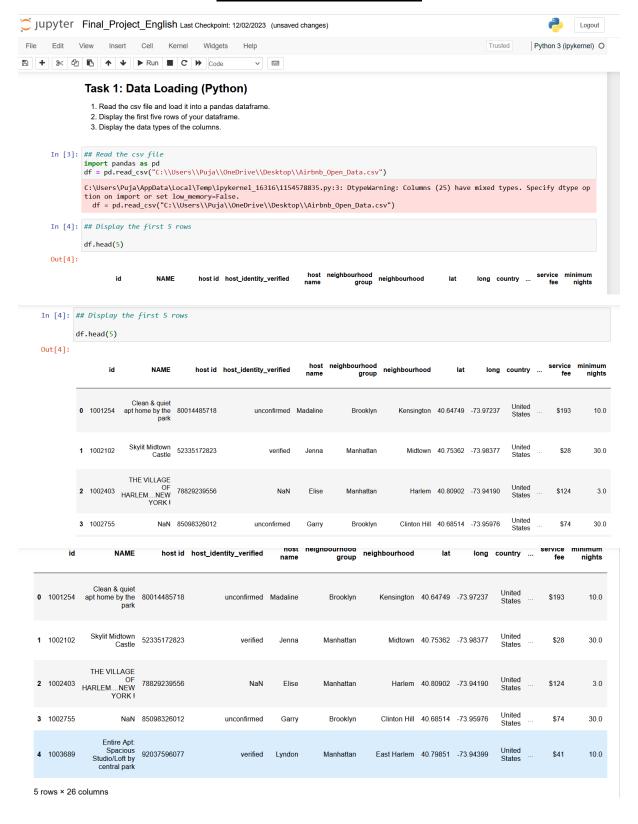
Final Project Snapshots



In [8]: ## Display the data types df.dtypes

dtype: object

Out[8]: id int64 NAME object int64 host id host_identity_verified object host name object neighbourhood group object neighbourhood object lat float64 long float64 country object country code object instant_bookable object cancellation_policy object room type object Construction year float64 price object service fee object minimum nights float64 number of reviews float64 object last review reviews per month float64 float64 float64 float64 review rate number calculated host listings count availability 365 house_rules object license object

```
In [9]: df.info()
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 102599 entries, 0 to 102598
       Data columns (total 26 columns):
           Column
                                           Non-Null Count
                                                           Dtype
        _ _ _
                                           _____
        0
            id
                                           102599 non-null int64
                                           102349 non-null object
        1
            NAME
                                           102599 non-null int64
        2
            host id
        3
                                          102310 non-null object
            host_identity_verified
        4
                                          102193 non-null object
            host name
        5
                                          102570 non-null object
            neighbourhood group
        6
            neighbourhood
                                          102583 non-null object
        7
            lat
                                          102591 non-null float64
        8
            long
                                          102591 non-null float64
        9
            country
                                          102067 non-null object
        10 country code
                                         102468 non-null object
        11 instant_bookable
                                          102494 non-null object
                                      102523 non-null object
        12 cancellation_policy
                                          102599 non-null object
        13 room type
        14 Construction year
                                          102385 non-null float64
        15 price
                                          102352 non-null object
                                          102326 non-null object
        16 service fee
        17 minimum nights
                                           102190 non-null float64
        18 number of reviews
                                          102416 non-null float64
        19 last review
                                          86706 non-null object
                                          86720 non-null
        20 reviews per month
                                                          float64
        21 review rate number
                                          102273 non-null float64
        22 calculated host listings count 102280 non-null float64
        23 availability 365
                                          102151 non-null float64
        24 house_rules
                                          50468 non-null
                                                          object
        25 license
                                          2 non-null
                                                          object
       dtypes: float64(9), int64(2), object(15)
       memory usage: 20.4+ MB
```

Task 2a: Data Cleaning (Any Tool)

- 1. Drop some of the unwanted columns. These include host id, id, country and country code from the dataset.
- 2. State the reason for not including these columns for your Data Analytics.

If using Python for this exercise, please include the code in the cells below. If using any other tool, please include screenshoots before and after the elimination of the columns.

```
In [10]: df = df.drop(columns= ["host id","id","country","country code"], axis=1, inplace=True)

In []: #Reason for dropping the columns: We do not need host id, id, country and country code for analysis on the bsais of mentioned col

In [36]: # Hence, dropped the unnecessary columnns.
```

Task 2b: Data Cleaning (Python)

- Check for missing values in the dataframe and display the count in ascending order. If the values are missing, impute the values as per the datatype of the columns.
- Check whether there are any duplicate values in the dataframe and, if present, remove them.
- Display the total number of records in the dataframe before and after removing the duplicates.

```
In [6]: ## Check for missing values in the dataframe and display the count in ascending order.
          df.isnull().sum().sort_values(ascending=False)
Out[6]: license
          house_rules
last review
                                                    52131
15893
          reviews per month
                                                    15879
          country
availability 365
                                                      532
448
          minimum nights
                                                      406
          host name
          review rate number
calculated host listings count
                                                      326
319
          host_identity_verified service fee
                                                      289
          NAME
                                                      250
                                                      247
214
          Construction year
```

```
Out[6]: license
                                           102597
        house_rules
                                            52131
        last review
                                             15893
                                            15879
        reviews per month
        country
                                               532
        availability 365
                                               448
        minimum nights
                                               409
        host name
                                               406
        review rate number
                                               326
        calculated host listings count
                                               319
                                               289
        host_identity_verified
        service fee
                                               273
        NAME
                                               250
        price
                                               247
        Construction year
                                               214
        number of reviews
                                               183
        country code
                                               131
        instant bookable
                                               105
        cancellation_policy
                                               76
        neighbourhood group
                                                29
        neighbourhood
                                                16
        long
                                                 8
        lat
                                                 8
        id
                                                 0
        host id
                                                 0
        room type
                                                 0
        dtype: int64
```

```
In [7]: ## Check whether there are any duplicate values in the dataframe and if present remove them.
df.duplicated().sum()
```

Out[7]: 541

```
In [28]: ## Display the total number of records in the dataframe after removing the duplicates.
            df.shape()
df.drop_duplicates()
Out[28]:
                                                                                             host
name
                             id
                                           NAME
                                                         host id host_identity_verified
                                                                                                                     neighbourhood
                                                                                                                                             lat
                                                                                                                                                       long country ...
                                  Clean & quiet apt home by the park
                                                                                                                                                               United
                      1001254
                                                   80014485718
                                                                             unconfirmed Madaline
                                                                                                            Brooklyn
                                                                                                                           Kensington 40.64749 -73.97237
                                                                                                                                                                             $193
                                    Skylit Midtown
                                                                                                                                                               United
                   1 1002102
                                                   52335172823
                                                                                 verified
                                                                                            Jenna
                                                                                                          Manhattan
                                                                                                                             Midtown 40.75362 -73.98377
                                                                                                                                                                              $28
                                            Castle
                                    THE VILLAGE
                                 OF
HARLEM....NEW
YORK!
                                                                                                                                                               United
                   2 1002403
                                                   78829239556
                                                                                    NaN
                                                                                             Elise
                                                                                                          Manhattan
                                                                                                                              Harlem 40.80902 -73.94190
                                                                                                                                                                             $124
                   3 1002755
                                             NaN 85098326012
                                                                                                                           Clinton Hill 40.68514 -73.95976
                                                                                                                                                               United
                                                                             unconfirmed
                                                                                             Garry
                                                                                                           Brooklyn
                                        Entire Apt
                      1003689
                                     Spacious
Studio/Loft by
                                                    92037596077
                                                                                                                         East Harlem 40.79851 -73.94399
                                                                                                                                                               United
                                                                                                                                                                              $41
                                                                                 verified
                                                                                                          Manhattan
                                                                                            Lyndon
                                                                                                                                                               States
                                       central park
                                  Cozy bright room
near Prospect 77326652202
Park
              102053 57365208
                                                                             unconfirmed
                                                                                           Mariam
                                                                                                            Brooklyn
                                                                                                                             Flatbush 40.64945 -73.96108
                                                                                                                                                                             NaN
                                   Private Bedroom
with Amazing 45936254757
Rooftop View
                                                                                                                                                               United
States
              102054 57365760
                                                                                 verified
                                                                                              Trey
                                                                                                            Brooklyn
                                                                                                                            Bushwick 40.69872 -73.92718
                                                                                                                                                                             NaN
                                   Pretty Brooklyn
One-Bedroom 23801060917
for 2 to 4 people
                                                                                                                             Bedford-
                                                                                                                                                               United
              102055 57366313
                                                                                  verified Michael
                                                                                                            Brooklyn
                                                                                                                                      40.67810 -73.90822
                                                                                                                                                                             NaN
                                   Room & private
bathroom in 15593031571
                                                                                                                                                               United
              102056 57366865
                                                                                                                              Harlem 40.81248 -73.94317
                                                                             unconfirmed
                                                                                           Shireen
                                                                                                          Manhattan
                                                                                                                                                                             NaN
                                                                                                                                                               States
                                    historic Harlem
                                                                                                                                                               United
              102057 57367417
                                   Rosalee Stewart 93578954226
                                                                                            Stanley
                                                                                                                              Harlem 40.81315 -73.94747
                                                                                  verified
                                                                                                          Manhattan
                                   Room & private
bathroom in
historic Harlem
              102056 57366865
                                                                             unconfirmed
                                                                                            Shireen
                                                                                                          Manhattan
                                                                                                                               Harlem 40.81248 -73.94317
                                                                                                                                                                              NaN
                                                                                                                                                               United
              102057 57367417 Rosalee Stewart 93578954226
                                                                                  verified Stanley
                                                                                                          Manhattan
                                                                                                                               Harlem 40.81315 -73.94747
                                                                                                                                                                              NaN
             102058 rows × 26 columns
```

Task 3: Data Transformation (Any Tool)

- \bullet Rename the column availability 365 to days_booked
- Convert all column names to lowercase and replace the spaces in the column names with an underscore "_".
- Remove the dollar sign and comma from the columns <code>price</code> and <code>service_fee</code>. If necessary, convert these two columns to the appropriate data type.

If using Python for this exercise, please include the code in the cells below. If using any other tool, please include screenshoots of your work.

```
In [23]: ## Rename the column.
    df.rename(columns={'availability 365' : 'days_booked'} , inplace=True)
```

Task 3: Data Transformation (Any Tool)

- Rename the column availability 365 to days_booked
- Convert all column names to lowercase and replace the spaces in the column names with an underscore "_".
- Remove the dollar sign and comma from the columns 'price' and service_fee'. If necessary, convert these two columns to the appropriate data type.

If using Python for this exercise, please include the code in the cells below. If using any other tool, please include screenshoots of your work.

Task 4: Exploratory Data Analysis (Any Tool)

- · List the count of various room types avaliable in the dataset.
- Which room type has the most strict cancellation policy?

In [23]: ## List the count of various room types available with Airbnb

• List the average price per neighborhood group, and highlight the most expensive neighborhood to rent from.

If using Python for this exercise, please include the code in the cells below. If using any other tool, please include screenshoots of your work.

```
import numpy as np
       import math
      df['room_type'].value_counts()
Out[23]: room_type
       Entire home/apt
                     53701
       Private room
                     46556
       Shared room
       Hotel room
                      116
       Name: count, dtvpe: int64
In [58]: ## Which room type adheres to more strict cancellation policy
          df1= df[df['cancellation_policy']=='strict']
          print(df1)
                                                                   host id \
                      id
                                                         name
          0
                  1001254
                            Clean & quiet apt home by the park 80014485718
          8
                  1005754
                             Large Furnished Room Near B'way 79384379533
          9
                  1006307
                            Cozy Clean Guest Room - Family Apt 75527839483
                            Central Manhattan/near Broadway 88136055909
                  1007964
          12
          24
                  1014591
                                        CBG Helps Haiti Rm #2 57827496420
                      . . .
                                                           . . .
          102586 6088019
                                Patio Oasis in Upper East Side 41141109074
          102587 6088571
                            Adorable One-Bed in Williamsburg! 20914678851
          102591 6090781
                          Columbus Ave Apt 1 block from Park 50908010324
          102593 6091885 Welcoming, Clean, Cheap on St Marks 33188605074
                             Big Studio-One Stop from Midtown 11160591270
          102597 6094094
                 host_identity_verified
                                          host_name neighbourhood_group \
          0
                            unconfirmed
                                           Madaline
                                                               Brooklyn
          8
                               verified
                                             Evelyn
                                                              Manhattan
          9
                            unconfirmed
                                               Carl
                                                              Manhattan
                               verified
          12
                                                NaN
                                                              Manhattan
    24
                      unconfirmed
                                     Charlotte
                                                            Brooklyn
                         verified
                                       Margaret
    102586
                                                           Manhattan
    102587
                         verified
                                         Taylor
                                                            Brooklyn
                         verified
    102591
                                       Lawrence
                                                           Manhattan
    102593
                         verified
                                         Felipe
                                                           Manhattan
    102597
                      unconfirmed Christopher
                                                              Queens
               neighbourhood
                                    lat
                                             long
                                                          country ... service_fee \
    0
                  Kensington 40.64749 -73.97237 United States ...
                                                                             $193
    8
              Hell's Kitchen 40.76489 -73.98493
                                                   United States ...
                                                                             $204
                                                   United States ...
    9
             Upper West Side 40.80178 -73.96723
                                                                              $58
    12
              Hell's Kitchen 40.76076 -73.98867
                                                   United States
                                                                             $143
    24
                  Park Slope 40.68001 -73.97865 United States ...
                                                                             $163
             Upper East Side 40.77703 -73.94773 United States ...
    102586
                                                                              $42
                                                   United States ...
    102587
                Williamsburg 40.71687 -73.94656
                                                                              $78
             Upper West Side 40.77408 -73.98181 United States ...
    102591
                                                                             $228
                East Village 40.72826 -73.98422 United States ...
    102593
                                                                             $220
```

Long Island City 40.74989 -73.93777 United States ...

\$109

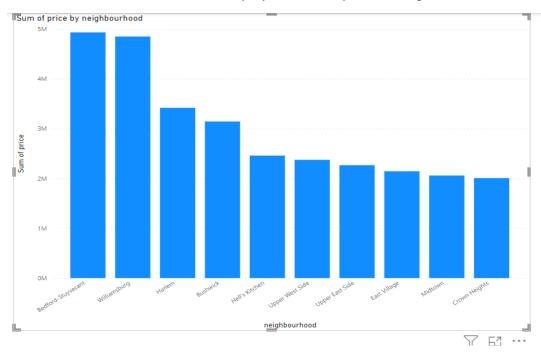
```
minimum_nights number_of_reviews last_review reviews_per_month \
                                   9.0 10/19/2021
0
                 10.0
                                                                   0.21
8
                  2.0
                                  430.0 6/24/2019
                                                                   3.47
                                  118.0 7/21/2017
                                                                   0.99
9
                  2.0
                                         6/23/2019
12
                  2.0
                                  188.0
                                                                   1.50
                  2.0
                                   71.0
                                          7/2/2019
                                                                   0.61
24
                                     ...
                                          8/8/2016
                  4.0
                                                                   0.17
102586
                                    8.0
102587
                  1.0
                                   66.0 6/16/2019
                                                                   1.35
102591
                  5.0
                                   17.0
                                          1/4/2019
                                                                   0.35
102593
                  1.0
                                    8.0
                                          9/6/2015
                                                                   0.16
102597
                  2.0
                                     5.0 10/11/2015
                                                                   0.10
       review_rate_number calculated_host_listings_count days_booked \
0
                      4.0
                                                      6.0
                                                                 286.0
                      3.0
8
                                                      1.0
9
                      5.0
                                                      1.0
                                                                 375.0
                                                                 258.0
12
                      4.0
                                                      1.0
24
                      4.0
                                                      6.0
                                                                 344.0
                       . . .
                                                      . . .
                      5.0
                                                                 159.0
102586
                                                      1.0
                                                                352.0
102587
                                                      1.0
                      2.0
                                               house_rules license
а
        Clean up and treat the home the way you'd like...
        - Please clean up after yourself when using th...
8
                                                                NaN
        NO SMOKING OR PETS ANYWHERE ON THE PROPERTY 1....
12
        - One of the bedroom closets is not accessible...
                                                                NaN
        We take great care of our home and expect you \dots
24
                                                                NaN
. . .
                                                                . . .
102586 - NO SMOKING (You'll have to do that outside i...
102587 • Check-in time is 2PM. Check-out time is 11am...
                                                                NaN
102591
                                                    #NAMF?
                                                                NaN
102593 * No smoking indoors. * No pets * No loud/la...
                                                                NaN
102597
                                                                NaN
[34106 rows x 26 columns]
  grp1= df['price'].groupby(df['neighbourhood_group']).sum().sort_values(ascending = False)
```

```
In [61]: ## List the prices by neighborhood group and also mention which is the most expensive neighborhood group for rentals
         # in descending order prices will display
```

neighbourhood_group \$966 \$368 \$71 \$1,060 \$778 \$656 \$1,057 \$777 \$1,000 \$718 \$916 \$606 \$355 \$724 \$1,084 \$938 \$405 \$907 \$1,1... \$966 \$368 \$71 \$1,060 \$778 \$656 \$1,095 \$477 \$13... Brooklyn Queens brookln Bronx \$483 \$538 \$643 \$55 \$811 \$101 \$73 \$581 \$361 \$64... manhatan \$460 \$142 \$620 \$204 \$577 \$1,018 \$291 \$319 \$606 \$714... Manhattan \$142 \$620 \$294 \$577 \$1,918 \$291 \$319 \$696 \$714... Staten Island \$1,915 \$725 \$784 \$593 \$1,932 \$629 \$675 \$459 \$1... Name: price, dtype: object

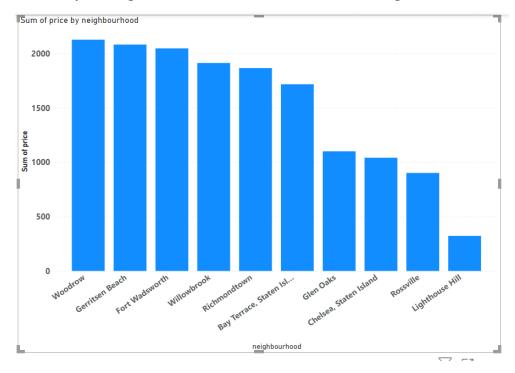
DATA VISUALIZATION IN POWER BI

Task 5a:- Horizontal bar chart to display 10 most expensive neighbourhoods in the dataset



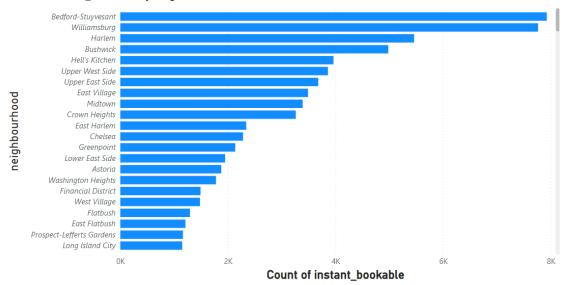
Expensive neighbourhoods for rentals are: Bedford-Stuyvesant > Williamsburg > Harlem

- 10 cheapest neighbourhoods in the dataset in decreasing order



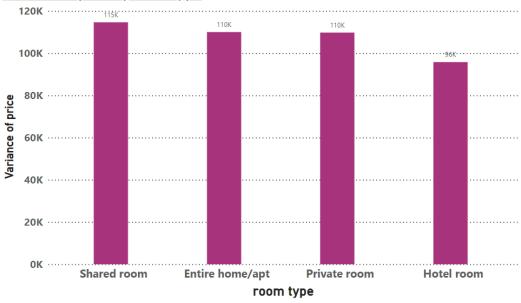
-List the neighborhoods which offer short term rentals within 10 days. Illustrate with a bar graph





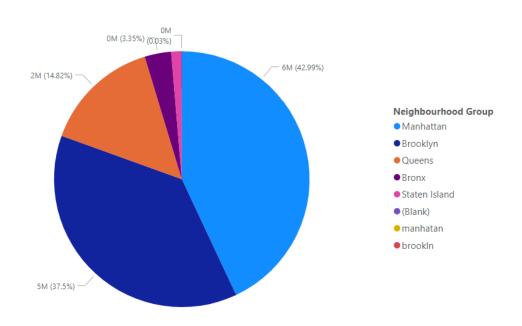
-List the prices with respect to room type using a bar graph and also state your inferences.

Variance of price by room type



-Create a pie chart that shows distribution of booked days for each neighborhood group .Which neighborhood has the highest booking percentage.

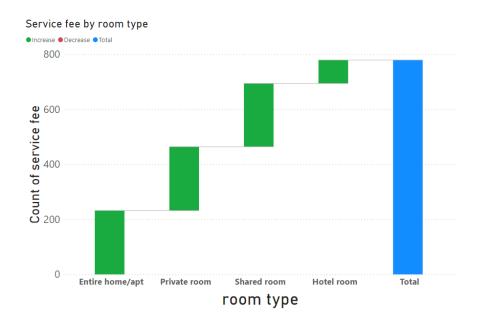
Distribution of booked days for each neighbourhood group



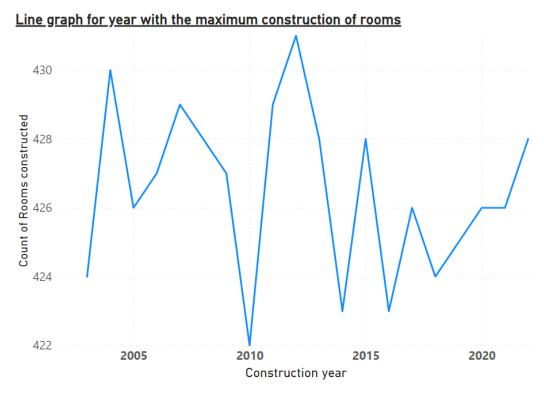
Manhattan has the highest booking percentage.

Task 5b: Data Visualization (Any Tool)

• Does service price and room price have an impact on each other. Illustrate this relationship with a scatter plot and state your inferences



-Create Using a line graph to show in which year the maximum construction of rooms took place.



Task 5c: Due to trial version Box plot in Power BI is not available.

With the help of box plots illustrate the following Effect of Review Rate number on price Count of neighbourhood by reviews per month

