

# Hotel Aggregator Analysis

A Power-BI Project

PRESENTATION BY : BOLORAM MITRA



# Table of contents

- 02
- 03 ABOUT THE PROJECT
- 04 PROBLEM STATEMENT
- 05 ABOUT THE DATASET
- 06 DELIVERABLES
- 07 ETL PROCESS IN POWER-BI
- 12 ANALYSIS & VISUALIZATION
- 17 FINDINGS & RECOMMENDATIONS
- 20 CONCLUSION



# About The Project

This internship project involves analyzing a dataset from a hotel aggregator using Power BI. The dataset includes a range of attributes related to listings, hosts, reviews, and availability. The goal is to develop detailed visualizations and insights to identify trends, patterns, and factors affecting the performance of these listings. Interns will use Power BI to explore important metrics such as pricing, availability, host characteristics, and review scores, aiming to generate actionable insights that can enhance the overall quality and competitiveness of the listings.

# Problem Statement

**Challenge:** The hotel aggregator platform is seeking to identify the key factors that contribute to the success of its listings in order to enhance overall quality and competitiveness

.

**Data:** A comprehensive dataset is available, containing detailed information about listings, hosts, reviews, and availability.

**Goal:** Utilize Power BI to analyze this dataset, uncovering trends and patterns that influence the performance of listings. The aim is to generate insights that can be used to improve the platform's offerings and better meet customer needs.



# About the Dataset



05

The dataset includes **23,185 listings** with **75 unique attributes**, covering aspects such as listings, hosts, reviews, and availability. **Key attributes includes:**

1. **id:** Unique identifier for each listing.
2. **listing\_url:** URL of the listing.
3. **scrape\_id:** Data scraping event identifier.
4. **last\_scraped:** Date of the last data scrape.
5. **source:** Source of the listing information.
6. **name:** Listing name.
7. **description:** Listing description.
8. **neighborhood\_overview:** Overview of the listing's neighborhood.
9. **picture\_url:** Listing picture URL.
10. **host\_id:** Unique identifier for the host.

Other columns provide details about **hosts, property types, room specifics, amenities, pricing, availability, and reviews.**

# Deliverables

## Interactive Dashboards:

- Geographical distribution
- Pricing trends
- Host characteristics
- Review scores

## Detailed Reports:

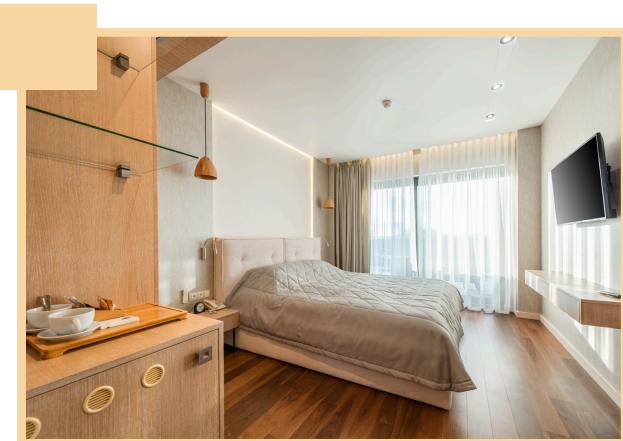
- Availability patterns
- Property and room preferences
- Factors influencing guest satisfaction

## Recommendations:

- Insights for hosts
- Strategies for improving the aggregator platform based on analysis results



# ETL Process in Power-BI



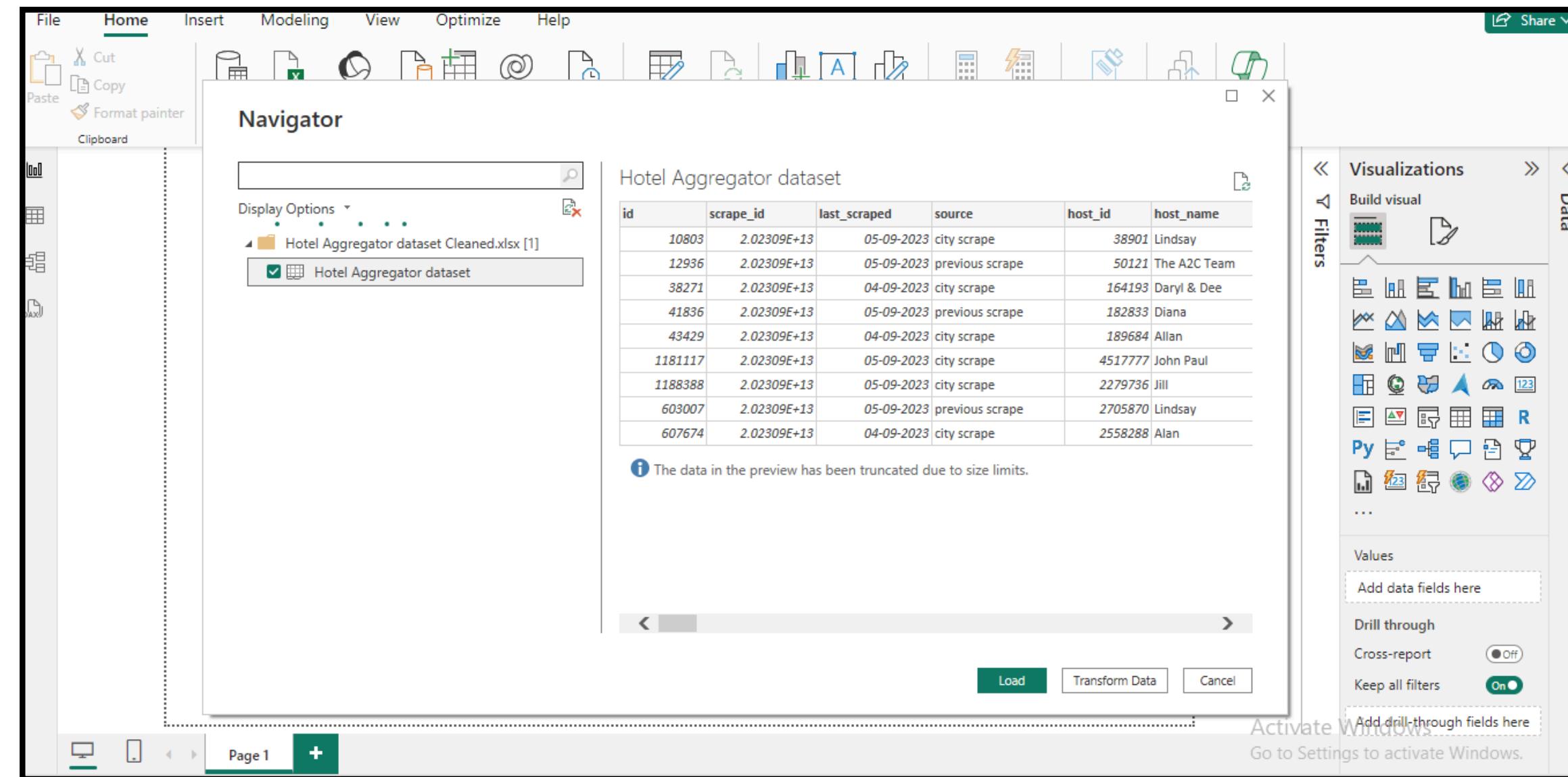
Connecting and  
Importing Data

Data Cleaning and  
Transformation

Data Analysis and  
Measures Creation

# Connecting and Importing Data

- **Data Source:** Imported the dataset from Excel into Power BI.
- **Initial Cleaning in Excel:** Performed basic data cleaning and column removal in Excel, addressing inconsistencies, formatting errors, and data quality issues.



# Data Cleaning and Transformation

- **Null and Irrelevant Data:** Checked for null, irrelevant, and duplicate columns or data. Replaced or removed as necessary to ensure data consistency.
- **Column Removal:** Removed irrelevant columns to streamline the dataset and focus on critical information for analysis.
- **Handling Missing Values:** Made informed decisions to handle missing values by imputing values based on criteria or excluding rows with excessive missing data.
- **Case Sensitivity:** Ensured data consistency by standardizing case sensitivity (e.g., converting 'f' to 'F' and 't' to 'T')

# Data Analysis and Measures Creation

- **DAX Functions and Measures:** Created DAX functions and measures for custom calculations within Power BI.
- **Calculated Metrics:** Developed measures to calculate total listings, total hosts, and available listings, providing a quick overview of the dataset's scope.
- **Host Listing Distribution:** Analyzed the distribution of host listings, identifying hosts with multiple listings for further analysis of host performance.

This ETL process ensures that the data is clean, consistent, and ready for detailed analysis and visualization in Power BI, allowing for more accurate and insightful reporting.

# Analysis & Visualization

- Geographic Distribution Analysis
- Pricing and Availability Analysis
- Host Performance Analysis
- Property and Room Type Analysis



# Geographic Distribution Analysis

Total Hosts

**23.19K**

Average Rating

**4.64**

Listed From

- Select all
- (Blank)
- 2009
- 2010
- 2011
- 2012
- 2013

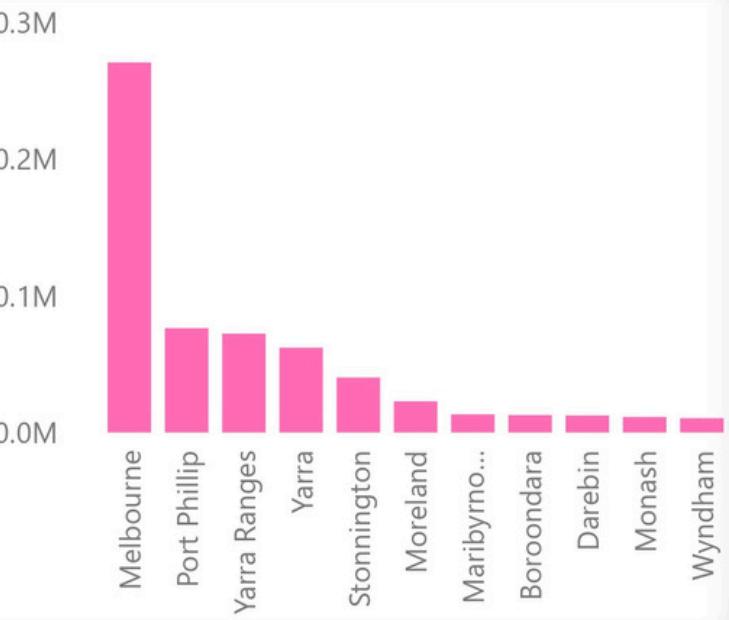
Host Verification

- Select all
- (Blank)
- No
- Yes

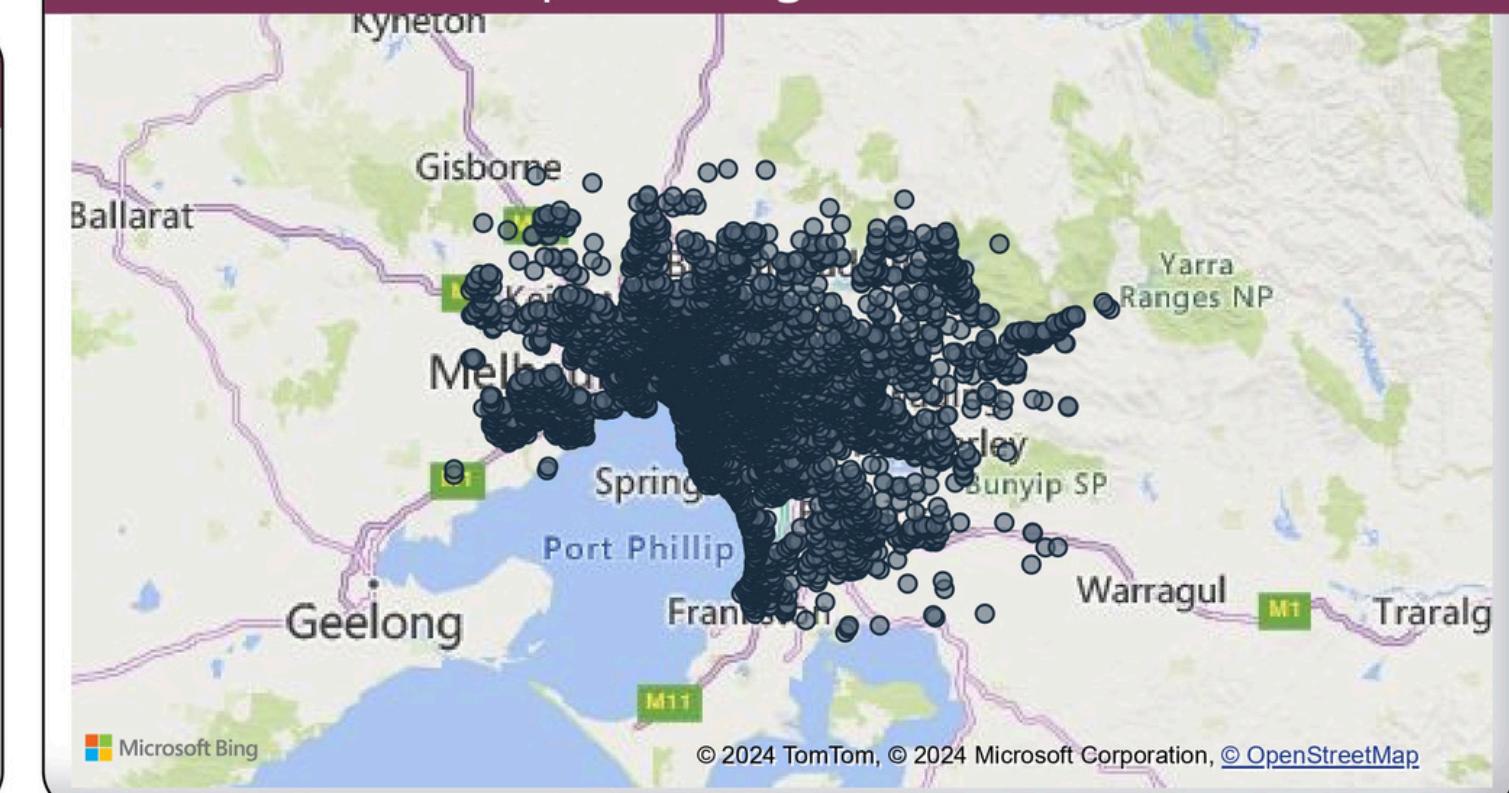
Neighbourhood

- Select all
- Banyule
- Bayside
- Boroondara
- Brimbank
- Cardinia
- Casey
- Darebin
- frankston
- Glen Eira
- Greater Dandenong
- Hobsons Bay
- Hume

Reviews availability



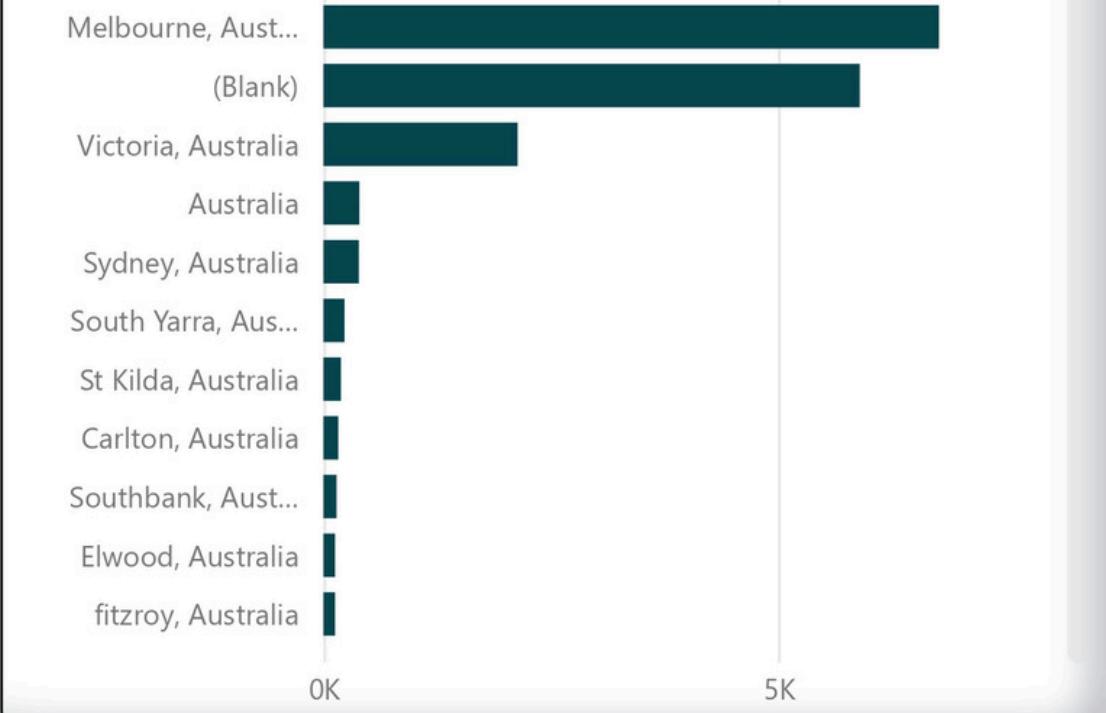
Popular Neighborhoods



Geographic Concentration



Host-Count in a Neighbourhood



# Pricing and Availability Analysis

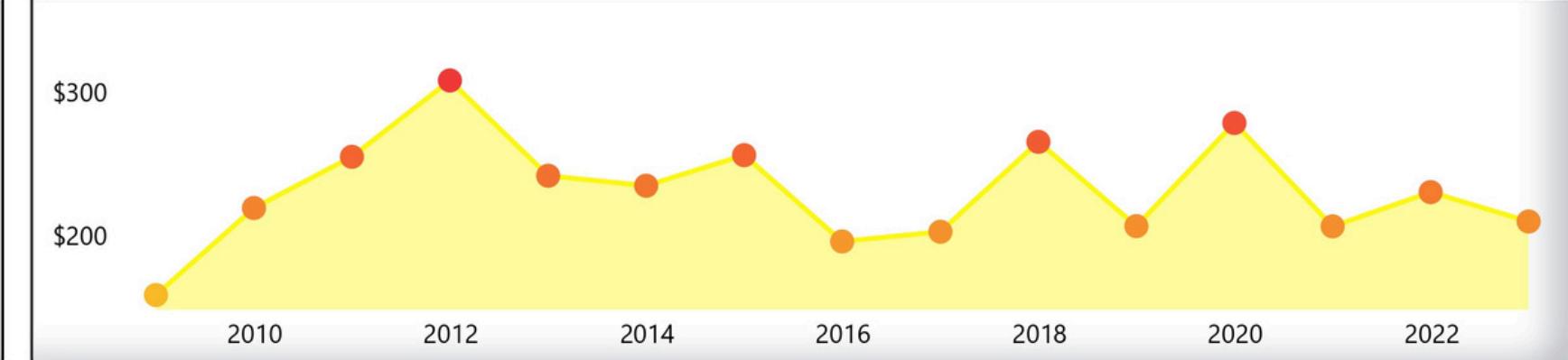
## Host Location

- Select all
  - (Blank)
  - Abbotsford, Australia
  - Abbotsford, Canada
  - Aberfeldie, Australia
  - Adelaide, Australia
  - Agnes Water, Australia
  - Aintree, Australia
  - Aireys Inlet, Australia
  - Airport West, Australia
  - Aix-en-Othe, France
  - Albanvale, Australia
  - Albert Park, Australia
  - Albion, Australia
  - Albury, Australia
  - Aldgate, Australia
  - Alexandra, Australia
  - Alexandria, Australia
  - Alice Springs, Australia
  - Alphington, Australia
  - Altona Meadows, Australia
  - Altona North, Australia

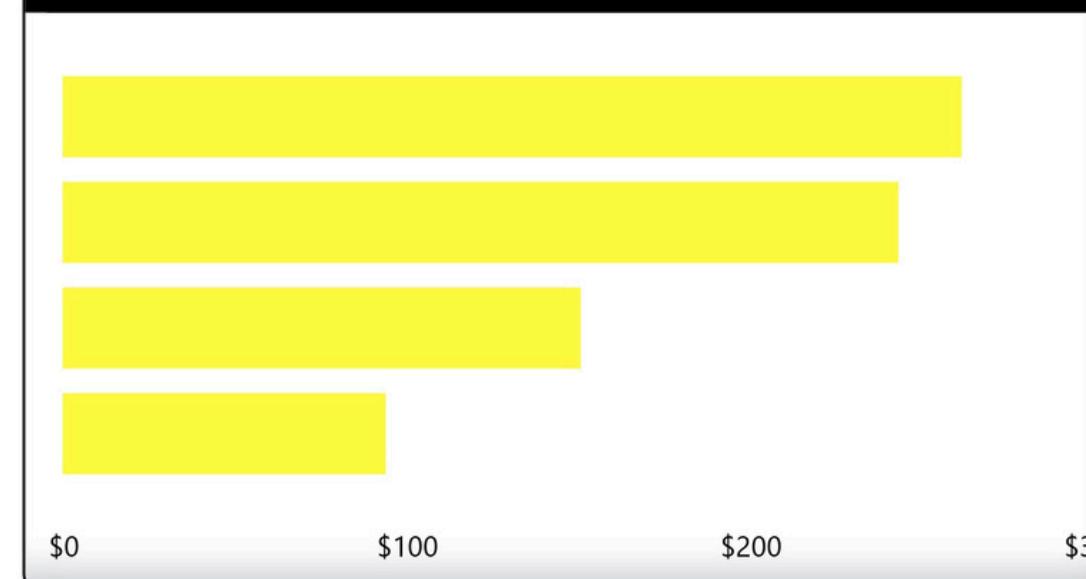
## Average Price by Property Type



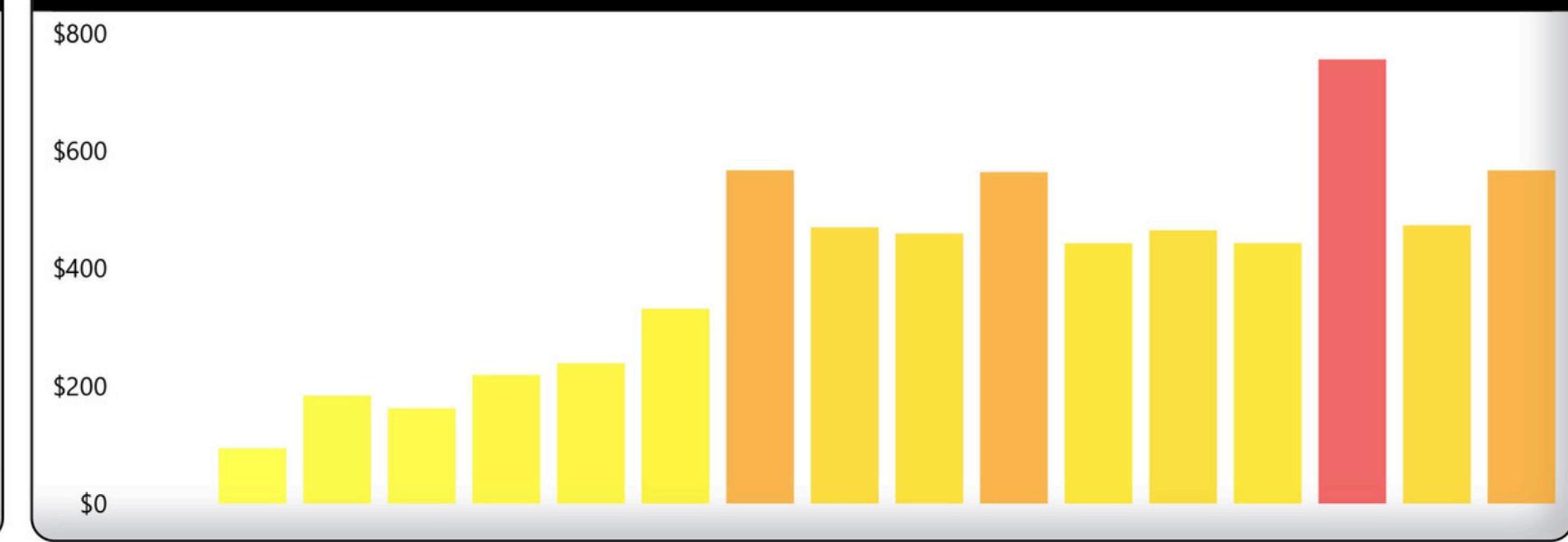
## Average of price by Year



## Average of price by Room Type



## Average of price by accommodates

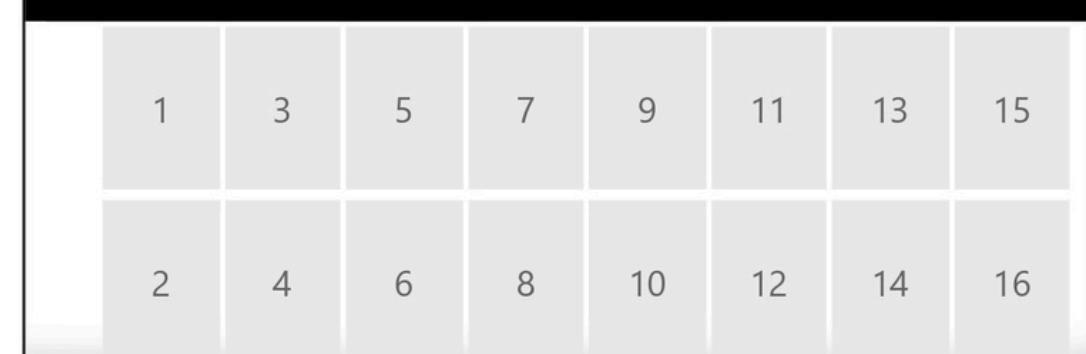


Select Year

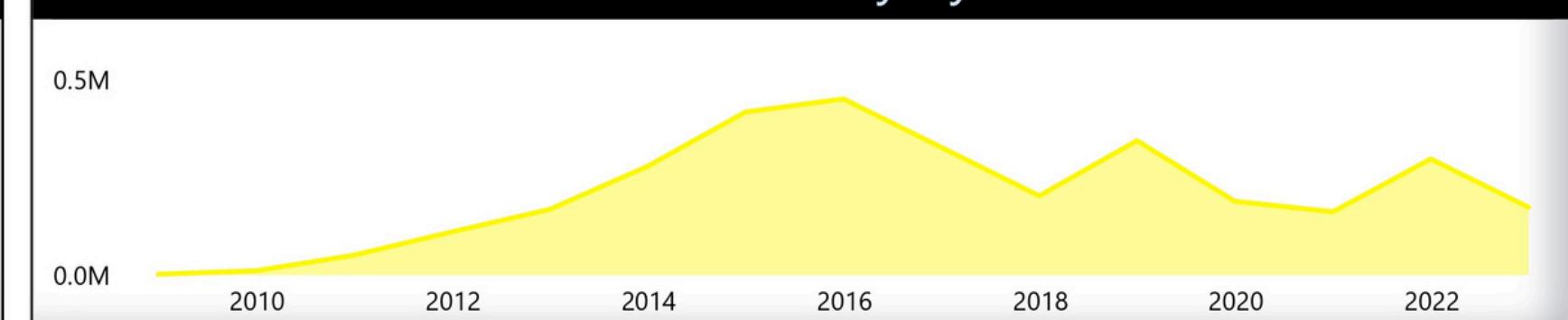
2009



### Select total accomodation



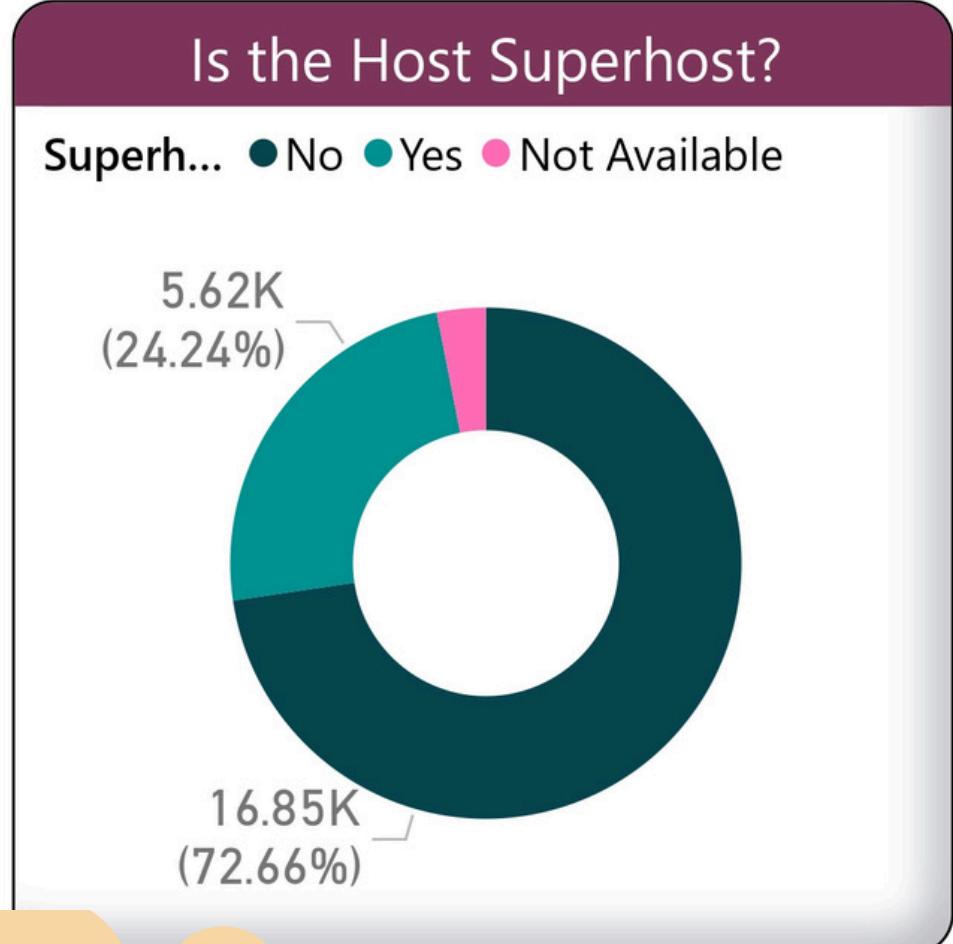
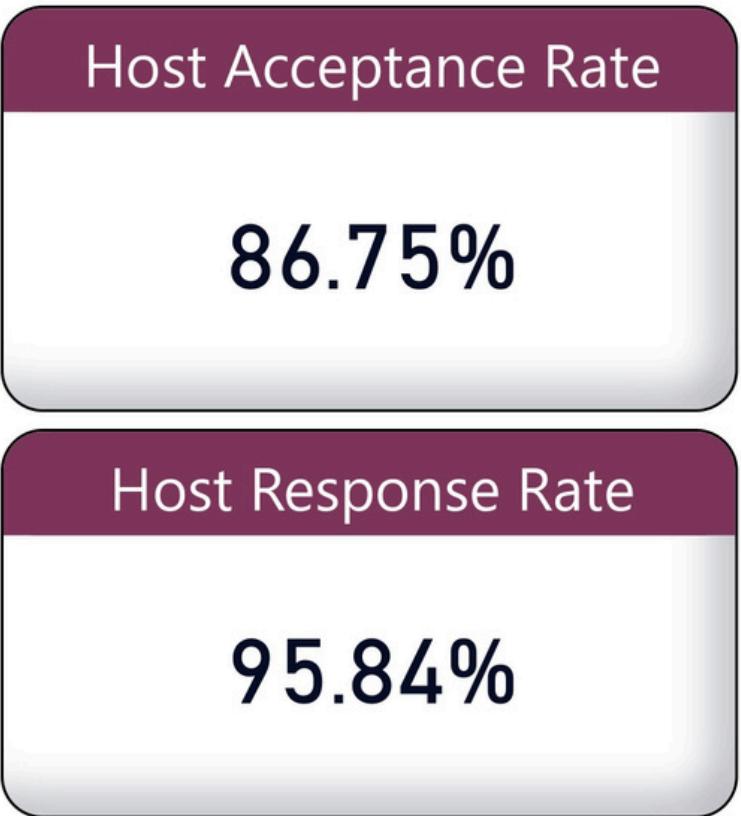
## Total availability by Year



# Host Performance Analysis

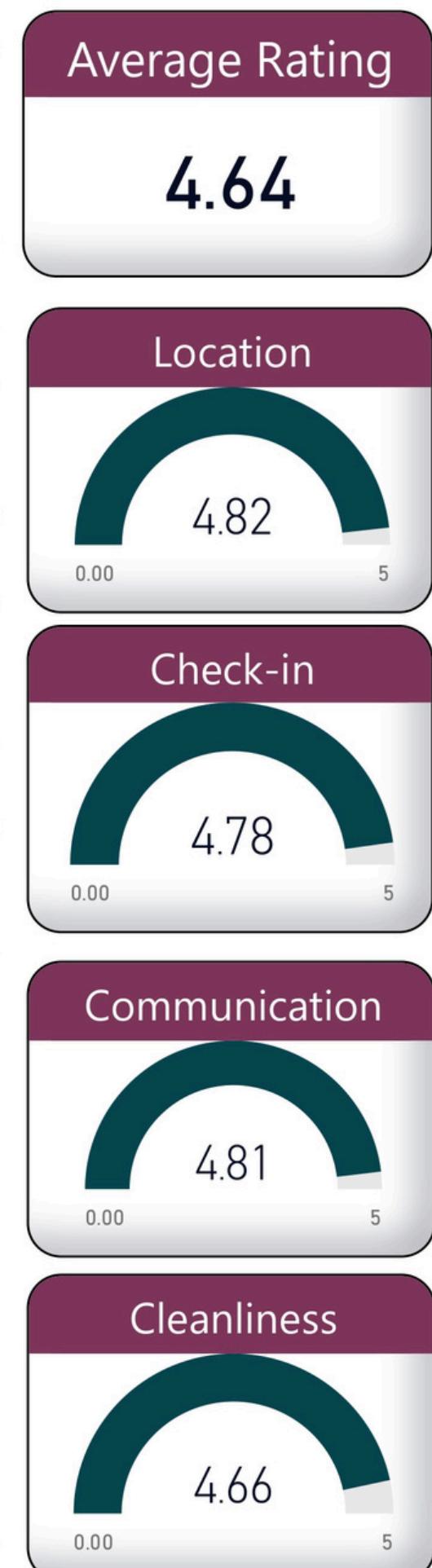
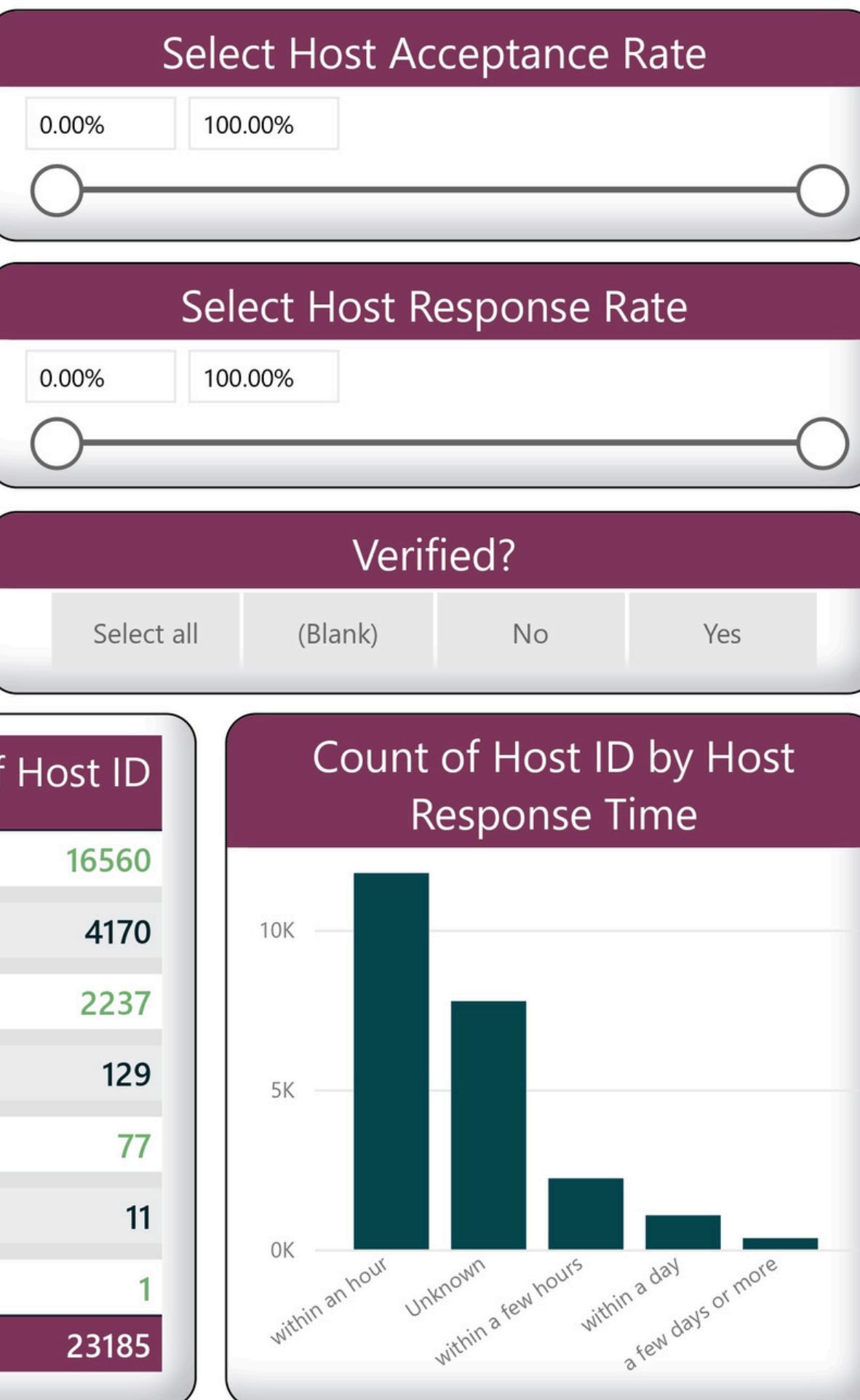
Neighbourhood
<input type="checkbox"/> Select all
<input type="checkbox"/> Banyule
<input type="checkbox"/> Bayside
<input type="checkbox"/> Boroondara
<input type="checkbox"/> Brimbank
<input type="checkbox"/> Cardinia
<input type="checkbox"/> Casey
<input type="checkbox"/> Darebin
<input type="checkbox"/> frankston
<input type="checkbox"/> Glen Eira
<input type="checkbox"/> Greater Dandenong
<input type="checkbox"/> Hobsons Bay
<input type="checkbox"/> Hume

Year
<input type="checkbox"/> Select all
<input type="checkbox"/> (Blank)
<input type="checkbox"/> 2009
<input type="checkbox"/> 2010
<input type="checkbox"/> 2011
<input type="checkbox"/> 2012
<input type="checkbox"/> 2013
<input type="checkbox"/> 2014
<input type="checkbox"/> 2015
<input type="checkbox"/> 2016
<input type="checkbox"/> 2017
<input type="checkbox"/> 2018
<input type="checkbox"/> 2019

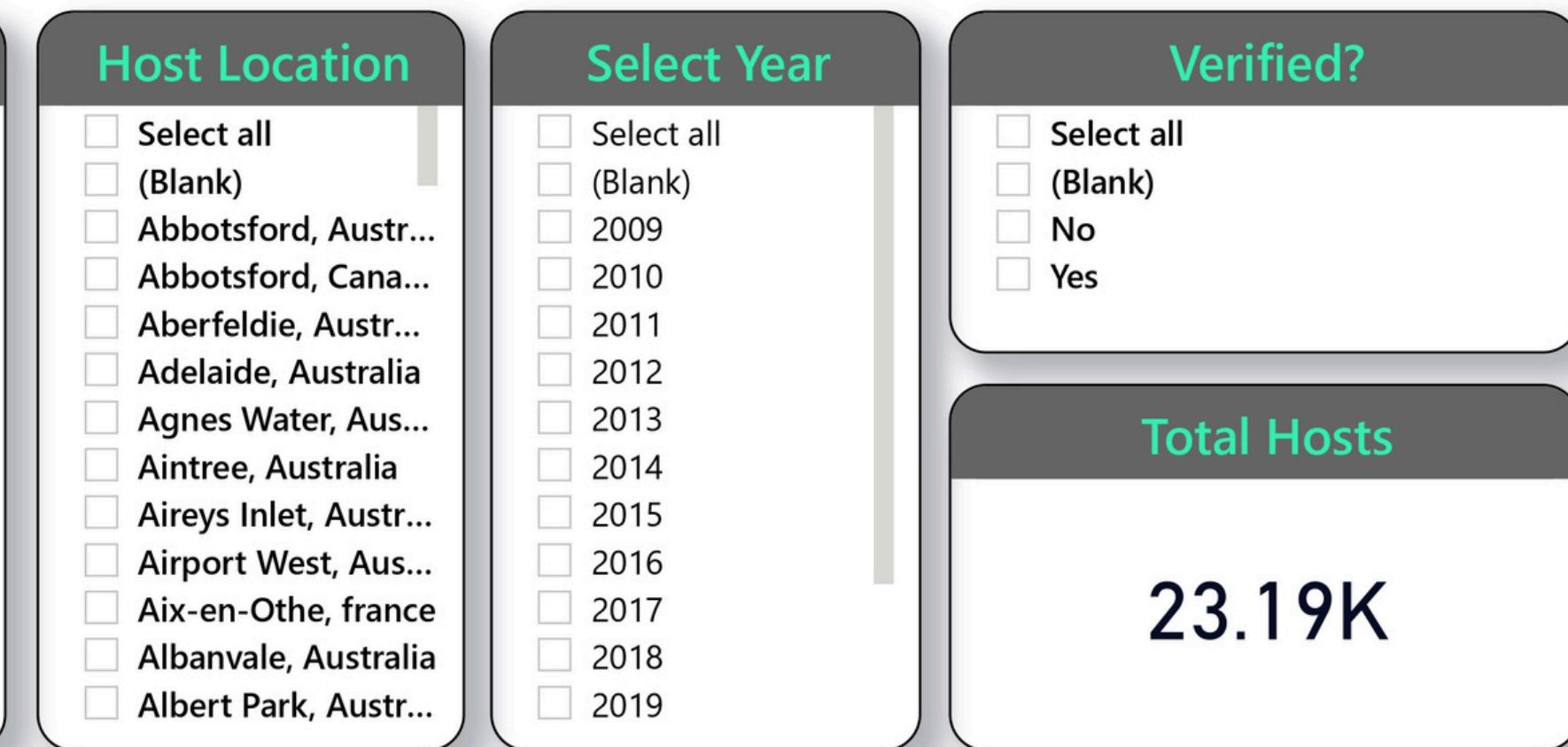
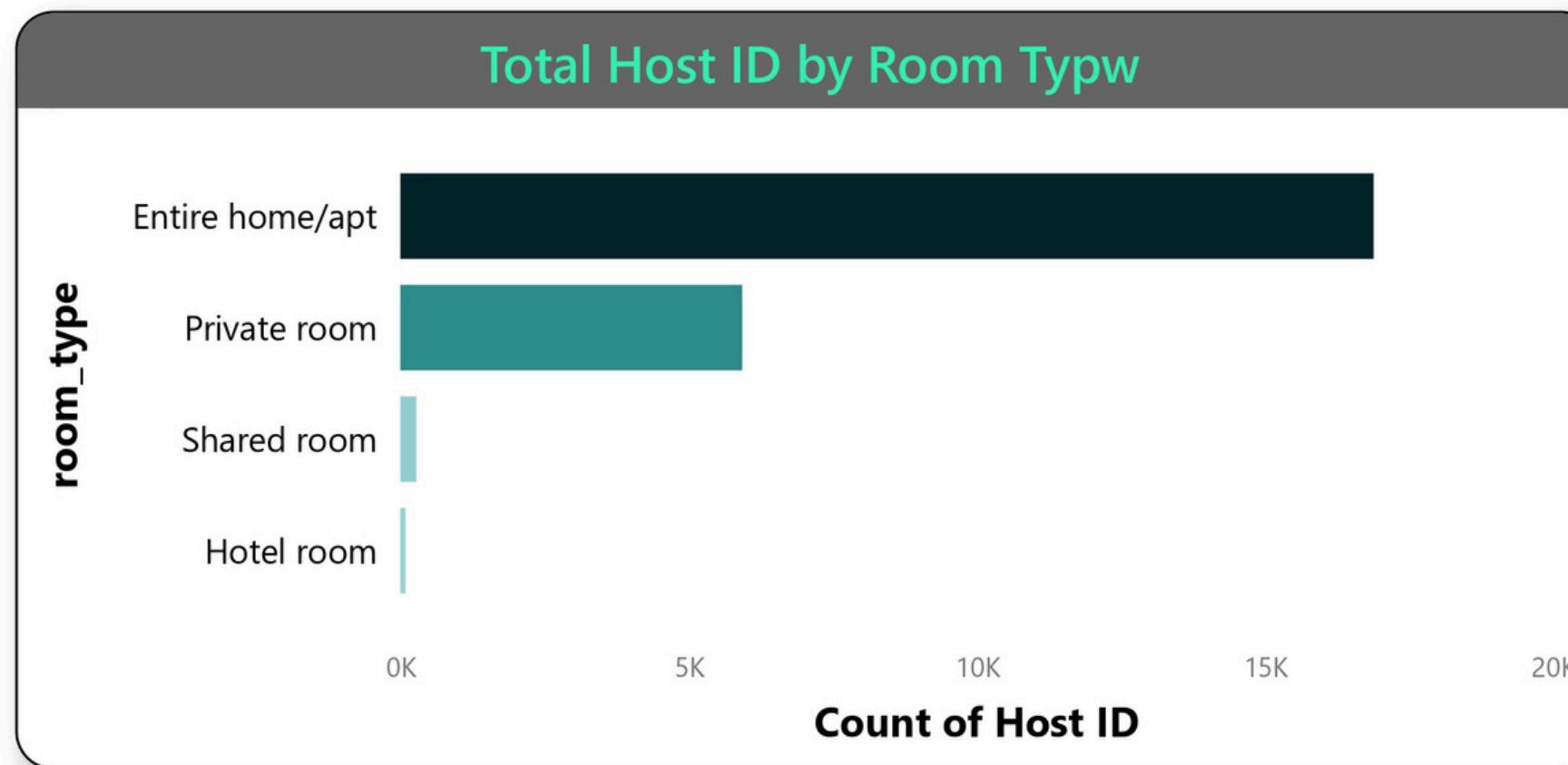
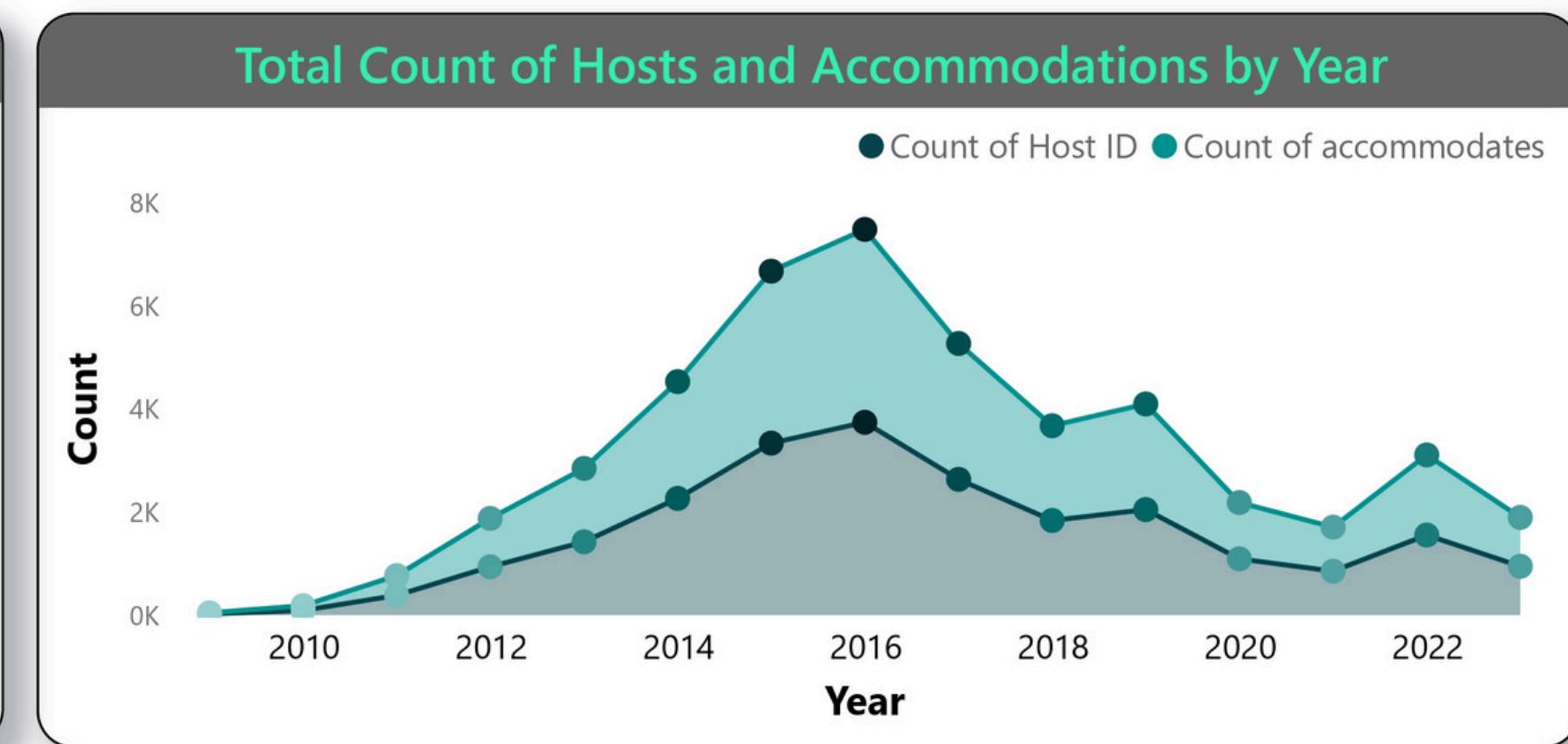
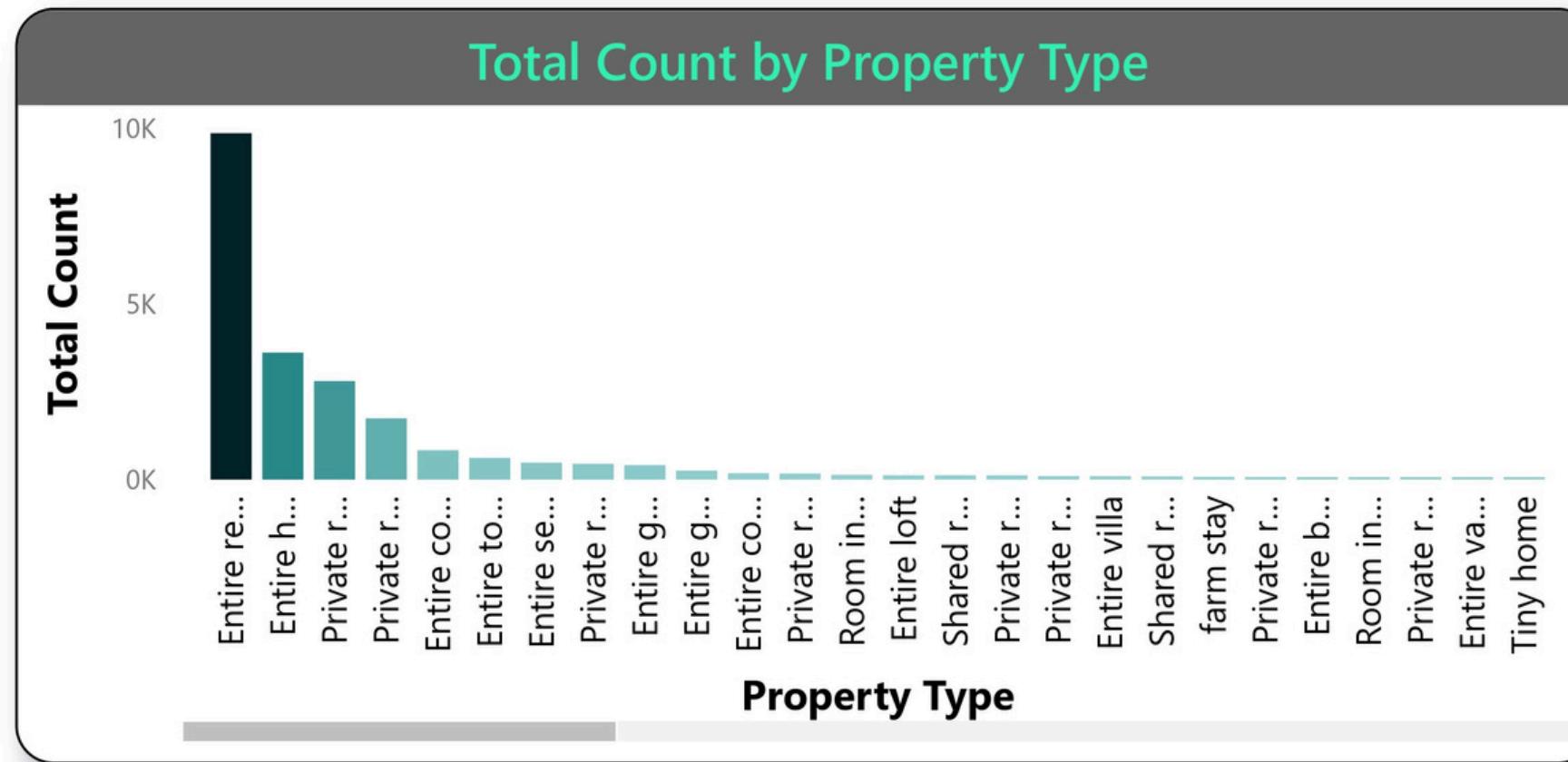


Verified With :

Method	Count of Host ID
Email, Phone	16560
Email, Phone, Work-Email	4170
Phone	2237
Phone, Work-Email	129
Email	77
Not Verified	11
Email, Work-Email	1
Total	23185



# Property and Room Type Analysis



# Findings

- **Geographical Insights:**

**Popular Locations:** Melbourne holds 31% of listings in Australia, followed by Port Phillip at 11%.

**Host Concentration:** The Newtown/Enmore neighborhood has a high concentration of hosts.

**Sydney Dominance:** Sydney has more hotel listings compared to other states in Australia.

- **Room Type Preferences:**

**Entire Room Preference:** The "Entire Room" type is overwhelmingly preferred, making it the most commonly chosen room type.

**Room Type Best Sellers:** The average price for an entire home/apartment is \$261.61.

- **Pricing and Availability Analysis:**

**Pricing Trends:** The highest prices are recorded for Casa Particular properties, with an average price of \$9652.50.

**Availability Patterns:** The highest availability was recorded in 2016.

# Findings

- **Host Performance :**

**Superhost Status:** The response rate does not significantly impact a host's Superhost status, with 72.6% indicating no influence on listing performance.

**Response Time and Verification:** These features positively affect performance, contributing to high ratings.

- **Review Scores and Guest Satisfaction :**

**Review Score Impact:** Higher response and acceptance rates lead to higher review scores, up to 5 stars.

**Areas for Improvement:** Cleanliness has the lowest rating at 4.66.

**Total Reviews and Listings:** There is a clear proportional relationship between the count of total reviews and total listing count.

- **Property Type and Room Analysis :**

**Property Distribution:** Entire rental units make up 42.5% of listings, followed by entire homes at 15%.

**Room Type Trends:** Entire home/apartment is the most popular, followed by private rooms.

**Popular Listings:** The most popular listings accommodate 2 guests.

# Recommendations



- Hosts should optimize their listings to cater to specific guest preferences in high-demand areas based on popular neighborhoods and property/room type trends.
- Develop pricing strategies that consider factors like property type, room capacity, and peak seasons to maximize competitiveness and profitability.
- The platform should provide resources or incentives to encourage faster response times and increased host verification methods, as these positively impact listing performance.
- Integrate guest feedback from reviews into actionable improvement suggestions for hosts, focusing on areas like cleanliness and amenities to enhance guest satisfaction.
- Ensure data completeness and integrity during scraping to maintain high-quality, reliable data for analysis and decision-making.

# Conclusion



This project's data analysis using Power BI has yielded valuable insights into the factors that influence hotel aggregator listing performance. By understanding geographical trends, pricing dynamics, host

characteristics, and guest satisfaction factors, we have equipped the platform with the knowledge needed to help hosts optimize their listings and attract more bookings. Utilizing these insights, the platform can enhance its overall listing quality and competitiveness in the market, ultimately leading to better performance and greater satisfaction for both hosts and guests.





# Thank You

PRESENTATION BY : BOLORAM MITRA



Github : [github.com/BoloramMitra](https://github.com/BoloramMitra)



LinkedIn : [linkedin.com/in/bolorammitra/](https://linkedin.com/in/bolorammitra/)



Email : Email : [boloram.mitra877@gmail.com](mailto:boloram.mitra877@gmail.com)