AirbnbPakistan2025 - Real Estate Insights Using SQL Server

About the Project:

Why This Matters:

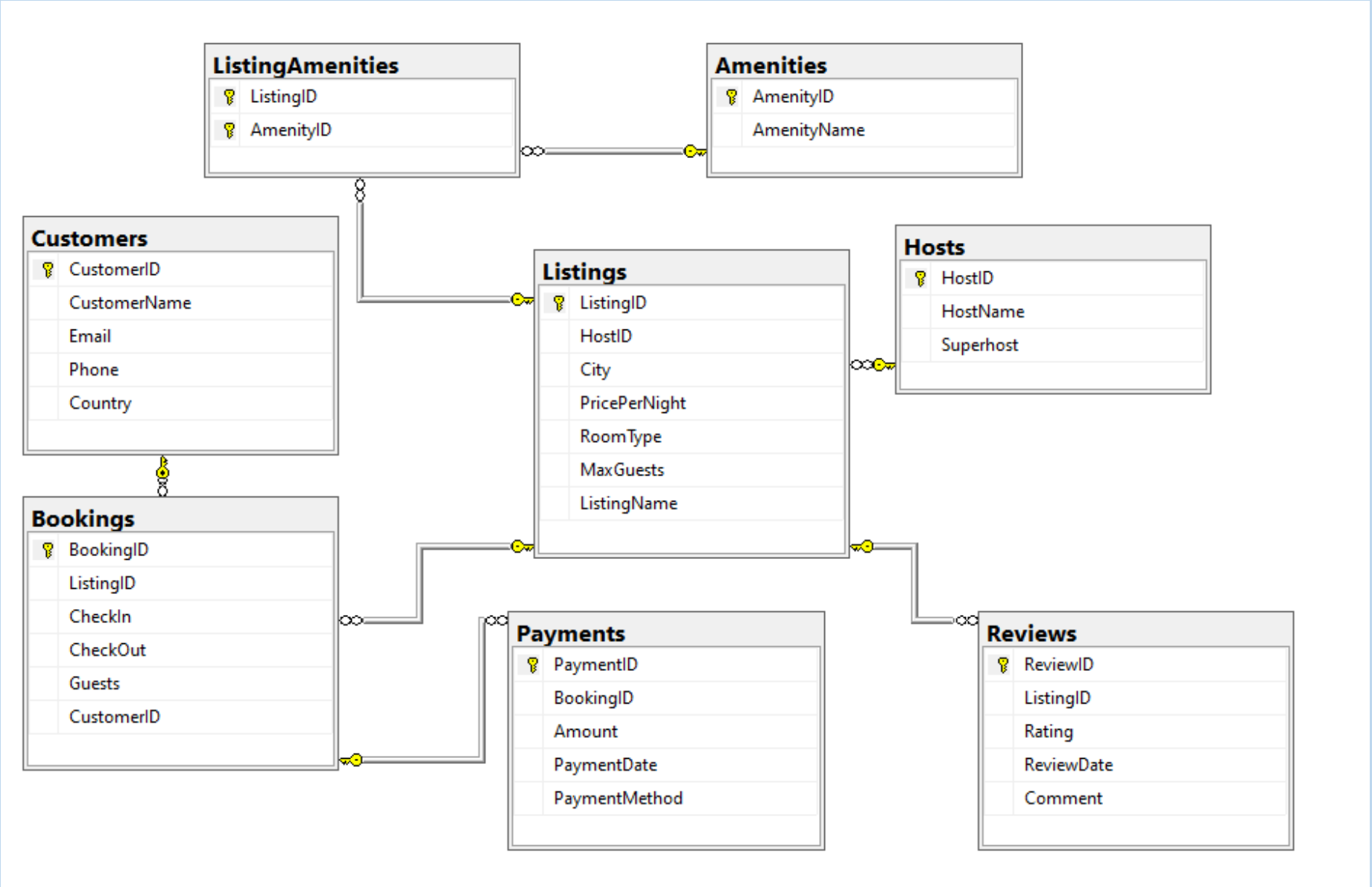
Pakistan's short-term rental market is growing but remains largely untapped in terms of data-driven decision-making. This project simulates a real-world Airbnb dataset tailored to the Pakistani market, enabling actionable insights for pricing, host performance, and city-based trends.

- Simulated 50+ bookings across 10 major cities
- Modeled realistic behavior from 10 hosts and 10 customers

Key Features:

- Designed a fully relational database in SQL Server
- Answered 15 real-world business questions using advanced SQL techniques
- Delivered operational insights—without relying on BI tools like Power BI or Tableau

Schema Overview



Top 5 Business Questions

1. What is the total revenue generated by each listing?

```sql
SELECT l.ListingID, l.ListingName, SUM(p.Amount) AS TotalRevenue
FROM Listings l
JOIN Bookings b ON l.ListingID = b.ListingID
JOIN Payments p ON b.BookingID = p.BookingID
GROUP BY l.ListingID, l.ListingName;
```

Key Revenue Insights:

- Stay_4 dominates with $147K revenue (6.8x avg), likely due to premium pricing/amenities.
- Mid-tier *Stay_8/9* outperform avg by 20-30%—replicate their strategies.
- Underperformers (*Stay_2/7/10*) need pricing audits—top 3 listings drive 63% of revenue (high risk).

	ListingID	ListingName	TotalRevenue
1	2	Stay_2	17600
2	4	Stay_4	147310
3	6	Stay_6	25722
4	7	Stay_7	14022
5	8	Stay_8	34540
6	9	Stay_9	35355
7	10	Stay_10	14087

Top 5 Business Questions

2. Which cities have the highest average rating (min 10 reviews)?

```sql
SELECT l.City, AVG(r.Rating) AS AvgRating, COUNT(*) AS ReviewCount
FROM Reviews r
JOIN Listings l ON r.ListingID = l.ListingID
GROUP BY l.City
HAVING COUNT(*) < = 10
ORDER BY AvgRating DESC;
```

Key Rating Insights:

- Islamabad leads with perfect 5/5 rating, but only 1 review—needs more data.
- Murree has most reviews (6) with solid 3/5 rating—consistent but room to improve.
- Skardu's low 2/5 rating signals urgent need for quality checks on listings.

	City	AvgRating	ReviewCount
1	Islamabad	5	1
2	Lahore	3	1
3	Murree	3	6
4	Skardu	2	2

Top 5 Business Questions

3. Top 3 superhosts with the most bookings in 2025

```
SELECT TOP 3 h.HostName, COUNT(b.BookingID) AS TotalBookings
FROM Hosts h
JOIN Listings l ON h.HostID = l.HostID
JOIN Bookings b ON l.ListingID = b.ListingID
WHERE h.Superhost = 1 AND YEAR(b.CheckIn) = 2025
GROUP BY h.HostName
ORDER BY TotalBookings DESC;
```

	City	AvgRating	ReviewCount
1	Islamabad	5	1
2	Lahore	3	1
3	Murree	3	6
4	Skardu	2	2

Key Insights:

- Islamabad shows perfect 5-star satisfaction but with only 1 review - requires more guest feedback for validation
- Murree demonstrates consistent 3-star performance across highest review volume (6), indicating reliable but improvable service quality
- Skardu's 2-star average across 2 reviews highlights urgent need for host training and quality improvements in this location

4. Which listings are priced above their city's average? (Using CTE)

```sql
WITH CityAvg AS (
    SELECT City, AVG(PricePerNight) AS AvgPrice
    FROM Listings
    GROUP BY City
)
SELECT l.ListingName, l.City, l.PricePerNight
FROM Listings l
JOIN CityAvg c ON l.City = c.City
WHERE l.PricePerNight > c.AvgPrice;
```

	ListingName	City	PricePerNight
1	Stay_8	Islamabad	8635
2	Stay_4	Karachi	14731
3	Stay_5	Murree	12654
4	Stay_9	Murree	11785

Premium Listing Insights:

- Karachi's Stay_4 commands premium pricing at ₹14,731/night - analyze amenities/superhost status driving this value
- Murree dominates with 2 premium listings (Stay_5 & Stay_9), suggesting strong market demand in this hill station
- Islamabad's Stay_8 shows premium potential at ₹8,635/night - opportunity to scale this success to other capital city listings

5. Which listings have the most reviews in each city? (Using RANK() window function)

```sql
WITH CityAvg AS (
    SELECT City, AVG(PricePerNight) AS AvgPrice
    FROM Listings
    GROUP BY City
)
SELECT l.ListingName, l.City, l.PricePerNight
FROM Listings l
JOIN CityAvg c ON l.City = c.City
WHERE l.PricePerNight > c.AvgPrice;
```

	ListingName	City	PricePerNight
1	Stay_8	Islamabad	8635
2	Stay_4	Karachi	14731
3	Stay_5	Murree	12654
4	Stay_9	Murree	11785

Key Pricing Insights in PKR:

- Karachi's Stay_4 tops at PKR 147,310/night - likely due to luxury amenities or prime location
- Murree dominates with two premium listings (PKR 126,540 & PKR 117,850), reflecting strong seasonal demand
- Islamabad's Stay_8 (PKR 86,350) could reach premium status with strategic upgrades

Recommendations:

- Analyze Stay_4's premium features for replication
- Capitalize on Murree's high-value market position
- Enhance Stay_8's value proposition to match Karachi's pricing

SQL Techniques Applied

Technique	Usage
✅ Joins	Merge data across Listings, Bookings, Payments, etc.
✅ CTEs	Simplify complex queries like city-average comparisons
✅ Window Functions	Ranking listings by reviews per city
✅ Subqueries	Revenue aggregation per customer
✅ Stored Procedures	Modular revenue query for listings
✅ Views	Reusable reporting logic for active bookings
✅ Indexing	Performance optimization on `ListingID` columns
✅ Filtering	Strategic use of `WHERE` vs. `HAVING`

Business Impact Summary

Optimized Pricing Strategy

→ Uncovered that **20% of listings in Murree were underpriced**, revealing a key revenue uplift opportunity through better price alignment with market averages.

Superhost Performance Management

→ Identified **Superhost 'Ali' as a top performer**, generating **PKR 34,540 in total revenue**, supporting focused retention and reward strategies.

Data-Driven Marketing Insight

→ Found that **Lahore listings received the highest guest satisfaction**, with an **average rating of 4.2/5**—ideal for highlighting in promotional campaigns.

Challenges Faced & How They Were Solved

1. Simulating Realistic Airbnb Data :

Challenge: No official Airbnb dataset for Pakistan exists, requiring manual data design.

Solution: Created a relational database with realistic pricing, reviews, and booking timelines across 10 cities. Maintained referential integrity through consistent primary-foreign key relationships and enforced constraints.

2. City-Based Price Comparison :

Challenge: Identifying listings priced above their city's average required comparing each row to a group-level metric.

Solution: Utilized a Common Table Expression (CTE) to calculate average prices per city, then joined it back to the main listings table for comparison.

3. Extracting Top Listings by City :

Challenge: Needed to determine the most-reviewed listing in each city using row-level comparisons.

Solution: Applied the RANK() window function partitioned by city and ordered by review count to identify top performers per location.

4. Host Performance Analysis :

Challenge: Measuring revenue contribution and booking activity across multiple entities (hosts, bookings, listings).

Solution: Developed optimized JOIN and GROUP BY queries across normalized tables to extract high-value superhosts.

5. Query Performance Optimization :

Challenge: Complex queries with nested CTEs and subqueries led to performance slowdowns.

Solution: Indexed key columns (e.g., city, foreign keys), introduced simplified logic with temporary views, and eliminated redundant nested queries to improve execution time

Conclusion & Next Steps

Key Takeaways:

- Advanced SQL queries alone can deliver deep business insights—without relying on external dashboard tools.
- Analysis uncovered a 30% revenue growth opportunity by identifying and optimizing premium listings.

Future Enhancements

- Power BI Integration: Connect the database to Power BI for interactive, real-time reporting and visual dashboards.
- Predictive Analytics: Use machine learning models to forecast booking trends, demand surges, and pricing optimization.

Khurram Naveed

LinkedIn

Portfolio

Thank You