

Project Report

Futuristic Airbnb – Exploring in London

Saranya Pandiaraj (015304497)
Data 230 Data Visualization
Department of Applied Data Science
San Jose State University
Section 11, Spring 2021

Project Details

Name	Saranya Pandiaraj
Student ID	015304497
GitHub ID	SaranyaPandiaraj
Project GitHub Link	https://github.com/SaranyaPandiaraj/Futuristic_Airbnb-Exploring_in_London

Project Goal

The main goal of this project is to design an Interactive dashboard using data visualization tools (Tableau) from the publicly available London Airbnb dataset.

Table of Contents

Abstract	4
Introduction & Motivation	5
Project Deliverables	6
Data Set Overview	8
Data Set Resources	8
Data Set Description	8
Dataset features	9
Sample Dataset.....	12
Data Cleaning	14
Determining the Null Value data	14
Dropping the features.....	15
Converting the Categorical features to Continuous	16
Dropping Similar Kind of Columns	18
Replacing Column values	18
Changing the Data Types	22
Removing special characters	22
Dropping Duplicates	22
Design Principles	23
Dashboard Visualization	24

Overview of Airbnb – London	24
Listing's Overview	27
When to Plan for Vacation?	31
Airbnb's Listing in London	35
Hosting's Overview and Improve Host Business in London.....	37
Price Analytics.....	44
Most Expensive Properties and Neighbourhoods	48
Story Telling	50
Takeaways/Outcomes	51
Conclusion & Future Work	52
Project Code.....	53
References	54

Abstract

Nowadays planning for a vacation is a very tedious task, it involves lots of processes from planning, deciding, and booking the best accommodation for a comfortable trip. Airbnb is an online platform for temporary housing for a vacation where the hosts in Airbnb list and rents out their properties. Based on these listings, this platform helps the user to locate a comfortable accommodation according to their requirements. You can access their portal via web app or mobile app which is very user-friendly. Airbnb is very popular in many places, but this project will concentrate on Airbnb London, England, United Kingdom. The project aims in discovering the London Airbnb data, performing the ETL process, and loading the cleansed data into Tableau. Once the cleansed data is uploaded in Tableau, interactional dashboards are created to help the users to discover the Airbnb Listing and Properties. This will help them to decide when to plan for a vacation and choose the best accommodation according to the user preference. They can also determine the host response and acceptance rate depending on the listings chosen. The dashboard will also help the hosts to improve their business in Airbnb and visually analyze the future predictions.

Introduction & Motivation

Airbnb is a vacation rental platform where the user can rent a home for tourism experiences. User can also list their properties in Airbnb for renting it out. Airbnb can see an exponential increase in the number of rentals listed on their platform every year. The platform has attracted more people with their home friendly services and hospitality. Thus, feels like staying in a home even on a vacation.

London is the most popular place having around 76534 listings as of December 16, 2020. Around 40% of their listings with the entire home types are being rented out to the tourists for more than 90 days term. On an average term, Airbnb guests stay for around an average of 4.5 nights in London.

The main purpose of this project is to analyze and visualize the Airbnb dataset. Our investigation will help us to determine the factors for the success of Airbnb business; places to explore for the hosts to expand the Airbnb business; covid impacts on Airbnb; listings managed by the hosts; Renting prices based on the accommodates; judging the host response rate; forecasting/predicting the listing price over the time, etc. To visually analyze the data, we will be creating an interactive dashboard & stories using Tableau. This dashboard will be helpful for the user to find a better accommodation based on customer satisfaction & requirements. And, will also help in improving the Airbnb business based on certain factors in London.

Project Deliverables

Table 1

Project Deliverables

Sr No.	Deliverable Name	Deliverable Description	Start Date	End Date	Status	Notes
1	Requirement Gathering	Project Goals - What needs to be done?	02/01/2021	02/02/2021	Completed	
2	Data Set Exploration	Identifying & Understanding the Data Sets	02/03/2021	02/08/2021	Completed	
3	Project Proposal	Project Abstract – What is to be done? - Research questions	02/09/2021	02/11/2021	Completed	
4	Data Extraction & Clean-up	Cleaning up unnecessary data from the dataset	02/22/2021	03/31/2021	Completed	Includes of Coding work
5	Final Data Analysis	Data Transformation (Any Logics to be applied)	02/22/2021	03/31/2021	Completed	Includes of Coding work

6	Data Visualizations (Includes of Creating Stories & Dashboards)	Listing Visualizations	04/01/2021	05/13/2021	Completed	Parallel Work on Visualization (plus coding work)
		Host Visualizations	04/01/2021	05/13/2021	Completed	
		Reviews Visualizations	04/01/2021	05/13/2021	Completed	
7	Outcomes	Take Away from the dashboards	05/01/2021	05/13/2021	Completed	
8	Future Work	Future Ideas to implement in the dashboards	05/13/2021	05/15/2021	Completed	
9	Presentation/ Project Report	To create a PPT & Project Report to include abstract, diagrams, snapshots; etc;	05/15/2021	05/23/2021	Completed	

Data Set Overview

Data Set Resources

Data is collected from the Insiderairbnb website.

Insideairbnb.com – London, England, United Kingdom

(<http://insideairbnb.com/get-the-data.html>)

- Sourced from publicly available information in the Airbnb Site

Data Set Description

The Data Files used for this project is as below.

- **listings.csv**

This file contains the Airbnb Host, listing, and Property Information's. It also contains the review scores & availability details of the property.

- **reviews.csv**

This file contains the reviews information of each property.

- **calendar.csv**

This file contains the Property availability and its price information for the future year. And, also it provides the minimum and maximum nights for the particular property.

- **Neighbourhoods.csv**

This file contains the London neighbourhoods having the Airbnb listings.

- **neighbourhoods.geojson**

This file contains the geographical features of the London neighbourhoods having the Airbnb listings.

Dataset features

Listings File

The Listing's file has around 76534 Airbnb listings in London. There are 74 features in this file as displayed below.

Figure 1

Listings features

```
Listings_df.columns
```

```
Index(['id', 'listing_url', 'scrape_id', 'last_scraped', 'name', 'description',
       'neighborhood_overview', 'picture_url', 'host_id', 'host_url',
       'host_name', 'host_since', 'host_location', 'host_about',
       'host_response_time', 'host_response_rate', 'host_acceptance_rate',
       'host_is_superhost', 'host_thumbnail_url', 'host_picture_url',
       'host_neighbourhood', 'host_listings_count',
       'host_total_listings_count', 'host_verifications',
       'host_has_profile_pic', 'host_identity_verified', 'neighbourhood',
       'neighbourhood_cleansed', 'neighbourhood_group_cleansed', 'latitude',
       'longitude', 'property_type', 'room_type', 'accommodates', 'bathrooms',
       'bathrooms_text', 'bedrooms', 'beds', 'amenities', 'price',
       'minimum_nights', 'maximum_nights', 'minimum_minimum_nights',
       'maximum_minimum_nights', 'minimum_maximum_nights',
       'maximum_maximum_nights', 'minimum_nights_avg_ntm',
       'maximum_nights_avg_ntm', 'calendar_updated', 'has_availability',
       'availability_30', 'availability_60', 'availability_90',
       'availability_365', 'calendar_last_scraped', 'number_of_reviews',
       'number_of_reviews_ltm', 'number_of_reviews_l30d', 'first_review',
       'last_review', 'review_scores_rating', 'review_scores_accuracy',
       'review_scores_cleanliness', 'review_scores_checkin',
       'review_scores_communication', 'review_scores_location',
       'review_scores_value', 'license', 'instant_bookable',
       'calculated_host_listings_count',
       'calculated_host_listings_count_entire_homes',
       'calculated_host_listings_count_private_rooms',
       'calculated_host_listings_count_shared_rooms', 'reviews_per_month'],
      dtype='object')
```

```
Listings_df.shape
```

```
(76534, 74)
```

Reviews File

The Reviews file has around 1163886 reviews for the Airbnb's listing in London. There are 6 features in this file as displayed below.

Figure 2

Reviews features

```
Reviews_df.columns
```

```
Index(['listing_id', 'id', 'date', 'reviewer_id', 'reviewer_name', 'comments'], dtype='object')
```

```
Reviews_df.shape
```

```
(1163886, 6)
```

Neighbourhoods File

The Neighbourhoods file has around 33 neighbourhoods having the Airbnb listings in London. There is only one feature in this file as displayed below.

Figure 3

Neighbourhoods features

```
Neighbourhoods_df.columns
```

```
Index(['neighbourhood'], dtype='object')
```

```
Neighbourhoods_df.shape
```

```
(33, 1)
```

Neighbourhoods Geographical File

The Neighbourhoods geographical file has around 33 neighbourhoods (in terms of latitude and longitude) having the Airbnb listings in London. There are two feature in this file as displayed below.

Figure 4

Neighbourhoods Geo features

```
Neighbourhoods_json_df.columns
```

```
Index(['type', 'features'], dtype='object')
```

```
Neighbourhoods_json_df.shape
```

```
(33, 2)
```

Calendar File

The Calendar file has around 27935194 rows of showing the price, availability, minimum and maximum nights of each calendar day. There are six feature in this file as displayed below.

Figure 5

Calendar features

```
Calendar_Summary_df.columns
```

```
Index(['date', 'available', 'price', 'adjusted_price', 'minimum_nights',
       'maximum_nights'],
      dtype='object')
```

```
Calendar_Summary_df.shape
```

```
(27935194, 6)
```

Sample Dataset

Below is the sample dataset for the Listings, Reviews, Neighbourhoods and Calendar data files.

Figure 6

Listings Sample Data

Sample Dataset

`Listings_df.head(5)`

	id	listing_url	scrape_id	last_scraped	name	description	neighborhood_overview	
0	13913	https://www.airbnb.com/rooms/13913	20210209201301	2021-02-12	Holiday London DB Room Let-on going	My bright double bedroom with a large window h...	Finsbury Park is a friendly melting pot commun...	https://a0.muscache.com/picture
1	15400	https://www.airbnb.com/rooms/15400	20210209201301	2021-02-12	Bright Chelsea Apartment. Chelseal!	Lots of windows and light. St Luke's Gardens ...	It is Chelsea.	https://a0.muscache.com/pictures
2	17402	https://www.airbnb.com/rooms/17402	20210209201301	2021-02-12	Superb 3-Bed/2 Bath & Wifi: Trendy W1	You'll have a wonderful stay in this superb mo...	Location, location, location! You won't find b...	https://a0.muscache.com/pictures
3	17506	https://www.airbnb.com/rooms/17506	20210209201301	2021-02-12	Boutique Chelsea/Fulham Double bed 5-star ensuite	Enjoy a chic stay in this elegant but fully mo...	Fulham is 'villagey' and residential – a real ...	https://a0.muscache.com/pictures
4	25123	https://www.airbnb.com/rooms/25123	20210209201301	2021-02-13	Clean big Room in London (Room 1)	Big room with double bed/ clean sheets/ clean ...	Barnet is one of the largest boroughs in Londo...	https://a0.muscache.com/pictures

Figure 7

Reviews Sample Data

`Reviews_df.head(5)`

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	13913	80770	2010-08-18	177109	Michael	My girlfriend and I hadn't known Alina before ...
1	13913	367568	2011-07-11	19835707	Mathias	Alina was a really good host. The flat is clea...
2	13913	529579	2011-09-13	1110304	Kristin	Alina is an amazing host. She made me feel rig...
3	13913	595481	2011-10-03	1216358	Camilla	Alina's place is so nice, the room is big and ...
4	13913	612947	2011-10-09	490840	Jorik	Nice location in Islington area, good for shor...

Figure 8

Neighbourhoods and Calendar Sample Data

```
Neighbourhoods_df.head(5)
```

neighbourhood	
0	Barking and Dagenham
1	Barnet
2	Bexley
3	Brent
4	Bromley

```
Neighbourhoods_json_df.head(5)
```

	type	features
0	FeatureCollection	{'type': 'Feature', 'geometry': {'type': 'Multi...
1	FeatureCollection	{'type': 'Feature', 'geometry': {'type': 'Multi...
2	FeatureCollection	{'type': 'Feature', 'geometry': {'type': 'Multi...
3	FeatureCollection	{'type': 'Feature', 'geometry': {'type': 'Multi...
4	FeatureCollection	{'type': 'Feature', 'geometry': {'type': 'Multi...

```
Calendar_Summary_df.head(5)
```

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
	100326	2021-02-13	f	33.0	33.0	2.0	10.0
	13913	2021-02-12	t	40.0	40.0	1.0	29.0
	13913	2021-02-13	t	40.0	40.0	1.0	29.0
	13913	2021-02-14	t	40.0	40.0	1.0	29.0
	13913	2021-02-15	t	40.0	40.0	1.0	29.0

Data Cleaning

Determining the Null Value data

The listings file has around 24 features having null values. The other data files (Reviews, Neighborhoods, Calendar) don't have the null value data. So, the first step is to fix these null value data.

Figure 9

Listings Percentage of Missing Values (Features)

```
: #Determining the Null Value Data
round(( Listings_df.isna().sum()/len(Listings_df) ) * 100).sort_values(ascending=False)

neighbourhood_group_cleansed      100.0
license                           100.0
bathrooms                         100.0
calendar_updated                  100.0
host_response_rate                56.0
host_response_time                56.0
host_acceptance_rate              45.0
host_about                         45.0
neighbourhood                     39.0
neighborhood_overview             39.0
review_scores_rating              31.0
review_scores_accuracy            31.0
review_scores_cleanliness          31.0
review_scores_checkin              31.0
review_scores_communication       31.0
review_scores_location             31.0
review_scores_value               31.0
last_review                        29.0
first_review                       29.0
reviews_per_month                 29.0
host_neighbourhood                24.0
bedrooms                           6.0
description                        4.0
beds                               2.0
host_has_profile_pic              0.0
host_name                          0.0
availability_60                   0.0
availability_90                   0.0
availability_365                  0.0
calendar_last_scraped             0.0
number_of_reviews                  0.0
number_of_reviews_ltm              0.0
number_of_reviews_l30d              0.0
host_is_superhost                 0.0
host_location                      0.0
host_since                         0.0
host_url                           0.0
host_identity_verified             0.0
host_id                            0.0
picture_url                        0.0
name                               0.0
last_scraped                       0.0
scrape_id                          0.0
instant_bookable                   0.0
calculated_host_listings_count    0.0
calculated_host_listings_count_entire_homes 0.0
calculated_host_listings_count_private_rooms 0.0
calculated_host_listings_count_shared_rooms 0.0
availability_30                   0.0
has_availability                  0.0
host_thumbnail_url                 0.0
maximum_nights_avg_ntm            0.0
host_verifications                 0.0
neighbourhood_cleansed             0.0
host_total_listings_count          0.0
latitude                           0.0
longitude                          0.0
property_type                      0.0
room_type                          0.0
accommodates                        0.0
host_listings_count                0.0
bathrooms_text                     0.0
host_picture_url                   0.0
listing_url                        0.0
amenities                          0.0
price                              0.0
minimum_nights                      0.0
maximum_nights                      0.0
minimum_minimum_nights              0.0
maximum_minimum_nights              0.0
minimum_maximum_nights              0.0
maximum_maximum_nights              0.0
minimum_nights_avg_ntm              0.0
id                                0.0
dtype: float64
```

Figure 10

Reviews, Neighbourhoods, Calendar - Percentage of Missing Values (Features)

```
#Determining the Null Value Data
round(( Reviews_df.isna().sum()/len(Reviews_df) ) * 100).sort_values(ascending=False)

listing_id      0.0
id              0.0
date            0.0
reviewer_id     0.0
reviewer_name   0.0
comments        0.0
dtype: float64

#Determining the Null Value Data
round(( Neighbourhoods_df.isna().sum()/len(Neighbourhoods_df) ) * 100).sort_values(ascending=False)

neighbourhood    0.0
dtype: float64

#Determining the Null Value Data
round(( Neighbourhoods_json_df.isna().sum()/len(Neighbourhoods_json_df) ) * 100).sort_values(ascending=False)

type          0.0
features      0.0
dtype: float64

#Determining the Null Value Data
round(( Calendar_Summary_df.isna().sum()/len(Calendar_Summary_df) ) * 100).sort_values(ascending=False)

date          0.0
available     0.0
price          0.0
adjusted_price 0.0
minimum_nights 0.0
maximum_nights 0.0
dtype: float64
```

Dropping the features

The below features have more than 80% of Null Values data (*Looks like there is no data at all for these features*) ; Hence dropping these features.

Figure 11

Listings Feature showing more than 80% of Null values

```
#Determining the Null Value Data
round(( Listings_df.isna().sum()/len(Listings_df) ) * 100).sort_values(ascending=False)

neighbourhood_group_cleansed           100.0
license                                100.0
bathrooms                               100.0
calendar_updated                        100.0
```

Data Cleaning – Dropping the features

```
#The below Columns has more than 80% of NA values, hence dropping it.
|
Listings_df.drop(['license', 'neighbourhood_group_cleansed', 'calendar_updated'], axis=1, inplace=True)
```

Converting the Categorical features to Continuous

The bathrooms_text is converted to a Continuous feature such that it will be easy to calculate the number of bathrooms. This is done by extracting only the numerical values from the bathroom_text feature.

Figure 12

Bathrooms text – aggregated counts

```
Listings_df['bathrooms_text'].value_counts()
```

```
1 bath           32188
1 shared bath    14038
2 baths          8910
1 private bath   6112
1.5 baths        5054
1.5 shared baths 3687
2.5 baths        1859
2 shared baths   1344
3 baths          1102
3.5 baths        434
2.5 shared baths 347
4 baths          230
0 shared baths   173
3 shared baths   161
0 baths          136
Shared half-bath 109
Half-bath         107
4.5 baths        95
5 baths          48
3.5 shared baths 37
Private half-bath 31
5.5 baths        21
6 baths          21
4 shared baths   20
6 shared baths   19
10 baths         19
6.5 baths        8
4.5 shared baths 8
5 shared baths   7
7 baths          5
8 baths          5
8 shared baths   3
7 shared baths   3
8.5 baths        2
7.5 baths        2
17 baths         1
10.5 baths       1
11.5 baths       1
9 shared baths   1
35 baths         1
9 baths          1
11 baths         1
12 baths         1
Name: bathrooms_text, dtype: int64
```

Figure 13

Data Cleaning – Splitting the Numerical and Categorical Features

```
# Replacing few column value texts in bathrooms_text column with the relevant values
# such that it will easy to calculate the no. of bathrooms

Listings_df.bathrooms_text.replace({
    'Shared half-bath' : '0.5 shared bath',
    'Half-bath': '0.5 bath',
    'Private half-bath': '0.5 Private bath'
}, inplace=True)

# Bathrooms has full of Null Values ; Hence Replacing the Null values in the bathroom column
# by extracting the numerical values in the bathroom_txt feature
Listings_df["bathrooms"] = Listings_df["bathrooms_text"].str.split(" ", 1, expand=True)
Listings_df["bathrooms"] = Listings_df["bathrooms"].astype(float)
Listings_df["bathrooms"].value_counts()

1.0      52338
2.0     10254
1.5      8741
2.5     2206
3.0      1263
3.5       471
0.0      309
4.0       250
0.5      247
4.5       103
5.0       55
6.0       40
5.5       21
10.0      19
6.5        8
7.0        8
8.0        8
9.0        2
8.5        2
7.5        2
12.0       1
11.0       1
17.0       1
35.0       1
11.5       1
10.5       1
Name: bathrooms, dtype: int64
```

Dropping Similar Kind of Columns

These features have similar kind of column values; Hence dropping these columns and retaining only one feature for the respective features.

Figure 14

Data Cleaning – Dropping Similar kind of column values

```
: ## Similar Values in minimum n maximum nights
Listings_df[['minimum_minimum_nights', 'maximum_minimum_nights','minimum_maximum_nights',
'maximum_maximum_nights', 'minimum_nights_avg_ntm', 'maximum_nights_avg_ntm']].head()

: minimum_minimum_nights maximum_minimum_nights minimum_maximum_nights maximum_maximum_nights minimum_nights_avg_ntm maximum_nights_avg_ntm
0 1.0 1.0 29.0 29.0 1.0
1 10.0 10.0 50.0 50.0 10.0
2 4.0 4.0 365.0 365.0 4.0
3 3.0 3.0 21.0 21.0 3.0
4 10.0 10.0 1125.0 1125.0 10.0

: ##Dropping the Similar kind of Columns
##host_listings_count and host_total_listings_count are the same in all
##These cases are those where the value is NaN.

Listings_df.drop(['calculated_host_listings_count_shared_rooms','minimum_minimum_nights', 'maximum_minimum_nights',
'minimum_maximum_nights', 'maximum_maximum_nights', 'minimum_nights_avg_ntm', 'maximum_nights_avg_ntm',
'host_total_listings_count', 'calculated_host_listings_count_entire_homes','calculated_host_listings_cc
axis=1, inplace=True)
```

Replacing Column values

A similar kind of column value is replaced with the respective other column values for a better plot. For example, LED TV is replaced with TV

Amenities

First, the array of object values is split into a single value column; this makes calculating the amenities count easier. Then irrelevant values were removed and similar kinds of values were replaced with their equivalent values. For example, you can HDTV is replaced with TV. This will make the plot for better visualization.

Figure 15

Data Cleaning – Replacing column values (Amenities)

```
Listings_df.amenities[1:5].values
array(['["Refrigerator", "Shampoo", "Long term stays allowed", "Hangers", "Hot water", "Kitchen", "Wifi", "Washer", "Essentials", "Hair dryer", "Smoke alarm", "Heating", "Dishes and silverware", "TV", "Dryer", "Iron", "Cooking basics", "Air conditioning", "Dedicated workspace", "Luggage dropoff allowed", "Coffee maker", "Cable TV", "Fire extinguisher", "Microwave"]', '["Oven", "Shampoo", "Refrigerator", "Hangers", "Long term stays allowed", "Hot water", "Kitchen", "Stove", "Lockbox", "Wifi", "Washer", "Essentials", "Hair dryer", "Paid parking off premises", "Smoke alarm", "Heating", "Dishes and silverware", "TV", "Dishwasher", "Dryer", "Iron", "Cooking basics", "Dedicated workspace", "Bed linens", "Coffee maker", "Microwave", "Elevator"]', '["Shampoo", "Dedicated workspace", "Hangers", "Essentials", "Hot water", "Iron", "Hair dryer", "Shower gel", "Breakfast", "Smoke alarm", "Lock on bedroom door", "Carbon monoxide alarm", "Heating", "Wifi", "Air conditioning", "TV"]', '["Shampoo", "Dedicated workspace", "Hangers", "Essentials", "Long term stays allowed", "Iron", "Hair dryer", "Kitchen", "Coffee maker", "Free street parking", "Smoke alarm", "Lock on bedroom door", "Heating", "Wifi", "Washer"]'], dtype=object)

#Data Cleaning the Amenities Column

Amenities = Listings_df.amenities

Amenities_List = []

for x in Amenities:
    x=x.replace("'", '')
    x=x.replace('{', '')
    x=x.replace('}', '')
    x=x.replace('[', '')
    x=x.replace(']', '')
    x=x.replace('Cable TV', 'TV')
    x=x.replace('Pocket wifi', 'Wifi')
    Amenities_List += x.split(',')

Amenities_List.remove(' 2-5 years old')
Amenities_List.remove(' 1 space')
Amenities_List.remove(' 5-10 years old')
Amenities_List.remove(' 5 spaces')
Amenities_List.remove(' and 10+ years old')
Amenities_List.remove('')
Amenities_New = pd.DataFrame(Amenities_List)
Amenities_New = Amenities_New.rename(columns = {0:"Amenities"})

Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*HDTV.*$)', 'TV')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*Mbps.*$)', 'Wifi')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*sound.*$)', 'Sound system')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*conditioner.*$)', 'Conditioner')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*shampoo.*$)', 'Shampoo')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*conditioning.*$)', 'Air conditioning')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*soap.*$)', 'Body soap')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*TV.*$)', 'TV')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*oven.*$)', 'Oven')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*refrigerator.*$)', 'Refrigerator')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*stove.*$)', 'Stove')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*storage.*$)', 'Clothing storage')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*space.*$)', 'Dedicated workspace')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*closet.*$)', 'closet')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*dresser.*$)', 'dresser')
Amenities_New.Amenities = Amenities_New.Amenities.str.replace(r'(^.*Ceiling fans.*$)', 'Ceiling fan')

Amenities_New.describe()
Amenities_New = Amenities_New.dropna()
Amenities_New.to_csv("Amenities.csv")
```

Property Type

Similarly, for the property type - similar kinds of values were replaced with their equivalent values. In the end, we just have four property type values; this will make a better-quality visualization.

Figure 16

Data Cleaning – Replacing column values (Property Type)

```
#Data Cleanup for Property type
Listings_df["property_type"] = Listings_df["property_type"].apply(lambda x: x.replace("Private room in ", ""))
Listings_df["property_type"] = Listings_df["property_type"].apply(lambda x: x.replace("Shared room in ", ""))
Listings_df["property_type"] = Listings_df["property_type"].apply(lambda x: x.replace("Entire ", ""))
Listings_df["property_type"] = Listings_df["property_type"].apply(lambda x: x.replace("Room in ", ""))
Listings_df['property_type'] = Listings_df['property_type'].str.capitalize()
Listings_df.property_type.value_counts(normalize=True)
```

Property Type	Percentage
Apartment	0.654010
House	0.218700
Townhouse	0.032600
Condominium	0.026877
Serviced apartment	0.016894
Loft	0.008820
Bed and breakfast	0.008428
Boutique hotel	0.006415
Guest suite	0.004704
Hotel	0.004207
Guesthouse	0.003854
Hostel	0.003358
Bungalow	0.001855
Aparthotel	0.001633
Private room	0.001372
Cottage	0.001084
Tiny house	0.000915
Boat	0.000888
Villa	0.000601
Houseboat	0.000353
Home/apt	0.000327
Place	0.000314
Earth house	0.000235
Cabin	0.000235
Camper/rv	0.000196
Chalet	0.000157
Hut	0.000131
Floor	0.000118
Yurt	0.000091
Casa particular	0.000078
Farm stay	0.000078
Shared room	0.000065
Barn	0.000065
Dome house	0.000052
Island	0.000039
Campsite	0.000039
Lighthouse	0.000039
Treehouse	0.000039
Tent	0.000026
Bus	0.000026
Castle	0.000026
Shepherd's hut	0.000013
Minsu	0.000013
Parking space	0.000013
Resort	0.000013
Name: property_type, dtype: float64	

```

#Data Cleanup for Property type
Listings_df.property_type.replace({
    'Condominium' : 'Apartment',
    'Loft' : 'Apartment',
    'Serviced apartment': 'Apartment',
    'Townhouse': 'House',
    'Bed and breakfast' : 'Hotel',
    'Guesthouse' : 'Guest suite',
    'Boutique hotel': 'Hotel',
    'Bungalow': 'House',
    'Cottage': 'House',
    'Boat': 'House',
    'Camper/RV': 'Other',
    'Earth house': 'House',
    'Houseboat': 'House',
    'Pension (South Korea)' : 'Other',
    'Tiny house' : 'House',
    'Aparthotel' : 'Hotel',
    'Cabin' : 'Other',
    'Casa particular (Cuba)' : 'Other',
    'Castle' : 'Other',
    'Barn' : 'Other',
    'Chalet' : 'Other',
    'Island' : 'Other',
    'Tipi' : 'Other',
    'In-law' : 'Other',
    'Cave' : 'Other',
    'Train' : 'Other',
    'Treehouse' : 'House',
    'Tent' : 'Other',
    'Villa' : 'House',
    'Resort' : 'Other',
    'Hostel' : 'House',
    'Guest suite' : 'Hotel',
    'Home/apt' : 'Apartment',
    'Private room' : 'Hotel',
    'Guest suite' : 'Hotel',
    'Shared room' : 'Hotel',
    'Dome house' : 'House',
    'Lighthouse': 'House',
    'Bus' : 'Other',
    'Minsu': 'Other',
    'Shepherd\'s hut': 'Other',
    'Parking space': 'Other',
    'Place': 'Other',
    'Hut': 'Other',
    'Floor': 'Other',
    'Yurt': 'Other',
    'Casa particular': 'Other',
    'Farm stay': 'Other',
    'Campsite': 'Other',
    'Camper/rv': 'Other',
}, inplace=True)

Listings_df.property_type.value_counts(normalize=True)

```

Apartment	0.706928
House	0.260721
Hotel	0.026825
Guest suite	0.003854
Other	0.001672

Name: property_type, dtype: float64

Changing the Data Types

First_Review and Host_Since

For these two features, the data type has been converted to DateTime.

Figure 17

Data Cleaning – Changing the Data Types

```
#Data Cleanup for First_Review, Host_Since  
Listings_df.first_review = pd.to_datetime(Listings_df.first_review)  
Listings_df.host_since = pd.to_datetime(Listings_df.host_since)
```

Removing special characters

Host Response Rate

The Host Response Rate column was converted to numeric values by removing the % value in this feature. By doing these it will be easy for calculating the host response rate.

Figure 18

Data Cleaning – Removing the Special Characters

```
#Data Cleanup for host_response_time  
#Listings_df.host_response_time.fillna("N/A", inplace=True)  
#Listings_df.host_response_time.value_counts(normalize=True)  
  
Listings_df['host_response_rate'] = pd.to_numeric(Listings_df.host_response_rate.str.strip('%'))  
Listings_df["host_response_rate"] = Listings_df["host_response_rate"].dropna()  
Listings_df["host_response_time"] = Listings_df["host_response_time"].dropna()
```

Dropping Duplicates

Dropping any duplicate rows based on the listing id.

Figure 19

Data Cleaning – Dropping Duplicates

```
Listings_df = Listings_df.drop_duplicates(subset=['id'], keep='last')
```

Remaining Null Values are filtered in Tableau directly while plotting

Design Principles

Below are the design principles followed while developing an interactive visualization

- Created the below Interactivity in the dashboards
 - Highlighting & Highlight Actions for the listing images
 - Web URL Actions for the listing images as seen in Figure 23 and 26
 - Filtering & Filter Actions for the Neighbourhoods, Room Types, Year and Price Range
 - Tooltips and Dashboard Navigations
- Removed all the Null and Duplicated values for a better plot.
- Followed the ETL process of extracting, transforming, and loading the cleansed data in Tableau.
- Used fixed size for all the dashboards (1000 x 800)
- Font size, boldness, borders are maintained the same in all the dashboards
- Tooltips are formatted and highlighted the important values
- For each dashboard a title is given properly followed by the subtitle
- Color is maintained throughout the dashboards (Red and White)
- Gradient color is used for the charts and maps visualization
(for higher values, it appears dark and for lower values, it appears light)
- Used different type of charts by following the chart guidelines
- Consistent with labeling all the axes and titles
- Formatting is done to look the dashboard look neater
- Dark grids were removed in all the graphs and applied transparent background in all the graphs

Dashboard Visualization

Overview of Airbnb – London

The first dashboard depicts the “*Overview of Airbnb*” in London as shown in Figure 21.

This dashboard gives the overall summary of the Airbnb listing dataset.

The dashboard displays the below statistics of the Listings in London. These values can be filtered based on each neighbourhood.

- Number of listings
- Average Price
- Average Rating
- Number of Reviews per month
- Hosts

From the visualization, we can infer that, (*for all the 33 neighbourhoods*)

- There are around 76534 listings in London
- The average rating of the listings is 4.79
- There are around 31064 reviews per month
- The average rate of the listings is £113.90
- There are around 50,764 unique hosts who listed their properties in Airbnb London
- There are 33 neighbourhoods in London where Airbnb has been listed

A map visualization has been plotted on Airbnb Neighbourhoods based on the Average Ratings. Even from the map, an action has been created for filtering the other values based on the selected Neighbourhoods.

- “Richmon upon Thames” borough has the highest average ratings of about 4.89

Figure 20

Map Visualization showing London Airbnb Neighbourhoods

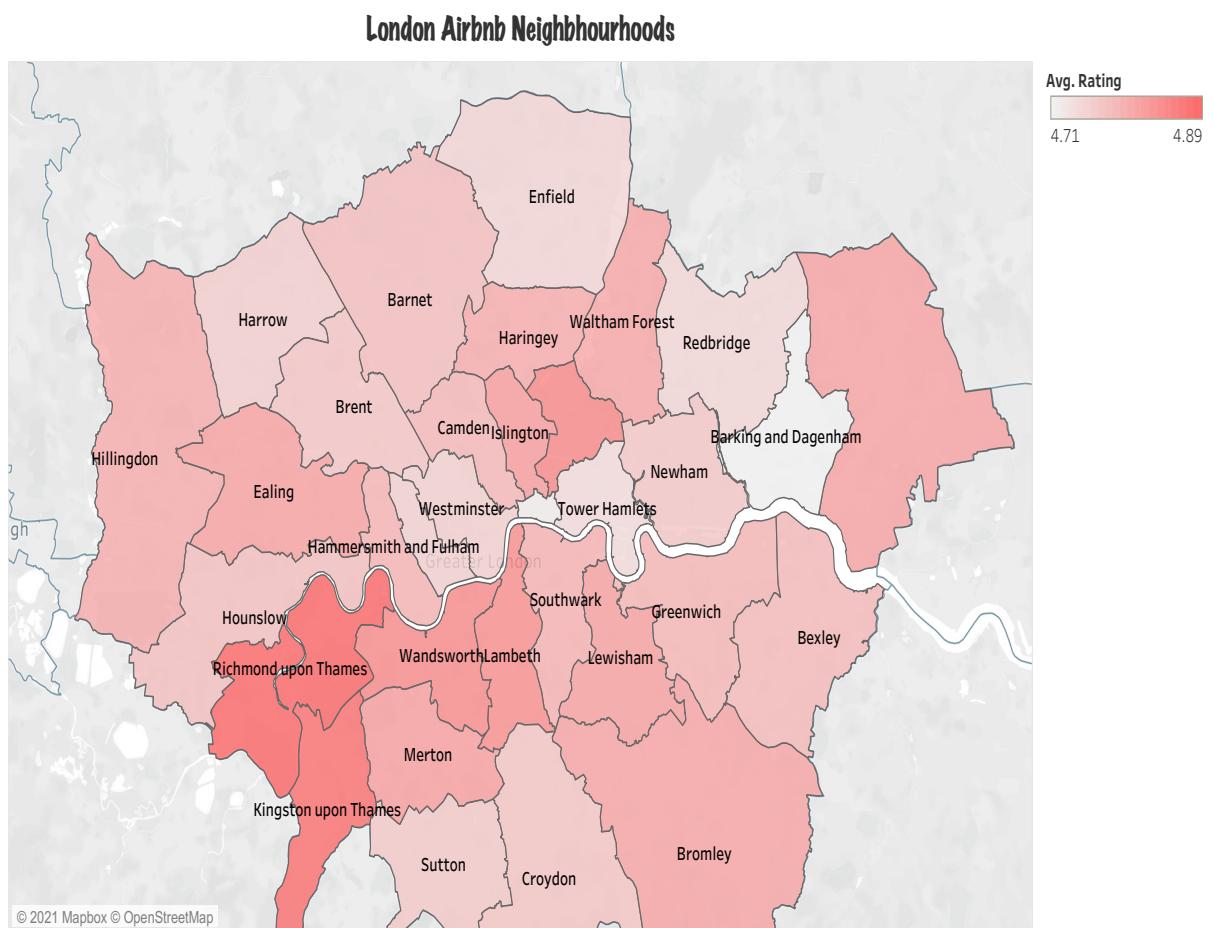
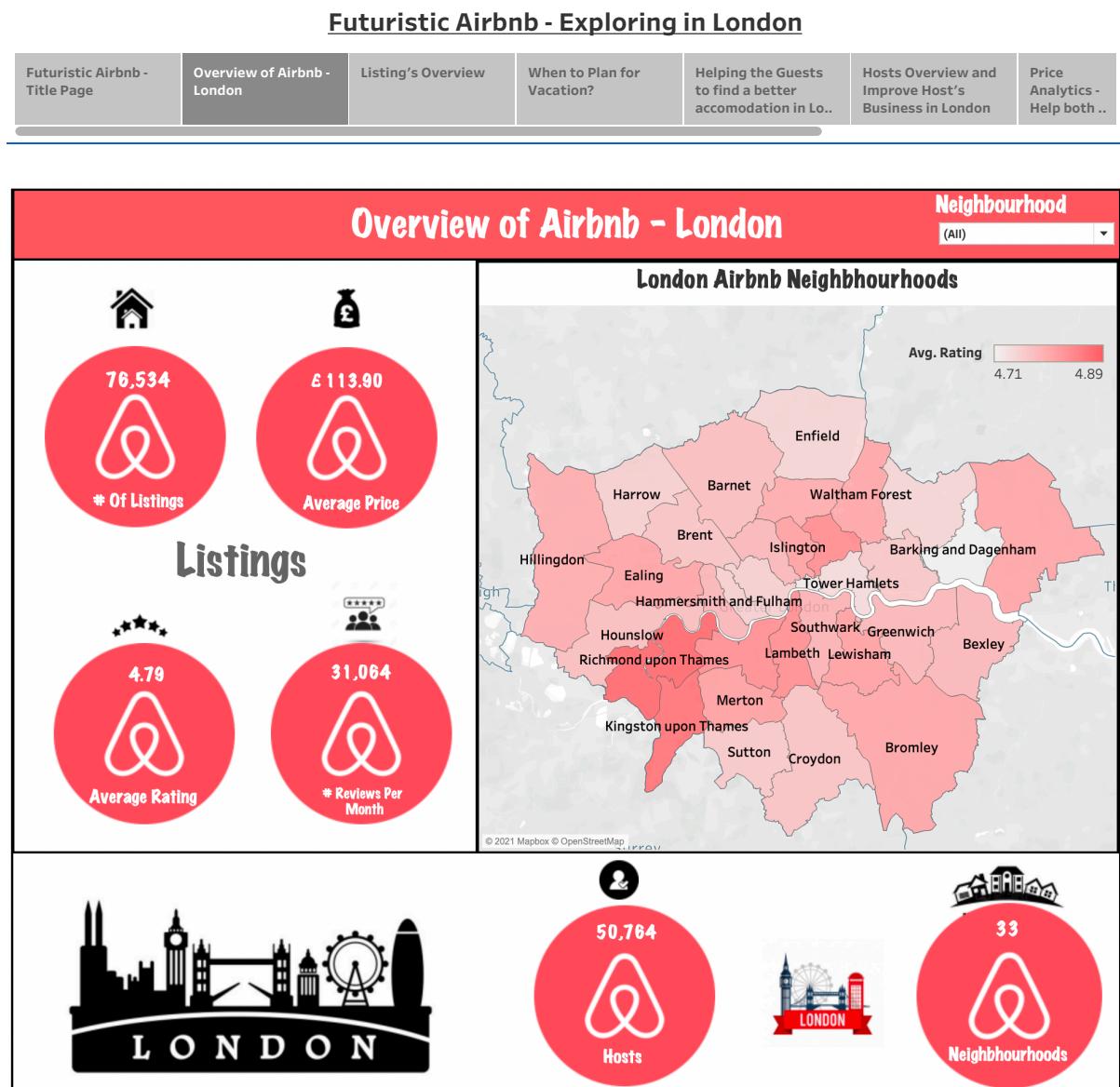


Figure 21

Dashboard – Overview of Airbnb London



[Listing's Overview](#)

The “*Listing’s Overview*” dashboard contains the below three visualizations as shown in Figure 25.

- Yearly Listing’s Count
- Property Type
- Room Type

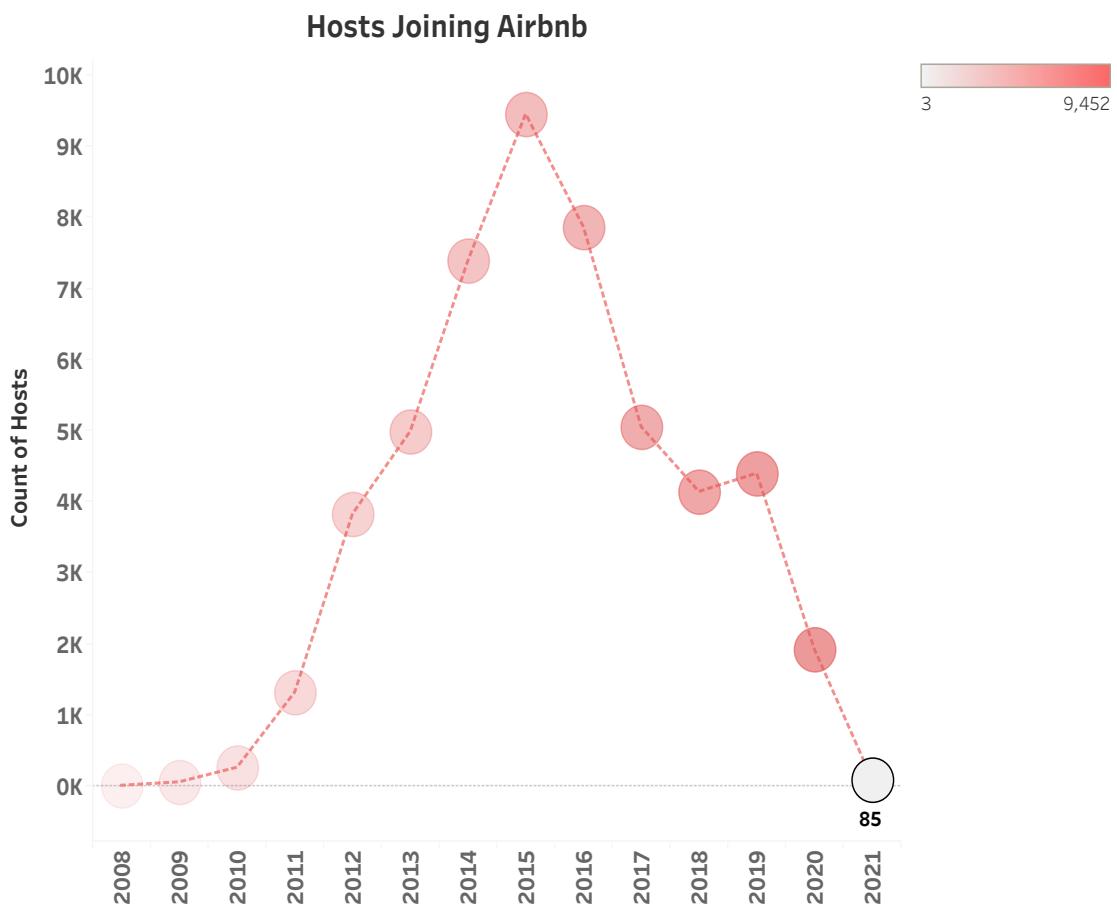
Yearly Listing’s Count

This visualization contains the listing count of London from 2008 to 2021. A line(discrete) graph has been plotted for visualizing this scenario. The x-axis represent the years and the y-axis represent the number of listings. The year column has been used in the Pages shelf to break the view into a series of pages by year-wise such that we can analyze the listing count year-wise. From the graph we can infer that,

- The listings count was increasing as we go from 2008 until 2015, then we see a drop in the listing’s count from 2016 onwards
- In 2020 & 2021, there was drastic drop down on the listing’s count due to the covid pandemic. People were not planning for any vacation during the covid crisis; hence the listings have been reduced drastically.

Figure 22

Visualization – Hosts Joining Airbnb



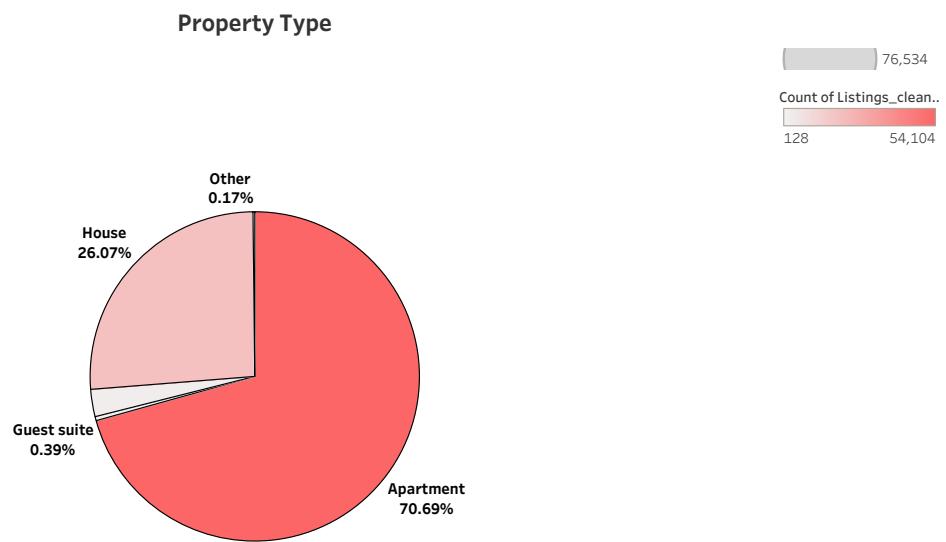
Property Type

This visualization contains the statistics details of the Property Type. A pie chart has been plotted to visualize the Property Type details. From the Graph, we can infer that,

- There are four categories of property types: Apartment, House, Guest Suite and Other
- Most of the Airbnb Property types in London belongs to Apartment of about 70.69% followed by House of about 26.07%

Figure 23

Visualization – Property Type



Room Type

This visualization contains the statistics details of the Room Type. A pie chart has been plotted to visualize the Room Type details. From the Graph, we can infer that,

- There are four categories of Room types: Entire home/apt, Private room, Shared room, Hotel room
- Most of the Airbnb Room types in London belongs to Entire home/apt of about 54.40% followed by Private room of about 44.08%

Figure 24

Visualization – Room Type

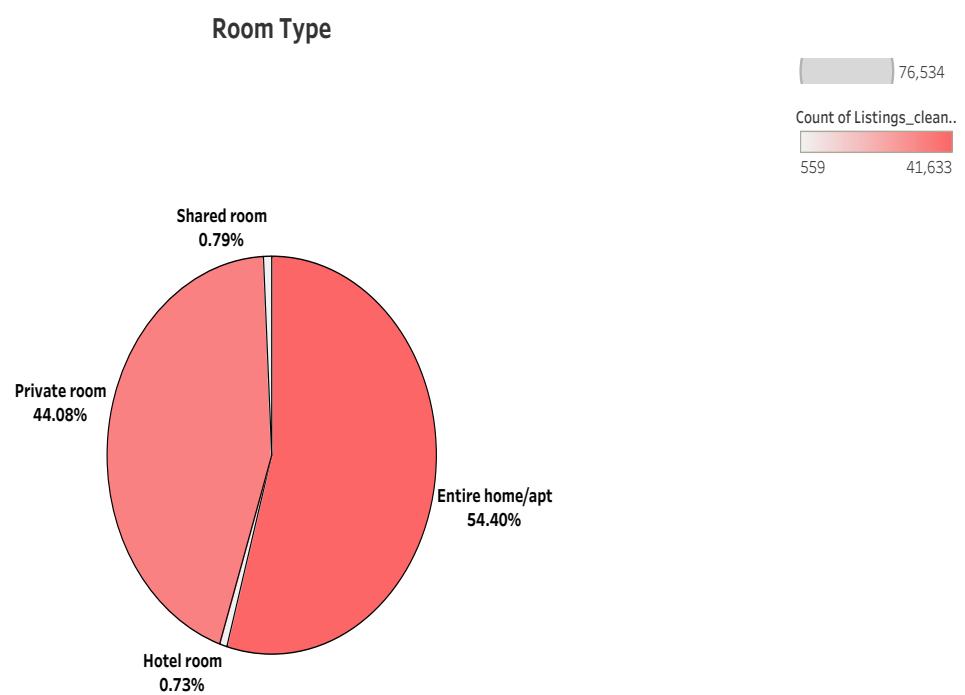
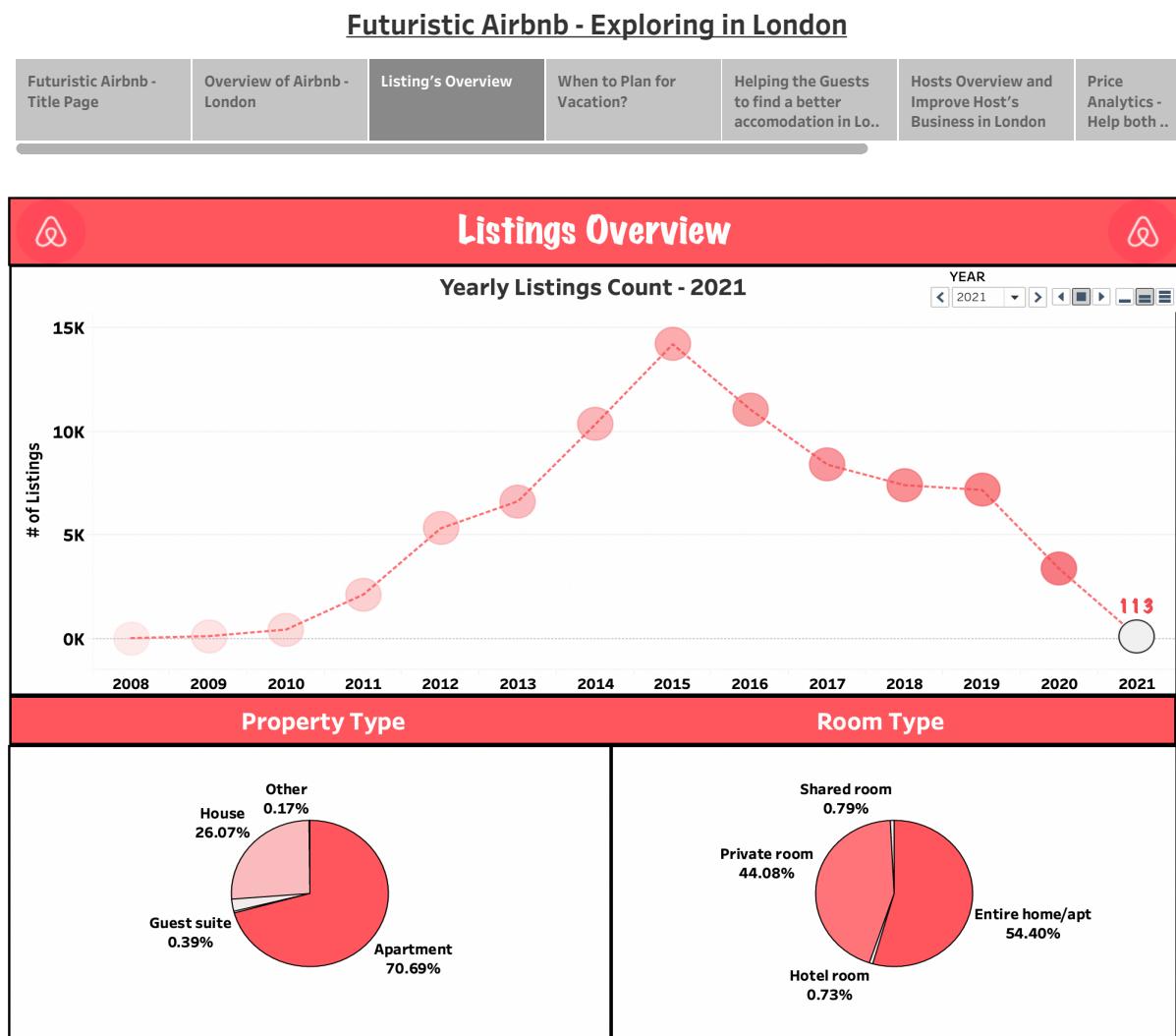


Figure 25

Dashboard – Listing's Overview



When to Plan for Vacation?

The “*When to Plan for Vacation?*” dashboard contains the below two visualizations as shown in Figure 28.

- Finding Busiest Month
- Most/Least Expensive Time to visit London

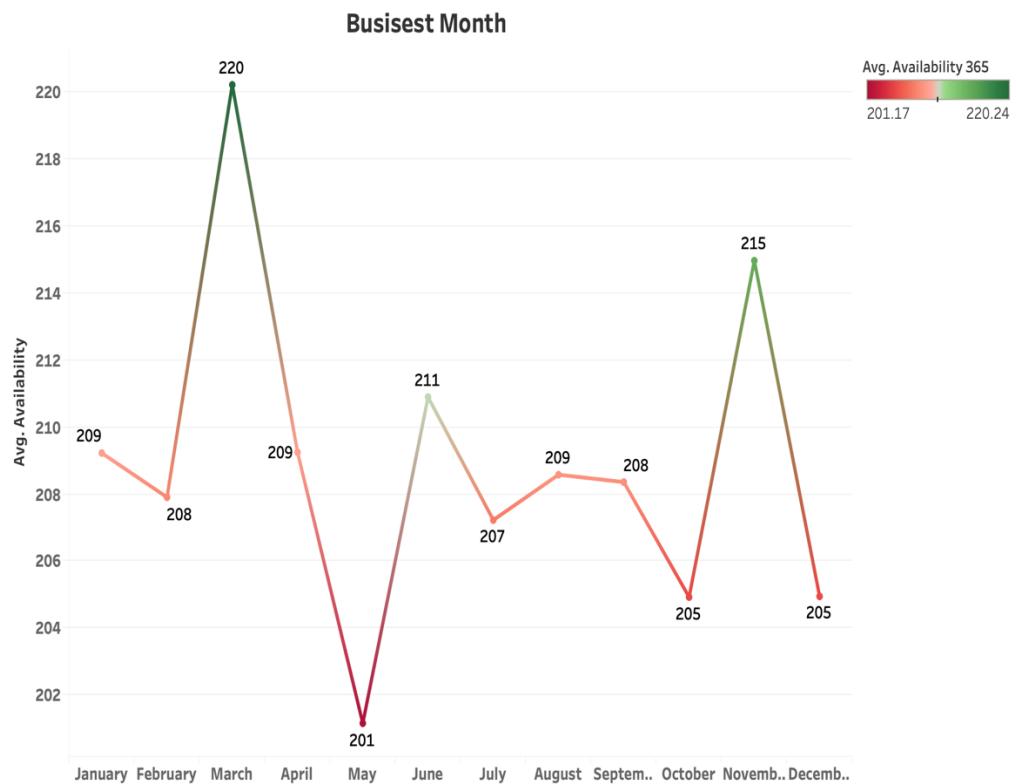
Finding the Busiest Month

This visualization helps in determining the peak vacation month in London. A line(discrete) chart has been plotted for visualizing this scenario. The x-axis represents the Month and the y-axis represent the Average availability. From the Graph, we can infer that,

- The peak vacation month is seems to be May since you can see the average availability of getting a property is very less.
- March month being the off-season to travel London
- Looks like the best time to visit London is during its summer season (April – June)
- To plan a trip for a cheaper accommodations (the time when fewer people travel)

Figure 26

Visualization – Busiest Month



Most/Least Expensive Time to Visit London

This visualization contains the average price range for each weekday – Month wise. A highlighting table has been plotted to visualize this scenario. From the tables, we can infer that,

- The rates start to increase from May throughout December month. December being the costliest month
- To plan a trip for a cheaper accommodations (the time when the average price is less)

Figure 27

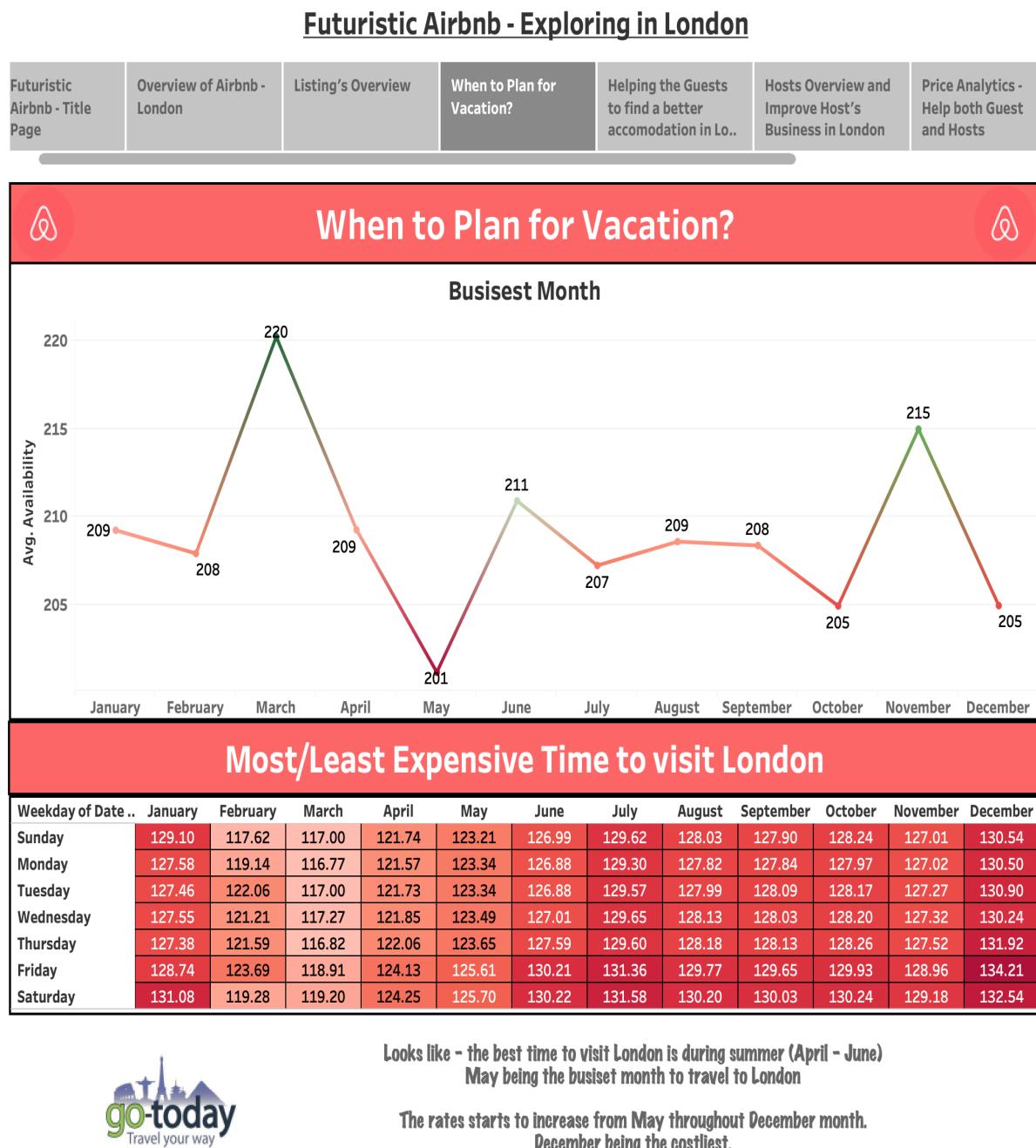
Visualization – Most/Least Expensive Time to Visit London

Plan

Weekday of Date..	January	February	March	April	May	June	July	August	September	October	November	December	Avg. Price (Calendar.Cs..)
Sunday	129.10	117.62	117.00	121.74	123.21	126.99	129.62	128.03	127.90	128.24	127.01	130.54	116.77 134.21
Monday	127.58	119.14	116.77	121.57	123.34	126.88	129.30	127.82	127.84	127.97	127.02	130.50	
Tuesday	127.46	122.06	117.00	121.73	123.34	126.88	129.57	127.99	128.09	128.17	127.27	130.90	
Wednesday	127.55	121.21	117.27	121.85	123.49	127.01	129.65	128.13	128.03	128.20	127.32	130.24	
Thursday	127.38	121.59	116.82	122.06	123.65	127.59	129.60	128.18	128.13	128.26	127.52	131.92	
Friday	128.74	123.69	118.91	124.13	125.61	130.21	131.36	129.77	129.65	129.93	128.96	134.21	
Saturday	131.08	119.28	119.20	124.25	125.70	130.22	131.58	130.20	130.03	130.24	129.18	132.54	

Figure 28

Dashboard - When to Plan for Vacation?



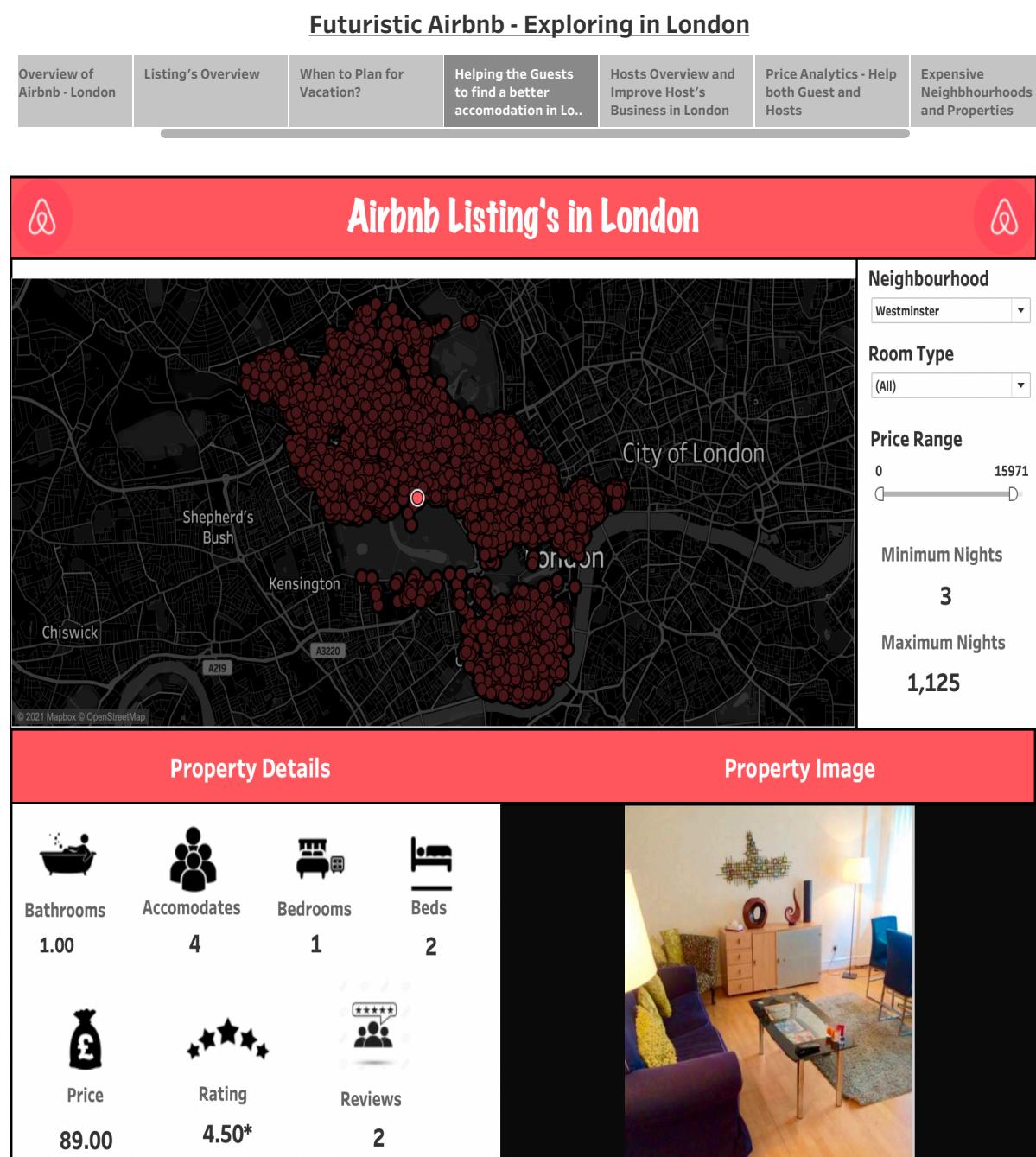
Airbnb's Listing in London

The dashboard depicts the “*Airbnb’s Listings*” in London which helps in finding better accommodation in London as shown in Figure 29. It visualizes the Listings using the map visualization – Neighbourhood wise (at a time only one neighbourhood listings will be displayed on the map for better visualization). The map can be filtered based on Neighbourhood, Room Type, and Price Range. Once the required listing is selected on the map; the dashboard will display the following statistics based on the property chosen.

- Minimum and Maximum Nights
- Bathrooms
- Accommodates
- Bedrooms
- Beds
- Price
- Ratings
- Review
- Property Image

Figure 29

Dashboard - Airbnb's Listings in London



Hosting's Overview and Improve Host Business in London

The “*Hosting’s Overview and Improve Host Business in London*” dashboard contains the below five visualizations as shown in Figure 35.

- Hosts joining Airbnb
- Host Response Time
- Judging Host Response and Host Acceptance Rate based on Super host
- Host Verification
- Top 10 Amenities

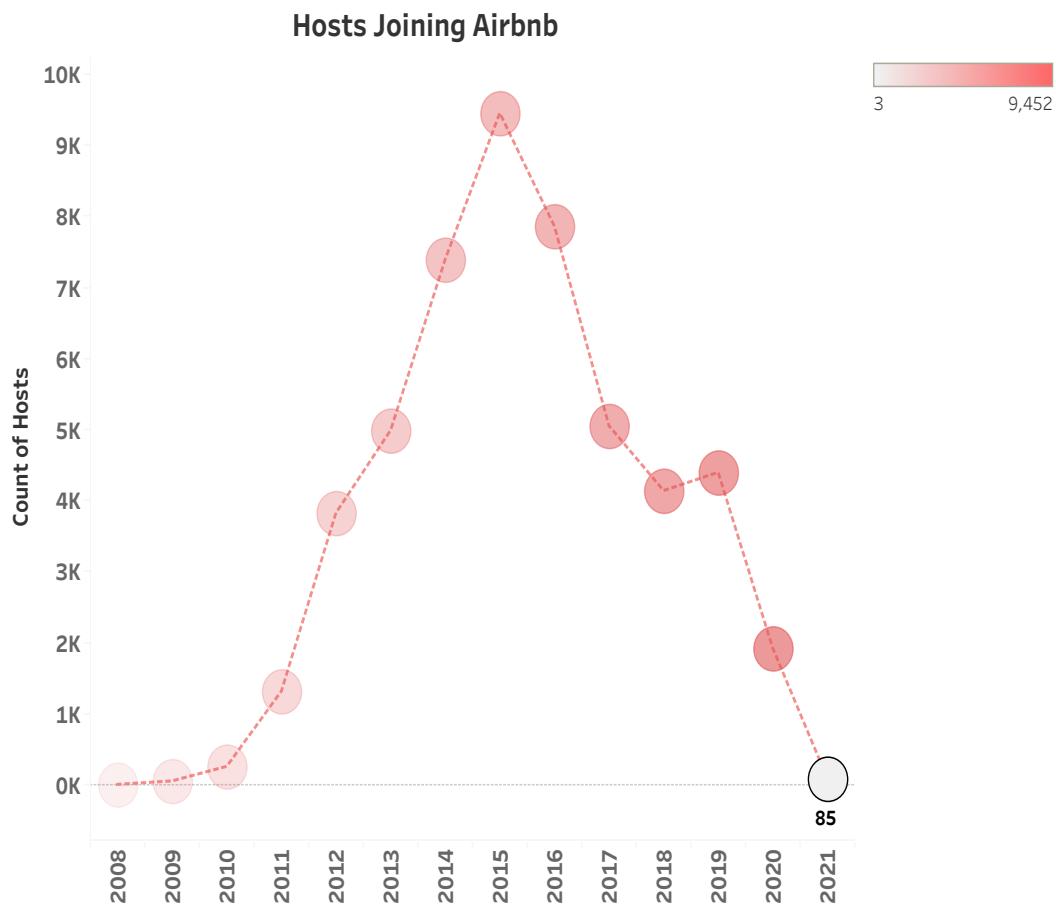
Host joining Airbnb

This visualization contains the number of hosts joining London Airbnb from 2008 to 2021. A line(discrete) graph has been plotted for visualizing this scenario. The x-axis represents the years and the y-axis represents the number of listings. The year column has been used in the Pages shelf to break the view into a series of pages by year-wise such that we can analyze the hosting count year-wise. From the graph we can infer that,

- The hosts joining Airbnb count was increasing as we go from 2008 until 2015, then we see a drop in the listing’s count from 2016 onwards
- In 2020 & 2021, there was a drastic drop down in the hosts count due to the covid pandemic. People were not planning for any vacation during the covid crisis; hence the hosts joining and putting listings on Airbnb has been reduced drastically.

Figure 30

Visualization – Hosts Joining Airbnb



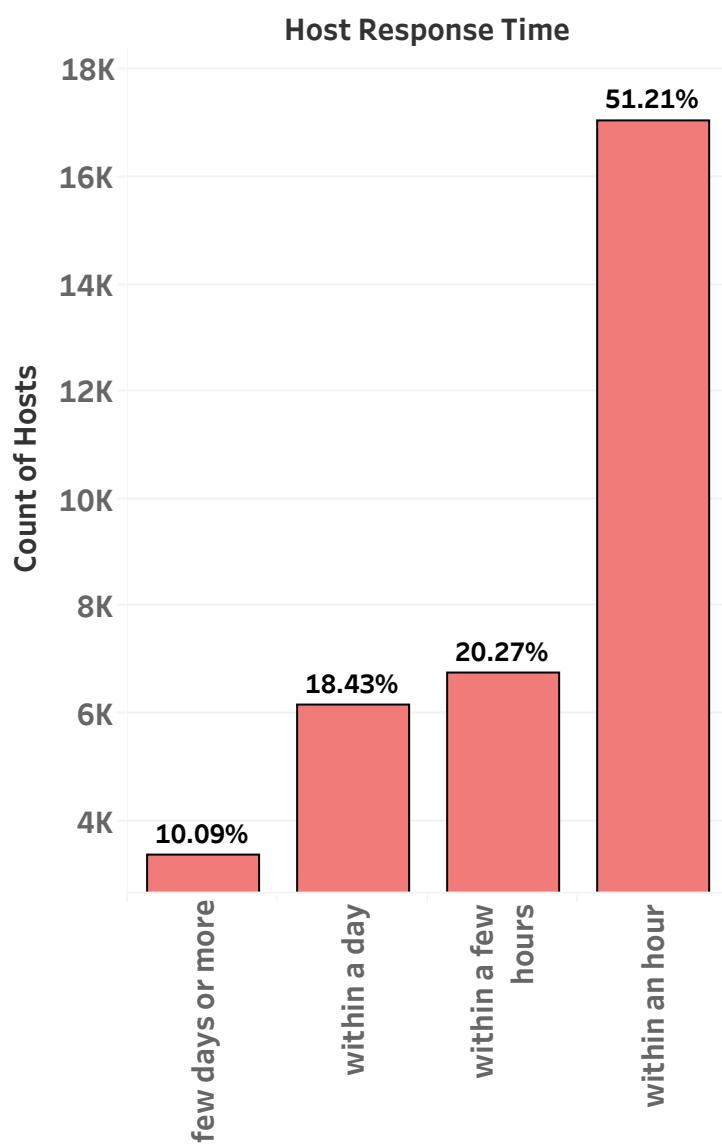
Host Response Time

This visualization contains the statistics on the Host Response Time. A bar chart has been plotted for visualizing this scenario. The x-axis represents the Host Response Time values and the y-axis represent the Count of hosts. From the Graph, we can infer that,

- About 51% of the hosts reply within an hour and 20% of the hosts reply within few hours.
- From the Guest Perspective, For knowing about the best hosts for ease of communication
- From the Host Perspective, For improving their response rate to further attract the customers

Figure 31

Visualization – Host Response Time



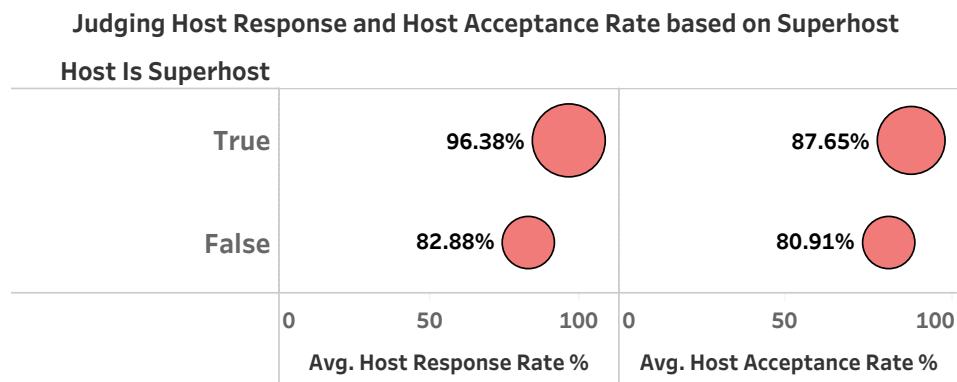
Judging Host Response and Host Acceptance Rate based on Super host

This visualization displays the Average Host Response and Host Acceptance Rate. A scatter plots have been designed to visualize these rates percentage based on the Super Host feature. From the Graph, we can infer that,

- Super Hosts Average Response Rate % (96.38%) is more than the normal hosts (82.88%)
- Super Hosts Average Acceptance Rate % (87.65%) is more than the normal hosts (80.91%)

Figure 32

Visualization – Judging Host Response and Host Acceptance Rate based on Superhost



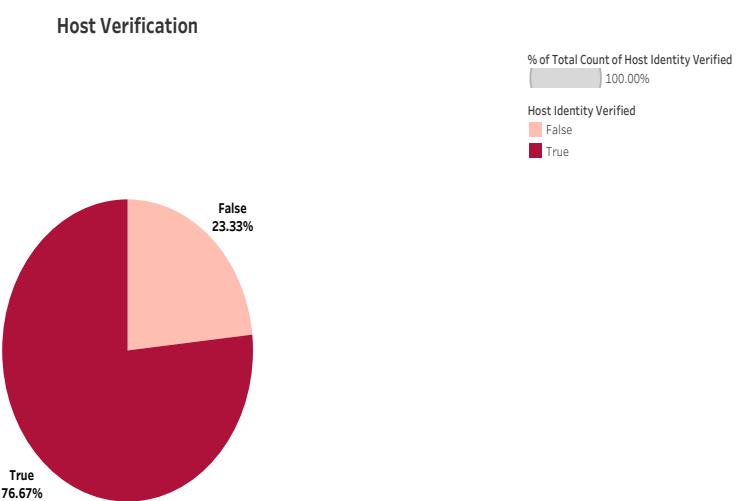
Host Verification

This visualization contains the status of the Host Verification whether the host has been verified or not. A pie chart has been plotted to visualize the Host Verification feature. From the Graph, we can infer that,

- Around 76.67% of hosts have been verified. Verifying the details is something hosts often think they are expected to do
- Guests might expect the additional trust commanded by verified hosts for their accommodation

Figure 33

Visualization – Host Verification



Top 10 Amenities

This visualization contains the Top 10 Amenities – Wifi, Washer, TV, Smoke alarm, Kitchen, Iron, Heating, Hangers, Hairdryers, and Essentials. A text table using the mark *Circle* has been plotted to visualize the Top 10 Amenities of the properties. From the Graph, we can infer that,

- Wi-Fi, Heating, Essentials, Kitchen and Washer are the most important amenities to be considered for the successful of listing.
- For those who are already in the business and don't have those amenities, make sure the facility of installing all or some of the amenities to grow your business

Figure 34

Visualization – Top 10 Amenities

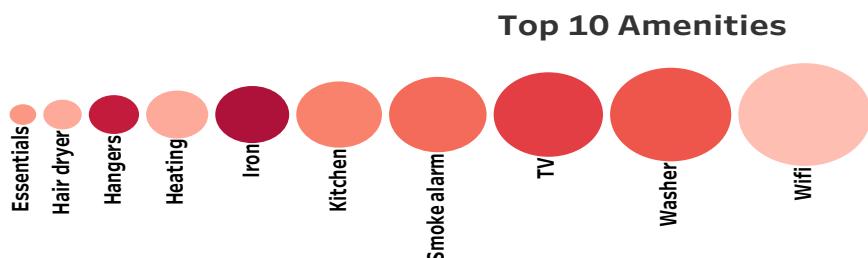
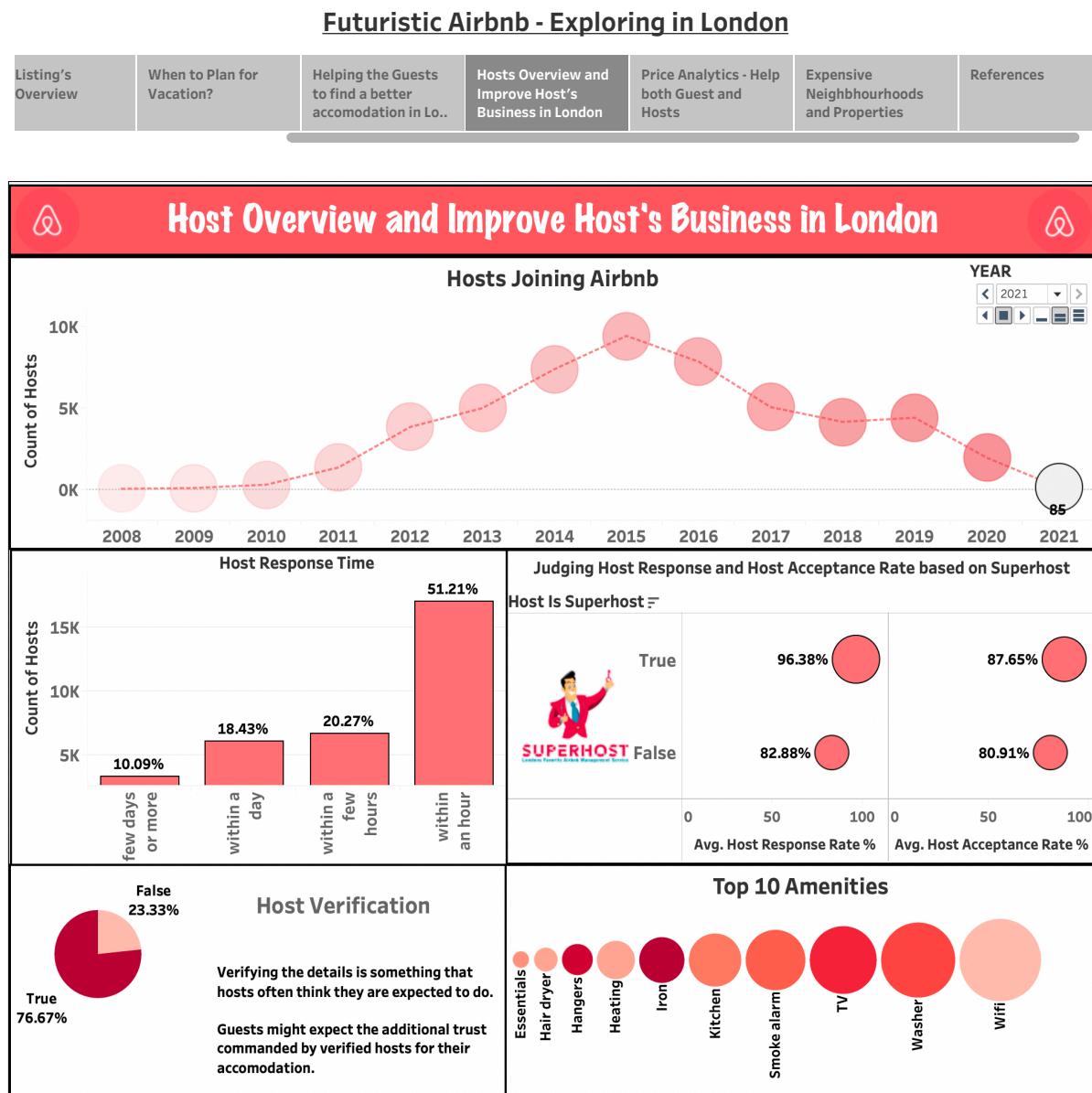


Figure 35

Dashboard - Host Overview and Improve Host's Business in London



Price Analytics

The “*Price Analytics*” dashboard contains the below three visualizations as shown in Figure 39.

- How have prices changed over time?
- Accommodates vs Average Price
- Future Price Prediction based on Room Type

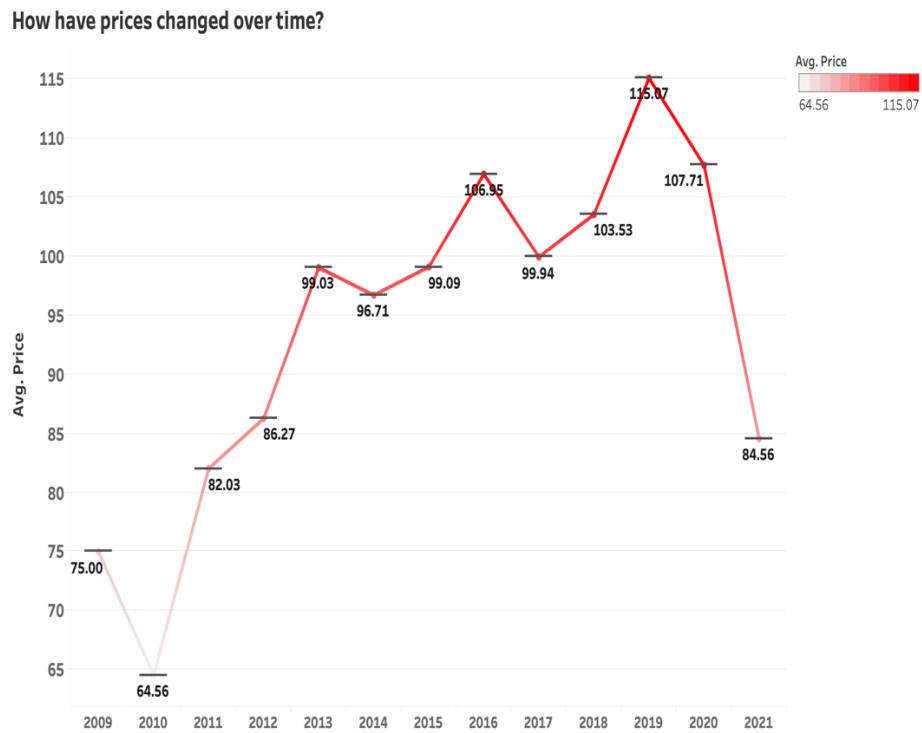
How have prices changed over time?

This visualization shows the average price range change from 2008 to 2021. A line(discrete) graph has been plotted for visualizing this scenario. The x-axis represent the years and the y-axis represent the average price. From the graph, we can infer that,

- The average price has been keeping on fluctuating – going up and down throughout the years
- In 2019, the average price of the listing was high around £115.07. But in 2021 the listing price dropped to £84.56 due to the covid pandemic. People were not planning for any vacation during the covid crisis; hence the listings price has been reduced drastically.

Figure 36

Visualization – How have prices changed over time?



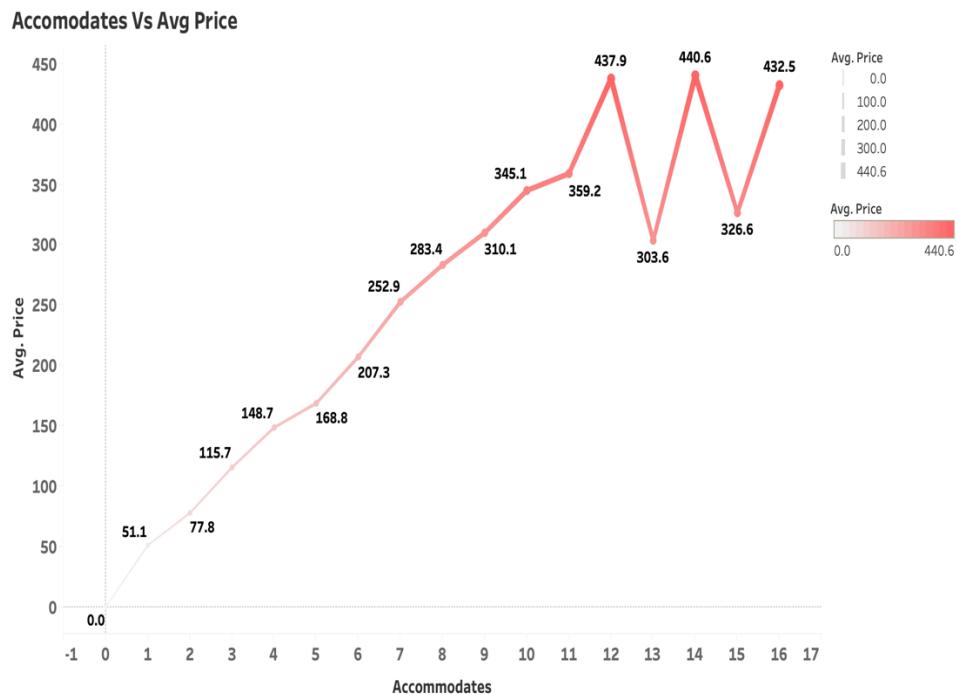
Accommodates Vs Average Price

This visualization shows the average price based on the accommodates. A line(discrete) graph has been plotted for visualizing this scenario. The x-axis represent the accommodates and the y-axis represents the average price. From the graph, we can infer that,

- It is evident that - As the accommodates increases, the average prices also increases
- Though, there was a drop in the average price for the accommodates 13 and 15. (almost £100 difference range)

Figure 37

Visualization – Accommodates vs Average Price



Future Price Prediction based on Room Type

This visualization shows the future price prediction based on the Room Type. A box and whisker plot has been designed for visualizing this scenario. The x-axis represent the current and the future year. The y-axis represents the average price. From the graph, we can infer that,

- For the Guest Suite and the House Room type, the future predicted price is almost the same for the years 2021 and 2022.

- For the Apartment Room type, there is a slight increase in the future predicted price of about £3
- For the Hotel Room type, the future predicted price is around £180.53 which is like a huge increase of about £17

Figure 38

Visualization – Future Price Prediction

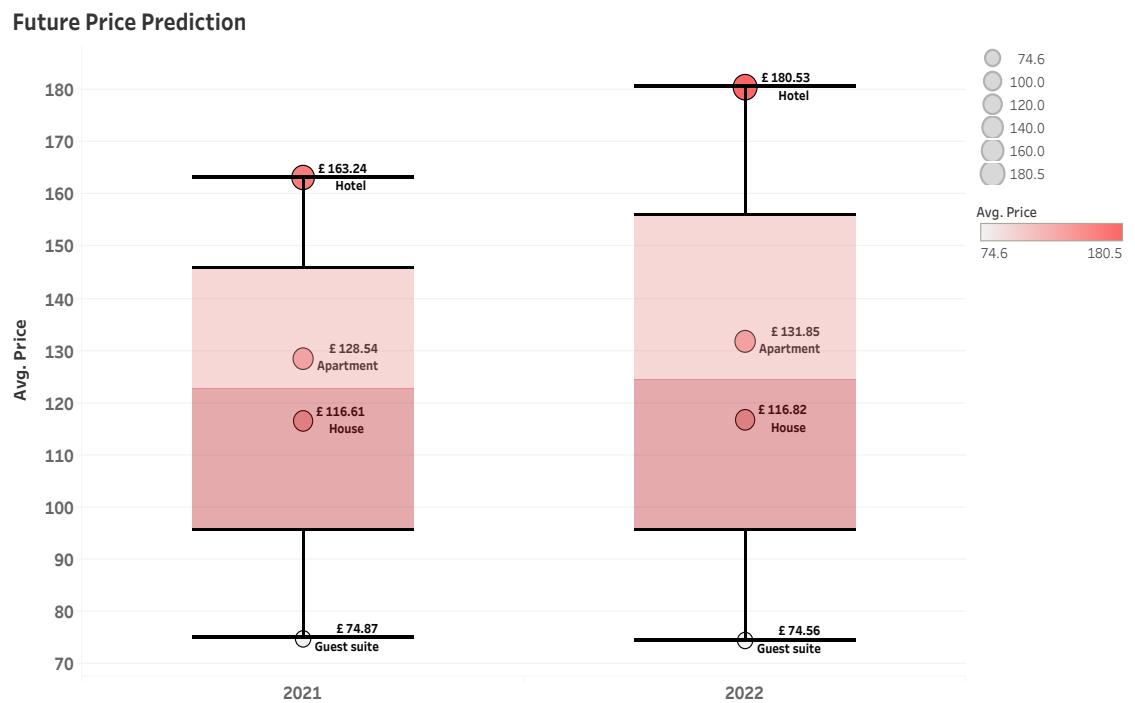
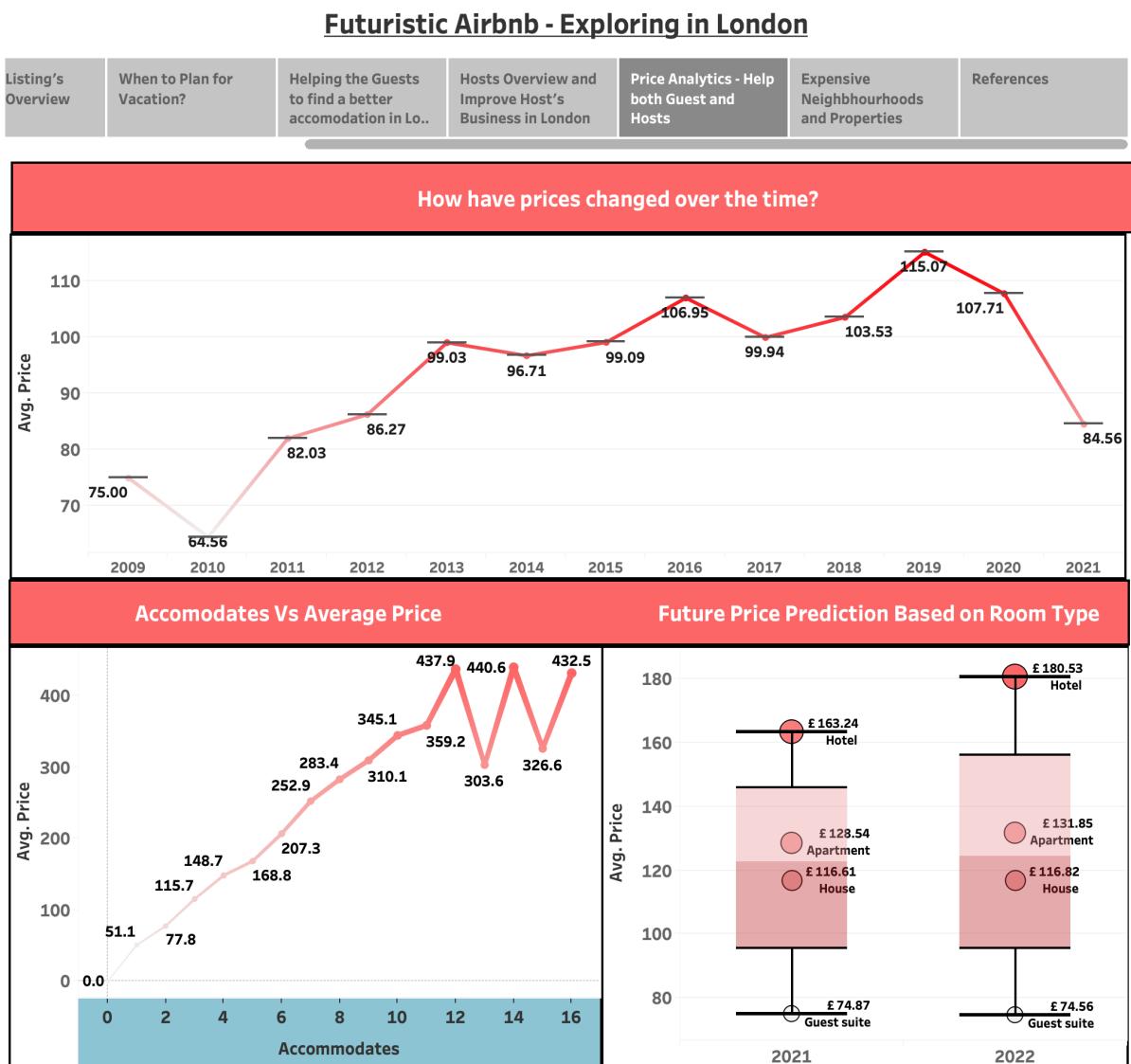


Figure 39

Dashboard - Price Analytics



Most Expensive Properties and Neighbourhoods

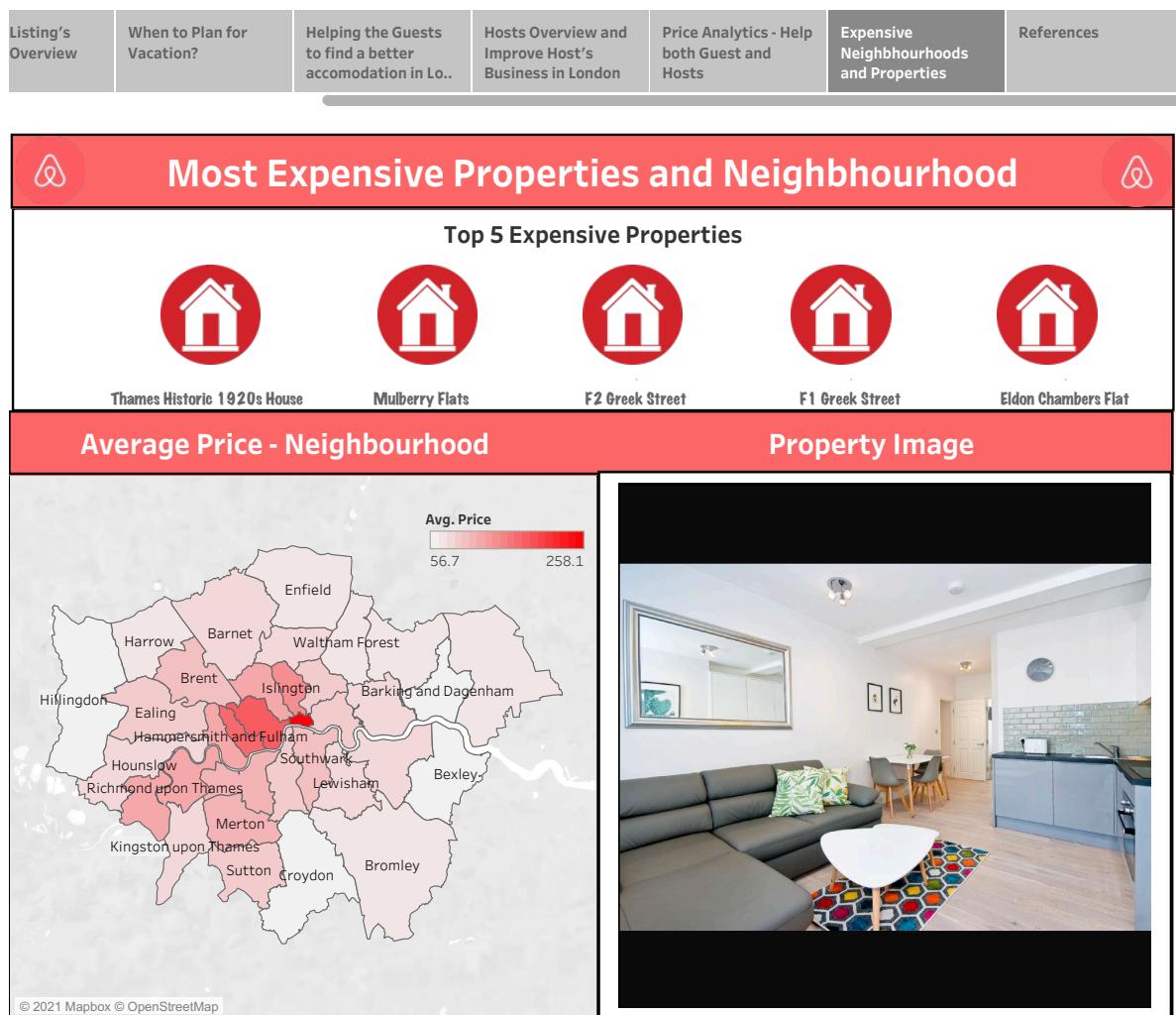
This dashboard displays the most expensive properties and Neighbourhoods of London as shown in Figure 40. The top 5 expensive properties are plotted using the bar charts. The x-axis represents the property's name and the y-axis represents the average price. From the Graph, we can infer that the top 5 properties are – *Thames Historic 1920s House, Mulberry Flats, F2 Greek Street, F1 Greek Street, and Eldon Chamber Flats*. The

Property images also displayed in the dashboard when we select the property. The expensive Neighbourhoods are plotted using the map visualizations based on the average price feature. The darker the place (highlighted in red); the more expensive the place is. From the Average Price – Neighbourhood Map visualization, we can infer that City of London is the most expensive neighbourhood with about £258.1 rate followed by Westminster of about £183.5 rate.

Figure 40

Dashboard - Expensive Neighbourhoods and Properties

Futuristic Airbnb - Exploring in London



Story Telling

The dashboard is arranged by creating a story of Airbnb. Below are the stories illustrated :

- The first story describes the Title Page of Futuristic Airbnb
- The second story illustrates the overview summary of the Airbnb – London
- The third story portrays the listings' and property overview
- The fourth story tells when to plan for vacation for the guests?
- The fifth story describes in helping the guests to find better accommodation in London
- The sixth story depicts the Host's overview and improve hosts business in London
- The seventh story performs the price analytics which will help both the guests and Hosts
- The eight story displays the expensive neighbourhoods and properties

The story depicts the title to introduce the topic of the story as Futuristic Airbnb – Exploring in London and then it illustrates the overview summary of the Airbnb in London in the second story. Then, it portrays the overview of the listing and property details in the third story. Once we get an idea of the Airbnb listings and property, we plan for the vacation based on our budget and decide when to travel based on the fourth story? Once you have decided on your travel, the fifth story will help the guest to choose the best accommodation in London based on their requirements. Once the listing has been selected, you can check for the sixth story to check the host acceptance and response rate accordingly. This story also helps the hosts to improve their business in London. The final story is just to display the expensive neighbourhoods and properties if you wanted to explore them.

Takeaways/Outcomes

- **From Guest Perspective :**

- ✓ Renting Price based on the Accommodates

Can decide on the rent prices based on the number of accommodates and Choosing the best neighbourhood place according to their budget by filtering the neighbourhood

- ✓ Finding the Busiest Month by Listing Availability? When to plan for vacation?

To plan a trip for cheaper accommodations (the time when fewer people travel)

Best time to visit London (April and May – Summer Season)

May being the Peak Season

- ✓ Judging the Host Response and Acceptance Rate?

Knowing about the best hosts for ease of communication

Super Hosts reply very fast than the normal hosts

Super Host accepts the requests sooner than the normal hosts

- ✓ Most and Least Expensive time to visit London

December is being the Costliest time to visit London

January to April (we can see some Minimal Rate)

- **From the Host Perspective,**

- ✓ Amenities

Wi-Fi, Heating, Essentials, Kitchen, and Washer are the most important amenities to be considered for the success of listing.

For those who are already in the business and don't have those amenities, make sure the facility of installing all or some of the amenities to grow your business

- ✓ Renting Price based on the Accommodates

Can decide on the rent prices or rent out the property based on neighbourhoods

- ✓ Judging the Host Response and Acceptance Rate based on Super host?

For improving their response rate to attract the customers in replying to them sooner.

- ✓ Future Price Prediction

For the Hosts to set their prices based on the predicted result

Conclusion & Future Work

From the London Airbnb Data we can conclude the below points :

- The average price of the listings in London are around the rate of £113.90
- The average rating of the listings in London is 4.79 and they get around 31084 reviews per month
- The listings count is very less in London considering the current situation because of the pandemic
- Most of the property and room type belong to Apartment and Entire Home/apt respectively
- The Peak season to travel in London is during their summer season (April and May), where May is the busiest month for travel
- The price range of the listings is very high during December month
- Using Airbnb's listing data, it helps the guests to find better accommodation in London

- The Hosts joining Airbnb count is very less in London considering the current situation because of the pandemic
- Host Response time, Host Acceptance and Response rate are actually good in London Airbnb in the margin of above 80%. Additionally, most of the hosts are verified in the margin of about 77%
- The Price change over time is keep on fluctuating in London. The future price prediction is also high in the case of apartments.

Future Work

- Covering more countries to know the actual Airbnb business strategy
- Work on reviews texts by generating word clouds to know what the customers are actually thinking about it?
- To determine the future predictions based on the neighbourhoods
- To implement more design principles as discussed in the class

Project Code

Github Link:

https://github.com/SaranyaPandiaraj/Futuristic_Airbnb-Exploring_in_London

References

Dataset Resource:

<http://insideairbnb.com/get-the-data.html>

Dashboard Design Resource:

https://help.tableau.com/current/pro/desktop/en-us/buildmanual_shelves.htm

<https://help.tableau.com/current/pro/desktop/en-gb/actions.htm>

https://help.tableau.com/current/blueprint/en-us/bp_visual_best_practices.htm

<https://senturus.com/blog/tableau-dashboard-design-10-best-practices/>

<https://www.tableau.com/learn/articles/data-visualization-tips>

https://help.tableau.com/current/pro/desktop/en-us/what_chart_example.htm

Image Source:

(Dashboard Title Page) <https://www.axisrooms.com/Airbnb-effect-hotel-industry/>

(Icons Used) Google Search In images

General References:

<https://www.airbnb.co.in/>

<https://en.wikipedia.org/wiki/Airbnb>