In [9]:

```
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
```

```
---------------------------------------------------------------------------
-
ModuleNotFoundError                       Traceback (most recent call las
t)
Cell In[9], line 5
      3 import seaborn as sns
      4 import matplotlib.pyplot as plt
----> 5 from sklearn import preprocessing ,svm
      6 from sklearn.model_selection import train_test_split
      7 from sklearn.linear_model import LinearRegression

ModuleNotFoundError: No module named 'sklearn'
```

In [10]:

```
pip install scikit-learn
```

```
     -----                                    1.1/8.3 MB 425.8 kB/s eta
0:00:17
     -----                                    1.1/8.3 MB 425.8 kB/s eta
0:00:17
     -----                                    1.1/8.3 MB 418.9 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.9 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.2 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.2 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.2 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.2 kB/s eta
0:00:18
     -----                                    1.1/8.3 MB 418.2 kB/s eta
0:00:18
     ------                                   1.3/8.3 MB 435.8 kB/s eta
0:00:16
```

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\91628\Downloads\bottle.csv.zip")
df
```

```
C:\Users\91628\AppData\Local\Temp\ipykernel_8584\2831184167.py:1: DtypeWar
ning: Columns (47,73) have mixed types. Specify dtype option on import or
set low_memory=False.
  df=pd.read_csv(r"C:\Users\91628\Downloads\bottle.csv.zip")
```

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 [17]:

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

In [18]:

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

Out[18]:

	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

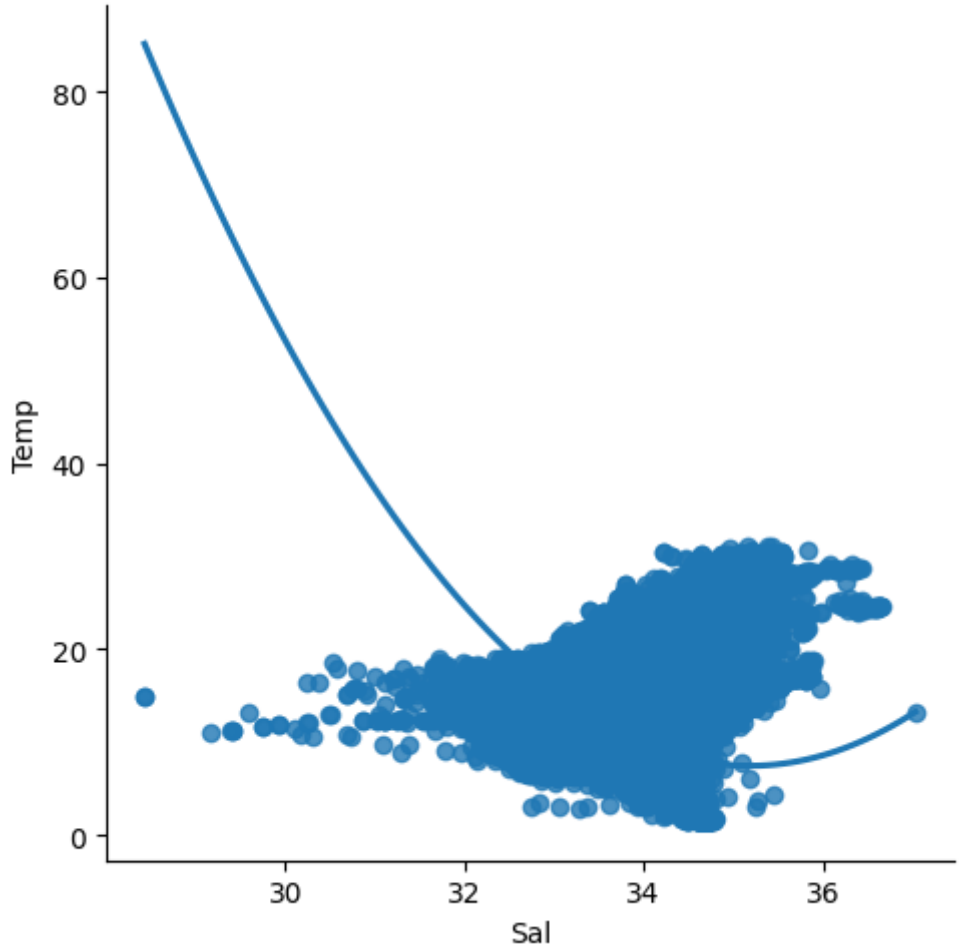
In [29]:

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

Out[29]:

	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-310-2339-09340264-	15	17.533	33.3880	5.774	24.15297

```
<seaborn.axisgrid.FacetGrid at 0x2a294ad3690>
```



In [30]:

```
df.describe()
```

Out[30]:

	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 [31]:

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

```
C:\Users\91628\AppData\Local\Temp\ipykernel_8584\4116506308.py:1: SettingW
ithCopyWarning:
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 [39]:

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

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

In [40]:

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

```
C:\Users\91628\AppData\Local\Temp\ipykernel_8584\1379821321.py:1: SettingW
ithCopyWarning:
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)
```

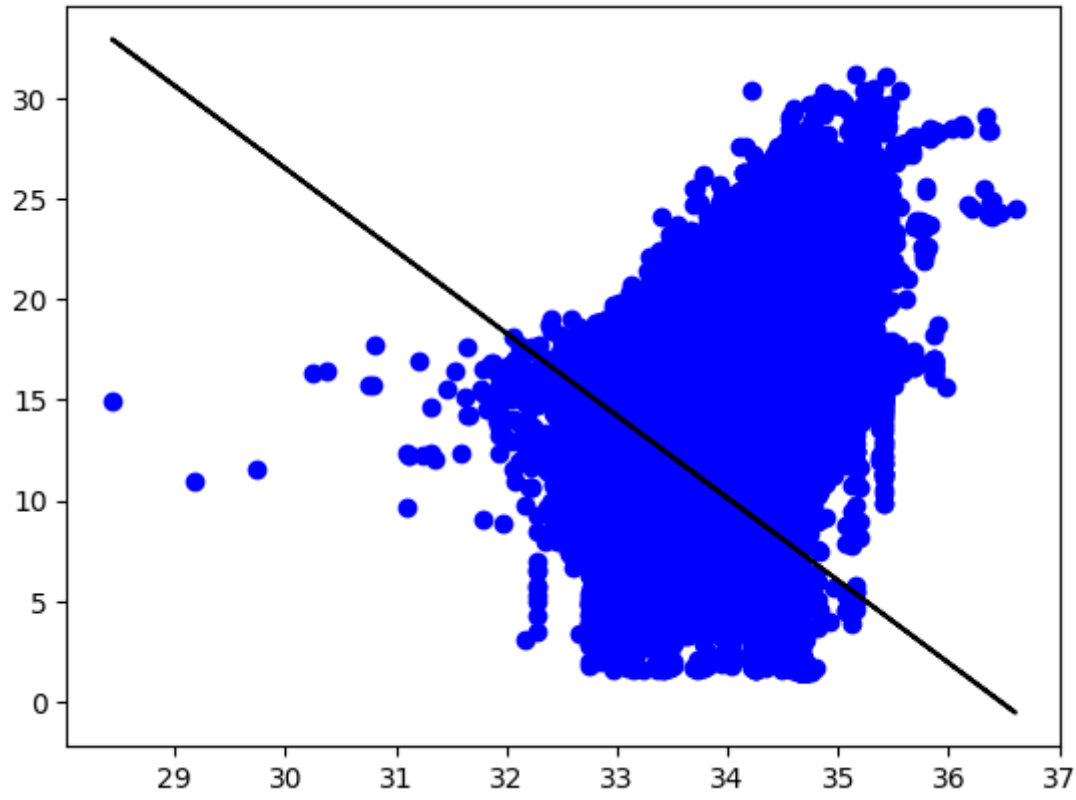
In [38]:

```
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))
```

```
0.205469915517976
```

In [42]:

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

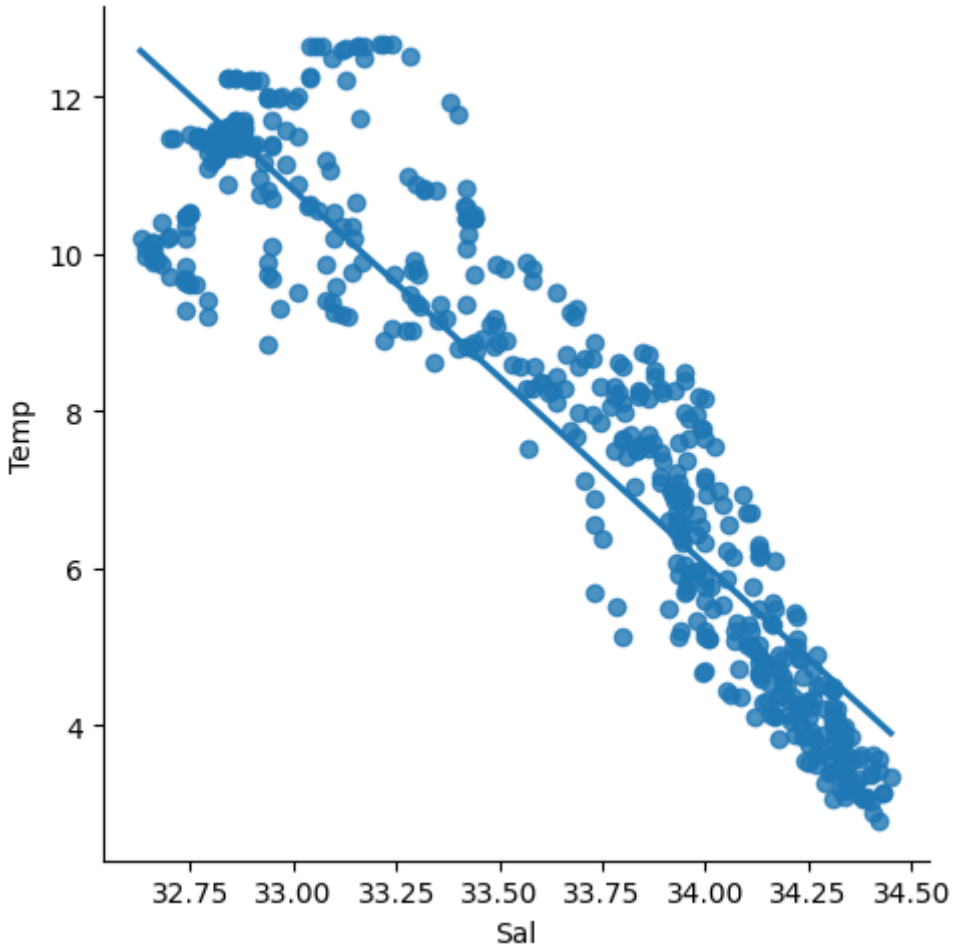


In [43]:

```
df500=df[:][:500]
sns.lmplot(x="Sal",y="Temp",data=df500,order=1,ci=None)
```

Out[43]:

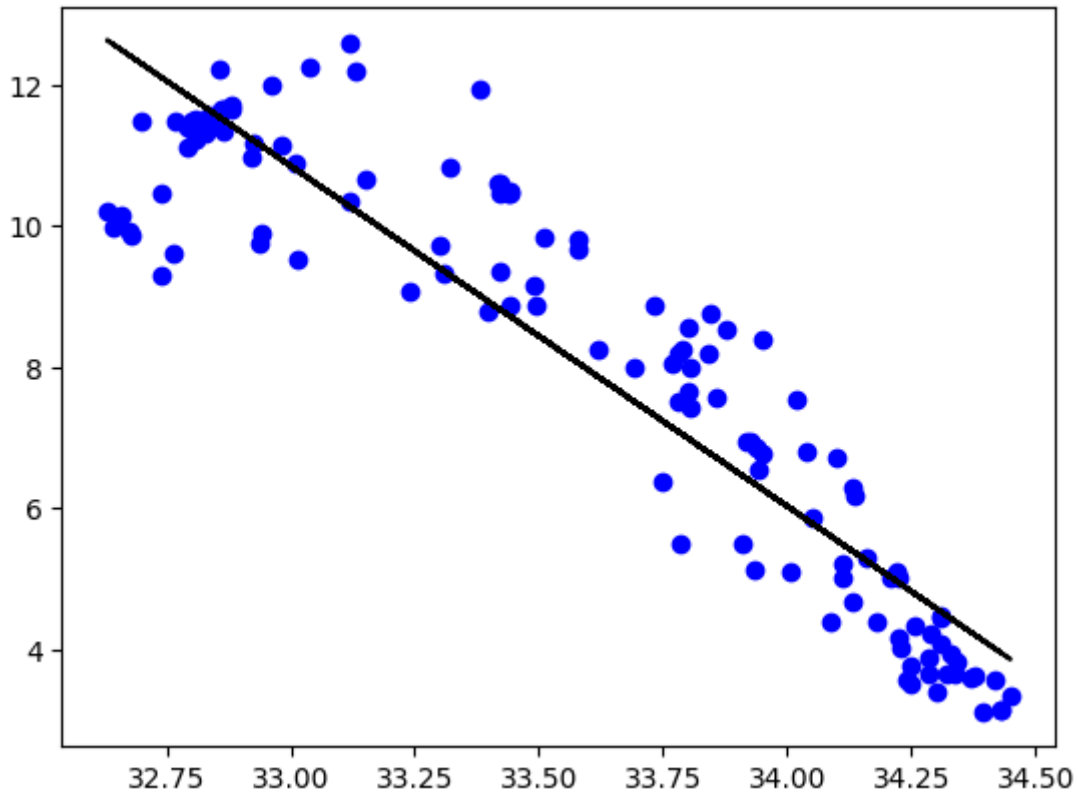
```
<seaborn.axisgrid.FacetGrid at 0x2a294c8d490>
```



In [46]:

```
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: 0.8464423428987112
```



In [3]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
mode1=LinearRegression()
mode1.fit(x_train,y_train)
y_pred=mode1.predict(x_test)
r2=r2_score(y_test,y_pred)
print("R2 score:",r2)
```

```
---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
Cell In[3], line 4
      2 from sklearn.metrics import r2_score
      3 mode1=LinearRegression()
----> 4 mode1.fit(x_train,y_train)
      5 y_pred=mode1.predict(x_test)
      6 r2=r2_score(y_test,y_pred)

NameError: name 'x_train' is not defined
```

In [ ]: