Q3

November 4, 2016

Group 19 Abhisek Mohanty Abhishek Nigam

In [24]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline
```

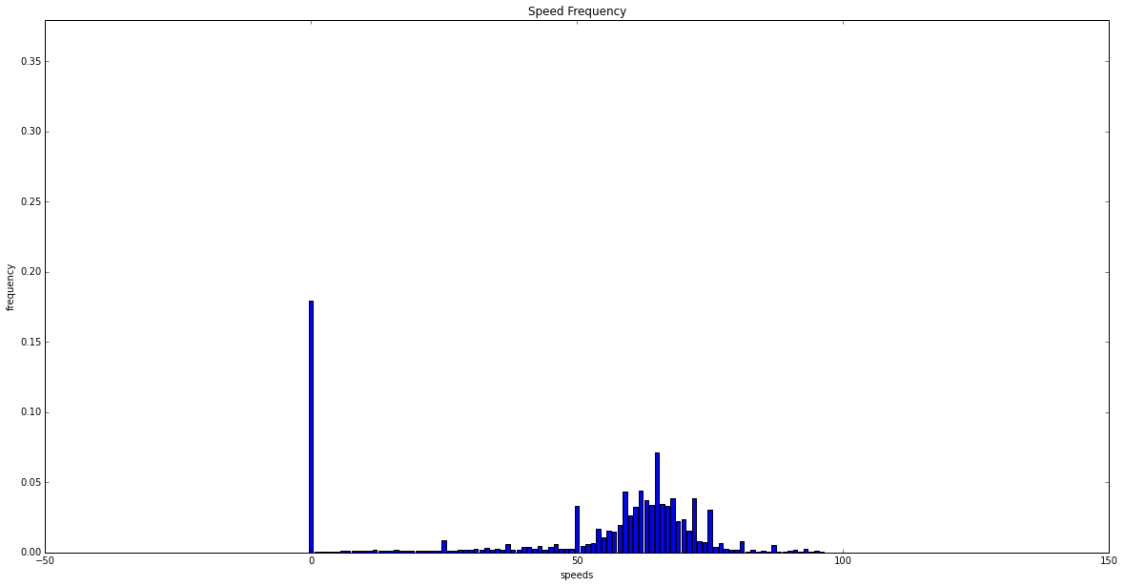
In [3]:
```
df = pd.read_csv('datasets/cleaning_test_06_09.tsv', delimiter='\t')
df.head()
```

Out[3]:

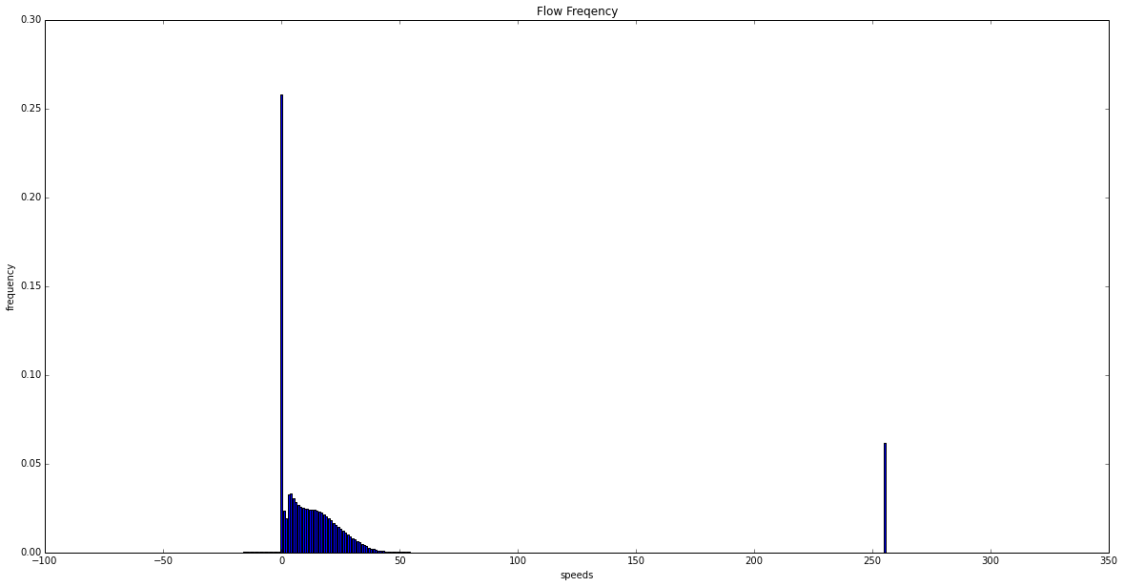
	trial_id	lane_id	measurement_start	speed	flow	occupancy	quality
0	c_06_09_000000000	12	2006-09-01T00:00:07-04:00	65	0	0	0
1	c_06_09_000000001	13	2006-09-01T00:00:07-04:00	63	3	2	0
2	c_06_09_000000002	14	2006-09-01T00:00:07-04:00	64	-2	1	0
3	c_06_09_000000003	15	2006-09-01T00:00:07-04:00	59	4	3	0
4	c_06_09_000000004	16	2006-09-01T00:00:07-04:00	66	5	1	0

In [18]:
```
# speed

test_df = df
total_count = test_df['speed'].count()
cols = ['value', 'count']
counts = pd.DataFrame(test_df['speed'].value_counts().reset_index())
counts.columns = cols
counts['count'] = counts['count']/float(total_count)
# counts.head()
plt.figure(figsize=(20,10))
plt.bar(counts['value'], counts['count'], align='center')
plt.xlabel('speeds')
plt.ylabel('frequency')
plt.xlim(-50, 150)
y_lim_up = np.max(counts['count']) + 0.2
y_lim_up = 1 if y_lim_up > 1 else y_lim_up
plt.ylim(0, y_lim_up)
plt.title("Speed Frequency")
```

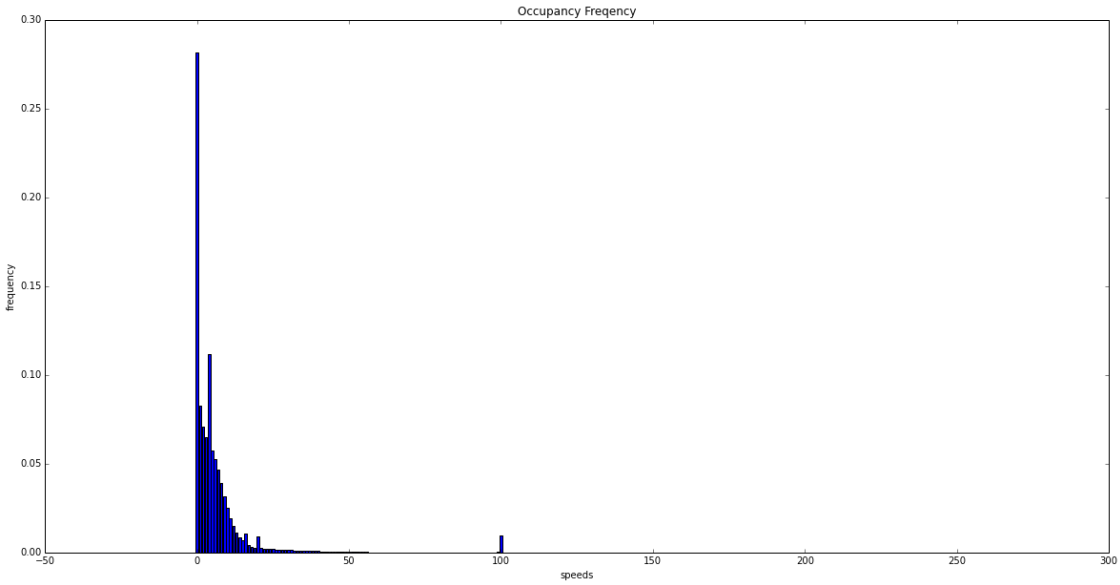
Out[18]:
```
<matplotlib.text.Text at 0x7f9f12409410>
```




```
In [21]: flow_df = df
         total_count = flow_df['flow'].count()
         cols = ['value', 'count']
         flow_counts = pd.DataFrame(flow_df['flow'].value_counts().reset_index())
         flow_counts.columns = cols
         flow_counts['count'] = flow_counts['count']/float(total_count)
         flow_counts.head()

         plt.figure(figsize=(20,10))
         plt.bar(flow_counts['value'], flow_counts['count'], align='center', alpha=1)
         plt.xlabel('speeds')
         plt.ylabel('frequency')
         plt.title('Flow Freqency')

Out[21]: <matplotlib.text.Text at 0x7f9f11a709d0>
```




```
In [20]: occupancy_df = df
         total_count = occupancy_df['occupancy'].count()
         cols = ['value', 'count']
         occupancy_counts = pd.DataFrame(occupancy_df['occupancy'].value_counts().reset_index())
         occupancy_counts.columns = cols
         occupancy_counts['count'] = occupancy_counts['count']/float(total_count)
         occupancy_counts.head()

         plt.figure(figsize=(20,10))
         plt.bar(occupancy_counts['value'], occupancy_counts['count'], align='center', alpha=1)
         plt.xlabel('speeds')
         plt.ylabel('frequency')
         plt.title('Occupancy Freqency')

Out[20]: <matplotlib.text.Text at 0x7f9f11ec5690>
```


Occupancy Freqency
frequency
speeds
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−50
0
50
100
150
200
250
300

In []: