

Airbnb Market Analysis

Introduction:-

The "Airbnb Market Analysis & Real Estate Sales Data (2019)" dataset provides a comprehensive view of the rental market and real estate trends within two popular areas in California: Big Bear and Joshua Tree. This dataset includes valuable insights into the performance of Airbnb listings and property sales over a 12-month period in 2019. The data is aggregated on a monthly basis, allowing for in-depth analysis of occupancy, pricing, and revenue metrics for Airbnb listings, as well as property sales trends in the specified regions. By incorporating information on property amenities, sales data, and geolocation, this dataset serves as a powerful tool for understanding the dynamics of both short-term rental and traditional real estate markets.

Datasets:- I am using the "Zillow Market Analysis and Real Estate Sales Data" dataset from Kaggle for my Power BI project. This dataset provides detailed information on real estate prices, sales trends, and market conditions across various U.S. regions, enabling comprehensive data visualization and insights into housing market dynamics.

Below is the link to the dataset:

[https://www.kaggle.com/datasets/computingvictor/zillow-market-analysis-and-real-estate-sales-d
ata](https://www.kaggle.com/datasets/computingvictor/zillow-market-analysis-and-real-estate-sales-data)

Dataset Information:-

This dataset consists of several key files, each providing a unique aspect of the Airbnb market and real estate trends in California:

1. Market Analysis:

- This file contains listing-level information for Airbnb properties, including metrics such as revenue, occupancy ratios, nightly rates, availability, lead times, and average length of stay. The data is aggregated monthly and offers valuable insights into the performance of short-term rentals.

2. Amenities:

- A file indicating whether a listing includes specific amenities (e.g., hot tub or pool), with values of 1 for presence and 0 for absence. This information helps

assess how amenities influence listing performance and pricing.

3. **Geolocation:**

- This file contains the latitude and longitude coordinates for each Airbnb listing, enabling precise spatial analysis and visualization of market trends by location.

4. **Sales Properties:**

- This dataset covers properties available for sale within the study areas. For the Joshua Tree region (zip codes 92284 and 92252), there are two files—one with general sales property data and another focusing specifically on properties with pools.

Potential Applications:-

This dataset is well-suited for analysts and researchers interested in exploring the intersection of Airbnb rentals and real estate sales in California. The rich data can be utilized for Power BI reporting to create dynamic dashboards that uncover key trends, such as:

- Performance optimization for Airbnb listings (e.g., pricing strategies based on occupancy rates, amenities, and location).
- Real estate investment strategies based on property sales trends and the availability of specific amenities.
- Geographic analysis of market hotspots using geolocation data.

With its broad coverage and diverse insights, this dataset serves as a valuable resource for data-driven decision-making in both the short-term rental and real estate sectors.

Steps Performed on Datasets:-

1) Amenities: "Added columns for Year, Quarter, Start of Quarter, Day, and Day Name to the Amenities table using Power BI's Transform Data functionality."

Explanation:

Using the Transform Data option in Power BI (Power Query Editor), the following columns were added to enrich the *Amenities* dataset:

- **Year:** Extracted the year from the date column to enable time-based analysis.
- **Quarter:** Identified the calendar quarter for each date to support quarterly reporting.
- **Start of Quarter:** Calculated the first day of the quarter for time intelligence measures.

- **Day:** Retrieved the day component of each date for granular insights.
- **Day Name:** Added the name of the weekday (e.g., Monday, Tuesday) for trend analysis by day.

These transformations help enhance the dataset for better filtering, grouping, and time-series analysis in Power BI reports.

Below is screenshot attached for reference:-

The screenshot displays the Power BI Desktop interface for a project named 'Airbnb Project'. The main view shows a data table with the following columns: Year, Quarter, Start of Quarter, Day, and Day Name. The data is filtered for the year 2022, quarter 4, and the start of the quarter on 01/10/2022. The 'Day' column shows the value 1, and the 'Day Name' column shows 'Thursday'. The table contains 23 rows of data.

The left pane shows the 'Queries' list with the following queries: market_analysis_2019, sales_properties_total_zipcode_92252, sales_properties_total_zipcode_92314, sales_properties_total_zipcode_92315, sales_properties_total_zipcode_92284, sales_properties_with_pool_zipcode_92252, sales_properties_with_pool_zipcode_92284, amenities, geolocation, and market_analysis. The 'amenities' query is selected.

The right pane shows the 'Query Settings' for the 'amenities' query. The 'APPLIED STEPS' list includes: Source, Promoted Headers, Changed Type, Filtered Rows, Inserted Year, Filtered Rows1, Inserted Quarter, Inserted Start of Quarter, Inserted Day, and Inserted Day Name. The 'Inserted Day Name' step is highlighted.

2) Geolocation: "Removed null and blank values from the Street Name column using Power BI's Transform Data feature."

Explanation:

Using the Transform Data option in Power BI (Power Query Editor), the *Street Name* column was cleaned to ensure data quality and consistency. Specifically:

- Null values (missing entries) were filtered out to eliminate incomplete data.
- Empty strings (rows with no visible value) were also removed.
- These steps help maintain accurate filtering and analysis, particularly when grouping or mapping data by street name.

The cleaning was done through the *Remove Rows* and *Text Filters* options in Power Query, ensuring only valid, meaningful street names remain in the dataset.

The attached screenshot serves as a visual reference:

The screenshot displays the Power BI Power Query Editor interface. The ribbon at the top includes options for File, Home, Transform, Add Column, View, Tools, and Help. The 'Transform' tab is active, showing options like Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, Azure Machine Learning, and AI Insights. The 'Query' section shows the formula bar with the M code: `= Table.SelectRows("#Changed Type", each ([street_name] <> null and [street_name] <> ""))`. The 'Properties' pane on the right shows the 'Name' property set to 'geolocation' and the 'Applied Steps' list including 'Removed Empty Rows'. The main area displays a table of data with columns: unified_id, month, street_name, latitude, and longitude. The 'geolocation' query is selected, showing 239 distinct values for unified_id, 5 distinct values for month, 160 distinct values for street_name, 234 distinct values for latitude, and 238 distinct values for longitude. The 'Properties' pane on the right shows the 'Name' property set to 'geolocation' and the 'Applied Steps' list including 'Removed Empty Rows'.

unified_id	month	street_name	latitude	longitude
1	01/12/2022	Heavenly Valley Road	3423948	-11687991
2	01/12/2022	Round Drive	3425095	-11694139
3	01/12/2022	Starvation Flats Road	3425341	-11688025
4	01/12/2022	Spruce Lane	3423844	-11682383
5	01/12/2022	La Crescenta Drive	342355	-1168483
6	01/12/2022	Waynoka Lane	3424263	-11692486
7	01/12/2022	Carmelita Way	3408019	-11637654
8	01/12/2022	Starvation Flats Road	3425264	-1168808
9	01/12/2022	Fenway Drive	342422	-1168618
10	01/12/2022	Mullins Drive	3427166	-11685132
11	01/12/2022	Summit Boulevard	3423979	-11688697
12	01/12/2022	Starvation Flats Road	342521	-11687999
13	01/12/2022	Eureka Drive	3424654	-11689829
14	01/12/2022	Mountain View Circle	3412838	-11631354
15	01/12/2022	Silver Tip Drive	34237736	-116862297
16	01/12/2022	Navajo Trail	3411093	-1163238
17	01/12/2022	Saddleback Road	34121407	-116328949
18	01/12/2022	Yosemite Drive	3422938	-11684423
19	01/12/2022	Deer Run Court	3423125	-11686125
20	01/12/2022	East Aeroplane Boulevard	3425887	-11684183
21	01/12/2022	Siskiyou Drive	3422742	-11685785
22	01/12/2022	Jeffries Road	3423966	-1169024
23	01/12/2022	Sunburst Drive	3412841	-11631947

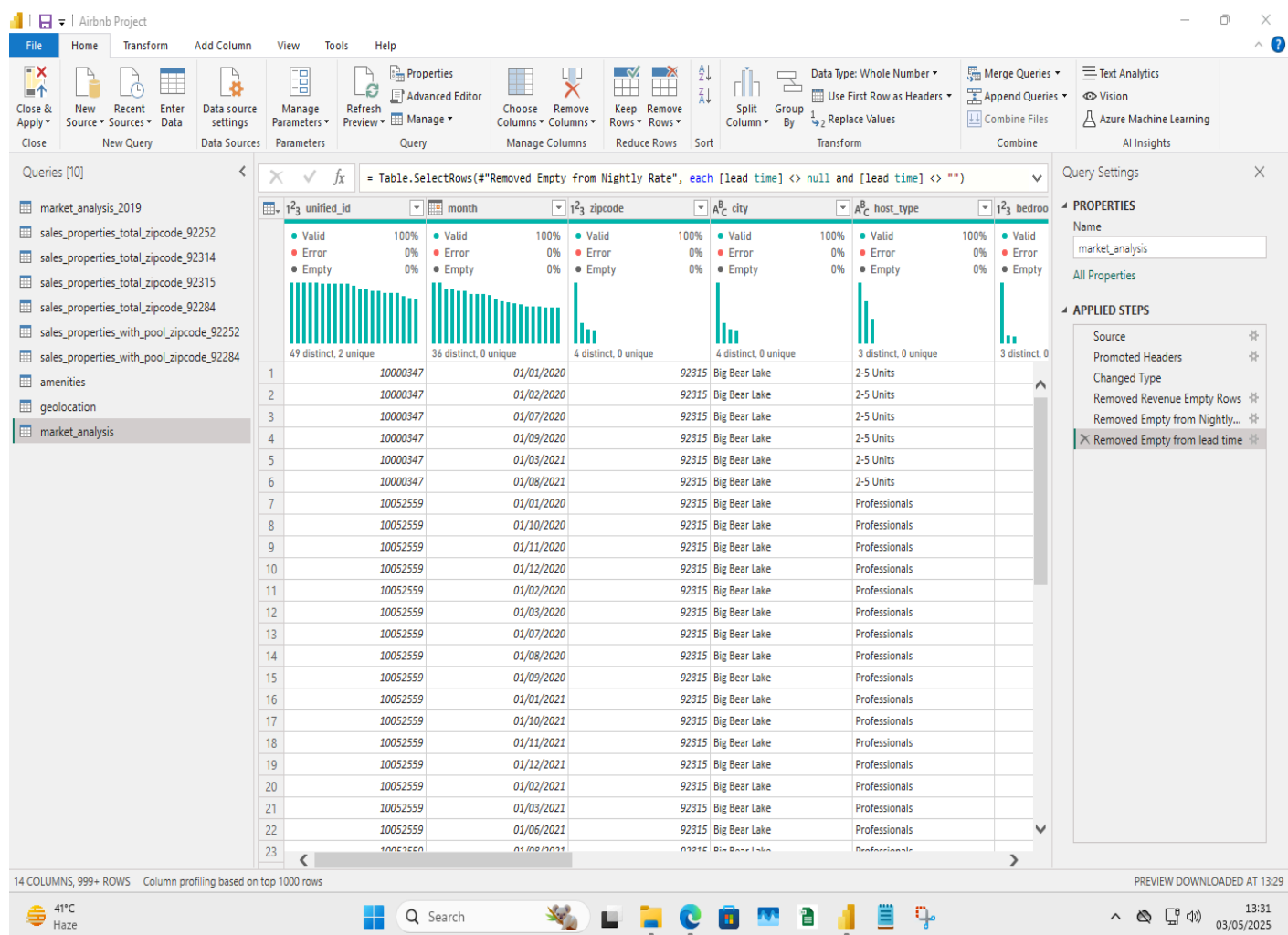
3) Market_analysis: "Removed null and blank values from the Revenue, Nightly Rate, and Lead Time columns using Power BI's Transform Data feature."

Explanation:

To ensure data accuracy and reliability in reporting, the following data cleaning steps were performed using Transform Data in Power BI (Power Query Editor):

- Null values and empty entries were removed from the Revenue, Nightly Rate, and Lead Time columns.
- These columns are essential for financial and operational analysis; therefore, maintaining complete and valid data is crucial.
- The cleaning was done using the *Remove Rows* and *Value Filters* options in Power Query, which ensures that only rows with meaningful numerical data are retained.
- This step prevents errors in calculations such as averages, totals, or trend analysis in Power BI visuals.

Below is screenshot attached for reference:-



4) Market_analysis_2019: "Removed empty values from the Nightly Rate and Lead Time columns using Power BI's Transform Data feature."

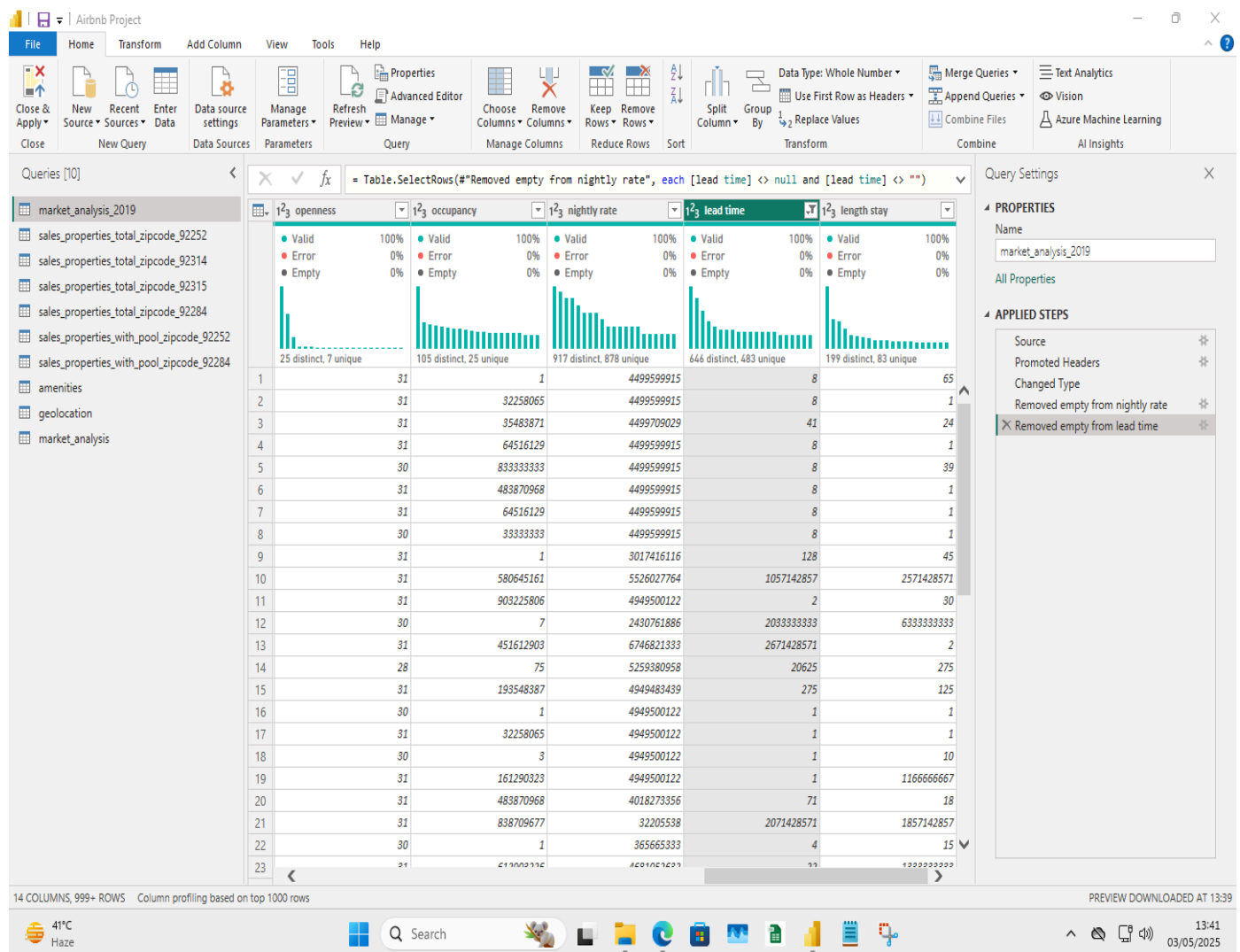
Explanation:

To improve data quality and ensure accurate analysis, empty rows in the Nightly Rate and Lead Time columns were removed using Transform Data in Power BI (Power Query Editor).

Specifically:

- The *Text Filters* or *Remove Blank Rows* functions were used to eliminate entries where these columns had no data.
- Cleaning these columns is essential to avoid calculation errors and to ensure that metrics like average rate or booking lead time are based on complete data.
- This step ensures the reliability of visuals and aggregations that depend on these numeric fields.

Below is screenshot attached for reference:-



5) Sales_Properties: "Appended four tables—*sales_properties_total_zipcode_92252*, *92284*, *92314*, and *92315*—into a single dataset. Removed empty values from the *Zestimate*, *Rent*, *Zestimate*, and *Broker Name* columns, and fully deleted the *Lot Size* column due to the absence of any data."

Explanation:

Using Power BI's Transform Data feature, the following data preparation steps were performed:

→ **Appended Tables:**

- ◆ Combined four datasets (*sales_properties_total_zipcode_92252*, *92284*, *92314*, and *92315*) into one consolidated table to enable unified analysis across zip codes.
- ◆ This was done using the *Append Queries* option in Power Query.

→ **Removed Empty Values:**

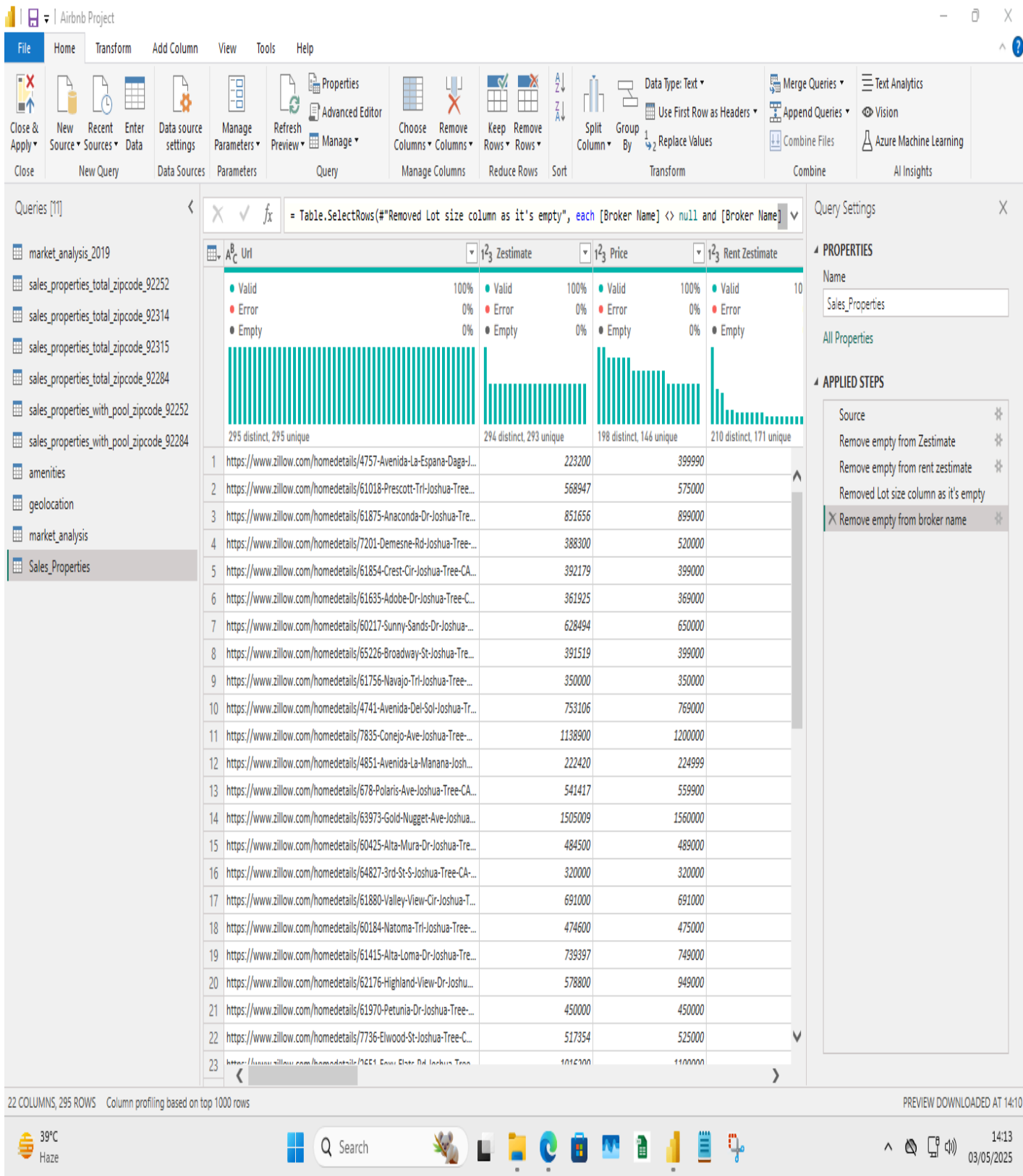
- ◆ Cleansed the combined dataset by removing rows where *Zestimate*, *Rent*, *Zestimate*, or *Broker Name* fields were empty.
- ◆ This ensures that only complete and reliable data points are retained for analysis and reporting.

→ **Deleted Unused Column:**

- ◆ The *Lot Size* column was completely removed since it contained no values across all records.
- ◆ Eliminating such columns helps streamline the dataset and improve performance during report generation.

These steps enhance data quality and ensure more accurate insights in Power BI dashboards and visuals.

The attached screenshot serves as a visual reference:



5) Sales_Properties_Pool: "Append two tables: *sales_properties_with_pool_zipcode_92252* and *sales_properties_with_pool_zipcode_92284*—into a single dataset. Removed empty values from the *Zestimate* column and completely deleted the *Lot Size* column due to the absence of any data."

Explanation:

Using Power BI's Transform Data feature (Power Query Editor), the following data transformation steps were carried out:

→ **Table Appending:**

- ◆ Combined the datasets *sales_properties_pool_zipcode_92252* and *sales_properties_pool_zipcode_92284* into a unified table using the *Append Queries* option.
- ◆ This consolidation allows for joint analysis of pool-related property sales data across both zip codes.

→ **Cleaning Empty Rows:**

- ◆ Removed rows where the *Zestimate* column contained no values.
- ◆ This step ensures that analyses involving property value estimates are based on complete and accurate data.

→ **Column Removal:**

- ◆ Fully deleted the *Lot Size* column as it contained no data in either table.
- ◆ Removing empty or redundant columns improves data model efficiency and maintains a clean dataset for reporting.

The attached screenshot serves as a visual reference:

Airbnb Project

FileHomeTransformAdd ColumnViewToolsHelp

Close & Apply

New Source

Recent Sources

Enter Data

Data source settings

Manage Parameters

Refresh Preview

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Split Column

Group By

Use First Row as Headers

Replace Values

Merge Queries

Append Queries

Combine Files

Text Analytics

Vision

Azure Machine Learning

Close

New Query

Data Sources

Parameters

Query

Manage Columns

Reduce Rows

Sort

Transform

Combine

AI Insights

Queries [12]

market_analysis_2019

sales_properties_total_zipcode_92252

sales_properties_total_zipcode_92314

sales_properties_total_zipcode_92315

sales_properties_total_zipcode_92284

sales_properties_with_pool_zipcode_92252

sales_properties_with_pool_zipcode_92284

amenities

geolocation

market_analysis

Sales_Properties

Sales_Properties_Pool

fx

= Table.RemoveColumns("#Remove empty from Zestimate",{"Lot Size"})

Url

Zestimate

Price

Rent Zestimate

Valid

Error

Empty

100%

0%

0%

27 distinct, 27 unique

Valid

Error

Empty

100%

0%

0%

27 distinct, 27 unique

Valid

Error

Empty

100%

0%

0%

24 distinct, 21 unique

Valid

Error

Empty

100%

0%

0%

26 distinct, 25 unique

3	https://www.zillow.com/homedetails/8552-Taft-Ct-Yucca-Valley-CA-9...	467500	530000	2.
4	https://www.zillow.com/homedetails/5088-Elata-Ave-Yucca-Valley-CA...	684150	699000	3.
5	https://www.zillow.com/homedetails/7478-Goleta-Ave-Yucca-Valley-...	525004	525000	2.
6	https://www.zillow.com/homedetails/7629-Rockaway-Ave-Yucca-Vall...	764822	775000	3.
7	https://www.zillow.com/homedetails/7354-Rubidoux-Ave-Yucca-Valle...	880671	900000	2.
8	https://www.zillow.com/homedetails/7399-Palomar-Ave-Yucca-Valley...	478372	489900	2.
9	https://www.zillow.com/homedetails/8210-Church-St-Yucca-Valley-C...	418600	479000	2.
10	https://www.zillow.com/homedetails/55960-Onaga-Tri-Yucca-Valley-...	277830	285000	1.
11	https://www.zillow.com/homedetails/58783-Santa-Barbara-Dr-Yucca-...	674473	699000	3.
12	https://www.zillow.com/homedetails/56763-Java-Dr-Yucca-Valley-CA-...	489600	489600	2.
13	https://www.zillow.com/homedetails/56839-Bonanza-Dr-Yucca-Valley-...	374800	545000	4.
14	https://www.zillow.com/homedetails/7731-Arrowhead-Dr-Yucca-Vall...	1076582	1100000	3.
15	https://www.zillow.com/homedetails/58345-Bonanza-Dr-Yucca-Valley...	478884	499950	3.
16	https://www.zillow.com/homedetails/61854-Crest-Cir-Joshua-Tree-CA...	392179	399000	1.
17	https://www.zillow.com/homedetails/4741-Avenida-Del-Sol-Joshua-Tr...	753106	769000	2.
18	https://www.zillow.com/homedetails/63973-Gold-Nugget-Ave-Joshua...	1505009	1560000	1.
19	https://www.zillow.com/homedetails/61880-Valley-View-Cir-Joshua-T...	691000	691000	2.
20				

FALSO

Query Settings

PROPERTIES

Name

Sales_Properties_Pool

All Properties

APPLIED STEPS

Source

Remove empty from Zestimate

Removed Lot Size Column

22 COLUMNS, 27 ROWS

Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 14:46

41°C

Light rain

Search

14:49

03/05/2025

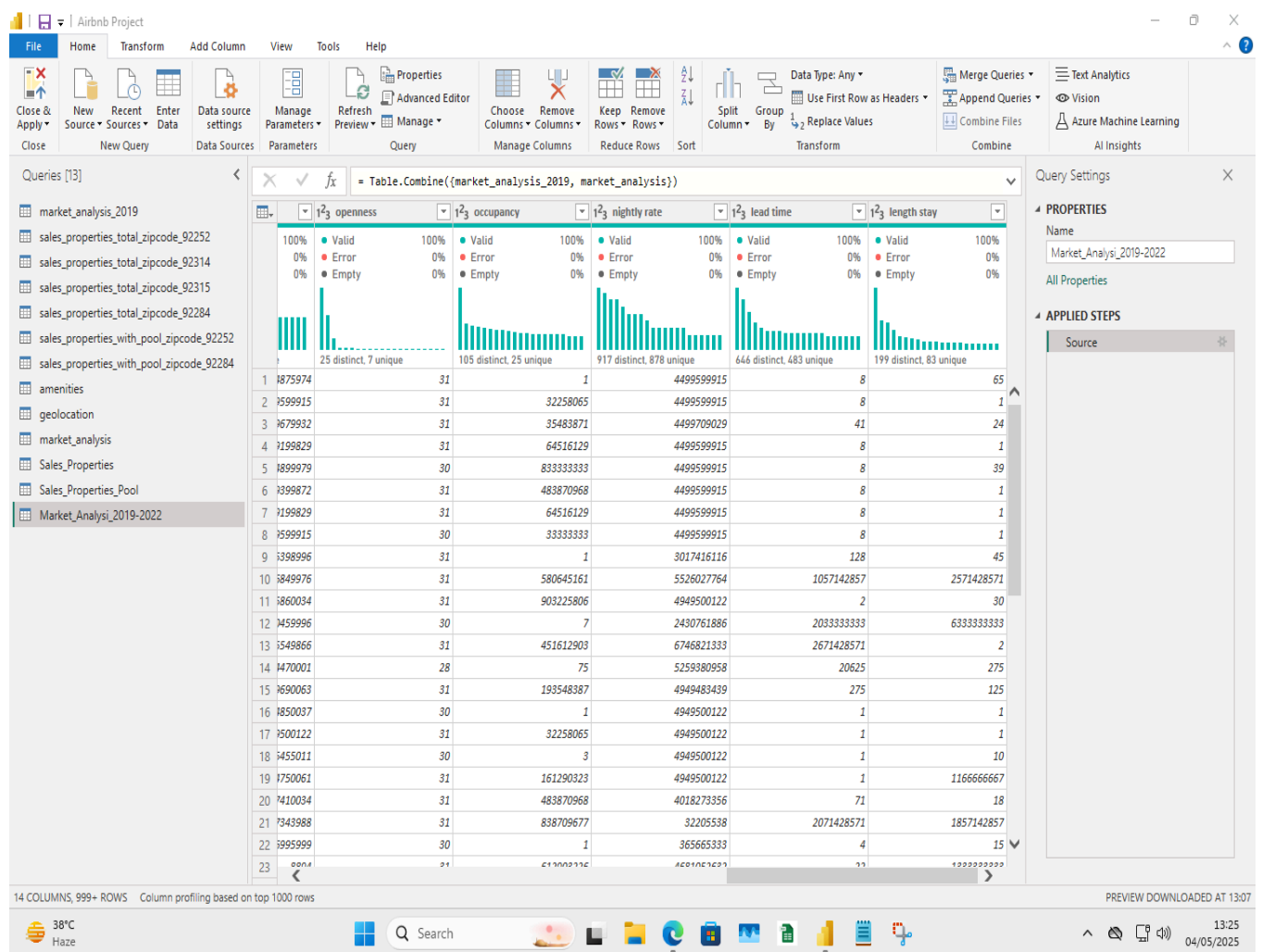
6) Market_analysis_2019-22: "Appended the *market_analysis* and *market_analysis_2019* tables into a single consolidated dataset."

Explanation:

Using Power BI's Transform Data feature (Power Query Editor), the following step was performed:

- The *market_analysis* and *market_analysis_2019* tables were combined using the Append Queries option.
- This action merged historical and current market data into a unified table, allowing for seamless analysis across multiple time periods.
- Appending these tables supports trend analysis, comparisons, and aggregated insights without switching between datasets.
- Data types and structures were aligned to ensure compatibility during the append process.

The attached screenshot serves as a visual reference:



Relationships between tables:-

1. amenities ↔ market_analysis_2019-22

Key field: unified_id

Relationship: Many-to-many

2. geolocation ↔ amenities

Key field: unified_id

Relationship: Many-to-many

3. market_analysis_2019-22 ↔ Sales_Properties

Key field: zip code

Relationship: Many-to-many

4. Sales_Properties_Pool ↔ Sales_Properties

Key field: zip

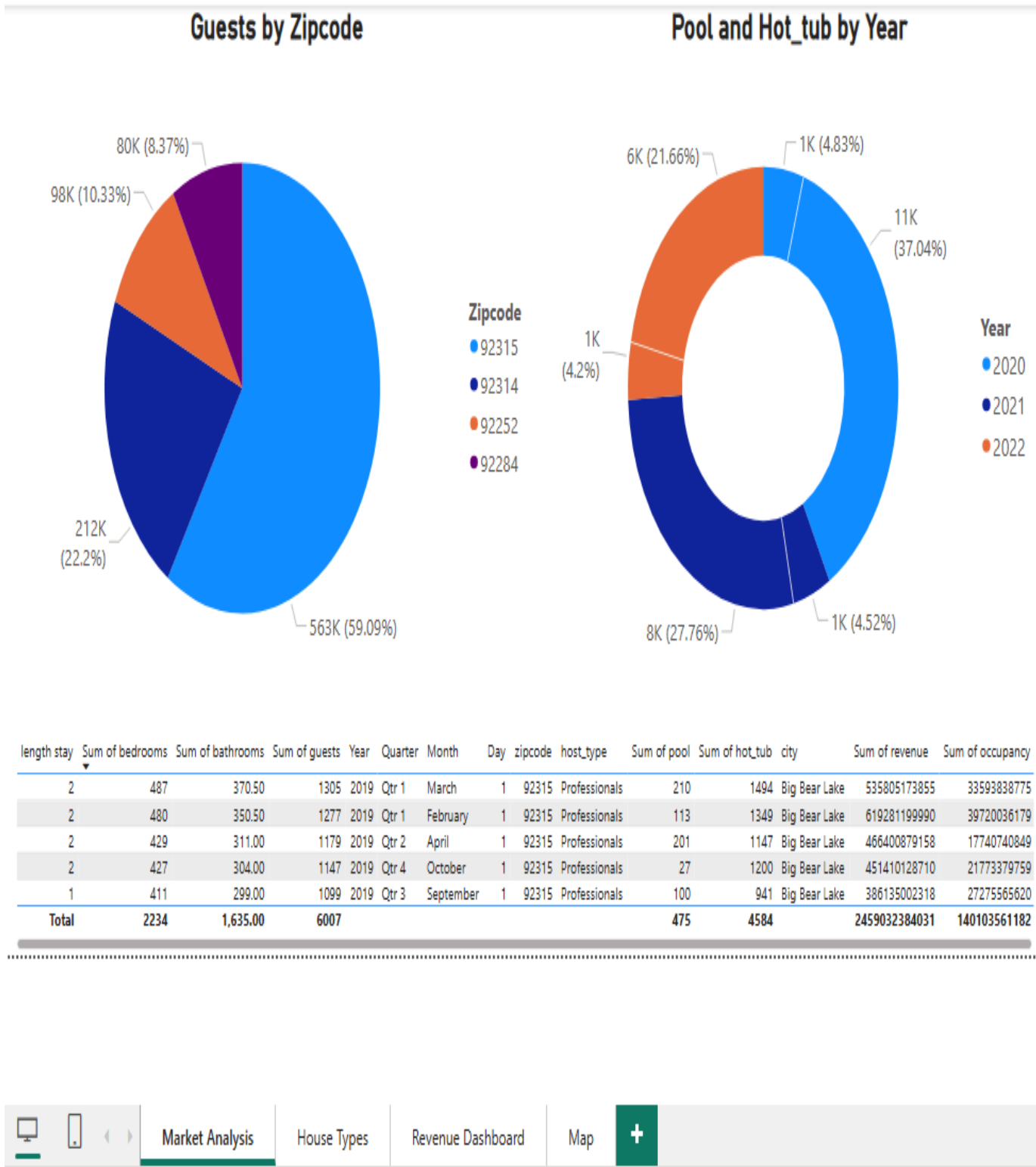
Relationship: Many-to-many

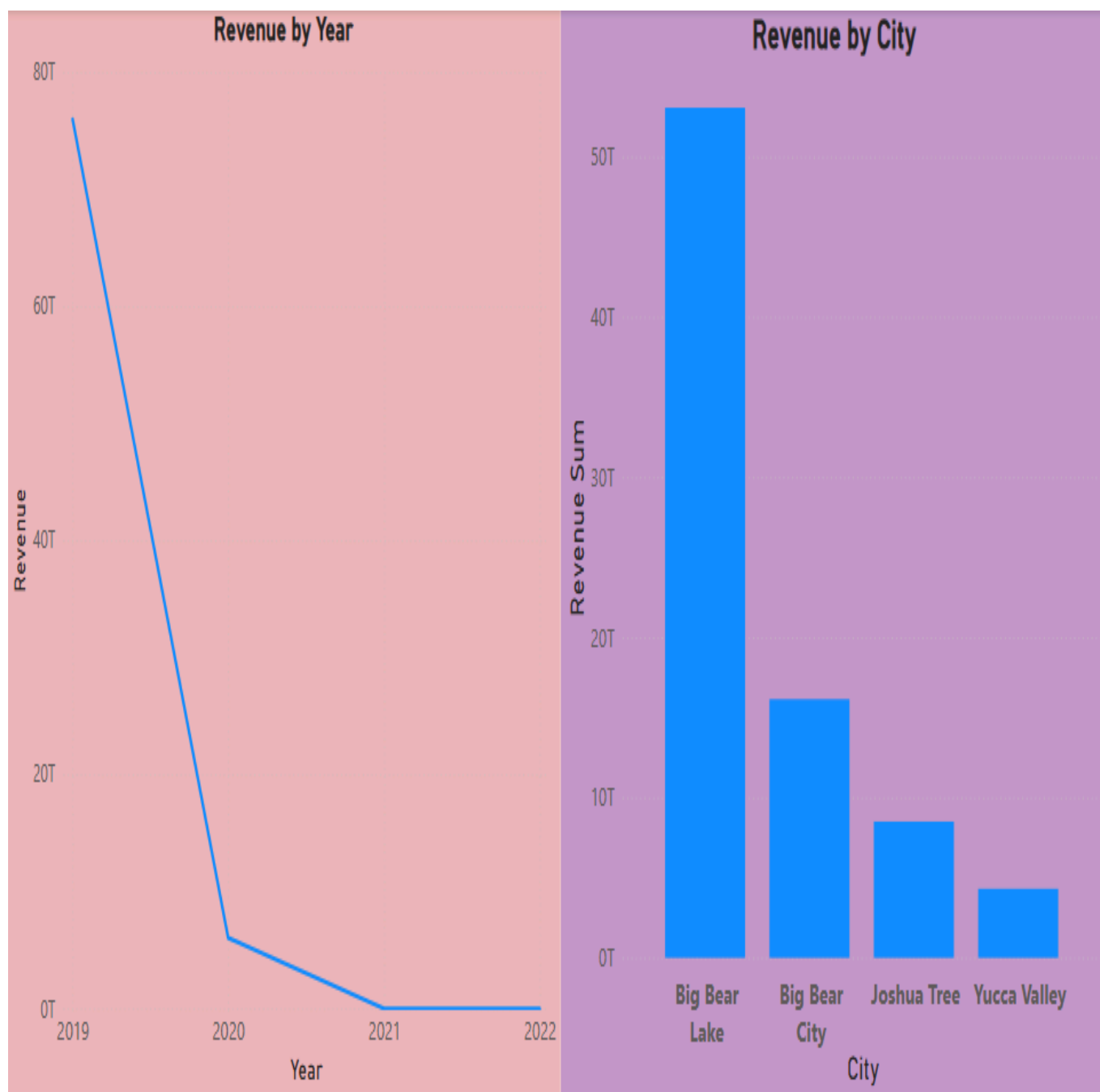
The attached screenshot serves as a visual reference:

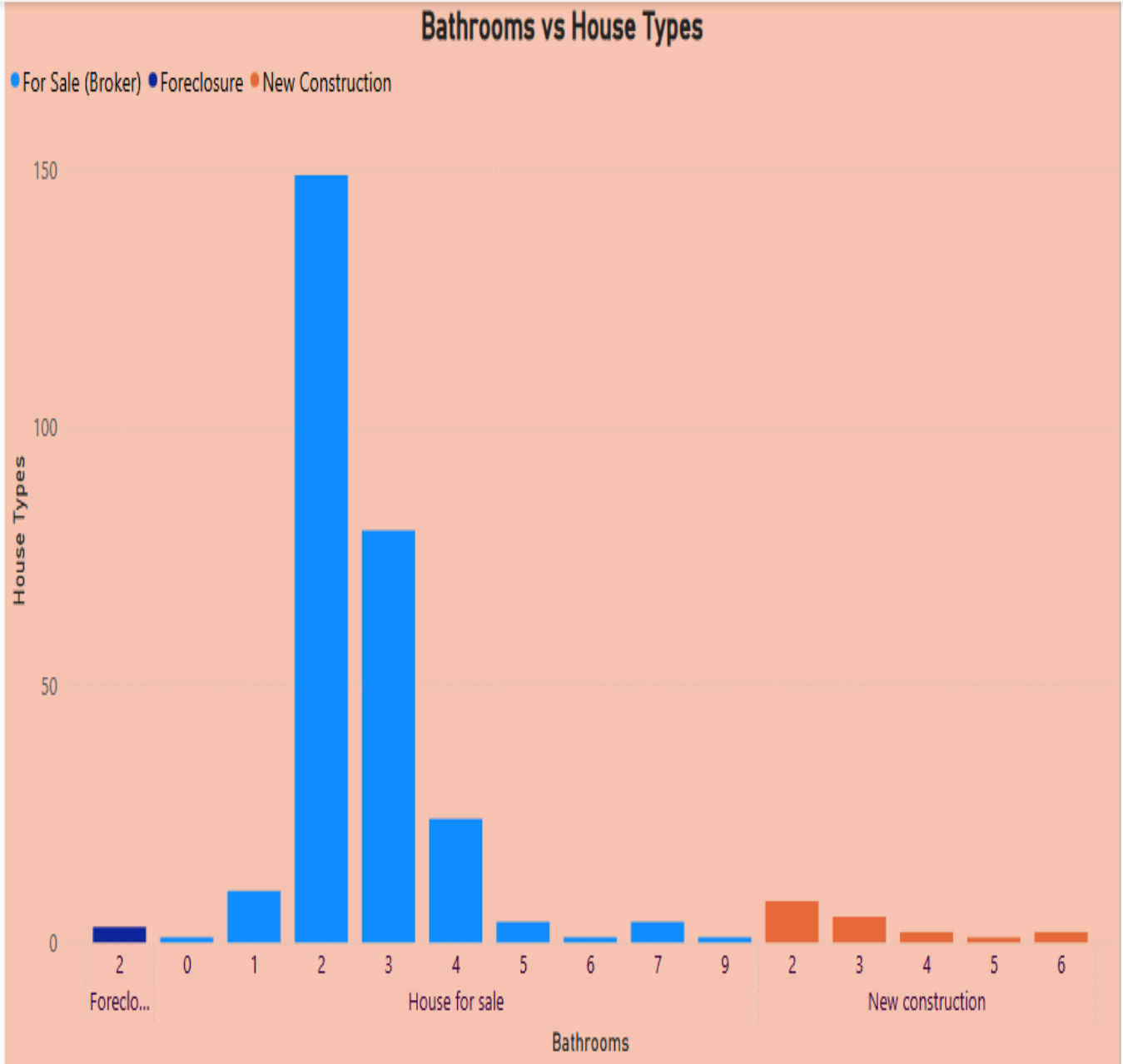
The screenshot displays the Microsoft Power BI Desktop application. The top ribbon shows the 'Home' tab with various toolbars including Clipboard, Data, Queries, Relationships, Calculations, Security, Q&A, and Sensitivity. The main workspace shows a data model with five tables: 'amenities', 'geolocation', 'Market_Analysis_2019-2022', 'Sales_Properties', and 'Sales_Properties_Pool'. The tables are connected by relationships, indicated by lines with cardinality symbols (1-to-many). The 'amenities' table is connected to 'geolocation' (1-to-many), 'Market_Analysis_2019-2022' (1-to-many), and 'Sales_Properties' (1-to-many). The 'geolocation' table is connected to 'Market_Analysis_2019-2022' (1-to-many). The 'Market_Analysis_2019-2022' table is connected to 'Sales_Properties' (1-to-many). The 'Sales_Properties' table is connected to 'Sales_Properties_Pool' (1-to-many). The right sidebar shows the 'Data' pane with a search bar and a list of tables. The bottom status bar shows the temperature as 38°C, the date as 04/05/2025, and the time as 13:23.

Dashboards:-

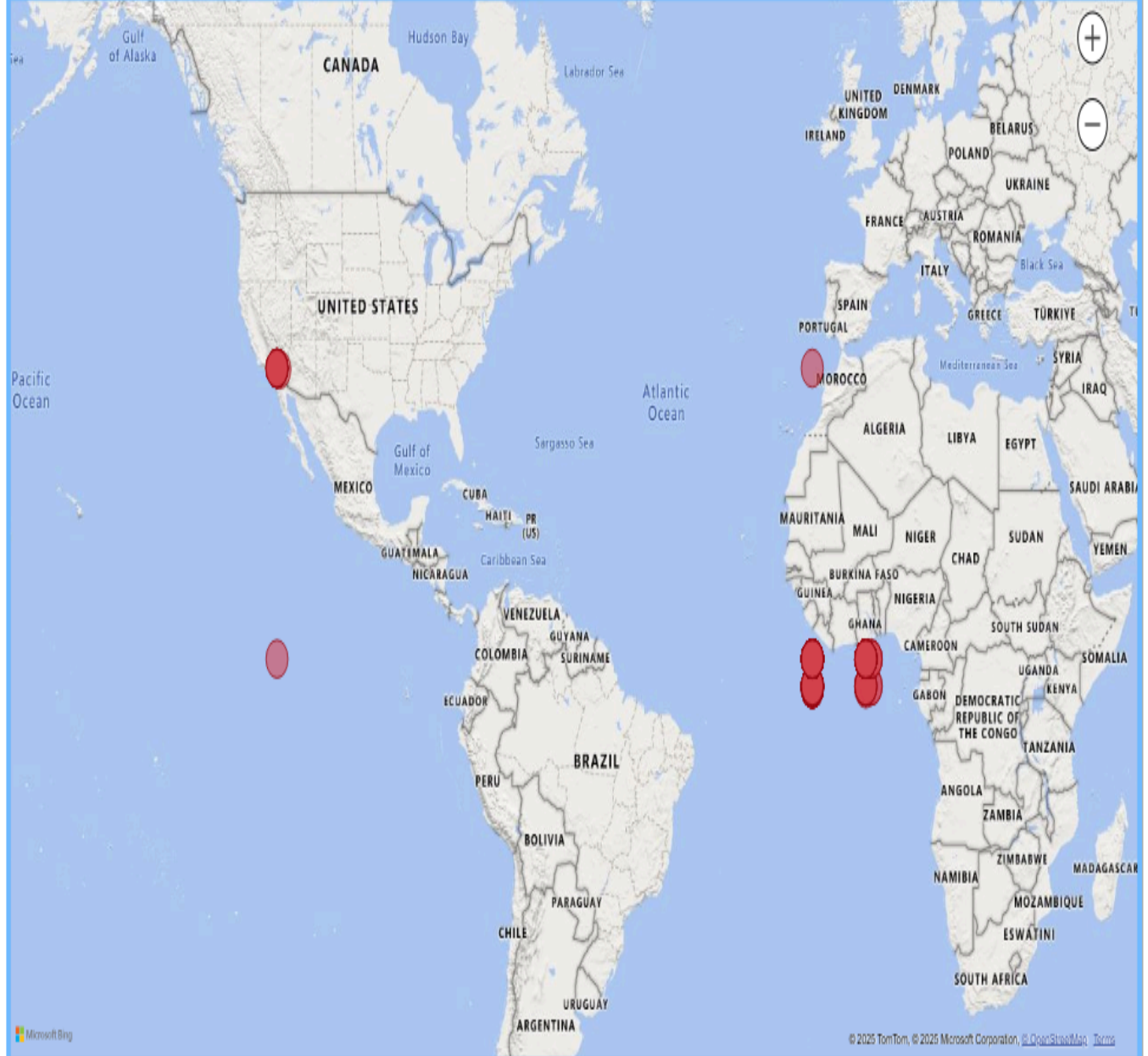
The attached screenshot serves as a visual reference:-







Map Locations



Conclusion:-

This analysis was conducted in Power BI with the objective of identifying the key factors influencing real estate pricing. While working with the dataset, several challenges emerged—most notably, the data was aggregated on a monthly basis rather than at the individual client or client group level, which limited the granularity of insights.

Some columns, such as "*Guests*" and "*Openness*", were found to be non-informative and lacked analytical value. However, one clear insight derived from the dashboard was the strong correlation between web traffic and revenue—higher traffic consistently led to increased income.

Through interactive visualizations, the analysis also revealed how COVID-19 significantly impacted the rental real estate sector, and how seasonal trends affect both demand and revenue, depending on the property's geographic location.

Recommendations:

To enhance the accuracy and depth of future analyses:

- Transition from aggregated monthly reports to client-level or group-level data. This would enable more targeted insights and support advanced analytics, including predictive modeling.
- Improve data collection processes, as approximately 25% of values were missing in several key columns. More complete and structured data will directly enhance the effectiveness of both manual and automated (e.g., machine learning) analysis.