



PREDICTING AIRBNB PRICE

How can we determine the optimal price?

Nurdin Sulaeman

AIRBNB SINGAPORE

TABLE OF CONTENT

- **Background**
- **Methodology**
- **Data Exploration**
- **Optimizing Model**
- **Conclusion**



BACKGROUND

PENDAHULUAN

Airbnb adalah sarana penyewaan untuk sebuah properti dan sejenisnya yang berbasis online yang memungkinkan seorang penggunanya mendaftarkan atau menyewa properti tersebut untuk digunakan dalam jangka pendek.

Aturan untuk nominal harga sewanya pun ditetapkan oleh sang pemilik properti. Airbnb sendiri telah berdiri sejak bulan Agustus 2008 dan berkantor pusat di San Francisco, California



"Pricing is very important for business."

HOW CAN WE DETERMINE THE OPTIMAL PRICE?



Menentukan harga sewa rental properti memainkan peran yang sangat penting dalam berbisnis. Strategi harga yang tepat dapat meningkatkan revenue, tapi strategi harga yang salah justru dapat menimbulkan kerugian.

- Too High Price : Customer akan berpindah ke kompetitor
- Too Low Price : Pemilik properti & Company akan mengalami kerugian



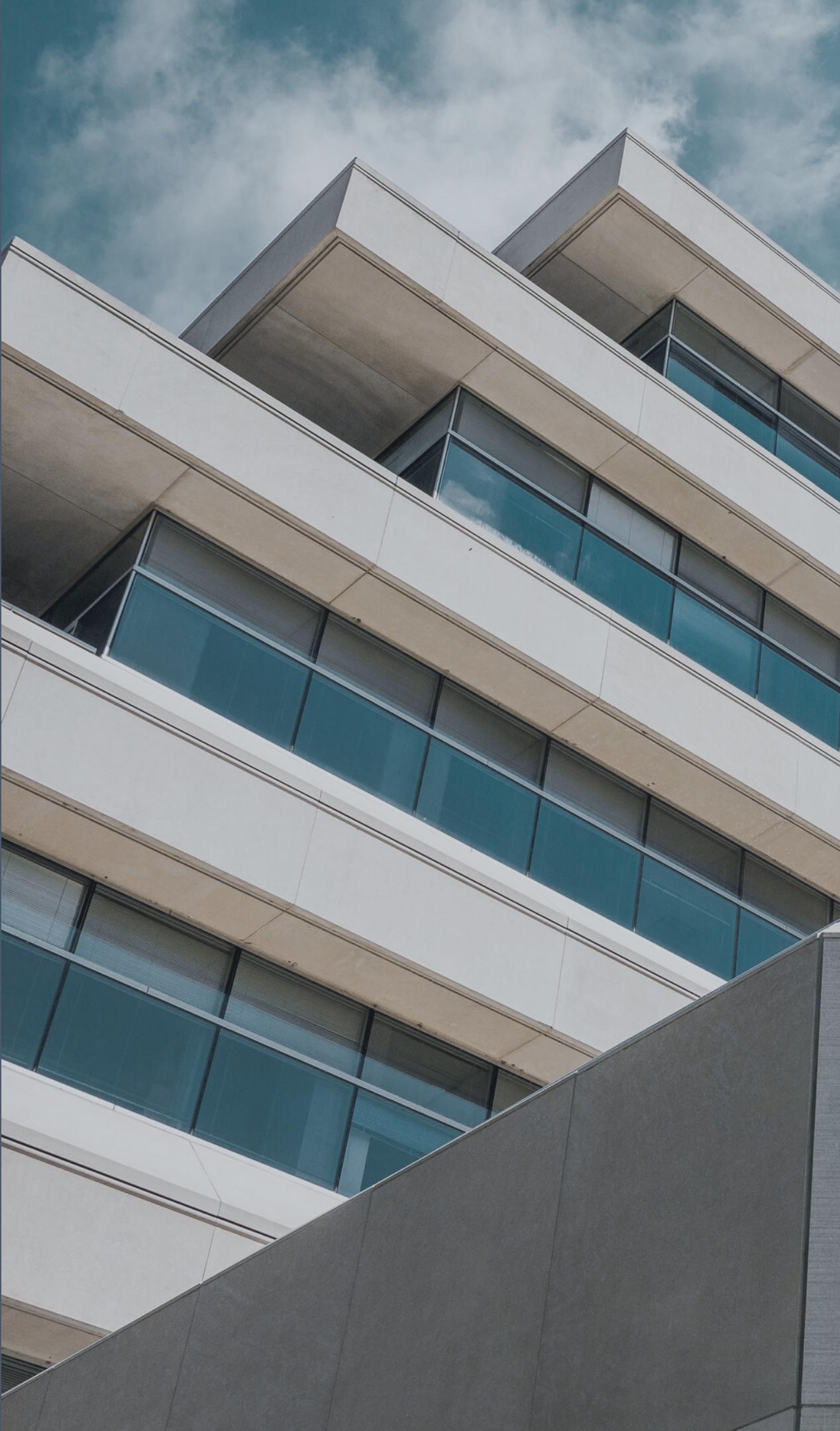
APPROACH

Membangun sebuah model Regresi untuk memprediksi harga sewa yang optimal berdasarkan kategori dan fasilitas properti AirBnB Singapore

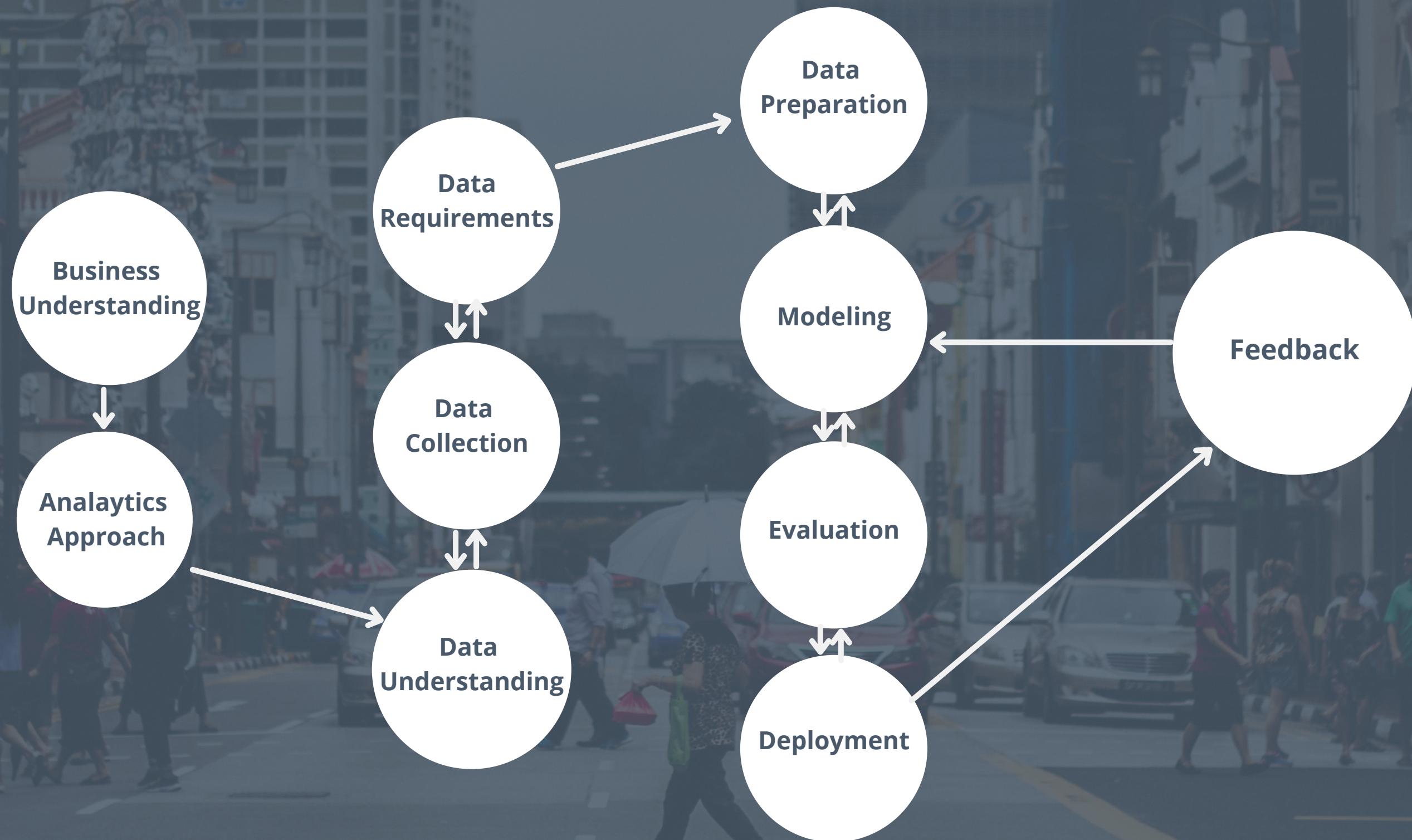


GOALS

Menentukan strategi harga yang tepat dan memaksimalkan potensi income dari rental properti di AirBnB Singapore



METHODOLOGY



column	Data Type	Nan	Nan(%)	N unique	Sample
dataset_id	int64	0	0.00	4221	[50646, 71509, 71896, 71903, 275343, 275344, 2...
listing_url	object	0	0.00	1	
scrape_id	int64	0	0.00	1	[20210929043603]
last_scraped	object	0	0.00	1	[2021-09-29]
name	object	0	0.00	1	
description	object	137	3.23%	1	• 61 Features • 4213 Rows
neighborhood_overview	object	1455	34.48%	1	The serenity & quiet surrounding makes it an ideal bedroom with a nice view on t...
picture_url	object	0	0.00	3885	[https://a0.muscache.com/pictures/36342984/670...
host_id	int64	0	0.00	1	
host_url	object	0	0.00	1	Inside AirBnB
host_name	object	8	0.19	800	http://insideairbnb.com/get-the-data.html
host_since	object	8	0.19	912	[2010-09-08, 2011-01-29, 2011-11-24, 2011-12-1...

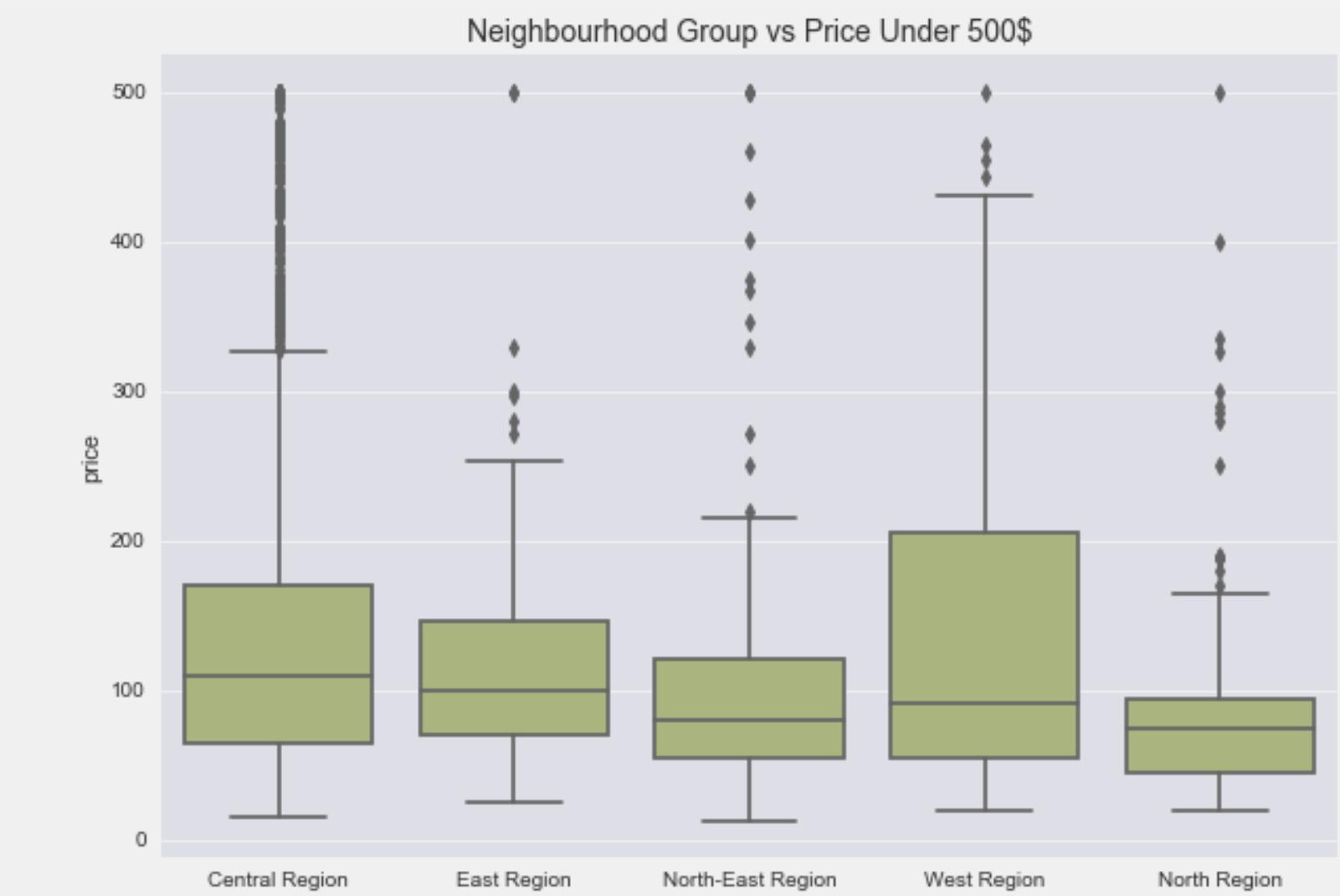
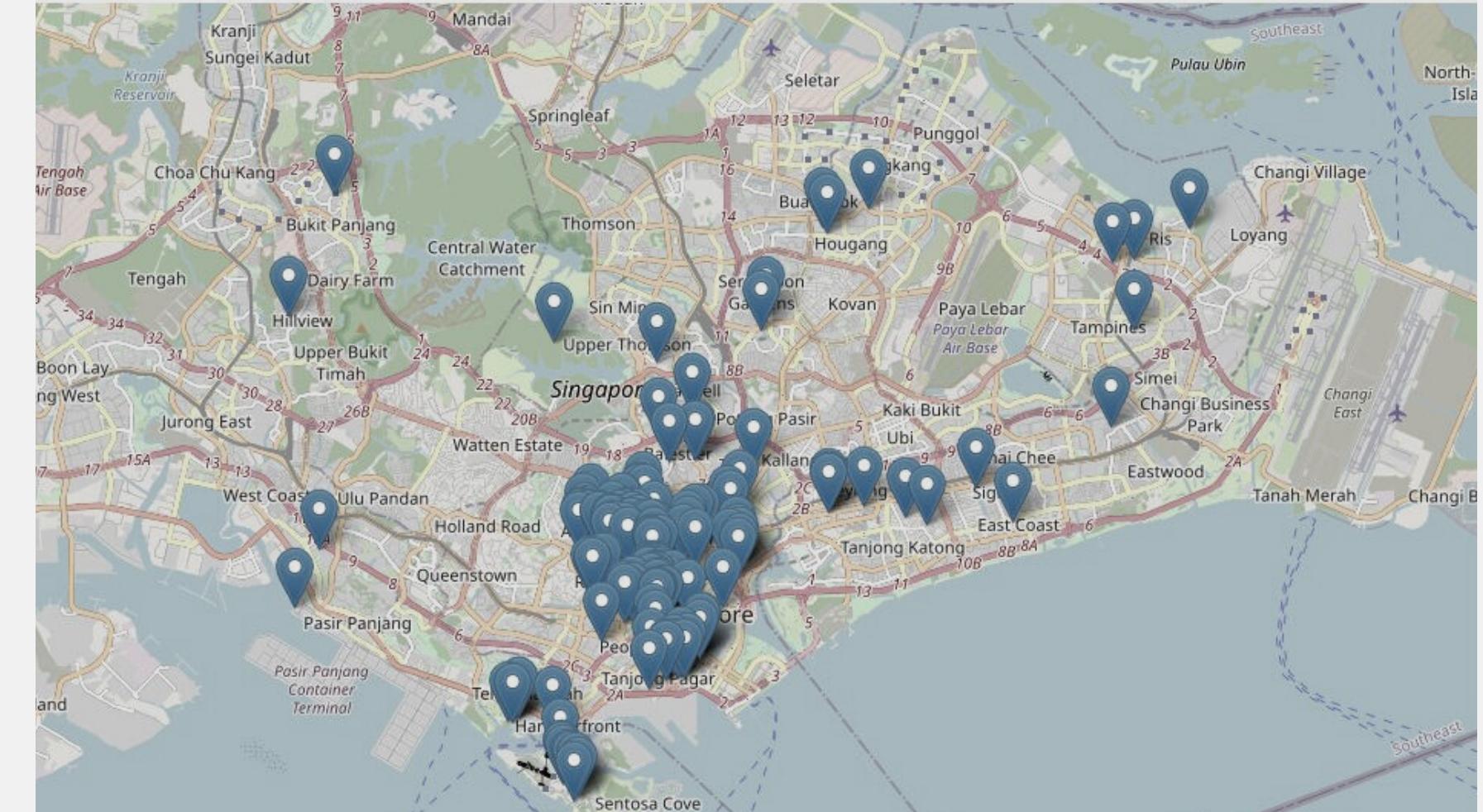
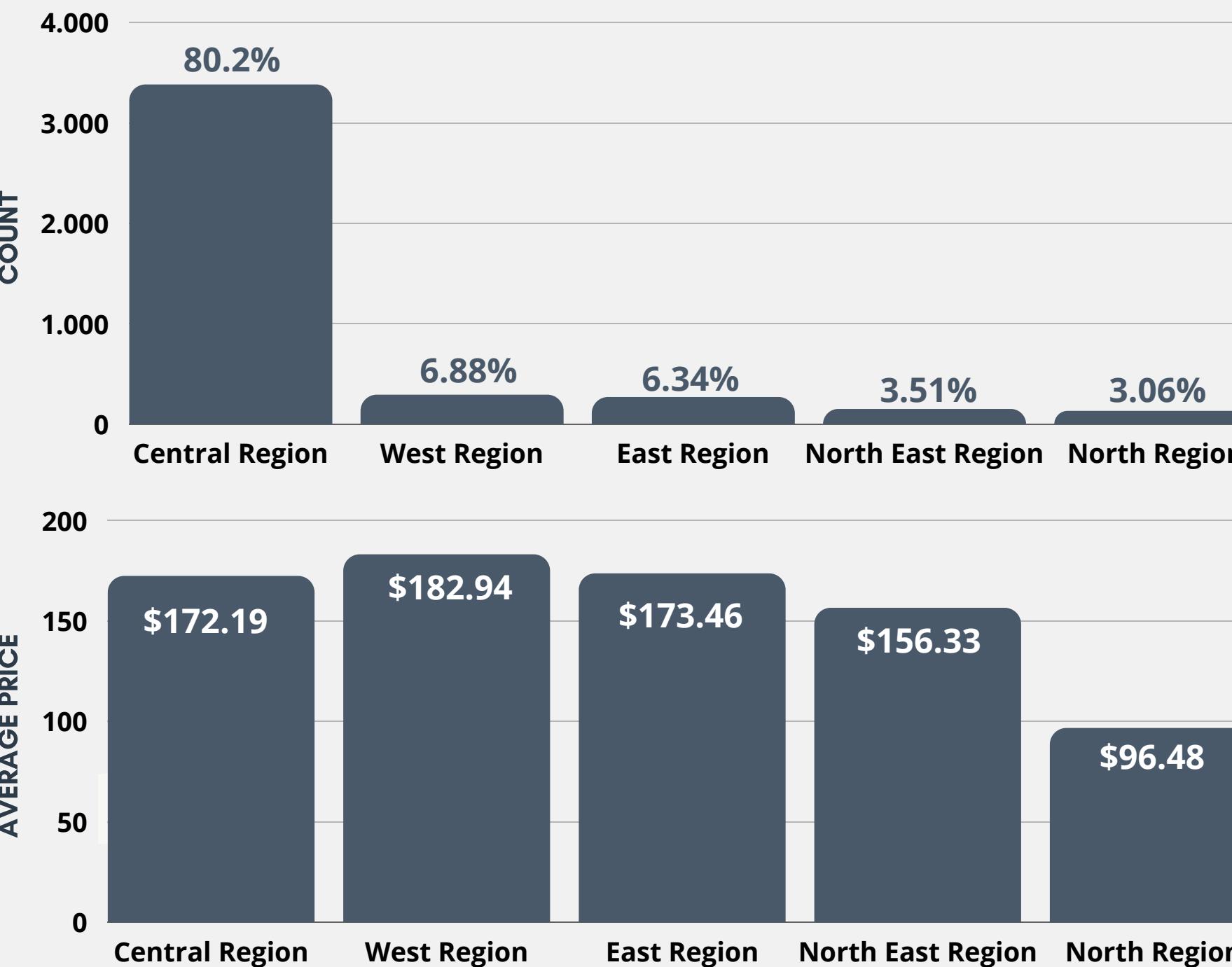
column	Data Type	Nan	Nan(%)	Nunique	Sample
id	int64	0	0.00	4221	[50646, 71609, 71896, 71903, 275343, 275344, 2...
listing_url	object	0	0.00	4221	[https://www.airbnb.com/rooms/50646, https://w...
scrape_id	int64	0	0.00	1	[20210929043603]
last_scraped	object	0	0.00	1	[2021-09-29]
name	object	0	0.00	3905	[Pleasant Room along...
description	object	1	34.47	1303	ly furnished bed...
neighborhood_overview	object	1455	34.47	1303	[The serenity & qu...
picture_url	object	0	0.00	3885	[https://a0.muscache.com/pictures/36342984/670...
host_id	int64	0	0.00	1128	[227796, 367042, 1439258, 1490000, 1521514, 15...
host_url	object	0	0.00	1128	[https://www.airbnb.com/users/show/227796, ht...
host_name	object	8	0.19	865	[Sujatha, Belinda, Joyce, B., Elizabeth, Rahul...
airbnb	host_since	8	0.19	912	[2010-09-08, 2011-01-29, 2011-11-24, 2011-12-1...

Data Exploration

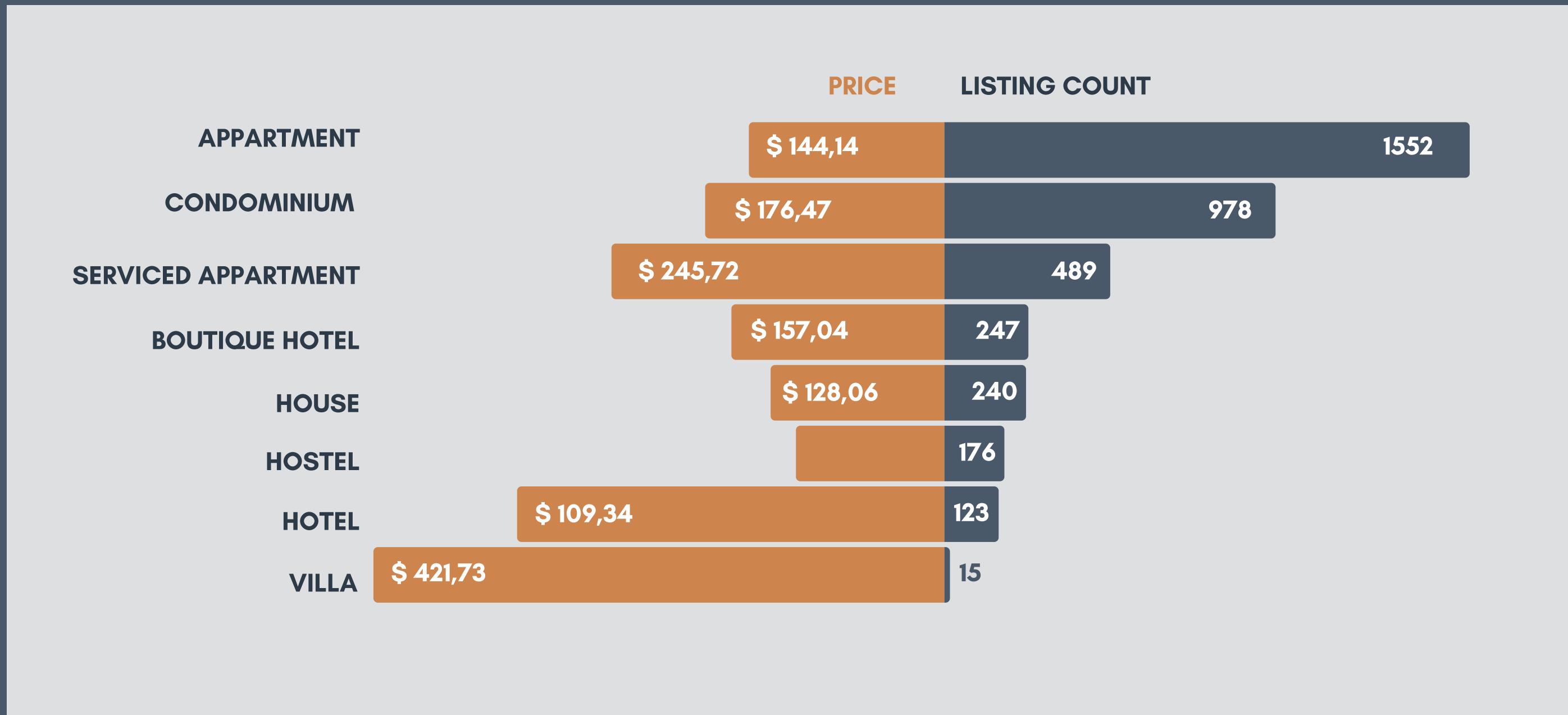


LOCATION

- 80% Listing AirBnB Singapore terletak di kawasan Central Region
- Kawasan Central Region, East Region dan West Region memiliki harga sewa yang cenderung lebih mahal



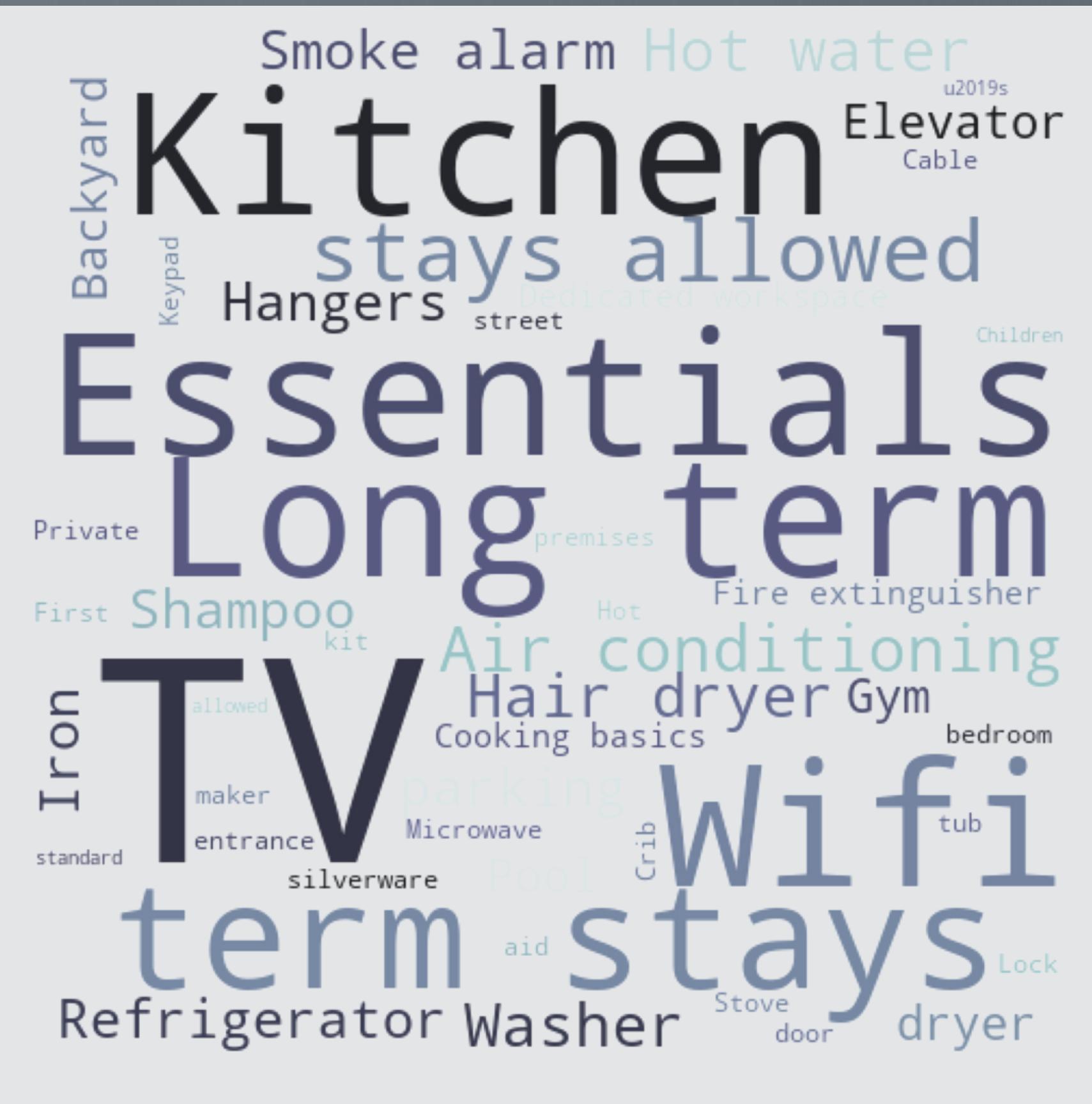
PROPERTY TYPE



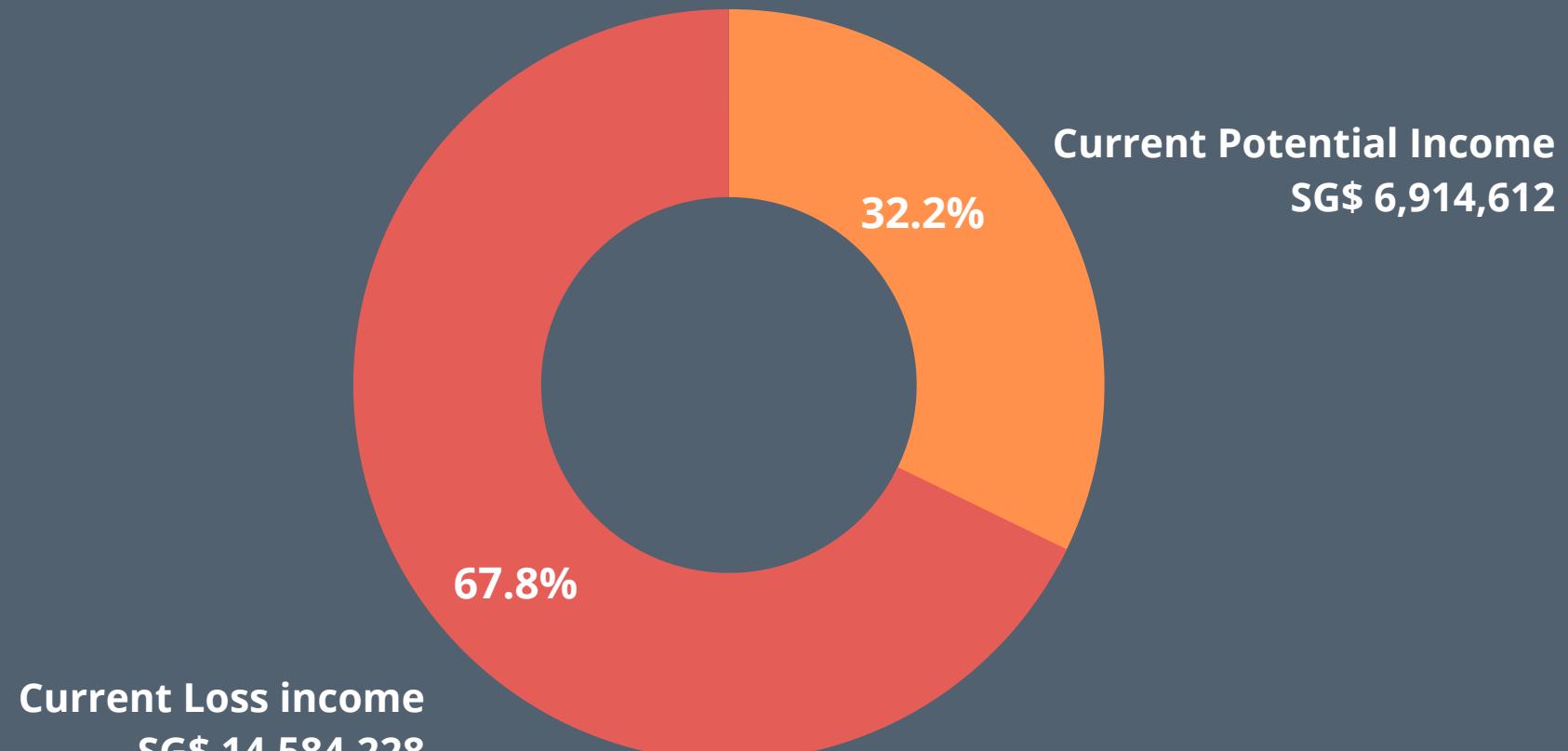
- Appartment, Condominium dan Serviced Appartment menjadi type properti listings yang mendominasi di AirBnB Singapore
- Jenis properti / listing Villa hanya ada sedikit di AirBnB Singapore, namun rata-rata harga yang ditawarkan cukup mahal untuk tipe listing tersebut

AMENITIES

Berikut adalah Top 20 dari fasilitas atau perlengkapan yang ditawarkan dari setiap listing/rental



WHAT HAPPENED?



Dari dataset kita mencoba menganalisa potensi income untuk semua listing selama 30 hari kedepan. Ditemukan bahwa masih ada 20.4% listing yang masih kosong selama 30 hari kedepan

Potensi **maksimal** income dalam 30 hari kedepan bisa mencapai SG\$ 21,498,840, namun potensi income saat ini hanya SG\$ 6,914,612 Ini sama dengan 32.2% dari maksimum potensi income yang bisa di dapatkan.

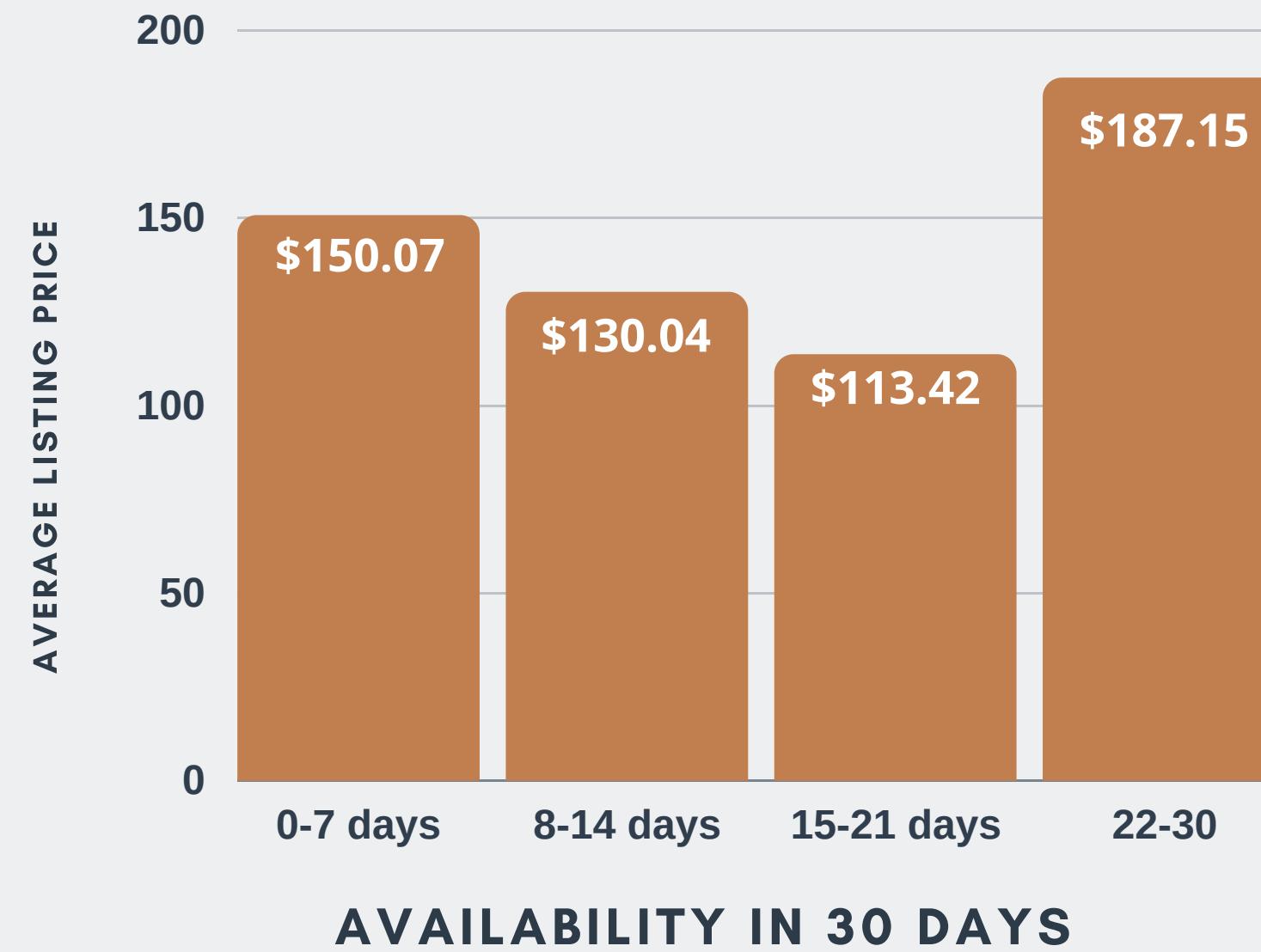
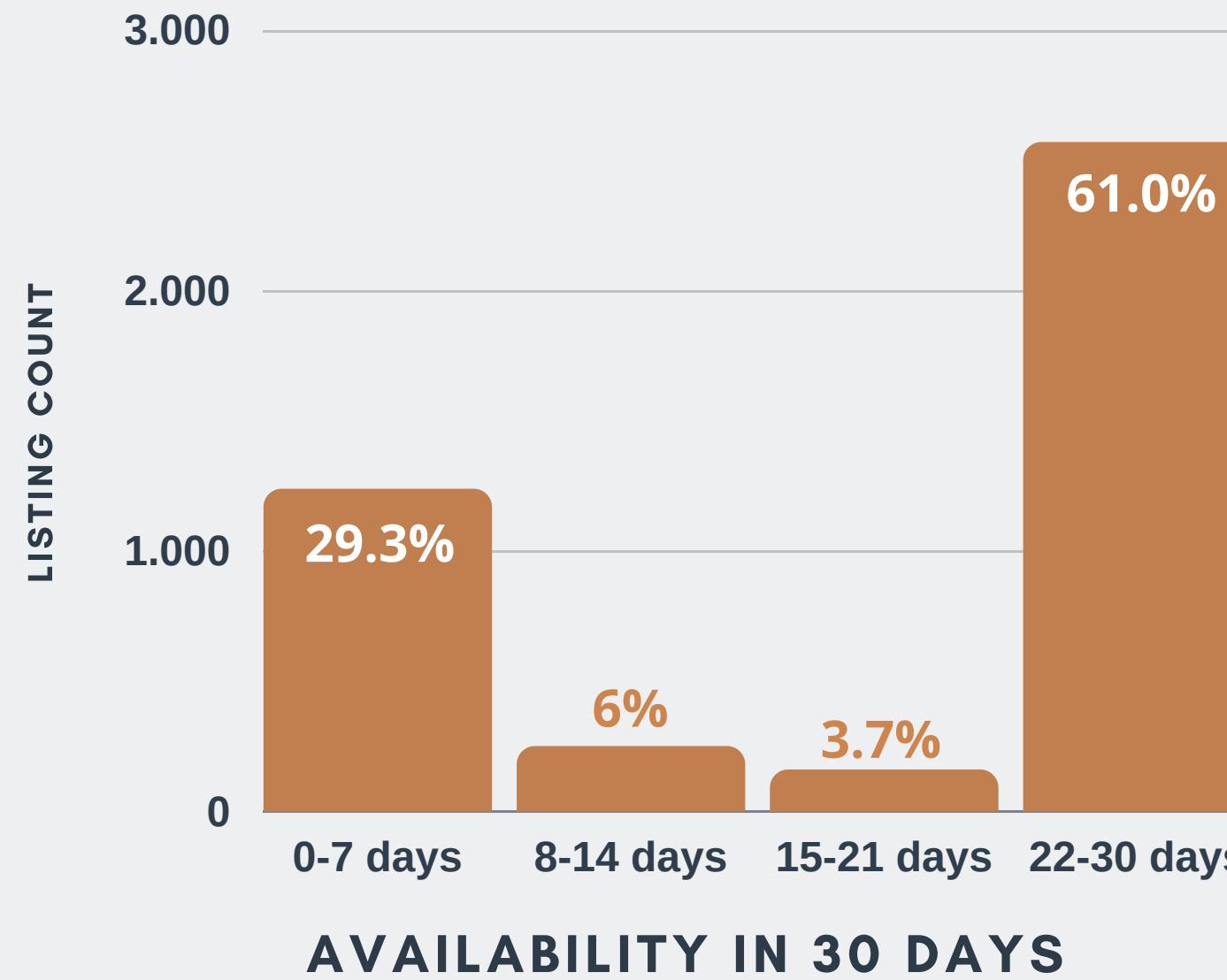
20.4%

RENTAL PROPERTI YANG TIDAK TERISI

WHY DID IT HAPPEN?

Selain dari rental kosong tadi. Ditemukan juga ada sebanyak 61.0% rental yang kosong dalam jangka waktu 22-30 hari kedepan, sayangnya rental inipun memiliki rata-rata harga yang cukup tinggi.

Kita perlu melakukan penyesuaian harga untuk rental tersebut, sehingga dapat mengurangi ketersediaan (kekosongan) dan akhirnya meningkatkan revenue



FEATURE ENGINEERING AND MODELING



DATA PREPARATION

- Setelah dilakukan feature selection, terpilih 49 feature dari 73 feature untuk digunakan dalam pemodelan
- Setelah dilakukan proses Feature Engineering, dataset yang siap untuk dilakukan pemodelan kini menjadi 67 feature

HANDLING MISSING VALUES

Feature yang masih memiliki missing value ada pada beberapa feature "review". Untuk metodenya sendiri kita menggunakan constant -1

DATA SPLITTING

75% Data training, 25% data testing

ONE HOT ENCODING

Feature "neighbourhood_group" & "room_type"

BINARY ENCODING

Feature yang memiliki kategori cukup banyak di encode menggunakan Binary Encoding untuk mencegah terlalu banyaknya feature ketika pemodelan

Model Selection

Model	R-square(%)
Decision Tree	37.557
Linear Regresion	45.566
Adaptive Boosting	39.005
Gradient Boosting	41.077
XGBoosting	64.898

Kita mencoba 5 based model untuk menentukan benchmark. Dan yang terpilih adalah model XGBoosting dengan score R-square 64.898%.

Best Parameter After Tuning

Parameter	Best
learning_rate	0.01
max_depth	5
n_estimators	4000
random_state	40

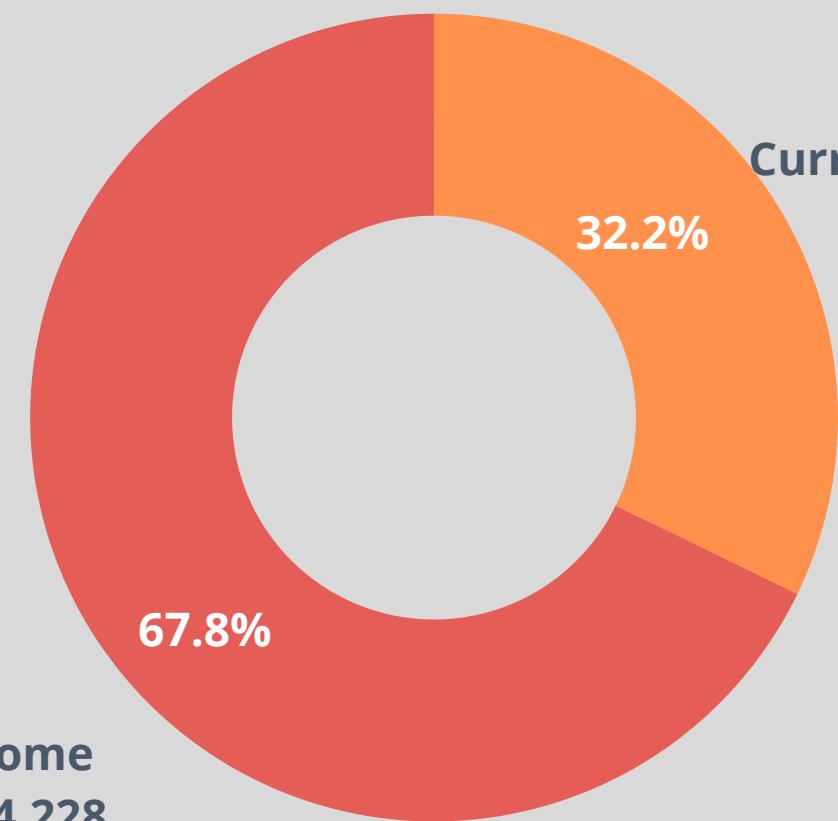
Best Model Result

XGBoosting	
R-square	75.433%
MAE	27.442
MSE	2101.261

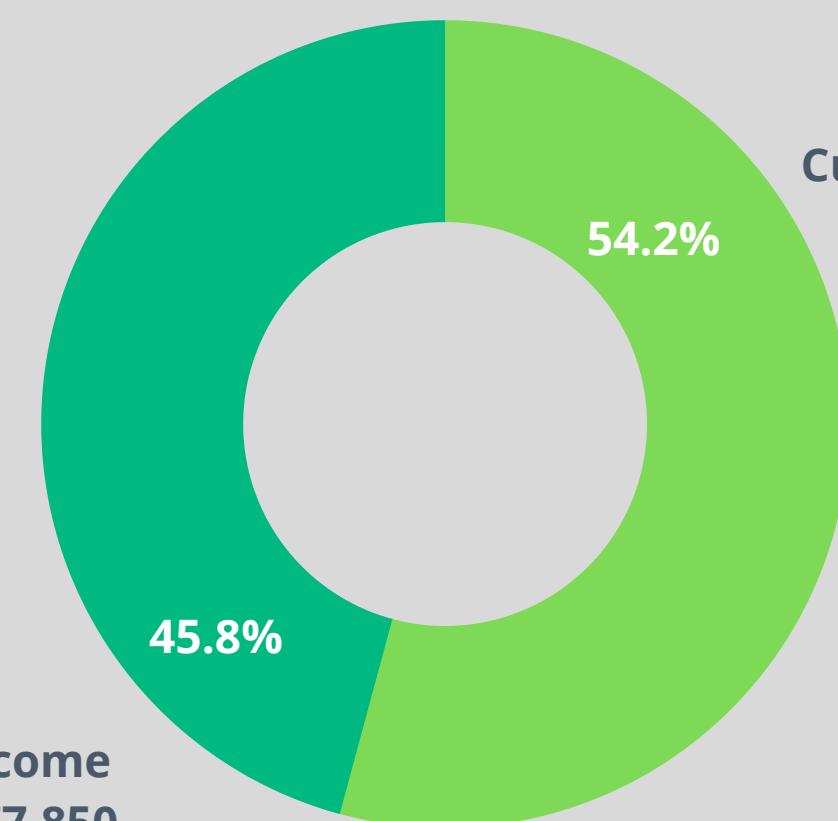
Model terbaik yang kita dapatkan setelah hyperparameter tuning. Dengan score R-square 75.433% dengan error (MAE)27.442

Model Xgboost kita lakukan hyperparameter tuning, dengan beberapa grid parameter menggunakan metode Cross Validation dengan n_splits = 5

RESULT



Current Potential Income
\$ 6,914,612



Current Potential Income
\$ 11,563,137

Max Omzet	\$ 21,498,840
Current Income	\$ 6,914,612
Current Loss Income	\$ 14,584,228

Max Omzet	\$ 21,337,988	⬇️ 0.007%
Current Income	\$ 11,563,137	⬆️ 40.020%
Current Loss Income	\$ 9,77,4850	⬇️ 32,977%

HOW TO MAKE IT HAPPEN?

- Cost, Location, Reviews, Length of Stay dan availability adalah group feature yang digunakan dalam pemodelan dan Price sebagai target dari model
- Kita bisa melakukan **Adjustment** (penyesuaian) berdasarkan length of stay dan availability menggunakan model yang telah kita bangun dalam hal ini kita menggunakan model Xgboosting untuk memprediksi harga yang optimal
- Kita melakukan penyesuaian harga terhadap listing/rental yang memiliki availability yang tinggi atau dalam hal ini listing yang availabilitinya 22-30 hari
- Penyesuaian yang dibuat hanya sebagai rekomendasi dalam menentukan harga listing/rental
- **Adjustment threshold**, harga minimal dari setiap listing/rental
- **Example Adjustment**, model memprediksi potensi income saat ini yang bisa di dapatkan naik dari **\$ 6,914,612** menjadi **\$ 11,563,137**

THANKYOU



<https://github.com/nurdinsulaemann>

View Process and Code:

- Data Cleaning : https://bit.ly/Airbnb_CleaningData
- Analyze and Modeling : https://bit.ly/Airbnb_Analyze_Modeling