Jason Wu November 3, 2020
HC8: Individual Analysis

Raw Data

To begin, we can take a look at some of the output data attached below.
-
Summary of price values for Allerton:
count 42.000000
mean 87.595238
std 73.048617
min 33.000000
25% 47.000000
50% 66.500000
75% 104.750000
max 450.000000

Summary of price values for all 4 neighborhoods:
count 225.000000
mean 79.191111
std 53.272999
min 21.000000
25% 45.000000
50% 65.000000
75% 95.000000
max 450.000000

Summary of price values in all of Bronx
count 1091.000000
mean 87.496792
std 106.709349
min 0.000000
25% 45.000000
50% 65.000000
75% 99.000000
max 2500.000000

Summary of price values in all of NYC
count 48895.000000
mean 152.720687

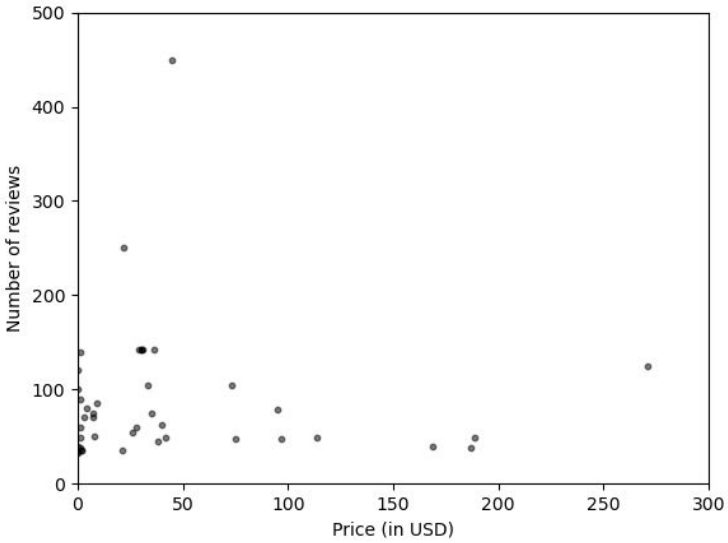
std 240.154170
min 0.000000
25% 69.000000
50% 106.000000
75% 175.000000
max 10000.000000

-

Data Analysis

The printed data here is relatively straightforward. The price values for airbnb locations in Allerton are similar to those of all 4 neighborhoods assigned to my group, however it appears that Allerton has one outlier in the form of the 450 datapoint that skews its mean and standard deviation slightly higher. Allerton's airbnb prices are around 70 to 100 dollars per day, which is similar to that of Bronx and lower than that of all of NYC (100 to 175). Considering our neighborhoods had a relatively poor safety rating (at around a C or so), it would make sense that the prices are lower. The owners in the Bronx and in Allerton are aware that customers are not as willing to rent an airbnb in less secure neighborhoods and will have to lower prices to make their offers seem more appealing.

Visualization

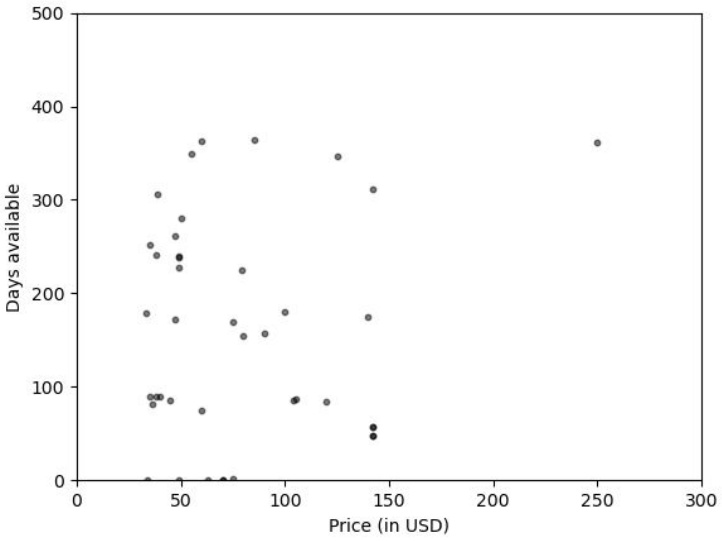


With this scatter graph, I tried to see if there is any correlation between the price of an airbnb location and the number of reviews it received. As shown in this graph, there is a weak negative correlation between price and reviews. As the price grows, the number of reviews tend to drop

down towards zero (ignoring the outlier at 275 price). This could be because more expensive airbnb locations receive less customers, especially in not particularly safe neighborhoods such as Allerton. Less customers mean less reviews.

In reviewing, I disregarded the data points that are at the y-axis. A price of 0 does not make any sense in the context of an airbnb and is likely an issue with uncleaned data.

With this simple bar graph, I looked for a difference between prices for the various (in this case, 2) room types. The graph shows that entire house room types tend to have a higher price than private room house types. Leasing out an entire apartment or house as an airbnb naturally results in higher upkeep (more rooms), which would explain the higher prices.



I also tried to see if the days available affected the price in any way, and as this graph shows there is no clear correlation between the two. There are airbnb locations of both high and low prices with both large and small numbers of available days. Once again, the values at the x-axis are disregarded since an availability of 0 days does not make sense in this context and seems to indicate uncleaned data. I have not implemented further code cleaning besides those used in the group analysis, so perhaps this may have been an issue.

Code

See next page.

```
#Modified by: Ze Hong Wu (Jason)
#Email: zehong.wu17@myhunter.cuny.edu
#Date: November 3, 2020
#Individual assignment for HC8.
#Uses code from the group assignment for HC8.

import pandas as pd
import matplotlib.pyplot as plt

airbnb = pd.read_csv("AirBnB_NYC_2019.csv")

#Cleaning the data
airbnb.fillna({'reviews_per_month':0, 'last_review':0},inplace=True)
airbnb.fillna({'host_name':"", 'name':""},inplace=True)
airbnb.drop(['host_id','host_name'], axis=1, inplace=True)

bronx = airbnb[airbnb['neighbourhood_group'] == 'Bronx']

mybronx = bronx[bronx['neighbourhood'].isin(['Allerton'])]
ourbronx = bronx[bronx['neighbourhood'].isin(['Fordham', 'Allerton', 'Kingsbridge',
'Concourse'])]

mybronxstat = mybronx.describe()
ourstat = ourbronx.describe()
totalstat = bronx.describe()
nycstat = airbnb.describe()
print("Summary of price values for Allerton: ")
print(mybronxstat['price'])
print("Summary of price values for all 4 neighborhoods: ")
print(ourstat['price'])
print("Summary of price values in all of Bronx")
print(totalstat['price'])
print("Summary of price values in all of NYC")
print(nycstat['price'])


allerton = airbnb[airbnb['neighbourhood'] == 'Allerton']
allerton_roomtypes = allerton.groupby(['room_type'])
price = allerton['price']
```

```
reviews = allerton['number_of_reviews']
y = price
x = reviews
colors = "black"

plt.xlim(0,300)
plt.ylim(0,500)

plt.scatter(x, y, s=10, c=colors, alpha=0.5)
plt.xlabel("Price (in USD)")
plt.ylabel("Number of reviews")
#fig1 = plt.gcf()
plt.savefig('PriceVsReviewsScatterplot.png')
plt.show()


rooms_index = list(mybronx['room_type'].unique())
rooms_index.sort()

allerton_roomtypes['price'].mean().plot.bar()
plt.ylabel('Price')
plt.gcf().subplots_adjust(bottom=0.4)
plt.title('mean price of listings in Allerton')
#fig1 = plt.gcf()
plt.savefig('RoomTypeVsPriceAllerton.png')
plt.show()

price = allerton['price']
avi = allerton['availability_365']
x = price
y = avi
colors = "black"

plt.xlim(0,300)
plt.ylim(0,500)

plt.scatter(x, y, s=10, c=colors, alpha=0.5)
plt.xlabel("Price (in USD)")
plt.ylabel("Days available")
#fig1 = plt.gcf()
```

```
plt.savefig('PriceVsAvailabilityScatterplot.png')
plt.show()
```