AIRBNB DATA ANALYSIS USING EXCEL: by: ARUNKUMAR links: github | 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;

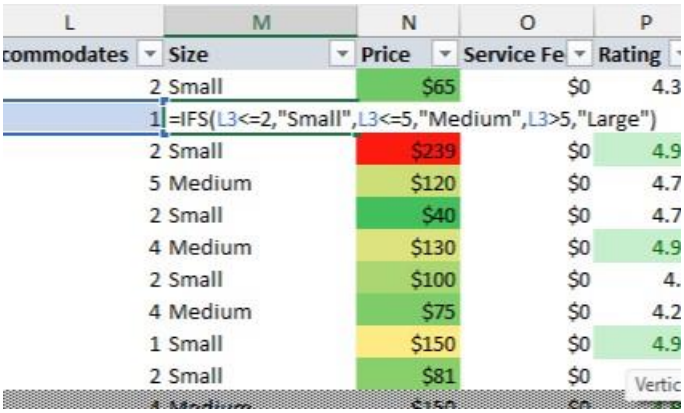
- Places -> contains Airbnb listing data Place name , host_id,Neighbourhood_id , room_type, rating , pricing etc;
- Hosts-> contains Airbnb host data including name, id, email, country, date_joined,response_time, response _rate, etc;
- Neighborhoods -> contains districts in New york and neighborhoods 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:

A) PLACES WORKSHEET:

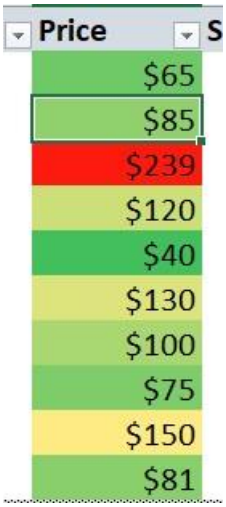
Inserting a new column size by using the

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



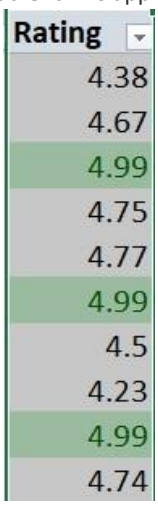
L	M	N	O	P
commodates	Size	Price	Service Fe	Rating
2	Small	$65	$0	4.3
1	=IFS(L3<=2,"Small",L3<=5,"Medium",L3>5,"Large")			
2	Small	$239	$0	4.9
5	Medium	$120	$0	4.7
2	Small	$40	$0	4.7
4	Medium	$130	$0	4.9
2	Small	$100	$0	4.
4	Medium	$75	$0	4.2
1	Small	$150	$0	4.9
2	Small	$81	$0	

- conditional formatting is applied for pricing with color scales



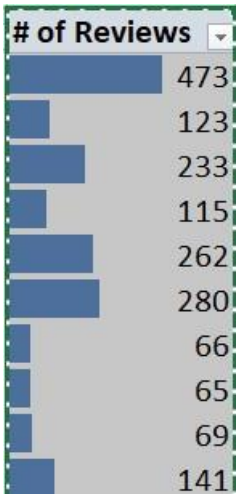
Price	S
$65	
$85	
$239	
$120	
$40	
$130	
$100	
$75	
$150	
$81	

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



Rating
4.38
4.67
4.99
4.75
4.77
4.99
4.5
4.23
4.99
4.74

- 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)
```

Response Time	Response Rate	Acceptance Rate	Superhost
Within a few hours	100.00%	100.00%	No
Within a day	100.00%	83.00%	No
Within an hour	100.00%	93.00%	Yes
Within an hour	100.00%	99.00%	No
Within a day	60.00%	73.00%	No
Within a day	100.00%	91.00%	Yes
Within a few hours	100.00%	97.00%	No
Within a few hours	83.00%	88.00%	No
Within a day	73.00%	93.00%	No
Within an hour	100.00%	97.00%	No

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

District	Neighborhood
Manhattan	Hell's Kitchen
Manhattan	Upper West Side
Manhattan	East Harlem
Brooklyn	Bedford-Stuyvesant
Manhattan	East Village
Queens	Long Island City
Brooklyn	Bedford-Stuyvesant
Queens	Jamaica
Manhattan	Chelsea
Brooklyn	Bedford-Stuyvesant

```
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.

Place Id	Place Name	Host ID	Response Time	Response Rate	Acceptance Rate	Superhost	Neighborhood ID	District	Neighborhood	Room Type	Accommodates
1	Large Furnished Room Near B'wa	8	Within a few hours	100.00%	100.00%	No	17	Manhattan	Hell's Kitchen	Private room	2
2	Wonderful Guest Bedroom in Ma	10	Within a day	100.00%	83.00%	No	20	Manhattan	Upper West Side	Private room	1
3	UES Beautiful Blue Room / 2 Bedr	14	Within an hour	100.00%	93.00%	Yes	14	Manhattan	East Harlem	Entire place	2
4	MAISON DES SIRENES1,bohemiar	24	Within an hour	100.00%	99.00%	No	7	Brooklyn	Bedford-Stuyvesant	Entire place	5
5	ENJOY Downtown NYC!	37	Within a day	60.00%	73.00%	No	15	Manhattan	East Village	Private room	2
6	1 Stop fr. Manhattan! Private Suit	38	Within a day	100.00%	91.00%	Yes	24	Queens	Long Island City	Private room	4
7	Charming Brownstone 3 - Near Pl	39	Within a few hours	100.00%	97.00%	No	7	Brooklyn	Bedford-Stuyvesant	Entire place	2
8	(E) RIGHT ON BUDGET	26	Within a few hours	83.00%	88.00%	No	23	Queens	Jamaica	Private room	4
9	Beautiful Attic bedroom Chelsea!	40	Within a day	73.00%	93.00%	No	13	Manhattan	Chelsea	Private room	1
10	Blue Room for 2 in Brownstone fc	42	Within an hour	100.00%	97.00%	No	7	Brooklyn	Bedford-Stuyvesant	Private room	2
11	Greenpoint Place...Has It All!	44	Within a few hours	75.00%	97.00%	Yes	11	Brooklyn	Greenpoint	Entire place	4
12	Williamsburg 1 bedroom Apartme	45	Within an hour	100.00%	100.00%	No	12	Brooklyn	Williamsburg	Entire place	2
13	Best Location in NYC! TIMES SQU/	49	Within a few hours	100.00%	93.00%	No	17	Manhattan	Hell's Kitchen	Private room	1
14	Sunny & Spacious Chelsea Apartn	50	Within a day	100.00%	19.00%	No	13	Manhattan	Chelsea	Entire place	4
15	Sunny, Modern room in East Villa	54	Within a day	50.00%	98.00%	Yes	15	Manhattan	East Village	Private room	2
16	Large Room in Amazing East Villag	4	Within a few hours	100.00%	57.00%	No	15	Manhattan	East Village	Private room	1
17	SpaHa Loft: Enormous and Bright	57	Within a day	100.00%	100.00%	Yes	14	Manhattan	East Harlem	Entire place	8
18	Beautiful Meatpacking District Lof	64	Within a day	100.00%	93.00%	Yes	13	Manhattan	Chelsea	Private room	3
19	Cozy 2 BR in Williamsburg	65	A few days or more	0.00%	100.00%	No	12	Brooklyn	Williamsburg	Private room	4
20	Spacious luminous apt Upper We:	66	Within a few hours	100.00%	100.00%	Yes	16	Manhattan	Harlem	Entire place	3
21	Cozy Bedroom in Williamsburg 3 I	65	A few days or more	0.00%	100.00%	No	12	Brooklyn	Williamsburg	Private room	2
22	Sunny room+Pvte office in huge l	343	A few days or more	43.00%	32.00%	No	8	Brooklyn	Bushwick	Private room	2
23	Williamsburgâ€"Steps To Subway	76	Within an hour	100.00%	100.00%	Yes	12	Brooklyn	Williamsburg	Private room	2
24	1,800 sq foot in luxury building	80	Within an hour	100.00%	100.00%	No	16	Manhattan	Harlem	Private room	2
25	Times Square, Safe, Clean and Co:	49	Within a few hours	100.00%	93.00%	No	17	Manhattan	Hell's Kitchen	Private room	1
26	2 story family home in Williamsbu	87	Within an hour	100.00%	100.00%	Yes	12	Brooklyn	Williamsburg	Entire place	6

Places | Hosts | Neighborhoods | Mortgage Calculator | 2019 Occupation | Pivot table Overview analysis | pt Price Per Guest in chelsea | Amortization Schedule | Visualizations

B) 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)
```

D	E	F	G	H
Date Joined	Date	Year Joined	Month Joined	Weekday Joine
20080827	2008-08-27	2008	8	3
20080907	2008-09-07	2008	9	7
20080910	2008-09-10	2008	9	3
20081119	2008-11-19	2008	11	3
20081212	2008-12-12	2008	12	5
20090129	2009-01-29	2009	1	4
20090210	2009-02-10	2009	2	2
20090303	2009-03-03	2009	3	2
20090321	2009-03-21	2009	3	6

- 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)

Host Location	Host State	Host Country
New York, New York, United States	New York	United States
Brooklyn, New York, United States	New York	United States
Brooklyn, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States
New York, New York, United States	New York	United States

LEFT SIDE: RIGHT SIDE:

Host ID: 49 *Name:* **Kyle P. Mcintyre**

	Totals	Entire place	Private room	Shared room
Places:	**2**	0	2	0
Reviews:	**699**	0	699	0
Rating:	**4.70**	-	4.70	-

Superhost?

1

First Review:	2010-01-02
Last Review:	2020-07-19
Days since Last Review:	983
Workdays since Last Review:	703

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.

C) NEIGHBOURHOODS WORKSHEET:

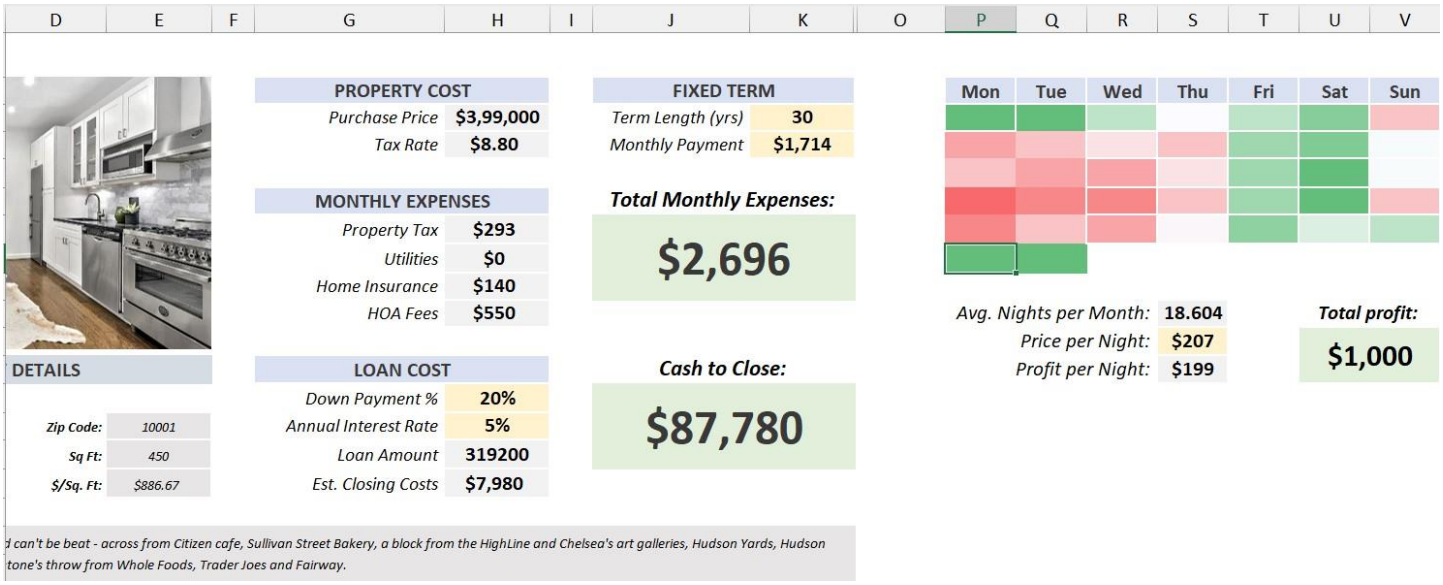
- In neighbourhood worksheet i have added 0nly two columns extra where it tells the no of places in the respective neighbourhood where i used the COUNTIF function.

- 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)
```

A	B	C	D	E
Neighborhood ID	**District**	**Neighborhood**	**count**	**No Of Places**
1	Bronx	Allerton	1	22
2	Bronx	Concourse	1	23
3	Bronx	Longwood	1	17
4	Bronx	Mott Haven	1	27
5	Bronx	Port Morris	1	23
6	Bronx	Wakefield	1	25
7	Brooklyn	Bedford-Stuyvesant	1	940
8	Brooklyn	Bushwick	1	451
9	Brooklyn	Crown Heights	1	341
10	Brooklyn	East Flatbush	1	183
11	Brooklyn	Greenpoint	1	180
12	Brooklyn	Williamsburg	1	640
13	Manhattan	Chelsea	1	146
14	Manhattan	East Harlem	1	230

D) MORTGAGE CALCULATOR:



PROPERTY COST	
Purchase Price	$3,99,000
Tax Rate	$8.80

MONTHLY EXPENSES	
Property Tax	$293
Utilities	$0
Home Insurance	$140
HOA Fees	$550

LOAN COST	
Down Payment %	20%
Annual Interest Rate	5%
Loan Amount	319200
Est. Closing Costs	$7,980

FIXED TERM	
Term Length (yrs)	30
Monthly Payment	$1,714

Total Monthly Expenses: $2,696

Cash to Close: $87,780

DETAILS	
Zip Code:	10001
Sq Ft:	450
$/Sq. Ft:	$886.67

Mon	Tue	Wed	Thu	Fri	Sat	Sun

Avg. Nights per Month:	18.604
Price per Night:	$207
Profit per Night:	$199

Total profit: $1,000

d can't be beat - across from Citizen cafe, Sullivan Street Bakery, a block from the HighLine and Chelsea's art galleries, Hudson Yards, Hudson tone's throw from Whole Foods, Trader Joes and Fairway.

- 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))
```

E) 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.

F) 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)
```

Year	Principal	Interest	Balance
2020	$1,544	$5,310	$3,17,656
2021	$4,788	$15,774	$3,12,868
2022	$5,033	$15,529	$3,07,835
2023	$5,291	$15,272	$3,02,544
2024	$5,562	$15,001	$2,96,982
2025	$5,846	$14,716	$2,91,136
2026	$6,145	$14,417	$2,84,991
2027	$6,460	$14,103	$2,78,531
2028	$6,790	$13,772	$2,71,741
2029	$7,137	$13,425	$2,64,604
2030	$7,503	$13,060	$2,57,101

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

	A	B	C	D	E	F	G	H
1	Room Type	Entire place						
2	District	Manhattan						
3	Status	Active						
4								
5	**Size**	**Neighborhood**	**No of Places**	**Average Price**	**Price Difference**	**Average Rating**	**Rating Rank**	
6	**Small**		**326**	**₹ 155**		**4.83**		
7		East Harlem	26	₹ 118	-26%	4.84	5	
8		Harlem	40	₹ 119	-26%	4.84	4	
9		Upper East Side	61	₹ 148	-8%	4.83	6	
10		East Village	54	₹ 160	0%	4.85	3	
11		Chelsea	35	₹ 160		4.89	1	
12		Hell's Kitchen	40	₹ 161	0%	4.74	8	
13		Upper West Side	44	₹ 178	11%	4.86	2	
14		Midtown	26	₹ 194	21%	4.82	7	
15	**Medium**		**626**	**₹ 197**		**4.77**		
16		East Harlem	40	₹ 160	-36%	4.76	4	
17		Harlem	130	₹ 164	-35%	4.81	3	
18		Upper East Side	94	₹ 181	-28%	4.73	7	
19		Hell's Kitchen	110	₹ 183	-27%	4.76	5	
20		East Village	69	₹ 188	-25%	4.74	6	
21		Upper West Side	94	₹ 246	-2%	4.82	2	
22		Chelsea	42	₹ 251		4.84	1	
23		Midtown	47	₹ 255	2%	4.72	8	
24	**Large**		**262**	**₹ 406**		**4.77**		
25		East Harlem	31	₹ 193	-48%	4.67	7	
26		Chelsea	13	₹ 368		4.87	1	
27		Midtown	31	₹ 369	0%	4.64	8	
28		Hell's Kitchen	45	₹ 372	1%	4.80	4	
29		Harlem	65	₹ 374	2%	4.80	5	
30		East Village	24	₹ 388	5%	4.82	3	
31		Upper West Side	37	₹ 578	57%	4.76	6	
32		Upper East Side	16	₹ 777	111%	4.87	2	
33	**Grand Total**		**1214**	**₹ 231**		**4.79**		
34								
35								

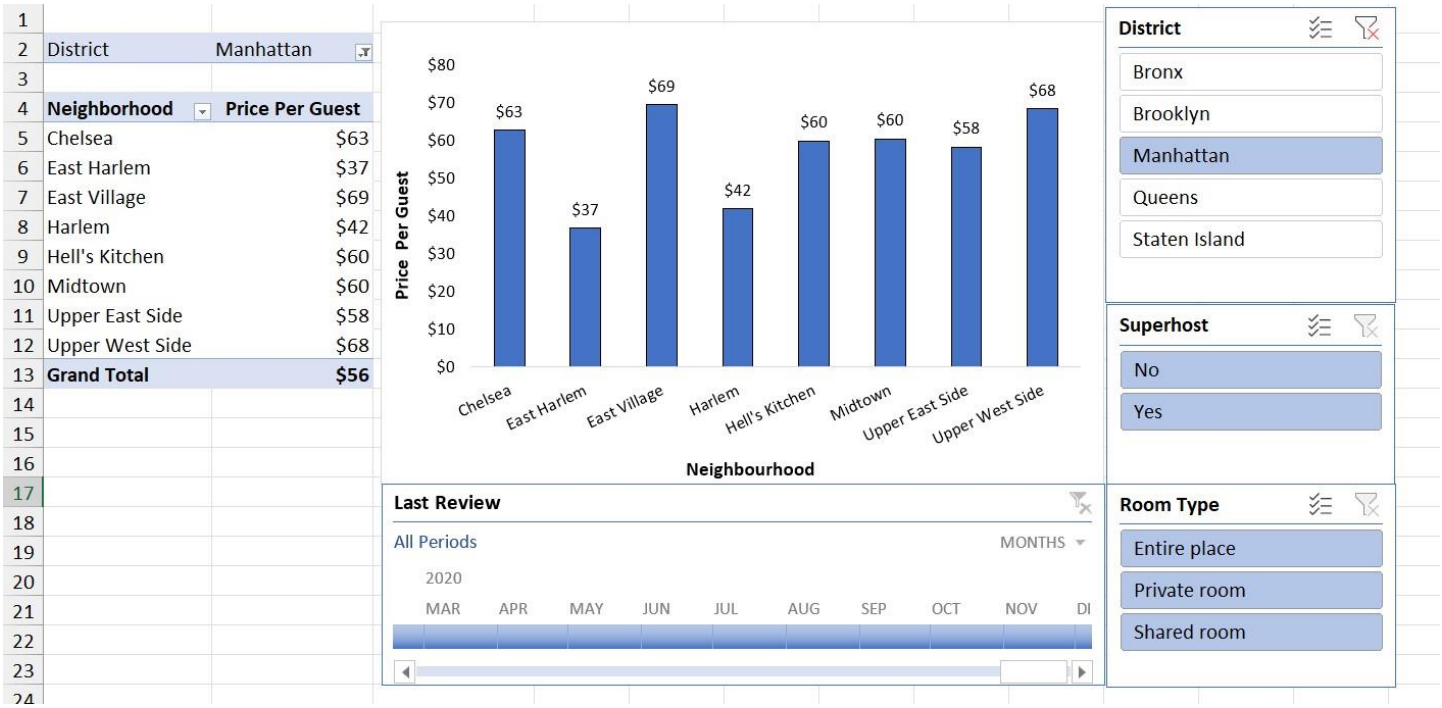
H) PT (PIVOT_TABLE) PRICE PER GUEST IN MANHATTAN NEIGHBOURHOOD:

Inserted the following ;

1) district into filter field in pivot table.

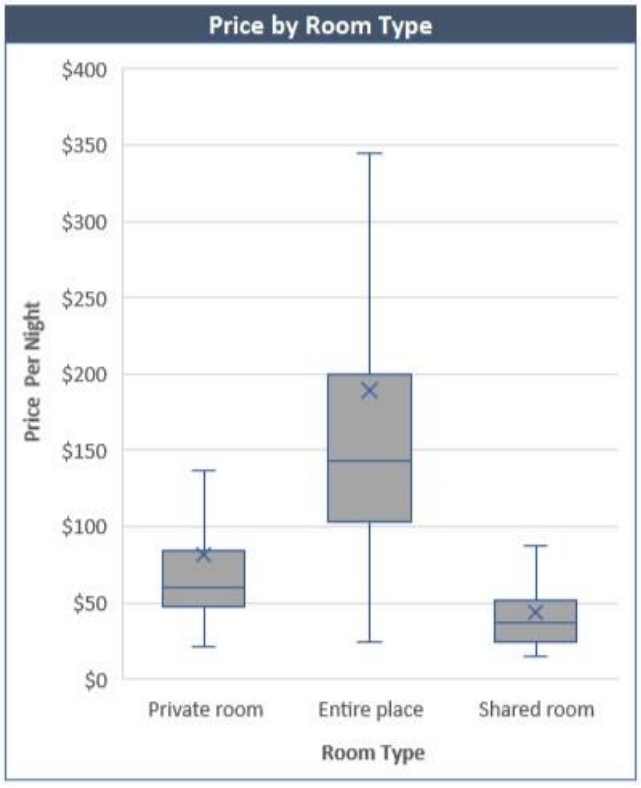
2) neighbourhood into rows field in pivot table

- 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:

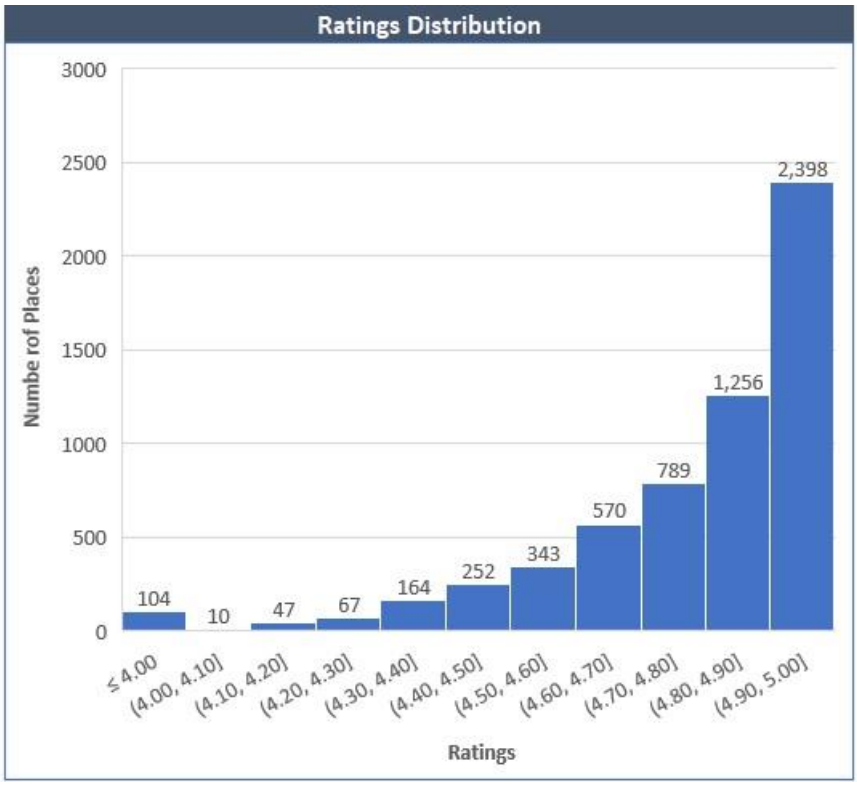
1) BOX AND WHISKER PLOT:



- 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 .

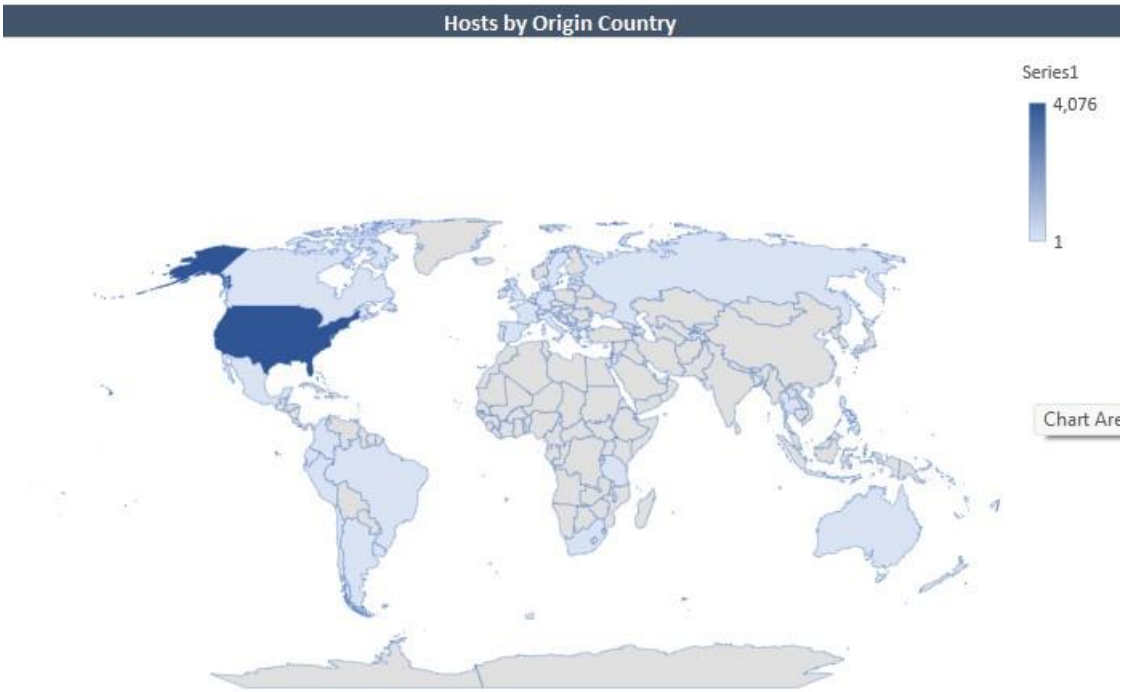
2) BAR CHART:

AIRBNB DATA ANALYSIS USING EXCEL:



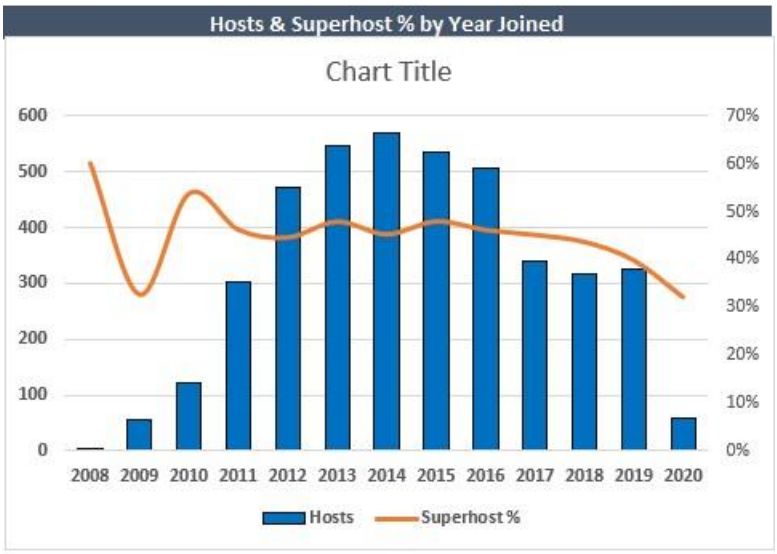
- Rating distribution visual is created using the bar chart based on the ratings and no of places falls on the respective ratings.

3) MAP CHART:



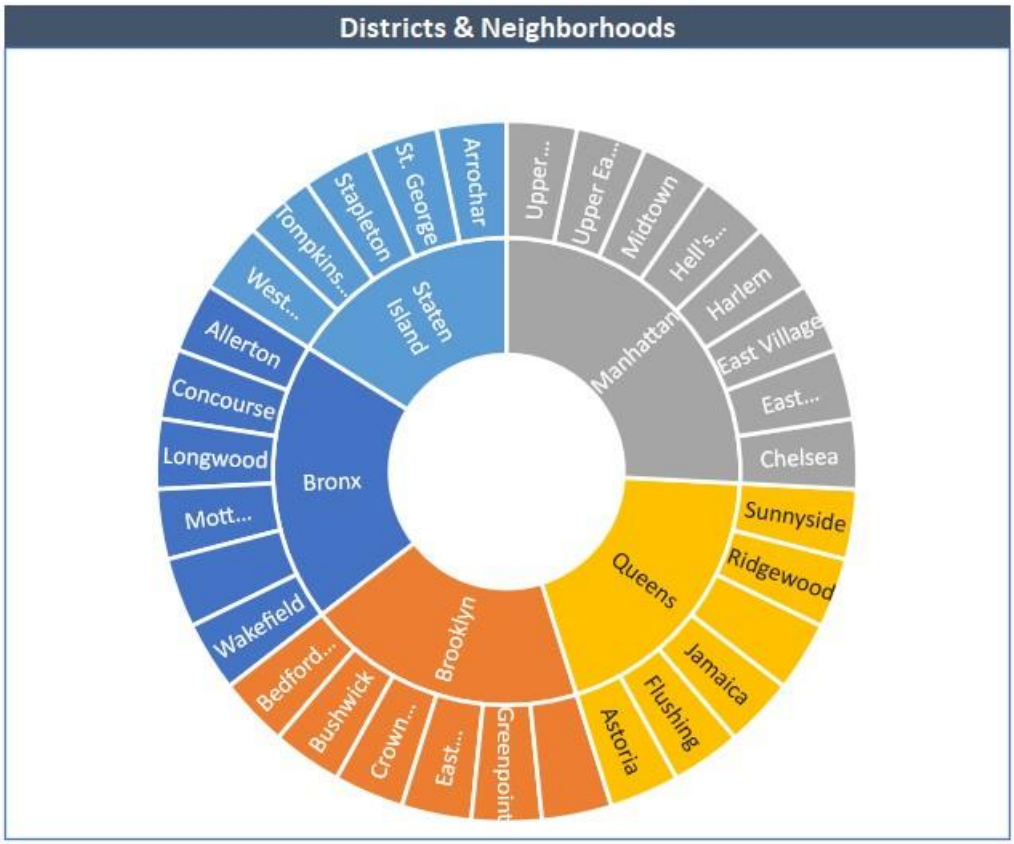
- Hosts by country is visualized by using or inserting the map chart and make sure to convert the data into geographical.

4) COMBO CHART (LINE AND BAR) :



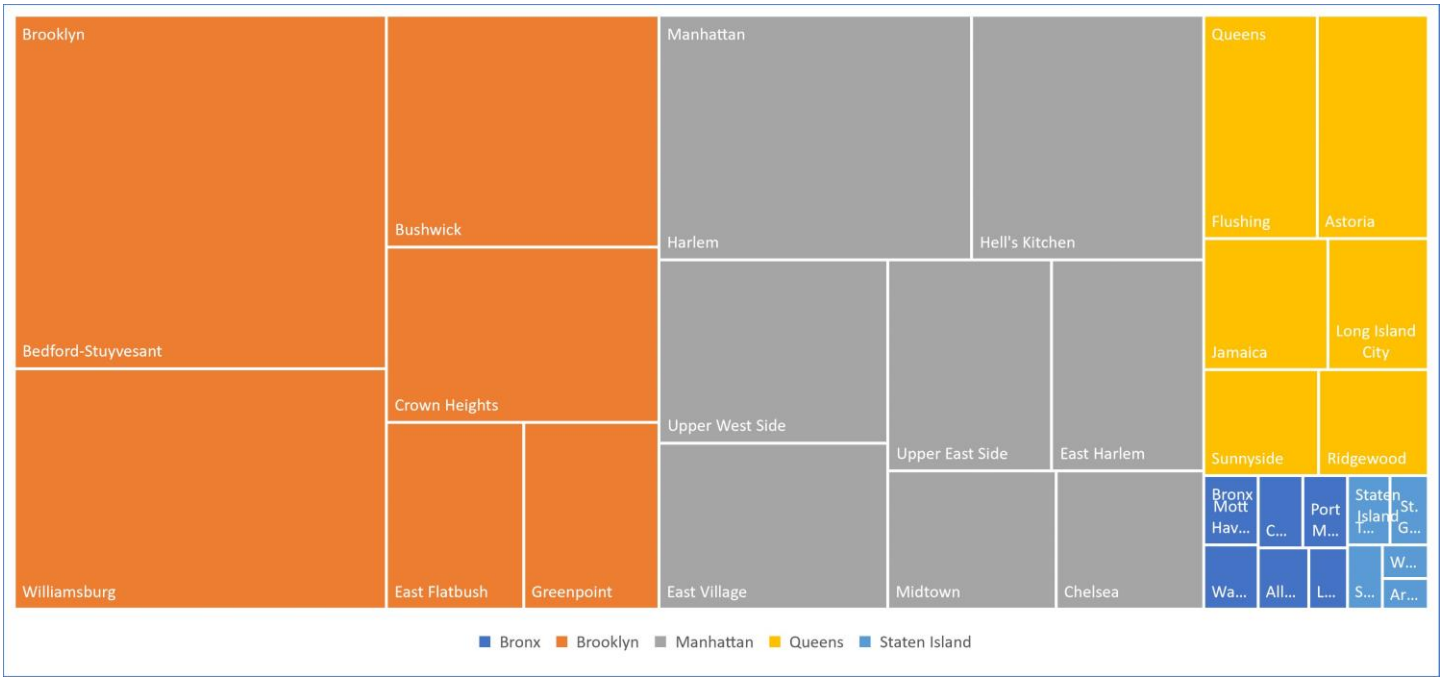
- Distribution of host to the superhost percentage is plotted using the combo chart of bar graph and line chart combined with respect to the years and no of hosts or superhosts in it.

5) SUNBURST CHART:



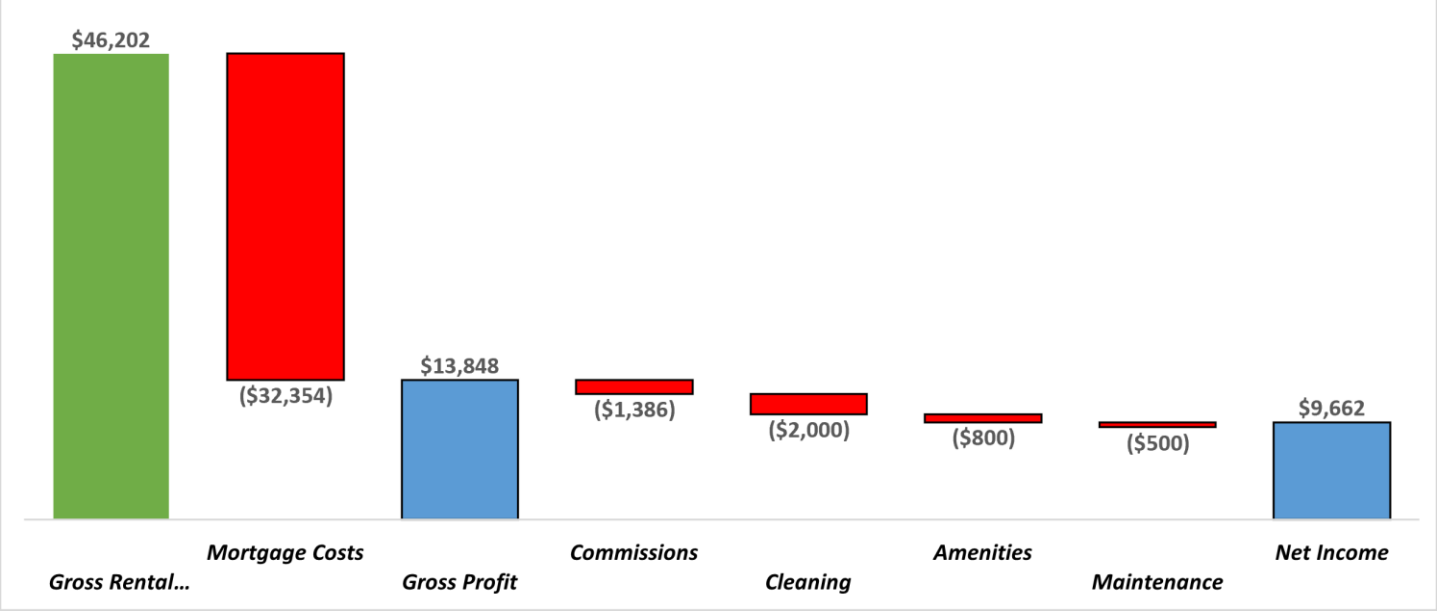
- Neighbourhoods in the each district is visually represented by using a one of the hierarchy map sunburst map .

6) TREEMAP CHART:



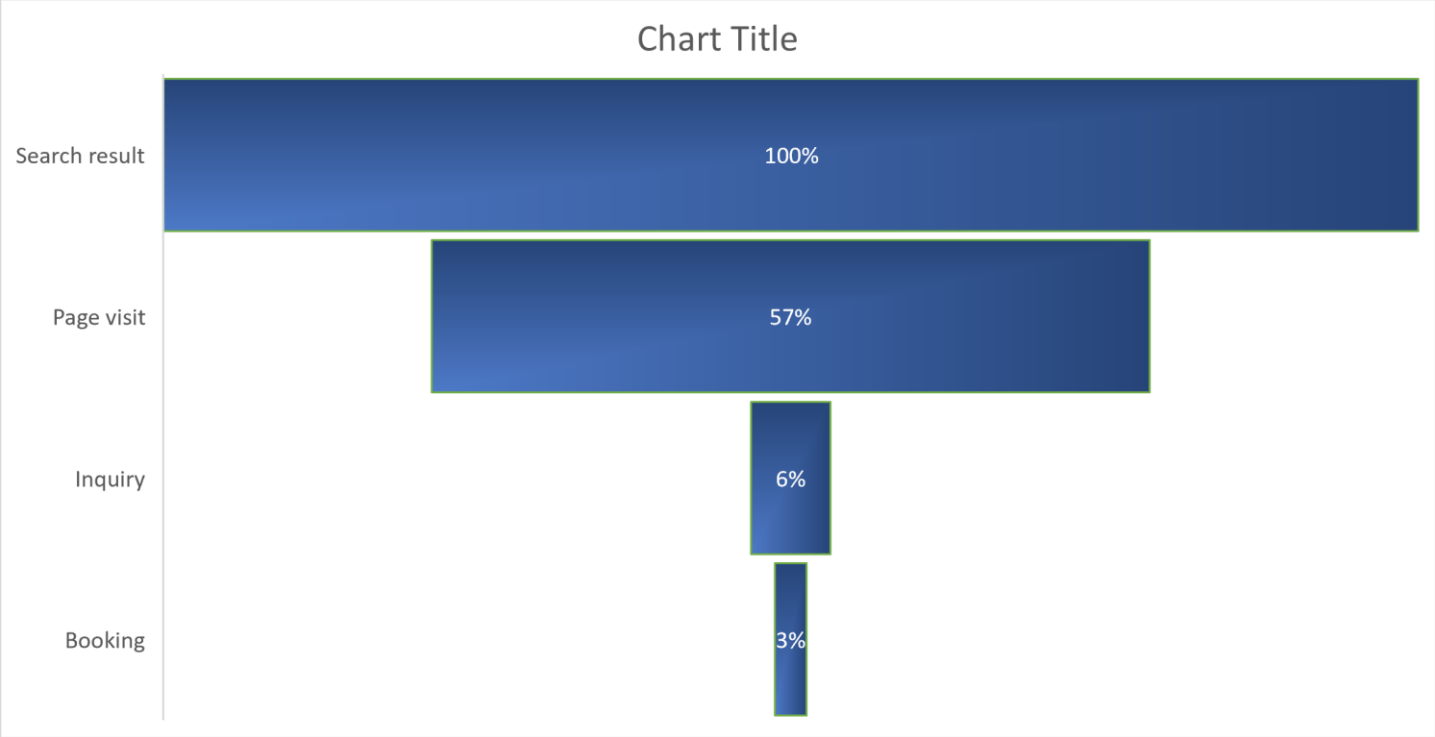
- No of places in the respective neighbourhoods in the district is visually represented using another Hierarchy map Treemap
- which is nested one another based on the population of the properties in the respective neighbourhoods of the districts.

7) WATERFALL CHART:



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

8) 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.