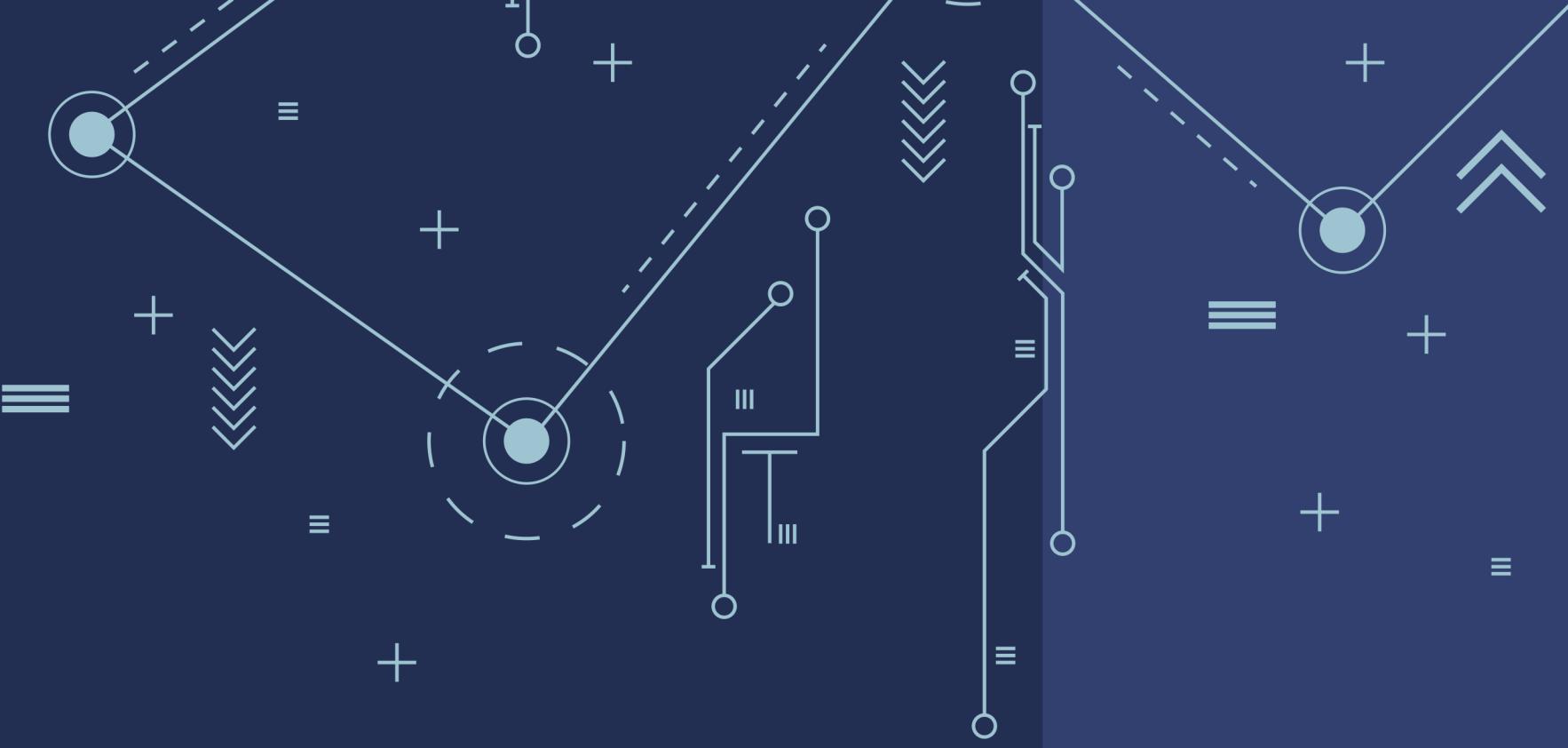


NULU VALUES



TORONTO LISTINGS WITH PRICE





WHATS NEXT!



DATA PROCESSING

- Null Values: Imputation Or Dropping
- Dummy Variables for nominal categorical columns



FEATURE ENGINEERING

- Use the 'Host_since' column to derive host tenure
- Word Embedding with Customer Reviews



BASE MODEL

- Linear Regression Model

Q and A

ASK AWAY!

