Methodology Document

Presentation- 1

Step 1 – Storyboarding

- Checking the data and getting familiarized with it, Noting down all the important fields
- Made a mind map of all the slides of presentation including which visualization needs to be prepared

Step 2 – Data Wrangling

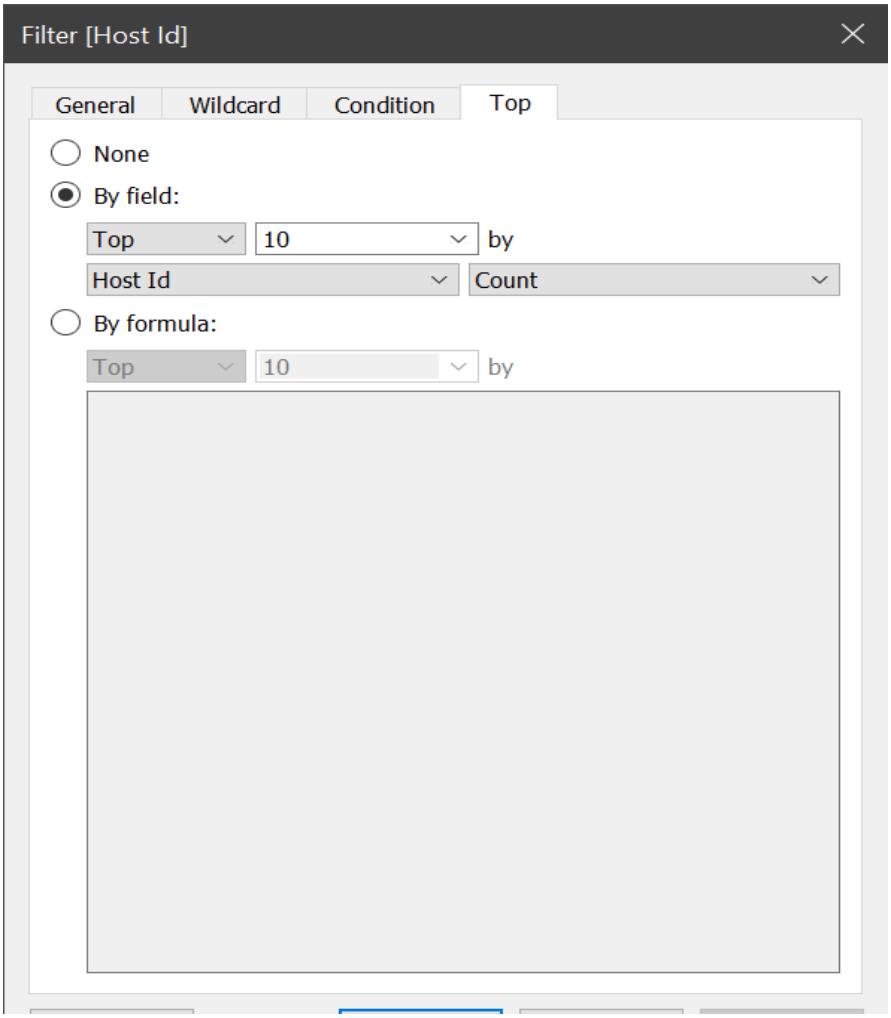
- Data Cleaning was done using python
- We imported the necessary libraries needed for analysis such as pandas and numpy
- Imported the necessary data from excel
- Checked the shape of the data, How many columns and Rows the Airbnb data contained
- Dropped 3 columns – id, name and last review as they were not required for analysis
- Checked for missing values, and filled them with 0
- Checked for duplicate rows, Data standardization, Validated data types, Inconsistent data
- Saved the data into a CSV file

Data Analysis and Visualization using Tableau

We utilized the use of Tableau to create visualizations for this case study, Below are the steps for each of the visualizations created

1. **Top 10 Hosts**

- Identified the top 10 host name along with Host ids by filtering the top 10 ten hosts by field



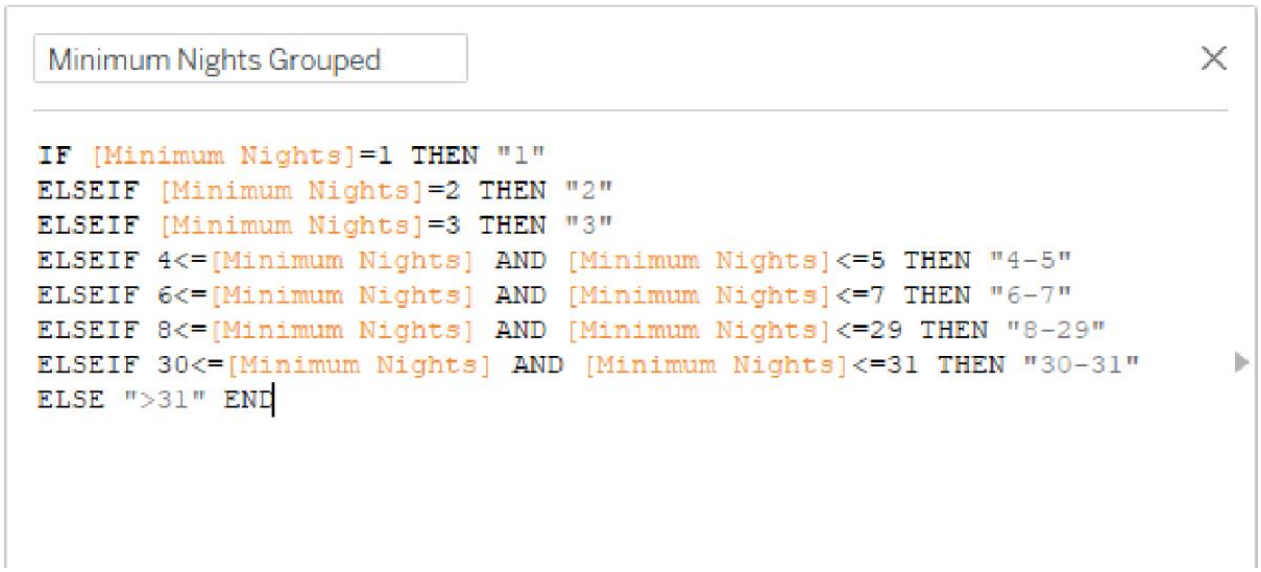
2. **Occupancy based on room type**
    - Created a pie chart in order to understand the occupancy of the customers based on their preferred room type w.r.t the neighborhood
    - Added the Room type to the colors card to highlight different room types and added the neighborhood to the count to check the count of customers
3. **Rooms type preferred by customers in different neighborhoods**
    - Performed the same visualization but now finding customers choice of room type percentage w.r.t neighborhood and creating 5 different pie charts
4. **Average Price of each neighborhood group**
    - Created a bubble chart for identifying which neighborhood group was costlier on an average

- Added the average price to the size marks label and added the neighborhood group to the colors marks label, Afterwards labelling their names to the bubble chart

5. **Average price variance w.r.t neighborhood group and room type (Map)**
    - Created a map to illustrate the price variance w.r.t neighborhood group and room type using the longitude and latitude
    - Added the Neighborhood group to color mark label and price bin Labels

Presentation- 2

6. **Neighborhood group box plot against price**
    - Optimized the use of box plots to plot a graph consisting which neighborhood were the costliest, the least costliest, their mean etc
    - Unticked Aggregate measures from the Analysis tab on the top left drop down menu and added the box plot to the workspace from analytics on the left panel to create a box plot chart
7. **Customer bookings with respect to neighborhood group & minimum nights**
    - Created a bin for minimum nights as shown in the figure below
    - The bins were used to identify customers through their ids and bin them in different categories on the basis of minimum nights for each neighborhood group



8. **Popular Neighborhood based on reviews**

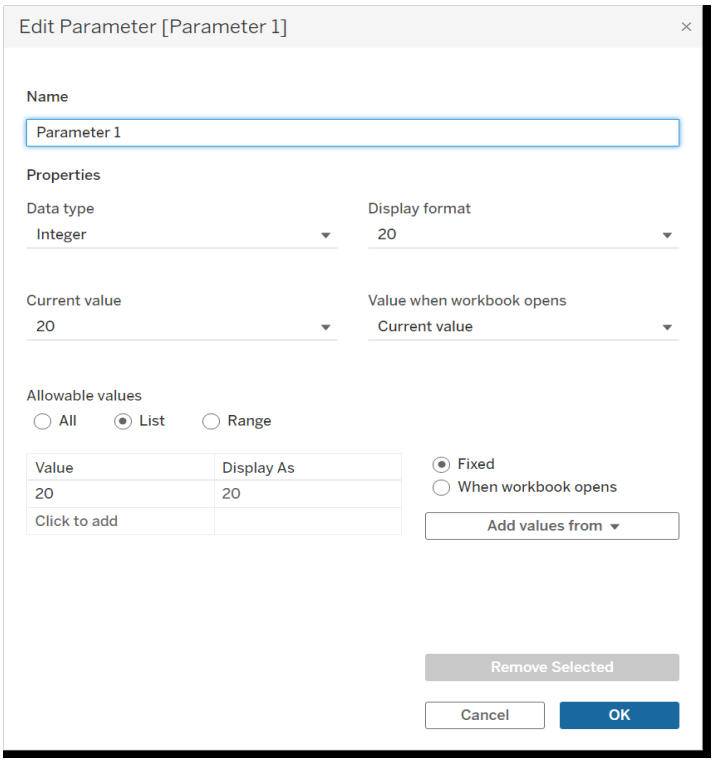
- Used neighborhood feature in rows and the sum of number of reviews to columns to plot which neighborhood had the highest reviews
- Adjusted the rows in descending order

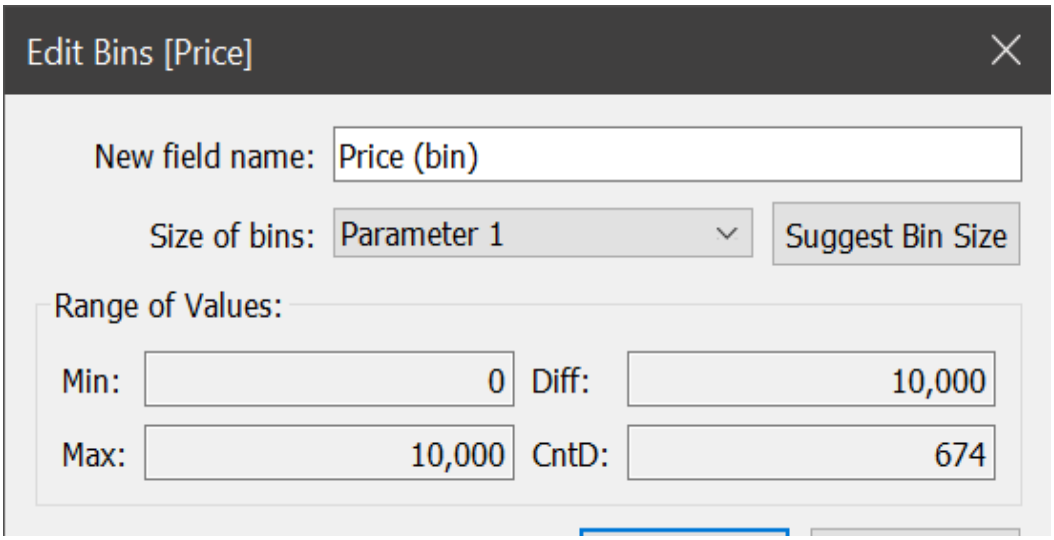
9. Availability against neighborhoods

- Created a dual axis chart to find the relationship between availability and price against the top 10 neighborhoods using filter
- Representing bar charts as availability and price as the line
- Added neighborhood to columns and sum of availability and sum of price to rows, tick marking on dual axis from drop down of price

10. Availability against neighborhood (Parameter)

- Created a bar chart by using price bin on columns and count of host ids on rows
- Also created a parameter of 20 to allow for ease of interaction with the visualization and created a price bin





11. **Average price variance w.r.t neighborhood group and room type**
    - Utilized a heat map or correlation chart to find average price variance between neighborhood group and room type
    - Added neighborhood group to columns and room type to rows and adding avg price to color mark label
12. **Top Neighborhood based on Reviews**
    - Utilized horizontal bar chart to illustrate top neighborhood based on the number of reviews each neighborhood received
    - Colored the Bars in correspondence with neighborhood groups

By

Rudwaan Vankar,

Sharadhi Kulkarni,

Vijayraj Poojary