In [22]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [23]:
```
#Import data files
```

In [24]:
```
d16_05=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2016/hos_revised_flatfiles_archive_05_2016/Timely and Effective Care - Hospital.csv",engine='python')
```

In [25]:
```
d16_08=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2016/hos_revised_flatfiles_archive_08_2016/Timely and Effective Care - Hospital.csv",engine='python')
```

In [26]:
```
d16_11=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2016/hos_revised_flatfiles_archive_11_2016/Timely and Effective Care - Hospital.csv",engine='python')
```

In [27]:
```
d16_12=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2016/hos_revised_flatfiles_archive_12_2016/Timely and Effective Care - Hospital.csv",encoding='cp1252')
```

In [28]:
```
d15_01=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2015/hos_revised_flatfiles_archive_01_2015/Timely and Effective Care - Hospital.csv",engine='python')
d15_04=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2015/hos_revised_flatfiles_archive_04_2015/Timely and Effective Care - Hospital.csv",engine='python')
d15_07=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2015/hos_revised_flatfiles_archive_07_2015/Timely and Effective Care - Hospital.csv",engine='python')
d15_10=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2015/hos_revised_flatfiles_archive_10_2015/Timely and Effective Care - Hospital.csv",engine='python')
d15_12=pd.read_csv("/Users/bikashadhikari/Desktop/GMU Assignment/HAP 725/1/Teach one /Assignment 1/2015/hos_revised_flatfiles_archive_12_2015/Timely and Effective Care - Hospital.csv",engine='python')
```

In [29]:
```
d16_12.head()
d16_12.shape
```

Out[29]:
```
(105754, 16)
```

In [30]:
```
d16_05.dtypes
```

Out[30]:
```
Provider ID           object
Hospital Name         object
Address               object
City                  object
State                 object
ZIP Code               int64
County Name           object
Phone Number           int64
Condition             object
Measure ID            object
Measure Name          object
Score                 object
Sample                object
Footnote              object
Measure Start Date    object
Measure End Date      object
dtype: object
```

In [31]:
```
d16_08.dtypes
```

Out[31]:
```
Provider ID           object
Hospital Name         object
Address               object
City                  object
State                 object
ZIP Code               int64
County Name           object
Phone Number           int64
Condition             object
Measure ID            object
Measure Name          object
Score                 object
Sample                object
Footnote              object
Measure Start Date    object
Measure End Date      object
dtype: object
```

In [32]:
```
#filtering for only southeast alabama medical center for each year
#select measure id "SCIP_INF 3"; this measure refers to prophylactic antibiotic use.
```

In [33]:
```
big_df=pd.concat([d16_05,d16_08,d16_11,d16_12,d15_01,d15_04,d15_07,d15_10,d15_12], ignore_index=True)
```

In [50]:
```
big_df.shape
```

Out[50]:
```
(1728141, 16)
```

In [51]:
```
big_dfs= big_df[['Provider ID','Hospital Name','State','Measure ID','Score','Sample' ,'Measure Start Date','Measure End Date']]
```

In [52]:
```
big_df_final=big_dfs.loc[big_dfs["Hospital Name"]=='SOUTHEAST ALABAMA MEDICAL CENTER']
```

In [53]:
```
big_df_finals=big_df_final.loc[big_df_final["Measure ID"]=="SCIP_INF_3"]
```

In [58]:
```
big_df_finals
```

Out[58]:

	Provider ID	Hospital Name	State	Measure ID	Score	Sample	Measure Start Date	Measure End Date
23	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99	340	7/1/2014	6/30/2015
180788	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	100	256	10/1/2014	9/30/2015
369502	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99	170	01/01/2015	09/30/2015
682434	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98	324	04/01/2013	03/31/2014
899747	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98	330	07/01/2013	06/30/2014
1117462	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98	329	10/01/2013	09/30/2014
1335283	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98	332	01/01/2014	12/31/2014
1547446	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99	337	4/1/2014	3/31/2015

In [55]:
```
big_df_finals.shape
```

Out[55]:
```
(8, 8)
```

In [66]:
```
big_df_finals.dtypes
```

Out[66]:
```
Provider ID            object
Hospital Name          object
State                  object
Measure ID             object
Score                 float64
Sample                 object
Measure Start Date     object
Measure End Date       object
dtype: object
```

Converting filtered data to the required data type

In [70]:
```
big_df_finals['Score']=big_df_finals['Score'].astype(float)
big_df_finals['Sample']=big_df_finals['Sample'].astype(float)
big_df_finals['Measure Start Date']=pd.to_datetime(big_df_finals['Measure Start Date'])
```

```
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/154143594.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Score']=big_df_finals['Score'].astype(float)
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/154143594.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Sample']=big_df_finals['Sample'].astype(float)
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/154143594.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Measure Start Date']=pd.to_datetime(big_df_finals['Measure Start Date'])
```

In [71]:
```
big_df_finals.dtypes
```

Out[71]:
```
Provider ID                    object
Hospital Name                  object
State                          object
Measure ID                     object
Score                         float64
Sample                        float64
Measure Start Date     datetime64[ns]
Measure End Date               object
dtype: object
```

Calculating "Overuse Rate", "Grand Rate", "Upper and Lower Limits"

In [72]:
```
big_df_finals['OverUse Rate']=big_df_finals['Score']/big_df_finals['Sample']
big_df_finals['Grand Rate']=sum(big_df_finals['Score'])/sum(big_df_finals['Sample'])
```

```
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/3168625771.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['OverUse Rate']=big_df_finals['Score']/big_df_finals['Sample']
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/3168625771.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Grand Rate']=sum(big_df_finals['Score'])/sum(big_df_finals['Sample'])
```

In [77]:
```
big_df_finals['Upper Limit']=big_df_finals['Grand Rate']+(1.96*np.sqrt((big_df_finals['Grand Rate']*(1-big_df_finals['Grand Rate']))/big_df_finals['Sample']))
big_df_finals['Lower Limit']=big_df_finals['Grand Rate']-(1.96*np.sqrt((big_df_finals['Grand Rate']*(1-big_df_finals['Grand Rate']))/big_df_finals['Sample']))
```

```
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/2208927651.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Upper Limit']=big_df_finals['Grand Rate']+(1.96*np.sqrt((big_df_finals['Grand Rate']*(1-big_df_finals['Grand Rate']))/big_df_finals['Sample']))
/var/folders/cy/wngl16rn4hg_0y7sls7vlf740000gn/T/ipykernel_1540/2208927651.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  big_df_finals['Lower Limit']=big_df_finals['Grand Rate']-(1.96*np.sqrt((big_df_finals['Grand Rate']*(1-big_df_finals['Grand Rate']))/big_df_finals['Sample']))
```

In [78]:
```
big_df_finals
```

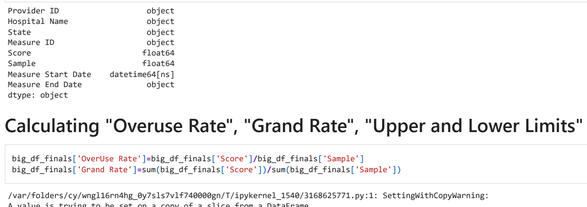
Out[78]:

	Provider ID	Hospital Name	State	Measure ID	Score	Sample	Measure Start Date	Measure End Date	OverUse Rate	Grand Rate	Upper Limit	Lower Limit
23	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99.0	340.0	2014-07-01	6/30/2015	0.291176	0.326303	0.376141	0.276465
180788	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	100.0	256.0	2014-10-01	9/30/2015	0.390625	0.326303	0.383738	0.268867
369502	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99.0	170.0	2015-01-01	09/30/2015	0.582353	0.326303	0.396784	0.255821
682434	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98.0	324.0	2013-04-01	03/31/2014	0.302469	0.326303	0.377356	0.275249
899747	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98.0	330.0	2013-07-01	06/30/2014	0.296970	0.326303	0.376890	0.275715
1117462	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98.0	329.0	2013-10-01	09/30/2014	0.297872	0.326303	0.376967	0.275639
1335283	010001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	98.0	332.0	2014-01-01	12/31/2014	0.295181	0.326303	0.376737	0.275868
1547446	10001	SOUTHEAST ALABAMA MEDICAL CENTER	AL	SCIP_INF_3	99.0	337.0	2014-04-01	3/31/2015	0.293769	0.326303	0.376362	0.276244

Plotting the Graph

In [85]:
```
plt.figure(figsize=(15,5))
ax=plt.gca()
big_df_finals.plot(kind='line',x='Measure Start Date',y='OverUse Rate', color='blue', marker='o', ax=ax)
big_df_finals.plot(kind='line',x='Measure Start Date',y='Lower Limit', color='red',  ax=ax)
big_df_finals.plot(kind='line',x='Measure Start Date',y='Upper Limit', color='red', ax=ax)
plt.title('Rate of Post-Op Antibiotic Overuse:  2015 -  2016')
plt.xlabel('Measure Start Date')
plt.ylabel('Rate of Overuser')
```

Out[85]:
```
Text(0, 0.5, 'Rate of Overuser')
```



In [ ]:
```
```