

```
In [42]: #step-1:problem statement
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 [43]: #step-2:reading the data set
df=pd.read_csv(r"C:\Users\chinta pavani\Downloads\bottle.csv.zip")
df
```

C:\Users\chinta pavani\AppData\Local\Temp\ipykernel_25960\685022416.py:1:
DtypeWarning: Columns (47,73) have mixed types. Specify dtype option on i
mport or set low_memory=False.

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

Out[43]:

	Cst_Cnt	Btl_Cnt	Sta_ID		Depth_ID	Depthm	T_degC	Salnty
0	1	1	054.0 056.0	19-4903CR- HY-060-0930-05400560-0000A-3		0	10.500	33.4400
1	1	2	054.0 056.0	19-4903CR- HY-060-0930-05400560-0008A-3		8	10.460	33.4400
2	1	3	054.0 056.0	19-4903CR- HY-060-0930-05400560-0010A-7		10	10.460	33.4370
3	1	4	054.0 056.0	19-4903CR- HY-060-0930-05400560-0019A-3		19	10.450	33.4200
4	1	5	054.0 056.0	19-4903CR- HY-060-0930-05400560-0020A-7		20	10.450	33.4210
...
864858	34404	864859	093.4 026.4	20-1611SR- MX-310-2239-09340264-0000A-7		0	18.744	33.4083
864859	34404	864860	093.4 026.4	20-1611SR- MX-310-2239-09340264-0002A-3		2	18.744	33.4083
864860	34404	864861	093.4 026.4	20-1611SR- MX-310-2239-09340264-0005A-3		5	18.692	33.4150
864861	34404	864862	093.4 026.4	20-1611SR- MX-310-2239-09340264-0010A-3		10	18.161	33.4062
864862	34404	864863	093.4 026.4	20-1611SR- MX-310-2239-09340264-0015A-3		15	17.533	33.3880

864863 rows × 74 columns

```
In [13]: df=df[['Salnty','T_degC']]
df.columns=['sal','Temp']
```

```
In [8]: df.head(10)
```

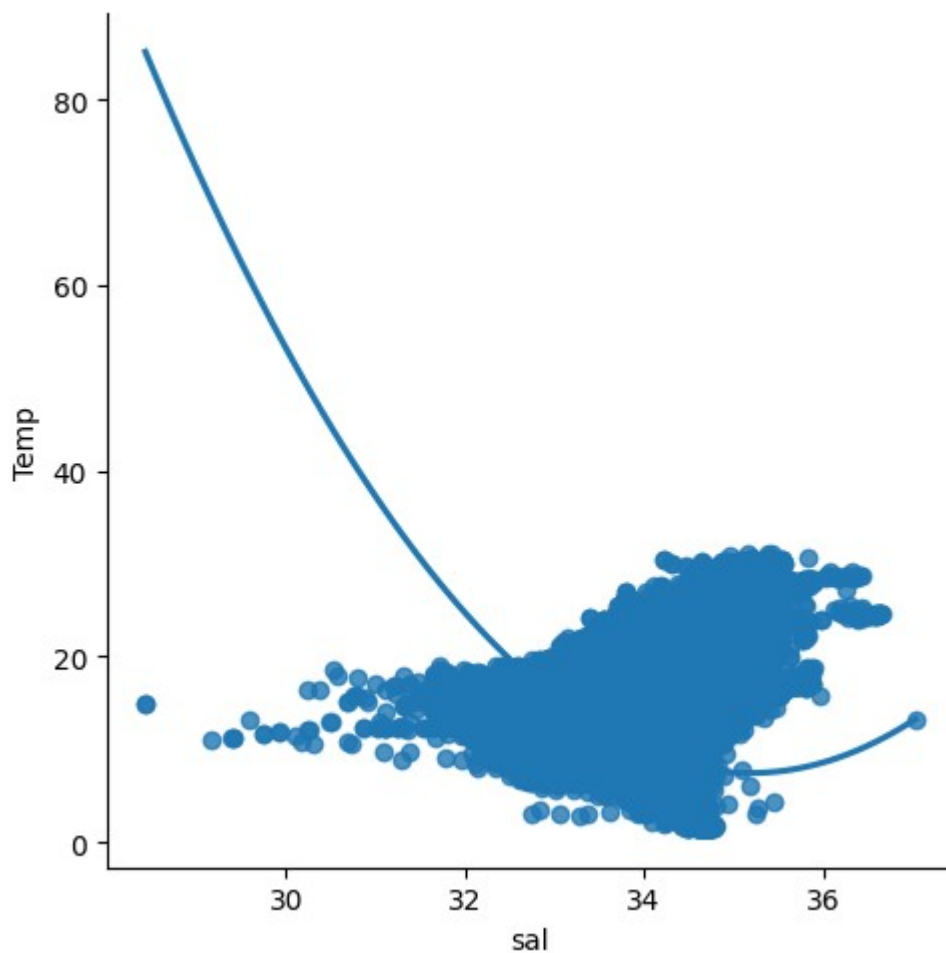
Out[8]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_
0	1	1	054.0 056.0	19-4903CR- HY-060-0930-05400560-0000A-3	0	10.50	33.440	Na
1	1	2	054.0 056.0	19-4903CR- HY-060-0930-05400560-0008A-3	8	10.46	33.440	Na
2	1	3	054.0 056.0	19-4903CR- HY-060-0930-05400560-0010A-7	10	10.46	33.437	Na
3	1	4	054.0 056.0	19-4903CR- HY-060-0930-05400560-0019A-3	19	10.45	33.420	Na
4	1	5	054.0 056.0	19-4903CR- HY-060-0930-05400560-0020A-7	20	10.45	33.421	Na
5	1	6	054.0 056.0	19-4903CR- HY-060-0930-05400560-0030A-7	30	10.45	33.431	Na
6	1	7	054.0 056.0	19-4903CR- HY-060-0930-05400560-0039A-3	39	10.45	33.440	Na
7	1	8	054.0 056.0	19-4903CR- HY-060-0930-05400560-0050A-7	50	10.24	33.424	Na
8	1	9	054.0 056.0	19-4903CR- HY-060-0930-05400560-0058A-3	58	10.06	33.420	Na
9	1	10	054.0 056.0	19-4903CR- HY-060-0930-05400560-0075A-7	75	9.86	33.494	Na

10 rows × 9 columns

```
In [18]: #step-3:exploring the data scatter-plotting the data scatter  
sns.lmplot(x="sal",y="Temp",data=df,order=2,ci=None)
```

Out[18]: <seaborn.axisgrid.FacetGrid at 0x1b928065110>



```
In [19]: df.describe()
```

Out[19]:

	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 [20]: ▶ 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 [21]: ▶ *#step-4: data cleaning-eliminating nan or missing input*
df.fillna(method='ffill', inplace=True)

C:\Users\chinta pavani\AppData\Local\Temp\ipykernel_25960\4116506308.py:
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-a-copy ([http s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni ng-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 [24]: ▶ *#step-5: training our model*
x=np.array(df['sal']).reshape(-1,1)
y=np.array(df['Temp']).reshape(-1,1)

In [25]: ▶ df.dropna(inplace=True)

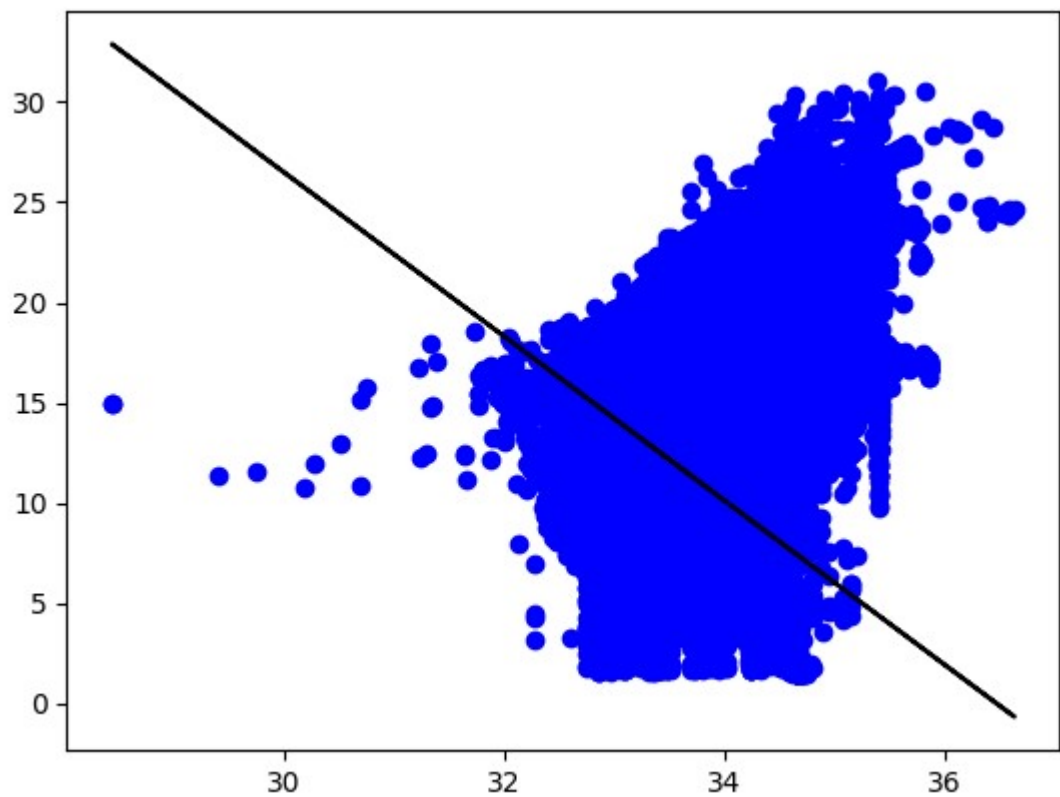
C:\Users\chinta pavani\AppData\Local\Temp\ipykernel_25960\1379821321.py:
1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy ([http s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni ng-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 [26]: ▶ 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))

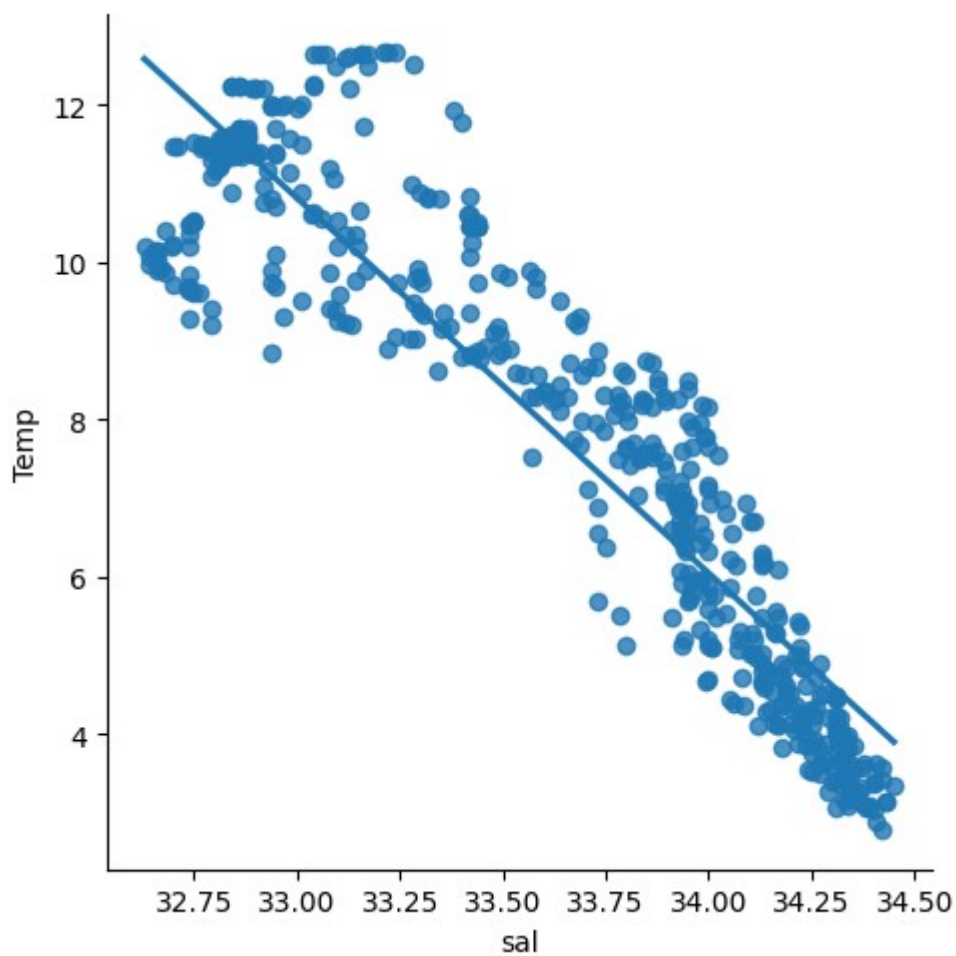
0.2072481345490751

```
In [27]: #step-6:exploring our results  
y_pred=regr.predict(x_test)  
plt.scatter(x_test,y_test,color='b')  
plt.plot(x_test,y_pred,color='k')  
plt.show()  
#data scatter of predicted values
```



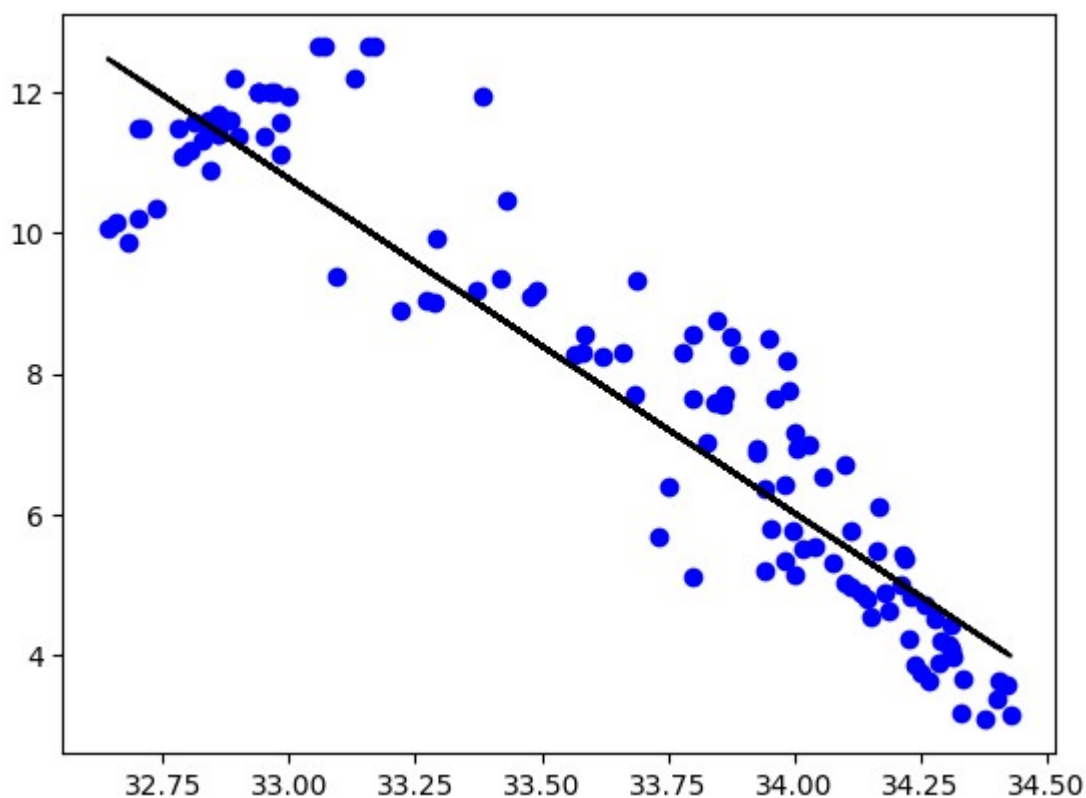
```
In [28]: #step-7:working with a smallest dataset  
df500=df[:][:500]  
sns.lmplot(x="sal",y="Temp",data=df500,order=1,ci=None)
```

Out[28]: <seaborn.axisgrid.FacetGrid at 0x1b92807eb90>



```
In [37]: ▶ 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: 0.8679243594971805



```
In [39]: ▶ #step-8:evaluation of model
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: 0.8679243594971805

```
In [ ]: ▶ #step-9:conclusion
#dataset we have taken is poor for linear model but with the smaller data i
```

