ANALYTICAL SOLUTION

FOR

AIRBNB DATASET OF SEATTLE

-AROHI VAGHELA

https://github.com/arohi-01/Articles

CONTENT

INTRODUCTION

Airbnb is a global online marketplace that connects hosts who have accommodations to rent with travelers seeking short-term lodging. Founded in 2008, the platform allows individuals to list, discover, and book unique spaces, ranging from spare rooms and apartments to entire homes and unconventional stays like treehouses or castles.

The service operates on a peer-to-peer model, enabling hosts to earn income by renting out their properties while providing guests with diverse and often more affordable lodging options compared to traditional hotels. Airbnb emphasizes community and shared experiences, allowing travelers to immerse themselves in local cultures. Over the years, it has grown into a significant player in the hospitality industry, with millions of listings in cities worldwide, impacting local economies and housing markets.

This open source data was founded from Kaggle.com.

PURPOSE AND GOALS

The primary purpose of analyzing Airbnb data in Seattle is to gain insights into the short-term rental market, assess its impact on local housing, and identify trends that can inform stakeholders, including policymakers, property managers, and potential hosts.

Goals may include:

1. Analyze Pricing Trends:

- **Objective**: Track the average nightly rates for different types of listings (entire homes, private rooms, shared rooms) across various neighborhoods over time.
- **Outcome**: Understand how pricing fluctuates seasonally and in response to local events or economic conditions.

2. Identify Location-Based Performance:

- **Objective**: Evaluate occupancy rates and average daily rates across key neighborhoods in Seattle to determine which areas are most profitable for hosts.
- **Outcome**: Identify high-demand neighborhoods that may warrant further investment or marketing efforts.

3.Assess Market Trends Over Time:

- **Objective**: Examine the growth of listings, changes in occupancy rates, and shifts in guest demographics over multiple years.
- **Outcome**: Determine how the Seattle Airbnb market is evolving and predict future trends based on historical data.

4.Investigate Competitive Pricing Strategies:

- **Objective**: Analyze how different properties within close proximity set their prices and what factors (amenities, ratings, etc.) influence pricing strategies.
- **Outcome**: Identify best practices for competitive pricing that can enhance occupancy and revenue.

5. Evaluate Impact of Location on Guest Experience:

- **Objective**: Correlate guest ratings and reviews with specific locations to determine how location affects overall guest satisfaction.
- **Outcome**: Provide insights for hosts on how to leverage their location to improve guest experiences and ratings.

WHY I CHOSE THIS DATASET?

The Airbnb dataset is important for several reasons:

•**Valuable Insights**: The dataset provides a complete view of Seattle's Airbnb market, including listings, pricing, and booking trends.

•**Real-world dynamics:** The data reflects real-world dynamics in a major city's short-term rental market, offering valuable Findings for Executives and policymakers.

•**Variety Of Metrics:** It contains information on property types, locations, reviews, and availability, enabling in-depth analysis.

•**Visualization Potential:** The dataset is well-Structured for building an interactive dashboard in Power BI to Investigate patterns and trends.

•**Open Source:** Accessible through Kaggle, making it a well-documented and reliable source for analysis.

HOW I CONNECTED THE DATASET?

We are utilising two primary CSV files to build our dashboard in Power BI.

- **Listing CSV:**
  This file contains detailed information about listing ID, Name of customers, host information, latitude and longitude, property type, room type, accommodate, bathrooms and bedrooms, price ranges per month and week.
- **Calendar CSV:**
  This file contains information regarding dates, listing ID as well as booking status.

These datasets were then integrated into power BI using ETL to create comprehensive and interactive dashboard.

KEY EXPECTATIONS

- **Pricing Insights:**

- Identify average nightly rates for various property and room types.

- Analyze how pricing affects booking rates.

- **Booking Insights:**

- Examine booking trends throughout the year and identify peak times.

- Analyze customer preferences for different property types.

- **Geographical Distribution**:

- Determine which neighborhoods have the highest concentration of listings and bookings.

- Compare listings and bookings with average prices in those areas.

- **Cancellation Rate:**

- Identify cancellation rates throughout the year.

- **Recognising the top 10 hosts of the year**

KEY QUESTIONS

Here are some of the crucial questions that rose while performing analysis:

1. **What are the peak booking times throughout the year, and how do they relate to local events or festivals?**
2. **How do different property types (e.g., entire home, shared room, private room) perform in terms of pricing and occupancy?**

3.**How do seasonal trends affect cancellation rates, and are there specific times of year when cancellations peak?**

METRICES AND KPIS

To meet the above expectations I have chosen certain measures which can assist in making the analytical solution more comprehensive. The measures are as follows:

1. **Total Listings:**
   Visualization: Card
   Description: Total listing of particular time period can be seen.
2. **Total Bookings:**
   Visualization: Card
   Description: Total number of bookings are displayed.
3. **Cancellation Rate:**
   Visualization: Card
   Description: Total number of cancellation rate is displayed.
4. **Sum of Average night price by room type:**
   Visualization: Stacked column chart
   Description: This chart gives overview of what type of room is more expensive.
5. **Bookings by room type:**
   Visualization: Pie Chart
   Description: This graph gives overview of what type of rooms were mostly preferred by customers.
6. **Total bookings and total cancellation by month:**
   Visualization: Clustered Column Chart
   Description: This chart gives information about the total bookings and total cancellation per month which will be helpful to know which month is the peak time.
7. **Top 10 host of the year:**
   Visualization: Tree map
   Description: It shows who were the top 10 hostess of that particular time.

8. **Average response rate by month:**
   Visualization: Gauge Chart
   Description: It shows what was the the average rate of response from Hosts.
9. **Total Bookings by property type:**
   Visualization: Donut Chart
   Description: This show which property type was mostly booked by Customers.
10. **Total Bookings by longitude and latitude:**

    Visualization: Map
    Description: This shows where people mostly liked to booked their stay.
11. **Sum of average night price by property type:**

    Visualization: Clustered column Chart
    Description: This shows how prices are different for different property Type.

DATA MODEL

I normalised both the csv files and ended up having a fact table , property type look up table, host lookup, location lookup, booking lookup as well as price lookup. I even created a date table as the data I got was of specific year. Also, to meet the expectations I created measures as well as group by.

Here are some of the measures:

-Average Response rate by host

-Bookings by room type

-Cancellation Rate

-Percentage of bookings by room type

-Total Listings

-Total Bookings

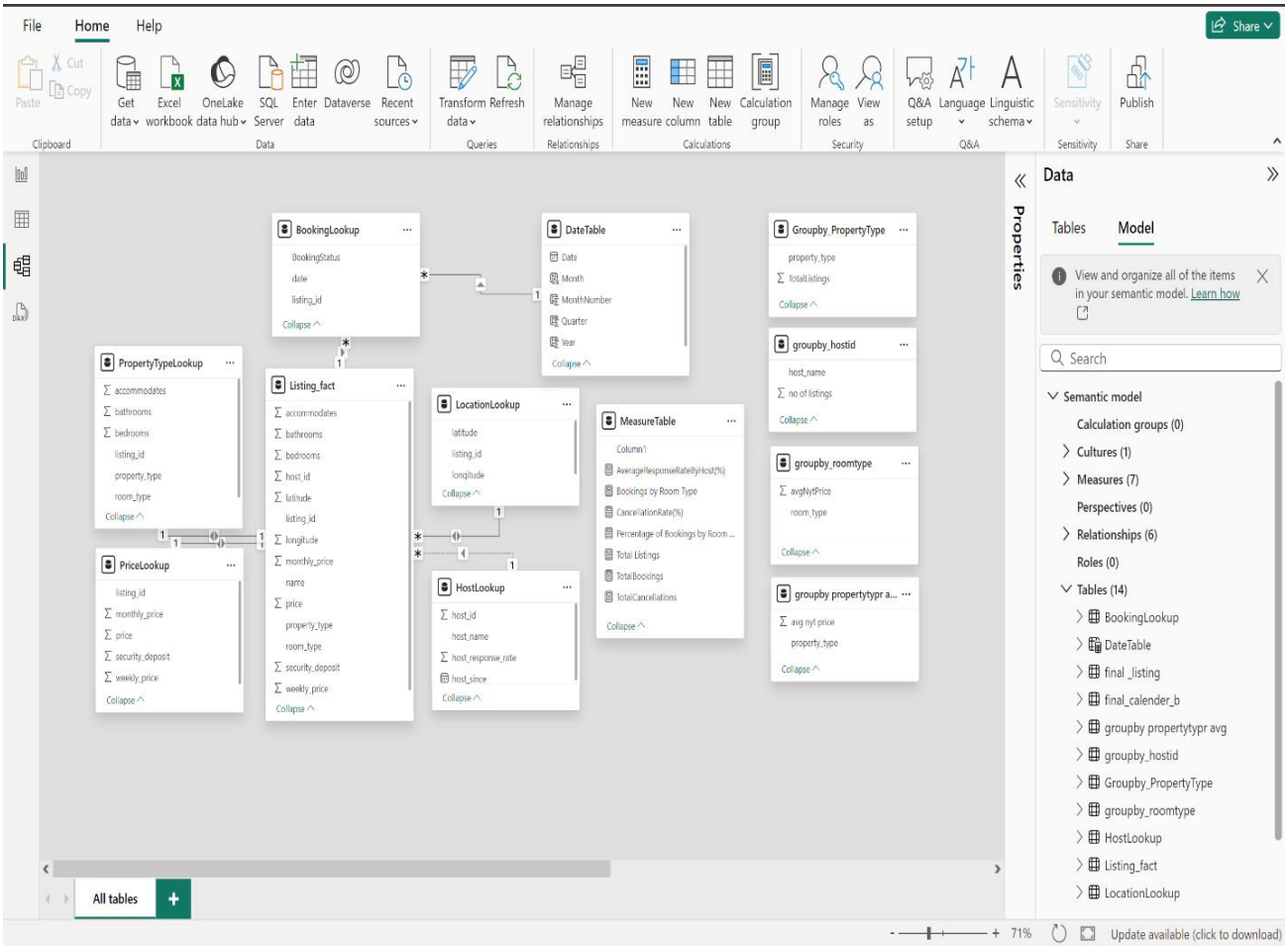
Here are some of the group by functions which I created:

-Group by property type

-Group by room type

-Group by host id

-Group by property type and average night price

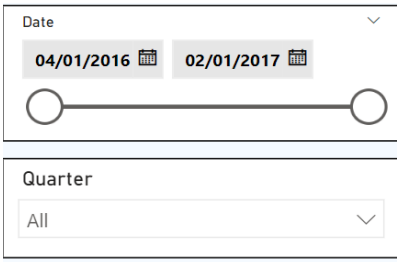


https://github.com/arohi-01/Articles

RESULT

1) **DATE AND QUARTER:**

This date card is introduced in the dashboard to show the time period of dataset for example of what year, month or week the data is. One can even change the dates to check the analysis of that particular day or month.
The quarter card divided the yearly data into four parts.
So that one can watch the productivity per four months.

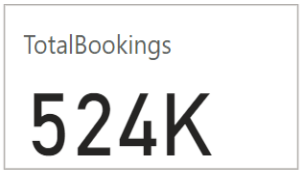


2) **TOTAL LISTINGS:**

This total listing card shows how many listings were listed in the particular year.

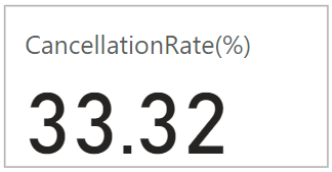
3) **TOTAL BOOKINGS:**

This total booking card shows what were the total bookings In the given year.



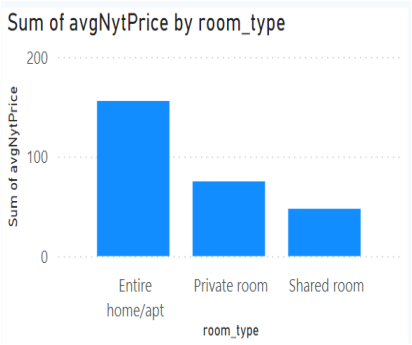
4) **CANCELLATION RATE:**

This cancellation rate card shows what was the cancellation rate during that year, like how many customers cancelled their booking and due to what reasons.



5) **SUM OF AVEARGE NIGHT PRICE BY ROOM TYPE:**

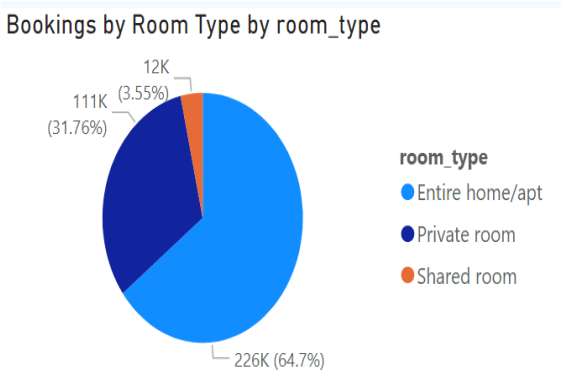
Here I chose stacked column chart to analyse prices per room types so that the comparison is clear. As in graph one can see the price of one night in entire home/apt is the highest while is the least for shared room, which is a obvious thing. This graph further assists in analysing which type of room were mostly preferred by customers.



https://github.com/arohi-01/Articles

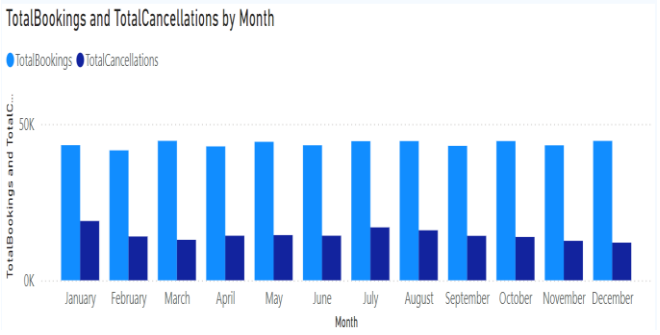
6) BOOKINGS BY ROOM TYPE:

This pie chart of bookings by room type gives a brief overview of which room was booked more. As in graph, one can see whole house or apartments were mostly booked, followed by private room and shared room respectively. From this graph I analysed that people on vacation with families might be in favour of booking entire home while alone travelling women usually prefer a private room rather than going for shared one and there are very few who are comfortable in sharing a room with stranger.



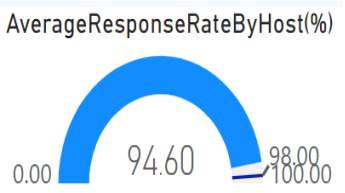
7) TOTAL BOOKINGS AND TOTAL CANCELLATION PER MONTH:

This clustered column chart provides valuable insights on total bookings and total cancellation per month. This chart is great if there are minor changes in numbers as one can easily see the difference. Through this graph I got to know that December had the least cancellation rate maybe because of Christmas while January has the highest cancellation rate. If I talk about highest bookings, then according to graph, March was the most booked month to travel to Seattle. This graph helped in knowing which was the peak time when people visited the most.



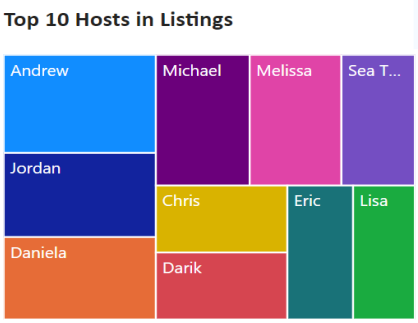
8) AVERAGE RESPONSE RATE:

This gauge chart helped in knowing what was the average rate of response by host. As in chart, it is patent that the hosts were very active in responding to customer's queries. Maybe in future they can try to convert rate of 94.60 to 98.00% so that no one can beat Airbnb.



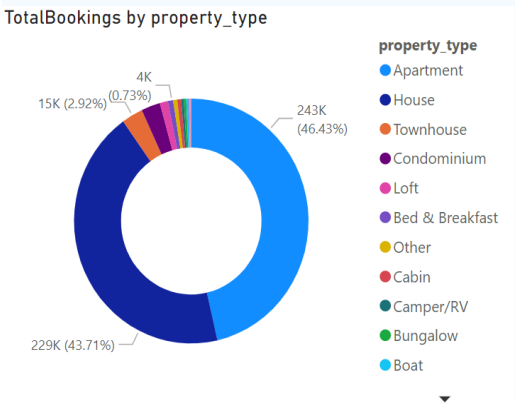
9) TOP 10 HOSTS:

This tree map was mainly focused just to appreciate top 10 hosts of the given year so that every host push themselves harder to be in the list which can indirectly help Airbnb in being the best.



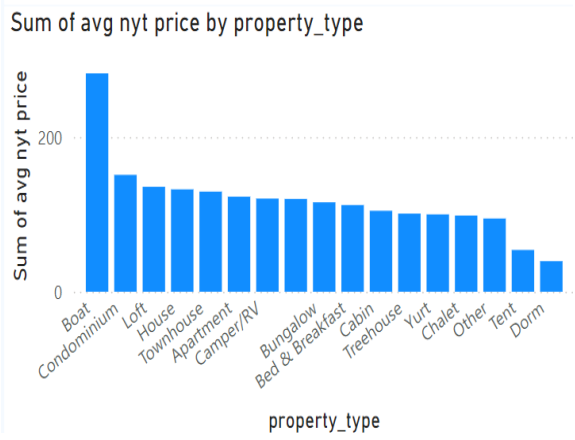
10) TOTAL BOOKINGS BY PROPERTY TYPE:

This donut chart aids in by providing information regarding which property type was the most liked by customer. And it is clear from the chart too that people more fond of booking apartments and as we know they are mostly located in downtown so maybe people preferred to live around facilities. Also, houses were popular among some of the customers. This chart also helped in letting know which locations were preferred along with that it also helped in knowing despite of different prices for distinct properties which was the most liked.



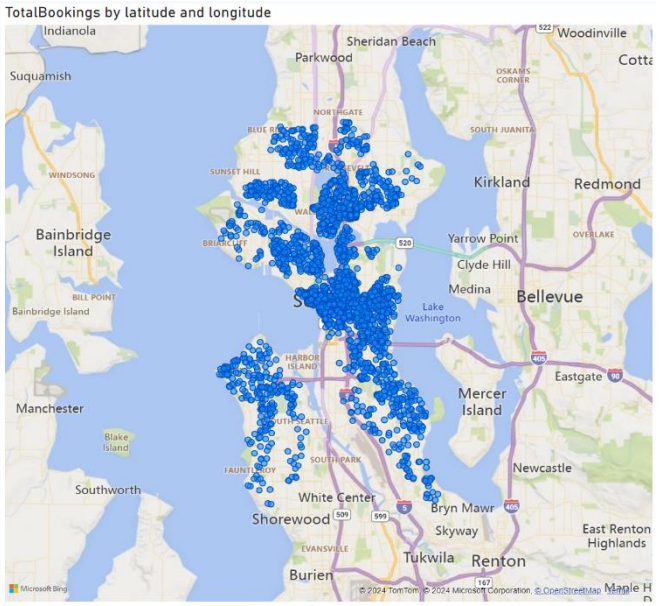
11) SUM OF AVERAGE NIGHT PRICE BY PROPERTY TYPE:

This clustered column chart describes how the average night prices vary by property type. As the chart suggests boats were the most expensive while dorm is the cheapest and the rest of the properties were average in terms of pricing they had minor differences.



12) TOTAL BOOKINGS BY LATITUDE AND LONGITUDE:

This map describes which locations were mostly booked in the given year. This map is linked with property type, if in the dashboard one selects apartments, it shows at which locations the apartments are located. This also helped in analysing which location were liked by customer if they booked a property which can be helpful in analysing trends. Ans it may help Airbnb to introduce such properties more which can indirectly help their profit margin too.



CONCLUSION

-This analysis provides executives with critical insights into Airbnb pricing trends, booking patterns, and neighborhood dynamics in Seattle.

-It enables informed decision-making and strategic planning.

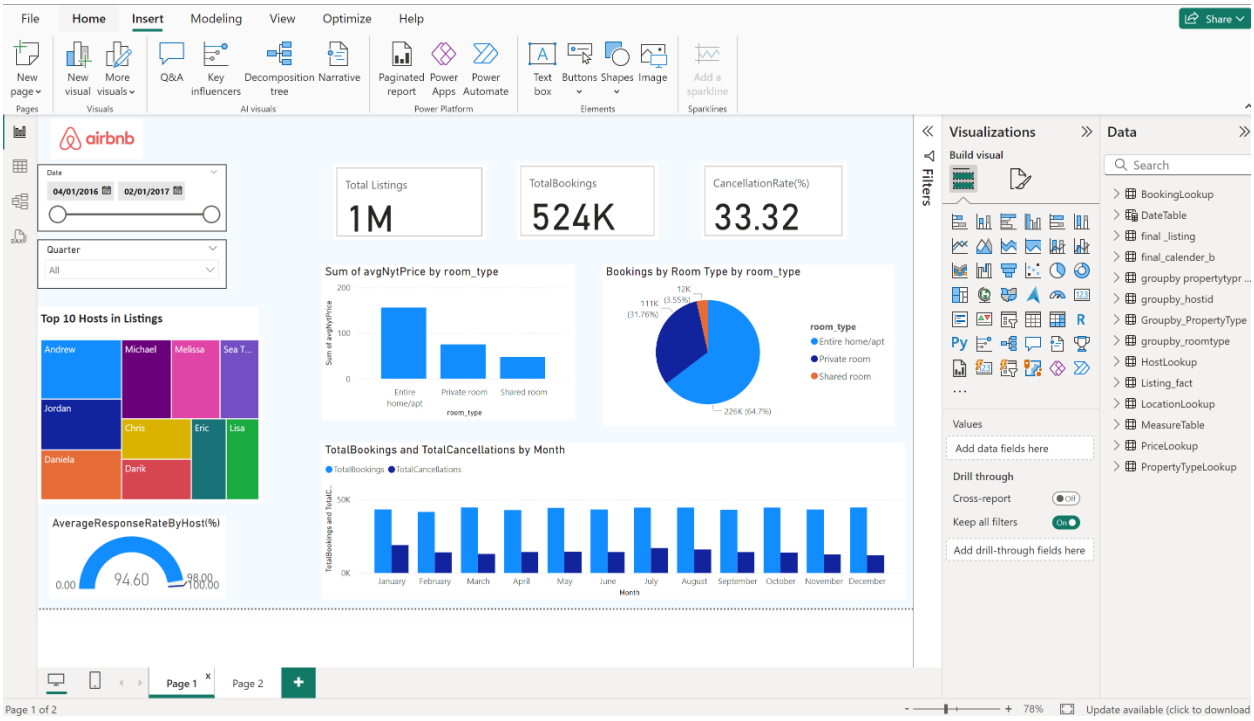
-Hosts can use these insights to optimize their listings and adjust pricing strategies.

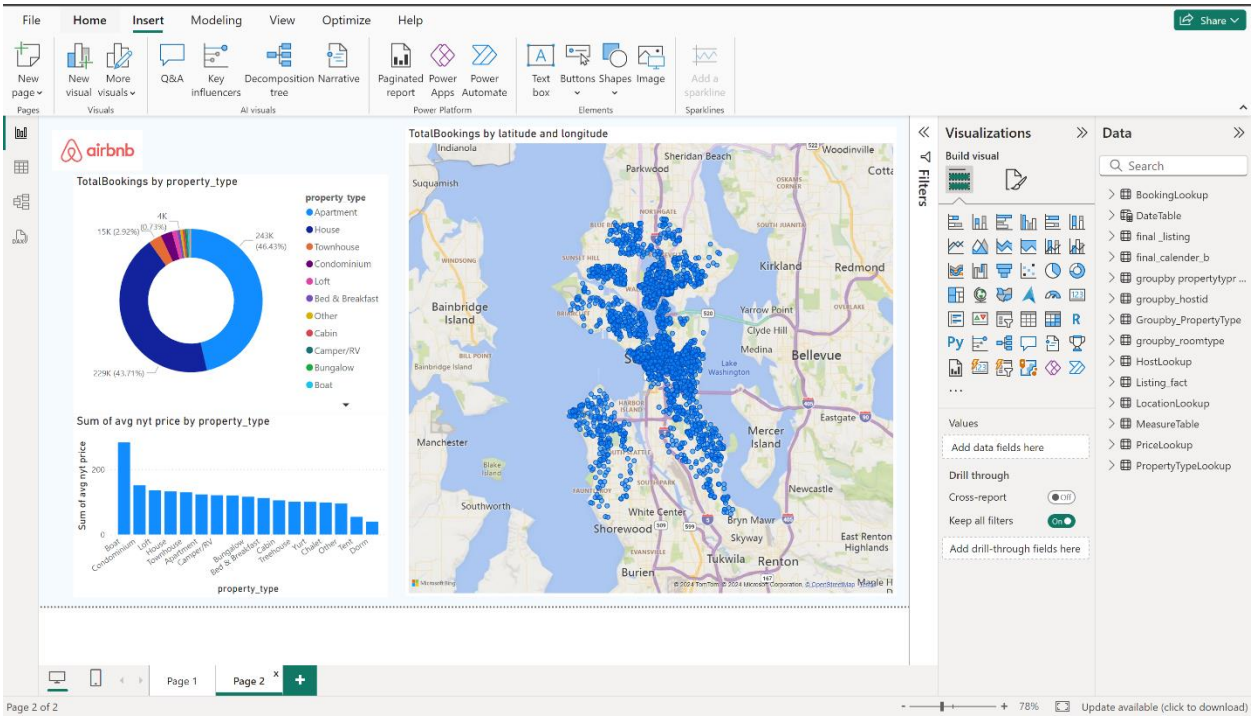
- The project showcases how data visualization can turn complex data into actionable insights.

https://github.com/arohi-01/Articles

APPENDIX

Here is the dashboard:





https://github.com/arohi-01/Articles

- **SUM OF AVEARGE NIGHT PRICE BY ROOM TYPE:**
  Visualization: Stacked column chart
  X axis: Room type
  Y axis: Sum of average night price

- **TOTAL BOOKINGS AND TOTAL CANCELLATION PER MONTH:**
  Visualization: Clustered column chart
  X axis: Month
  Y axis: Total bookings and total cancellation rate

- **SUM OF AVERAGE NIGHT PRICE BY PROPERTY TYPE:**
  Visualization: Clustered column chart
  X axis: Property type
  Y axis: Sum of average night price

➢ **Listing table description:**

Listing ID- ID of property listed
Name- To know the customer's name
Host ID-This ID is given to the person to whom the home belongs whether it is any business, individual or any property manager.
Host Name- The name of host
Host since- since when the host is on Airbnb
Host response – how well the host responds to customers
Latitude and Longitude- to know the location of where the properties are located.
Property Type – It includes what type of property it is whether it is a studio, apartment or house.
Room type- it includes what type pf room is available whether it is entire house, private room or shared room.
Accommodate- it gives the the data of how many person are staying
Bedroom and bathrooms – this tells us that how many bathrooms and bedrooms are included in the property.
Price- Gives the price on basis of for how many weeks or months one needs the property as well as how much security deposit one needs to pay.

- **Calender table description:**

  Listing ID- Id of property listed

  Date- Date of property listed

  Booking Status: describes status of booking whether the property is booked, not booked or pending.

- **Measure Table syntaxes:**

  These were the measure syntaxes created in order to meet expectation.

  - AverageResponseRateByHost(%) = AVERAGE(HostLookup[host_response_rate])*100
  - Bookings by Room Type =
    CALCULATE(
    COUNT('BookingLookup'[listing_id]),
    'BookingLookup'[BookingStatus] = "booked",
    TREATAS(VALUES('Listing_fact'[listing_id]), 'BookingLookup'[listing_id])
)
  - CancellationRate(%) =
    DIVIDE(
    [TotalCancellations],
    [TotalBookings],
    0 // This is the alternate result if the denominator is zero to avoid errors
) *100
  - Percentage of Bookings by Room Type =
    DIVIDE([Bookings by Room Type], [TotalBookings]) * 100
  - Total Listings = COUNT(BookingLookup[listing_id])
  - TotalBookings =
    CALCULATE(
    COUNT(BookingLookup[listing_id]),
    bookinglookup[BookingStatus] IN {"booked", "cancelled"}
)
  - TotalCancellations =
    CALCULATE(
    COUNT(bookinglookup[BookingStatus]),
    bookinglookup[BookingStatus] = "cancelled"
)