P1

June 21, 2020

1 Project 1 - Udacity DSND

For this project the following libraries have being imported

```
In [1]: #Importing libraries
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
```

I have selected the Seattle AirBNB Dataset (https://www.kaggle.com/airbnb/seattle/data?select=listings.csv

```
In [2]: # Gathering data
        df = pd.read_csv("listings.csv")
```

```
In [4]: # Overview dataset
        df.head()
```

```
Out[4]:         id                           listing_url        scrape_id last_scrap
        0   241032   https://www.airbnb.com/rooms/241032  20160104002432   2016-01-
        1   953595   https://www.airbnb.com/rooms/953595  20160104002432   2016-01-
        2  3308979  https://www.airbnb.com/rooms/3308979  20160104002432   2016-01-
        3  7421966  https://www.airbnb.com/rooms/7421966  20160104002432   2016-01-
        4   278830   https://www.airbnb.com/rooms/278830  20160104002432   2016-01-

                                       name  \
        0         Stylish Queen Anne Apartment
        1    Bright & Airy Queen Anne Apartment
        2  New Modern House-Amazing water view
        3                  Queen Anne Chateau
        4        Charming craftsman 3 bdm house

                                                summary  \
        0                                           NaN
        1  Chemically sensitive? We've removed the irrita...
        2  New modern house built in 2013.  Spectacular s...
        3  A charming apartment that sits atop Queen Anne...
        4  Cozy family craftman house in beautiful neighb...
```

1

```
                                                  space  \
0  Make your self at home in this charming one-be...
1  Beautiful, hypoallergenic apartment in an extr...
2  Our house is modern, light and fresh with a wa...
3                                                NaN
4  Cozy family craftman house in beautiful neighb...

                                         description experiences_offered  \
0  Make your self at home in this charming one-be...                none
1  Chemically sensitive? We've removed the irrita...                none
2  New modern house built in 2013.  Spectacular s...                none
3  A charming apartment that sits atop Queen Anne...                none
4  Cozy family craftman house in beautiful neighb...                none

                               neighborhood_overview  ... review_scores_val
0                                                NaN  ...                10
1  Queen Anne is a wonderful, truly functional vi...  ...                10
2  Upper Queen Anne is a charming neighborhood fu...  ...                10
3                                                NaN  ...                 N
4  We are in the beautiful neighborhood of Queen ...  ...                 9

  requires_license license jurisdiction_names instant_bookable  \
0                f     NaN         WASHINGTON                f
1                f     NaN         WASHINGTON                f
2                f     NaN         WASHINGTON                f
3                f     NaN         WASHINGTON                f
4                f     NaN         WASHINGTON                f

  cancellation_policy  require_guest_profile_picture  \
0            moderate                              f
1              strict                              t
2              strict                              f
3            flexible                              f
4              strict                              f

  require_guest_phone_verification calculated_host_listings_count  \
0                                f                              2
1                                t                              6
2                                f                              2
3                                f                              1
4                                f                              1

  reviews_per_month
0               4.07
1               1.48
2               1.15
3                NaN
4               0.89
```

```
        [5 rows x 92 columns]

In [5]: #List of Columns
        df.columns

Out[5]: Index(['id', 'listing_url', 'scrape_id', 'last_scraped', 'name', 'summary',
               'space', 'description', 'experiences_offered', 'neighborhood_overvie
               'notes', 'transit', 'thumbnail_url', 'medium_url', 'picture_url',
               'xl_picture_url', 'host_id', 'host_url', 'host_name', 'host_since',
               'host_location', 'host_about', 'host_response_time',
               'host_response_rate', 'host_acceptance_rate', 'host_is_superhost',
               'host_thumbnail_url', 'host_picture_url', 'host_neighbourhood',
               'host_listings_count', 'host_total_listings_count',
               'host_verifications', 'host_has_profile_pic', 'host_identity_verifie
               'street', 'neighbourhood', 'neighbourhood_cleansed',
               'neighbourhood_group_cleansed', 'city', 'state', 'zipcode', 'market'
               'smart_location', 'country_code', 'country', 'latitude', 'longitude'
               'is_location_exact', 'property_type', 'room_type', 'accommodates',
               'bathrooms', 'bedrooms', 'beds', 'bed_type', 'amenities', 'square_fe
               'price', 'weekly_price', 'monthly_price', 'security_deposit',
               'cleaning_fee', 'guests_included', 'extra_people', 'minimum_nights',
               'maximum_nights', 'calendar_updated', 'has_availability',
               'availability_30', 'availability_60', 'availability_90',
               'availability_365', 'calendar_last_scraped', 'number_of_reviews',
               'first_review', 'last_review', 'review_scores_rating',
               'review_scores_accuracy', 'review_scores_cleanliness',
               'review_scores_checkin', 'review_scores_communication',
               'review_scores_location', 'review_scores_value', 'requires_license',
               'license', 'jurisdiction_names', 'instant_bookable',
               'cancellation_policy', 'require_guest_profile_picture',
               'require_guest_phone_verification', 'calculated_host_listings_count'
               'reviews_per_month'],
              dtype='object')
```

Missing Data:

The dataset indeed presents missing data. I have decided to not proceed with data cleaning since the data take for this notebook contains all the data without missing data. Based on the following there are several missing data but this does not impact my questions so I have left these unchanged.

```
In [11]: df.count().sort_values(ascending=True)

Out[11]: license                           0
         square_feet                      97
         monthly_price                  1517
         security_deposit               1866
         weekly_price                   2009
         notes                          2212
```

```
neighborhood_overview            2786
cleaning_fee                     2788
transit                          2884
host_about                       2959
host_acceptance_rate             3045
review_scores_checkin            3160
review_scores_accuracy           3160
review_scores_value              3162
review_scores_location           3163
review_scores_cleanliness        3165
review_scores_communication      3167
review_scores_rating             3171
last_review                      3191
first_review                     3191
reviews_per_month                3191
space                            3249
host_response_time               3295
host_response_rate               3295
neighbourhood                    3402
xl_picture_url                   3498
thumbnail_url                    3498
medium_url                       3498
host_neighbourhood               3518
summary                          3641
                                 ...
id                               3818
price                            3818
listing_url                      3818
scrape_id                        3818
last_scraped                     3818
name                             3818
description                      3818
experiences_offered              3818
picture_url                      3818
host_id                          3818
host_url                         3818
host_verifications               3818
street                           3818
neighbourhood_cleansed           3818
guests_included                  3818
neighbourhood_group_cleansed     3818
state                            3818
market                           3818
smart_location                   3818
country_code                     3818
country                          3818
calculated_host_listings_count   3818
longitude                        3818
```

```
is_location_exact                3818
room_type                        3818
accommodates                     3818
bed_type                         3818
amenities                        3818
city                             3818
latitude                         3818
Length: 92, dtype: int64
```

1.0.1 Question 1

What is the proportion of room_type per property_type?

As we can see below from the plot, there are 3 categories of room_type: * Entire home/apt *
Private room * Shared room

On the other hand, there are 16 different property_type, however for this exercise I have
selected the top 5: * House * Apartment * Townhouse * Condominium * Loft

Following observation from the plot: * **House** (property_type) tend to have similar percentages
of **Entire home/apt** (room type) and **Private room**(room type) * **Apartment** (property_type)
has greater percentage of **Entire home/apt** (room type) compare to **Private room**(room type)
* **Townhouse** (property_type) has greater percentage of **Private room** (room type) compare to
Entire home/apt(room type) * **Condominium** (property_type) has greater percentage of **Entire
home/apt** (room type) compare to **Private room**(room type)

Regarding **Shared Room**, there is a low percentage so it is not useful to discuss this

```
In [12]: # Counting instances in order
         order = df['property_type'].value_counts(ascending=False).index[:5]
         #Plot
         sns.countplot(x="property_type", hue="room_type",  data=df, order=order)
         #Plot Show in notebook
         plt.show()
```
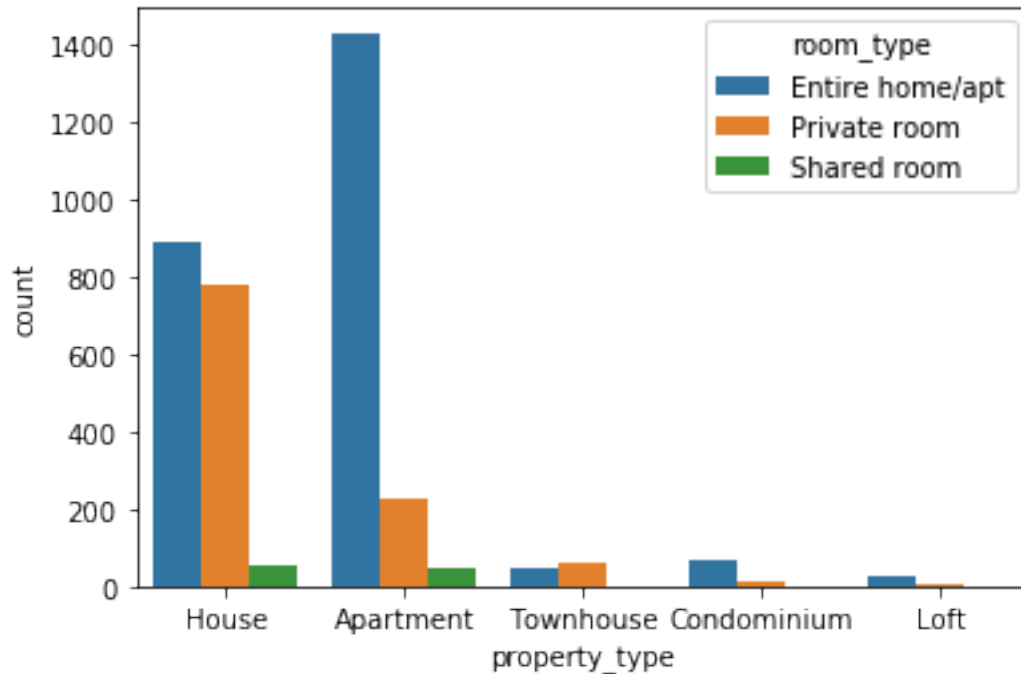
1.0.2 Question 2

Who is the host in the dataset which posses the highest number of properties?

Steps and answer: * I have found below the number of properties per host * Got the first and highest host possessing greater number of properties (8534462) * Filtered data on host_id = 8534462 * Identified and reported properties the host is placing on AirBnB * Her name is Daniela and owns 46 properties.

```
In [14]:  # Count of distinct host_id
          df['host_id'].value_counts()

Out[14]:  8534462     46
          4962900     39
          1243056     37
          430709      36
          3074414     34
          74305       33
          26967583    21
          7354306     18
          42537846    16
          1623580     12
          658155      12
          2911360     11
          862329      11
          754810      10
```

```
31148752      10
50893         10
14980831       9
5325329        9
4411144        9
1787819        8
10558142       7
5820099        7
1121532        7
6372203        6
22764472       6
5177328        6
50550045       6
12796487       6
2409388        5
32807985       5
              ..
45917623       1
41319913       1
17928378       1
21149225       1
42273230       1
23372237       1
45437387       1
9565783        1
34605512       1
295367         1
2161126        1
9060841        1
27761130       1
15512043       1
205325         1
36706828       1
7182858        1
27129222       1
13042181       1
1845764        1
13959773       1
15436286       1
2294267        1
33931765       1
36882932       1
38090170       1
11839982       1
668141         1
1691632        1
10420224       1
Name: host_id, Length: 2751, dtype: int64
```

```
In [16]: # Filter on highest host_id having greater number of properies compare to
         df.loc[df['host_id'] == 8534462][['host_id', 'name', 'host_name', 'street'

Out[16]:        host_id                                name  host_name  \
         1077  8534462          Pike's Place. 100 Walkscore    Daniela
         1149  8534462      Westlake Ave Downtown Condos 1BD    Daniela
         1152  8534462    Westlake Ave Downtown Condos 1BD 4    Daniela
         1155  8534462    Westlake Ave Downtown Condos 1BD 2    Daniela
         1159  8534462    Westlake Ave Downtown Condos 1BD 3    Daniela
         1230  8534462             100 Walk Score - Lux 2Bed    Daniela
         1237  8534462           Great Location. 98 Walkscore 2    Daniela
         1238  8534462                   Belltown Condos 1BD 4    Daniela
         1278  8534462     Downtown 2Bed Western Sunset Views    Daniela
         1298  8534462       Sky High Water View. 99 Walkscore    Daniela
         1312  8534462       Sky High Penthouse. 99 Walkscore    Daniela
         1323  8534462                   Belltown Condos 1BD 3    Daniela
         1340  8534462                   Belltown Condos 1BD 2    Daniela
         1366  8534462    Westlake Ave Downtown Condos 1BD 5    Daniela
         1376  8534462                     Belltown Condos 1BD    Daniela
         1377  8534462           Great Location. 98 Walkscore    Daniela
         1400  8534462      Cozy 1BD on Pike's Place Downtown    Daniela
         1404  8534462        99 1Bed, CapHill. Walk Downtown    Daniela
         1424  8534462                   Belltown Condos 1BD 5    Daniela
         1460  8534462           Pike's Place. 98 Walkscore 2    Daniela
         1466  8534462         Pike's Place 1Bed Sunset View    Daniela
         1469  8534462           Convention Center Lux 1BD    Daniela
         1478  8534462           Pike's Place. 98 Walkscore 3    Daniela
         1481  8534462         Convention Center Lux 2BD 2    Daniela
         1482  8534462         Convention Center Lux 1BD 2    Daniela
         1500  8534462             99 Lux Pike's Market 2BD    Daniela
         1519  8534462  99 Walkscore Sky-Hi. Convention Ctr    Daniela
         1532  8534462             Convention Center Lux 2BD    Daniela
         1537  8534462        99 Downtown Walkscore Lux 1BD    Daniela
         1548  8534462    Ultra-Modern 1Bed w/ Amazing Views    Daniela
         1557  8534462             Pike's Place. 98 Walkscore    Daniela
         1559  8534462           Pike's Place. 100 Walkscore 2    Daniela
         1575  8534462           8th Ave Condos Modern 1BD 3    Daniela
         1606  8534462       98 Walkscore. Swedish Suites 1BD    Daniela
         1613  8534462             8th Ave Condos Modern 1BD    Daniela
         1631  8534462               99 Amazing View Downtown    Daniela
         1637  8534462       98 Walkscore. Swedish Suites 2BD    Daniela
         1645  8534462             8th Ave Condos Modern 2BD    Daniela
         1652  8534462  98 Walkscore. Swedish Suites Studio    Daniela
         1654  8534462     98 Walkscore. Swedish Suites 1BD 2    Daniela
         1657  8534462           8th Ave Condos Modern 1BD 2    Daniela
         1660  8534462                         99 Modern 1Bed    Daniela
         1667  8534462         99 Open Layout in Cap Hill 2    Daniela
         1695  8534462    Ultra-Modern 2Bed w/ Amazing Views    Daniela
```

8

```
2748  8534462      Lux 1BD in CapHill. 97 Walkscore   Daniela
2892  8534462          99 Comfy 1Bed in Cap Hill   Daniela


                                       street property_type
1077     Union Street, Seattle, WA 98101, United States     Apartment
1149  Westlake Avenue North, Seattle, WA 98109, Unit...     Apartment
1152  Westlake Avenue North, Seattle, WA 98109, Unit...     Apartment
1155  Westlake Avenue North, Seattle, WA 98109, Unit...     Apartment
1159  Westlake Avenue North, Seattle, WA 98109, Unit...     Apartment
1230     Terry Avenue, Seattle, WA 98101, United States   Condominium
1237       Bell Street, Seattle, WA 98121, United States     Apartment
1238      Cedar Street, Seattle, WA 98121, United States     Apartment
1278       1st Avenue, Seattle, WA 98121, United States     Apartment
1298   Stewart Street, Seattle, WA 98101, United States   Condominium
1312   Stewart Street, Seattle, WA 98101, United States   Condominium
1323      Cedar Street, Seattle, WA 98121, United States     Apartment
1340      Cedar Street, Seattle, WA 98121, United States     Apartment
1366  Westlake Avenue North, Seattle, WA 98109, Unit...     Apartment
1376      Cedar Street, Seattle, WA 98121, United States     Apartment
1377       Bell Street, Seattle, WA 98121, United States     Apartment
1400       1st Avenue, Seattle, WA 98101, United States     Apartment
1404  Melrose Avenue East, Seattle, WA 98102, United...     Apartment
1424      Cedar Street, Seattle, WA 98121, United States     Apartment
1460      Western Ave, Seattle, WA 98104, United States     Apartment
1466       1st Avenue, Seattle, WA 98101, United States     Apartment
1469       9th Avenue, Seattle, WA 98101, United States     Apartment
1478      Western Ave, Seattle, WA 98104, United States     Apartment
1481       9th Avenue, Seattle, WA 98101, United States     Apartment
1482       9th Avenue, Seattle, WA 98101, United States     Apartment
1500   Stewart Street, Seattle, WA 98101, United States     Apartment
1519       3rd Avenue, Seattle, WA 98121, United States     Apartment
1532       9th Avenue, Seattle, WA 98101, United States     Apartment
1537       9th Avenue, Seattle, WA 98101, United States     Apartment
1548   Stewart Street, Seattle, WA 98101, United States     Apartment
1557  Western Avenue, Seattle, WA 98104, United States     Apartment
1559     Union Street, Seattle, WA 98101, United States     Apartment
1575       8th Avenue, Seattle, WA 98104, United States     Apartment
1606    Minor Avenue, Seattle, WA 98104, United States     Apartment
1613       8th Avenue, Seattle, WA 98104, United States     Apartment
1631    Minor Avenue, Seattle, WA 98104, United States     Apartment
1637    Minor Avenue, Seattle, WA 98104, United States     Apartment
1645       8th Avenue, Seattle, WA 98104, United States     Apartment
1652    Minor Avenue, Seattle, WA 98104, United States     Apartment
1654    Minor Avenue, Seattle, WA 98104, United States     Apartment
1657       8th Avenue, Seattle, WA 98104, United States     Apartment
1660       8th Avenue, Seattle, WA 98104, United States     Apartment
1667    Minor Avenue, Seattle, WA 98101, United States     Apartment
1695  Western Avenue, Seattle, WA 98104, United States     Apartment
```

```
2748     Belmont Avenue, Seattle, WA 98122, United States      Apartment
2892  Harvard Avenue East, Seattle, WA 98102, United...     Apartment
```

1.0.3   Question 3

Which properti on the selected database has the worst rating?

Steps and Answer: * Create copy of intrested columns * Dropped NA values * Plotting BoxPlot showing distribution of values * Identified one properfity having rating about 20 * Sorted ascending by rating column and identifed the worst rating and id of the property * Filtered properti id on original dataframe * The property_id on dataset with the worst rating is 9183838

```
In [26]: # removing missing data from review_scores_rating
         reviewdf = df[['id','review_scores_rating']].copy()
         reviewdf.dropna(inplace=True)
```

```
In [27]: #Conunting missing values
         reviewdf.isnull().sum().sum()
```

```
Out[27]: 0
```

```
In [23]: sns.boxplot(x=reviewdf['review_scores_rating'])
         plt.show()
```
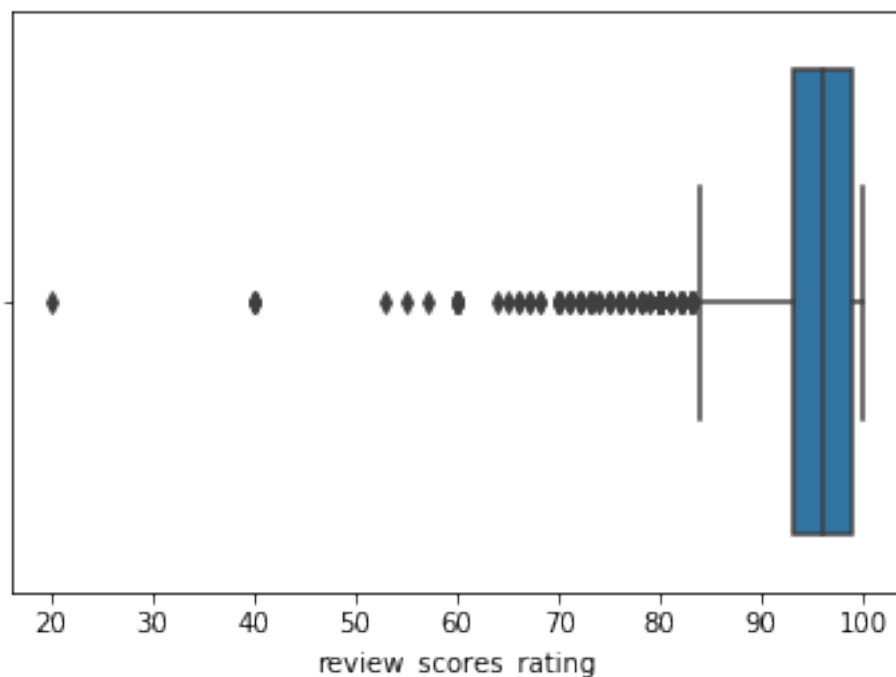


```
In [28]: reviewdf.sort_values(by=['review_scores_rating']).head()
```

```
Out[28]:            id  review_scores_rating
         1222  9183838                  20.0
         3148  5639238                  40.0
         2320  8474294                  40.0
         1162  9151865                  40.0
         1666  9151374                  40.0

In [30]: df.loc[df['id']==9183838]

Out[30]:            id                        listing_url      scrape_id  \
         1222  9183838  https://www.airbnb.com/rooms/9183838  20160104002432


               last_scraped                          name  \
         1222   2016-01-04  Belltown / Downtown Apartment


                                                     summary  \
         1222  This 700 square foot apartment is close to res...


                                                       space  \
         1222  The Space This is a 700 square foot, one bedro...


                                                 description experiences_offere
         1222  This 700 square foot apartment is close to res...                nor


               neighborhood_overview  ... review_scores_value requires_license licen
         1222                    NaN  ...                 2.0                f      N


               jurisdiction_names instant_bookable cancellation_policy  \
         1222          WASHINGTON                f              strict


                require_guest_profile_picture require_guest_phone_verification  \
         1222                              f                                f


               calculated_host_listings_count reviews_per_month
         1222                               2               1.0

         [1 rows x 92 columns]

In [ ]:
```