Airbnb Market Intelligence: A Comprehensive Data-Driven Analysis

Introduction

This comprehensive analysis leverages a robust dataset spanning ten major European cities to decode the complex dynamics of the Airbnb marketplace. By employing advanced SQL analytics and statistical methods, we transform granular property data into strategic insights that illuminate market opportunities, consumer behavior patterns, and optimal pricing strategies. Our methodology combines sophisticated data engineering with practical business intelligence to deliver actionable recommendations for property managers, investors, and market strategists.

Strategic Objectives

1) Market Intelligence Integration

Develop a unified analytical framework by consolidating disparate city-level datasets

Create a standardized methodology for cross-market comparison and trend identification

Enable real-time market performance monitoring through automated data pipelines

2) Performance Analytics

Decode revenue patterns and price elasticity across different market segments

Map the correlation between guest satisfaction metrics and property characteristics

Identify key drivers of property success across diverse urban environments

3) Competitive Benchmarking

Analyze market positioning and pricing strategies across metropolitan areas

Evaluate the impact of property attributes on guest satisfaction and booking rates

Assess the influence of location-based factors on property performance

4) Strategic KPI Framework

Establish comprehensive performance metrics for market evaluation

Develop predictive models for revenue optimization and guest satisfaction

Create benchmarks for property performance across different market segments

Key Questions Addressed

1. Which city generates the highest revenue from Airbnb rentals?

2. Are there significant price differences between weekdays and weekends in different cities?

3. How does proximity to attractions affect room prices?

4. Which city offers the best value for money in terms of guest satisfaction and pricing?

5. What are the most popular room types and their distribution across cities?

6. Which cities have the highest capacity to accommodate large groups of travelers?

Data Preparation

Creating a Unified View for All City Data:

Combine Airbnb data from multiple cities and time periods into a single view for comprehensive analysis.

```
CREATE OR REPLACE VIEW all_city_data AS
SELECT
    'Amsterdam' AS city,
    'weekdays' AS day_type,
    s_no, realsum, room_type, room_shared, room_private, person_capacity,
    host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall,
    bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, rest_index_norm,
    lng, lat
FROM airbnb.amsterdam_weekdays

UNION ALL

SELECT
    'Amsterdam' AS city,
    'weekends' AS day_type,
```

```
    s_no, realsum, room_type, room_shared, room_private, person_capacity,
    host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall,
    bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, rest_index_norm,
    lng, lat
FROM airbnb.amsterdam_weekends

UNION ALL
```

This unified view consolidates data from all city tables, categorizing records by city and day_type (weekdays or weekends).

Analysis and Insights

Key Performance Indicators (KPI):

1. **Average Guest Satisfaction (KPI)**
2. **Total Revenue Generated by Each City (KPI)**
3. **Average Price Per Night (KPI)**
4. **Average Revenue Per Capacity (KPI)**

```
--Average Guest Satisfaction (KPI)
SELECT
    cast(ROUND(AVG(guest_satisfaction_overall), 2)as float) AS avg_guest_satisfaction
FROM all_city_data

--Total Revenue Generated by Each City (KPI)
SELECT
    cast(SUM(realsum)as float) AS total_revenue
FROM all_city_data

--Average Price Per Night (KPI)
SELECT
```

```
    cast(ROUND(AVG(realsum), 2)as float) AS avg_price_per_night
FROM all_city_data

--Average Revenue Per Capacity (KPI)
SELECT
   cast(ROUND(AVG(realsum / NULLIF(person_capacity, 0)), 2)as float) AS
avg_revenue_per_capacity
FROM all_city_data
```

Total Room Count by City:

Identify the total number of rooms available in each city.

```
--Total Room Count by City
SELECT city, cast(COUNT(*)as float) AS total_rooms
FROM all_city_data
GROUP BY city
ORDER BY total_rooms DESC;
```

Average Room Price by City:

Find the average price of rooms in each city.

```
--Average Room Price by City
SELECT city, cast(ROUND(AVG(realsum), 2)as float) AS avg_price
FROM all_city_data
GROUP BY city
ORDER BY avg_price DESC;
```

Total Person Capacity by City:

Determine the total accommodation capacity of each city.

```
--Total Person Capacity by City
SELECT city, cast(SUM(person_capacity)as float) AS total_capacity
FROM all_city_data
GROUP BY city
ORDER BY total_capacity DESC;
```

Rooms Near Attractions by City:

Count rooms located near attractions (high attraction index) for each city.

```
-- Rooms Near Attractions by City
SELECT city, cast(COUNT(*)as float) AS rooms_near_attractions
FROM all_city_data
WHERE attr_index > 80 -- Adjust the threshold as needed
GROUP BY city
ORDER BY rooms_near_attractions DESC;
```

Room Prices Based on Proximity to Attractions:

Analyze how room prices vary with distance from nearby attractions.

```
--Analyze Room Prices Based on Proximity to Attractions
SELECT
    CASE
        WHEN dist <= 5 THEN 'Within 1km'
        WHEN dist <= 10 THEN '1-5km'
        ELSE 'More than 5km'
    END AS distance_category,
    cast(ROUND(AVG(realsum), 2)as float) AS avg_price,
    cast(COUNT(*)as float) AS property_count
FROM all_city_data
```

```
GROUP BY distance_category
ORDER BY avg_price DESC;
```

Average Room Price per Bedroom by City:

Determine the average price per bedroom in each city to analyze cost-efficiency.

```
--Calculate the Average Room Price per Bedroom in Each City
SELECT
   city,
   cast(ROUND(SUM(realsum) / NULLIF(SUM(bedrooms), 0), 2)as float) AS
avg_price_per_bedroom
FROM all_city_data
GROUP BY city
ORDER BY avg_price_per_bedroom DESC;
```

Most Popular Room Types Across Cities:

Identify the top 5 room types preferred by guests across all cities.

```
--Find the Most Popular Room Types Across All Cities
SELECT
   room_type,
   cast(COUNT(*)as float) AS room_count
FROM all_city_data
GROUP BY room_type
ORDER BY room_count DESC
LIMIT 5;
```

Guest Satisfaction Rankings by City:

Rank cities based on average guest satisfaction to identify top-performing locations.

```
--Rank Cities by Guest Satisfaction Using a Dense Rank
SELECT
   city,
   cast(ROUND(AVG(guest_satisfaction_overall), 2)as float) AS avg_guest_satisfaction,
   cast(DENSE_RANK() OVER (ORDER BY AVG(guest_satisfaction_overall) DESC)as float) AS satisfaction_rank
FROM all_city_data
GROUP BY city
ORDER BY satisfaction_rank;
```

Compare Weekday and Weekend Prices by City:

Compare the average prices of rooms on weekdays vs weekends for each city.

```
--Compare Weekday and Weekend Prices by City
SELECT
   city,
   cast(ROUND(AVG(CASE WHEN day_type = 'weekdays' THEN realsum END), 2)as float) AS avg_weekday_price,
   cast(ROUND(AVG(CASE WHEN day_type = 'weekends' THEN realsum END), 2)as float) AS avg_weekend_price
FROM all_city_data
GROUP BY city
ORDER BY avg_weekday_price DESC;
```

Find the Most Popular Room Types Across All Cities:

Identify the top 5 most common room types across all cities.

```
--Calculate the Distribution of Room Count Across Room Types
SELECT
   room_type,
```

```
    cast(COUNT(*)as float) AS room_count
FROM all_city_data
GROUP BY room_type
ORDER BY room_count DESC;
```

Regression Analysis of Price Differences: Weekends vs Weekdays:

Predict the impact of weekends on rental prices across cities using regression analysis.

```
--Predict Weekend vs Weekday Price Differences Using Regression
SELECT
    city,
    REGR_SLOPE(realsum, CASE WHEN day_type = 'weekends' THEN 1 ELSE 0 END) AS
weekend_price_increase,
    REGR_INTERCEPT(realsum, CASE WHEN day_type = 'weekends' THEN 1 ELSE 0 END) AS
base_price,
    cast(COUNT(*)as float) AS data_points
FROM all_city_data
GROUP BY city
HAVING COUNT(*) > 100 -- Ensure sufficient data points for reliable regression
ORDER BY weekend_price_increase DESC;
```

Room Type Distribution Analysis by City:

Determine the percentage of shared and private rooms in each city to understand room type preferences.

```
--Analyze the Percentage of Shared vs Private Rooms by City
SELECT
    city,
    cast(ROUND(100.0 * SUM(CASE WHEN room_shared = 't' THEN 1 ELSE 0 END) / COUNT(*),
2)as float) AS percent_shared_rooms,
```

```
    cast(ROUND(100.0 * SUM(CASE WHEN room_private = 't' THEN 1 ELSE 0 END) / COUNT(*), 2)as float) AS percent_private_rooms
FROM all_city_data
GROUP BY city
ORDER BY percent_shared_rooms DESC;
```

Find Cities with the Highest Capacity for Large Groups:

Identify cities with the highest number of properties for large groups and their percentage of total listings.

```
--Find Cities with the Highest Capacity for Large Groups
SELECT
    city,
    cast(COUNT(*)as float) AS large_group_properties,
    cast(ROUND(100.0 * COUNT(*) / SUM(COUNT(*)) OVER (), 2)as float) AS percent_of_total_properties
FROM all_city_data
WHERE person_capacity > 5
GROUP BY city
ORDER BY large_group_properties DESC;
```

Find the Cities with the Best Ratio of Price to Guest Satisfaction:

Identify cities with the best balance between price and guest satisfaction.

```
-- Find the Cities with the Best Ratio of Price to Guest Satisfaction
SELECT
    city,
    cast(ROUND(AVG(realsum) / NULLIF(AVG(guest_satisfaction_overall), 0), 2)as float) AS price_to_satisfaction_ratio,
    cast(ROUND(AVG(realsum), 2)as float) AS avg_price,
    cast(ROUND(AVG(guest_satisfaction_overall), 2)as float) AS avg_satisfaction
```

```
FROM all_city_data

GROUP BY city

ORDER BY price_to_satisfaction_ratio ASC

LIMIT 5;
```

Insights

1. The $13.6M revenue, $96 revenue per capacity, and $283 average price per night highlight strong market performance, supported by an excellent 93% guest satisfaction.

2. Premium cities like Amsterdam, Paris, and London have the highest prices, driven by proximity to attractions (within 1km boosts prices to $289.43 vs. $189.6 for 5km+).

3. London (9,942 rooms) and Rome (9,027 rooms) lead in availability near attractions, while Amsterdam has the highest weekend price increase (59.8%). Athens (95%) and Budapest (94.6%) top guest satisfaction, suggesting opportunities for hosts to optimize pricing and service.

4. Athens offers the best price-to-satisfaction ratio (1.6), with an average price of $151.74. Weekend prices ($2,687.65) are slightly higher than weekday prices ($2,594.91). Room prices are lowest in Athens ($119.35) and highest in Amsterdam ($443.48).

5. Private rooms dominate (353.45%), while shared rooms are minimal (7.71%). Rome has the highest capacity (30,307) and large group properties (21.05%), with Amsterdam having the lowest (5,786) and just 0.99% for large groups.

Conclusion

This project demonstrates how SQL-based data consolidation and analysis can transform raw Airbnb data into actionable insights. Key findings include:

1. Revenue and satisfaction leaders among cities.

2. Trends in pricing relative to weekdays, weekends, and proximity to attractions.

3. Popular room types and their distribution.

By s these insights, Airbnb hosts and managers can refine pricing strategies, enhance guest satisfaction, and optimize property utilization for long-term profitability.

```
--Identify the Most Profitable Cities for Each Room Type

/*WITH city_revenue AS (

   SELECT

      city,
```

```
        room_type,
        cast(SUM(realsum)as float) AS total_revenue
    FROM all_city_data
    GROUP BY city, room_type
),
ranked_revenue AS (
    SELECT
        city,
        room_type,
        total_revenue,
        RANK() OVER (PARTITION BY room_type ORDER BY total_revenue DESC) AS rank
    FROM city_revenue
)
SELECT
    room_type,
    city,
    total_revenue
FROM ranked_revenue
WHERE rank = 1
ORDER BY total_revenue DESC;*/

--Distribution of Room Types Across Cities
/*SELECT city, room_type, cast(COUNT(*)as float) AS room_count
FROM all_city_data
GROUP BY city, room_type
ORDER BY city, room_type;*/
```