Source Of data : AirBnb Listing Statistics in Boston 2017

url: https://www.kaggle.com/airbnb/boston (https://www.kaggle.com/airbnb/boston)

```
In [66]:  import matplotlib.pyplot as plt
          import matplotlib
          import pandas as pd
          import numpy as np
          import seaborn as sns
          %matplotlib inline
```

```
In [68]:  #to read csv file from desktop
          Airbnb= pd.read_csv("C:/Users/Louis/Desktop/listings.csv")
```

In [65]:
```python
fig = plt.figure(figsize=(12,12))
# to add more space between subplots
fig.subplots_adjust(hspace=.5)

ax1 = fig.add_subplot(3, 2, 1)
ax2 = fig.add_subplot(3, 2, 2)
ax3 = fig.add_subplot(3, 2, 3)
ax4 = fig.add_subplot(3, 2, 4)
ax5 = fig.add_subplot(3, 2, 5)

#Preparation for line chart, to see how many  active Boston AirBnb Host registered over the
Airbnb['host_since'] = pd.to_datetime(Airbnb['host_since'])
Airbnb['year'] = Airbnb['host_since'].dt.year
Register_year= Airbnb['year'].value_counts().sort_index(ascending=True).index
Register_total = Airbnb['year'].value_counts().sort_index(ascending=True).values

#to plot a line chart
ax1.plot(Register_year, Register_total,'k-')
ax1.set_title('Host registration Year')
ax1.set_xlabel('Year')
ax1.set_ylabel('Frequency')
#ax2.pie(Film_permit_counts_category, labels = set( Film_permit['Category']), autopct='%1.1
#ax2.set_title('Film Permit category by Percentage')

# Preparation for a scatter plot which I want to see if the score rating is reliable. by pl
Rating = Airbnb['review_scores_rating']
reviews_per_month = Airbnb['reviews_per_month']
ax2.scatter(Rating , reviews_per_month)
ax2.set_title('Reliability of Reviews')
ax2.set_xlabel ('Rating')
ax2.set_ylabel ('reviews_per_month')


#to plot a pie about the cancellation policy
counts_of_policy = Airbnb['cancellation_policy'].value_counts()
types_of_policy = Airbnb['cancellation_policy'].value_counts().index
ax3.pie(counts_of_policy,autopct='%1.1f%%')
ax3.legend(types_of_policy, bbox_to_anchor=(.1, .6))
ax3.set_title('Cancellation Policy Categories')

#to create a bar chart to see what kind of rooms you can rent in Boston Via AirBnb
ax4.bar(Airbnb['room_type'].value_counts().index,Airbnb['room_type'].value_counts(),color=[
ax4.set_title('Rooms provided by AirBnB Host in Boston')
ax4.set_xlabel('Room Catgories')
ax4.set_ylabel('Frequecny')

#To create a histagram to show the price distribution
#to check data type of price column
Airbnb['price'].dtype
#to convert object data type by removing the dollar sign and convert to float
price = Airbnb['price'].replace('[\$,]', '', regex=True).astype(float)
# to plot axes.histogram
ax5.hist(price,bins=20)
ax5.axvline(price.mean(), color='k', linestyle='--', linewidth=1)
ax5.set_title('Price Range Histogram')
ax5.set_xlabel('Price in Dollars')
```
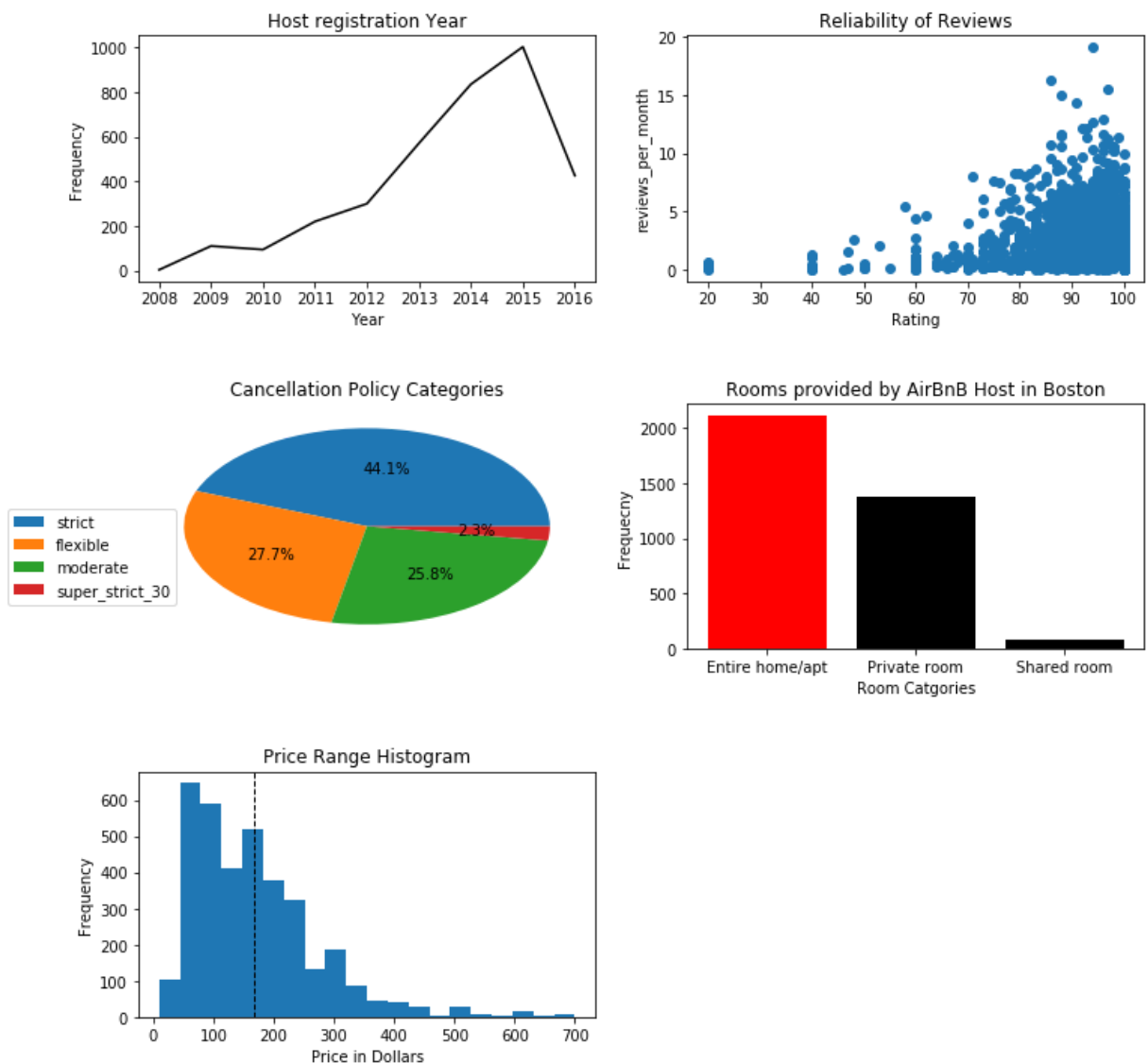
```
ax5.set_ylabel('Frequency')


def reset(event):
    circle1.remove()
```


This Dashboard reprensting the statistics for AirBnb host information who rent their residential property in Boston in 2017 The first line chart shows the year these hosts registered. There was a progressive growing trend since 2008 and reach the "Host Craze" in 2015 and it fell back in 2016, It also indicates that mostly like the host are relatively new to Airbnb with around 2 years experience

The second graph is trying find out if the reviews are reliable by plotting the frequency of review on certain host and the rating level AirBnb claims that their review accuracy is more than 90%, so , I'm expecting the host are having large volume of reviews and high ratings, however,as we can see from the graph the 90% of the data point locates at the right bottom corner, meaning, most of the host has high rating but with only a few guests, the accuracy cannot demonstrate if the rating is reliable or not.

The thrid graph representing the cancellation policy adopted by the host about 44% of the host chose strict policy, meaning full refund can only be issued witin 24 hours in advance.

The forth graph is telling people which kind of room type they can easily find in Boston Airbnb, it turns out more than 2000 host will rent the entire or apatment while less than a 100 of them would rent out a shared house which makes total since it's a lot easier to rent out the entire house due to privacy reason, cleaning purposes and safety concern as well.

The last graph is simply showing the distribution of price with a mean slight above $150 and the shape skewed to the right.