In [2]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt,seaborn as sns
```

In [3]:
```
df=pd.read_csv(r"C:\Users\pappu\Downloads\Salnity.csv")
df
```

```
C:\Users\pappu\AppData\Local\Temp\ipykernel_9212\1810579436.py:1: DtypeWarnin
g: Columns (47,73) have mixed types. Specify dtype option on import or set lo
w_memory=False.
  df=pd.read_csv(r"C:\Users\pappu\Downloads\Salnity.csv")
```

Out[3]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2S
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900	Na
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600	Na
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400	Na
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300	Na
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300	Na
...	...	...	...	...	...	...	...	...	...	
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	108.
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	108.
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	108.
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2S
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.

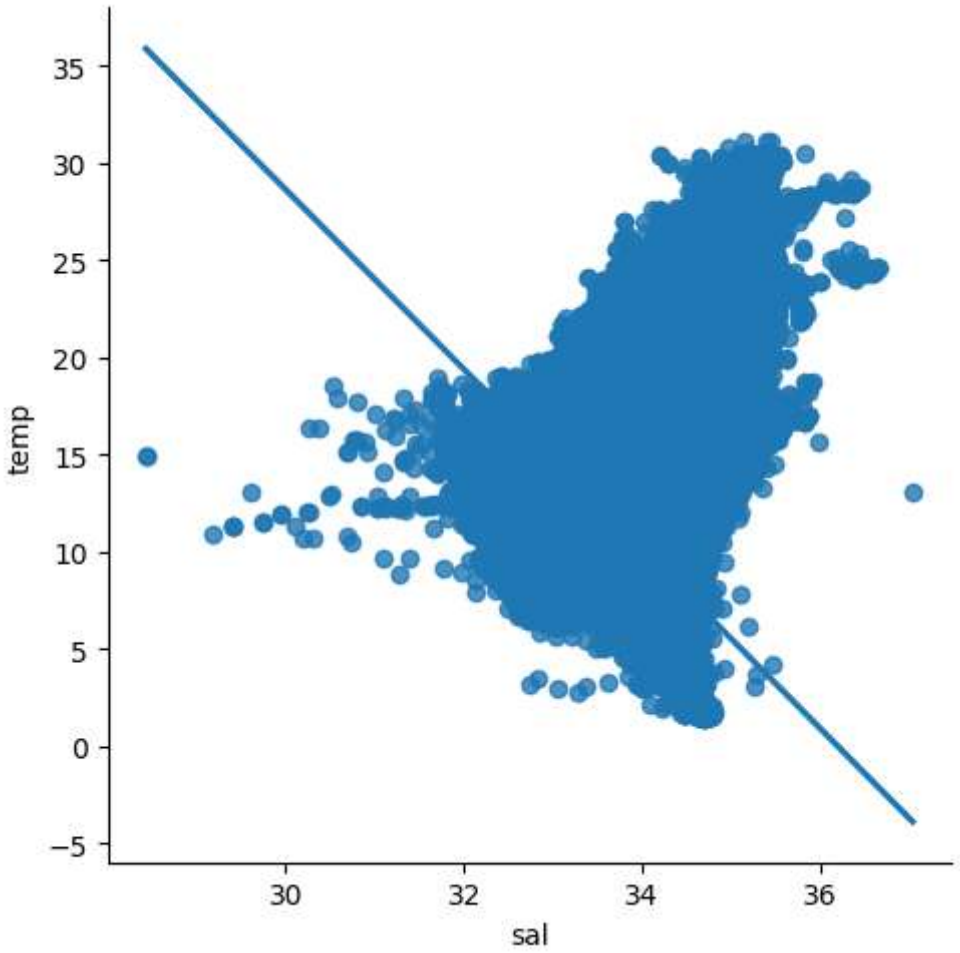
864863 rows × 74 columns

In [4]:
```
df=df[['Salnty','T_degC']]
```

In [5]:
```
df.columns=['sal','temp']
```

In [7]:
```
#plt.scatter(df['sal'],df['temp'])
sns.lmplot(x='sal',y='temp',data=df)
```

Out[7]: <seaborn.axisgrid.FacetGrid at 0x275a2af7460>



In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 864863 entries, 0 to 864862
Data columns (total 2 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   sal     817509 non-null  float64
 1   temp    853900 non-null  float64
dtypes: float64(2)
memory usage: 13.2 MB
```

In [9]:
```
df.dropna()
```

Out[9]:

	sal	temp
0	33.4400	10.500
1	33.4400	10.460
2	33.4370	10.460
3	33.4200	10.450
4	33.4210	10.450
...	...	...
864858	33.4083	18.744
864859	33.4083	18.744
864860	33.4150	18.692
864861	33.4062	18.161
864862	33.3880	17.533

814247 rows × 2 columns

In [12]:
```
import numpy as np
```

In [14]:
```
df.dropna(inplace=True)
```

```
C:\Users\pappu\AppData\Local\Temp\ipykernel_9212\1379821321.py:1: SettingWith
CopyWarning:
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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  df.dropna(inplace=True)
```

In [21]:
```
x=np.array(df['sal']).reshape(-1,1)
y=np.array(df['temp']).reshape(-1,1)
```

In [22]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

In [23]:
```
from sklearn.linear_model import LinearRegression
```

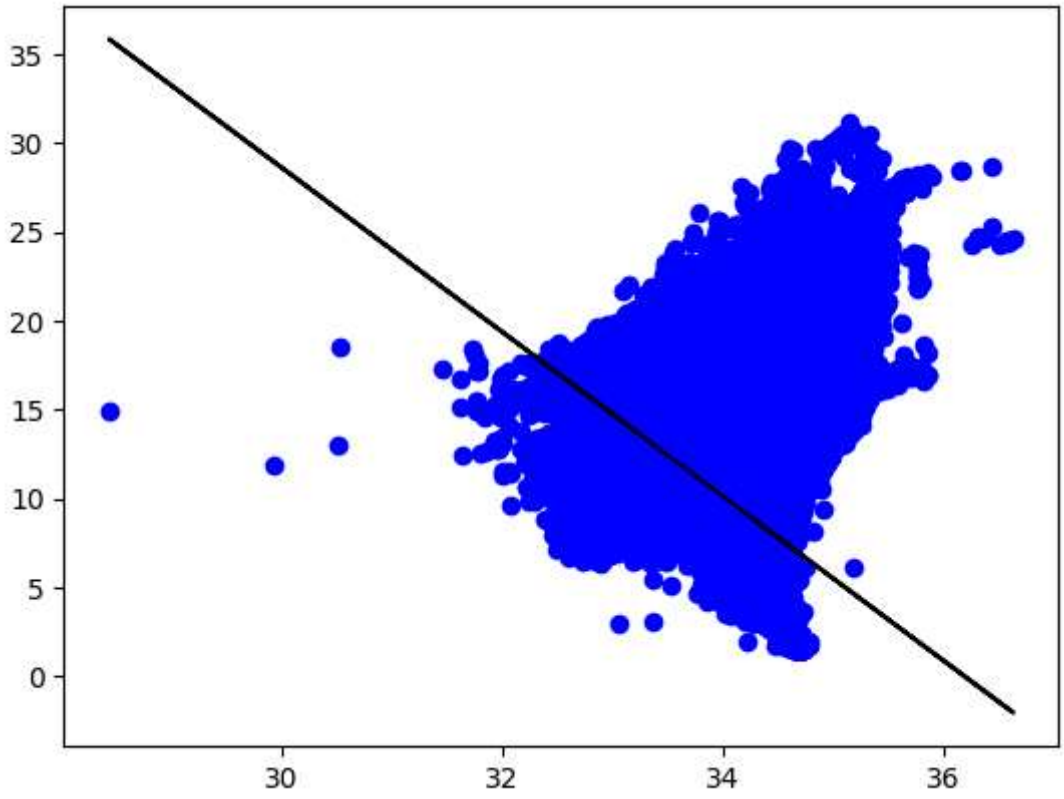
In [24]:
```
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[24]:
```
LinearRegression
LinearRegression()
```

In [25]:
```
print(lr.score(x_test,y_test))
```

```
0.2551517938515231
```

In [29]:
```
y_pred=lr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```



In [ ]:
```
```

In [ ]: