BHT Berliner Hochschule
für Technik

Individual Report for Data Visualization

Ahmed Dider Rahat (**916146**)
s40183@bht-berlin.de

Submitted on: 11th July 2022

Introduction: In our group report we made some interesting insight from the dataset where we divided the analysis into three broad categories such as basic statistics of the dataset, Univariate feature analysis, and Bivariate feature analysis.

In my individual report I will explain the implementation of "***room type vs. neighborhood group (fig-31)***" one of the figure from **Bivariate feature analysis**.

Intended Figure: The below figure represent the relationship between neighborhood group and the type of Airbnb.
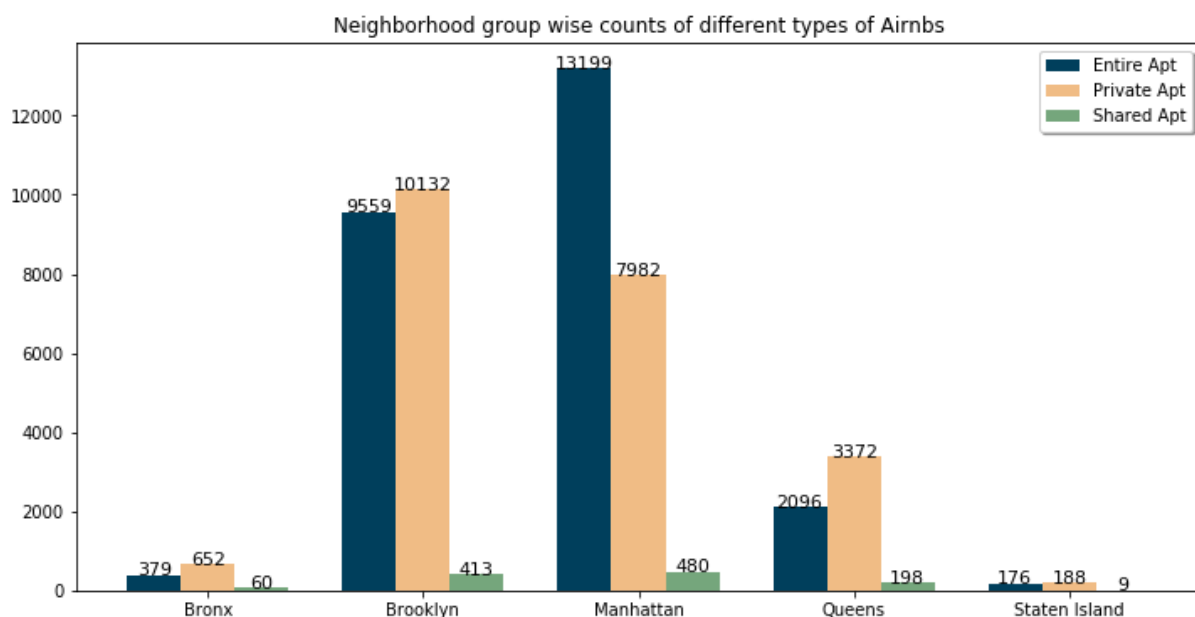

Figure- Bar plot of the relationship between neighborhood group and Airbnb type

Target Person for this analysis: The analysis convey the counts of each Airbnb type in different neighborhood group. Both type of user such as end user and hosts will be benefited from this analysis. The host could get the overview of how the apt type are vary in different type of neighborhood group. Also the end user get a clear picture which area is suitable for which type of Airbnbs.

Implementation: There are two parts in the implementation section. First one is data preparation and second one is plotting.

1. **Data Preparation:** For this analysis, I calculated the grouped counts of 'room_type' and 'neighborhood group'.

```python
# neighbourhood_group vs apt.type

apt_nhg_df = df[['id', 'room_type', 'neighbourhood_group']]
apt_nhg_df = apt_nhg_df.groupby(['room_type','neighbourhood_group'], as_index=False)['id'].count()

apt_nhg_df.rename(columns={'id':'ID_Counts','neighbourhood_group': 'Neighbourhood_Group', \
                           'room_type': 'Room_Type' }, inplace=True)
apt_nhg_df
```

2. **Plotting:** To plot the data I used a group bar plot. For that, I make three different list for each of the categories. Then I set the name of the neighborhood group as the label of X-axis. After that, I made three bar for each of the groups. First bar, represent 'entire home/apt', second bar represent 'private apt' and last one stand for 'shared apt'. Finally, I made a function that add the number of counts on the top of the each bar.

```python
labels = list(apt_nhg_df['Neighbourhood_Group'].unique())

entire_home = list(apt_nhg_df['ID_Counts'][apt_nhg_df['Room_Type'] == 'Entire home/apt'])
private_apt = list(apt_nhg_df['ID_Counts'][apt_nhg_df['Room_Type'] == 'Private room'])
shared_apt = list(apt_nhg_df['ID_Counts'][apt_nhg_df['Room_Type'] == 'Shared room'])

width = 0.25

fig, ax = plt.subplots(figsize=(12,6))

ax.set_title('Neighborhood group wise counts of different types of Airnbs')

x = np.arange(len(labels))

ax.set_xticks(np.arange(len(labels)))
ax.set_xticklabels(labels, rotation = 0)

rects1 = ax.bar(x-width, entire_home, width, label='Entire Apt', color='#003f5c')
rects2 = ax.bar(x, private_apt, width, label='Private Apt', color='#f0bc85')
rects3 = ax.bar(x + width, shared_apt, width, label='Shared Apt', color='#75a67c')

ax.legend(bbox_to_anchor=(1, 1),fancybox=True, shadow=True, ncol=1)

def addlabels(x, e, p, s):
    for i in range(len(x)):
        plt.text(i+.25, s[i], f'{str(round(s[i], 1))}', ha = 'center', fontsize = 11)
        plt.text(i, p[i], f'{str(round(p[i], 1))}', ha = 'center', fontsize = 11)
        plt.text(i-.25, e[i], f'{str(round(e[i], 1))}', ha = 'center', fontsize = 11)


addlabels(labels, entire_home, private_apt, shared_apt)

plt.savefig('../report/fig/plot_15.png',  bbox_inches='tight')
plt.show()
```

Insight of the figure: The figure is a simple bar-diagram which show the visual relationship between two variables i.e. neighborhood group and room/Airbnb type. From the figure we could made some simple interpretations. Firstly, largest number of entire home/apt, and shared apartment are situated in Manhattan. Secondly, Private apartments are mostly situated at Brooklyn. Thirdly, Queens and Bronox are at 3rd and 4th respectively in all types of apartments. And lastly, Staten Island has lowest number of count in all three categories.

Reason for choosing grouped bar plot: For visualize the data, I used a grouped bar plot for some interesting reason. Firstly, in a grouped bar chart we could look at how the second category variable changes within each level of the first, and/or vice-versa. Secondly, from a grouped bar chart we could draw the counts of each bar on top of it. And lastly, we could easily get the comparison of each group of data using this type of chart.

Design Decisions for the graph: Without proper design, we couldn't get meaningful information from a chart. So I use some important design decisions.

1. **Figure Size:** To plot the graph I used a rectangular figure with size (12, 6). The width of each bar is 0.25.

2. **Color:** I used a white background and three different colors for each types of Airbnb. My chosen bar colors are not too much bright or dark. The colors I used to represent the bars are Midnight Green (#003F5C) for entire apt., Gold (#F0BC85) for private apt., and Oxley (#75a67c) for shared apt.

3. **Axis ranges and labels:** In X-axis I set the name of the neighborhood group. For that I also make sure that the name should be in the middle of the bars. In Y-axis I placed the counts for each type. As the default range is 2,000 and there are total 6 labels present in the graph, I though it is perfect. For that I could not change the default labels and ranges.

4. **Legend:** I added a legend that represent the color for each type. I used a shadow to make the legend more attractive. I make the legend as 3x1 table, where each data are present into a separate row.

5. **Set the counts of the bar:** I set the counts of each bar on top of them. To placed them, I used plt.text() function. By using the function I set the counts as bar labels. I also set the numbers in the center position and set the font size as 11.