Airbnb Analysis Project Report

---

1. Introduction

Airbnb is one of the leading platforms in the hospitality sector, offering short-term stays to travelers around the world. Understanding the market requires analyzing pricing, customer preferences, location factors, and performance metrics.

This project uses real Airbnb listing data from multiple European cities and converts raw data into business insights using Excel, Python, MySQL, and Power BI. An interactive visualization dashboard was developed to support informed decision-making for hosts, investors, and stakeholders.

2. Project Objectives

- Analyze Airbnb performance based on revenue, ratings, and location
- Identify best-performing cities and room types
- Understand customer satisfaction and cleanliness levels
- Evaluate revenue impact based on distance from metro stations
- Build and showcase a professional interactive dashboard

3. Dataset Information

Attribute	Description
realSum	Revenue of listing
room_type	Type of stay (Entire home, Private room, Shared room)
guest_satisfaction_overall	Customer rating out of 100
cleanliness_rating	Cleanliness score
metro_dist	Distance from nearest metro station
host_is_superhost	If host is a Superhost (0/1)
lat, lng	Geographical coordinates
city	City where the listing is located

4. Tools and Technologies Used

Tool	Usage
MS Excel (Power Query)	Combined multiple files, cleaned columns, datatype corrections
Python (Jupyter Notebook)	Null value detection, data quality checks
MySQL	KPI calculations through SQL queries
Power BI	Visualization and dashboard analysis

5. Data Pre-Processing Workflow

Excel

- Merged multiple CSV datasets
- Removed unwanted columns
- Standardized data types

Python

- Checked and evaluated missing or abnormal values

MySQL

- Performed aggregation queries:

6. KPI Analysis Using SQL Queries

KPI	SQL Query Used	Result
Total Listings	SELECT COUNT(*) ...	25,500
Total Revenue	SELECT ROUND(SUM(realSum),2)...	7,029,888.56
Average Revenue per Listing	SELECT ROUND(AVG(realSum),2)	275.68
Avg. Guest Rating	AVG function	92.59
Avg. Cleanliness	AVG function	9.39
Superhost Percentage	SUM/COUNT calculation	25.85%
Top Revenue City	GROUP BY city	London
Best Room Type by Revenue	GROUP BY room_type	Entire Home/Apt

```
mysql> select COUNT(*) AS total_listings FROM air_bnb;
+----------------+
| total_listings |
+----------------+
|          25500 |
+----------------+
```

```
mysql> select ROUND(SUM(realSum),2) AS total_revenue FROM air_bnb;
+---------------+
| total_revenue |
+---------------+
|    7029888.56 |
+---------------+
```

```
mysql> SELECT ROUND(AVG(realSum),2) AS average_revenue FROM air_bnb;
+-----------------+
| average_revenue |
+-----------------+
|          275.68 |
+-----------------+
```

```
mysql> SELECT room_type, COUNT(*) AS total_rooms FROM air_bnb GROUP BY room_type ORDER BY total_rooms DESC;
+-----------------+-------------+
| room_type       | total_rooms |
+-----------------+-------------+
| Entire home/apt |       15994 |
| Private room    |        9326 |
| Shared room     |         180 |
+-----------------+-------------+
```

```
mysql> SELECT ROUND(AVG(guest_satisfaction_overall),2) AS avg_guest_rating FROM air_bnb;
+--------------------+
| avg_guest_rating |
+--------------------+
|              92.59 |
+--------------------+
```

```
mysql> SELECT (SUM(host_is_superhost) / COUNT(*)) * 100 AS superhost_percentage FROM air_bnb;
+-----------------------+
| superhost_percentage |
+-----------------------+
|               25.8510 |
+-----------------------+
```

```
mysql> SELECT room_type, ROUND(SUM(realSum),2) AS total_revenue FROM air_bnb GROUP BY room_type ORDER BY total_revenue DESC;
+-----------------+---------------+
| room_type       | total_revenue |
+-----------------+---------------+
| Entire home/apt |    5119413.38 |
| Private room    |    1885436.25 |
| Shared room     |      25038.93 |
+-----------------+---------------+
```

```
mysql> SELECT city,ROUND(SUM(realSum),2) AS revenue_by_city FROM air_bnb GROUP BY city ORDER BY revenue_by_city DESC;
+-----------+-----------------+
| city      | revenue_by_city |
+-----------+-----------------+
| london    |      1662102.82 |
| paris     |       1248202.3 |
| rome      |        905668.3 |
| lisbon    |       675238.98 |
| amsterdam |       601157.64 |
| barcelona |       448449.04 |
| vienna    |       417788.84 |
| athens    |        413515.1 |
| budapest  |       349322.51 |
| berlin    |       308443.02 |
+-----------+-----------------+
```

```
mysql> SELECT ROUND(AVG(cleanliness_rating),2) AS avg_cleanliness,ROUND(AVG(guest_satisfaction_overall),2) AS avg_satisfaction FROM air_bnb;
+----------------+-----------------+
| avg_cleanliness | avg_satisfaction |
+----------------+-----------------+
|           9.39 |           92.59 |
+----------------+-----------------+
```

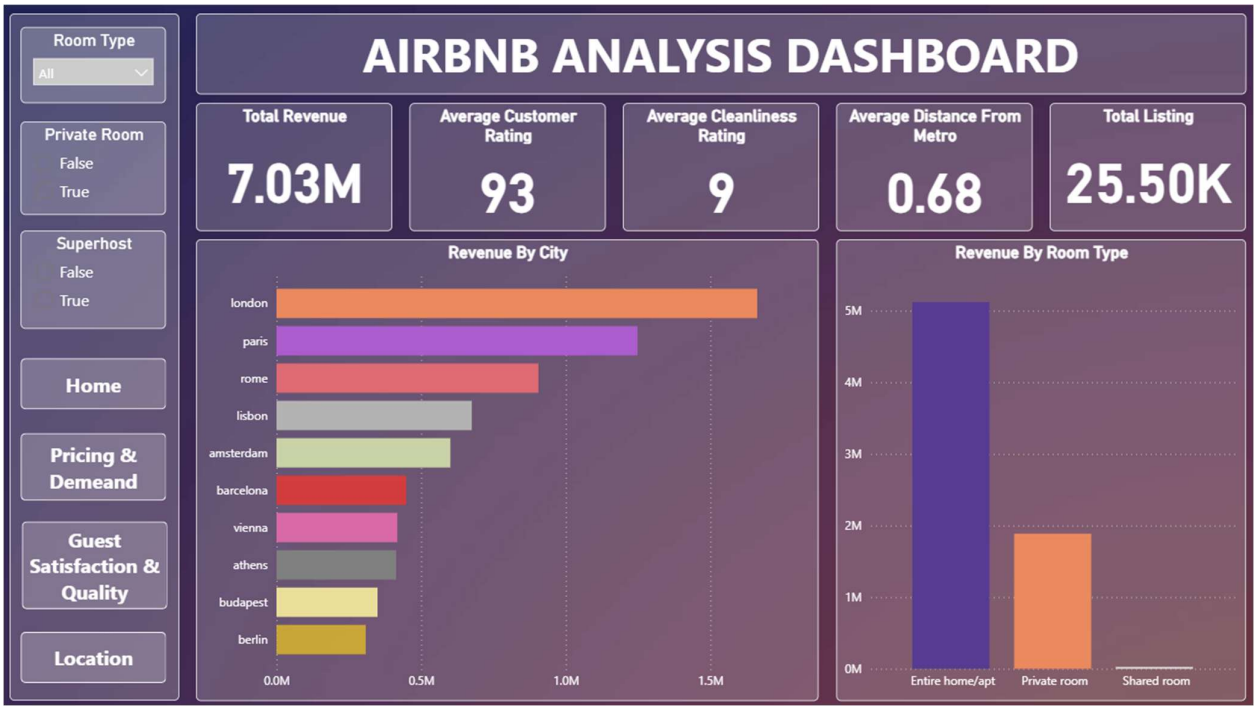
```
mysql> SELECT CASE WHEN metro_dist <= 1 THEN '0-1 km' WHEN metro_dist <= 3 THEN '1-3 km' WHEN metro_dist <= 5 THEN '3-5 km' ELSE '>5 km' END AS metro_range,
 ROUND(AVG(realSum),2) AS avg_price FROM air_bnb GROUP BY metro_range ORDER BY avg_price DESC;
+------------+-----------+
| metro_range | avg_price |
+------------+-----------+
| 0-1 km     |    281.56 |
| 1-3 km     |    255.55 |
| 3-5 km     |     214.2 |
| >5 km      |     177.9 |
+------------+-----------+
```

7. <u>Dashboard Development in Power BI</u>

The dashboard consists of **4 key report pages**:

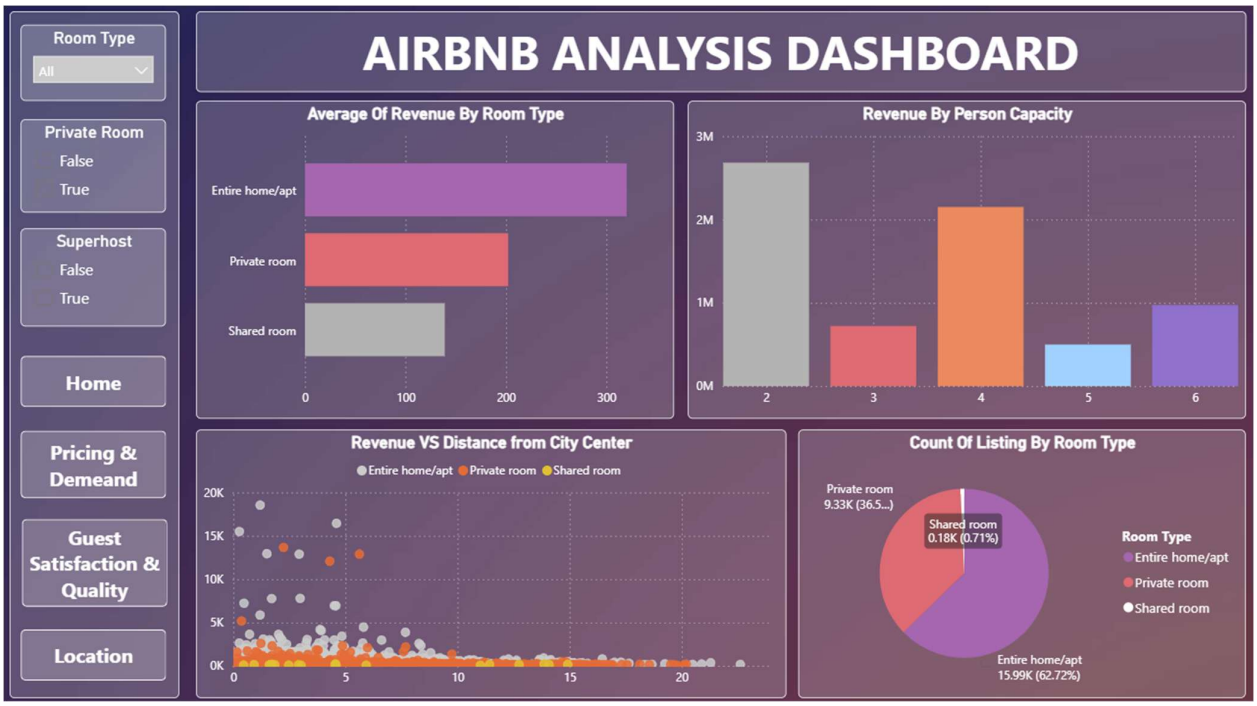
Page 1 — Home / Overall KPI Summary

- Total Revenue: 7.03M
- Avg. Customer Rating: 93
- Avg. Cleanliness Rating: 9
- Total Listings: 25.50K



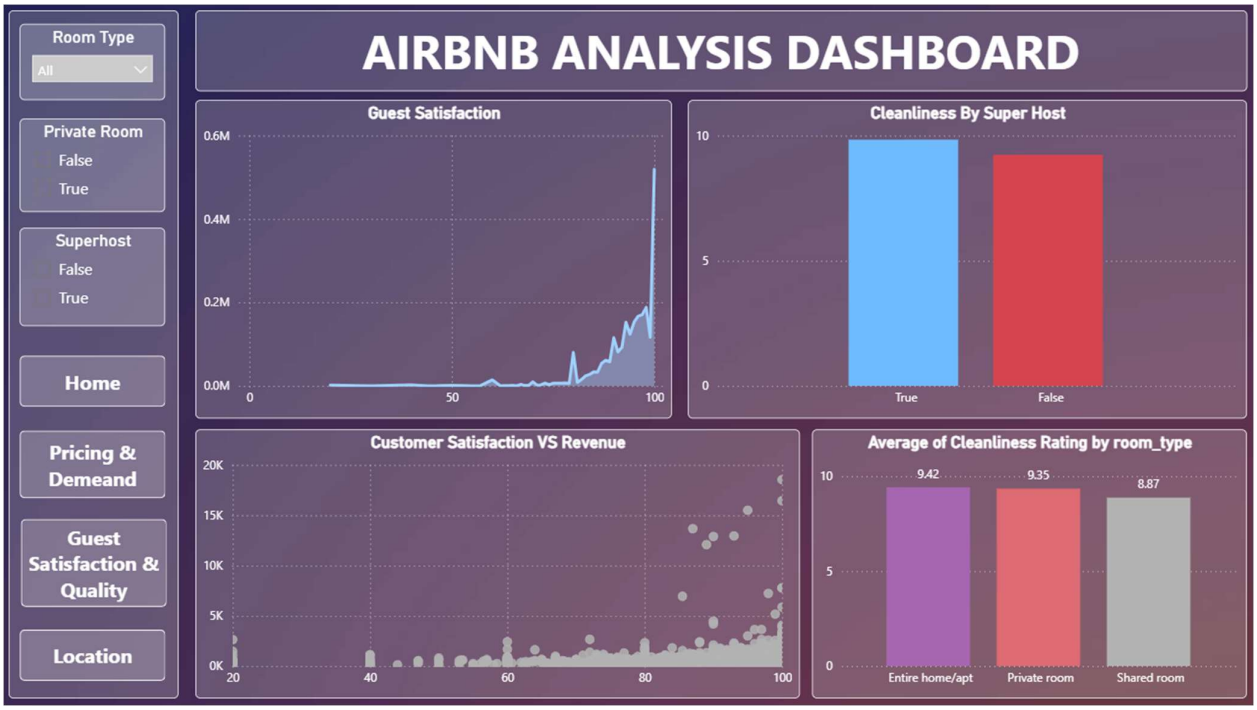
Page 2 — Pricing & Demand

- Revenue by Room Type
- Revenue by Person Capacity
- Revenue vs Distance Scatter Plot
- Count of Listings by Room Type



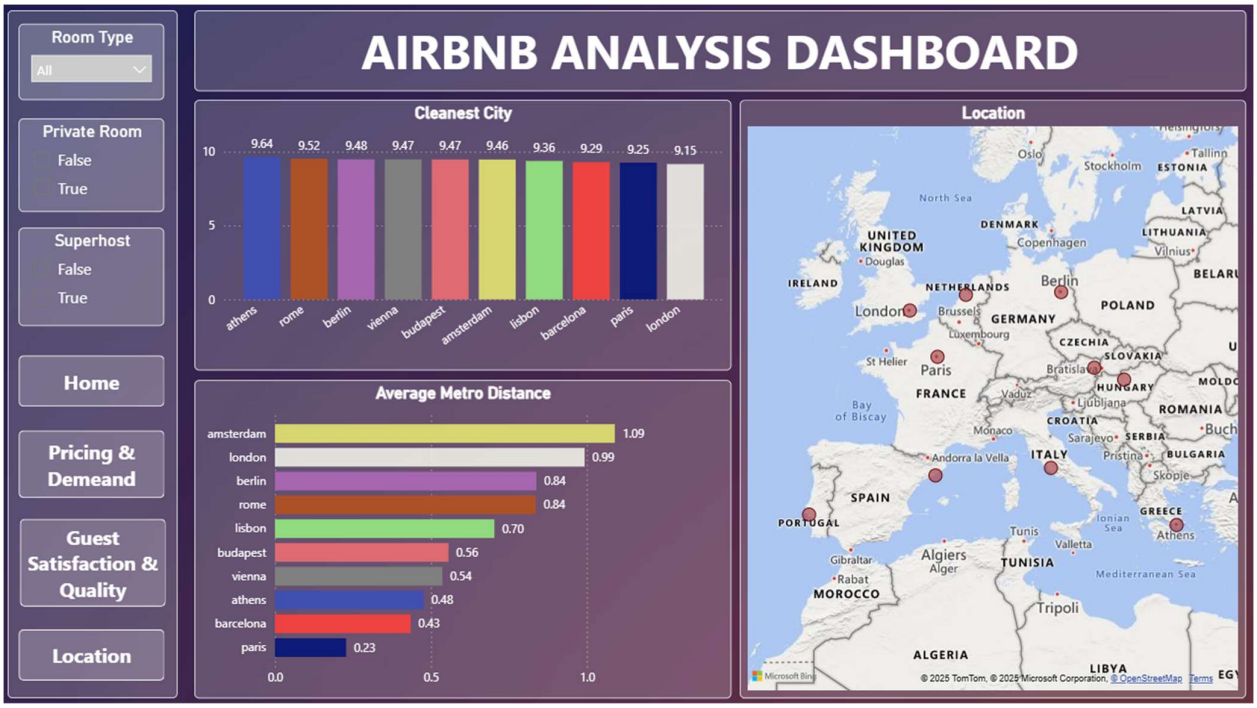
Page 3 — Guest Satisfaction & Quality

- Cleanliness by Superhost
- Guest Satisfaction Distribution
- Satisfaction vs Revenue Scatter
- Cleanliness by Room Type



Page 4 — Location Insights

- Cleanest City Ranking
- Average Metro Distance by City
- Map showing City-wise Listings



8. Insights & Findings

Revenue & Demand

- London generates the highest revenue (1.66M)
- Entire Homes contribute the largest revenue share (5.12M)
- Shared rooms are least profitable

Guest Satisfaction & Cleanliness

- Superhosts maintain slightly higher cleanliness
- Better cleanliness correlates with better satisfaction

Location Impact

- Listings within **0–1 km** distance from metro have **highest average pricing**
- Metro access is a crucial demand factor

9. Business Recommendations

- Encourage investment in **Entire home/apt** properties
- Improve metro accessibility for better pricing
- Increase host quality to achieve **Superhost** badge
- Focus expansion in **London, Paris, Rome**
- Maintain cleanliness to improve booking rates

10. Conclusion

This project successfully showcases the use of data analytics in the hospitality sector.
The analysis helped uncover revenue trends, satisfaction behaviour, and geographical factors influencing Airbnb success.

The interactive dashboard enables Airbnb hosts and investors to:

- Track performance
- Improve operational strategies
- Optimize location-based pricing

It demonstrates complete proficiency in:
Data Collection → Cleaning → SQL Analysis → Interactive Dashboard Storytelling

11. Future Scope

- Prediction of pricing using ML models
- Automated data refresh using Power BI Gateway
- Include seasonal & event-based demand patterns
- Add more listing attributes like amenities and reviews