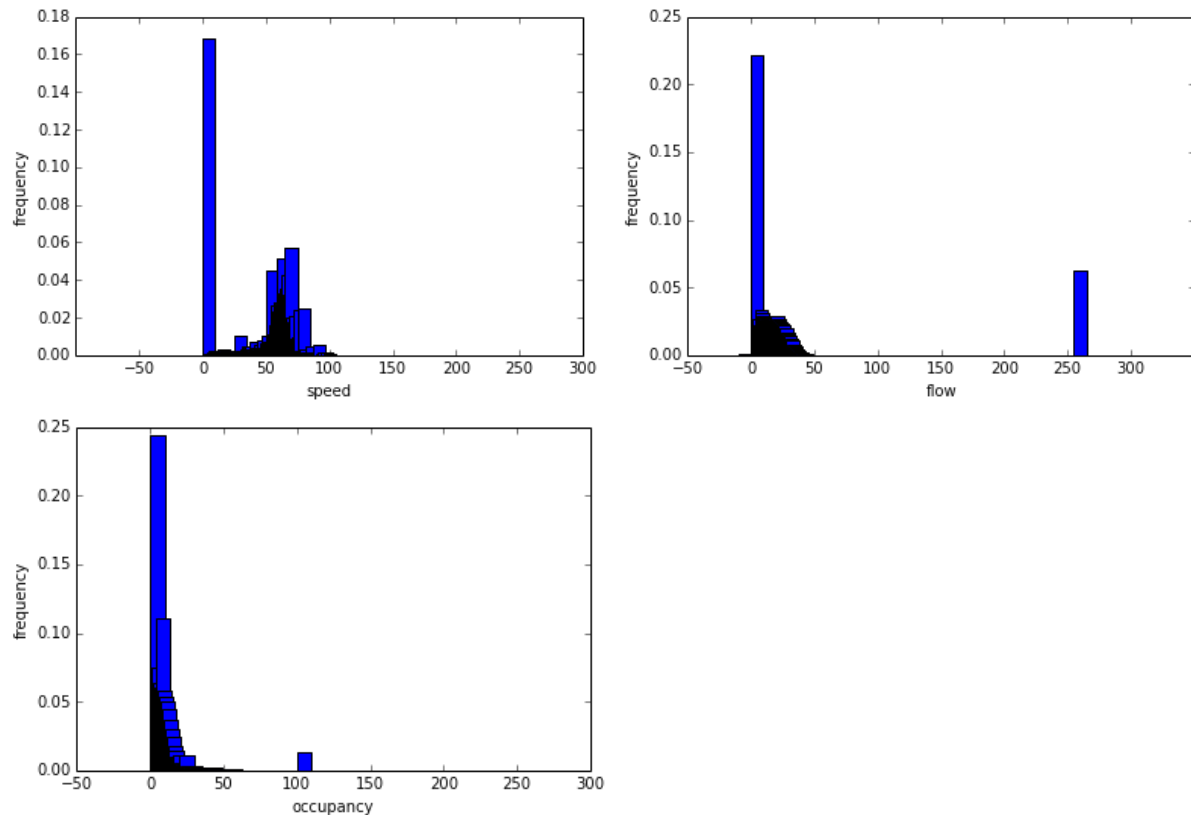Q3. Plotting the distributions of flow, speed and occupancy



Inference:

Looking at the graphs we can determine the distribution of speed, flow and occupancy of traffic. The speed frequency ranges from 0.0 to 0.18 whereas both the occupancy and the flow fall in the range 0.0 to 0.25. The speed of the traffic is mainly concentrated between the speeds of 0-100. Highest frequencies in all the three plots are at 0. This might be because of some missing values where they have put up the default value to be 0, or these might be places where there is no traffic, or stagnant traffic. These values need further analysis.

**Code**

```
import pandas as pd
import numpy as np
%matplotlib inline
```

```python
import matplotlib.pyplot as plt
df = pd.read_csv("/home/datascience/cleaning_test_06_09.tsv", sep = "\t")
speed_grouped  = df.groupby(lambda x: df['speed'][x])
speed_users = speed_grouped['trial_id'].nunique()
a = un_users.tolist()
type(a)
b = sum(a)
speed_users2 = speed_users*1./b
z = speed_grouped.size().index
xticks = [-50,0,50,100,150,200,250,300]
fig,ax= plt.subplots()
ax.bar(z, speed_users2, color = 'b', width = 10)
ax.set_xticks(xticks)
ax.set_xlabel('speed')
ax.set_ylabel('frequency')
plt.show()

flow_grouped  = df.groupby(lambda x: df['flow'][x])
flow_users = flow_grouped['trial_id'].nunique()
a = flow_users.tolist()
type(a)
b = sum(a)
flow_users2 = flow_users*1./b
z = flow_grouped.size().index
xticks = [-50,0,50,100,150,200,250,300]
fig,ax= plt.subplots()
ax.bar(z, flow_users2, color = 'b', width = 10)
ax.set_xticks(xticks)
ax.set_xlabel('flow')
ax.set_ylabel('frequency')
plt.show()

occupancy_grouped  = df.groupby(lambda x: df['occupancy'][x])
occupancy_users = occupancy_grouped['trial_id'].nunique()
a = occupancy_users.tolist()
type(a)
b = sum(a)
occupancy_users2 = occupancy_users*1./b
z = occupancy_grouped.size().index
xticks = [-50,0,50,100,150,200,250,300]
```

```python
fig,ax= plt.subplots()
ax.bar(z, occupancy_users2, color = 'b', width = 10)
ax.set_xticks(xticks)
ax.set_xlabel('occupancy')
ax.set_ylabel('frequency')
plt.show()
```