DATA201 Group Project

Taking an inside look at AirBNBs around the world!

By Yazeed, Heran and Chaarvee

Our research question and what we wanted to achieve.

- Comparing AirBNB prices around the world and if the countries population and GDP per capita has any relation to the AirBNB prices?
- We also wanted to compare the number of bedrooms and bathrooms for each AirBNB and to see if that also has any relation to the price of AirBNBs.

Why we chose AirBNBS?

- As a group we were all interested in this idea and felt like it was something unique, and maybe it could help us choose our next travel destinations.

How we acquired the data for AirBNBs?

- We downloaded CSVs for each of the countries we wanted to scrape from the following website http://insideairbnb.com/get-the-data/
- We then read in the CSVs for the selected countries
- Here is an example of the raw data that the CSVs spat out

```
Rows: 46,412
Columns: 80
$ id                                  <dbl> 6113, 35325, 46071, 48443…
$ listing_url                         <chr> "https://www.airbnb.com/r…
$ scrape_id                           <dbl> 2.02309e+13, 2.02309e+13,…
$ last_searched                       <date> NA, NA, 2023-09-03, NA, …
$ last_scraped                        <date> 2023-09-04, 2023-09-04, …
$ name                                <chr> "Place to stay in Otaki ·…
$ description                         <chr> "<b>The space</b><br />La…
$ neighborhood_overview               <chr> NA, "We are located in th…
$ picture_url                         <chr> "https://a0.muscache.com/…
$ host_id                             <dbl> 12177, 152089, 202747, 22…
$ host_url                            <chr> "https://www.airbnb.com/u…
$ host_name                           <chr> "Dianne", "Chika", "Donna…
$ host_since                          <date> 2009-04-03, 2010-06-25, …
$ host_location                       <chr> "Otaki, New Zealand", "Qu…
$ host_about                          <chr> "I have a great interest …
$ host_response_time                  <chr> "N/A", "within a few hour…
$ host_response_rate                  <chr> "N/A", "100%", "100%", "N…
$ host_acceptance_rate                <chr> "N/A", "94%", "100%", "N/…
$ host_is_superhost                   <lgl> FALSE, TRUE, FALSE, FALSE…
$ host_thumbnail_url                  <chr> "https://a0.muscache.com/…
$ host_picture_url                    <chr> "https://a0.muscache.com/…
$ host_neighbourhood                  <chr> NA, NA, NA, NA, NA, NA, N…
$ host_listings_count                 <dbl> 1, 2, 1, 2, 2, 1, 1, 1, 1…
$ host_total_listings_count           <dbl> 1, 3, 2, 3, 3, 1, 3, 1, 1…
$ host_verifications                  <chr> "['email', 'phone']", "['…
$ host_has_profile_pic                <lgl> TRUE, TRUE, TRUE, FALSE, …
$ host_identity_verified              <lgl> FALSE, TRUE, TRUE, FALSE,…
$ neighbourhood                       <chr> NA, "Queenstown, Otago, N…
$ latitude                            <dbl> -40.75807, -45.00532, -38…
$ longitude                           <dbl> 175.1564, 168.7771, 175.7…
$ property_type                       <chr> "Private room", "Entire h…
$ room_type                           <chr> "Private room", "Entire h…
$ accommodates                        <dbl> 2, 6, 8, 1, 2, 4, 1, 4, 2…
$ bathrooms                           <lgl> NA, NA, NA, NA, NA, NA, N…
$ bathrooms_text                      <chr> NA, "1.5 baths", "2 baths…
$ bedrooms                            <dbl> 1, 3, 5, 1, 1, 1, 1, NA, …
$ beds                                <dbl> 1, 6, 7, NA, 1, 2, NA, 2,…
$ amenities                           <chr> "[\"Breakfast\", \"Wifi\"…
$ price                               <chr> "$109.00", "$250.00", "$2…
$ minimum_nights                      <dbl> 1, 13, 2, 1, 1, 2, 1, 1, …
$ maximum_nights                      <dbl> 21, 149, 10, 730, 730, 73…
$ minimum_minimum_nights              <dbl> 1, 13, 2, 1, 1, 2, 1, 1, …
$ maximum_minimum_nights              <dbl> 1, 13, 2, 1, 1, 2, 1, 1, …
$ minimum_maximum_nights              <dbl> 21, 149, 10, 730, 730, 73…
$ maximum_maximum_nights              <dbl> 21, 149, 10, 730, 730, 73…
$ minimum_nights_avg_ntm              <dbl> 1.0, 13.0, 2.0, 1.0, 1.0,…
$ maximum_nights_avg_ntm              <dbl> 21, 149, 10, 730, 730, 73…
$ calendar_updated                    <lgl> NA, NA, NA, NA, NA, NA, N…
$ has_availability                    <lgl> FALSE, TRUE, TRUE, FALSE,…
```

Wrangling the data

- The CSVs for each AirBNB location had thousands of variables we could select from.
- For this project we only wrangled the price, number of bathrooms and the number of bedrooms
- Here is a small example of the wrangled data

A tibble: 87946 × 3

price	bathrooms_text	bedrooms
<chr>	<chr>	<dbl>
$42.00	1.5 shared baths	NA
$175.00	1 bath	2
$79.00	1 shared bath	NA
$150.00	1 bath	1
$46.00	1 shared bath	NA
$476.00	2 baths	3
$371.00	2 baths	2
$250.00	1.5 baths	1
$75.00	1 bath	1
$29.00	1.5 shared baths	NA
$120.00	1 bath	2
$50.00	1 shared bath	NA
$151.00	1 bath	1
$35.00	1 shared bath	NA
$140.00	1 private bath	NA
$90.00	1 bath	1

How we acquired the data for world population and GDP per capita?

- We scraped this website https://www.worldometers.info/gdp/gdp-by-country/
- We selected the data we wanted from the website which was the Countries name, population and GDP per capita.

Wrangling the data

- We then created a function which scrapes the entire webpage and only selects the countries we wanted their population and GDP per capita.

[44]:
```r
scrape_and_filter <- function(url) {
  # Scrape data
  webpage <- read_html(url)

  countries <- webpage %>%
    html_nodes('#example2 a') %>%
    html_text()

  population <- webpage %>%
    html_nodes('td:nth-child(6)') %>%
    html_text() %>%
    gsub("[^0-9.]", "", .) %>%
    as.numeric()

  gdp_per_capita <- webpage %>%
    html_nodes('td:nth-child(7)') %>%
    html_text() %>%
    gsub("[^0-9.]", "", .) %>%
    as.numeric()

  # Convert to data frame
  data <- data.frame(
    Country = countries,
    Population = population,
    GDP = gdp,
    GDP_per_Capita = gdp_per_capita
  )

  # Filter for the required countries
  filtered_data <- data %>%
    filter(Country %in% c('United Kingdom', 'Japan', 'New Zealand', 'Greece', 'Belgium', 'Netherlands', 'United States', 'Spain', 'Thailand', 'Australia'))

  return(filtered_data)
}

# Call the function
url <- "https://www.worldometers.info/gdp/gdp-by-country/"
result <- scrape_and_filter(url)
print(result)
```

	Country	Population	GDP	GDP_per_Capita
1	United States	338289857	$25,462,700,000,000	75269
2	Japan	123951692	$4,231,140,000,000	34135
3	United Kingdom	67508936	$3,070,670,000,000	45485
4	Australia	26177413	$1,675,420,000,000	64003
5	Spain	47558630	$1,397,510,000,000	29385
6	Netherlands	17564014	$991,115,000,000	56429
7	Belgium	11655930	$578,604,000,000	49640
8	Thailand	71697030	$495,341,000,000	6909
9	New Zealand	5185288	$247,234,000,000	47680
10	Greece	10384971	$219,066,000,000	21095

Difficulties we had during the project

- Something that was out of our control was some of the inputs for the number of bathrooms and bedrooms showed up as NA. We assume this is because the people who have listed up their homes on AirBNB have just left these sections blank for some odd reason.
- The pricing of AirBNBs being in the local currencies of the countries we had scraped (for example Japanese AirBNB data was in JPY and not NZD) which would make it much more difficult for us to answer our research question - however we did have a solution for this
- Lots and lots of raw data!

How we overcame the issue pricing being in different currencies?

- After doing lots of research we found this website https://fixer.io/ which has an API with the live exchange rates and helped us exchange the pricing to $NZD which made it much easier for us to answer the research question.
- Here is the code for how we did this

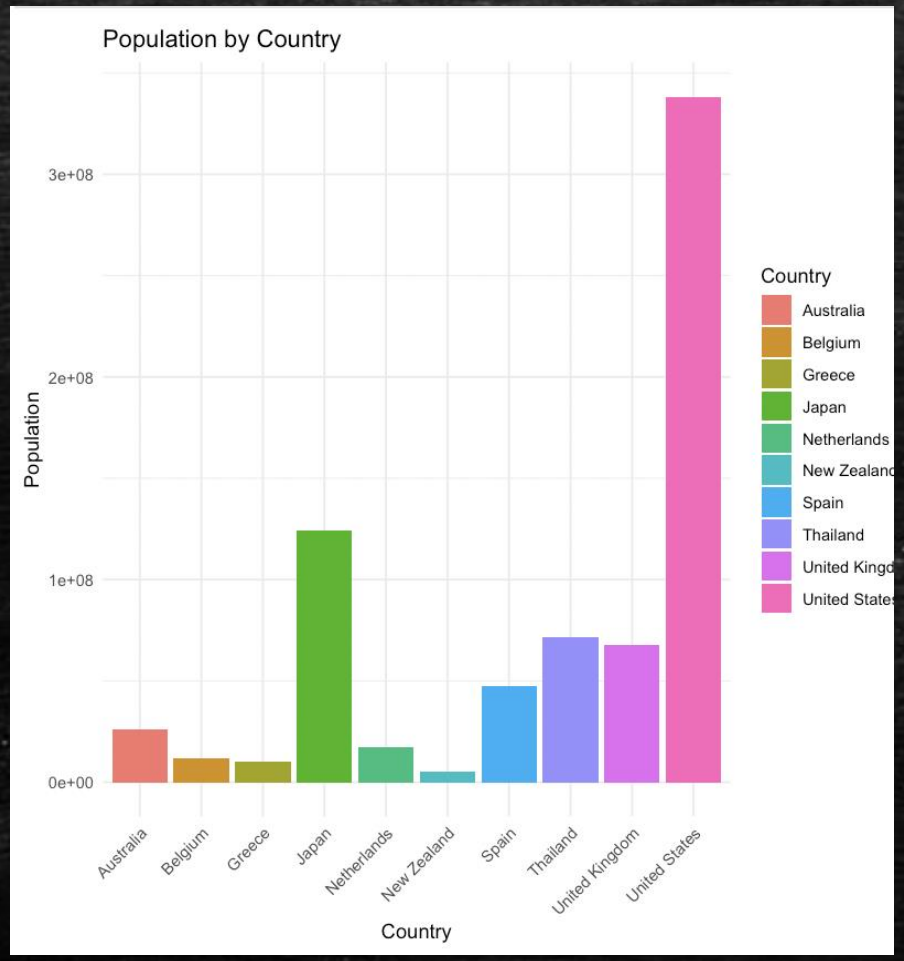
```
[42]: # API key for fixer.io a live exchange rate site so we can convert the price column from JPY to NZD

      api_key <- "58c8eb71a98bc0e70380918c4ec1041c"

[43]: response <- GET(paste0("http://data.fixer.io/api/latest?access_key=", api_key, "&symbols=JPY,NZD"))

[44]: data <- fromJSON(content(response, as = "text"))

[45]: exchange_rate_jpy_to_nzd <- data$rates$NZD / data$rates$JPY

[46]: japan_airbnb$price_nzd <- japan_airbnb$price * exchange_rate_jpy_to_nzd

[47]: japan_airbnb_nzd <- japan_airbnb %>%
        mutate(price_nzd = price * exchange_rate_jpy_to_nzd) %>%
        select(price_nzd, bathrooms_text, bedrooms)

      japan_airbnb_nzd
```

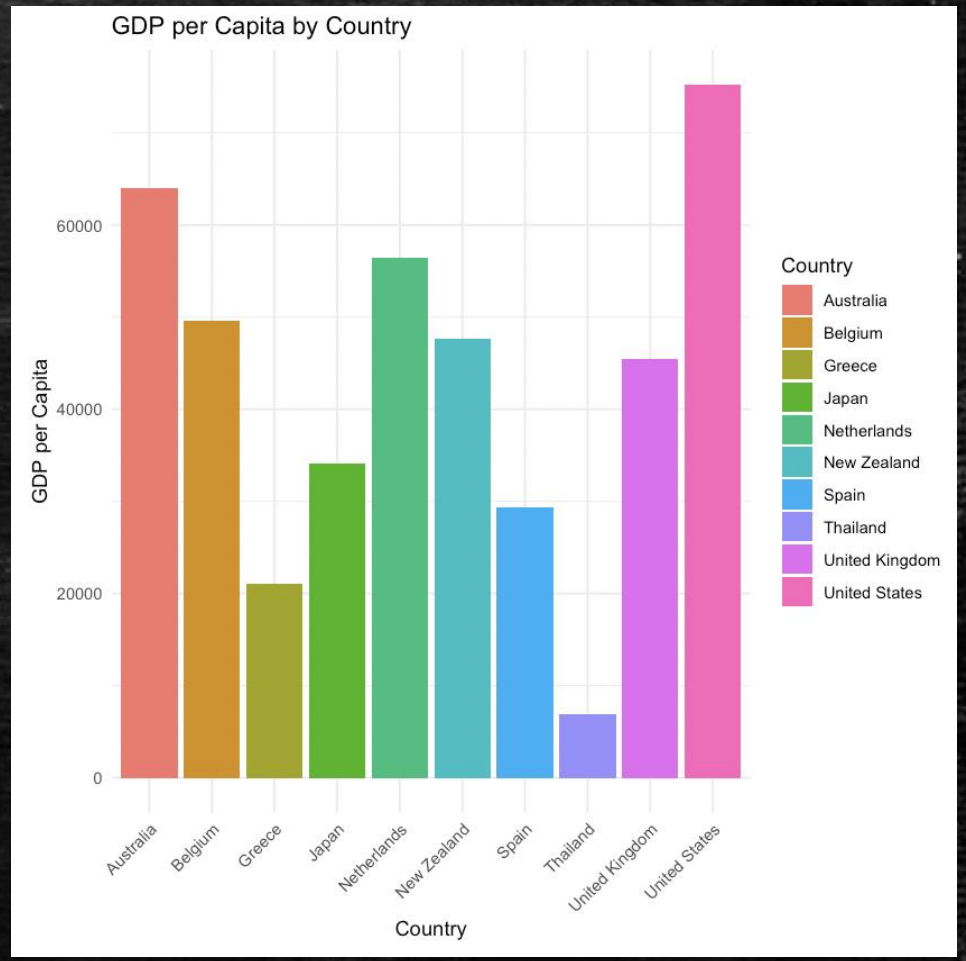
A tibble: 12234 × 3

price_nzd	bathrooms_text	bedrooms
<dbl>	<chr>	<dbl>
133.77839	1 bath	1
103.27691	1 bath	NA
82.70849	1 bath	1
99.24127	1 shared bath	NA

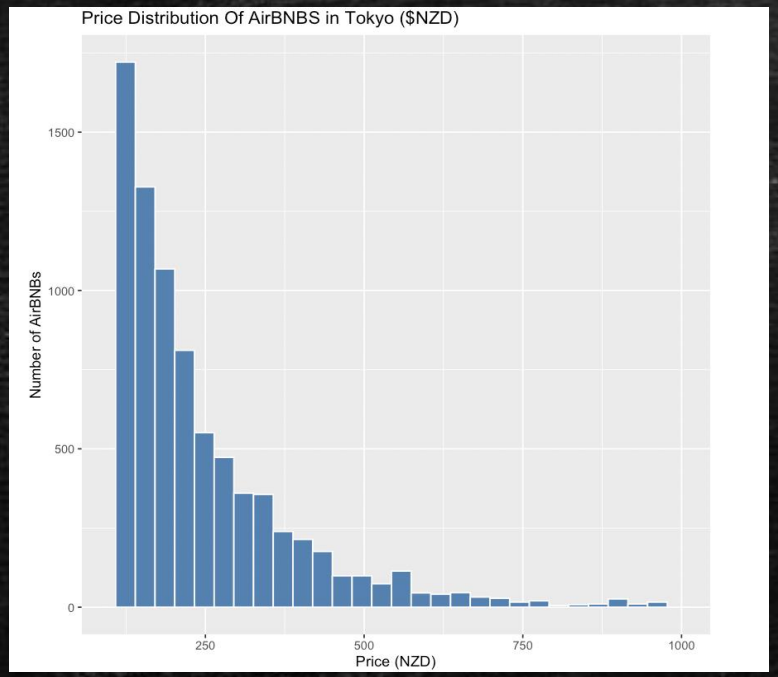
Our findings!

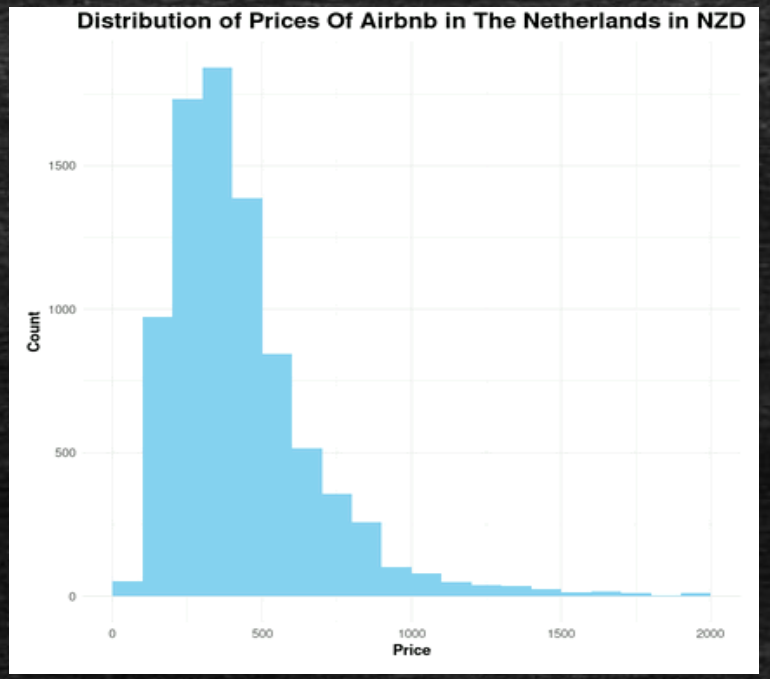
Before we look into the actual AirBNB data here are the graphs for the population and GDP per capita for each of the countries we scraped AirBNB data for

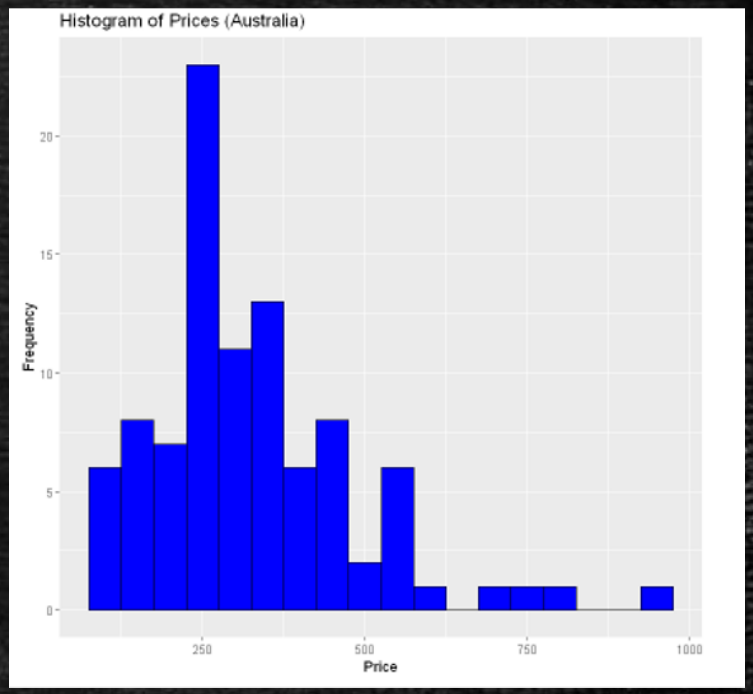




AirBNB Price Distribution Graphs

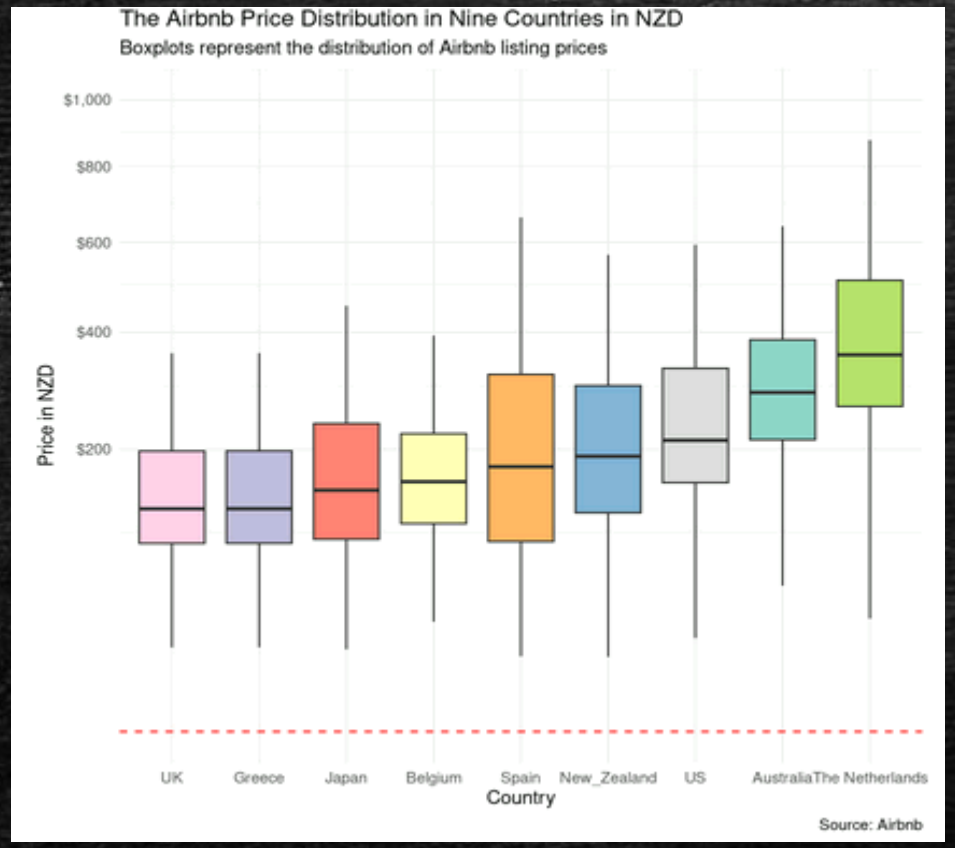






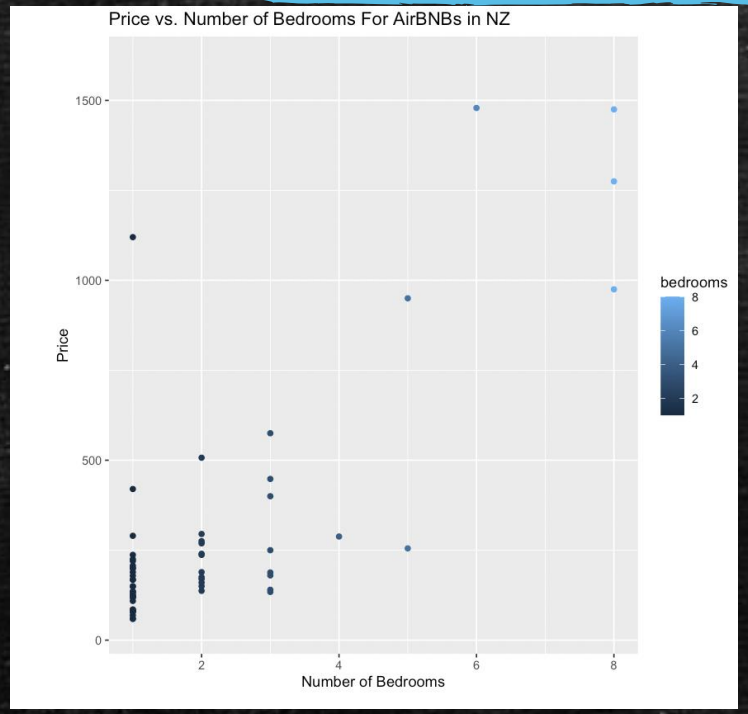
In these graphs we can see that the majority of AirBNB prices are at the lower end being between $0-$250, and the number of listings significantly drop at price points above $500. When comparing these graphs to previous graphs of population and GDP per capita there is no direct comparison, that could help us answer our research question.

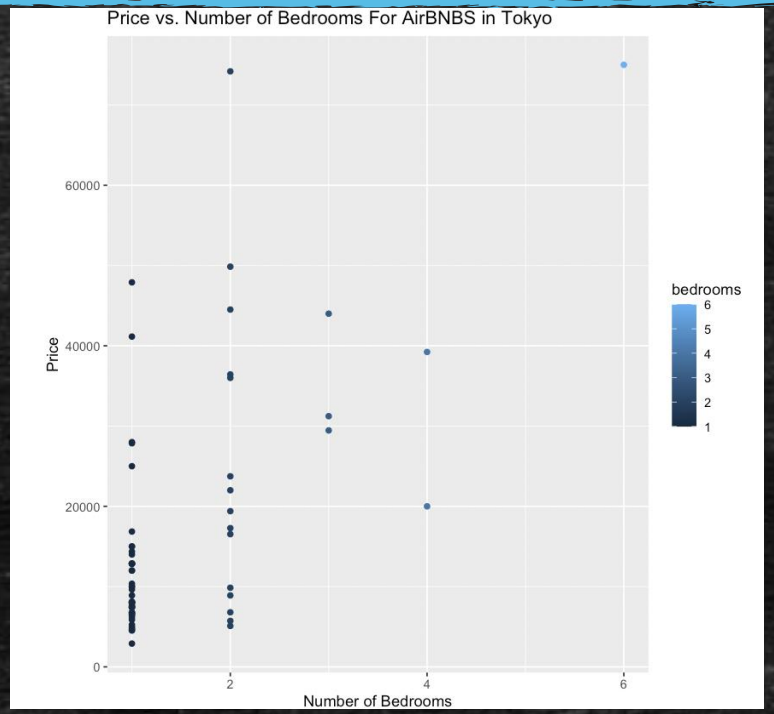
BOX PLOT OF PRICE DISTRIBUTION IN THREE COUNTRIES

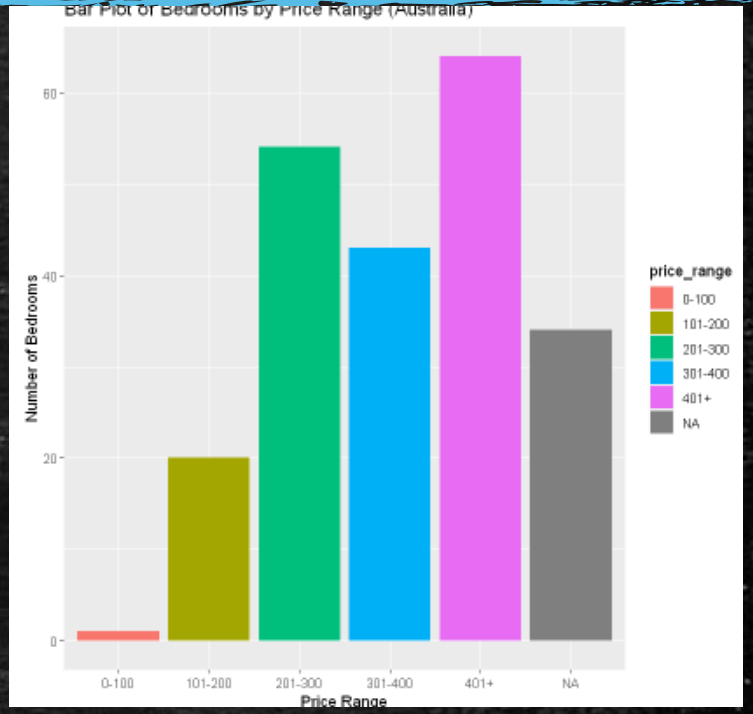


This graph shows us that The Netherlands, Australia and USA have the highest price distribution of AirBNBs. These three countries also have the highest GDP per capita. This graph also shows us that Greece has one of lowest price distributions for AirBNBs, and Greece had the lowest GDP per capita between these countries. This answers a part of our research question where the higher the GDP per capita the higher the price of the AirBNBs and vice versa. However, there was no relation to the country's population and AirBNB pricing

Comparing AirBNB Prices With The Number Of Bedrooms

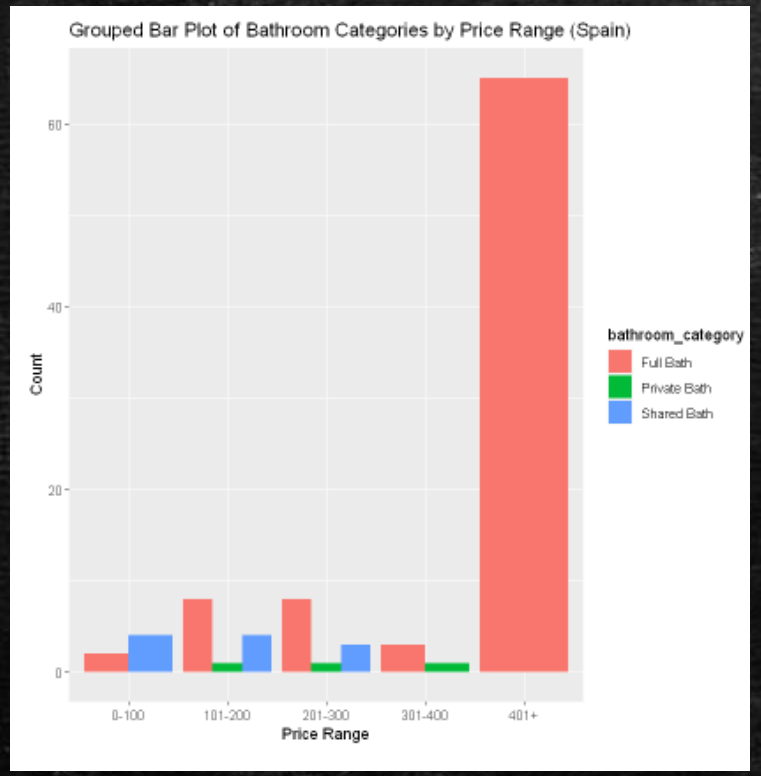


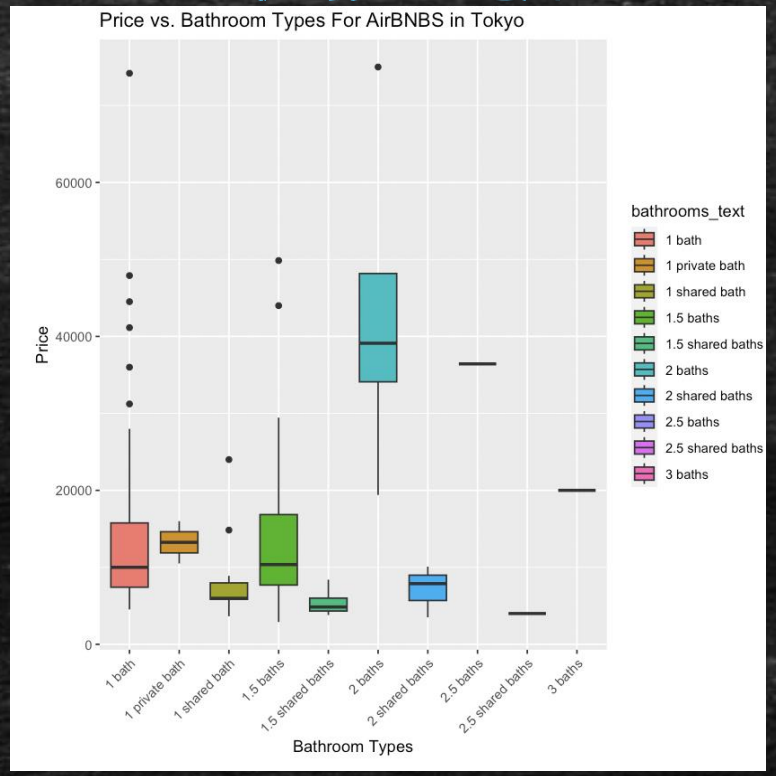


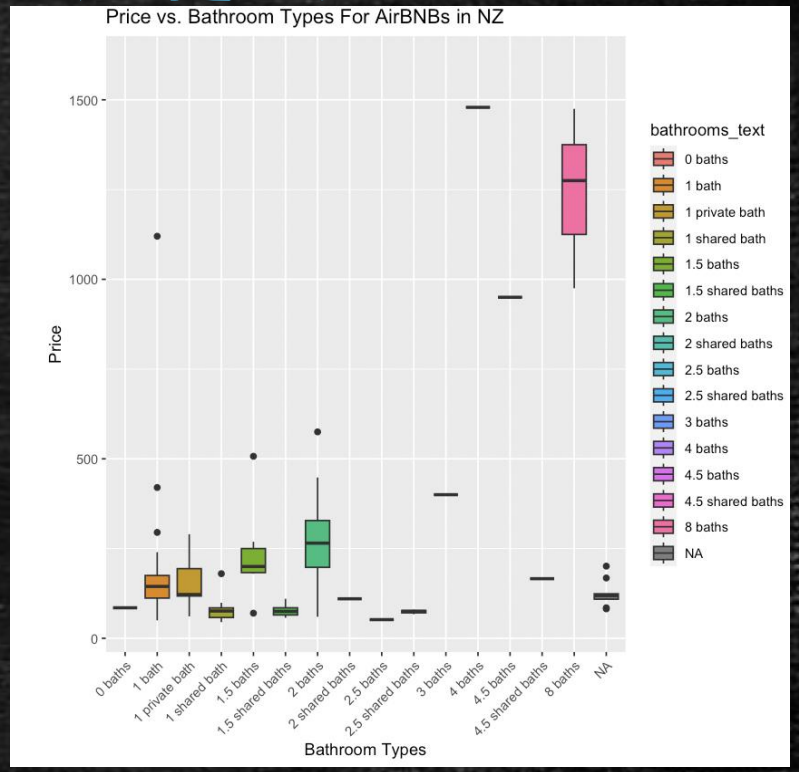


These graphs show us that the more bedrooms AirBNBs have the higher priced they tend to be. For example, the highest priced AirBNB in Tokyo has the most bedrooms which is 6.

Comparing AirBNB Prices With The Number and Type Of Bathrooms







These graphs shows us that the more bathrooms the higher the AirBNB Price tends to be especially the price increase between 1-2 bathrooms. However the price increase when there was more than 2 bathrooms wasn't as great as the price difference between 1 and 2 bathrooms

Conclusion

- Our research question was if population and GDP per capita had any correlation with the pricing of AirBNBs. However, we could not conclude that population has any effects on AirBNB pricing.
- GDP per capita did have correlation with the prices of AirBNBs as seen in slide 10 and 12, where the 3 countries with the highest GDP per capita had the highest price distributions of AirBNBs and the countries with the lowest GDP per capita had the lowest price distribution of AirBNBs.
- The second part of our question was if the AirBNBs had a higher bedrooms and bathrooms does that mean the pricing would be higher, and yes from our findings we can conclude that this was true.

Thanks for listening!

Q & A