Airbnb_Project

December 7, 2023

[2]:
```
'''
Project Title: Airbnb - IBM Skills
Author: Pilar Wilches
Date: '2023-12-04'
Source data: https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata/data
Goal: Practice data analysis by performing the process of cleaning,␣
 ↪transforming, and visualizing data.

Título del proyecto: Airbnb - IBM Skills
Autora: Pilar Wilches
Fecha: '2023-12-04'
Fuente de Datos: https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata/
 ↪data
Objetivo: Practicar el análisis de datos realizando proceso de limpieza,␣
 ↪transformación y visualización de datos.
'''
```

[2]:
```
#First of all, I'm going to list the necessary libraries for the entire process␣
 ↪of loading and cleaning data
#Primero que todo voy a listar las librerias necesarias para todo el proceso de␣
 ↪cargue y limpieza de datos

# Importing pandas for data manipulation and analysis
# Importo pandas para manipulación de datos y analisis
import pandas as pd

# Importing numpy for numerical operations
# importo numpy para cálculos numéricos
import numpy as np
```

[81]:
```
## Task 1. Load Data / Cargue de datos
df = pd.read_csv('/Users/maria/Python/Proyecto final/Airbnb_Open_Data.csv')
```

[82]:
```
# Task 1.1 Verify data upload / Verificar la carga de datos
df.head()
```

[82]:
```
        id                                              NAME      host id  \
0  1001254                Clean & quiet apt home by the park  80014485718
1  1002102                           Skylit Midtown Castle  52335172823
2  1002403              THE VILLAGE OF HARLEM…NEW YORK !  78829239556
3  1002755                                           NaN  85098326012
4  1003689  Entire Apt: Spacious Studio/Loft by central park  92037596077

  host_identity_verified host name neighbourhood group neighbourhood  \
0            unconfirmed  Madaline            Brooklyn    Kensington
1               verified     Jenna           Manhattan       Midtown
2                    NaN     Elise           Manhattan        Harlem
3            unconfirmed     Garry            Brooklyn  Clinton Hill
4               verified    Lyndon           Manhattan   East Harlem

        lat      long        country  … service fee minimum nights  \
0  40.64749 -73.97237  United States  …       $193            10.0
1  40.75362 -73.98377  United States  …        $28            30.0
2  40.80902 -73.94190  United States  …       $124             3.0
3  40.68514 -73.95976  United States  …        $74            30.0
4  40.79851 -73.94399  United States  …        $41            10.0

  number of reviews last review  reviews per month review rate number  \
0               9.0  10/19/2021                0.21                4.0
1              45.0   5/21/2022                0.38                4.0
2               0.0         NaN                 NaN                5.0
3             270.0    7/5/2019                4.64                4.0
4               9.0  11/19/2018                0.10                3.0

  calculated host listings count  availability 365  \
0                            6.0             286.0
1                            2.0             228.0
2                            1.0             352.0
3                            1.0             322.0
4                            1.0             289.0

                                      house_rules license
0  Clean up and treat the home the way you'd like…     NaN
1  Pet friendly but please confirm with me if the…     NaN
2  I encourage you to use my kitchen, cooking and…     NaN
3                                              NaN     NaN
4  Please no smoking in the house, porch or on th…     NaN

[5 rows x 26 columns]
```

[83]:
```
df.tail()
```

[83]:

```
             id                              NAME      host id  \
102594  6092437       Spare room in Williamsburg  12312296767
102595  6092990     Best Location near Columbia U  77864383453
102596  6093542    Comfy, bright room in Brooklyn  69050334417
102597  6094094  Big Studio-One Stop from Midtown  11160591270
102598  6094647              585 sf Luxury Studio  68170633372

       host_identity_verified    host name neighbourhood group  \
102594               verified         Krik            Brooklyn
102595             unconfirmed        Mifan           Manhattan
102596             unconfirmed        Megan            Brooklyn
102597             unconfirmed  Christopher              Queens
102598             unconfirmed      Rebecca           Manhattan

              neighbourhood       lat      long        country  …  \
102594         Williamsburg  40.70862 -73.94651  United States  …
102595  Morningside Heights  40.80460 -73.96545  United States  …
102596           Park Slope  40.67505 -73.98045  United States  …
102597    Long Island City  40.74989 -73.93777  United States  …
102598     Upper West Side  40.76807 -73.98342  United States  …

       service fee minimum nights number of reviews last review  \
102594      $169             1.0               0.0         NaN
102595      $167             1.0               1.0    7/6/2015
102596      $198             3.0               0.0         NaN
102597      $109             2.0               5.0  10/11/2015
102598      $206             1.0               0.0         NaN

        reviews per month review rate number calculated host listings count  \
102594                NaN                3.0                             1.0
102595               0.02                2.0                             2.0
102596                NaN                5.0                             1.0
102597               0.10                3.0                             1.0
102598                NaN                3.0                             1.0

        availability 365                                        house_rules  \
102594             227.0  No Smoking No Parties or Events of any kind Pl…
102595             395.0  House rules: Guests agree to the following ter…
102596             342.0                                                NaN
102597             386.0                                                NaN
102598              69.0                                                NaN

       license
102594     NaN
102595     NaN
102596     NaN
102597     NaN
```

```
102598     NaN

[5 rows x 26 columns]
```

[ ]:
```
# At this point, I see different fields with the value NaN
# En este punto veo diferentes campos con el valor NaN
```

[84]:
```
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
 1   NAME                            102349 non-null  object
 2   host id                         102599 non-null  int64
 3   host_identity_verified          102310 non-null  object
 4   host name                       102193 non-null  object
 5   neighbourhood group             102570 non-null  object
 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
 12  cancellation_policy             102523 non-null  object
 13  room type                       102599 non-null  object
 14  Construction year               102385 non-null  float64
 15  price                           102352 non-null  object
 16  service fee                     102326 non-null  object
 17  minimum nights                  102190 non-null  float64
 18  number of reviews               102416 non-null  float64
 19  last review                     86706 non-null   object
 20  reviews per month               86720 non-null   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
```

[85]:
```
print(df.columns)
```

```
Index(['id', 'NAME', 'host id', 'host_identity_verified', 'host name',
       'neighbourhood group', 'neighbourhood', 'lat', 'long', 'country',
       'country code', 'instant_bookable', 'cancellation_policy', 'room type',
```

```
       'Construction year', 'price', 'service fee', 'minimum nights',
       'number of reviews', 'last review', 'reviews per month',
       'review rate number', 'calculated host listings count',
       'availability 365', 'house_rules', 'license'],
      dtype='object')
```

[86]:
```
## 2. Data Cleaning / Limpieza de datos
# 2.1 Remove unwanted columns from the dataframe, including host id, id,␣
 ↪country, and country code.
# 2.2 Specify the reason for removing these columns for your data analysis.

count_values_host_id = df['host id'].nunique()
count_values_id = df['id'].nunique()
count_values_country = df['country'].nunique()
count_values_country_code = df['country code'].nunique()

print("Number host id':", count_values_host_id)
print("Number id':", count_values_id)
print("Number country':", count_values_country)
print("Number country code':", count_values_country_code)
```

```
Number host id': 102057
Number id': 102058
Number country': 1
Number country code': 1
```

[87]:
```
# The previous results show that the IDs are unique for each record; each one␣
 ↪has a different ID,
# which is why it is not useful for the analysis.
# On the other hand, the 'country' field has a unique value, as well as the␣
 ↪'country code'.

# Los anteriores resultados presentan que los id son únicos para cada registro,␣
 ↪cada uno tiene un id diferente,
# razón por la cual no es útil para el análisis. De otra parte el campo country␣
 ↪tiene un único valor al igual
# que el country code.
```

[88]:
```
df.drop(columns = ['host id', 'id', 'country', 'country code'], axis = 1,␣
 ↪inplace = True)
```

[89]:
```
df.head()
```

[89]:

	NAME	host_identity_verified
0	Clean & quiet apt home by the park	unconfirmed
1	Skylit Midtown Castle	verified
2	THE VILLAGE OF HARLEM…NEW YORK !	NaN
3	NaN	unconfirmed

```
4  Entire Apt: Spacious Studio/Loft by central park               verified

  host name neighbourhood group neighbourhood       lat      long  \
0  Madaline            Brooklyn    Kensington  40.64749 -73.97237
1     Jenna           Manhattan       Midtown  40.75362 -73.98377
2     Elise           Manhattan        Harlem  40.80902 -73.94190
3     Garry            Brooklyn  Clinton Hill  40.68514 -73.95976
4    Lyndon           Manhattan   East Harlem  40.79851 -73.94399

  instant_bookable cancellation_policy        room type  …  service fee  \
0            False              strict     Private room  …         $193
1            False            moderate  Entire home/apt  …          $28
2             True            flexible     Private room  …         $124
3             True            moderate  Entire home/apt  …          $74
4            False            moderate  Entire home/apt  …          $41

  minimum nights number of reviews  last review  reviews per month  \
0           10.0               9.0   10/19/2021               0.21
1           30.0              45.0    5/21/2022               0.38
2            3.0               0.0          NaN                NaN
3           30.0             270.0    7/5/2019               4.64
4           10.0               9.0   11/19/2018               0.10

  review rate number  calculated host listings count  availability 365  \
0                4.0                             6.0             286.0
1                4.0                             2.0             228.0
2                5.0                             1.0             352.0
3                4.0                             1.0             322.0
4                3.0                             1.0             289.0

                                        house_rules  license
0  Clean up and treat the home the way you'd like…      NaN
1  Pet friendly but please confirm with me if the…      NaN
2  I encourage you to use my kitchen, cooking and…      NaN
3                                              NaN      NaN
4  Please no smoking in the house, porch or on th…      NaN

[5 rows x 22 columns]
```

[90]:
```
# In the new dataframe visualization, 22 columns are displayed. There were␣
 ↪originally 26.
# En la nueva visualización del dataframe aparecen 22 columnas. Eran 26␣
 ↪originalmente.
```

[91]:
```
## Task 2b: Data Cleaning (Python)
# Check for null values and display the count in ascending order.
# If missing values are present, impute the values as you see fit.
```

```
## Tarea 2b: Limpieza de datos (Python)
# Comprueba si hay valores nulos y muestra el recuento en orden ascendente.
# **Si faltan valores, imputa los valores como consideres.**
```

[92]:
```
df.isnull().head()
```

[92]:
```
    NAME  host_identity_verified  host name  neighbourhood group  \
0  False                   False      False                False
1  False                   False      False                False
2  False                    True      False                False
3   True                   False      False                False
4  False                   False      False                False

   neighbourhood    lat   long  instant_bookable  cancellation_policy  \
0          False  False  False             False                False
1          False  False  False             False                False
2          False  False  False             False                False
3          False  False  False             False                False
4          False  False  False             False                False

   room type  …  service fee  minimum nights  number of reviews  \
0      False  …        False           False              False
1      False  …        False           False              False
2      False  …        False           False              False
3      False  …        False           False              False
4      False  …        False           False              False

   last review  reviews per month  review rate number  \
0        False              False               False
1        False              False               False
2         True               True               False
3        False              False               False
4        False              False               False

   calculated host listings count  availability 365  house_rules  license
0                           False             False        False     True
1                           False             False        False     True
2                           False             False        False     True
3                           False             False         True     True
4                           False             False        False     True

[5 rows x 22 columns]
```

[93]:
```
null_values = df.isnull().sum().sort_values(ascending = True)
print(null_values)
```

```
room type                         0
```

```
lat                                8
long                               8
neighbourhood                     16
neighbourhood group               29
cancellation_policy               76
instant_bookable                 105
number of reviews                183
Construction year                214
price                            247
NAME                             250
service fee                      273
host_identity_verified           289
calculated host listings count   319
review rate number               326
host name                        406
minimum nights                   409
availability 365                 448
reviews per month              15879
last review                    15893
house_rules                    52131
license                       102597
dtype: int64
```

[94]:
```
# I am checking the 'lat' and 'long' fields; there are 8 null records that can
 ↪be reviewed at a glance.
# Inicio revisando los campos lat y long, hay 8 registros nulos que se pueden
 ↪revisar a vista.
print(df[df['lat'].isnull()])
```

```
                                  NAME host_identity_verified host name  \
779  Large, furnished room in a 2 bedroom!            unconfirmed    Gibson
785      Authentic NY Charming Artist Loft            unconfirmed    Bailey
799         Huge room with private balcony               verified      Hunt
814    Decorators 5-Star Flat West Village               verified    Watson
843     Nice Private Room Beauty in Queens               verified   Roberts
885            Cute Room in Historic Loft!            unconfirmed     Jones
926        21 day Chelsea Apartment rental            unconfirmed     Owens
986         New York City for All Seasons!            unconfirmed   Douglas

    neighbourhood group      neighbourhood  lat  long instant_bookable  \
779            Brooklyn      Crown Heights  NaN   NaN            False
785            Brooklyn         Greenpoint  NaN   NaN            False
799           Manhattan       East Village  NaN   NaN            False
814           Manhattan       West Village  NaN   NaN             True
843              Queens           Elmhurst  NaN   NaN             True
885            Brooklyn         Greenpoint  NaN   NaN             True
926           Manhattan  Flatiron District  NaN   NaN            False
986           Manhattan    Upper West Side  NaN   NaN             True
```

```
    cancellation_policy        room type  …  service fee minimum nights  \
779              strict     Private room  …        $108             1.0
785              strict  Entire home/apt  …        $212             5.0
799            flexible     Private room  …        $101             6.0
814              strict  Entire home/apt  …         $76            20.0
843              strict     Private room  …         $45             1.0
885            flexible     Private room  …        $105            14.0
926              strict     Private room  …        $125            21.0
986            flexible     Private room  …         $83             1.0

    number of reviews  last review  reviews per month review rate number  \
779                1.0    3/18/2017                0.04                2.0
785               14.0    6/19/2019                0.16                5.0
799                1.0     5/6/2013                0.01                1.0
814              157.0    8/11/2016                1.71                4.0
843               63.0    5/18/2019                0.89                3.0
885               22.0     5/2/2019                0.25                1.0
926                0.0          NaN                 NaN                2.0
986               25.0    6/22/2013                0.28                2.0

     calculated host listings count  availability 365  \
779                             1.0              41.0
785                             1.0             226.0
799                             1.0             240.0
814                             1.0              61.0
843                             2.0              70.0
885                             1.0             266.0
926                             1.0             104.0
986                             1.0             259.0

                                        house_rules  license
779  - Weekly and monthly prices are much lower - P…      NaN
785  We live and let live - hoping that you'd be re…      NaN
799  Expect respect for the family and the space--t…      NaN
814  Please keep it clean, thats all we really ask …      NaN
843                                              NaN      NaN
885  Pets are cool (just clean up after them!), smo…      NaN
926                                              NaN      NaN
986                                No Smoking No Pets      NaN

[8 rows x 22 columns]
```

[95]:
```
search_df = df[(df['neighbourhood group'] == 'Brooklyn') & (df['neighbourhood']␣
  ↪== 'Crown Heights')][['lat', 'long']]
print(search_df)
```

```
            lat      long
```

```
19       40.67592 -73.94694
269      40.67780 -73.94339
271      40.67610 -73.95290
272      40.67586 -73.95155
284      40.67150 -73.94808
…           …        …
102435   40.67182 -73.94446
102527   40.67627 -73.94775
102533   40.67259 -73.95831
102582   40.66743 -73.94712
102590   40.66673 -73.96127

[3262 rows x 2 columns]
```

[96]:
```
# The values are very close, but not identical, so the median value will be␣
 ↪imputed.
# This criterion will be applied to all numeric columns.
# For categorical variables, the value of 0 will be assigned.

# Los valores son muy cercanos, pero no identicos, por lo tanto se imputara el␣
 ↪valor de la mediana.
# Este criterio se aplicará para todas las columnas numéricas.
# Para las variables categóricas se colocará el valor de 0
```

[97]:
```
for col in df.columns:
    # Si la columna es de tipo object ('O'), se asume que es categórica o de␣
 ↪texto
    if df[col].dtype == 'O':
        # categorical columns
        print("categorical columns:", col)
        # Rellenar valores nulos con la moda (el valor más frecuente)
        df[col].fillna(value=df[col].mode()[0], inplace=True)
    else:
        # numeric columns
        print("categorical numeric:", col)
        df[col].fillna(value=df[col].median(), inplace=True)
```

```
categorical columns: NAME
categorical columns: host_identity_verified
categorical columns: host name
categorical columns: neighbourhood group
categorical columns: neighbourhood
categorical numeric: lat
categorical numeric: long
categorical columns: instant_bookable
categorical columns: cancellation_policy
categorical columns: room type
categorical numeric: Construction year
```

```
categorical columns: price
categorical columns: service fee
categorical numeric: minimum nights
categorical numeric: number of reviews
categorical columns: last review
categorical numeric: reviews per month
categorical numeric: review rate number
categorical numeric: calculated host listings count
categorical numeric: availability 365
categorical columns: house_rules
categorical columns: license
```

[98]:
```
## Task 2b: Data Cleaning (Python)
# Check for duplicate values and remove them.
# Display the total number of records before and after removing duplicates.

## Tarea 2b: Limpieza de datos (Python)
# Comprueba si hay valores duplicados y elimínalos.
# Muestra el número total de registros antes y después de eliminar los␣
 ↪duplicados.
```

[99]:
```
df.duplicated().sum() # Before removing duplicates / Antes de eliminar␣
 ↪duplicados
```

[99]:
```
3461
```

[100]:
```
df.drop_duplicates(inplace = True)
```

[101]:
```
df.duplicated().sum() # After removing duplicates / Después de eliminar␣
 ↪duplicados
```

[101]:
```
0
```

[102]:
```
# Task 3: Data Transformation (any tool)
# Change the name of the column availability 365 to days_booked.
# Convert all column names to lowercase and replace spaces in column names with␣
 ↪an underscore "_".
# Remove the dollar sign and comma from the price and service_fee columns. If␣
 ↪necessary, convert
# these two columns to the appropriate data type.

#Tarea 3: Transformación de datos (cualquier herramienta)
# Cambia el nombre de la columna `availability 365` a `days_booked`.
# Convierte todos los nombres de columna a minúsculas y sustituye los espacios␣
 ↪en los nombres de columna por un guión bajo "_".
# Elimina el signo de dólares y la coma de las columnas `price` y `service_fee`.
 ↪ Si es necesario,
# convierte estas dos columnas al tipo de datos adecuado.
```

[103]:
```
df.rename(columns={"availability 365": "days_booked"}, inplace = True)
```

[104]:
```
df.columns = [col.lower().replace(" ", "_") for col in df.columns]
```

[105]:
```
df.columns
```

[105]:
```
Index(['name', 'host_identity_verified', 'host_name', 'neighbourhood_group',
       'neighbourhood', 'lat', 'long', 'instant_bookable',
       'cancellation_policy', 'room_type', 'construction_year', 'price',
       'service_fee', 'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'review_rate_number',
       'calculated_host_listings_count', 'days_booked', 'house_rules',
       'license'],
      dtype='object')
```

[106]:
```
df["price"].head()
```

[106]:
```
0    $966
1    $142
2    $620
3    $368
4    $204
Name: price, dtype: object
```

[107]:
```
df['price'] = df['price'].replace('[\$,]', '', regex=True).astype(float)
```

[108]:
```
df["price"].head()
```

[108]:
```
0    966.0
1    142.0
2    620.0
3    368.0
4    204.0
Name: price, dtype: float64
```

[109]:
```
df["service_fee"].head()
```

[109]:
```
0    $193
1     $28
2    $124
3     $74
4     $41
Name: service_fee, dtype: object
```

[110]:
```
df['service_fee'] = df['service_fee'].replace('[\$,]', '', regex=True).
 ↪astype(float)
```

[111]:
```
df["service_fee"].head()
```

[111]:
```
0    193.0
1     28.0
2    124.0
3     74.0
4     41.0
Name: service_fee, dtype: float64
```

[112]:
```
# Utilicé una función regular para cambiar comas y signo dolar en los campos␣
 ↪service_fee y price.
# Además cambié el tipo de dato a float para poder realizar operaciones
```

[113]:
```
# Task 4: Exploratory Data Analysis (any tool)
# List the types of available rooms in the dataset
# Which type of room has the strictest cancellation policy?
# List the average price per neighborhood and identify the set of neighborhoods␣
 ↪that is the most expensive to rent.

# Tarea 4: Análisis exploratorio de datos (cualquier herramienta)
# Enumera los tipos de habitaciones disponibles en el dataset
# ¿Qué tipo de habitación tiene la política de cancelación más estricta?
# Enumera el precio medio por barrio y señala cuál es el conjunto de barrios␣
 ↪más caro para alquilar.
```

[114]:
```
df.columns
```

[114]:
```
Index(['name', 'host_identity_verified', 'host_name', 'neighbourhood_group',
       'neighbourhood', 'lat', 'long', 'instant_bookable',
       'cancellation_policy', 'room_type', 'construction_year', 'price',
       'service_fee', 'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'review_rate_number',
       'calculated_host_listings_count', 'days_booked', 'house_rules',
       'license'],
      dtype='object')
```

[115]:
```
df['room_type'].value_counts() #types of available rooms
```

[115]:
```
room_type
Entire home/apt    51987
Private room       44887
Shared room         2149
Hotel room           115
Name: count, dtype: int64
```

[116]:
```
df['cancellation_policy'].value_counts() #types of cancellation
```

[116]:
```
cancellation_policy
moderate    33264
```

```
flexible    32948
strict      32926
Name: count, dtype: int64
```

[117]:
```
strict_room = df.loc[df['cancellation_policy']=='strict', 'room_type'].
 ↪value_counts().idxmax()
print(strict_room) # The entire home/apt has a cancellation policy more strict
```

```
Entire home/apt
```

[118]:
```
average_price = df.groupby('neighbourhood_group')['price'].mean().
 ↪sort_values(ascending = False)
print(average_price)
```

```
neighbourhood_group
Queens            628.668822
Brooklyn          625.471627
Bronx             625.271511
Staten Island     625.060870
Manhattan         621.666140
brookln           580.000000
manhatan          460.000000
Name: price, dtype: float64
```

[119]:
```
neighbourhood_groups_ny = df['neighbourhood_group'].unique()
print(neighbourhood_groups_ny)
```

```
['Brooklyn' 'Manhattan' 'brookln' 'manhatan' 'Queens' 'Staten Island'
 'Bronx']
```

[120]:
```
# At this point, I notice a typographical error in 'brookln' and 'manhatan', so␣
 ↪I proceed to update these values.
# En este punto noto que hay un error de tipografia en 'brookln' y 'manhatan,␣
 ↪procedo a actualizar estos valores
```

[121]:
```
data = df.copy()
data['neighbourhood_group'] = data['neighbourhood_group'].replace('brookln',␣
 ↪'Brooklyn')
```

[125]:
```
print(data['neighbourhood_group'].unique())
```

```
['Brooklyn' 'Manhattan' 'manhatan' 'Queens' 'Staten Island' 'Bronx']
```

[126]:
```
data['neighbourhood_group'] = data['neighbourhood_group'].replace('manhatan',␣
 ↪'Manhattan')
```

[127]:
```
print(data['neighbourhood_group'].unique())
```

```
['Brooklyn' 'Manhattan' 'Queens' 'Staten Island' 'Bronx']
```

[ ]:
```
#Here, I don't know what happened, and when I performed the replacement, it␣
 ↪updated everything to `None`.
# So, luckily, you can quickly rerun all the Python code to be cautious. At␣
 ↪this point,
# I made a copy of the DataFrame.

# Aquí no se que sucedio y cuando realicé el reemplazo me actualizo todo a␣
 ↪None, así que por suerte se puede volver
# a ejecutar todo el código python de manera rápida, para tomar mis␣
 ↪precauciones, realicé en este punto una
# copia del dataframe
```

[128]:
```
data.head()
```

[128]:

	name	host_identity_verified	host_name	neighbourhood_group	neighbourhood	lat	long	instant_bookable	cancellation_policy	room_type	…	service_fee	minimum_nights	number_of_reviews	last_review	reviews_per_month	review_rate_number	calculated_host_listings_count	days_booked
0	Clean & quiet apt home by the park	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	False	strict	Private room	…	193.0	10.0	9.0	10/19/2021	0.21	4.0	6.0	286.0
1	Skylit Midtown Castle	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	False	moderate	Entire home/apt	…	28.0	30.0	45.0	5/21/2022	0.38	4.0	2.0	228.0
2	THE VILLAGE OF HARLEM…NEW YORK !	unconfirmed	Elise	Manhattan	Harlem	40.80902	-73.94190	True	flexible	Private room	…	124.0	3.0	0.0	6/23/2019	0.74	5.0	1.0	352.0
3	Home away from home	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	True	moderate	Entire home/apt	…	74.0	30.0	270.0	7/5/2019	4.64	4.0	1.0	322.0
4	Entire Apt: Spacious Studio/Loft by central park	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	False	moderate	Entire home/apt	…	41.0	10.0	9.0	11/19/2018	0.10			

```
4                3.0                                1.0        289.0

                                                house_rules   license
0  Clean up and treat the home the way you'd like…  41662/AL
1  Pet friendly but please confirm with me if the…  41662/AL
2  I encourage you to use my kitchen, cooking and…  41662/AL
3                                                #NAME?  41662/AL
4  Please no smoking in the house, porch or on th…  41662/AL

[5 rows x 22 columns]
```

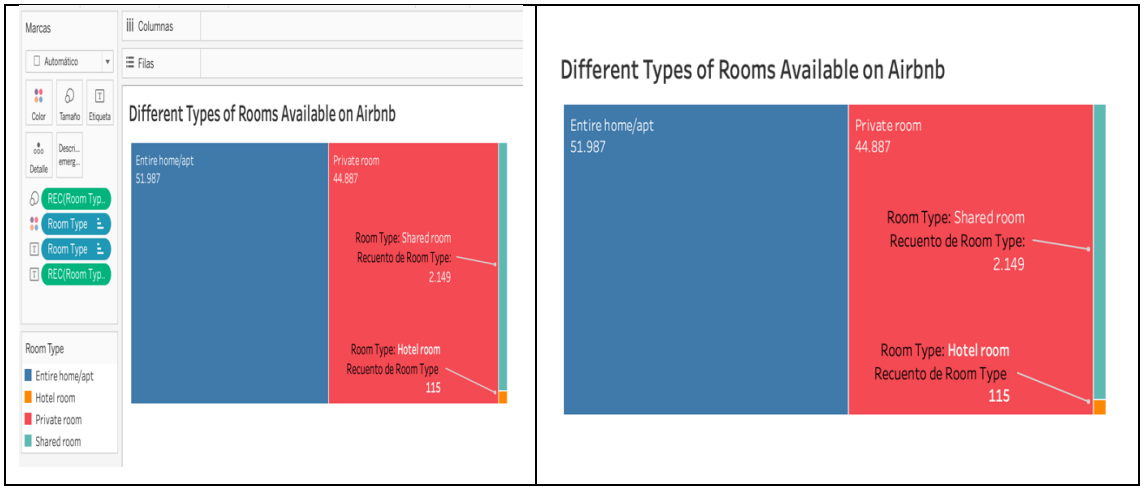
[130]:
```
grpl= data['price'].groupby(data['neighbourhood_group']).sum().
↪sort_values(ascending=False)
grpl # List the average price per neighborhood, Manhattan is the most expensive␣
↪to rent.
```

[130]:
```
neighbourhood_group
Manhattan        26354131.0
Brooklyn         25252746.0
Queens            8069593.0
Bronx             1635085.0
Staten Island      575056.0
Name: price, dtype: float64
```

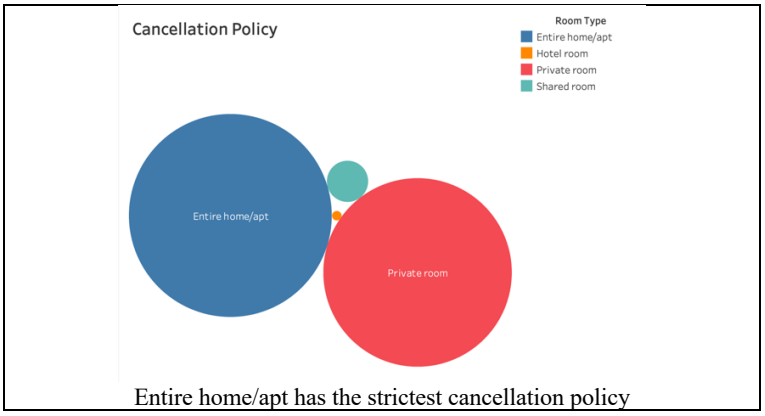
[131]:
```
data.to_csv('airbnb.csv') # I export to CSV because I will start visualizing in␣
↪Tableau.
```

Data Visualization in Tableau / Visualización de Datos en Tableau

5a. List the different types of rooms available on Airbnb.
* Enumerar los distintos tipos de habitaciones disponibles en Airnb



What type of room adheres to a stricter cancellation policy?
* Qué tipo de habitación se adhiere a una política de cancelación más estricta.



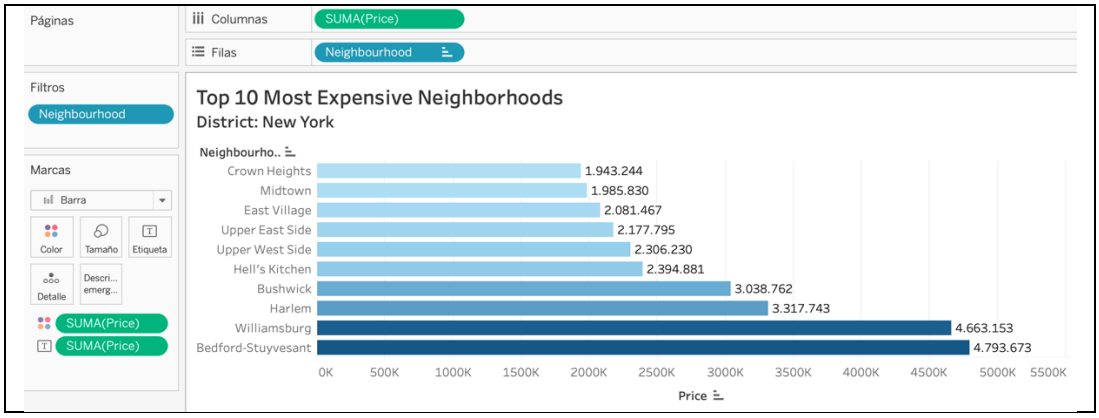
Entire home/apt has the strictest cancellation policy

List the prices by neighborhood group and also mention which neighborhood group is the most expensive for rentals.
* Enumere los precios por grupo de barrios y mencione también cuál es el grupo de barrios más caro para los alquileres.
datos.

List the average price per neighborhood group, Manhattan is the most expensive to rent

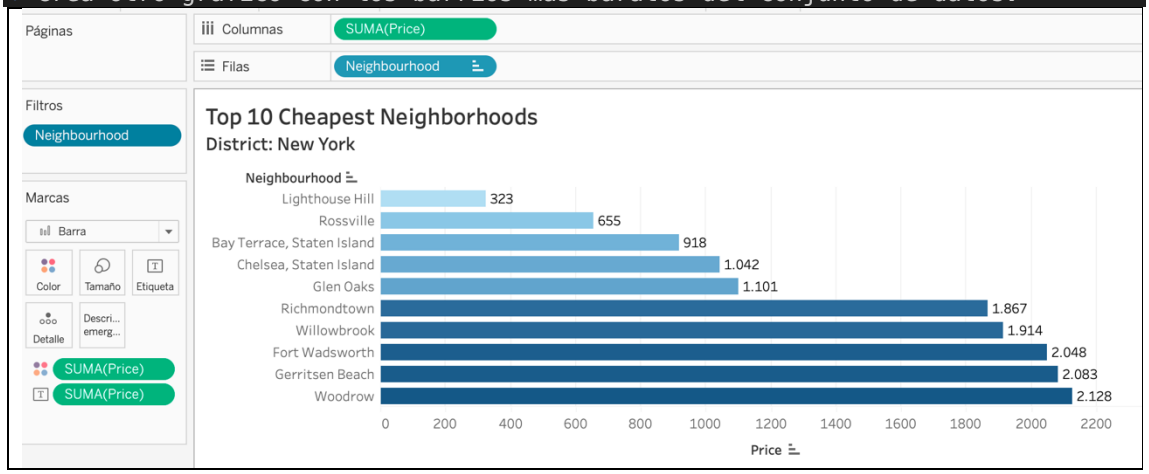
List the top 10 most expensive neighborhoods in ascending order of price using a horizontal bar chart.

* Enumere los 10 barrios más caros por orden creciente de precio con la ayuda de un gráfico de barras horizontales.

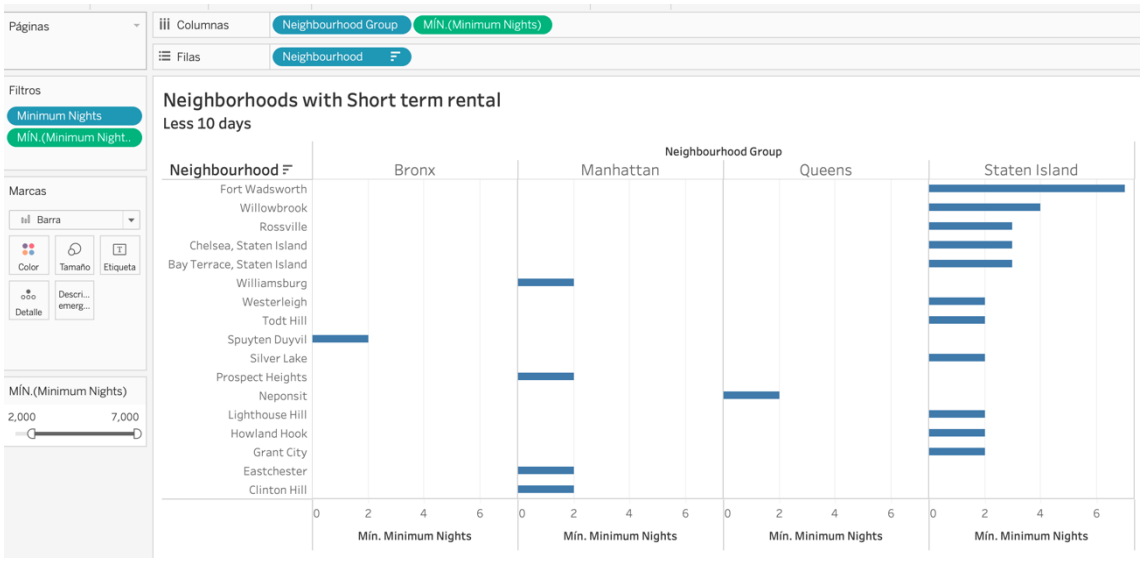


Create another chart with the cheapest neighborhoods in the dataset.

* Crea otro gráfico con los barrios más baratos del conjunto de datos.

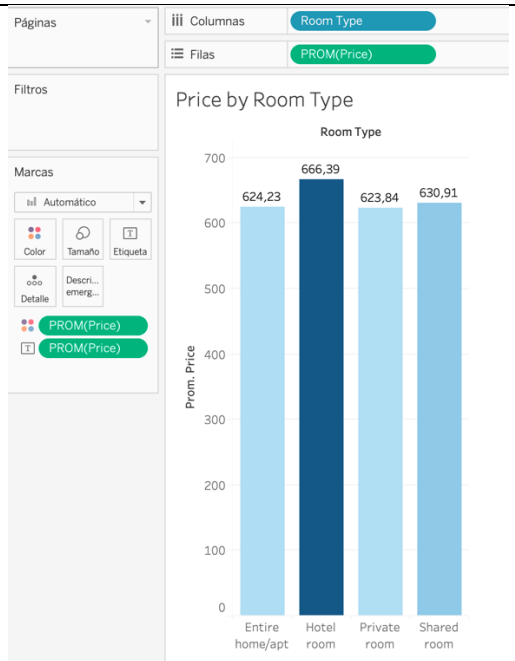


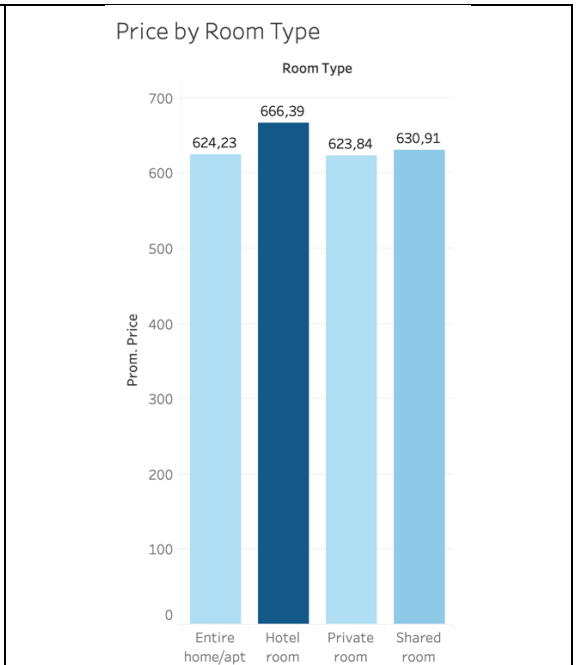
List the neighborhoods that offer short-term rentals of less than 10 days. Illustrate with a bar chart.
* Enumere los barrios que ofrecen alquileres a corto plazo de menos de 10 días. Ilustrar con un gráfico de barras



Se generan 224 registros, por lo cual para la visualización apliqué un filtro, de 2 a 7 noches y me visualiza 17 registros, más fácil de leer el gráfico.

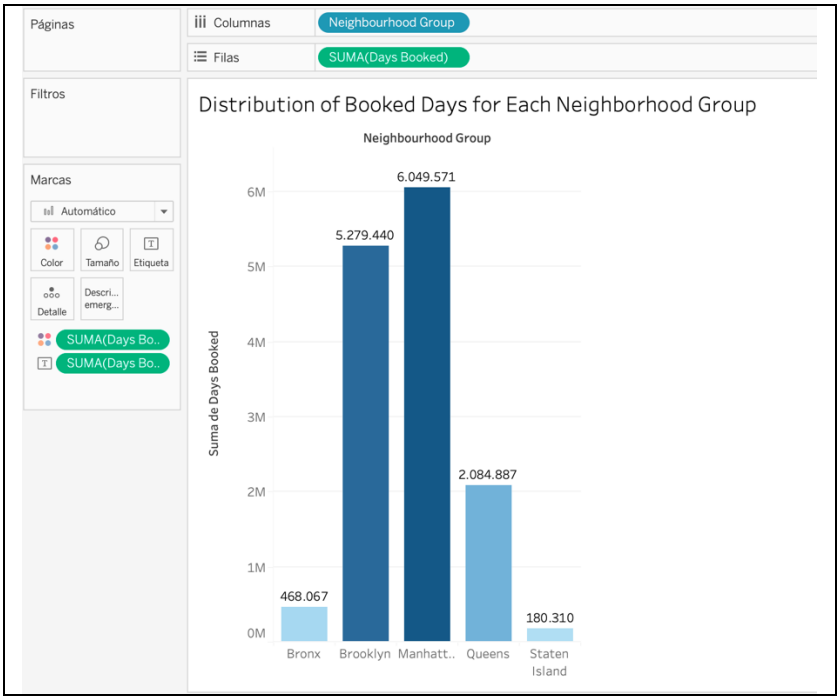
* List the prices by room type using a bar chart and also present your insights.
* Enumere los precios con respecto al tipo de habitación utilizando un gráfico de barras y exponga también sus inferencias.



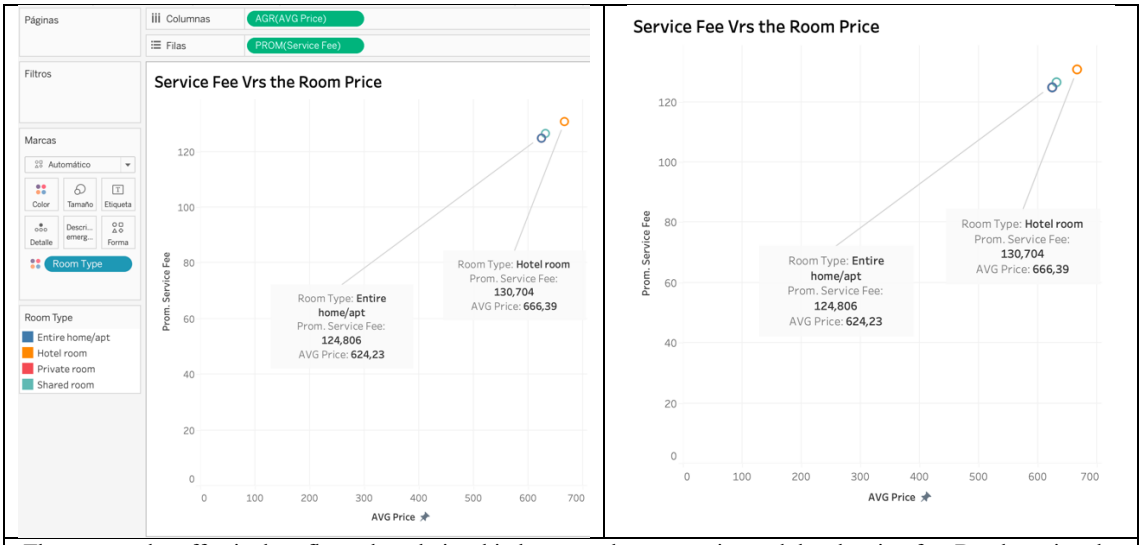


Se halló el promedio de precio por tipo de habitación y se encontró que la mejor opción es rentar un apartamento completo (puesto que tiene electrodomésticos incluidos por ejemplo) contrario a alquilar una habitación de hotel.

Create a pie chart that shows the distribution of booked days for each neighborhood group.
* Cree un gráfico circular que muestre la distribución de los días reservados para cada grupo de barrios.



Does the service fee impact the room price, and viceversa? Illustrate this relationship with a scatter plot and provide your insights.
* ¿El precio del servicio y el precio de la habitación tienen un impacto mutuo? Ilustre esta relación con un gráfico de dispersión e indique sus inferencias



The scatter plot effectively reflects the relationship between the room price and the cleaning fee. By observing the two extremes on the graph, a clear trend can be noted: as the room price increases, so does the cleaning fee. This indicates a positive correlation between both variables. Therefore, it can be inferred that there is a mutual relationship between them.

El gráfico de dispersión efectivamente refleja la relación entre el precio de la habitación y la tarifa de servicio. Al observar las dos marcas extremas del gráfico, se puede notar la tendencia: a medida que el precio de la habitación

aumenta, también lo hace la tarifa de limpieza. Se trata de una relación positiva entre ambas variables. Por lo tanto, se puede inferir que hay una relación mutua entre estas.