



Northeastern University
College of Professional Studies

Airbnb New York Analysis (2011 - 2019) :

Data Analysis On Hospitality Industry

Name: Abhilash Kumar Dikshit

Student ID: 2702209

Course: MPS Analytics- ALY 6000

Campus: Vancouver, Canada

Airbnb New York Analysis (2011 - 2019)

Airbnb New York Analysis (2011 - 2019)

This dataset contains information about Airbnb host listings, neighbourhoods, price, minimum nights, review counts/rate and the availability throughout the year.

Limitations:

No data on customer stay duration, timeline, review score for each listing and no. of tourist attractions nearby.

Acknowledgements:

Dataset taken from [Kaggle](#). This is a public dataset of Airbnb, and the original source can be found in their website.

Data Analysis Presented:

Part 1- Dataframe cleanup:

- Sorted as per price
- Removed blank cells
- Resized the data frame to rows< 6000 and attributes> 8

Part 2- Created new attributes based on the data:

- Extracted Year from column *last_review*
- Added extra column *Review_Year* to identify number of reviews based on the room type from year 2011-2019.
- Computed the mean and median for each of the variables.

Part 3- Data Storytelling:

- Data storytelling using descriptive statistics and visualizations

```
[1] "Uncleaned Dataframe Info:"
```

```
Number of Rows: 48895
```

```
Number of Columns: 16
```

```
Blank cells count: 10052
```

```
[1] "Removed Blank cells:"
```

```
Number of Rows after removing blank cells: 38843
```

```
Number of Columns after removing blank cells: 16
```

```
Blank cells count after removal: 0
```

```
[1] "Resized dataframe:"
```

```
Number of Rows: 4890
```

```
Number of Columns: 16
```

```
Blank cells count in cleaned dataframe: 0
```



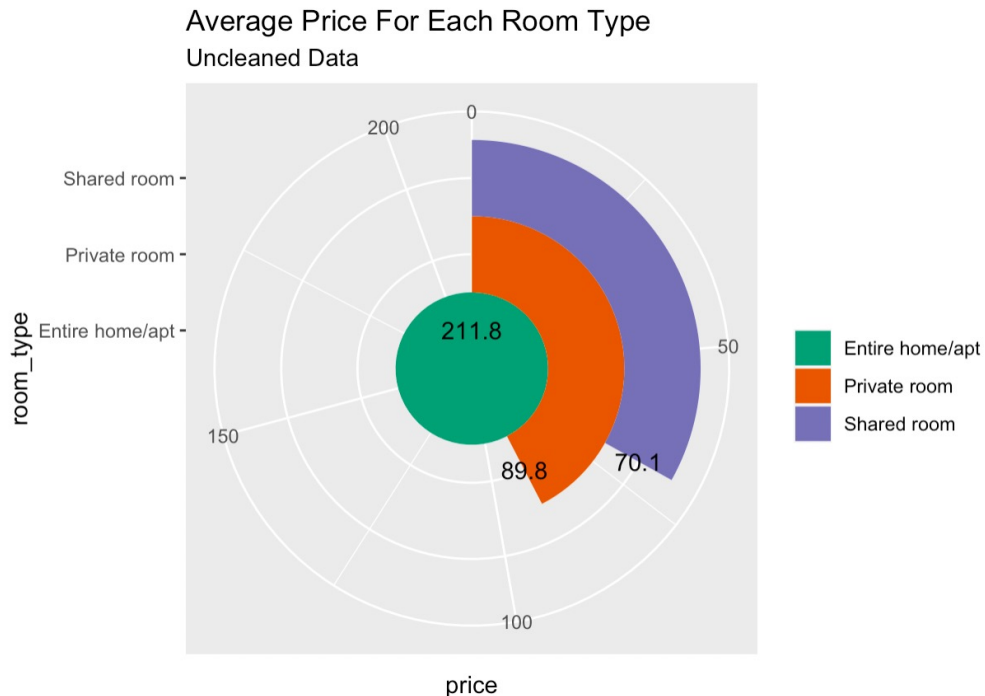
Uncleaned & Cleaned Dataset: Average Price For Each Room Type

Business Question:

What is the average cost for each room type and which area is the most expensive/cheap?

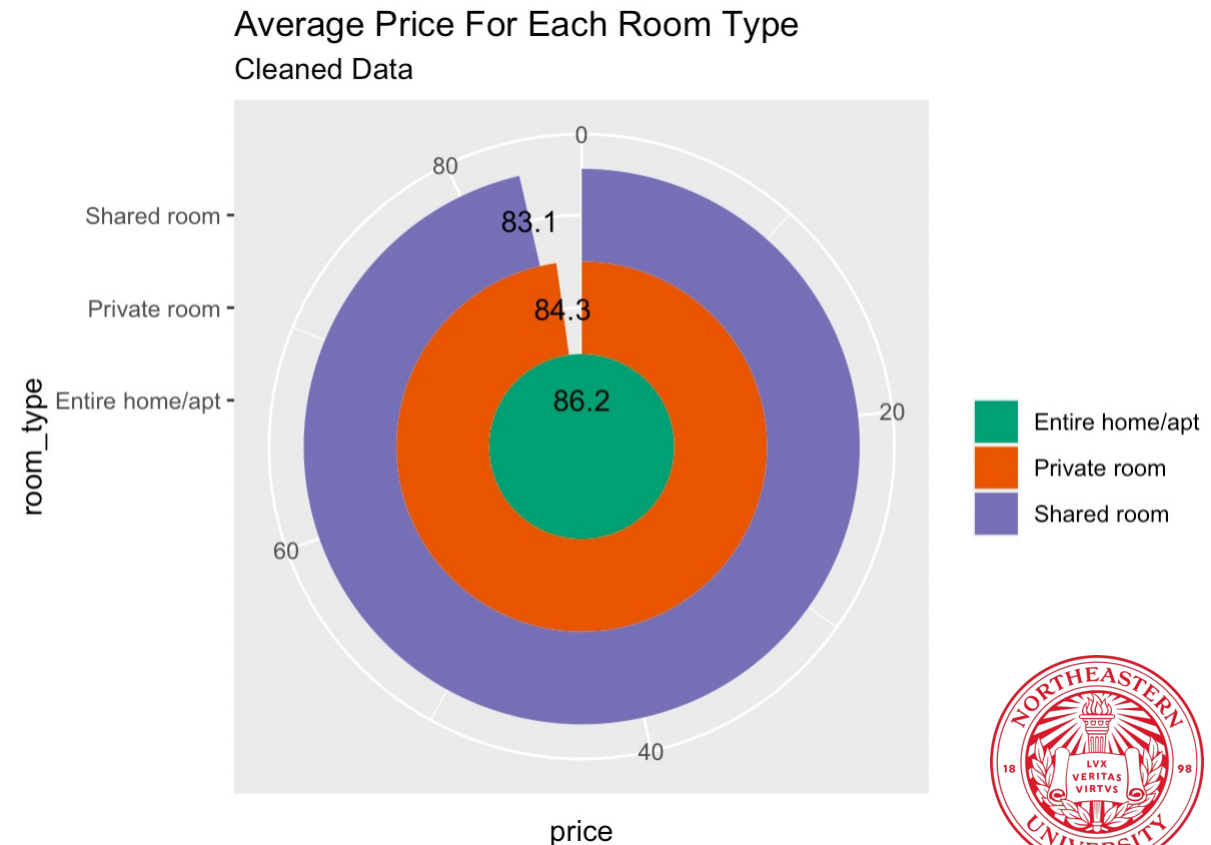
Analysis For Uncleaned Data:

- a) Entire home/apartment average cost is 212 dollars
- b) Private rooms average cost is 90 dollars
- c) Shared rooms average cost is 70 dollars



Analysis For Cleaned Data:

- a) Entire home/apartment average cost is 86 dollars
- b) Private rooms average cost is 84 dollars
- c) Shared rooms average cost is 83 dollars



Show Popular Room Type In New York City

Business Question:

Describe the distribution of room type in New York city? Which is the most popular?

Analysis:

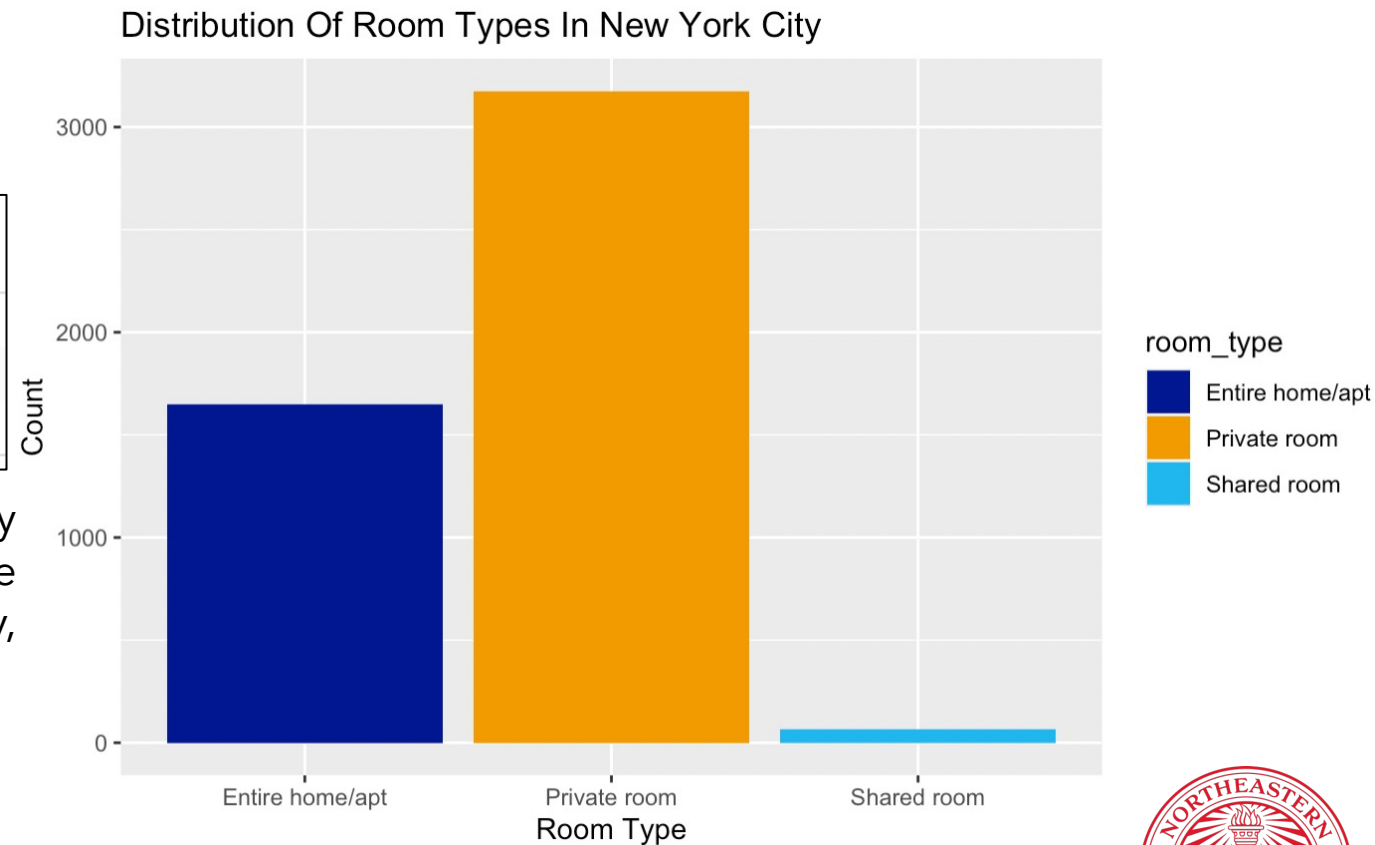
We have 3 room types in New York City, USA i.e., Entire home/apt, Private room and Shared room.

Room Type <chr>	freq <int>
Entire home/apt	1647
Private room	3176
Shared room	67

Customers prefer to book the private room type stay and it is the most popular as compared to Entire home or apartment and shared room types in New York city, USA as depicted in the plot.

Reason:

- Vacation place
- Business Travel
- Entertainment Industry



Neighborhood Groups & Percentage Of Airbnb's

Business Question:

What is the distribution of the neighbourhood groups variable and percentage of Airbnb's located in each area?

Analysis:

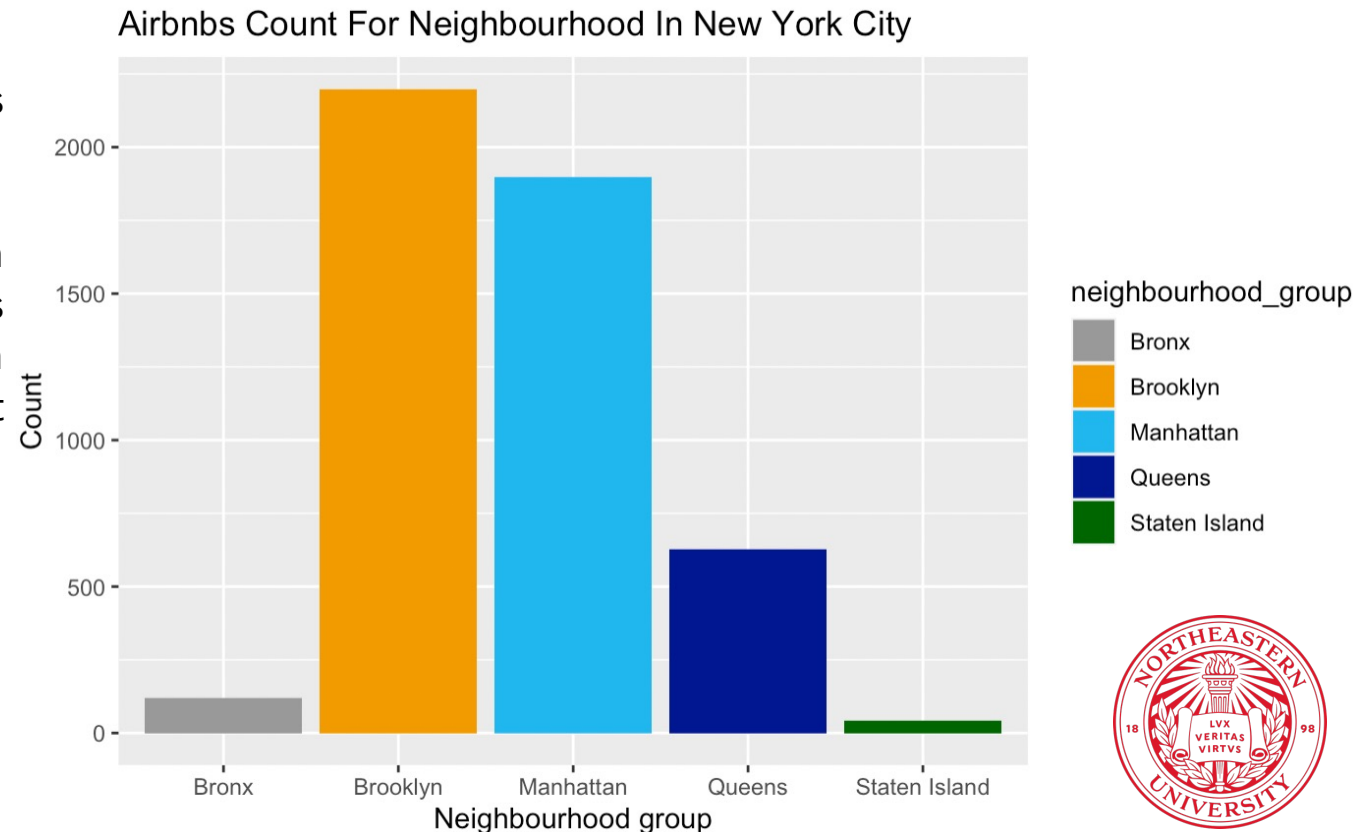
We have differentiated the neighbourhood in 5 groups as shows in the given table.

Out of Bronx, Brooklyn, Manhattan, Queens, and Staten Island, the percentage of Airbnb's located in Brooklyn is 45% which is the highest among all whereas Manhattan is at 39% followed by Queens which is at 13%, Bronx at 2% and Staten Island is below 1%.

Reason:

- Close by nature, parks and archeological sites.
- Good weather and food.
- Good commute and less traffic.

Neighbourhoods_Group <chr>	freq <int>	Percentage <dbl>
Bronx	121	2.4744376
Brooklyn	2199	44.9693252
Manhattan	1899	38.8343558
Queens	629	12.8629857
Staten Island	42	0.8588957



Distribution Of Neighborhood Group & Room Type

Business Question:

What is the distribution of the neighbourhood groups and room types located in each area?

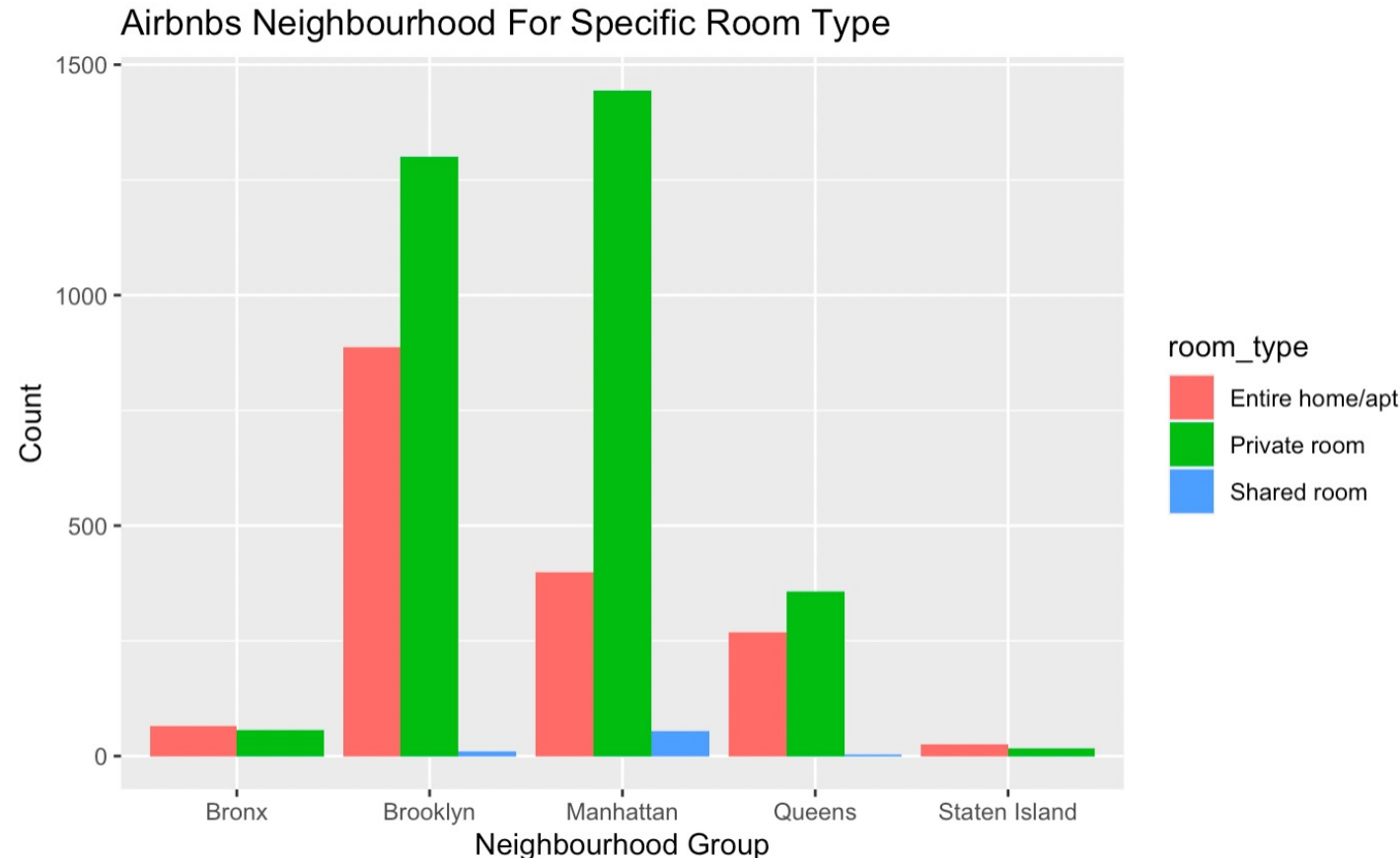
Analysis:

Manhattan and Brooklyn have the most rooms in general. They have the most entire private room. Manhattan and Brooklyn are basically split up between private room and entire home/apartment. The three other places have little rooms in general.

Reason:

- High demand of hotel industry.
- Number of customer are high.
- High lifestyle
- Top company executives and business market.

Mean & Median



	vars	n	mean	sd	median	trimmed	mad	min	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
id	1	4890	17029615.56	10563203.71	16715757.50	16931158.05	13552499.23	3831.00	10563203.71	16715757.50	16931158.05	13552499.23	3831.00	36238485.00	36234654.00	0.05	-1.20	151057.13
name*	2	4890	2437.62	1404.80	2437.50	2438.18	1801.36	1.00	1404.80	2437.50	2438.18	1801.36	1.00	4868.00	4867.00	0.00	-1.20	20.09
host_id	3	4890	56800400.11	69933126.99	24786381.00	43034251.97	32621057.73	2438.00	69933126.99	24786381.00	43034251.97	32621057.73	2438.00	272308792.00	272306354.00	1.46	1.05	1000065.67
host_name*	4	4890	1201.04	701.25	1199.00	1198.95	901.42	1.00	701.25	1199.00	1198.95	901.42	1.00	2477.00	2476.00	0.00	-1.17	10.03
neighbourhood_group*	5	4890	2.65	0.77	3.00	2.58	1.48	1.00	0.77	3.00	2.58	1.48	1.00	5.00	4.00	0.48	-0.21	0.01
neighbourhood*	6	4890	89.39	57.15	84.00	88.26	71.16	1.00	57.15	84.00	88.26	71.16	1.00	182.00	181.00	0.28	-1.23	0.82
latitude	7	4890	40.73	0.06	40.72	40.73	0.06	40.51	0.06	40.72	40.73	0.06	40.51	40.90	0.39	0.29	-0.29	0.00
longitude	8	4890	-73.95	0.05	-73.95	-73.95	0.03	-74.24	0.05	-73.95	-73.95	0.03	-74.24	-73.73	0.51	1.12	5.47	0.00
room_type*	9	4890	1.68	0.50	2.00	1.70	0.00	1.00	0.50	2.00	1.70	0.00	1.00	3.00	2.00	-0.42	-1.02	0.01
price	10	4890	84.90	5.94	85.00	84.83	7.41	75.00	5.94	85.00	84.83	7.41	75.00	95.00	20.00	0.07	-1.10	0.09
	vars	n	mean	sd	median	trimmed	mad	min	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
minimum_nights	11	4890	5.32	19.82	2.00	2.61	1.48	1.00	19.82	2.00	2.61	1.48	1.00	999.00	998.00	30.75	1366.54	0.28
number_of_reviews	12	4890	32.85	50.80	11.00	21.04	14.83	1.00	50.80	11.00	21.04	14.83	1.00	439.00	438.00	2.77	9.74	0.73
last_review*	13	4890	802.97	289.95	977.00	856.25	68.20	1.00	289.95	977.00	856.25	68.20	1.00	1028.00	1027.00	-1.24	0.23	4.15
reviews_per_month	14	4890	1.41	1.69	0.73	1.10	0.95	0.01	1.69	0.73	1.10	0.95	0.01	19.75	19.74	2.02	6.59	0.02
calculated_host_listings_count	15	4890	2.13	4.18	1.00	1.42	0.00	1.00	4.18	1.00	1.42	0.00	1.00	103.00	102.00	10.89	156.32	0.06
availability_365	16	4890	105.06	125.83	42.00	87.37	62.27	0.00	125.83	42.00	87.37	62.27	0.00	365.00	365.00	0.89	-0.72	1.80



Price Comparison Across Different Neighborhood And Room Types

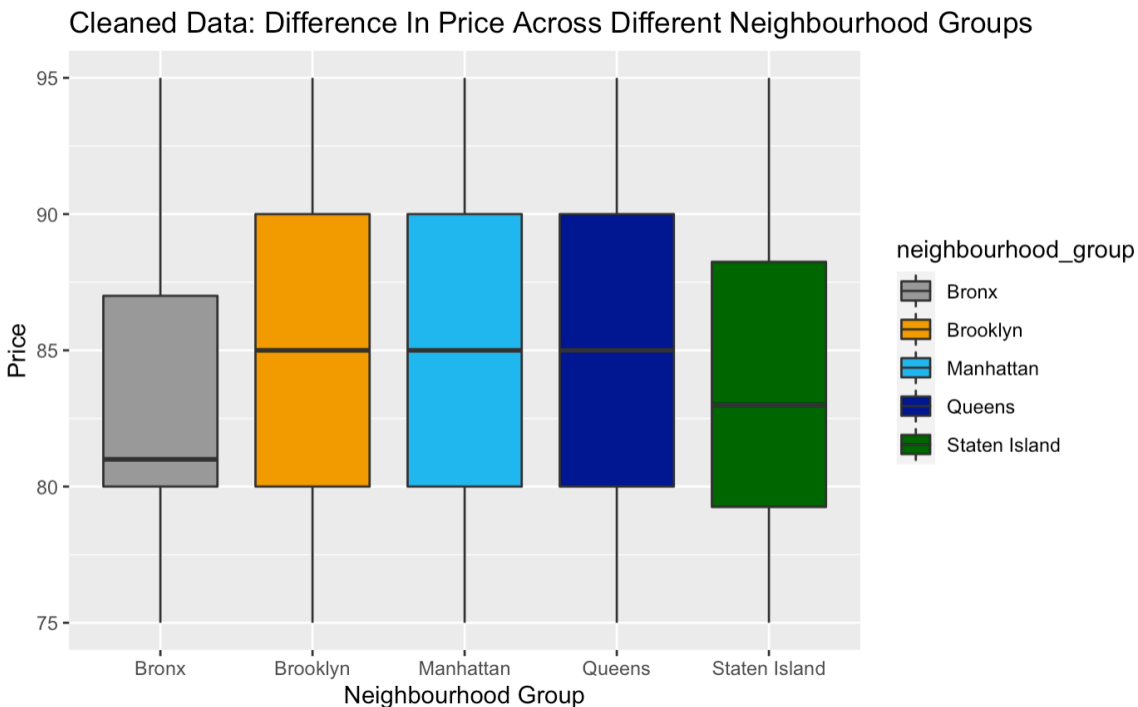
Business Question:

Is there a difference in price across different neighbourhood groups?

Analysis:

Brooklyn and Manhattan has the highest median price out of all of them i.e., 85. It also has the least number of outliers.

Bronx region has the lowest median price i.e., 82.



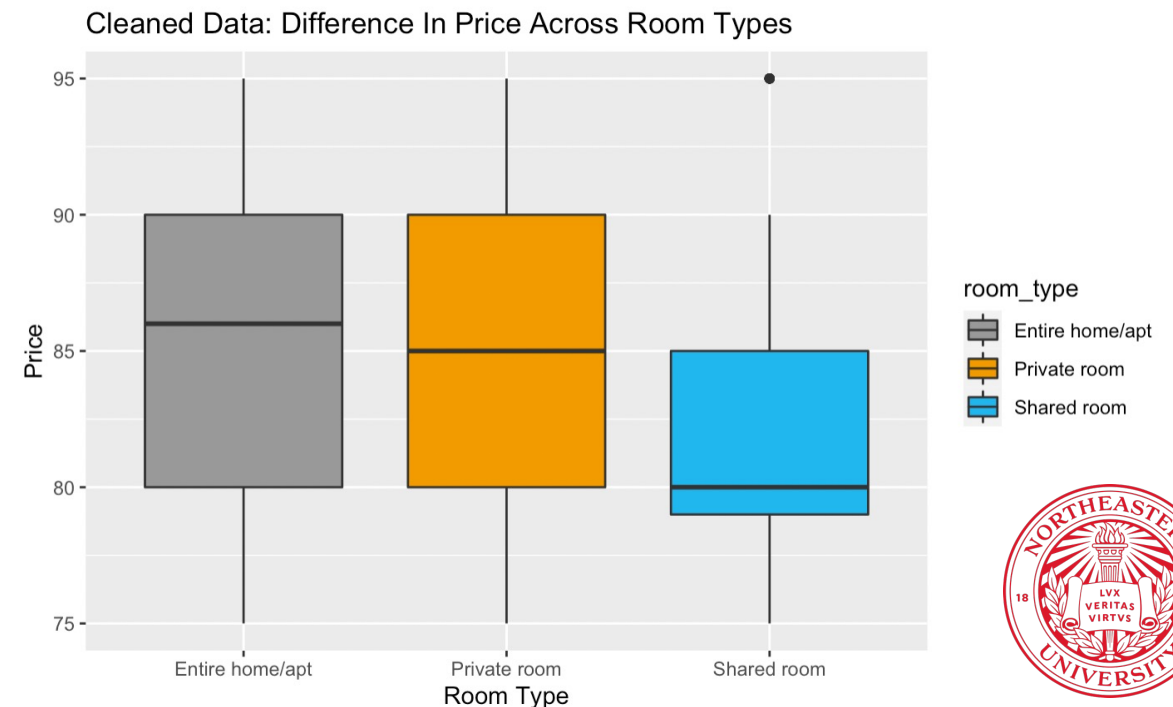
Business Question:

Is there a difference in price across different room types?

Analysis:

Entire home or apartment has the highest median price out of all of them i.e., 86. It also has the least number of outliers whereas Private room has 85 median price.

Shared room has the lowest median price i.e., 80.



Latitude and Longitude Density For Specific Room Types As Per Price

Business Question:

How much is the price density as per room types across a region?

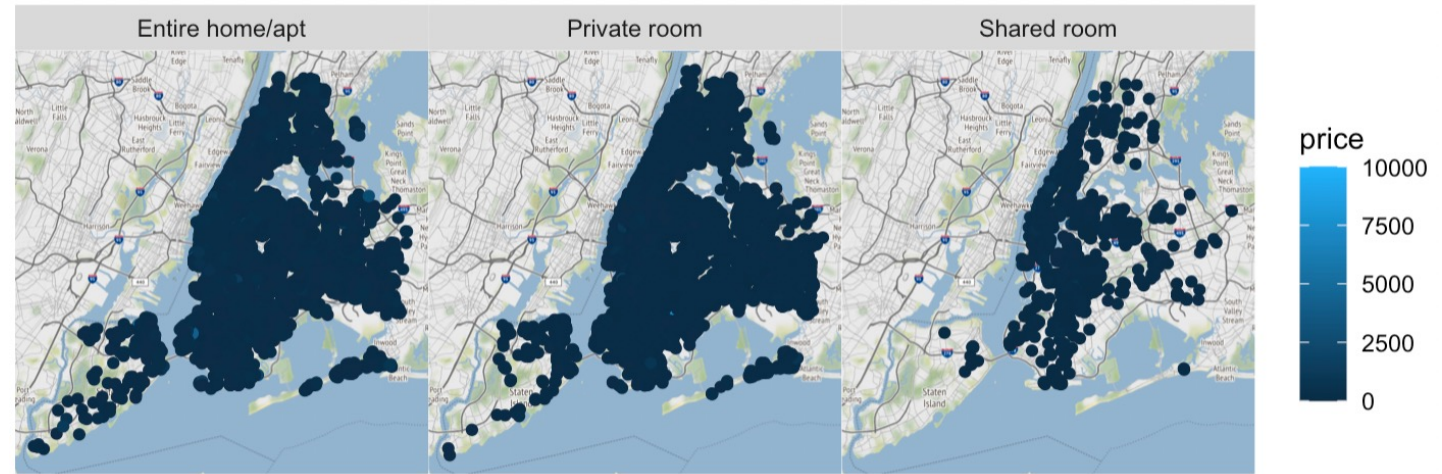
Analysis:

- For **Uncleaned Dataset**, there is lesser density of price in New York city due to lower booking rates. When compared the same to Private room and Entire home or apartment, we see the graph gets denser. However, we need further analysis to get the correct data as it's an uncleaned data set.
- For **Cleaned Dataset**, we can clearly analyse the density of prices across all 3 room types and Private rooms has the highest denser plot in the region.

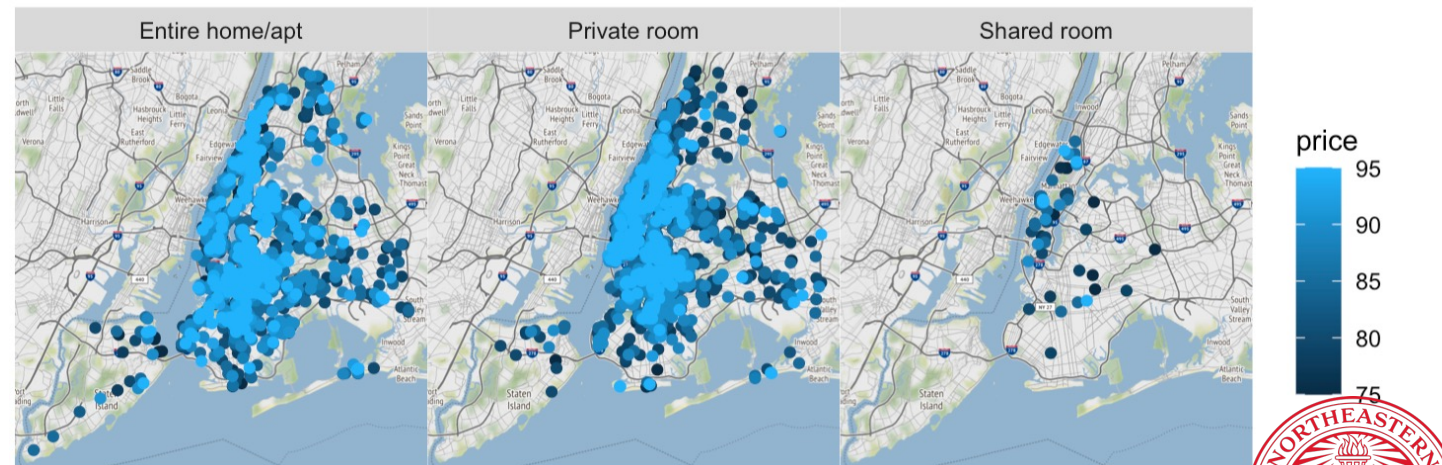
Reason:

- Good hospitality service and corporate discounts.
- Better options on room availability.

Uncleaned Data: Density of Latitude and Longitude



Cleaned Data: Density of Latitude and Longitude



Reviews In Specific Years Based On Room Types From Year 2011-2019

Business Question:

What is the number of reviews for each room type from the year 2011 to 2019?

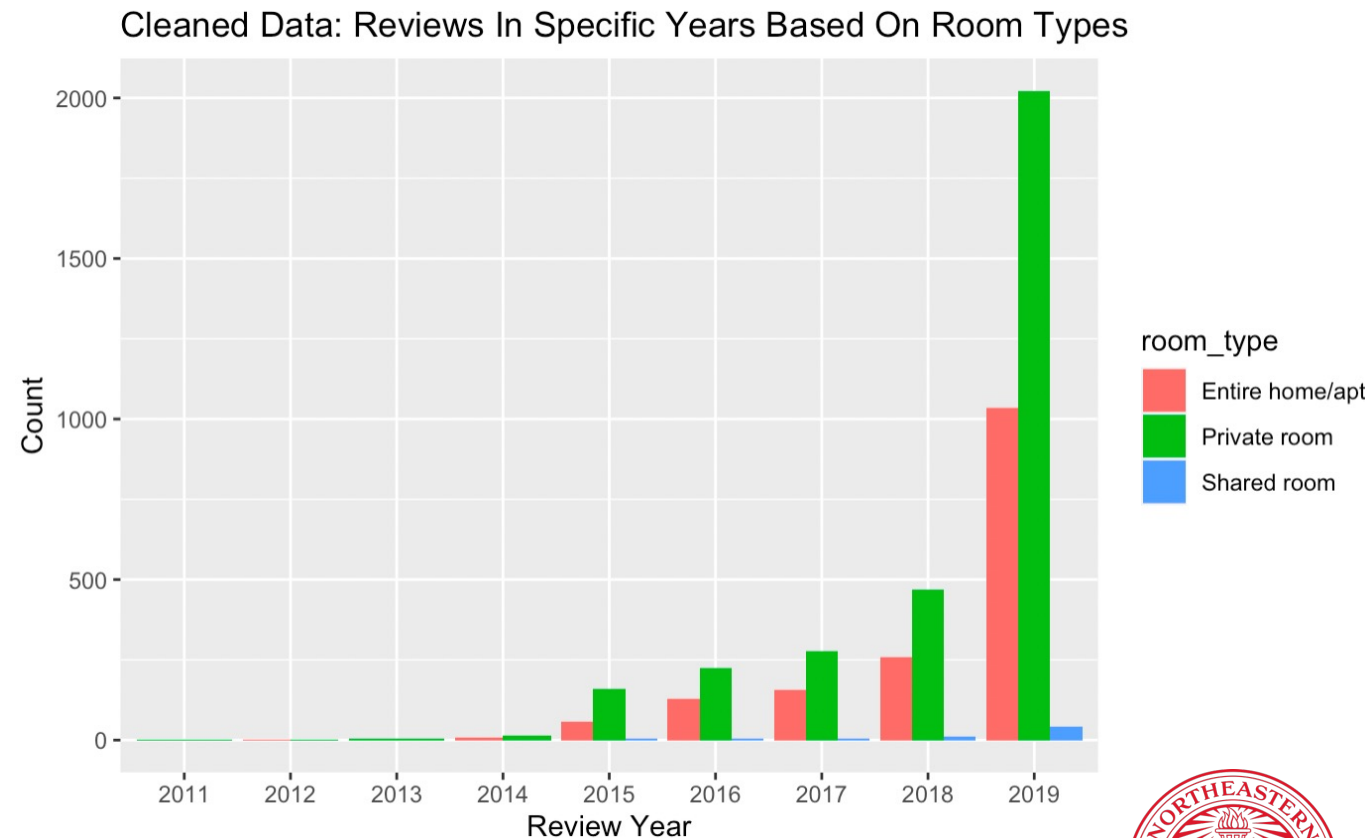
Analysis:

As per the given bar chart, we can clearly depict there is an exponential growth of the reviews received from year 2011-2019. Moreover, we see a huge jump of reviews received in 2019 as compared to the previous year 2018.

For the year 2011-2019, Private room were always in demand and there was an overflowing of reviews received from the customers as compared to other room types after 2015.

Reason:

- Good hospitality service and corporate discounts.
- Breakfast, parking, airport pickup/drop facility.
- Better transport and infrastructure.



References

Data Visualization Tips For More Effective And Engaging Design. (n.d.). Tableau. Retrieved October 29, 2022, from <https://www.tableau.com/learn/articles/data-visualization-tips>

Descriptive Statistics and Graphics - Easy Guides - Wiki - STHDA. (n.d.). Retrieved October 29, 2022, from <http://www.sthda.com/english/wiki/descriptive-statistics-and-graphics>

Kervizic, J. (2021, December 13). *Data Science Career Path & Progression - Hacking Analytics*. Medium. <https://medium.com/analytics-and-data/data-science-career-path-progression-ab5140cfbc84>

Mapping in R: Latitude and Longitude Coordinates. (n.d.). Mapping in R. Retrieved October 29, 2022, from https://map-rfun.library.duke.edu/01_georeference.html

New York City Airbnb Open Data. (2019, August 12). Kaggle. <https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data?resource=download>

R in Action, Second Edition. (n.d.). Manning Publications. Retrieved October 29, 2022, from <https://www.manning.com/books/r-in-action-second-edition>





```
1 def gratitude():  
2     print("Thank you.")  
3
```

Any Questions?



UML Diagrams

Structural

- Profile diagram
- Package diagram
- Class diagram
- Deployment diagram
- Object diagram

Behavioral

- Sequence
- Activity diagram
- Use case diagram
- Communication diagram
- Interaction overview diagram