**Report on AirBnB Listings in NYC**

Zara Atiq

BANA 6380

11/12/2020

Dr. Herbert Remidez

**Introduction:**

Airbnb has been used by guests and hosts worldwide for more than a decade to expand on traveling possibilities to offer an exclusive and personalized way of experiencing the world. The dataset refers to the Airbnb activity in NYC in year 2019. The data used for this report contains information to find out about hosts, prices, geographical availability, and all other necessary metrics to make predictions and draw conclusions about Airbnb listings in New York City. This report would interest specifically those who work in the hospitality business and also those who are frequent travelers.

**Data:**

Data used for this report has been sourced from [Kaggle](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data) and has 48,896 observations. It includes information and metrics from Airbnb listings in New York City by guest and host IDs. The dataset gives insight into prices for different kinds of Airbnb spaces available in the city: entire homes/ apartments, private rooms and shared rooms. The dataset desegregates this information by the five boroughs of city referred to in the records as neighborhood groups. All observations from the dataset have been used to run reports and to provide with a complete picture of the Airbnb listings. To gauge Airbnb activity so that both hosts and guest can make better decisions on how to market their rent spaces. While it also helps guests in identifying which neighborhood provide the most cost effective Airbnb/rental spaces. The public data available is part of Airbnb and the original dataset can be found on this [website](http://insideairbnb.com/).

Raw information gathered through Kaggle consisted of 16 columns and eight of them have mainly been used for this report, which includes:

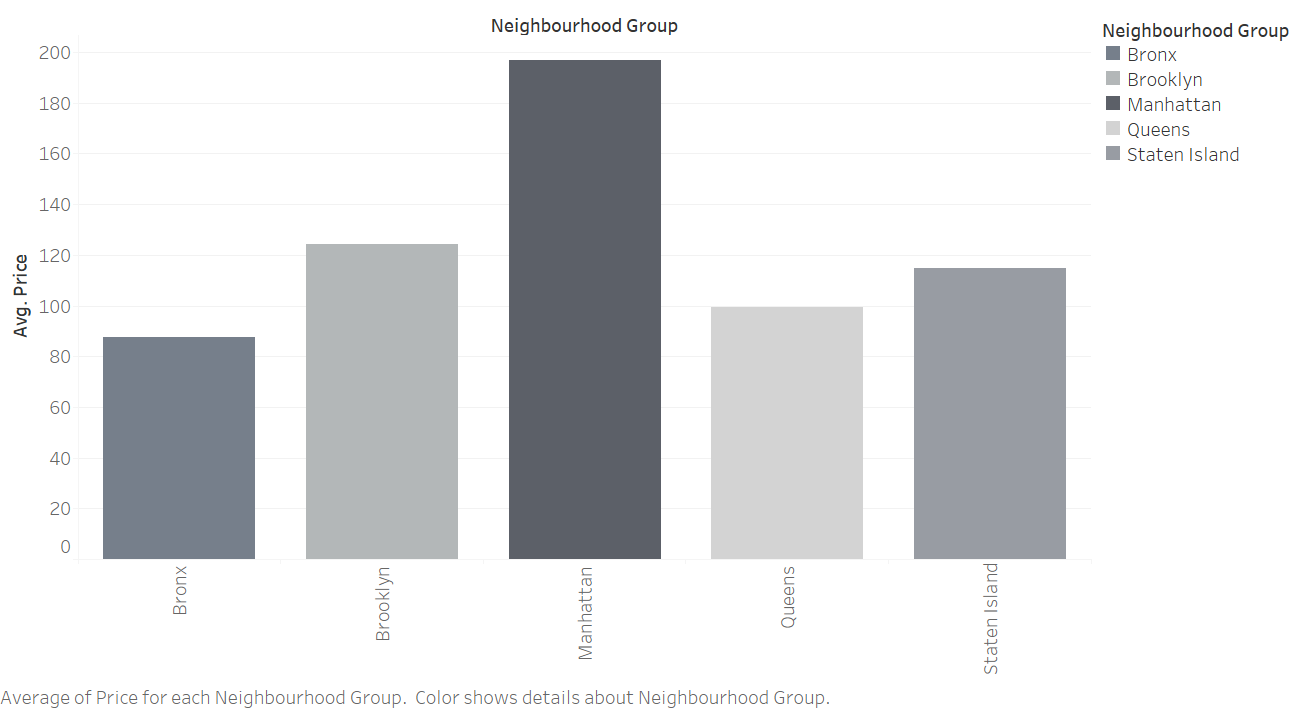
* Neighborhood group- Referred to the Boroughs in NYC
* Calculated host listings- Number of listings per host
* Listing- Refers to the ad posted on Airbnb
* Host names/IDs- Name if the hosts and their IDs
* Price- All prices in US dollars
* Room type- Refers to the listing space type (entire home/apartment, shared room, and private room)
* Availability 365- Refers to the number of days listing is available for.
* Number of reviews- Refers to the amount of reviews a certain location or host has received from a customer after their stay.

**Analysis and Results:**

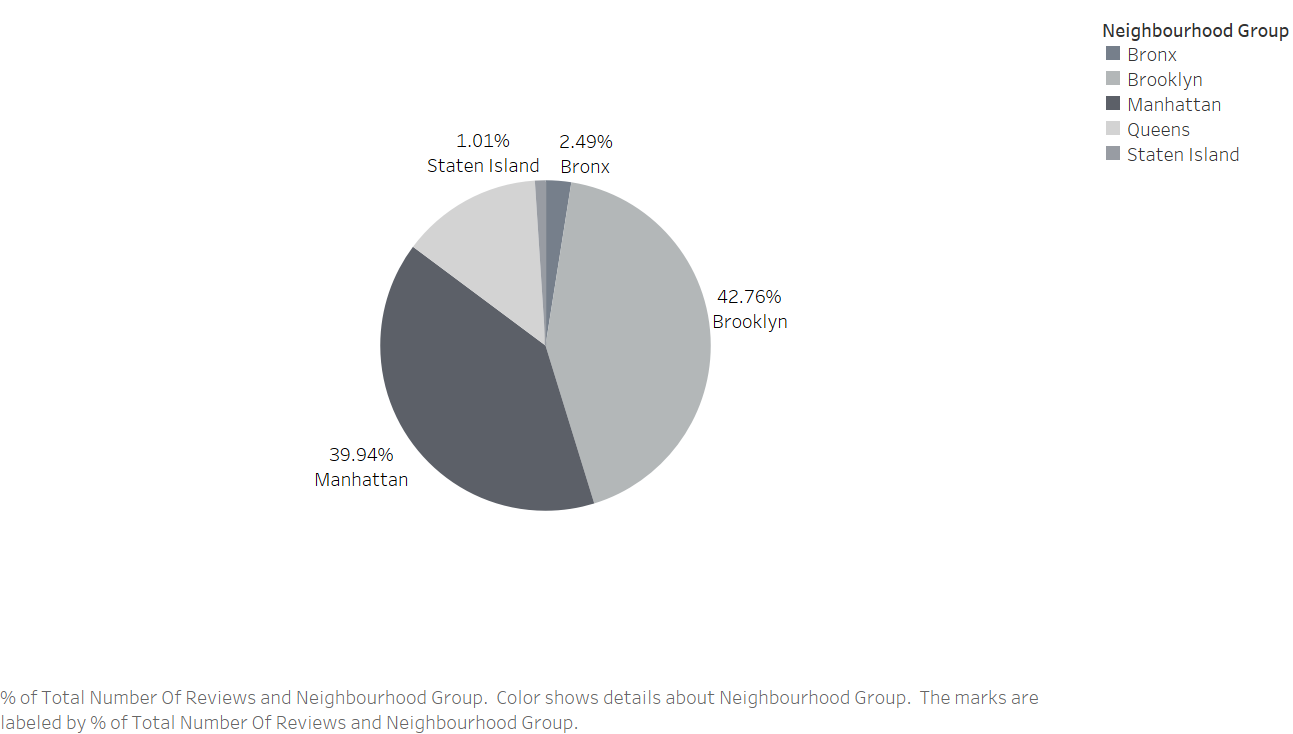
The raw data has been filtered and analyzed in SAS Studio by running SQL queries, while the visual analysis and assertions have been made using Tableau. The first part of the SQL process mainly consists of creating library, uploading the data from an excel file, running reports with joining tables, and using set operators in the SQL procedure. The SQL procedure also consists of statements that generate summary reports with joining tables, construction of sub setting data with and without correlated subqueries, inserting and deleting rows from a table. The final part of the SQL procedure solves problems using SQL process and running of traditional SAS programming techniques.

In addition to the SQL procedures Macros have been created and the technique has been used to automate programs by defining and calling macros. An understanding of macro functions has been portrayed in the later part of the code sheet. Data driven programs have also been created using SAS Macro language. The code sheet also demonstrates the usage of different system options available for macro debugging and displaying values of user-defined and automatic macro variables in SAS. Tableau has been used specifically to create visualizations to analyze Airbnb activity in New York City. The price differences, availability of the accommodations based on locations all year round, defining busy neighborhoods based on customer reviews have been illustrated through treemaps, horizontal bars, bar charts, and pie chart.

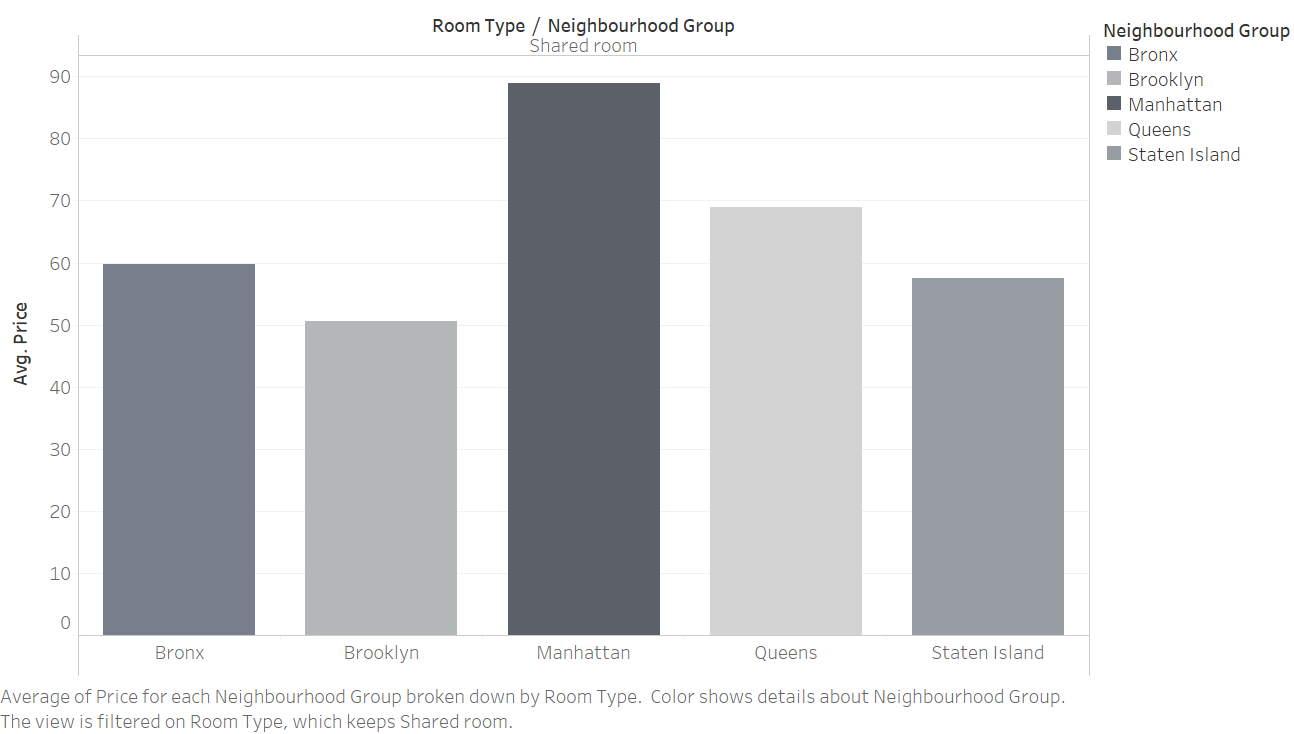
**Figure.1** below illustrates that across all room types Manhattan is the most expensive where the average cost per night is $196. According to the bar chart below the second most expensive boroughs is Brooklyn, followed by Staten Island, and Queens.



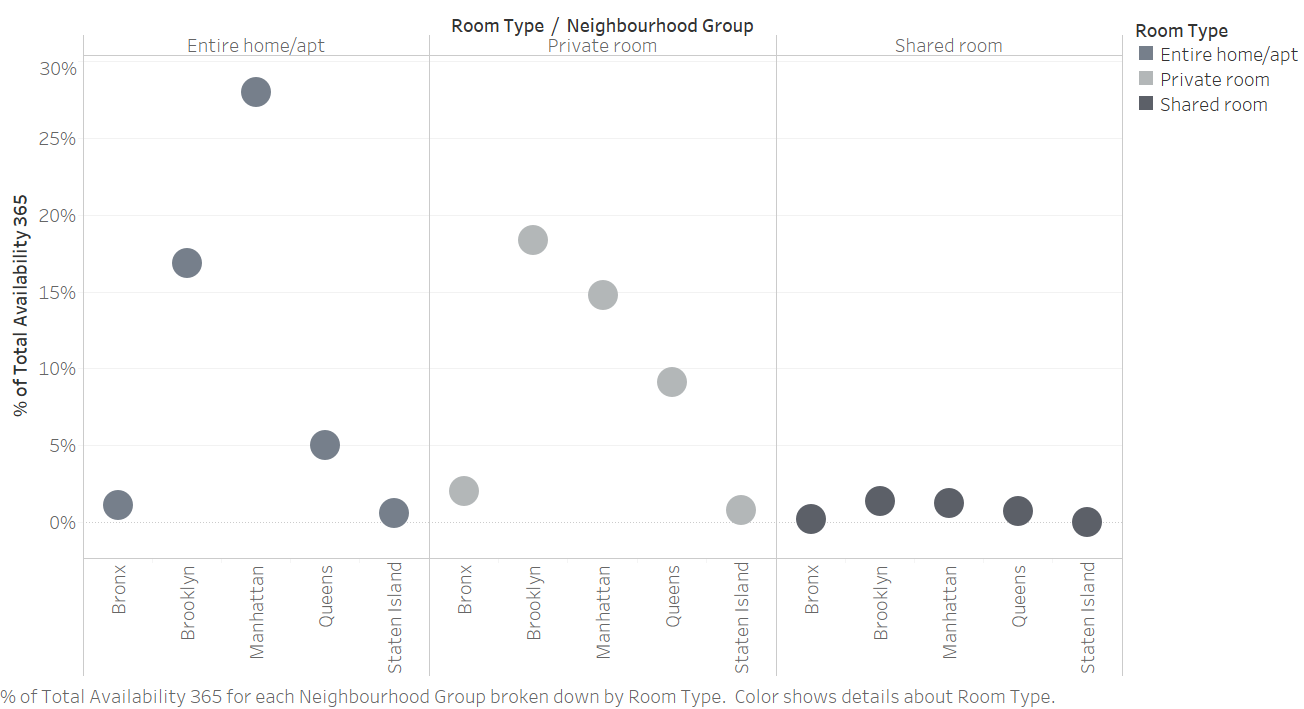
**Figure.2** below shows that Brooklyn Airbnb rentals are the busiest in the city. According to the pie chart 42.76% of the Airbnb business in NYC is generated from Brooklyn. The illustration is based on the number of reviews received by customers after their stay. This may also be because of the area being less expensive compared to Manhattan as indicated figure. 1 above. To describe the findings precisely, Airbnb rentals in Brooklyn form 42.76% of the total reviews received in the city followed by 39.9% from Manhattan, 13.8% from Queens, 2.49% from Bronx, and 1% from Staten Island.



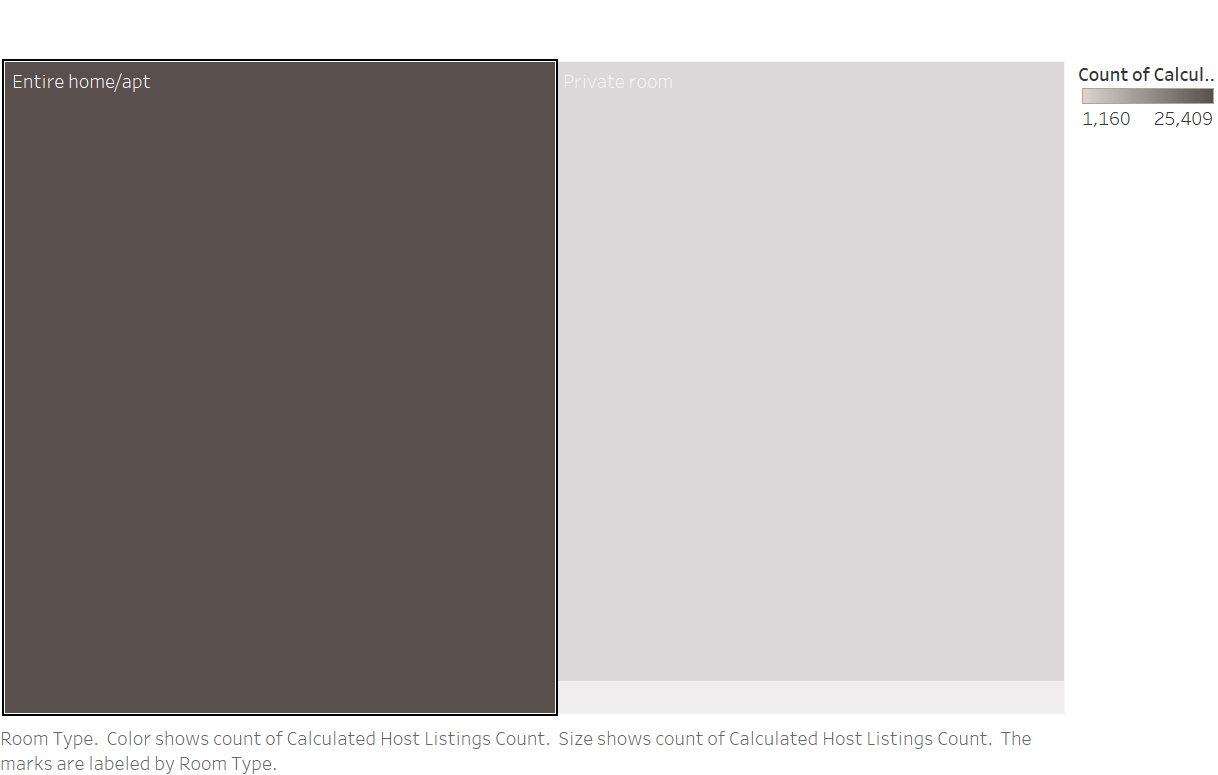
**Figure 3.** showsthe price difference of a shared room on average in all neighborhoods. From the bar chart below we can deduce that Brooklyn has the cheapest accommodation of shared rooms in NYC. The average price of a shared room in this neighborhood is $50. Offering the cheapest accommodation is also one of the reasons for this borough to be busiest in terms of Airbnb activity.

****

**Figure 4.** below showcases the frequency with which each type of Airbnb space- entire homes/apartments, private rooms, and shared rooms- are available for renting through the website. It illustrates that entire home/apartment in Manhattan are most widely available for people to make bookings all year round. The price and bigger space factors could also be one of the many reasons to conclude that these spaces are most widely available and less busy as compared to the rest.



The hosts put up their advertisements, also called “Host Listings”. These listings provide information about the rooms, amenities and services provided by the host as well as the associated price. These listings can be updated multiple times to either remain active on the website or to attract more customers. **Figure. 5** below shows that hosts of entire home/apartments post the most listings all year round.



**Conclusion:**

The report is based on a dataset of 48,896 observations with sixteen variables sourced from Kaggle which focuses on the price difference according to room types and locations in New York City. The data gathered is from the year 2019 and interests specifically those in hospitality business and those who travel frequently. Insights gained from the dataset has helped in creating a quantitative report. To delve further into how Airbnb market operates in the New York City and to provide multiple marketing strategies to hosts in respective areas to improve more data would be required. Host and guest demographics and qualitative measure of guest reviews and feedback would give a better understanding of the Airbnb activity in NYC.

**Appendix:**

*/\*Creating Library and Uploading Data\*/*

%let path=/home/u45075787/termproject;

libname Term "/home/u45075787/termproject";

proc import datafile= "/home/u45075787/termproject/AB\_NYC\_2019\_2\_V1.xlsx"

out= term.projectV1

dbms=xlsx;

proc import datafile= "/home/u45075787/termproject/AB\_NYC\_2019\_2\_V2.xlsx"

out= term.projectV2

dbms=xlsx;

proc import datafile= "/home/u45075787/termproject/AB\_NYC\_2019\_H1.xlsx"

out= term.projectH1

dbms=xlsx;

proc import datafile= "/home/u45075787/termproject/AB\_NYC\_2019\_H2.xlsx"

out= term.projectH2

dbms=xlsx;

*///\*Points 1-8\*///*

*/\*Joining Tables and Using Set Operators\*/*

proc sql;

create table term.NYC as

select\*

from term.projectV1 as a inner join term.projectV2 as b on a.host\_id=b.host\_id;

quit;

proc sql;

create table term.NYC2 as

select\*

from term.projectH1 as a inner join term.projectH2 as b on a.host\_id=b.host\_id;

quit;

proc sql;

create table term.NYCUNION as

select\*

from term.NYC

union corr

select\*

from term.NYC2;

quit;

*/\*Using the Select Statement to Select Columns in the Table\*/*

proc sql inobs=100;

select name, host\_name, neighbourhood\_group, neighbourhood, room\_type, price, number\_of\_reviews, calculated\_host\_listings\_count, availability\_365

from term.NYCUNION;

quit;

*/\*Creating a New Column\*/*

proc sql;

alter table term.NYCUNION add SerialNumber INTEGER;

quit;

*/\*Retrieving Rows that Satisfy a Condition\*/*

proc sql;

title 'Reviews of AirBnB Hosts in Brooklyn';

select name, host\_name, number\_of\_reviews, neighbourhood\_group

from term.NYCUNION

where neighbourhood\_group ='Brooklyn';

quit;

*/\*Sorting the Number of Reviews Received by each Host in Descending Order\*/*

proc sql;

title 'AirBnB Hosts in NYC Sorted by Reviews';

select name, host\_name, number\_of\_reviews, neighbourhood\_group

from term.NYCUNION

order by number\_of\_reviews desc;

quit;

*/\*Points 9-12 and 42\*/*

*/\*Summarizing Data- Prices in Manhattan for Private Rooms\*/*

options symbolgen;

%macro N\_G;

%local neighbourhood\_group;

%let neighbourhood\_group=Manhattan;

proc sql;

title 'Prices of Private Rooms in Manhattan';

select name, neighbourhood\_group, room\_type, price

from term.NYCUNION

where room\_type='Private room' and neighbourhood\_group='Manhattan'

order by price desc;

quit;

%mend N\_G;

options nosymbolgen;

*/\*Grouping Data\*/*

proc sql;

title 'Total Number of Hosts in Every NYC Neighbourhood Group';

select neighbourhood\_group, count(host\_name) as TotalHosts

from term.NYCUNION

where host\_name is not missing

group by neighbourhood\_group;

quit;

*/\*Filtering and Subsetting Grouped Data\*/*

proc sql;

title 'Least Number of AirBnB Hosts in NYC';

select neighbourhood\_group, count (\*) as TotalHosts

from term.NYCUNION

where host\_name is not missing

group by neighbourhood\_group

having TotalHosts not >1000;

quit;

*/\*Points 13 & 14\*/*

*/\*Subsetting data using correlated queries\*/*

proc sql;

SELECT neighbourhood\_group, room\_type, price

FROM term.NYCUNION as out WHERE room\_type in (SELECT room\_type

FROM term.NYCUNION

WHERE price=out.price);

quit;

*/\*In line views with other tables or views\*/*

proc sql;

create table reviews as

select reviews(timepart(last\_review)) as last\_review, count(\*) as TotalReviews from (select last\_review from term.NYCUNION)

group by neighbourhood\_group;

quit;

*/\*Points 15-21\*/*

*/\*Insert rows into tables\*/*

proc sql;

create table term.NYCUNION2

like term.NYCUNION;

quit;

proc sql;

insert into term.NYCUNION2

select \* from term.NYCUNION

where number\_of\_reviews > 50;

quit;

*/\*Updating data values in a table\*/*

proc sql;

create table term.NYCUNION2 like term.NYCUNION;

insert into term.NYCUNION2

select \* from term.NYCUNION

where number\_of\_reviews >50;

proc sql;

update term.NYCUNION2

set reviews\_per\_month=reviews\_per\_month\*30;

title "Updated Number of Reviews per Month";

select reviews\_per\_month format=comma10.0

from term.NYCUNION2;

quit;

*/\*Delete rows\*/*

proc sql;

delete from term.NYCUNION2

where number\_of\_reviews >5;

quit;

*/\*Altering Columns\*/*

proc sql;

alter table term.NYCUNION2 add neighbourhood\_group2 char(5);

quit;

/\*Creating an Index\*/

proc sql;

create index neighbourhood\_group2 on term.NYCUNION2(neighbourhood\_group2);

quit;

*/\*Deleting a table\*/*

proc sql;

drop table term.NYCUNION2;

quit;

*/\*Describing table\*/*

proc sql;

describe table term.NYCUNION;

quit;

*/\*Points 22, 23, 36\*/*

%let foot=%str(footnote "Report Generation Date: &SYSDATE Time: &systime";);

%let clear=%str(title;footnote;);

*/\* Point 24 & 26\*/*

*/\*Using %GLOBAL statement\*/*

%global dummy\_table;

%let dummy\_table=term.NYCUNION2;

*/\*Using INTO clause\*/*

proc sql noprint;

select count(\*), price as total\_revenue

into :price

from term.NYCUNION

where price=1

group by price;

select count(\*), minimum\_nights as total\_revenue

into :minimum\_nights

from term.NYCUNION

where minimum\_nights=1

group by minimum\_nights;

quit;

proc sql;

create table total\_revenue

(revenue char(20),

total\_price num format=7.);

quit;

*/\*Points 27,43,35\*/*

%macro dummy(t\_name);

%local tb\_name;

%let tb\_name=%upcase(&t\_name);

proc sql inobs=50;

create table &tb\_name as

select \* from term.NYCUNION;

quit;

%put &=tb\_name;

%mend dummy;

*/\*Point 28, 29, 38\*/*

*/\*28\*/*

data total\_revenue;

set total\_revenue;

call symput(price, total\_price);

run;

data total\_revenue1;

set total\_revenue;

call symputx(price, total\_price);

run;

*/\*29,38\*/*

data total\_revenue;

set total\_revenue;

price=symget(price);

run;

*/\* Point 30, 31\*/*

%macro neighbourhood\_group (Brooklyn, Queens);

*/\*Point 32\*/*

/\*This macro selects neighbourhood\_group locations Brooklyn and Queens\*/

*/\*Point 33\*/*

%macro test(var1,var2,var3);

%put &=var1;

%put &=var2;

%put &=var3;

%mend test;

*/\*Point 34\*/*

%macro price(level);

%if &level=1 %then %do;

proc sql;

update term.NYCUNION2

set price="Cheap" where price=&level;

quit;

%end;

%else %if &level=50 %then %do;

proc sql;

update term.NYCUNION2

set price="Expensive" where price=&level;

quit;

%end;

%else %if &level=250 %then %do;

proc sql;

update term.NYCUNION2

set price="Medium" where price=&level;

quit;

%end;

%else %if &level=100 %then %do;

%end;

%mend price;

*/\* Point 37\*/*

%let price=50\*2;

%let reviews\_per\_month=2\*30;

%let eval\_price=%eval(&price);

%let eval\_reviews\_per\_month=%eval(&reviews\_per\_month);

%put &price is &eval\_price;

%put &reviews\_per\_month is &eval\_reviews\_per\_month;

*/\* Points 39,40\*/*

options mlogic;

%dummy(&dummy\_table);

options nomlogic;

proc sql;

alter table term.NYCUNION2 drop number\_of\_reviews;

alter table term.NYCUNION2 add NUMBER\_OF\_REVIEWS char(10);

quit;

*/\*41\*/*

options mprint;

%poi(&amenity, &bump, &crossing, &give\_way, &junction, &no\_exit, &railway, &roundabout, &station, &stop, &traffic\_calming, &traffic\_signal);

options nomprint;

*/\* Point 44 \*/*

proc sql noprint;

select distinct number\_of\_reviews

into :number\_of\_rev separated by "," from term.NYCUNION;

quit;

%put &=number\_of\_rev;

*/\* Point 45\*/*

proc sql noprint;

select \* from &&tb&price;

quit;

*/\*Point 46\*/*

%macro loop(finish);

%let price=1;

%DO %while (&price<&finish);

%price(&price);

%let price=%eval(&price+1);

%END;

%mend loop;

%loop(5);

**References:**

Dgomonov. (2019, August 12). New York City Airbnb Open Data. Retrieved November 12, 2020, from <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

Inside Airbnb. Adding data to the debate. (n.d.). Retrieved November 12, 2020, from http://insideairbnb.com/