**AIRBNB DATA ANALYSIS USING EXCEL: by: ARUNKUMAR** links**:** [github](https://github.com/arunkumarsp-ds) **|** linkedin

Introduction:

* This Airbnb data analysis project is to find the optimal property to purchase and list for rental in Airbnb marketplace.
* Along with this i need to determine the return on investment or to forecast the potential earnings.

TO DO THIS I AM USING THE FOLLOWING TOOLS ;

* Advanced Formulas & Functions.
* Advanced Formatting.
* Analytical Tools
* Pivot Tables
* charts

The steps followed are:

* Structure of the dataset
* Cleaning and Processing
* Analysing and Sharing our analysis through Visualizations.
* Findings of my Analysis in conclusion.

Structure of the dataset:

In this Airbnb dataset, we have 7 Worksheets namely;

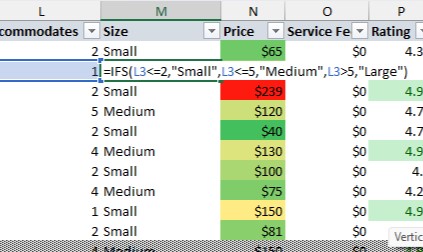
* Places -> contains Airbnb listing data Place name , host\_id,Neighbourhood\_id , room\_type, rating , pricing etc;
* Hosts-> contains Airbnb host data including name, id, email, country, date\_joined,response\_time, response \_rate, etc;
* Neighborhoods -> contains districts in New york and neighborhoods in each district.
* Mortgage calculator-> it contains a ready-made template that includes the property purchase price, tax rate, and monthly expenses.
* 2019\_occupation -> A binary Calendar for 2019 which tells that 1 for booked and 0 for not booked on that day.
* Amortization schedule -> Amortization schedule can be done by using the property used in mortgage calculator.
* Visualizations-> Annual Pro it and Loss data and airbnb conversion funnel

cleaning and Processing the data:

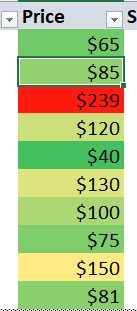
1. PLACES WORKSHEET:

Inserting a new column size by using the

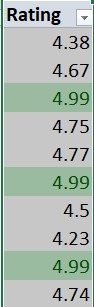
IFS function =IFS(L2<=2,"Small",L2<=5,"Medium",L2>5,"Large")



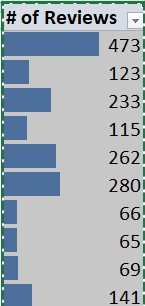
* + conditional formatting is applied for pricing with color scales



* + gradient ill is applied for above average rating by conditional formatting rule .

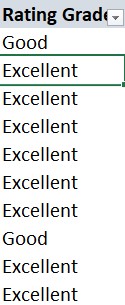


* + Data bar is applied for reviews by conditional formatting.



* + Switch function is used to create a rating grade column.

=SWITCH(ROUND(P2,0),5,"Excellent",4,"Good",3,"OK","Bad")



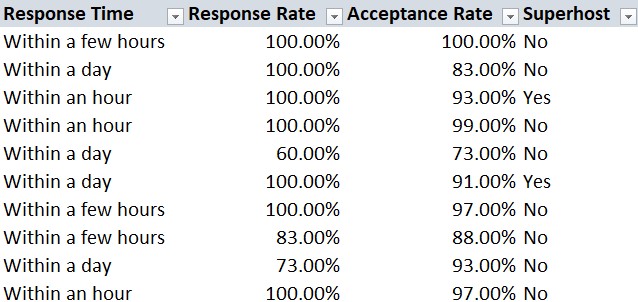
* + response time, response rate, acceptance rate and super host columns are populated from the Host worksheet by using the VLOOKUP function coupled with MATCH function by using host\_id to utilize the referencing.
  + so this means we don't want to change the formulas separately which will be taken care of by the MATCH function by just dragging our formula to another column.

Response\_time: =VLOOKUP(Places!$C2,Hosts!$A:$P,MATCH(D$1,Hosts!$1:$1,0),0)

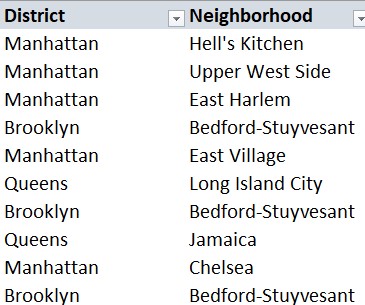
Response\_rate: =VLOOKUP(Places!$C2,Hosts!$A:$P,MATCH(E$1,Hosts!$1:$1,0),0)

Acceptance\_rate: =VLOOKUP(Places!$C2,Hosts!$A:$P,MATCH(F$1,Hosts!$1:$1,0),0)

Superhost: =VLOOKUP(Places!$C2,Hosts!$A:$P,MATCH(G$1,Hosts!$1:$1,0),0)



* + District and neighbourhood columns are populated from Neighbourhood worksheet by using INDEX AND MATCH function combined together by using neighbour id.



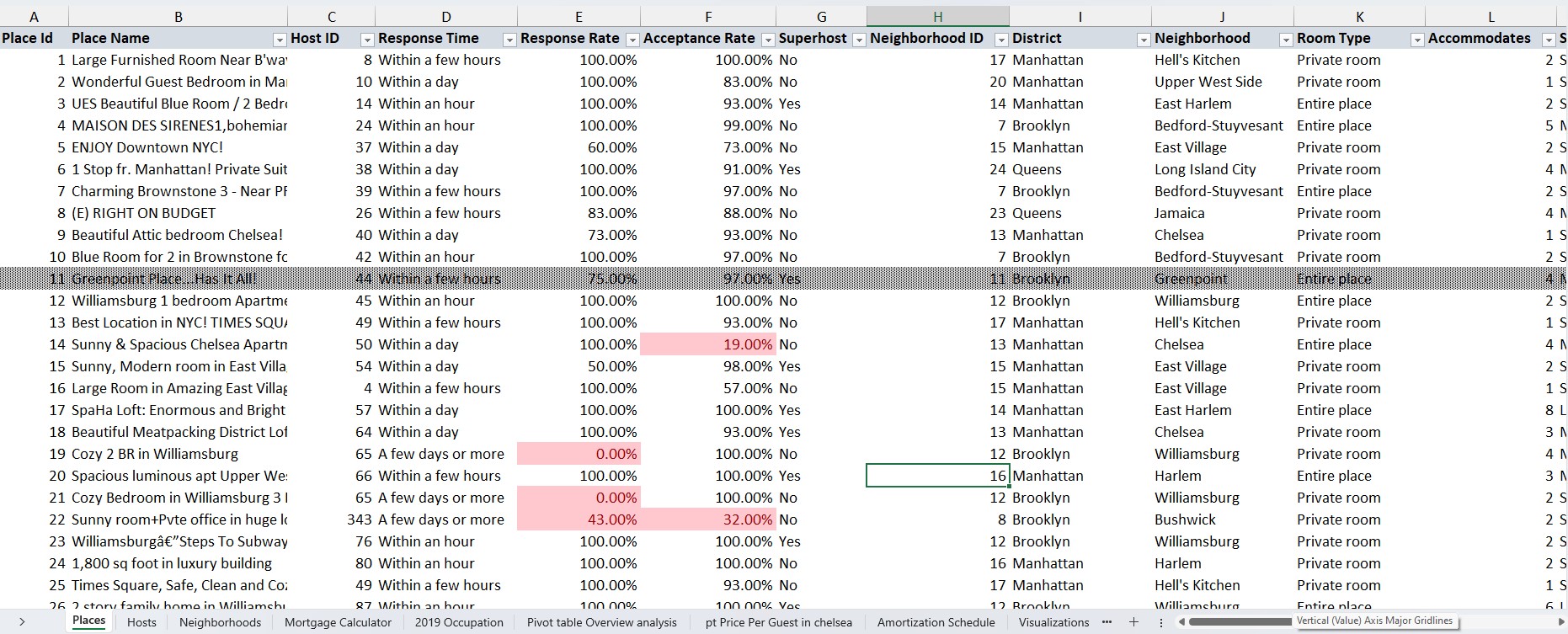
Neighbourhood: =INDEX(Neighborhoods!$A:$C,MATCH(Places!$H2,Neighborhoods!$A:$A,0),MATCH(Places!J$1,Neighborhoods!$1:$1,0))

District: =INDEX(Neighborhoods!$A:$C,MATCH(Places!$H2,Neighborhoods!$A:$A,0),MATCH(Places!I$1,Neighborhoods!$1:$1,0))

By Using conditional formatting on :

* + Red fill is used on the response rate and acceptance rate columns when it is below 50 %

● Grey fill is used in the rows when the place status is inactive.



1. HOSTS WORKSHEET:

* Date column is created by applying date function coupled with text extract functions (LEFT, MID, RIGHT) on date joined column.

=DATE(LEFT(D2,4),MID(D2,5,2),RIGHT(D2,2))

* Year\_Joined column is created using Year function applying on date column.

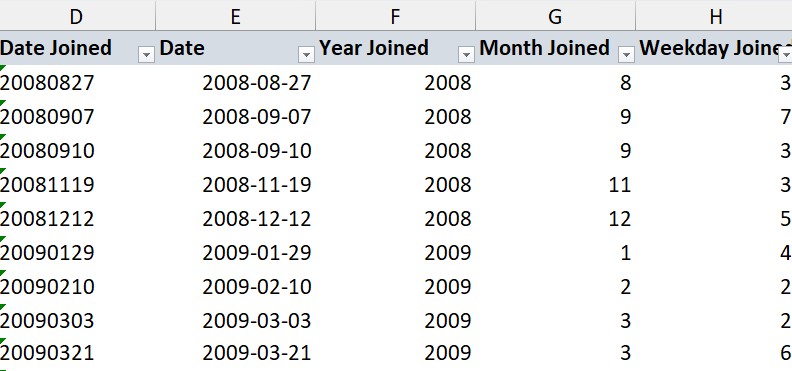
=YEAR(E2)

* Month\_joined column is created by using Month function on date column.

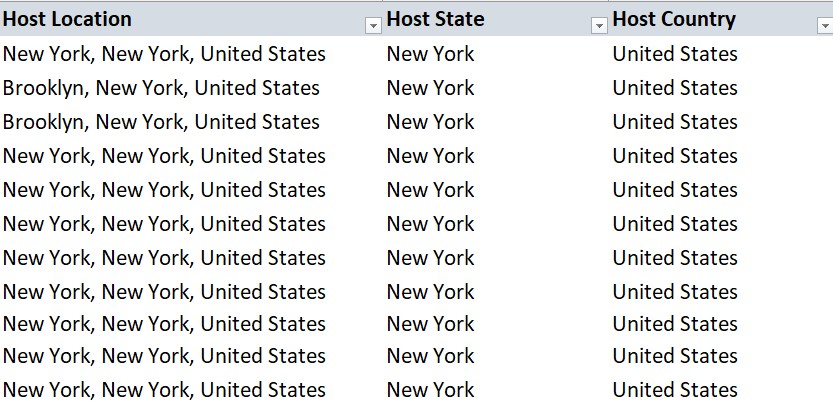
=MONTH(E2)

* Weekday joined column is created using Weekday function on date column.

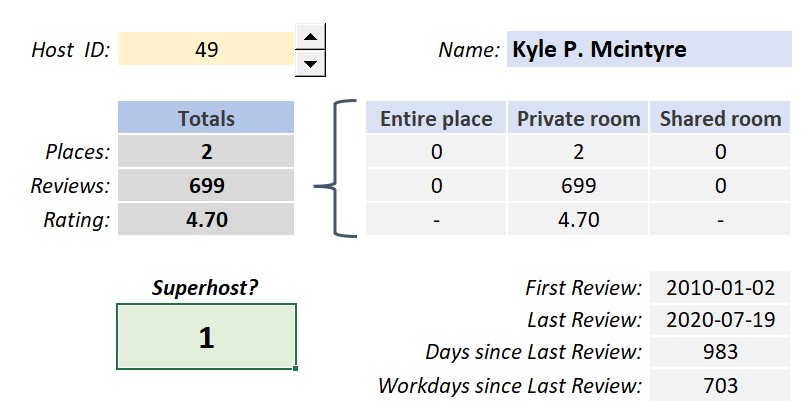
=WEEKDAY(E2,2)



* Host state and host country columns are created by using Flash Fill function which will recogonize the pattern in host location column .To recogonize it we have to to type 2 or more rows in the respective columns .
* This can also be done using Text to columns function by using the delimiters present in it ( here , eg : comma)



LEFT SIDE: RIGHT SIDE:



LEFT SIDE:

Totals

PLACES : =COUNTIF( Places!C:C,Hosts!R2)

Reviews: =SUMIF(Places!C:C,Hosts!R2,Places!R:R)

Ratings : =SUMIF(Places!C:C,Hosts!R2,Places!R:R)

superhost: =VLOOKUP($R$2,A:O,13,0)

RIGHT SIDE:

Entire Place

Places : =COUNTIFS(Places!$C:$C,Hosts!$R$2,Places!$K:$K,Hosts!T4)

Reviews : =SUMIFS(Places!$R:$R,Places!$C:$C,Hosts!$R$2,Places!$K:$K,T4)

Ratings: =IFERROR(AVERAGEIFS(Places!$P:$P,Places!$C:$C,Hosts!$R$2,Places!$K:$K,T4),"-")

ON LEFT SIDE IN ABOVE HOST DASHBOARD, I USED THE ;

* COUNTIF function in places row to count the no of places.
* SUMIF function in reviews to sum up the reviews .
* AVERAGEIF function in Ratings to get the average reviews. ON RIGHT SIDE IN ABOVE HOST DASHBOARD, I USED THE

Entire place:

* COUNTIFS function in places row to count the no of places with 2 criteria (Host id , room type)
* SUMIFS function in reviews to sum up the reviews with 2 criteria (Host id , room type)
* AVERAGEIFS function in Ratings to get the average reviews with 2 criteria (Host id , room type)
* I actually attached the formula for only entire place but it can be applied for other room types just tweaking the formula by changing the roomtype according to the room type will get the desired results of the respective column. (Private and shared)

First review: =MINIFS(Places!S:S,Places!C:C,Hosts!R2)

* MINIFS will return the first review date

Last review: =MAXIFS(Places!T:T, Places!C:C,Hosts!R2)

* MAXIFS function will return the last review date

days since last review: =TODAY()-V10

* TODAY function is used to get the today's date which is current date, by subracting with the last date of review we can get the above one.

workdays since last review : =NETWORKDAYS(V10,TODAY())

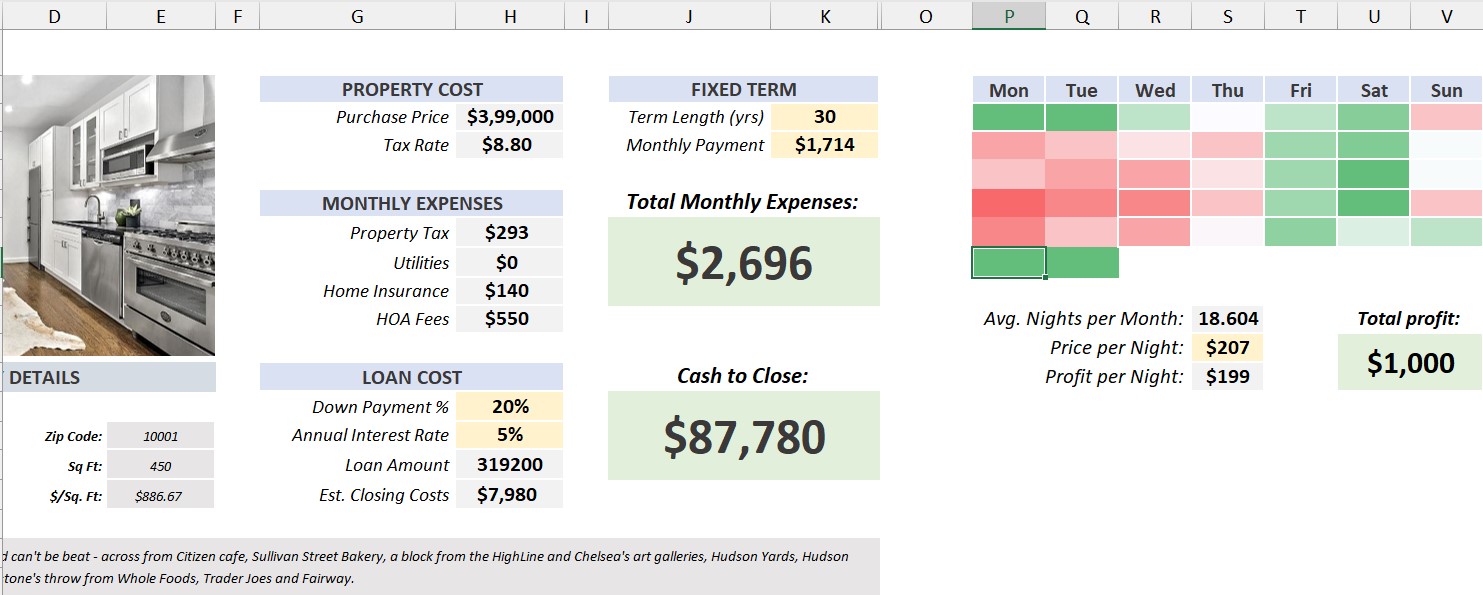
* This is very similar to above one but this will exclude the weekends and public holidays.

1. NEIGHBOURHOODS WORKSHEET:
   * In neighbourhood worksheet i have added 0nly two columns extra where it tells the no of places in the respective neighbourhood where i used the COUNTIF function.
   * To use the COUNTIF function i have assigned a count of 1 value to all the neighbourhoods.

No of places : =COUNTIF(Places!$H:$H,Neighborhoods!$A2)



1. MORTGAGE CALCULATOR:



* + Data validation is applied on both term length (years ) in Fixed payments and Fixed term column.
  + data validation is also applied on down payment %.
  + PMT function is used on the Fixed term column to get the periodic monthly payment

=PMT(H14/12,K3\*12,H15)\*-1

* + NPER function is applied on Fixed payments column to get the no of periods for the respective payment we can use anyone of them depending on the constraints.

=NPER(H14/12,N4\*-1,H15)/12

* + so grouped the Fixed payment to column which will minimize it.
  + On the right side the calendar is done by using consolidate function to get the average values of the year per day which tells

the average occupancy per day.

* + And it is coloured using conditional formatting of 3 color grade scale option .
  + To make it invisible without hiding i have used the custom number format (;;;)
  + Average nights per month is calculated using average occupancy \* total days in month .

Avg nights per month : =AVERAGE(P3:V8)\*30.5

* + at First Price per night is just entered randomly to proceed the calculation and after we calculated the total pro it per month we can use the what if analysis goal seek method to get the desired pro it and get the price per night.
  + Pro it per night is calculated by using VLOOKUP function to subract the service fee from price per night.

=S11\*(1-VLOOKUP(S11,Places!Z2:AA5,2))

1. 2019 OCCUPATION :
   * In this sheet i have just used the format painter to copy the format from the previous mortgage calendar and applied it other than this nothing i have done it.



1. AMORTIZATION SCHEDULE:
   * Year column is populated using the ill series function .
   * Principal column is populated using the SUMIF function.

=SUMIF( $B:$B,$M2,C:C)

* + Interest Column is populated using the SUMIF function.

=SUMIF( $B:$B,$M2,D:D)

* + Balance column is created using the MINIFS function.

=MINIFS(H:H,B:B,M2)

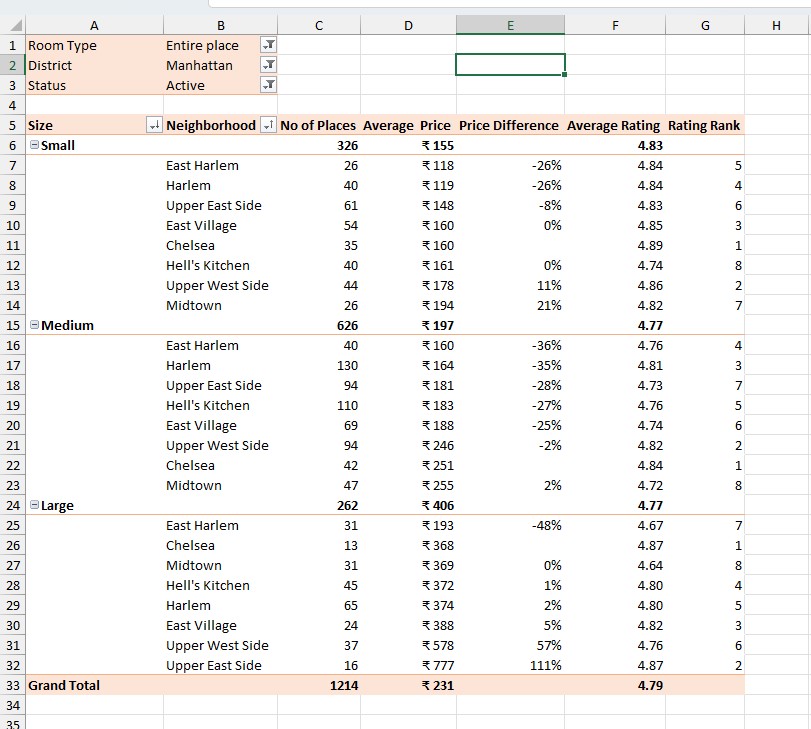


Analysing and sharing or visualizing the data: G) PT (PIVOT\_TABLE)-OVERVIEW ANALYSIS:

* + I have used the PIVOT-TABLE for analysis which is dynamic and efficient.
  + I have chosen the status, district, and room\_type as filters in pivot table.
  + And used the size of the room and neighborhood as the rows in pivot table.
  + And inserted the following ;

1. The places column to summarise as COUNT in values field.
2. The price and rating columns to summarise as AVERAGE in the values field.
3. Rating rank column to show as rank from larger to smallest based on the neighborhood field in values field .
4. Price difference column to show as percentage difference from Chelsea neighbourhood price in values field.

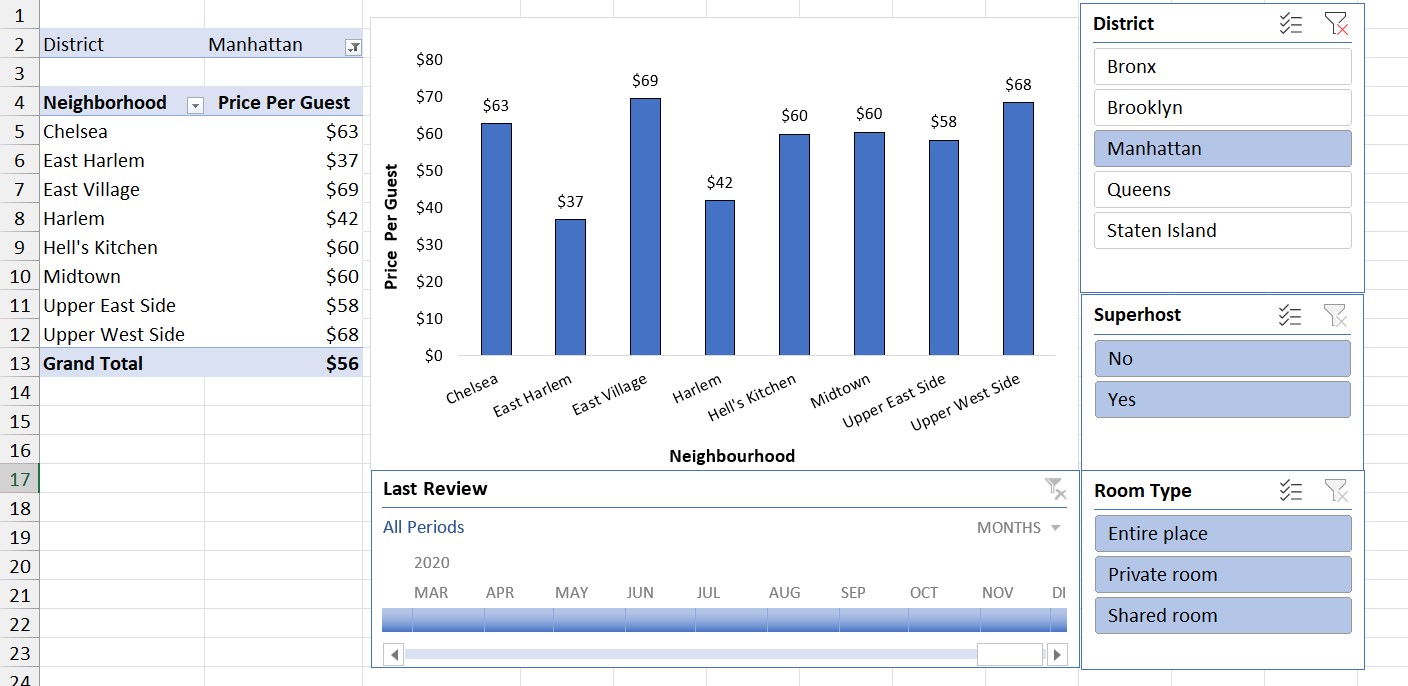
● And atlast selected the report layout to show in a outline form from design tab



H) PT (PIVOT\_TABLE) PRICE PER GUEST IN MANHATTAN NEIGHBOURHOOD:

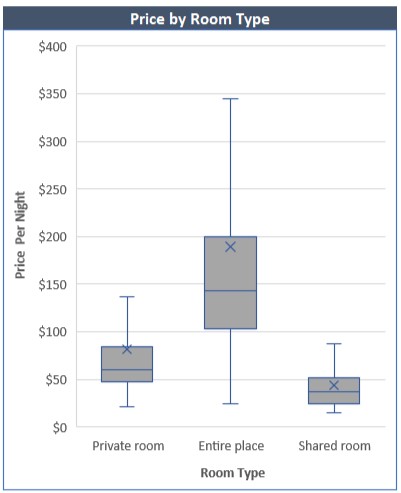
Inserted the following ;

1. district into filter field in pivot table.
2. neighbourhood into rows field in pivot table
   * created the calculated field for the price per guest column in pivot table by using the formula =price/ accommodates
   * used Pivot chart for creating the bar chart for price per guest with respect to district and neighbourhood.
   * also used slicers for choosing our desired ones for superhost, district, room type .
   * timeline used for choosing the last review of the respective property.



I) VISUALIZATIONS:

1. BOX AND WHISKER PLOT:

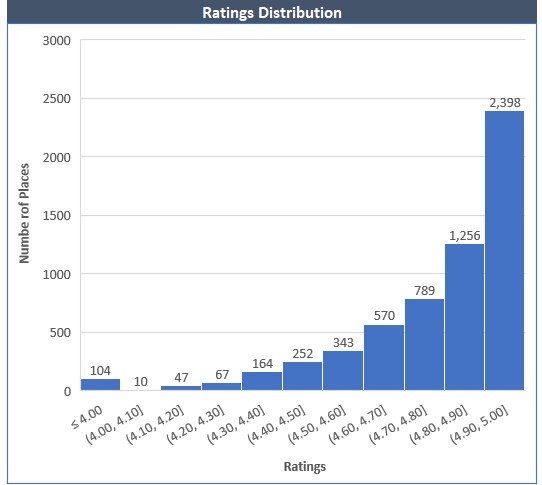


* + Created a Box and whisker plot the price per night with respect to the room type , which will give the percentiles , mean

and the distribution of the pricing .

1. BAR CHART:

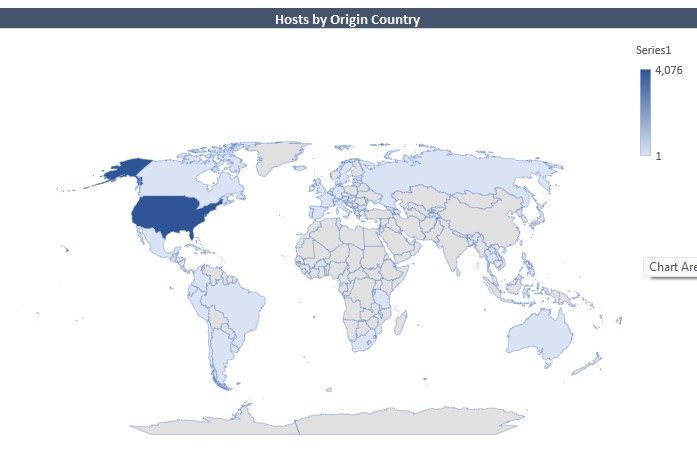
**AIRBNB DATA ANALYSIS USING EXCEL:**



* + Rating distribution visual is created using the bar chart based on the ratings and no of places falls on the respective

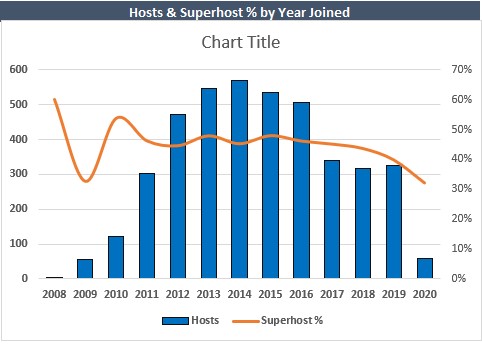
ratings.

1. MAP CHART:



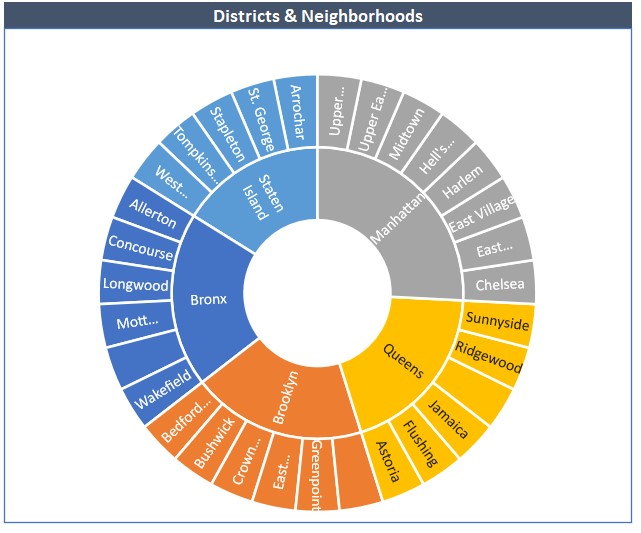
* + Hosts by country is visualized by using or inserting the map chart and make sure to convert the data into geographical.

1. COMBO CHART (LINE AND BAR) :



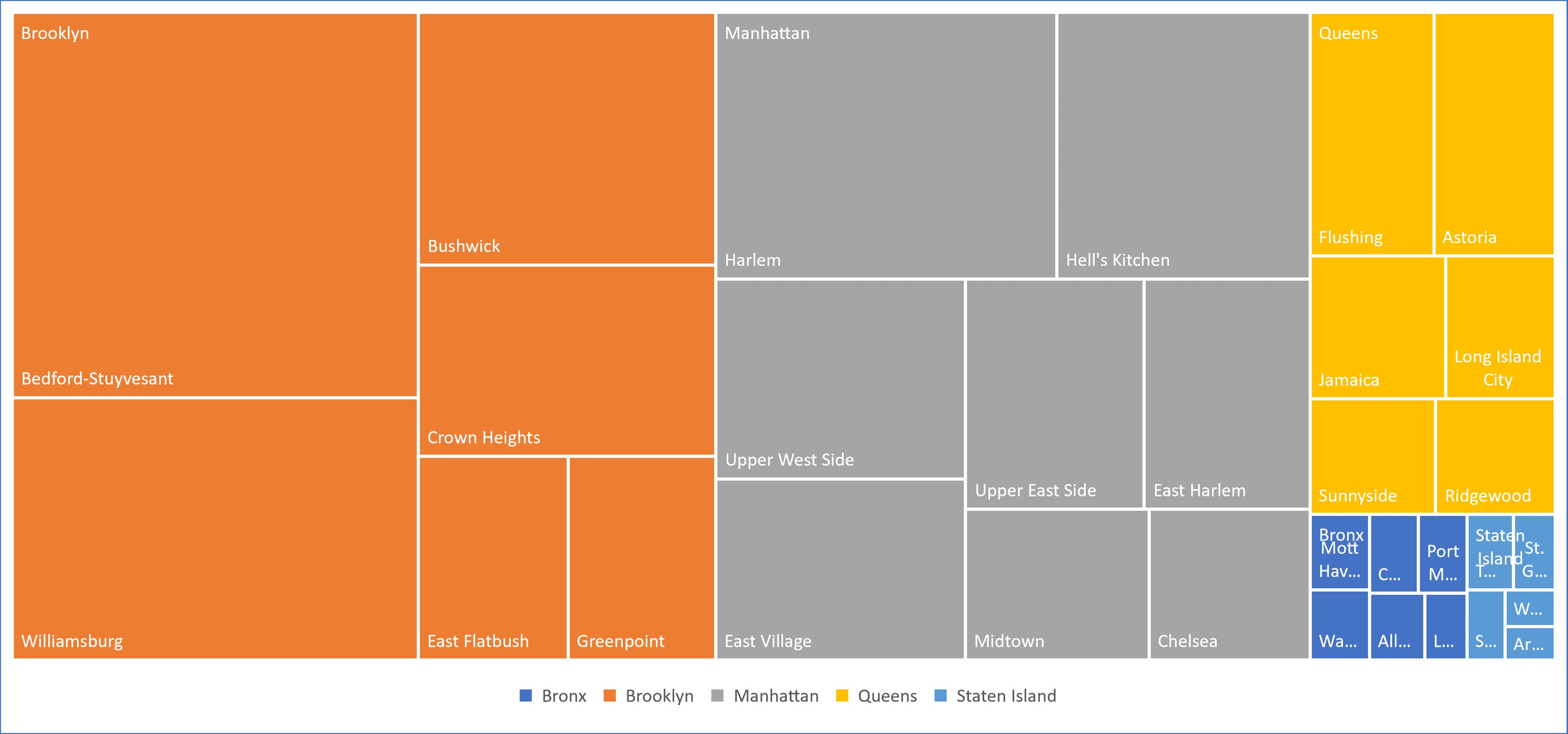
* + Distribution of host to the superhost percentage is plotted using the combo chart of bar graph and line chart combined with respect to the years and no of hosts or superhosts in it.

1. SUNBURST CHART:



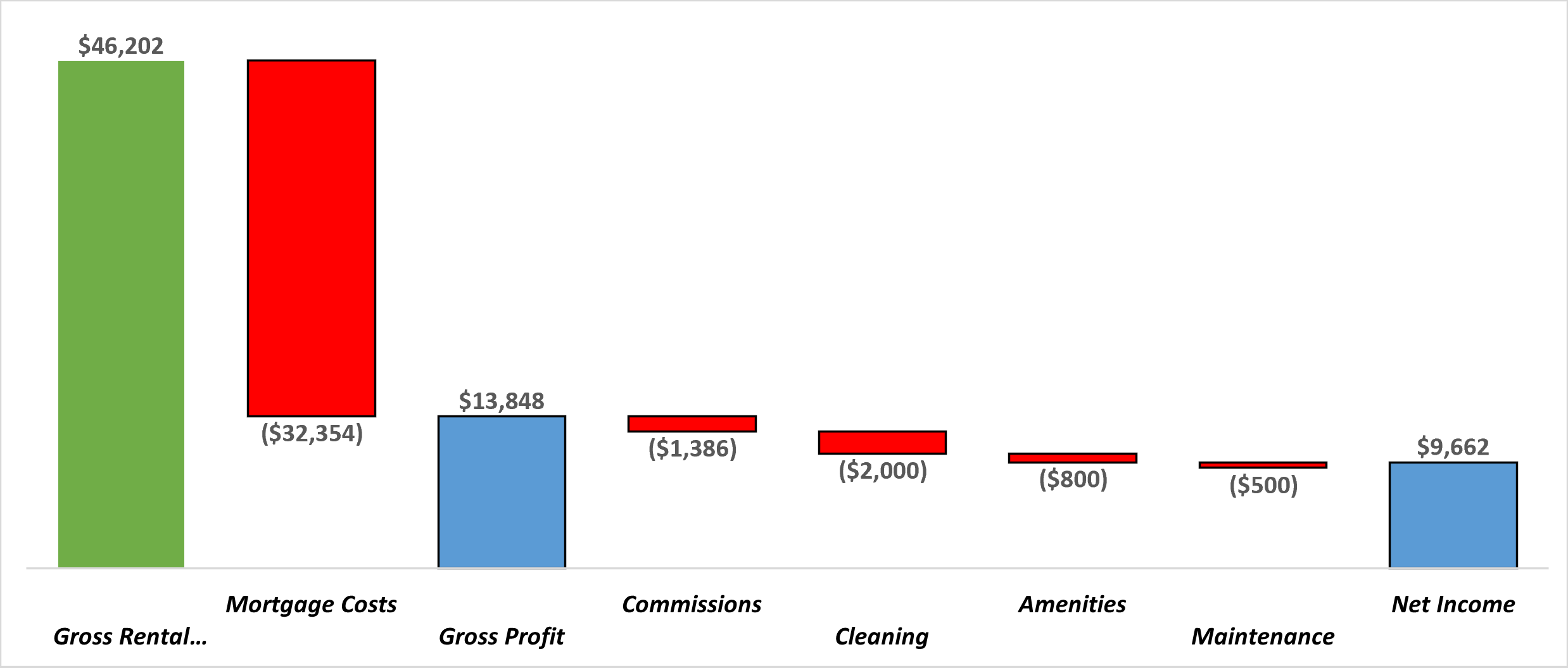
* + Neighbourhoods in the each district is visually represented by using a one of the hierarchy map sunburst map .

1. TREEMAP CHART:



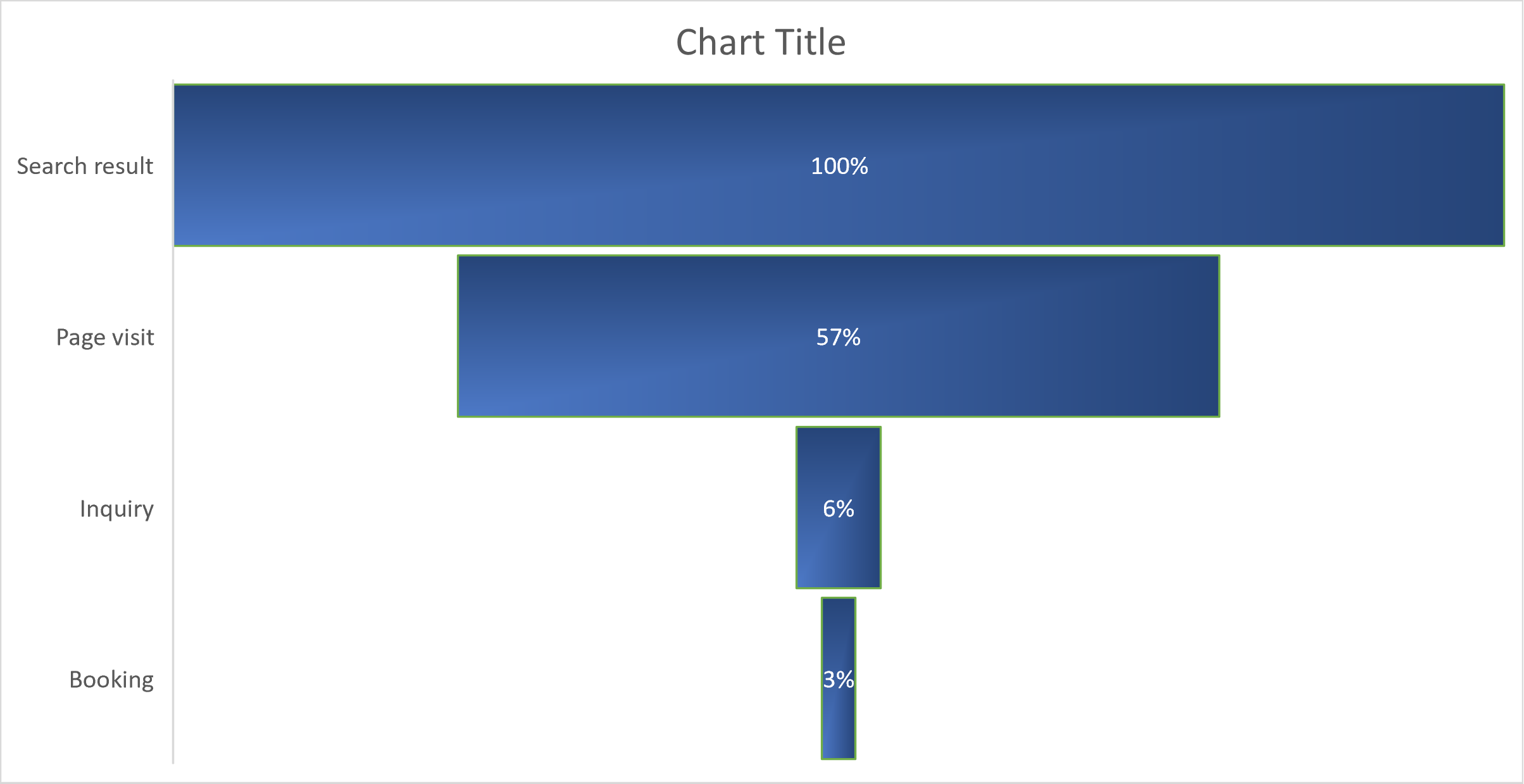
* + No of places in the respective neighbourhoods in the district is visually represented using another Hierarchy map Treemap
  + which is nested one another based on the population of the properties in the respective neighbourhoods of the districts.

1. WATERFALL CHART:



* + Balance sheet is visually represented using the Waterfall chart which is easy to get the hold of the positive and negative trends.

1. FUNNEL CHART:



* + Airbnb conversion funnel is created using the Funnel chart which will be used to know the efficiency of the conversion of our property listed on the airbnb to actual customers booking our property.

Conclusion :

FINDINGS OF MY ANALYSIS:

* + 50% of the properties are priced between $100 to $200 per night for the entire place.
  + Superhost percentage is gradually decreased from 60% to 32% from the year 2008 to 2020.
  + By analyzing i have chosen the optimal neighborhood for purchasing the property is Chelsea.
  + By earning $1000 pro it per month approximately after every expenses annual net income can be clocked around $9662 approximately.
  + 3 % of the people only booking the property from searching the property.