

ZONE 3451

```
import pandas as pd
zone_3451_flow_df = pd.read_csv("/home/datascience/Downloads/cleaning1/3451/flow.tsv",
                                na_values=['-'], sep='\t', names=['flow'])
zone_3451_speed_df = pd.read_csv("/home/datascience/Downloads/cleaning1/3451/speed.tsv",
                                na_values=['-'], sep='\t', names = ['speed'])
zone_3451_occupancy_df =
pd.read_csv("/home/datascience/Downloads/cleaning1/3451/occupancy.tsv",
            na_values=['-'], sep='\t', names = ['occupancy'])
zone_3451_vector = pd.concat([zone_3451_flow_df, zone_3451_speed_df, zone_3451_occupancy_df],
axis=1)
def isNaN(x):
    return (x == x) == False
is_3451_f_Nan = ~isNaN(zone_3451_vector['flow'])
zone_3451_vector = zone_3451_vector[is_3451_f_Nan]
is_3451_s_Nan = ~isNaN(zone_3451_vector['speed'])
zone_3451_vector = zone_3451_vector[is_3451_s_Nan]
is_3451_o_Nan = ~isNaN(zone_3451_vector['occupancy'])
zone_3451_vector = zone_3451_vector[is_3451_o_Nan]
from scipy.stats import multivariate_normal
import numpy as np
mean = np.mean(zone_3451_vector)
cov = zone_3451_vector.cov()
zone_3451_vector['probability'] = multivariate_normal.pdf(zone_3451_vector, mean, cov)
zone_3451_vector['flow'] = zone_3451_vector['flow'].astype(int)
zone_3451_vector['speed'] = zone_3451_vector['speed'].astype(int)
zone_3451_vector['occupancy'] = zone_3451_vector['occupancy'].astype(int)
zone_3451_vector = zone_3451_vector.reset_index()
is_occ_neg = zone_3451_vector['occupancy'] < 0
zone_3451_vector_is_neg_occ = zone_3451_vector[is_occ_neg]
is_speed_neg = zone_3451_vector['speed'] < 0
zone_3451_vector_is_neg_speed = zone_3451_vector[is_speed_neg]
is_flow_neg = zone_3451_vector['flow'] < 0
zone_3451_vector_is_neg_flow = zone_3451_vector[is_flow_neg]
zone_3451_vector_is_neg =
zone_3451_vector_is_neg_flow.append(zone_3451_vector_is_neg_speed.append(zone_3451_vector_is
_neg_occ))
zone_3451_vector_is_neg = zone_3451_vector_is_neg.drop_duplicates()
zone_3451_vector_is_neg['probability'] = 0
zone_3451_vector = zone_3451_vector.append(zone_3451_vector_is_neg)
zone_3451_vector['index'] = zone_3451_vector.index
zone_3451_vector = zone_3451_vector.drop_duplicates(['index'], take_last = True)
zone_3451_vector.set_index = zone_3451_vector['index']
```

```
zone_3451_vector = zone_3451_vector.drop('index', axis =1)
zone_3451_vector = zone_3451_vector.sort_index()
zone_3451_sorted_vector = zone_3451_vector.sort(['probability'])
zone_3451_sorted_vector = zone_3451_sorted_vector.reset_index(drop=True)
is_100th_row = zone_3451_sorted_vector.index % 100 == 0
zone_3451_sorted_vector_is_100th = zone_3451_sorted_vector[is_100th_row]
zone_3451_sorted_vector_is_100th['probability'] =
zone_3451_sorted_vector_is_100th['probability'].apply('{:.8f}'.format)
zone_3451_sorted_vector_is_100th.to_csv("/home/datascience/Downloads/cleaning1/3451/3451.txt",
sep='\t', index=False)
```