In [6]:

```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\prathyusha\appdata\loca
l\programs\python\python310\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\prathyusha
\appdata\local\programs\python\python310\lib\site-packages (from seaborn)
(1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\users\prathyusha\appdata
\local\programs\python\python310\lib\site-packages (from seaborn) (2.0.1)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\prathyu
sha\appdata\local\programs\python\python310\lib\site-packages (from seabor
n) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\prathyusha\app
data\local\programs\python\python310\lib\site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (1.0.7)
Requirement already satisfied: cycler>=0.10 in c:\users\prathyusha\appdata
\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.
1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\prathyusha\ap
pdata\local\programs\python\python310\lib\site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (4.39.4)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\prathyusha\ap
pdata\local\programs\python\python310\lib\site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\prathyusha\appd
ata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.
6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\prathyusha\appdat
a\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.
1,>=3.1->seaborn) (9.5.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\prathyusha\app
data\local\programs\python\python310\lib\site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\prathyusha
\appdata\local\programs\python\python310\lib\site-packages (from matplotli
b!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\prathyusha\appdata
\local\programs\python\python310\lib\site-packages (from pandas>=0.25->sea
born) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\prathyusha\appda
ta\local\programs\python\python310\lib\site-packages (from pandas>=0.25->s
eaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\prathyusha\appdata\loc
al\programs\python\python310\lib\site-packages (from python-dateutil>=2.7-
>matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [7]:

```
!pip install sklearn
```

```
Requirement already satisfied: sklearn in c:\users\prathyusha\appdata\loca
l\programs\python\python310\lib\site-packages (0.0.post5)
```

In [8]:

```
pip install matplotlip
```

```
Requirement already satisfied: matplotlip in c:\users\prathyusha\appdata\l
ocal\programs\python\python310\lib\site-packages (0.2)
Note: you may need to restart the kernel to use updated packages.
```

In [9]:

```
pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\users\prathyusha\appdata
\local\programs\python\python310\lib\site-packages (1.2.2)Note: you may ne
ed to restart the kernel to use updated packages.

Requirement already satisfied: numpy>=1.17.3 in c:\users\prathyusha\appdat
a\local\programs\python\python310\lib\site-packages (from scikit-learn)
(1.24.3)
Requirement already satisfied: scipy>=1.3.2 in c:\users\prathyusha\appdata
\local\programs\python\python310\lib\site-packages (from scikit-learn) (1.
10.1)
Requirement already satisfied: joblib>=1.1.1 in c:\users\prathyusha\appdat
a\local\programs\python\python310\lib\site-packages (from scikit-learn)
(1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\prathyusha
\appdata\local\programs\python\python310\lib\site-packages (from scikit-le
arn) (3.1.0)
```

In [10]:

```
pip install sklearn
```

```
Requirement already satisfied: sklearn in c:\users\prathyusha\appdata\loca
l\programs\python\python310\lib\site-packages (0.0.post5)Note: you may nee
d to restart the kernel to use updated packages.
```

In [11]:

```
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
```

In [12]:

```
df=pd.read_csv(r"C:\Users\Prathyusha\Downloads\bottle.csv")
df
```

```
C:\Users\Prathyusha\AppData\Local\Temp\ipykernel_10044\3580292764.py:1: Dt
ypeWarning: Columns (47,73) have mixed types. Specify dtype option on impo
rt or set low_memory=False.
  df=pd.read_csv(r"C:\Users\Prathyusha\Downloads\bottle.csv")
```

Out[12]:

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

In [13]:

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

In [14]:

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

Out[14]:

	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

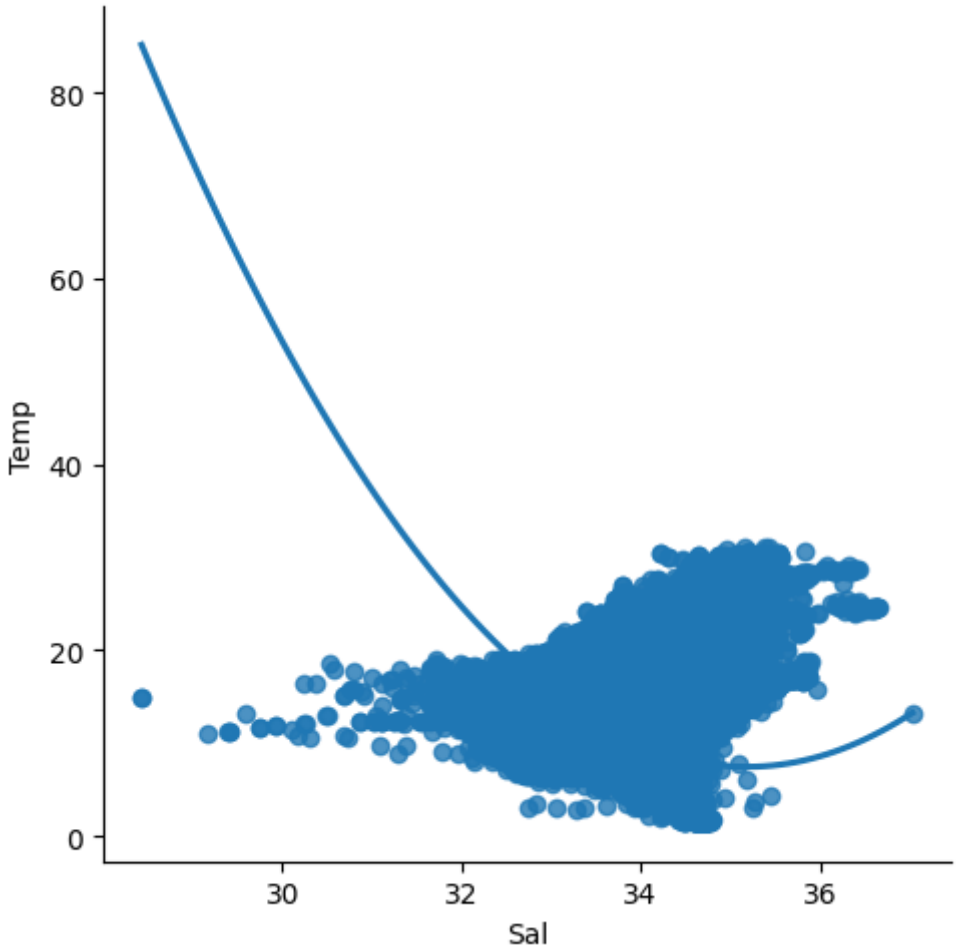
In [15]:

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

Out[15]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2230-09340264-	15	17.533	33.3880	5.774	24.15297

```
<seaborn.axisgrid.FacetGrid at 0x17e4caac220>
```



In [16]:

```
df.describe()
```

Out[16]:

	Sal	Temp
count	817509.000000	853900.000000
mean	33.840350	10.799677
std	0.461843	4.243825
min	28.431000	1.440000
25%	33.488000	7.680000
50%	33.863000	10.060000
75%	34.196900	13.880000
max	37.034000	31.140000

In [17]:

```
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 [18]:

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

```
C:\Users\Prathyusha\AppData\Local\Temp\ipykernel_10044\1844562654.py:1: Se
ttingWithCopyWarning:
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-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df.fillna(method="ffill",inplace=True)
```

In [19]:

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

In [20]:

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

```
C:\Users\Prathyusha\AppData\Local\Temp\ipykernel_10044\1379821321.py:1: Se
ttingWithCopyWarning:
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-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df.dropna(inplace=True)
```

In [13]:

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

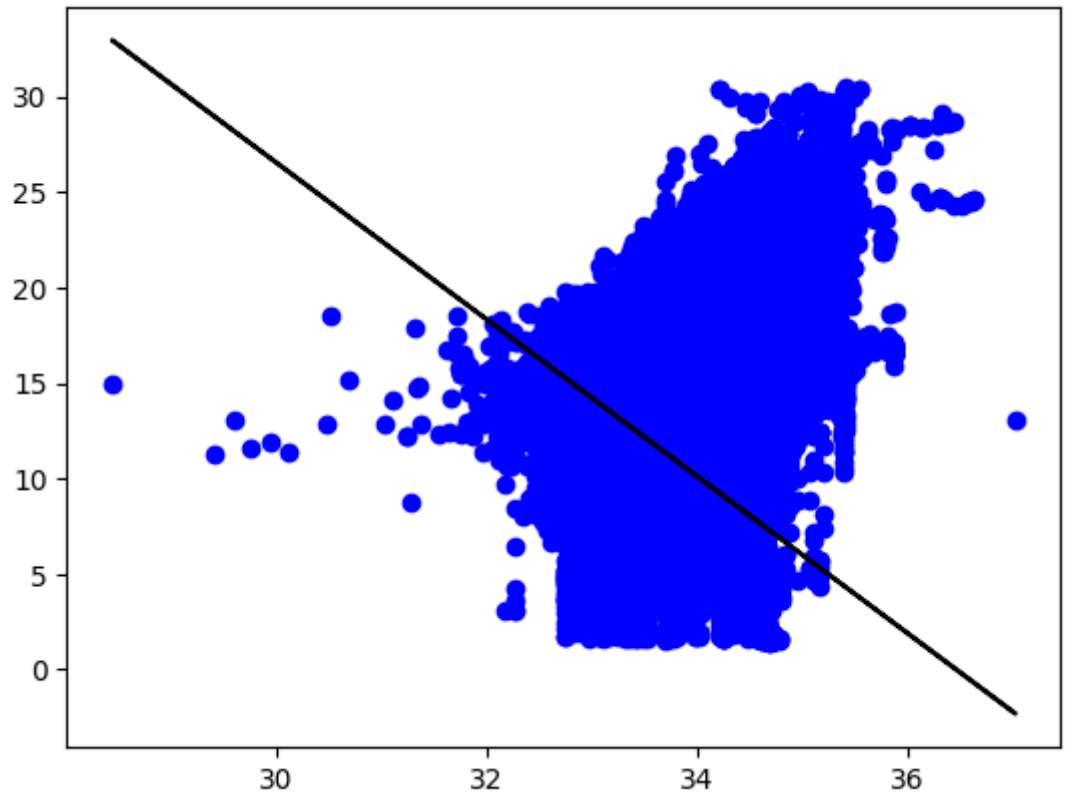
In [14]:

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

```
0.20596216315708216
```

In [15]:

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



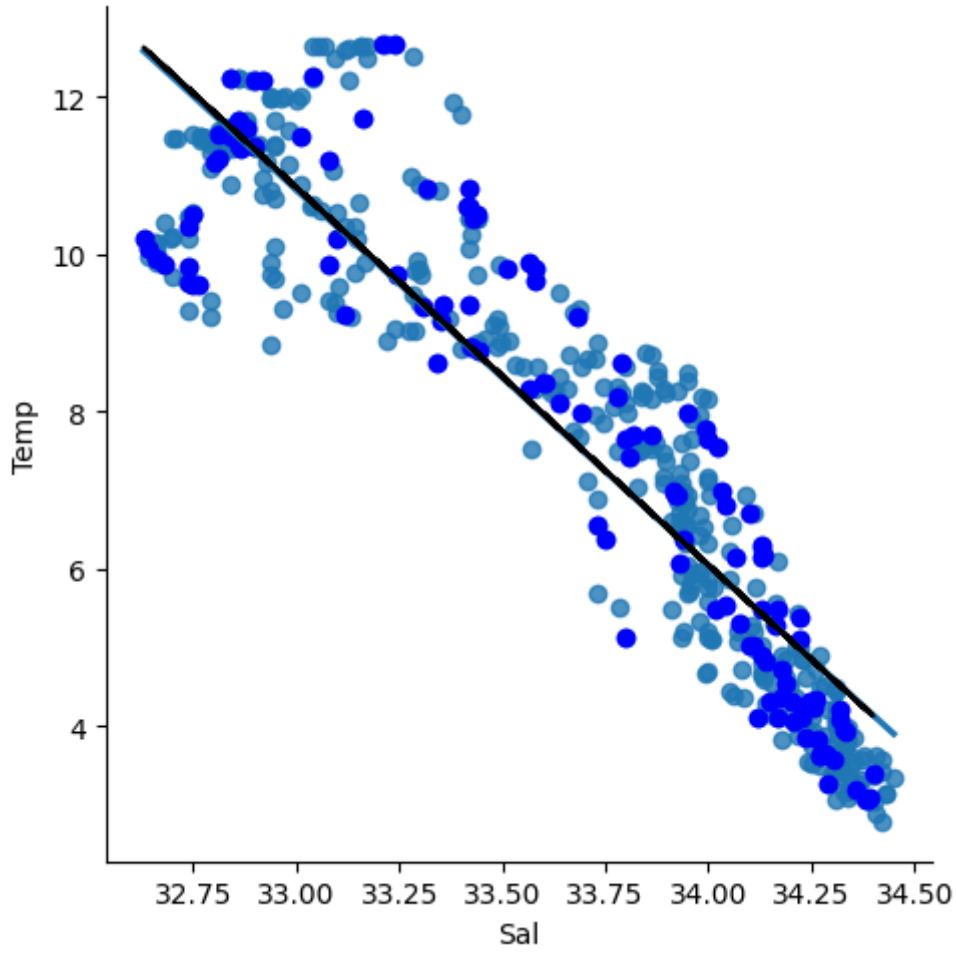
In [16]:

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

In [17]:

```
sns.lmplot(x="Sal",y="Temp",data=df500,order=1,ci=None)
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))
y_pred=regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```

```
Regression: 0.8311447195168511
```



In [24]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
model=LinearRegression()
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
r2=r2_score(y_test,y_pred)
print("R2 score:",r2)
```

```
R2 score: 0.8311447195168511
```

In [ ]:

```
conclusion:
    this model is fit.
```

In [ ]:

```

```