

Business Analysis on AirBnb Inc.

MSIS 2621: Business Intelligence and Data Warehouse.



Submitted By:

Bhagyashri Badgujar

Saira Gurung

Aasavari Kaley

Kshitija Sankhe

Rujuta Tamhankar

| | |
|-----|---|
| 1. | Business Description |
| 2. | Problem statement, user story and Business processes |
| 3. | Swim Lane Diagram |
| 4. | Performance Indicators |
| 5. | Data Modelling and Designing |
| 6. | Data Visualization Using Tableau |
| 7. | Data Assumptions |
| 8. | Data Insights |
| 9. | Dashboard 1: Descriptive Analysis |
| 10. | Dashboard 2: Predictive Analysis |

Business Description:

Airbnb, Inc, is a platform for customers to find and offer options for accommodations and experiences, It serves as an interchange marketplace for customers who want to offer up their place for lodging, or for customers to book other people's homes or rooms, without actually owning any of the spaces, or hosting any of the events. The company receives revenue from these transactions in the form of a commission.[1]

The lodging provided by Airbnb is commonly for guests seeking short-term accommodation for tourism purposes, although longer terms are also allowed. The guests are able to find the place of their choice by using several filters and selections such as, the type of lodging, whether it is a room or an entire home, or get into the specifics of the type of the home, such as, bed and breakfast, hostels, or unique homes such as a treehouse, and other filters including the dates, location of home, and the price per night for the place

Before proceeding with their booking, guests are required to verify their personal information, as well as their payment information. Hosts have the ability to request for guests to provide a scan of their government-issued identification before they accept the guests' reservation request. A secure messaging system within the platform allows potential guests to communicate with hosts, enabling them to make a better decision regarding their potential reservation by making enquiries. When the hosts list their place, they provide the price for the listing, the amenities provided, and additional details and rules pertaining to the house or an event that was listed. While Airbnb provides recommendations for pricing, the ultimate decision lies in the host for the determination of price. The customers of Airbnb, both hosts, and guests can write a review on their experience with the host, accommodation, or event. [2]

Problem Statement:

The project aims to analyze the business trends pertaining to accommodation industry in New York City, one of the most expensive cities in the United States, which is comparable to the bay area. By utilizing techniques of descriptive and predictive analysis, as well as multiple data sources, an evaluation of the lead and lag measures in the property lending business space will be made with a goal that this business analysis would foster the growth of such businesses in the market. This ultimate goal can be achieved by employing the results of the analysis to provide a data solution to Airbnb's potential hosts wanting to invest in properties in NYC as end users.

User Story:

As a customer, I want to create an account on the Airbnb website so that I can make a reservation. The aim of the customer is to do a reservation for his/her stay in the least cost and time possible.

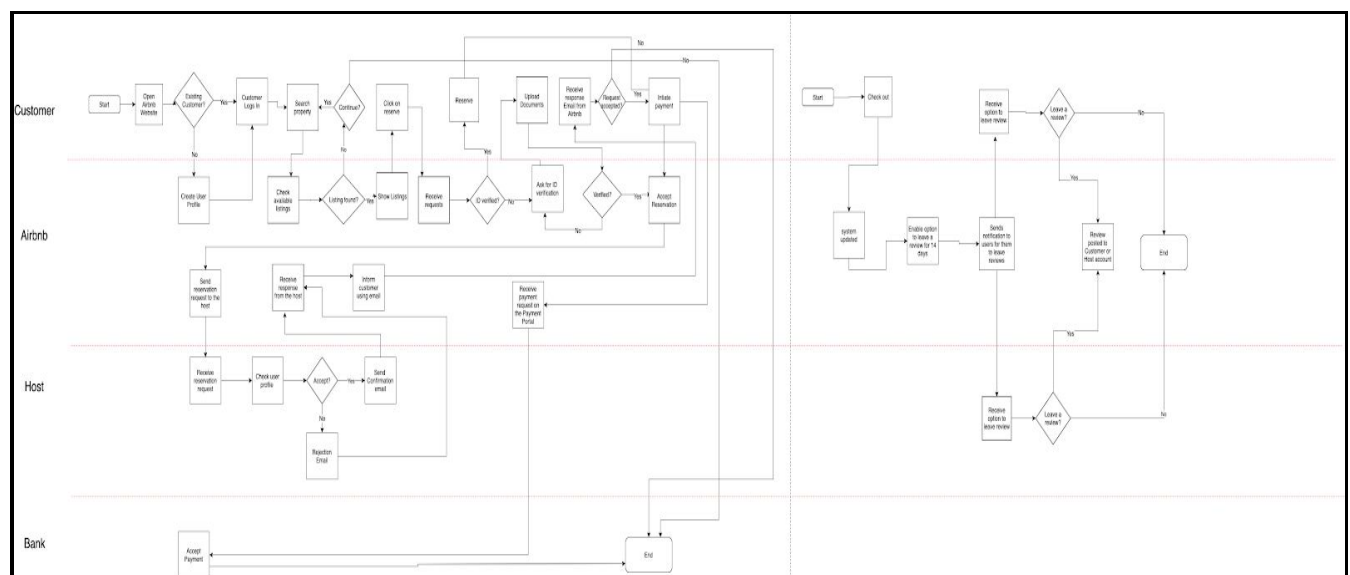
Acceptance Criteria: When the reservation is made, the customer will receive a booking confirmation number.

Business Processes:

- The customer visits the Airbnb website application and registers for an account if an existing account has not yet been created.
- After registration of an account, the customer logs into the account using 'Username' and 'Password'.
- In the case of an existing customer, the customer logs in to the account directly by using 'Username' and 'Password'.
- When logged in, the customer inputs values for the location, the desired dates of check-in and check-out, the number of guests, and clicks on 'Stays'
- The system displays the available listings based on the customer's input criteria. If the listings are not available, the system prompts the customer to modify the search.
- To proceed with reservation, the customer selects a preferred listing and clicks on 'Reserve'.
- Airbnb receives the request and checks if the identity of the customer has been verified.
- If the customer has not yet been verified, system prompts customer for verification.
- Once the verification is completed via text or email, the customer gains ability to reserve the property.
- Airbnb will send the booking request to the property host.
- The host checks the user profile and confirms/rejects the booking.

- Once confirmed, user will receive a confirmation email with property and host details and it will give confirmation details along with host details to the customer.
- Customer is led to a payment page.
- The customer initiates the payment.
- Airbnb website receives the payment information and forwards it to the bank for further processing.
- Once the payment is accepted by the bank, a confirmation email is generated and sent out.
- Once the check out date passes, hosts and guests get the option to leave a review as well as a star rating up to 5, for each other.
- This review will only be posted after both parties have completed it, or after 14 days has passed from the check out date.

Swim-lane diagram:



Performance Indicators:

For an organization heavily based upon the internet as well as real estate infrastructure, both Lead and Lag measures would prove to be imperative.

Lag Measures:

Revenue per neighborhood group

We have 5 neighborhood groups in the database. Based on the neighborhood group we are getting the revenue per neighborhood group.

Revenue per room type

Airbnb database has a table room with different room types. Namely Private room, Entire home/apt and shared room. Based on these types we are calculating revenue per room type.

Revenue per holiday season and non holiday season

To get the seasonality, we have divided the number of months into holiday season and non holiday season. Holiday season comprises of May, June, July, August, December and January. Rest of the months are considered as non holiday season.

We are calculating revenue from each of the season.

Top neighborhood by no. of hosts for each neighborhood group

We are finding the top neighborhood by considering number of hosts for each neighborhood group.

Cheapest property per neighborhood group by room type

We are finding the cheapest property per neighborhood group by room type.

Cheapest property per neighborhood group by price range.

We are also finding the cheapest property per neighborhood group by considering the price range.

Lead Measures:

Revenue by neighborhood group per month

Considering the neighborhood groups we are calculating the revenue earned from each group on a monthly basis.

Estimated revenue forecast per month per neighborhood group

We are also forecasting the estimated revenue on a monthly basis considering neighborhood group.

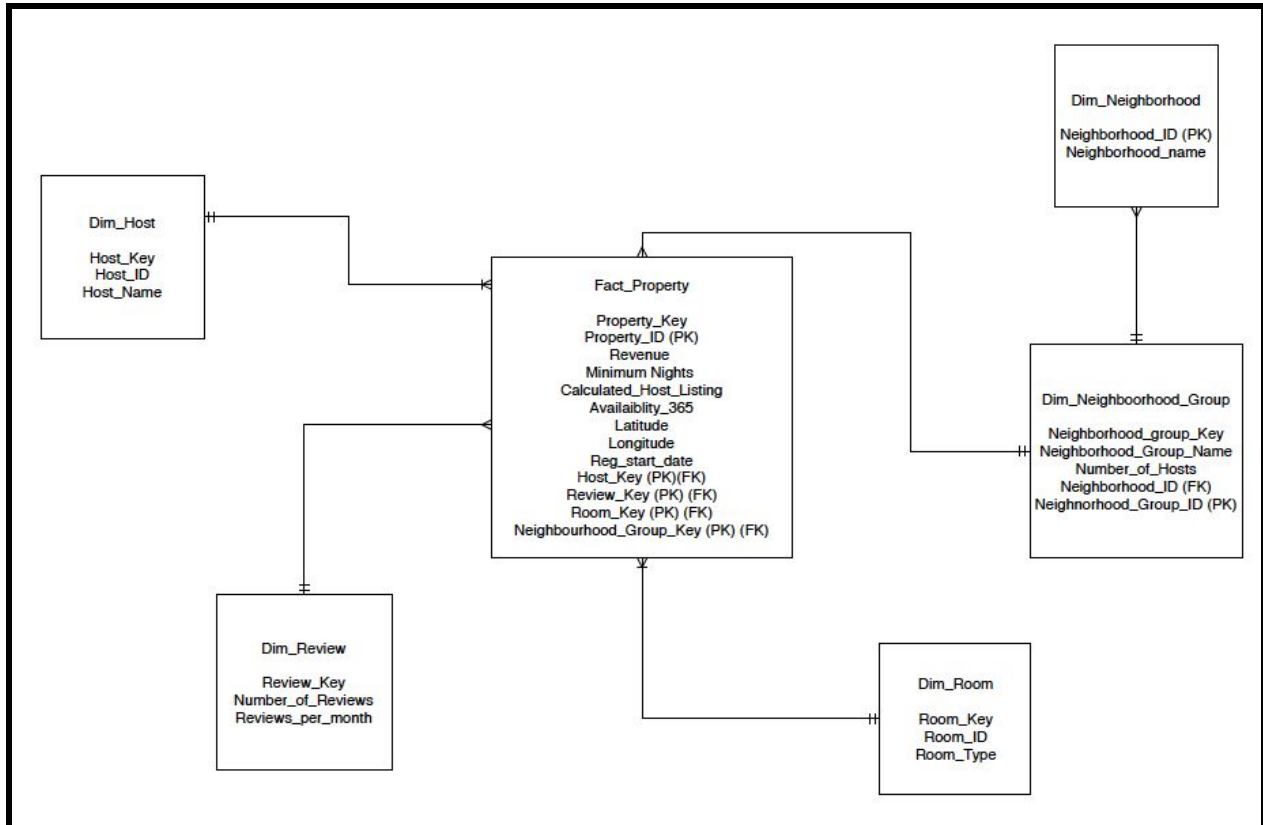
Dataset:

The dataset utilized for this project contains 16 columns and 48,900 rows. It is an open dataset from Kaggle, with original source being a website which utilizes data extracted from Airbnb's website that are public information, as well as the availability calendar for 365 days in the future, and the reviews for the listings. This data source was chosen for the project as the data from the original source was verified, cleansed, aggregated, and analyzed. The dataset also all the necessary attributes to draw conclusions from further analysis. Lastly, the data was also chosen as, due to the similarities of the team member's current area of residence, it would be interesting to draw similarities and patterns between the two cities as comparison.

Data Modelling and Designing

The snowflake schema below models data warehouse systems. The dimension tables are normalized as the data split into several dimension tables. The centralized fact table connects to the surrounding dimension tables in the multidimensional schema.

Tool used for creating SnowFlake Schema for data warehouse: Draw.io



Tools used for creating ETL transformation: Pentaho

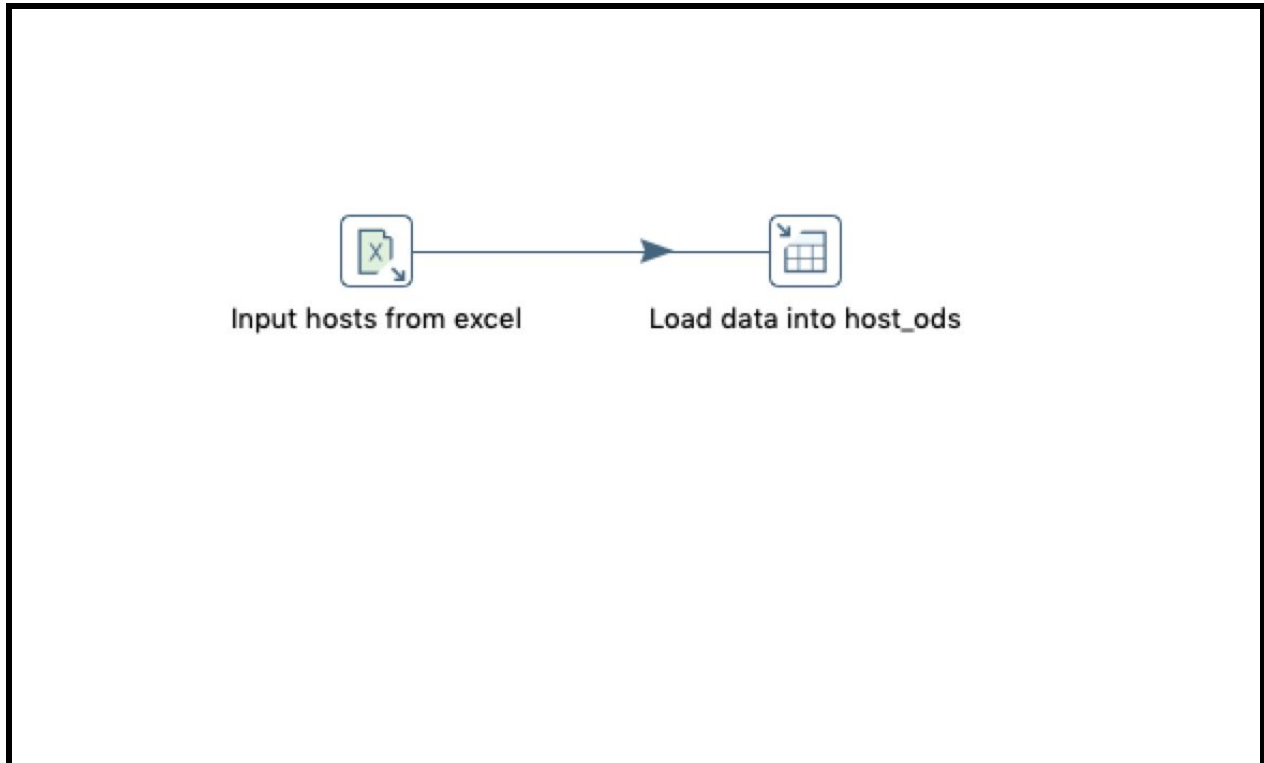
Database created in : MySql

We are using different data sources like sql, excel sheets to build the ODS transformation.

ODS Schema

Host_ods

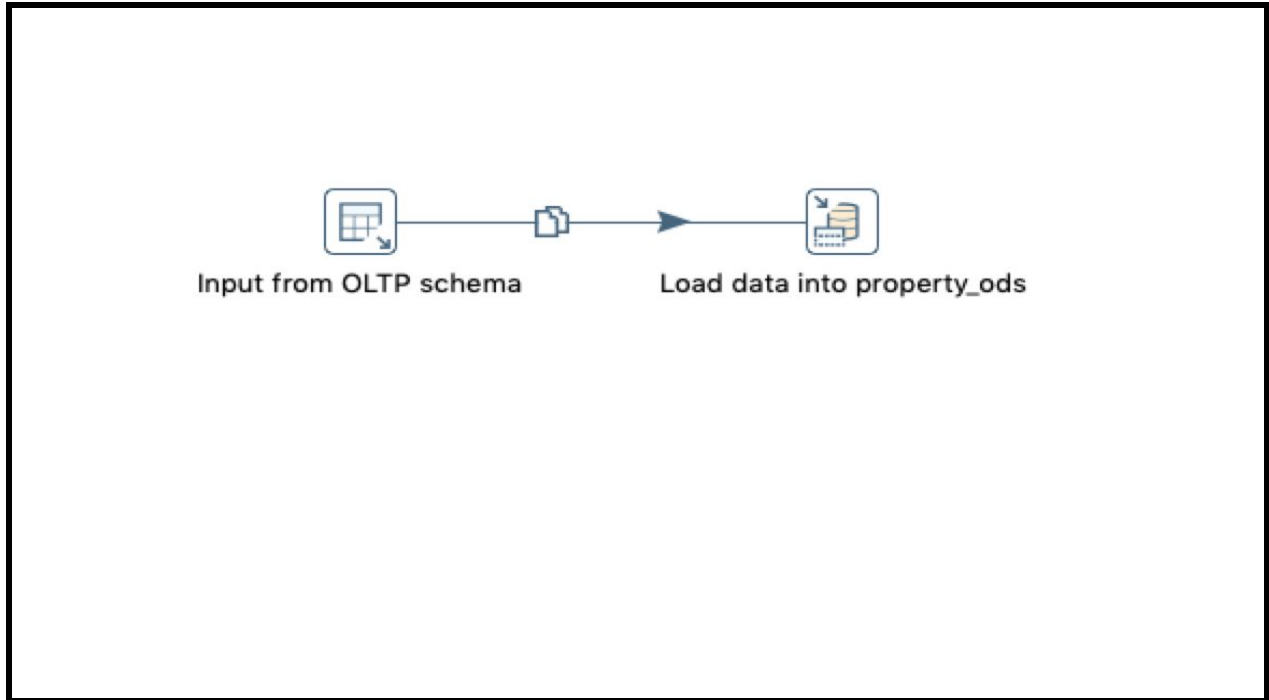
Data source of Host_ods is excel sheet. We have created pentaho transformation shown below to load data into host_ods.



Property_ods

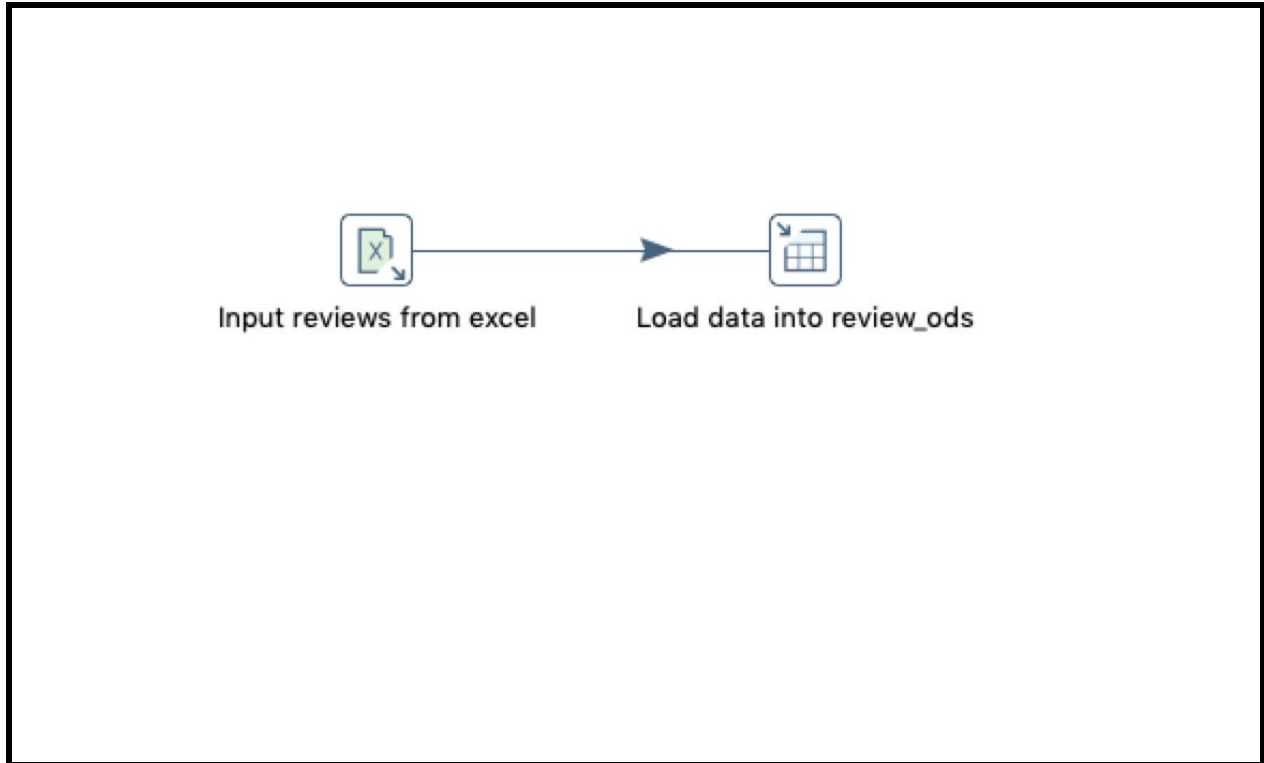
To load property_ods table in ODS schema, we are using OLTP schema as our data source.

Below is a screenshot of the pentaho transformation for property_ods.



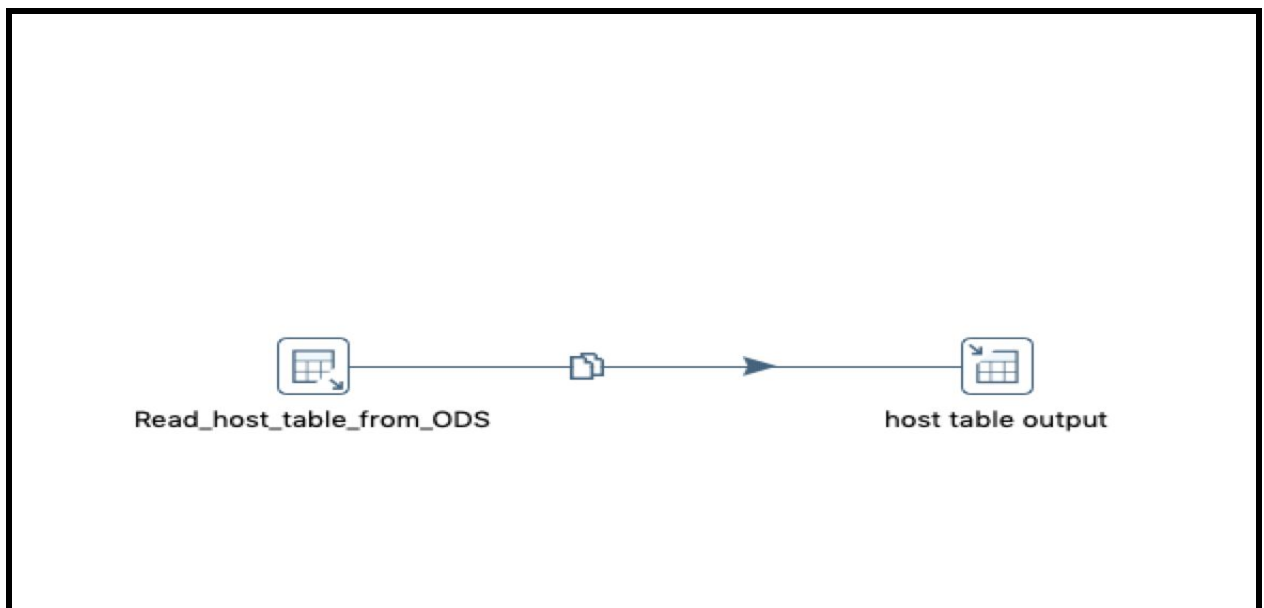
Review_ods:

Data source selected for Review_ods table is excel sheet. A transformation was created to input the data from excel sheet to ODS database.



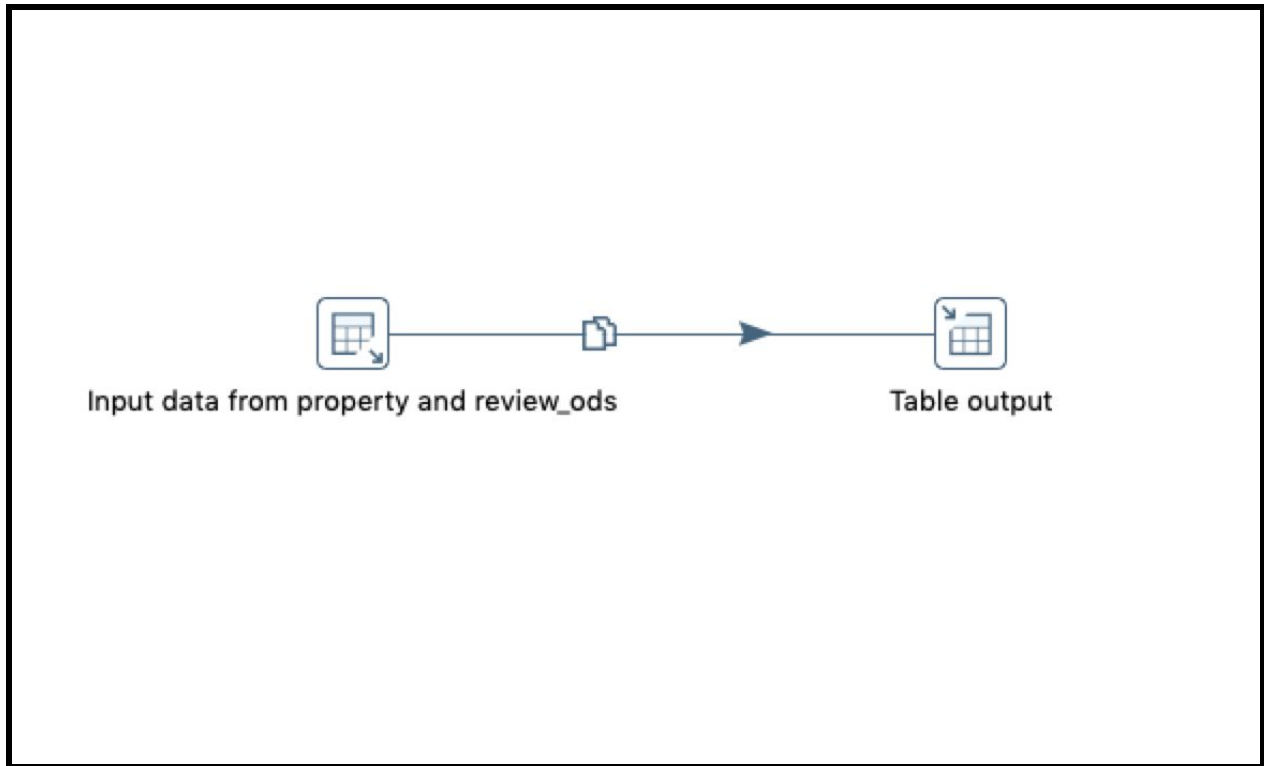
Dim Database:

Dim_host table is created using host_ods table.



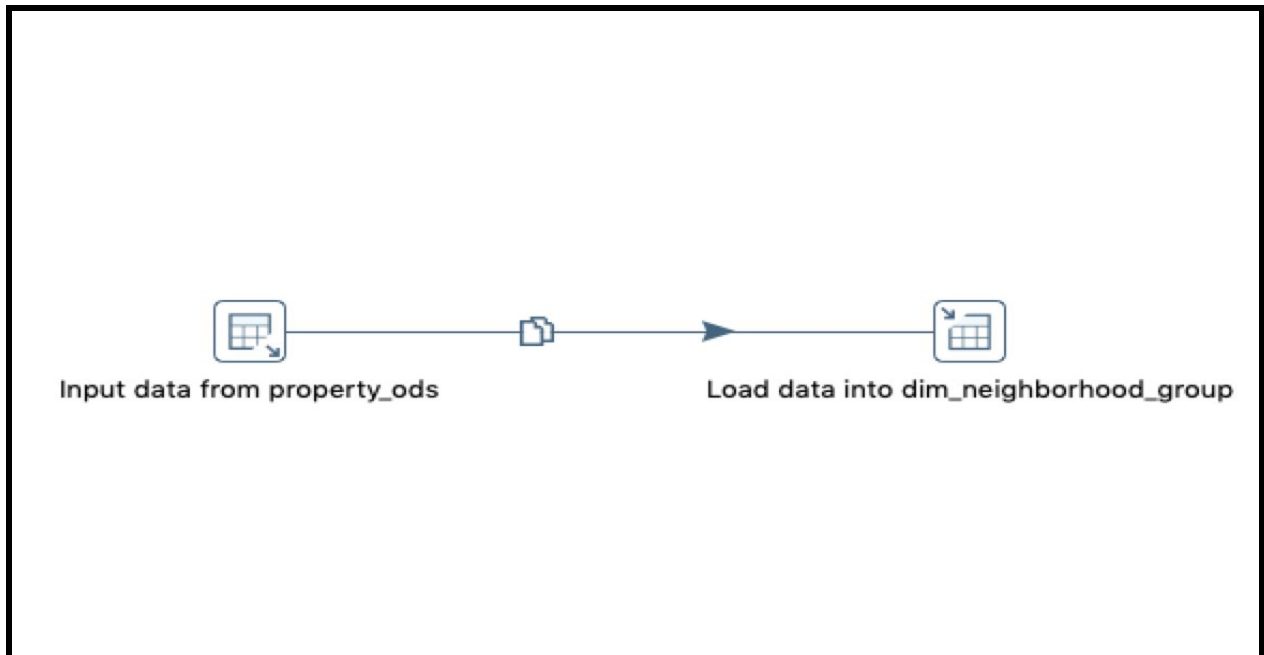
Dim_neighborhood

Dim_neighborhood table is built using review_ods table.

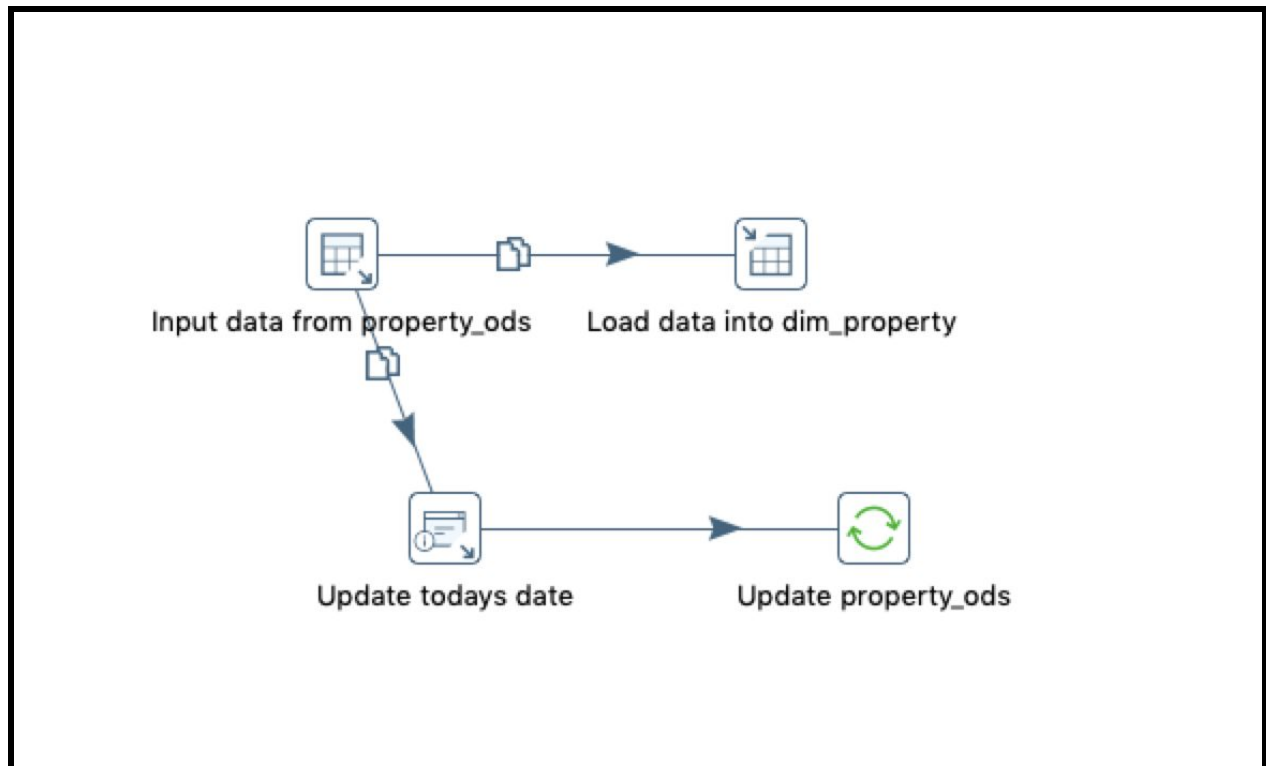


Dim_neighborhood_group

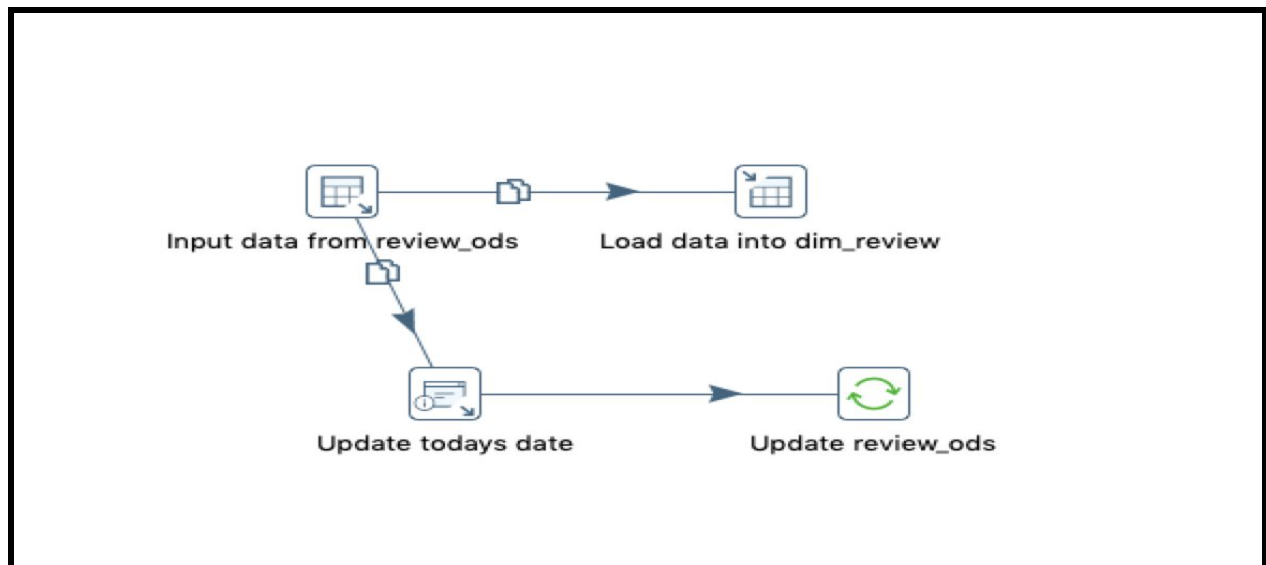
Dim_neighborhood_group is built from property_ods table



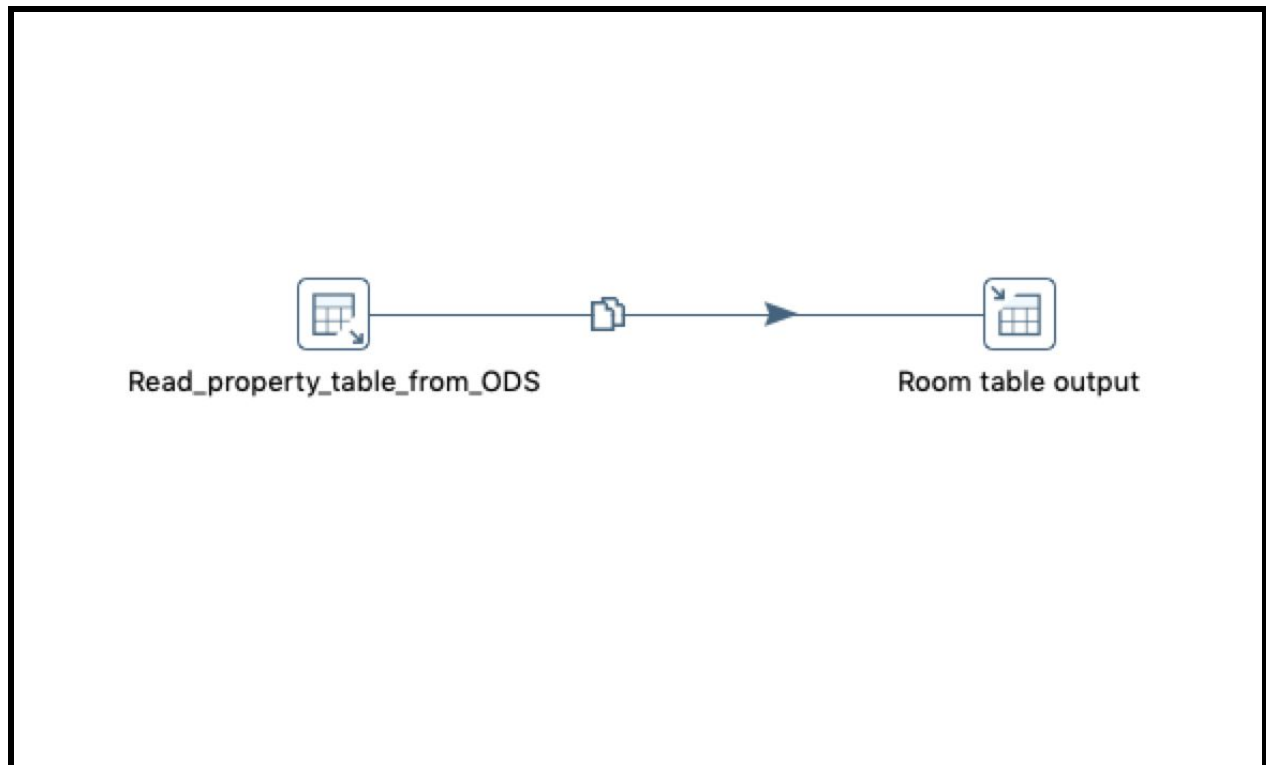
Dim_property



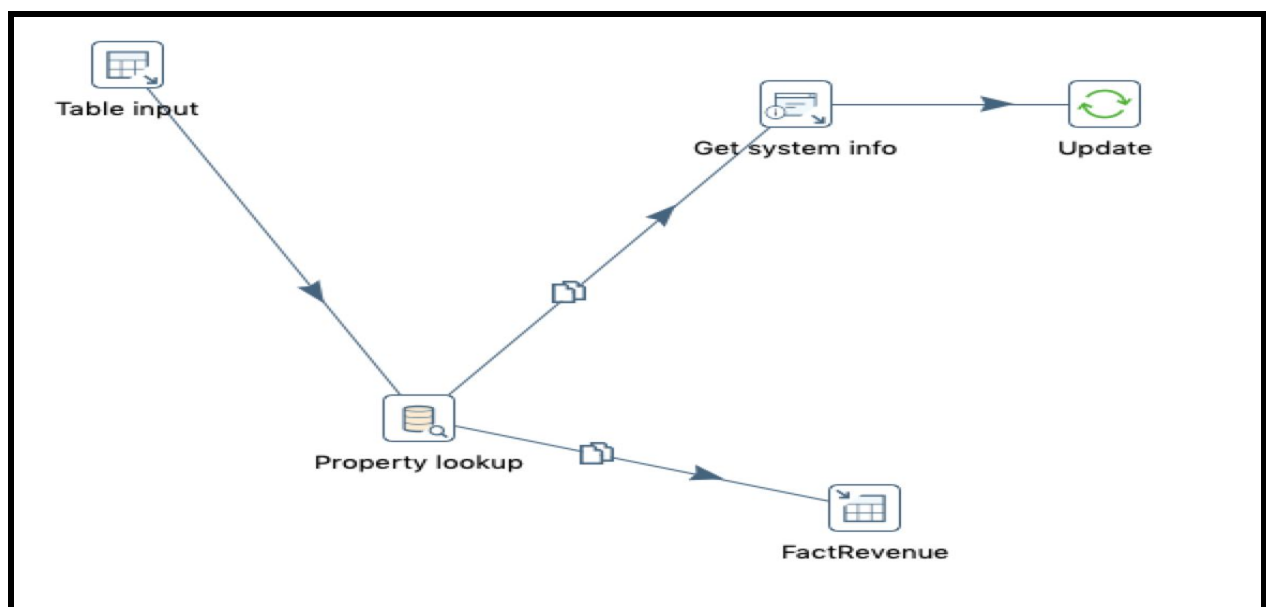
Dim_review



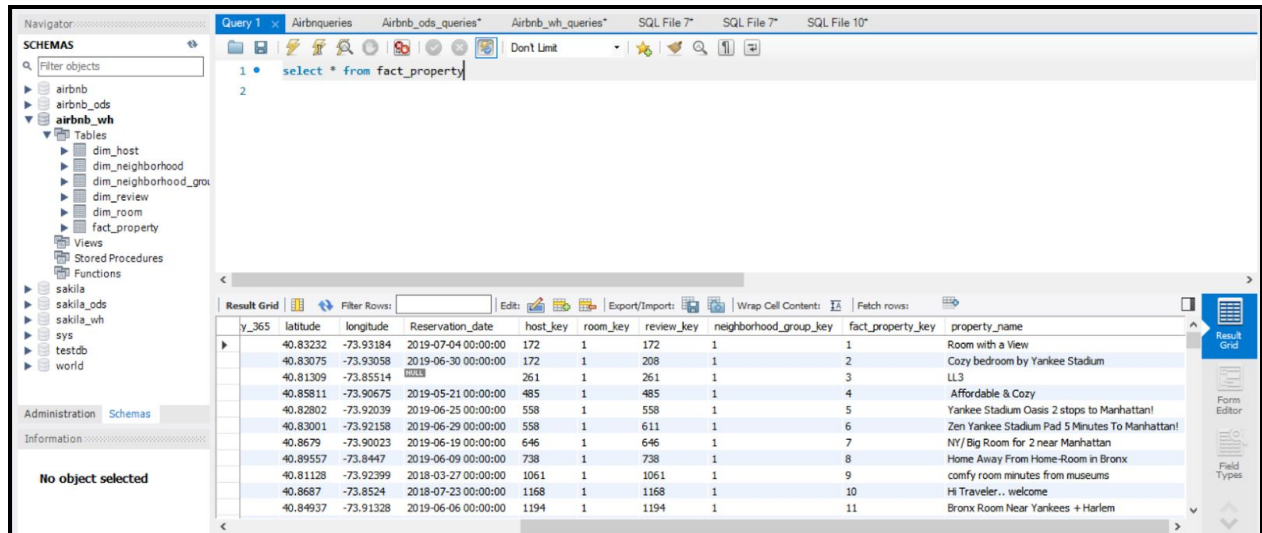
Dim_room



Fact_property



Snippet of Fact table:



The screenshot shows a database query tool interface. On the left is a 'SCHEMAS' pane with a tree view containing 'airbnb', 'airbnb_ods', and 'airbnb_wh'. Under 'airbnb_wh', there are 'Tables' (dim_host, dim_neighborhood, dim_neighborhood_group, dim_review, dim_room, fact_property) and 'Views'. The main area shows a query editor with the text 'select * from fact_property'. Below the editor is a 'Result Grid' displaying 11 rows of data. The columns are: y_365, latitude, longitude, Reservation_date, host_key, room_key, review_key, neighborhood_group_key, fact_property_key, and property_name. The data includes various property listings with their respective keys, coordinates, reservation dates, and names.

| y_365 | latitude | longitude | Reservation_date | host_key | room_key | review_key | neighborhood_group_key | fact_property_key | property_name |
|----------|-----------|---------------------|------------------|----------|----------|------------|------------------------|-------------------|--|
| 40.83232 | -73.93184 | 2019-07-04 00:00:00 | 172 | 1 | 172 | 1 | 1 | 1 | Room with a View |
| 40.83075 | -73.93058 | 2019-06-30 00:00:00 | 172 | 1 | 208 | 1 | 2 | 2 | Cozy bedroom by Yankee Stadium |
| 40.81309 | -73.85514 | 2019-05-21 00:00:00 | 261 | 1 | 261 | 1 | 3 | 3 | LL3 |
| 40.85811 | -73.90675 | 2019-06-25 00:00:00 | 485 | 1 | 485 | 1 | 4 | 4 | Affordable & Cozy |
| 40.82802 | -73.92039 | 2019-06-29 00:00:00 | 558 | 1 | 558 | 1 | 5 | 5 | Yankee Stadium Oasis 2 stops to Manhattan! |
| 40.83001 | -73.92158 | 2019-06-29 00:00:00 | 558 | 1 | 611 | 1 | 6 | 6 | Zen Yankee Stadium Pad 5 Minutes To Manhattan! |
| 40.8679 | -73.90023 | 2019-06-19 00:00:00 | 646 | 1 | 646 | 1 | 7 | 7 | NY/ Big Room for 2 near Manhattan |
| 40.89557 | -73.8447 | 2019-06-09 00:00:00 | 738 | 1 | 738 | 1 | 8 | 8 | Home Away From Home-Room in Bronx |
| 40.81128 | -73.92399 | 2018-03-27 00:00:00 | 1061 | 1 | 1061 | 1 | 9 | 9 | comfy room minutes from museums |
| 40.8687 | -73.8524 | 2018-07-23 00:00:00 | 1168 | 1 | 1168 | 1 | 10 | 10 | Hi Traveler.. welcome |
| 40.84937 | -73.91328 | 2019-06-06 00:00:00 | 1194 | 1 | 1194 | 1 | 11 | 11 | Bronx Room Near Yankees + Harlem |

Data Visualization using tableau:

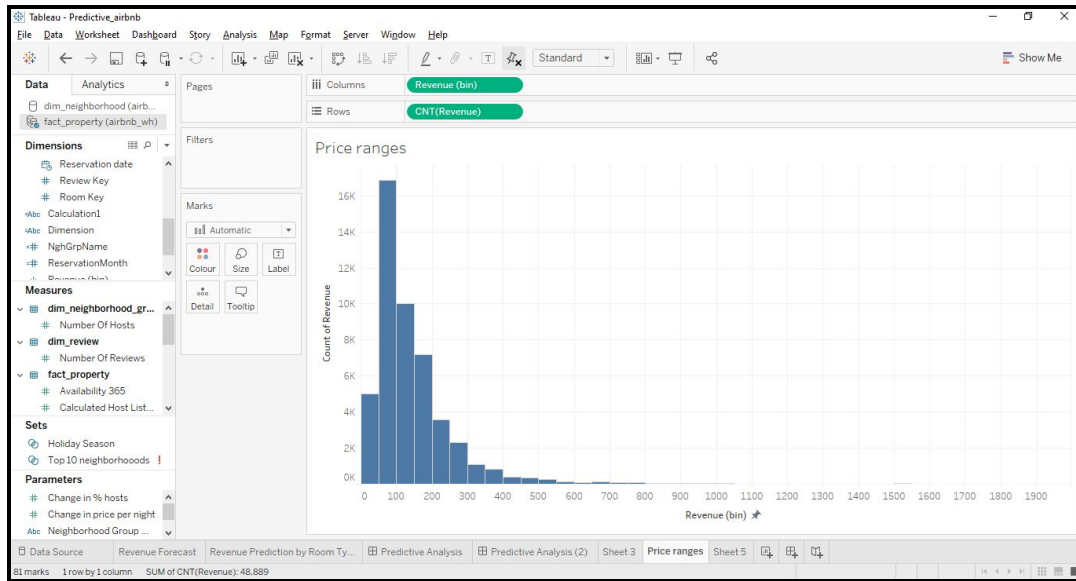
Data Assumptions:

- In the column of number of reviews, all the reviews were considered to be positive.
- Holiday season is considered for the months of May, June, July, August, December, January.
- Non-Holiday season is considered for the months of February, March, April, September, October, and November.
- The revenue is calculated based on the field 'price per night' and it is assumed that the property is occupied for 1 night
- The accuracy of the address of listing in the data or map will be from 0-450 feet of the actual address
- As listings can be deleted at any time, the data present is a snapshot of the listing at a particular time
- Some reviews may be spam, however, Airbnb claims it does not affect the statistics.
- Availability metric may be understated as as popular listings will be booked rather than being black out by host. In the dataset, there is no difference between a booked night and an unavailable night.

- **Data Insights:**

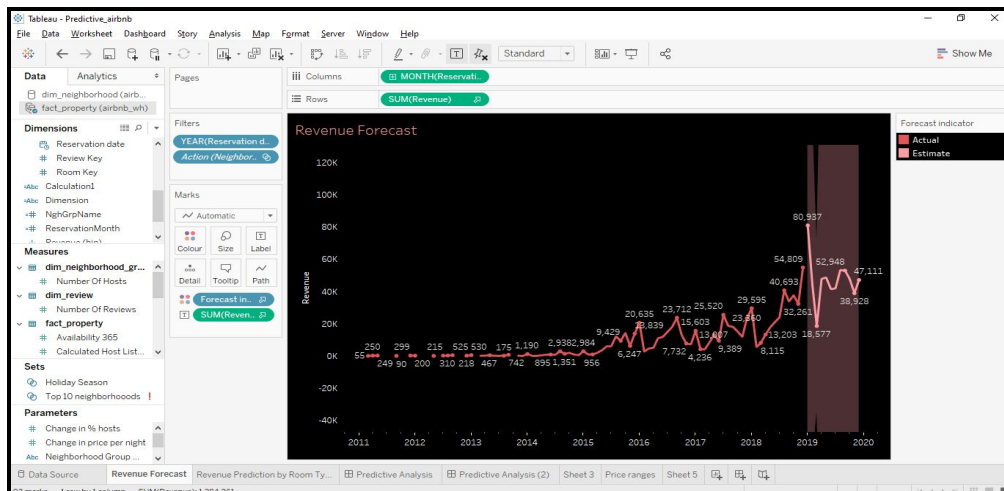
1. *What price range contains most number of Airbnb properties?*

- Most number of Airbnb properties belongs to the price range of \$50 - \$100.
- This insight enables competitive advantage to potential hosts when pricing their properties. They may choose a lower rate per night to acquire more guests reservations.

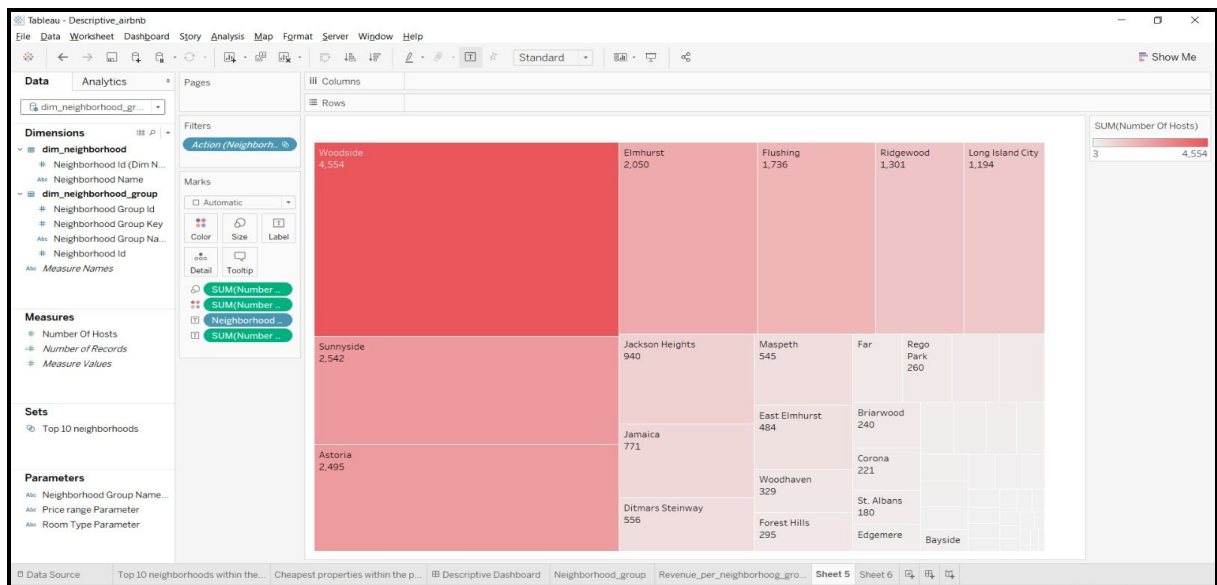


2. *What is the pattern of revenue over the period from 2011 to 2018?*

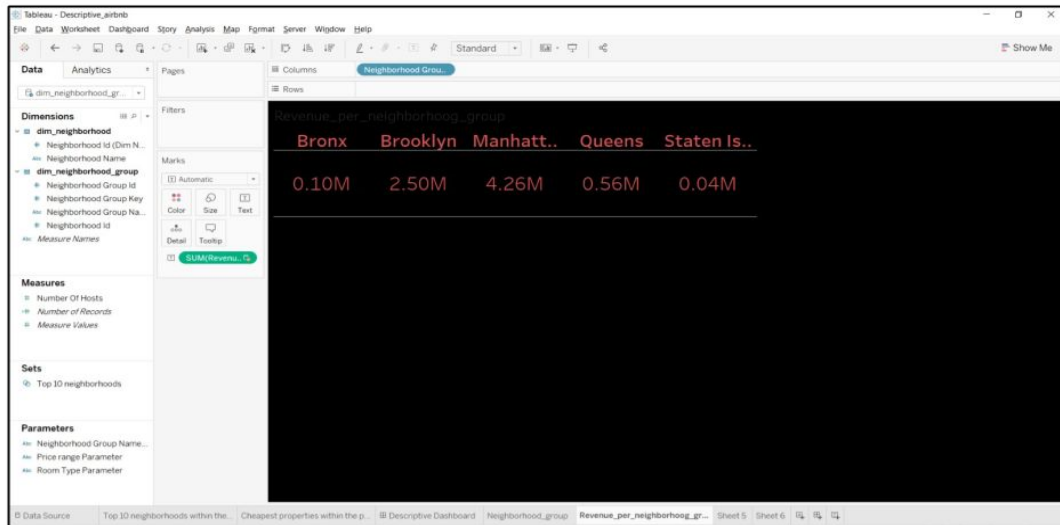
- Revenue of Airbnb kept on gradually increasing over a period of time from 2011 to 2018
- This insight could prove useful to Airbnb when presented to potential investors of the company or a potential host who may be deciding between what platform to use for the purpose of renting out their place.



3. *What are the top 10 neighborhoods wrt each neighborhood group containing the maximum number of hosts?*
 - Woodside contains most number of hosts in its area which are around 4554, next being Sunnyside with a count of 2542 hosts in its area.
 - This information will give insight to potential hosts when deciding where to avoid renting out a place due to the high competition. This could be used by airbnb to see where most of their user base in terms of hosts is coming from as well. If there are any changes in the structure(e.g big fire damaging properties or other public policies that may affect housing market) of the location such as woodside, Airbnb can derive that they may suffer potential revenue loss, or shortage of listing in that area, and reformulate their business strategies based on the information.

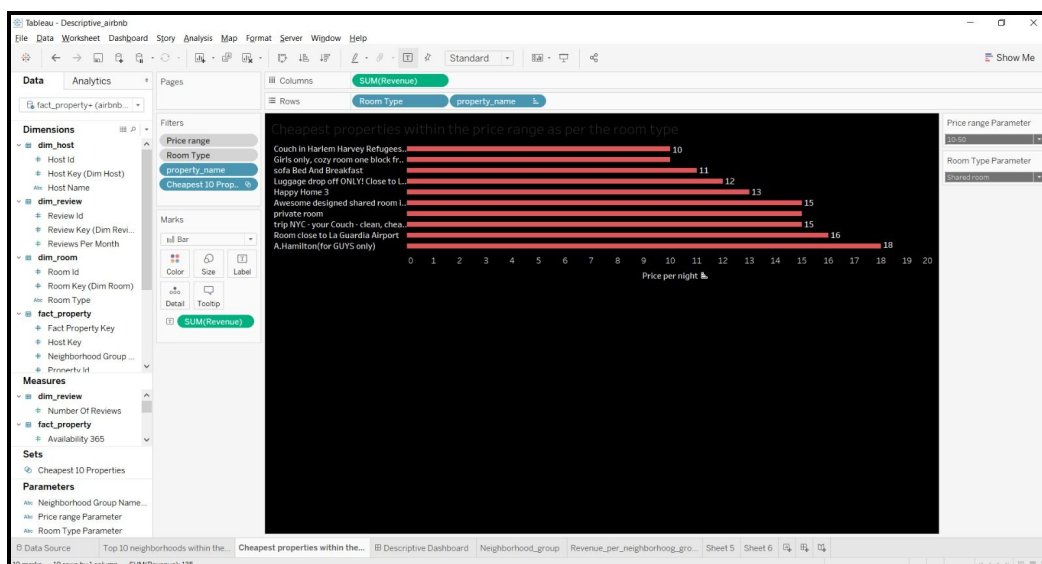


4. *Which neighborhood group grossed the highest revenue over the period from 2011 to 2018?*
 - Manhattan has the highest gross revenue collection of 4.26M
 - This allows Airbnb to know where most of their business is coming from in New York City. This information allows airbnb to prioritize the geographical area in some situations and allow them to strategize accordingly.



5. Which are the most economical properties within a particular neighborhood group in NY considering the price range and room type?

- Couch in Harlem Harvey Refugees is the most economical place.
- This information allows potential guests with budgets to make their reservations accordingly. It also allows for hosts to see their competitors' prices to allow the hosts to make decisions about the price they want to list their properties at.



Dashboard 1: Descriptive Analysis:

Descriptive Analysis: (<http://statulator.com/blog/descriptive-analysis-take-it-easy/>)

Descriptive analysis is an important first step for conducting statistical analyses. It gives you an idea of the distribution of your data, helps you detect outliers and typos, and enable you to identify associations among variables, thus preparing you for conducting further statistical analyses.

Business Value: The customer can directly use this interactive dashboard to find out the most cost effective and good rated airbnb in New York.

In the descriptive analysis of our project the Stakeholder was Customer/Host.

In the top part of the dashboard., top 10 neighborhoods were found in each neighborhood group to find out if they want to buy a place to rent out to Airbnb or rent a place to Airbnb which area contains more number of host.

In the bottom part of the dashboard cheapest properties within the price range as per the room type depicted from customer's perspective to find out the best room possible in the selected neighborhood was shown.

To summarize the overall dashboard, it allows you to find according to which neighborhood which one contains most number of host and helps the customer to find out the cheapest airbnb available.



Dashboard 2: Predictive Analysis:

Predictive Analysis:(https://en.wikipedia.org/wiki/Predictive_analytics)

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events.

In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision-making for candidate transactions.

Business Value: A future prediction of the revenue is shown for the Airbnb helping them to make or generate strategies accordingly. Along with this, a revised revenue prediction is shown envisioning that if some amount of hosts change and price per airbnb changes what the revised revenue will be.

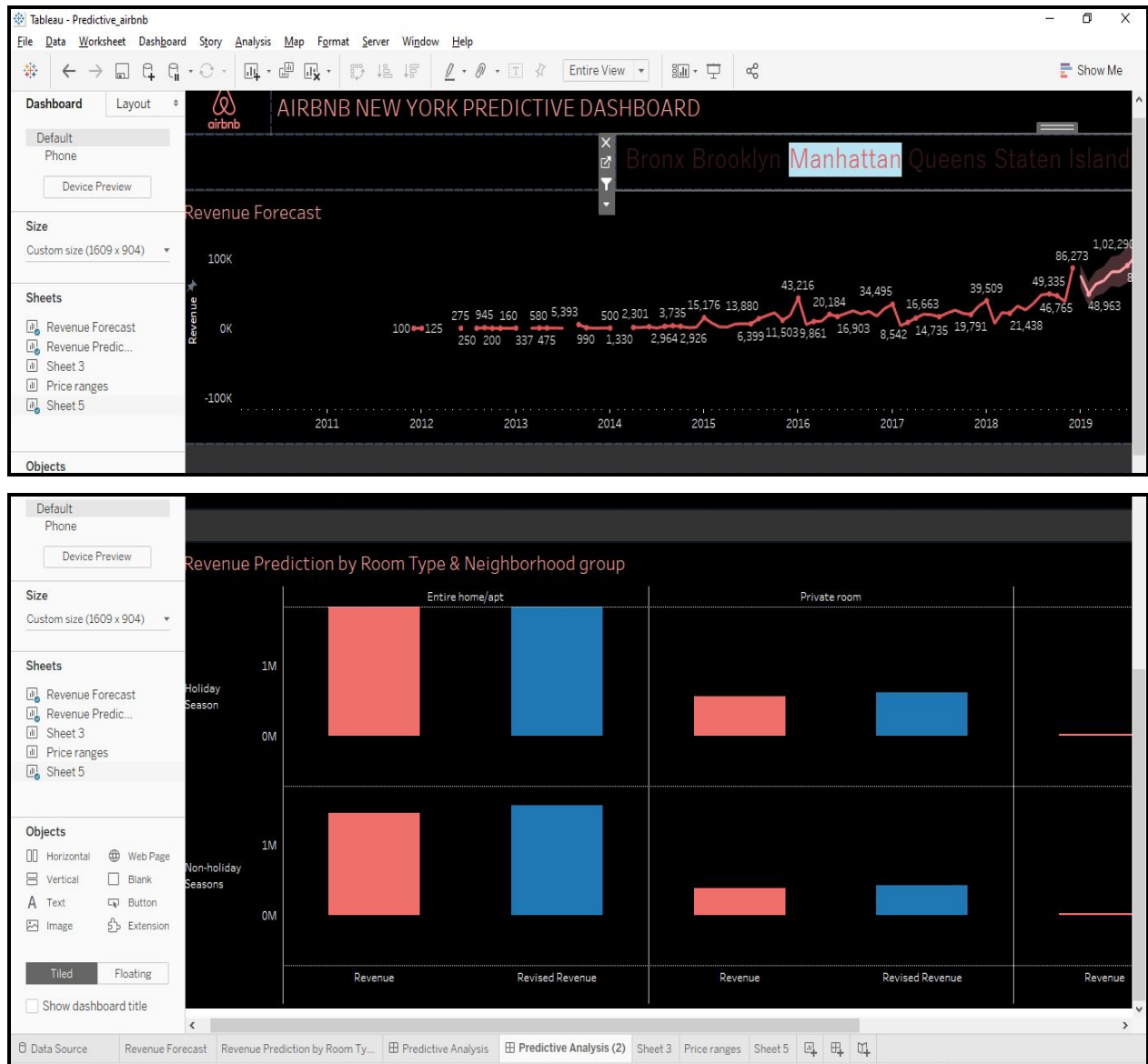
In the descriptive analysis of our project the Stakeholder was Airbnb. For all the companies around the world the revenue turnover is the most important factor. Therefore, the main idea behind doing this prediction was to find out how this business will perform in the coming year if some of the factors are changed.

Therefore, in the first upper half of the dashboard, a prediction of the revenue forecast is done for the year of 2019 by using the revenue collection data from 2011 to 2018.

Two sets were made of the data based on assumptions of the months which include holiday season and non-Holiday Seasons.

A prediction is done of the revenue actually collected and revised revenue collected by changing the percentage of host renting their property and change in price per night filtered on the type of room.

Overall to summarize this dashboard, it depicts the revenue forecast per neighborhood group and change in it based on holiday and non-holiday season with respect to the room type available.



References:

- [1] *Airbnb*, www.airbnb.com/a/?af=43720035&c=.pi0.pk25650614176_308605552029_c_12026464216&sem_position=1t1&sem_target=aud-428566875742:kwd-12026464216&location_of_interest=&location_physical=1013659&ghost. Accessed 9 Oct. 2019.
- [2] *Airbnb*, news.airbnb.com/about-us/. Accessed 9 Oct. 2019.
- [3] <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>