Importing Libraries

In [1]:
```
import pandas as pd
import numpy as np
```

Load the Data

In [2]:
```
df=pd.read_csv("https://raw.githubusercontent.com/dsrscientist/dataset1/master/hrdata.csv")
df.head()
```

Out[2]:

	Name	HireDate	Salary	SickDaysremaining
0	Graham Chapman	03/15/14	50000.0	10
1	John Cleese	06/01/15	65000.0	8
2	Eric Idle	05/12/14	45000.0	10
3	Terry Jones	11/01/13	70000.0	3
4	Terry Gilliam	08/12/14	48000.0	7

Lets check our data shape: Dataset has 6 rows and 4 columns.

In [3]:
```
df.shape
```

Out[3]:
```
(6, 4)
```

In [4]:
```
df.columns   # This will print the names of all columns
```

Out[4]:
```
Index(['Name', 'HireDate', 'Salary', 'SickDaysremaining'], dtype='object')
```

In [5]:
```
df.head()  # Will give you first 5 records
```

Out[5]:

	Name	HireDate	Salary	SickDaysremaining
0	Graham Chapman	03/15/14	50000.0	10
1	John Cleese	06/01/15	65000.0	8
2	Eric Idle	05/12/14	45000.0	10
3	Terry Jones	11/01/13	70000.0	3
4	Terry Gilliam	08/12/14	48000.0	7

In [6]:
```
df.tail()   # This will print the last n rows of the Data Frame
```

Out[6]:

	Name	HireDate	Salary	SickDaysremaining
1	John Cleese	06/01/15	65000.0	8
2	Eric Idle	05/12/14	45000.0	10
3	Terry Jones	11/01/13	70000.0	3
4	Terry Gilliam	08/12/14	48000.0	7
5	Michael Palin	05/23/13	66000.0	8

In [7]:
```
df.info()  # This will give Index, Datatype and Memory information
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 4 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               6 non-null      object 
 1   HireDate           6 non-null      object 
 2   Salary             6 non-null      float64
 3   SickDaysremaining  6 non-null      int64  
dtypes: float64(1), int64(1), object(2)
memory usage: 320.0+ bytes
```

In [9]:
```
# You can get idea about which column has missing values using this
df.describe()
```

Out[9]:

	Salary	SickDaysremaining
count	6.000000	6.000000
mean	57333.333333	7.666667
std	10838.204033	2.581989
min	45000.000000	3.000000
25%	48500.000000	7.250000
50%	57500.000000	8.000000
75%	65750.000000	9.500000
max	70000.000000	10.000000

Checking the NULL values

In [10]:
```
df.isnull().sum()
```

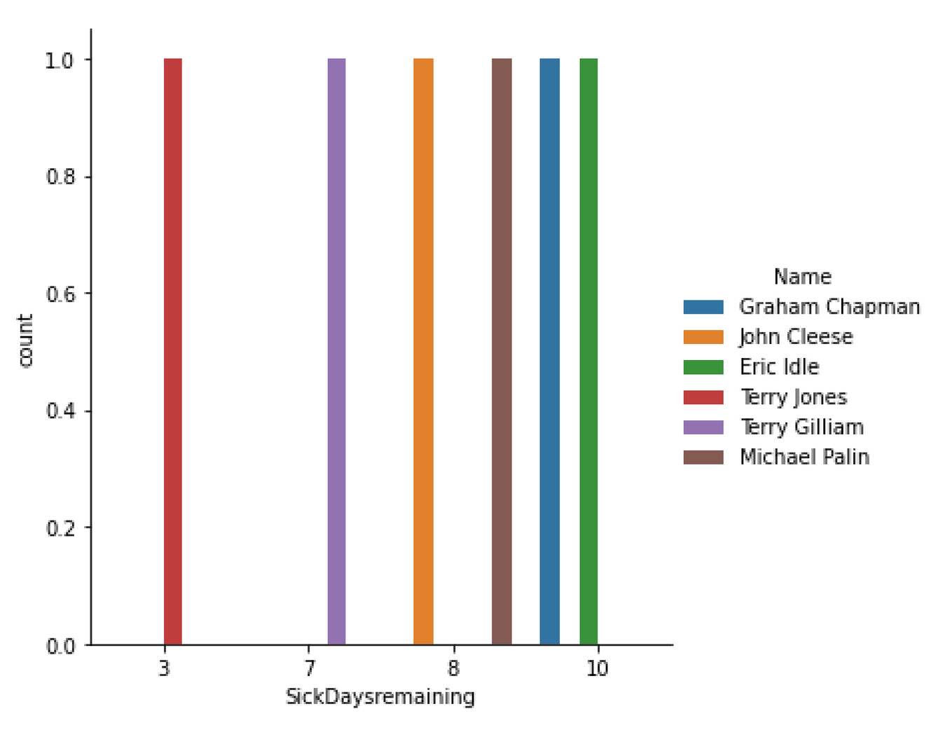
Out[10]:
```
Name                 0
HireDate             0
Salary               0
SickDaysremaining    0
dtype: int64
```

We can see that no missing values exist in dataset

Graphical Analysis

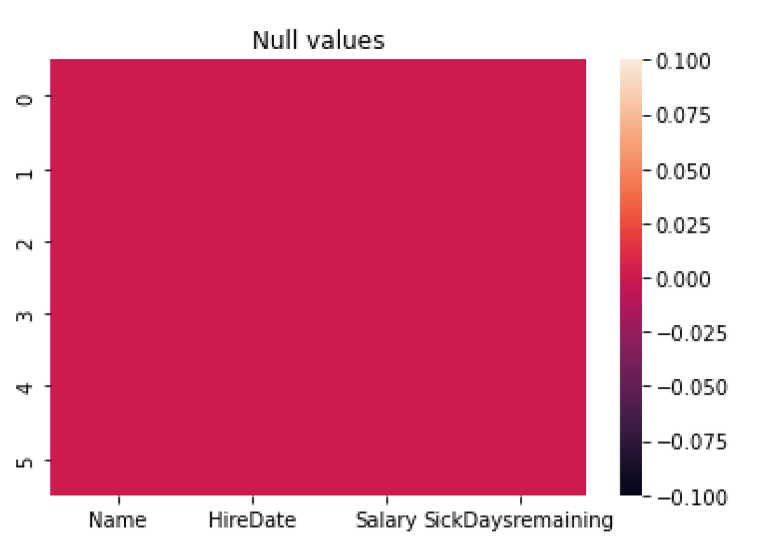
In [22]:
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.catplot(x ="SickDaysremaining", hue ="Name",
kind ="count", data = df)
```

Out[22]:
```
<seaborn.axisgrid.FacetGrid at 0x281f83bbf10>
```



In [29]:
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.heatmap(df.isnull())
plt.title('Null values')
plt.show
```

Out[29]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



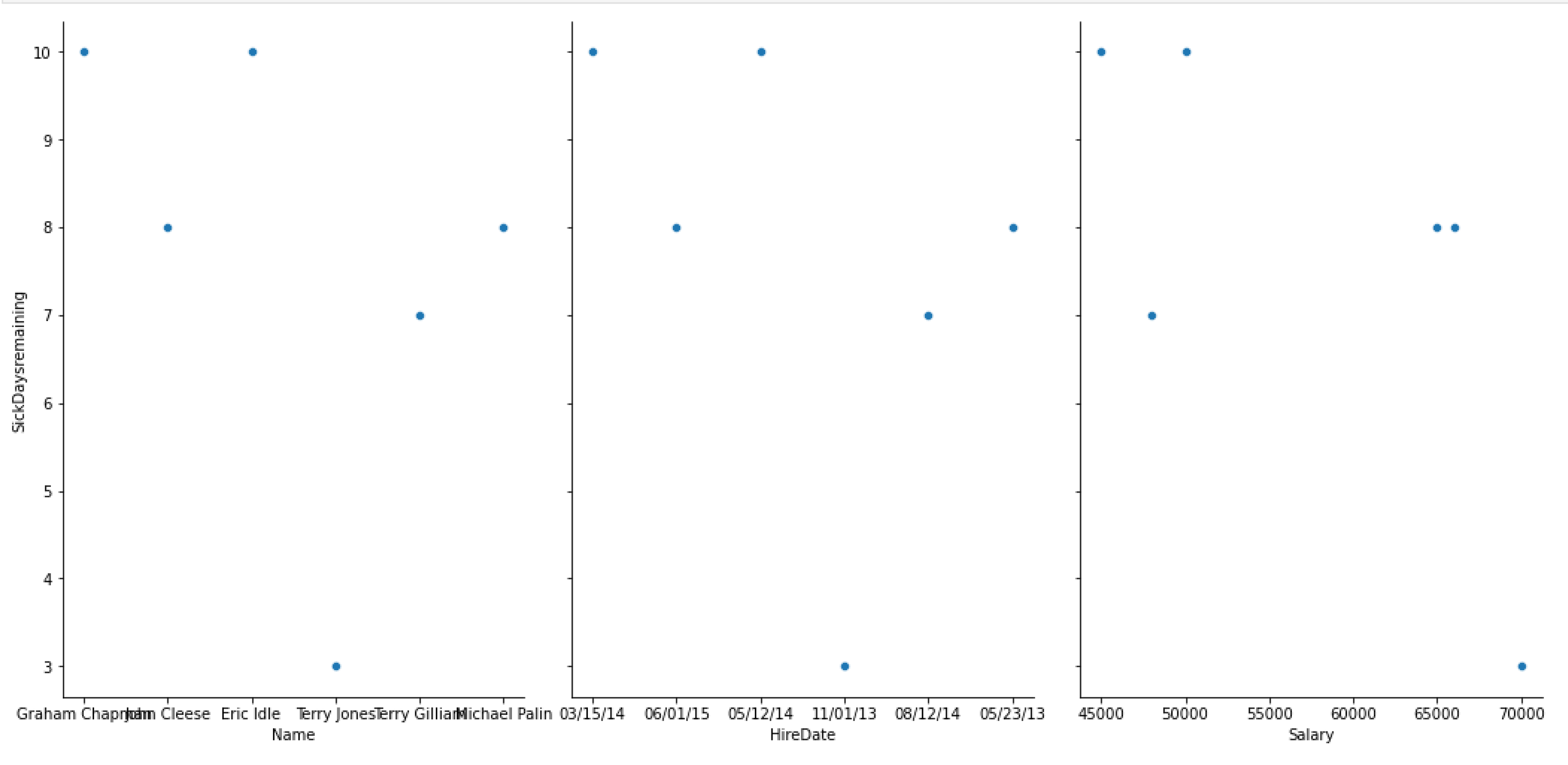
In [30]:
```
corr_matrix=df.corr()
corr_matrix
```

Out[30]:

	Salary	SickDaysremaining
Salary	1.000000	-0.667045
SickDaysremaining	-0.667045	1.000000

Relationship between Features and Response

In [39]:
```
import seaborn as sns
sns.pairplot(df, x_vars=['Name','HireDate','Salary'], y_vars='SickDaysremaining', height=7, aspect=0.7);
```



In [ ]:

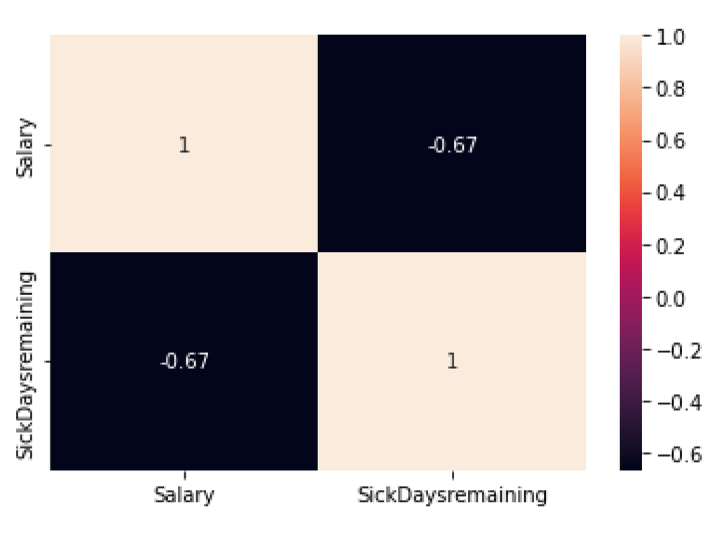
In [ ]:

In [ ]:

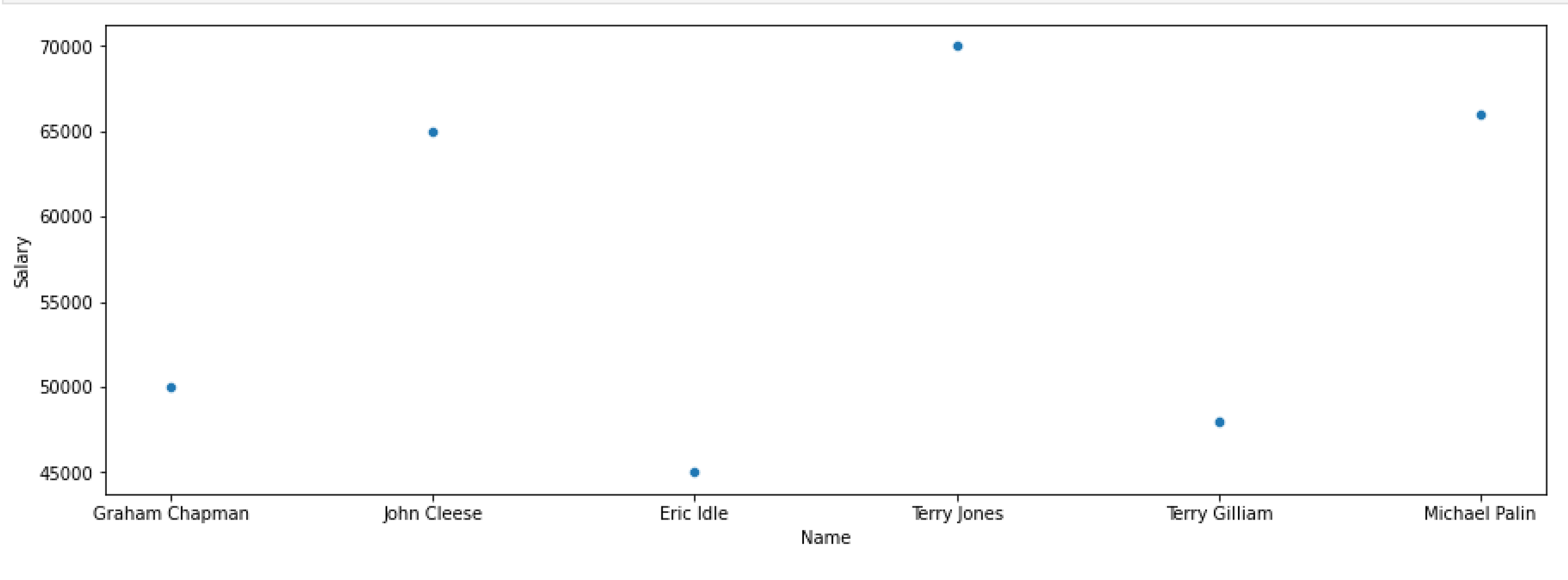
In [ ]:

In [40]:
```
sns.heatmap(df.corr(), annot=True)
```

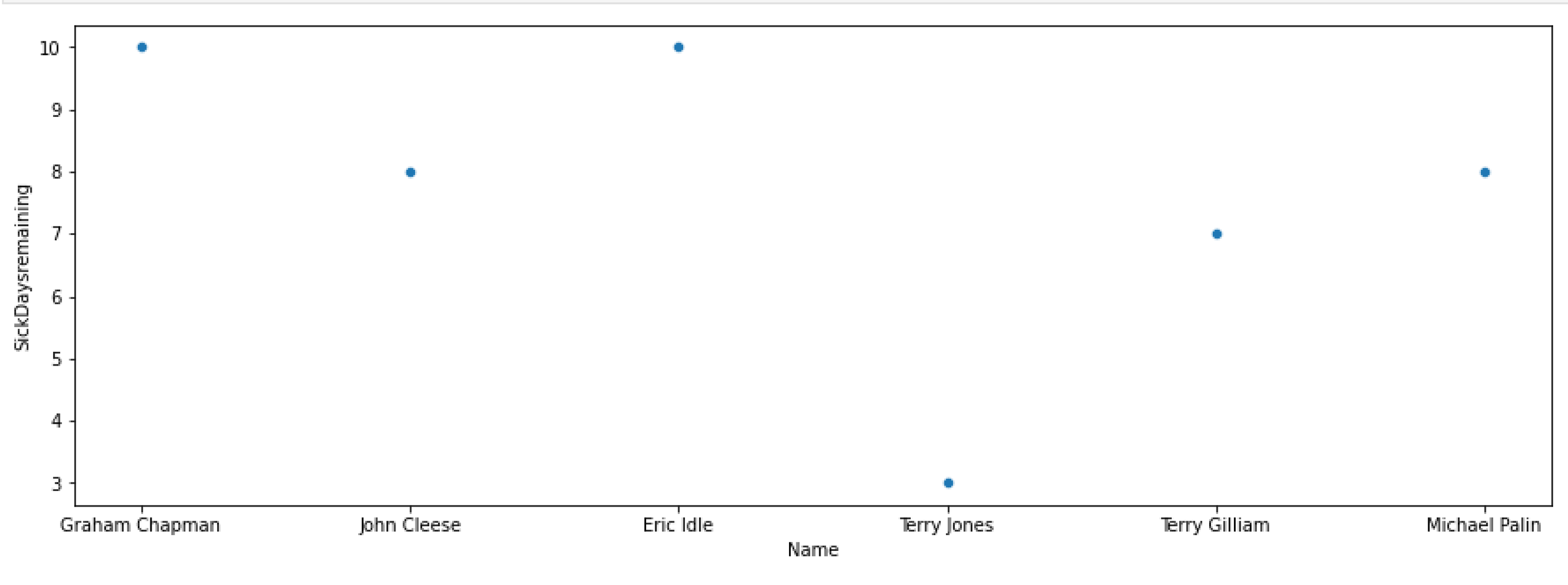
Out[40]:
```
<AxesSubplot:>
```



In [41]:
```
plt.figure(figsize=(15,5))
sns.scatterplot(x = df['Name'], y = df['Salary'])
plt.show()
```



In [42]:
```
plt.figure(figsize=(15,5))
sns.scatterplot(x = df['Name'], y = df['SickDaysremaining'])
plt.show()
```



Visualization of the Data

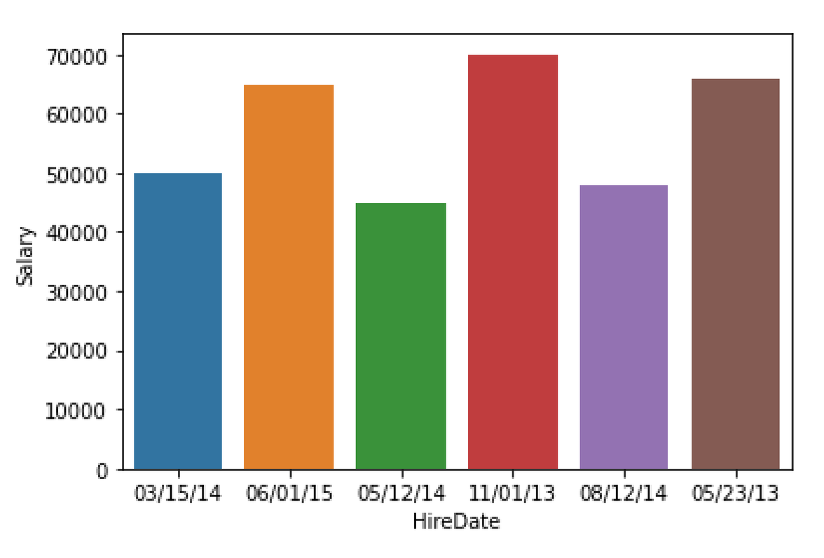
Making DataFrame of the orignal data

In [54]:
```
df_visualization_nominal=df[['Name', 'HireDate', 'Salary', 'SickDaysremaining']].copy()
df_visualization_nominal.columns
```

Out[54]:
```
Index(['Name', 'HireDate', 'Salary', 'SickDaysremaining'], dtype='object')
```

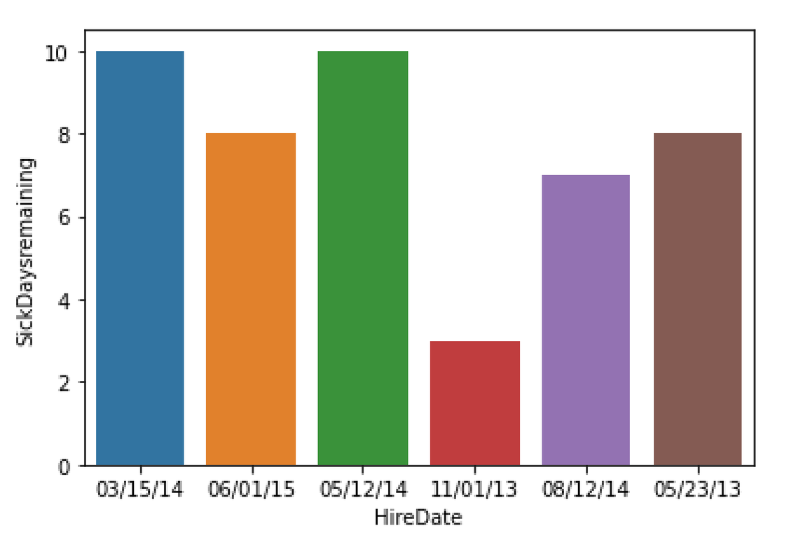
In [55]:
```
sns.barplot(x="HireDate",y="Salary",data=df)
```

Out[55]:
```
<AxesSubplot:xlabel='HireDate', ylabel='Salary'>
```



In [58]:
```
sns.barplot(x="HireDate",y="SickDaysremaining",data=df)
```

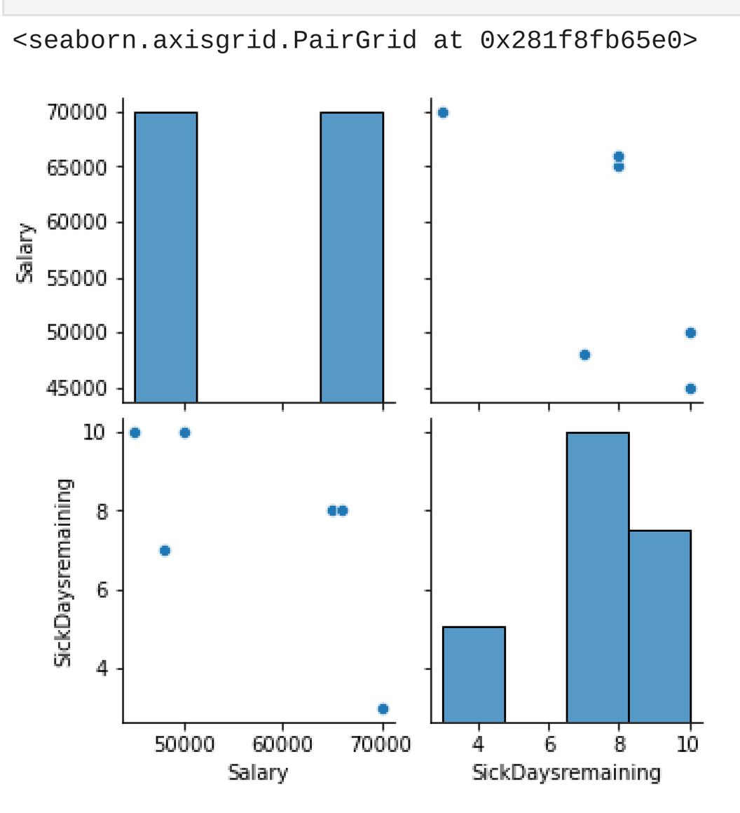
Out[58]:
```
<AxesSubplot:xlabel='HireDate', ylabel='SickDaysremaining'>
```



EDA

In [61]:
```
import seaborn as sns
sns.pairplot(df)
```

Out[61]:
```
<seaborn.axisgrid.PairGrid at 0x281f8fb65e0>
```



Conclusion

The highest level of the HR analytics maturity model is defined by making predictions. HR departments functioning at Level 4 are gathering data and using it not only to predict what will happen in the future, but also to plan for it. An example operations is "using turnover, promotion, and market data to model scenarios that help with workforce planning,"

And the given data is used for future employees progress and using some models and library we come to know the current status of the Dataset information

In [ ]:
```
THANK YOU
```