DATE:31-5-23_____**RELATIONSHIP B/W SALINITY & WATER TEMP.(ML)**

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing,svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

->READ THE DATA SET

```
df=pd.read_csv("/content/bottle.csv")
df
```

Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2S

```
df=df[['Salnty','T_degC']]
```

TAKING ONLY THE SELECTED TWO ATTRIBUTES FROM DATASET

```
df.columns=['Sal','Temp']
```

```
df.head(10)
```

	Sal	Temp
0	33.440	10.50
1	33.440	10.46
2	33.437	10.46
3	33.420	10.45
4	33.421	10.45
5	33.431	10.45
6	33.440	10.45
7	33.424	10.24
8	33.420	10.06
9	33.494	9.86

```
df.isna().any()
```

```
Sal     True
Temp    True
dtype: bool
```

```
df
```

	Sal	Temp
0	33.440	10.50
1	33.440	10.46
2	33.437	10.46
3	33.420	10.45
4	33.421	10.45
...	...	...
7975	33.609	11.92
7976	33.600	11.04
7977	33.647	10.71
7978	33.930	9.42
7979	33.964	9.22

7980 rows × 2 columns

```
df1=df.dropna()
```

```
df1
```

	Sal	Temp
0	33.440	10.50
1	33.440	10.46
2	33.437	10.46
3	33.420	10.45
4	33.421	10.45
...	...	...

```
fill_null=["Sal","Temp"]
for column in fill_null:
  mean=df[column].mean()
  df[column].fillna(mean,inplace=True)
```

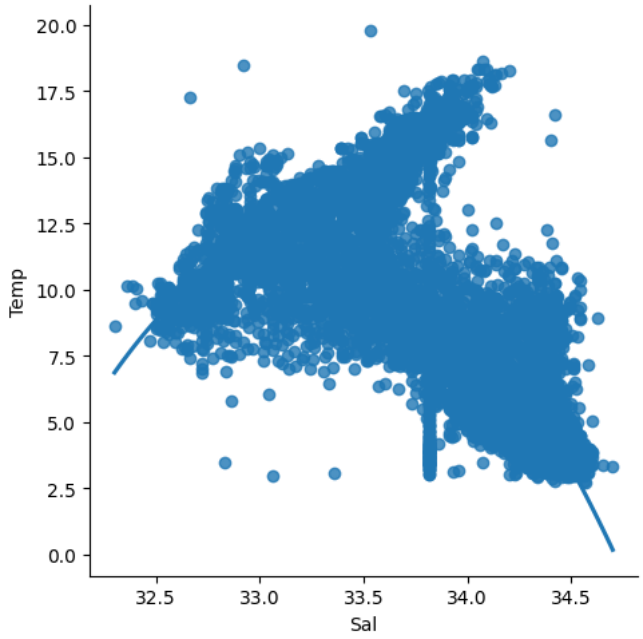
```
<ipython-input-33-85095d992cc0>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
  df[column].fillna(mean,inplace=True)
```

```
df.isna().any()
```

```
Sal     False
Temp    False
dtype: bool
```

```
sns.lmplot(x='Sal',y='Temp',data=df,order=2,ci=None)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f9346e2fee0>
```




```
df.describe()
```

	Sal	Temp
count	7980.000000	7980.000000
mean	33.812703	8.869087
std	0.512919	3.915512
min	32.300000	2.700000
25%	33.492000	5.350000
50%	33.880000	8.420000
75%	34.251000	11.920000
max	34.700000	19.760000

```
df.fillna(method='ffill',inplace=True)
```

```
<ipython-input-37-ae2c85fc64fc>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
  df.fillna(method='ffill',inplace=True)
```

```
x=np.array(df['Sal']).reshape(-1,1)
y=np.array(df['Temp']).reshape(-1,1)
```

```
df.dropna(inplace=True)
```

```
<ipython-input-39-c64f9f573c18>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus
  df.dropna(inplace=True)
```

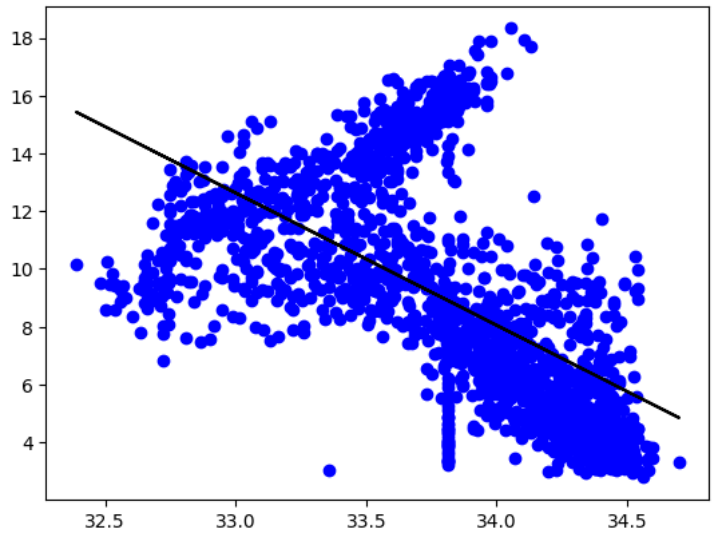
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
```

```
regr=LinearRegression()
regr.fit(x_train,y_train)
print(regr.score(x_test,y_test))
```

```
0.39274942375956456
```

```
y_pred=regr.predict(x_test)
```

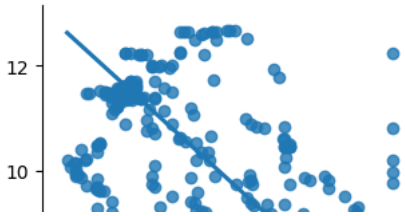
```
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```




```
df500=df[:][:500]
```

```
sns.lmplot(x="Sal",y="Temp",data=df500,order=1,ci=None)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f933ffbb400>
```

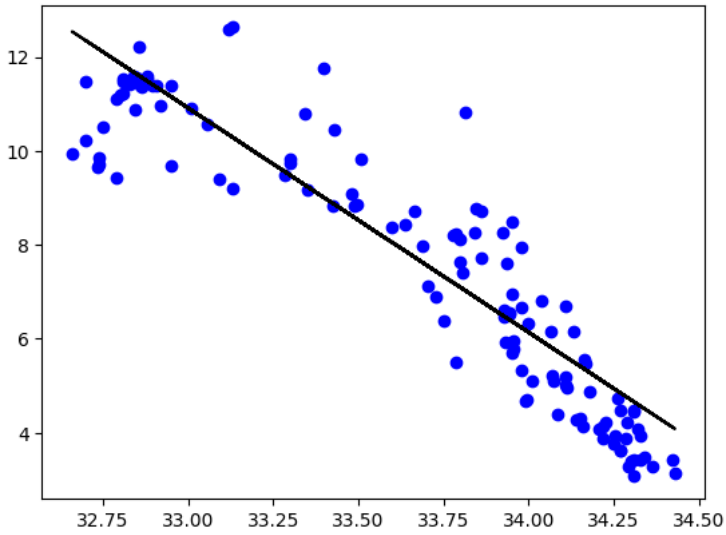



```
df500.fillna(method='ffill',inplace=True)
```

```
x=np.array(df500['Sal']).reshape(-1,1)
y=np.array(df500['Temp']).reshape(-1,1)
df500.dropna(inplace=True)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
regr=LinearRegression()
regr.fit(x_train,y_train)
print("Regression:",regr.score(x_test,y_test))
```

```
Regression: 0.8431299742395267
```

```
y_pred=regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```




```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

```
model=LinearRegression()
model.fit(x_train,y_train)
```

```
▾ LinearRegression
LinearRegression()
```