In [1]:
```
from sqlalchemy import create_engine
import pymysql
import pandas as pd
import datetime as dt
import seaborn as sns
import matplotlib.pyplot as plt

pd.set_option('display.max_columns', None)

sqlEngine       = create_engine('mysql+pymysql://root:@127.0.0.1', poo
l_recycle=3600)
dbConnection    = sqlEngine.connect()
#data            = pd.read_sql("SELECT * FROM atx_traffic.tmsr WHERE o
rigin_reader_identifier = 'benwhite_riverside' AND destination_reader_
identifier = 'fm973_tx71';", dbConnection)

data            = pd.read_sql("SELECT * FROM atx_traffic.tmsr WHERE or
igin_reader_identifier = 'fm973_tx71' AND destination_reader_identifie
r = 'benwhite_riverside';", dbConnection)
#data            = pd.read_sql("SELECT * FROM atx_traffic.tmsr;", dbCo
nnection)
#pd.set_option('display.expand_frame_repr', False)
#df
dbConnection.close()
```

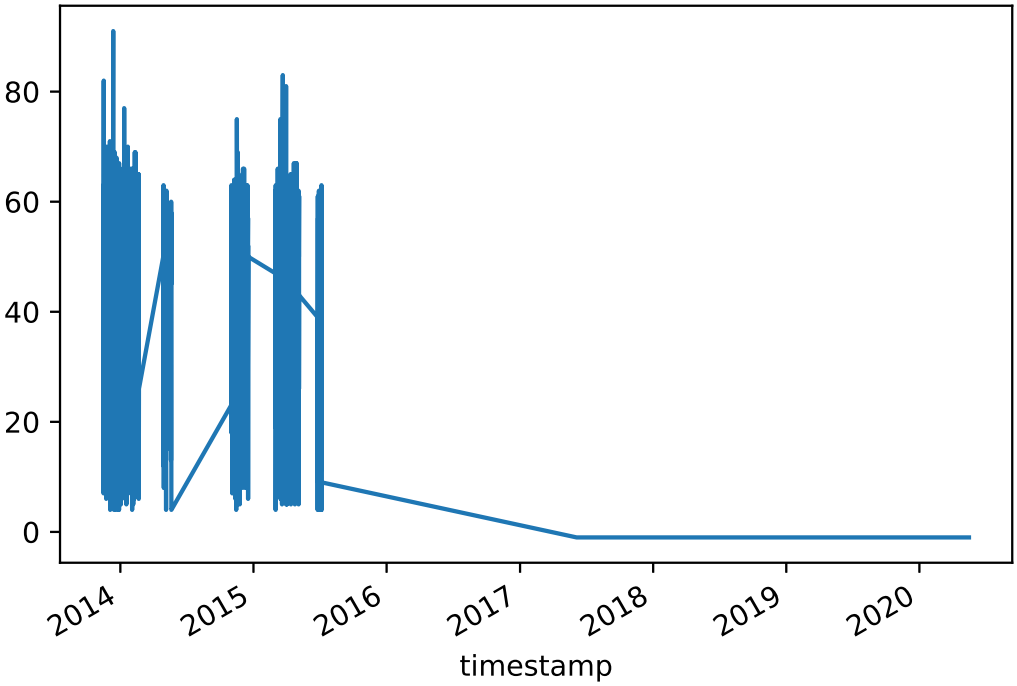
In [2]:
```
df = data[['segment_length_miles','timestamp','average_travel_time_sec
onds','average_speed_mph','number_samples','standard_deviation']]
```

In [3]:
```
df['timestamp'] = pd.to_datetime(df['timestamp'],infer_datetime_format
=True)
df['time'] = df['timestamp'].dt.time
df['weekday'] = df['timestamp'].dt.weekday
```

In [4]:
```
df.set_index('timestamp',inplace=True)
df.sort_index(inplace=True)
```

In [5]:
```
df.average_speed_mph.plot.line()
```

Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fadbb8403d0>



In [6]:
```
df = df[df.average_speed_mph != -1]
```

In [7]:
```
df['totalspeed'] = df.average_speed_mph * df.number_samples
df['totaltraveltime'] = df.average_travel_time_seconds * df.number_sam
ples
```

In [8]:
```
df['pooled_variance'] = (df.number_samples - 1) * (df.standard_deviati
on) ** 2
```

In [9]:
```
df
```

Out[9]:

timestamp	segment_length_miles	average_travel_time_seconds	average_speed_mph	number_s
2013-11-15 05:15:00	3.5	343	37	
2013-11-15 05:30:00	3.5	211	60	
2013-11-15 05:45:00	3.5	218	58	
2013-11-15 06:00:00	3.5	251	50	
2013-11-15 06:15:00	3.5	251	50	
...	...	...	...	
2015-07-07 01:30:00	3.5	1428	9	
2015-07-07 02:00:00	3.5	240	52	
2015-07-07 02:45:00	3.5	2865	4	
2015-07-07 04:00:00	3.5	241	52	
2015-07-07 04:30:00	3.5	1421	9	

15193 rows × 10 columns

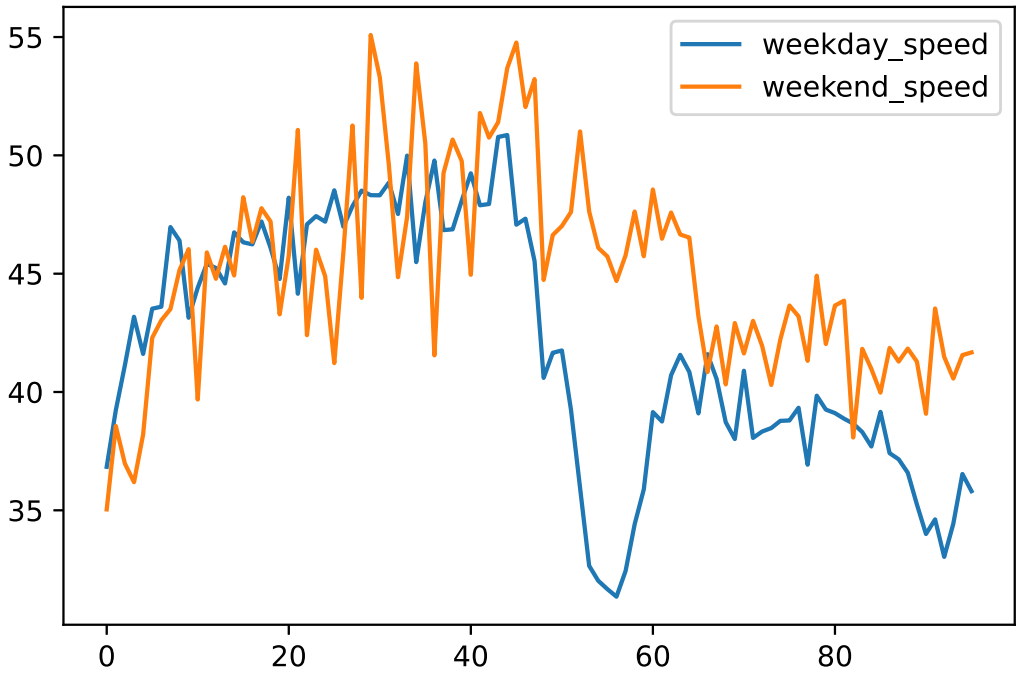
In [10]:
```
#df.to_csv('benwhite_riverside-fm973_tx71.csv')
df.to_csv('fm973_tx71-benwhite_riverside.csv')
```

In [11]:
```
times = []
timechunks = []
weekday_num_samples = []
weekend_num_samples = []
weekday_speed = []
weekend_speed = []
weekday_travel_time = []
weekend_travel_time = []
weekday_pooled_sd_mph = []
weekend_pooled_sd_mph = []

timechunk = 0
for j in range(24):
    for i in range(4):
        df_temp_weekday = df[(df.time == dt.time(j,15*i)) & (df.weekda
y < 5)]
        weekday_speed.append(df_temp_weekday.totalspeed.sum() / df_tem
p_weekday.number_samples.sum())
        weekday_travel_time.append(df_temp_weekday.totaltraveltime.sum
() / df_temp_weekday.number_samples.sum())
        weekday_pooled_sd_mph.append(df_temp_weekday.pooled_variance.s
um()/(df_temp_weekday.number_samples.sum() - df_temp_weekday.shape[0])
)
        weekday_num_samples.append(df_temp_weekday.number_samples.mean
())

        df_temp_weekend = df[(df.time == dt.time(j,15*i)) & (df.weekda
y >= 5)]
        weekend_speed.append(df_temp_weekend.totalspeed.sum() / df_tem
p_weekend.number_samples.sum())
        weekend_travel_time.append(df_temp_weekend.totaltraveltime.sum
() / df_temp_weekend.number_samples.sum())
        weekend_pooled_sd_mph.append(df_temp_weekend.pooled_variance.s
um()/(df_temp_weekend.number_samples.sum() - df_temp_weekend.shape[0])
)
        weekend_num_samples.append(df_temp_weekend.number_samples.mean
())

        timechunks.append(timechunk)
        timechunk += 1
        times.append(dt.time(j,15*i))
plt.plot(timechunks, weekday_speed, label = 'weekday_speed')
plt.plot(timechunks, weekend_speed, label = 'weekend_speed')
plt.legend()
plt.show()
```



In [12]:
```
df_timechunk = pd.DataFrame(list(
    zip(times,
    timechunks,
    weekday_num_samples,
    weekend_num_samples,
    weekday_speed,
    weekend_speed,
    weekday_travel_time,
    weekend_travel_time,
    weekday_pooled_sd_mph,
    weekend_pooled_sd_mph)), columns=[
    'time',
    'timechunk',
    'weekday_num_samples',
    'weekend_num_samples',
    'weekday_speed',
    'weekend_speed',
    'weekday_travel_time',
    'weekend_travel_time',
    'weekday_pooled_sd_mph',
    'weekend_pooled_sd_mph'
])
```

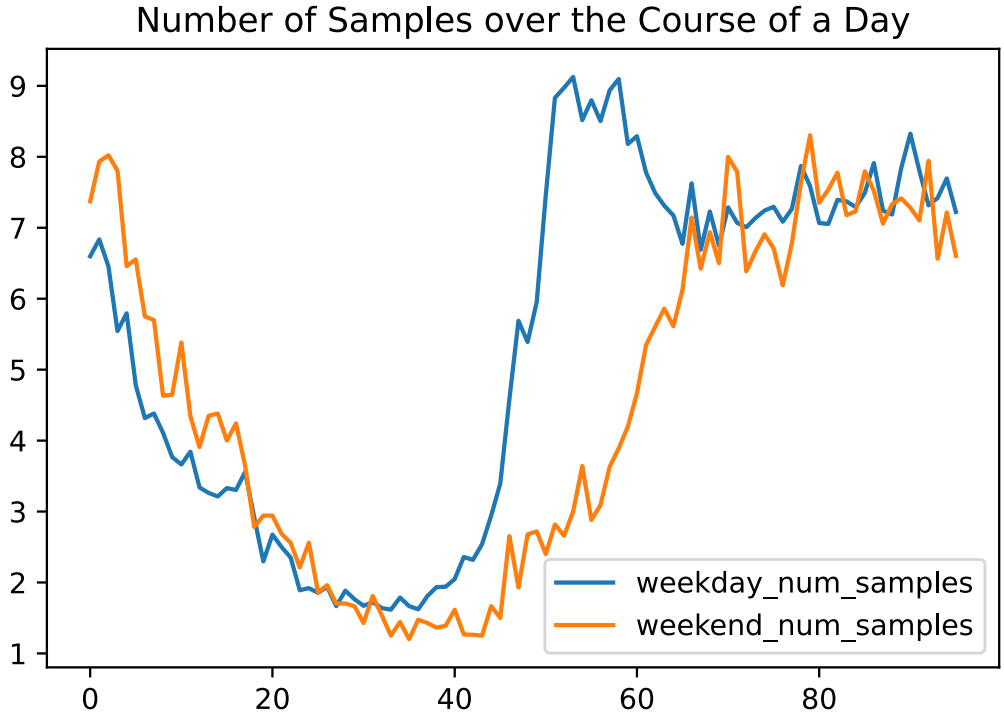
In [13]:
```
df_timechunk
```

Out[13]:

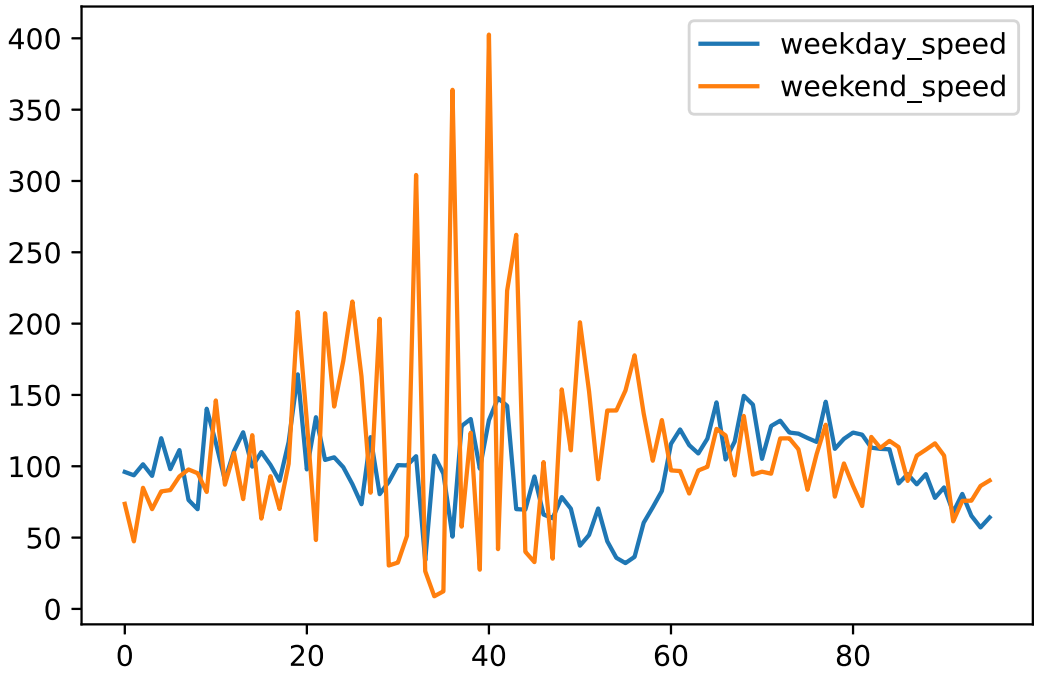
	time	timechunk	weekday_num_samples	weekend_num_samples	weekday_speed	wee
0	00:00:00	0	6.598540	7.372549	36.838496	
1	00:15:00	1	6.835714	7.934783	39.226750	
2	00:30:00	2	6.457143	8.021277	41.132743	
3	00:45:00	3	5.543478	7.804348	43.179085	
4	01:00:00	4	5.793651	6.460000	41.608219	
...	...	...	...	...	...	
91	22:45:00	91	7.798658	7.102041	34.617040	
92	23:00:00	92	7.315068	7.943396	33.031835	
93	23:15:00	93	7.420690	6.562500	34.431227	
94	23:30:00	94	7.695364	7.215686	36.534423	
95	23:45:00	95	7.223776	6.604167	35.810261	

96 rows × 10 columns

In [18]:
```
plt.plot(timechunks, weekday_num_samples, label = 'weekday_num_samples
')
plt.plot(timechunks, weekend_num_samples, label = 'weekend_num_samples
')
plt.legend()
plt.title('Number of Samples over the Course of a Day')
plt.show()
```



In [15]:
```
plt.plot(timechunks, weekday_pooled_sd_mph, label = 'weekday_speed')
plt.plot(timechunks, weekend_pooled_sd_mph, label = 'weekend_speed')
plt.legend()
plt.show()
```



In [16]:
```
df.average_speed_mph / df.average_travel_time_seconds
```

Out[16]:
```
timestamp
2013-11-15 05:15:00    0.107872
2013-11-15 05:30:00    0.284360
2013-11-15 05:45:00    0.266055
2013-11-15 06:00:00    0.199203
2013-11-15 06:15:00    0.199203
                           ...
2015-07-07 01:30:00    0.006303
2015-07-07 02:00:00    0.216667
2015-07-07 02:45:00    0.001396
2015-07-07 04:00:00    0.215768
2015-07-07 04:30:00    0.006334
Length: 15193, dtype: float64
```

In [17]:
```
df_timechunk.to_csv('fm973_tx71-benwhite_riverside-timechunks.csv', index=False)
```