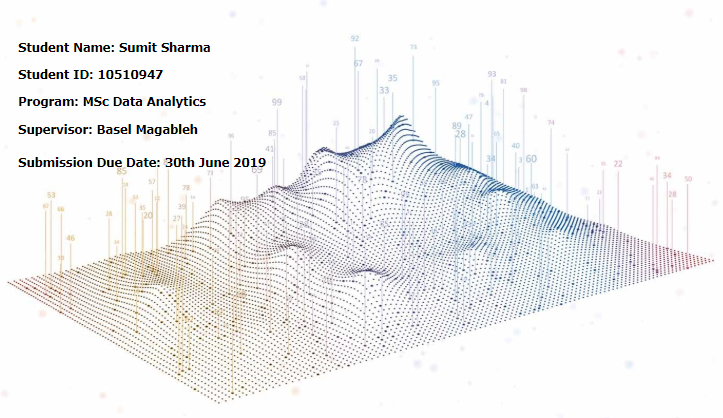
**Data Visualisation**

**Assignment 1**

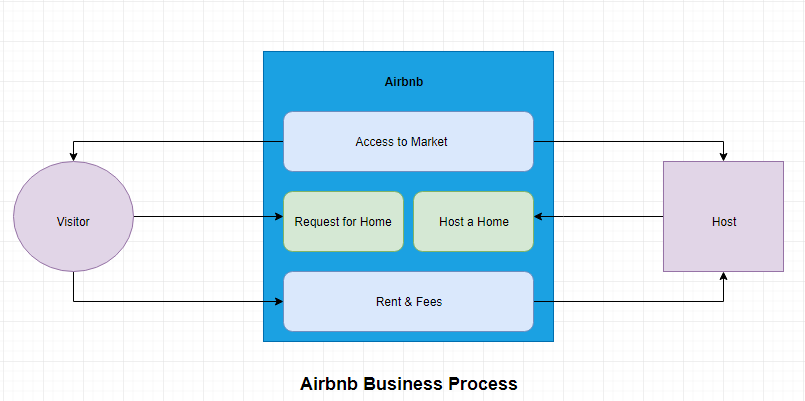
**Exploratory Data Analysis**



# Introduction

**Airbnb** is a platform for renting homes and catalogue homes in details online where it links the local hosts with travellers. Greatest part is traveller can connect host (without owning any homes itself) and have the renting facilities to the core.

Airbnb Flow Diagram

****

[**http://insideairbnb.com/get-the-data.html**](http://insideairbnb.com/get-the-data.html)

[**https://www.dublincity.ie/councilmeetings/documents/s9670/4%20Airbnb\_OverviewforDCC.pdf**](https://www.dublincity.ie/councilmeetings/documents/s9670/4%20Airbnb_OverviewforDCC.pdf)

Above is the link from where the dataset is driven and used for analysing the Airbnb in details, and it is actually expressions Airbnb in details that in which cities around the world with all the features like reviews about house with negative & positive feedback, price range, property type, room type in details, availability and more.

# Details about Datasets and Fields

Dataset ‘**roomAirBnb’** includes the data about host who is hosting a house and the details about property.

**Fields**: host\_id, host\_name, availability\_365, property\_type, room\_type, airbnb\_room\_id, name, bathrooms, bedrooms, price, number\_of\_reviews

Dataset **‘calendarAirBnb’** includes the data about home, amount paid for renting and number of day availability throughout the year.

**Fields**: listing\_id, price, minimum\_nights, maximum\_nights, calendar\_id, date, available.

Dataset **‘reviewsAirBnb’** includes the data about comments given by visitor on property and host and also reviewer details.

**Fields**: listing\_id, id, reviewer\_name, comments, date, reviewer\_id.

Dataset ‘**ListingsPerHostEntireHomes\_Apartment**’ includes the data about host who have listed the entire home and apartment in Dublin.

**Fields:** Host Name, Number of listings.

Dataset ‘**ListingsRecentFrequentBooking’** includes the data of host and about the recent and frequent rooms.

**Fields:** Host Name, Number of listings.

Dataset ‘**ListingsRecentHighlyavailable’** includes the data of host and about the highly available rooms in Dublin.

**Fields:** Host Name, Number of listings.

Dataset ‘**TopHostListingRooms’** includes the data about the host and how many listing each host has.

**Fields:** Host Name, Number of listings, Type.

# Reasons for selecting this subject

The goal is to select this subject area (Airbnb) is to develop a data visualisation and to study and learn about the Airbnb present data for its performance improvement and the services provided by them and how it can be helpful with my investigations.

# Vision & Goals

Vision is to have business perfections which will help the Airbnb organisation and also the quality services to the customers from Airbnb for better choices to selecting their homes as they can set up the areas to investigate and develop a case studies and by this the reports which can clarify improvements and qualities in details.

# Requirements for developing the Visualisations

Below are the tools used to develop the, reports, datasets, design diagrams and visualisations:

1. Microsoft Word
2. Microsoft Excel (.csv)
3. Draw.io
4. Tableau

# Data visualisations based on below analytical questions

1. Which is the most preferred property type booked by visitors based on different selling factors?
2. Which selling factor of the property affects the booking decision of the visitors?
3. Which visitor has spent money more on Airbnb?
4. Most available property in Dublin throughout the year?
5. Which property type has raised more revenue for host in 2019?
6. Who is the top ten host listing of different types of rooms in Dublin?
7. Host who has listed the maximum number of house property (Entire home) in Dublin?
8. Host who has most Recent and Frequent Home?
9. Which host have Highly Available Rooms?
10. Which host with maximum reviews and feedback in Dublin?

# Data quality and issues

As per dataset selection and analysis the shape & structure of my dataset is almost proper in format and quality data driven from the inside-Airbnb site. In driven dataset, variables or field contains data about homes in Dublin, name and details of their host who listed the homes, home details, review of most of the homes. They distributed are in different excel and pdf files.

Pdf file data that is table which was required for my query was extracted into excel file. No, there was not any notable data quality issues found for my dataset as such but there were some surprising relationships among the variables which help in explaining my investigation in proper way.

As researching I got the PDF file in which it contains a data in tabular format and that I must extract into excel file using r script for visualization investigation on Airbnb. [Click here for script](#_Script_for_Pdf).

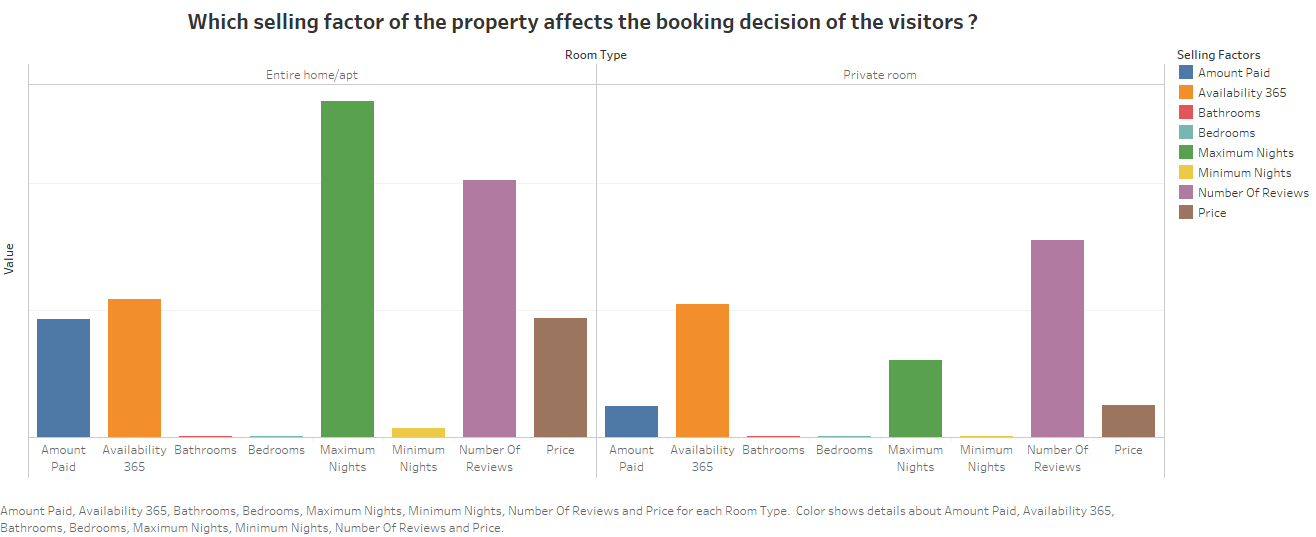
# Discoveries and Insights

# Which is the most preferred property type booked by visitors based on different selling factors?

**Fig.1**

In Fig.1, tell about the property type preferred by the visitor based on the factor or features like availability, number of maximum nights stayed by visitor, number of reviews, amount paid, price and more. Here its shows that apartment and the loft are most preferred in Dublin and least is cabin property type by visitors for stay. By this visitor will be helped in preferring which property type more for staying as per their budget and features requirement.

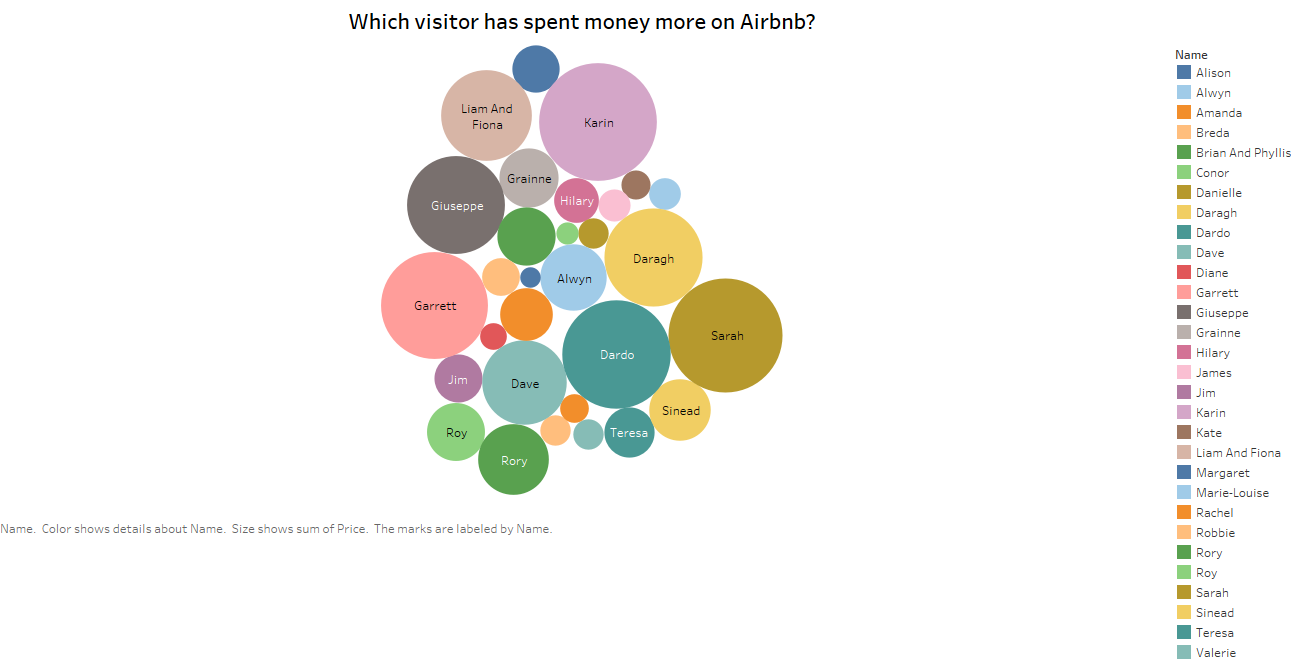
# Which selling factor of the property affects the booking decision of the visitors?



**Fig.2**

In Fig. 2 it’s showing in the factors which affects the booking decision of the visitors to rent a property is price, number of reviews, Availability 365, maximum nights and many more. As per the graph, the factor which affects the most in case of apartment or entire home is the maximum nights and the second most factor affecting is number of reviews whereas in case of private room, the factor which affects the most while booking decision of the visitor is number of reviews and then availability 365 with bathroom and bedrooms are the least affecting factors while booking the property. Therefore, visitors can make decision which room type to book depending on the given features.

# Which visitor has spent money more on Airbnb?

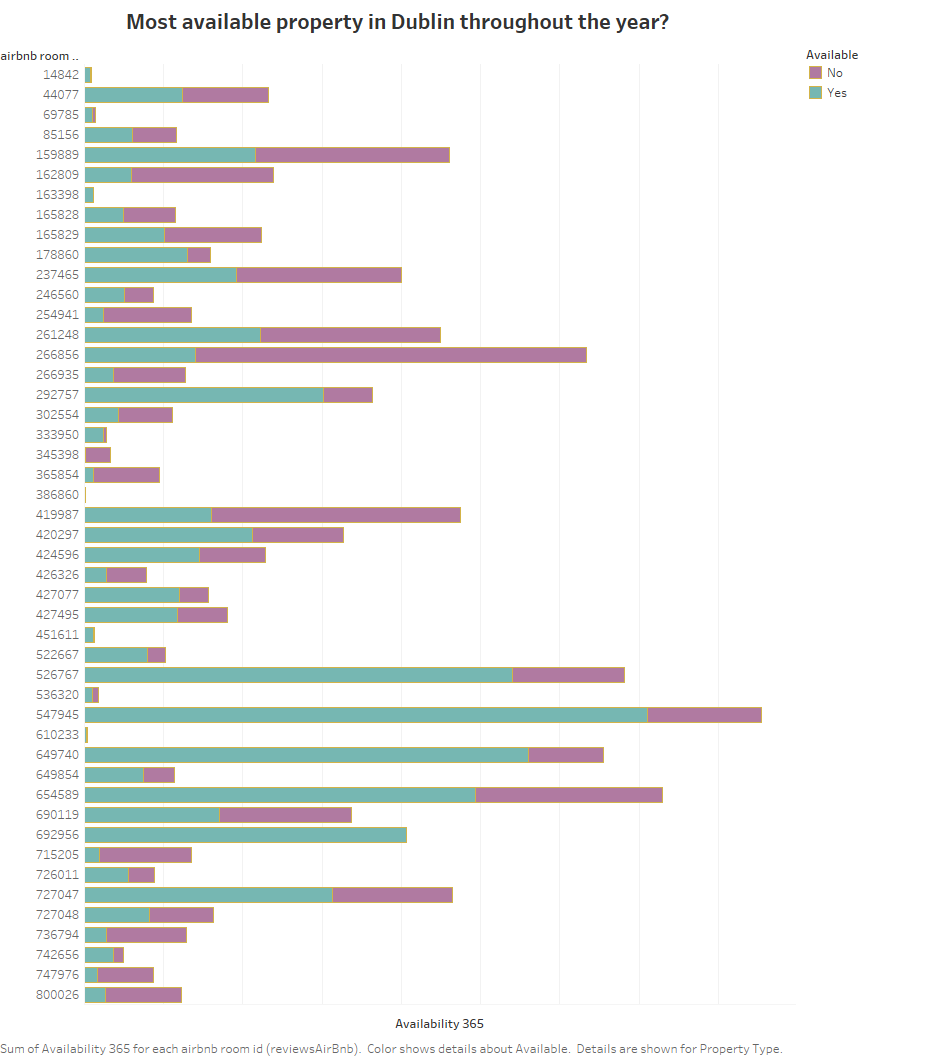


**Fig.3**

In Fig. 3, the given data helps in identifying the visitor who has spent the most on Airbnb. As per the data, Karen has spent the most on the Airbnb followed by Sarah, Dardo and Garrett.

This type of data also helpful for the Airbnb, as they can target their loyal and repetitive customers with more benefits and the least spending visitors with more discounts and offers.

# Most available property in Dublin throughout the year?



**Fig.4**

Visitors decision of booking a property depends on various factors. One of the the factor which is very important while making that decision is Availability of the property throught out the year. In the Fig. 4, will help the visitors to see the properties available and plan their trips.

# Which property type has raised more revenue for host in 2019?

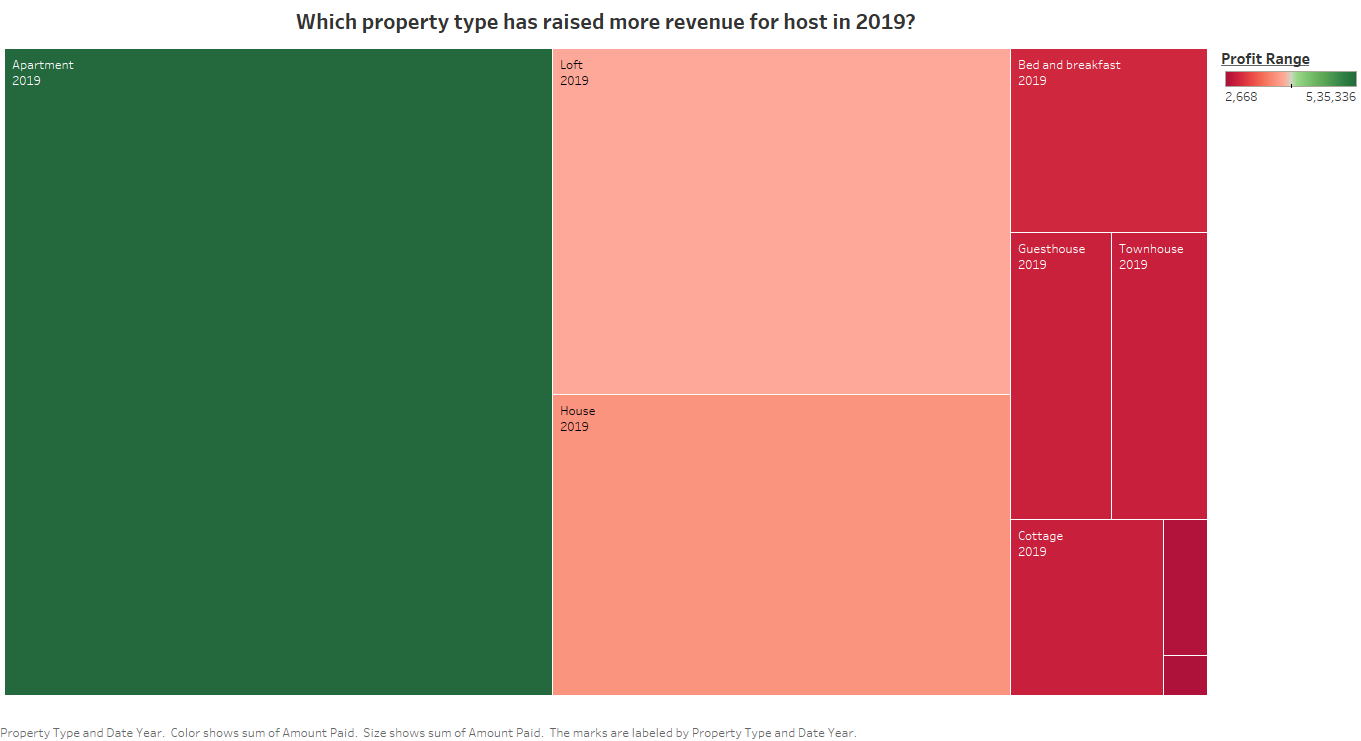


Fig.5

Fig 5 shows the property which has raised more revenue for the host in the year 2019. As per the data, apartments have raised the highest revenue for the host followed by the lofts and the houses. The lowest revenues raised by the cottages. This will help the hosts to decide which type of properties will earn them more revenues and will be beneficial for them. Also, this shows the visitors preferences towards the property types.

# Who is the top ten host listing of different types of rooms in Dublin?

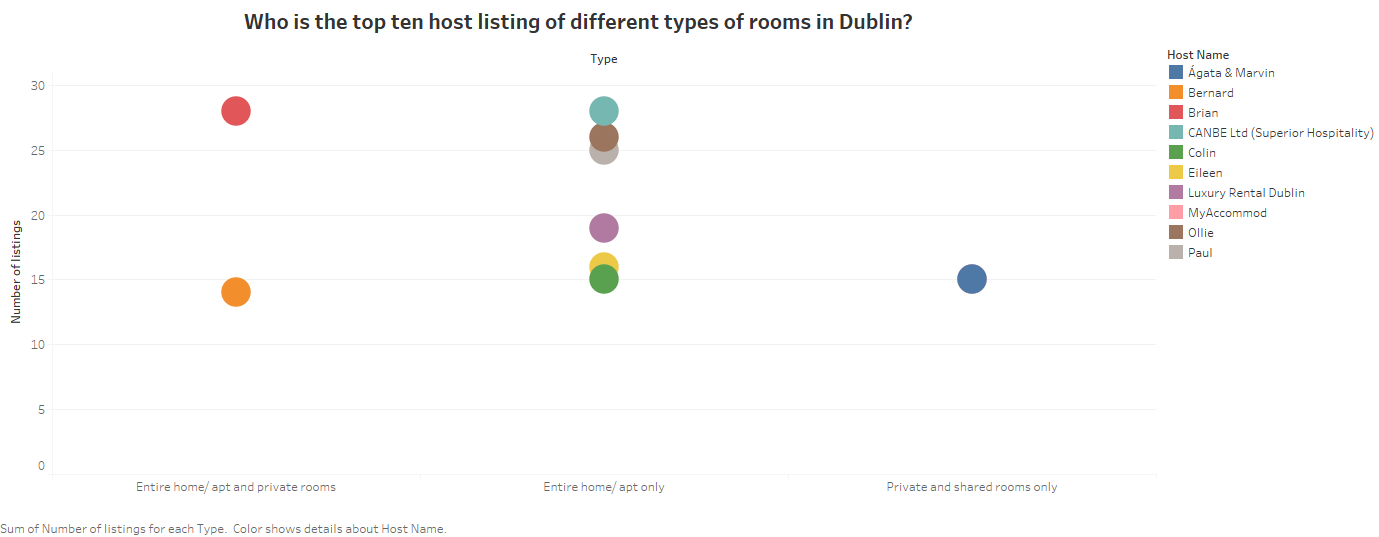


Fig.6

In Fig.6 the top ten host listing of different types of room in Dublin is given. Brian is at the top with the listing of 27 different types of rooms and Canbee Ltd. with almost the same number of listings. We can see the top 4 listers with more than 25 listings. There are other four hosts with listing between 15-20 and one host with 14. Hence, the visitors can see the top hosts in terms of different types of property.

# Host who has listed the maximum number of house property (Entire home) in Dublin?

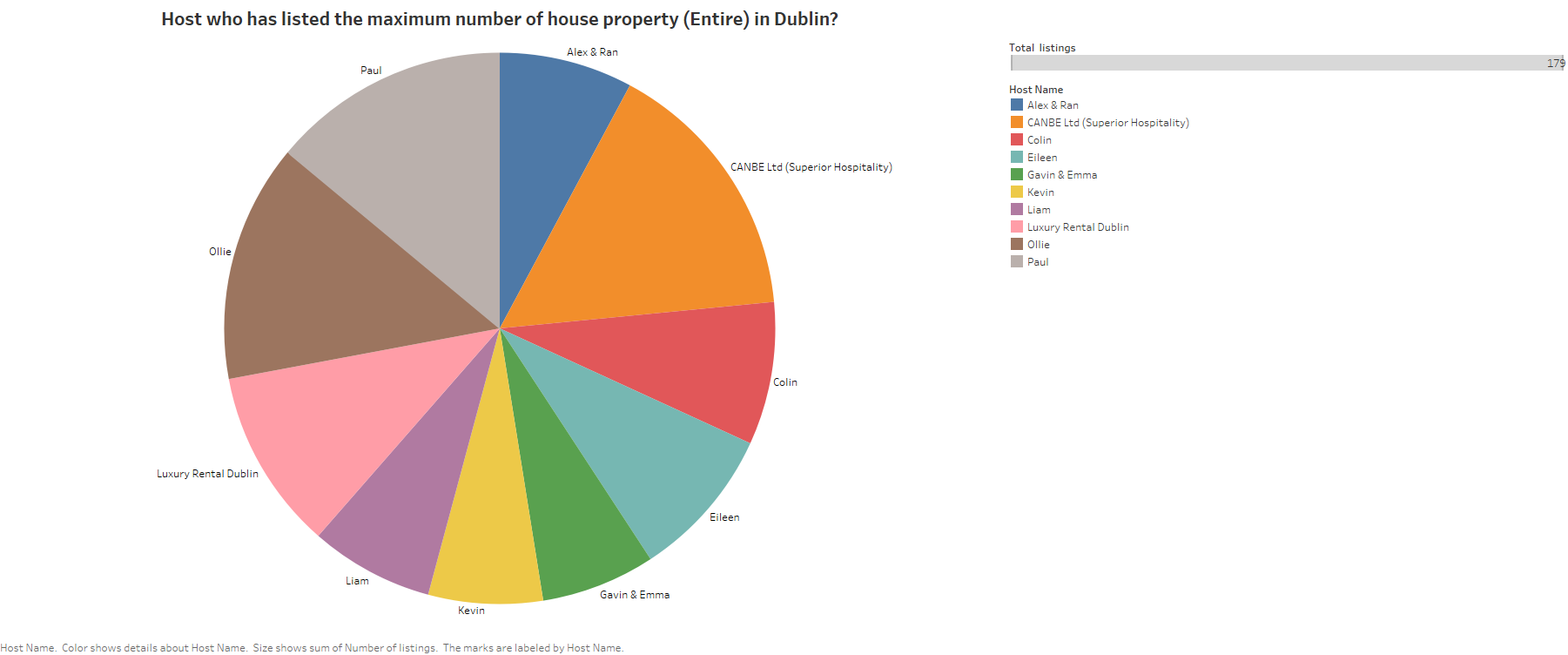


Fig.7

In the given Fig. 7, it can be identified that Canbee Ltd. listed the maximum number of house property in Dublin. Olie being the second highest lister. Liam, Kevin, Gavin & Emma have listed the minimum number of house property. This data will encourage other hosts to list their properties on Airbnb and earn more profits by listing more properties.

# Host who has most Recent and Frequent Home?

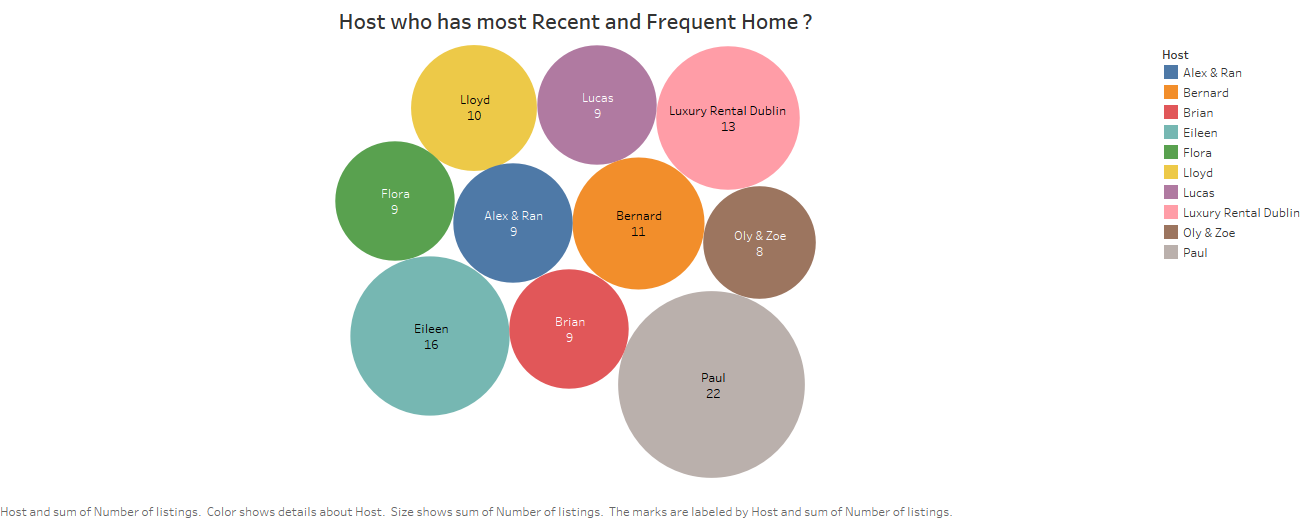


Fig.8

In Fig. 8, We can see the most recent and frequent homes listing. The highest number of listing is of Paul, with the sum of 22 followed by Eileen with the sum of 16. This data can show that which hosts is most frequent in listing the house properties. This data also tells us about the hosts who have recently listed their property. Visitors like to explore new and beautiful properties to stay and plan their trips too. This gives them vast choices in renting the properties with more variety.

# Which host have Highly Available Rooms?

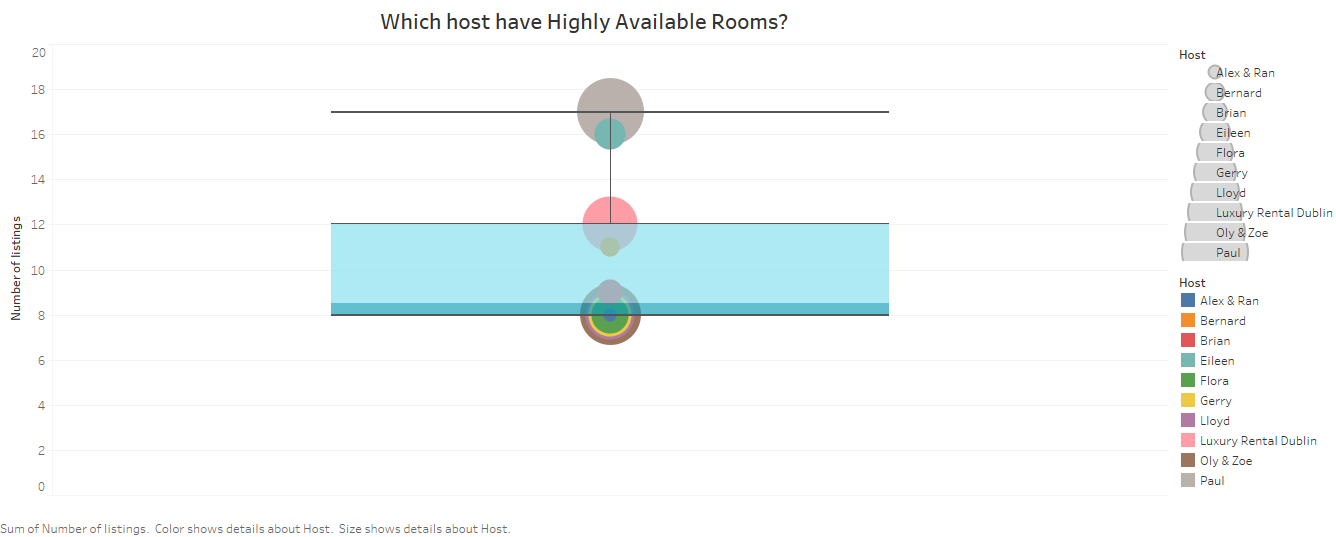


Fig.9

In Fig. 9, Boxplot show the rooms availability level low to high in Dublin city. Here upper whisker is 17 and lower whisker is 8 and median is 8.5 on this range the rooms are available for visitor who visit Dublin for stay. Visitors like to travel stay in properties as soon as possible they get the availability of house in Dublin as getting a house in Dublin is bit difficult. So this would help the visitor and host as well.

# Which host with maximum reviews and feedback in Dublin?

Fig.10

The other important factor which help the visitors to decide the which property to book is Number of reviews and feedbacks. In fig .10, the data shows the host with maximum number of reviews and feedbacks in Dublin is Alwyn with more than 40,000 reviews and feedbacks.

Such data can be very helpful for the visitors to book the property.

# Script for Pdf table scraping to excel files

library(tabulizer)  
library(dplyr)  
# storing pdf file  
pdfFile <-'https://www.dublincity.ie/councilmeetings/documents/s9670/4%20Airbnb\_OverviewforDCC.pdf'  
# Extracting the table  
outVariable <- extract\_tablesa(pdfFile)  
finalValues <- do.call(rbind, outVariable[-length(outVariable)])

# table headers get extracted as rows with bad formatting. Dump them.  
finalValues <- as.data.frame(final[3:nrow(finalValues), ])  
# Column names in vector  
ColHeaders <- c('Host Name', 'Number of listings', 'Type')  
# Applying customs column names  
names(finalValues) <- ColHeaders  
finalValues <- finalValues %>%  
# Convert Number of listings to numeric  
mutate(Number of listings = as.numeric(levels(Number of listings)[Number of listings]))  
# Writing final table  
write.csv(finalValues, file='TopHostListingRooms.csv', row.names=FALSE)

# Tableau Workbook Link

<https://public.tableau.com/profile/sumit.sharma6296#!/vizhome/CA_ONE_VIZ_SUMIT_SHARMA/MyDetails>

# Workbook Attached Below



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*End of the Assignment\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*