Problem Statement

Linear Regression

Import Libraries

In [30]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [31]:

```
a=pd.read_csv("iris.csv")
a
```

Out[31]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

To display top 10 rows

In [32]:

```
c=a.head(15)
c
```

Out[32]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa

To find Missing values

In [33]:

```
c.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             15 non-null     int64  
 1   SepalLengthCm  15 non-null     float64
 2   SepalWidthCm   15 non-null     float64
 3   PetalLengthCm  15 non-null     float64
 4   PetalWidthCm   15 non-null     float64
 5   Species        15 non-null     object 
dtypes: float64(4), int64(1), object(1)
memory usage: 848.0+ bytes
```

To display summary of statistics

In [34]:

```
a.describe()
```

Out[34]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

To display column heading

In [35]:

```
a.columns
```

Out[35]:

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

Pairplot

In [36]:

```
s=a.dropna(axis=1)
s
```

Out[36]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

In [38]:

```
s.columns
```

Out[38]:

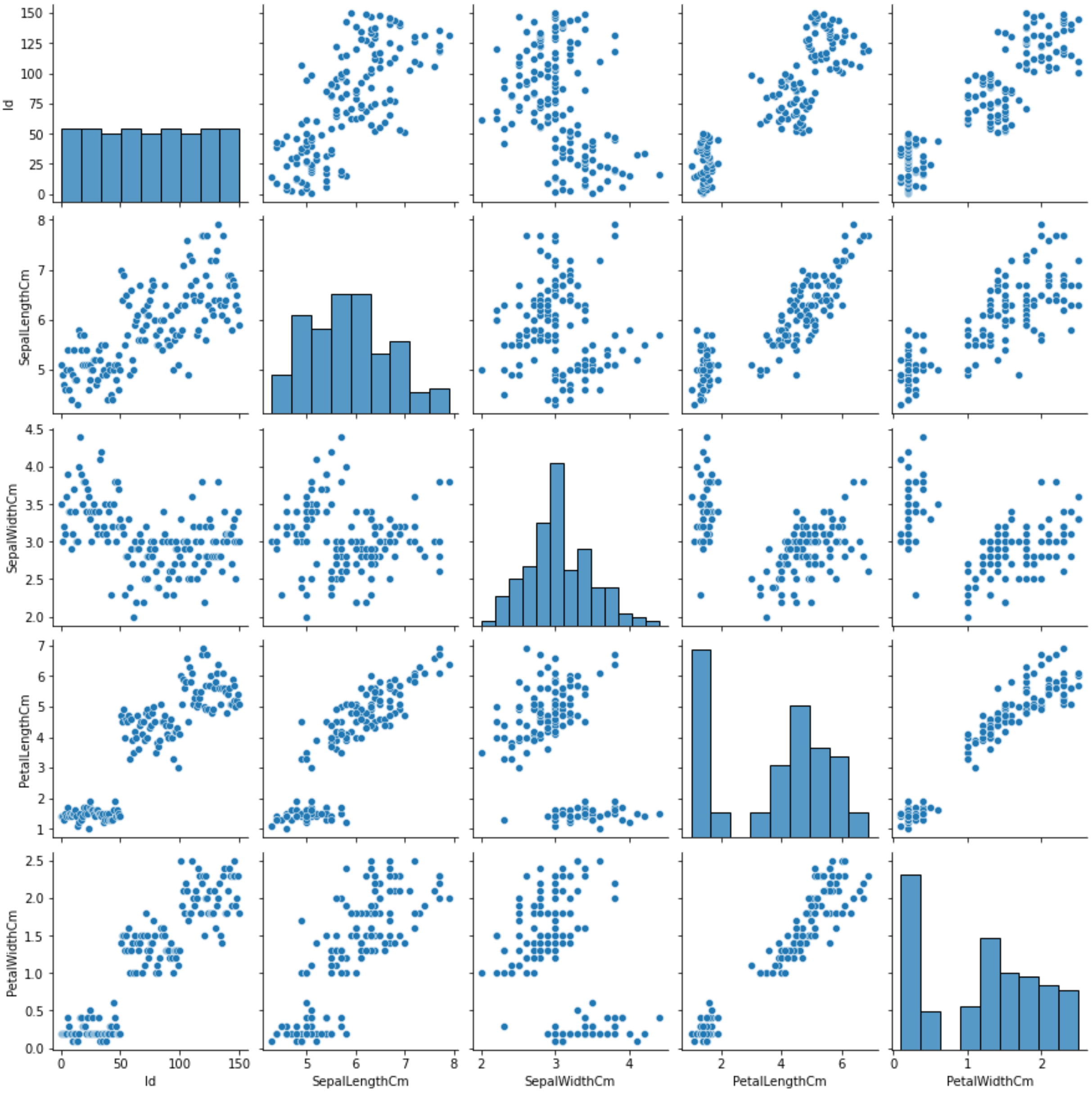
```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

In [39]:

```
sns.pairplot(a)
```

Out[39]:

```
<seaborn.axisgrid.PairGrid at 0x1a23f0ffbe0>
```



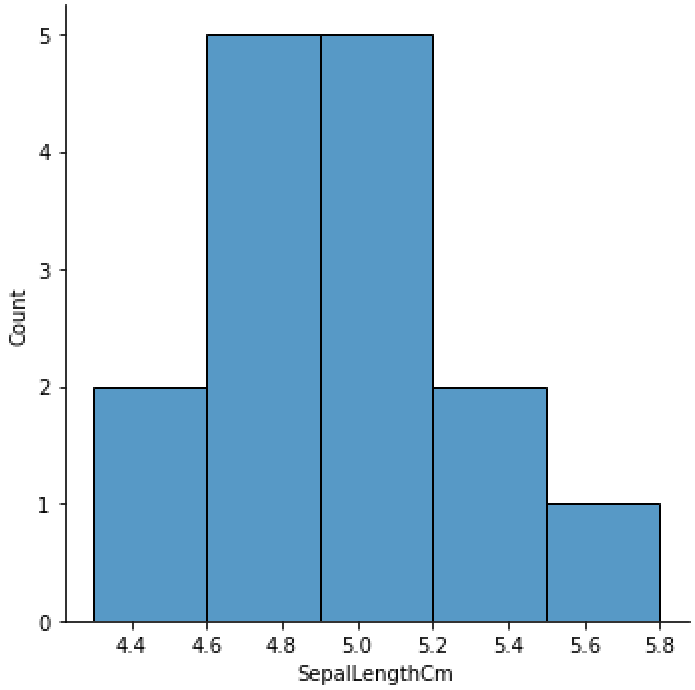
Distribution Plot

In [40]:

```
sns.displot(c['SepalLengthCm'])
```

Out[40]:

```
<seaborn.axisgrid.FacetGrid at 0x1a23f0d2c40>
```



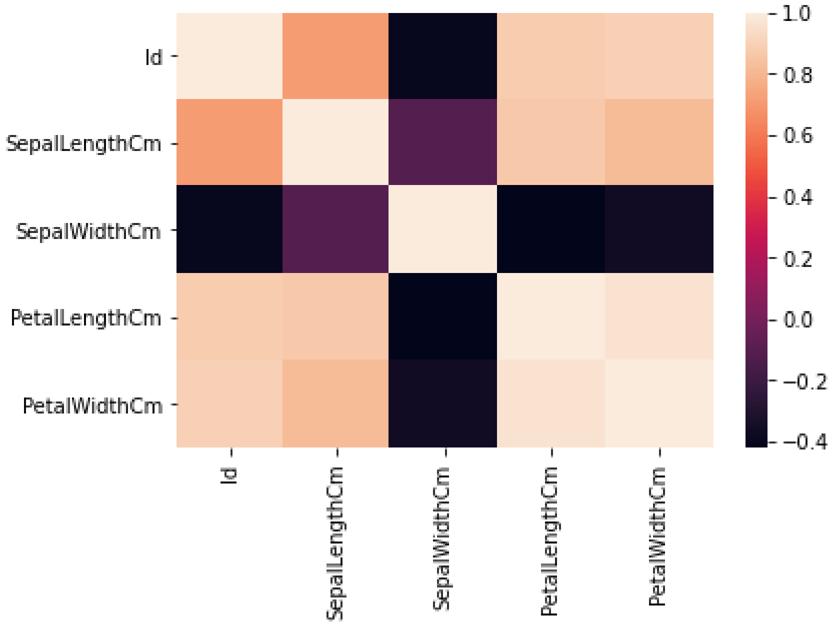
Correlation

In [41]:

```
b=a[['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species']]
sns.heatmap(b.corr())
```

Out[41]:

```
<AxesSubplot:>
```



Train the model - Model Building

In [42]:

```
g=c[['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm']]
h=c['PetalWidthCm']
```

To split dataset into training end test

In [43]:

```
from sklearn.model_selection import train_test_split
g_train,g_test,h_train,h_test=train_test_split(g,h,test_size=0.6)
```

To run the model

In [44]:

```
from sklearn.linear_model import LinearRegression
```

In [45]:

```
lr=LinearRegression()
lr.fit(g_train,h_train)
```

Out[45]:

```
LinearRegression()
```

In [46]:

```
print(lr.intercept_)
```

```
-1.675380878014923
```

Coeffecient

In [47]:

```
coeff=pd.DataFrame(lr.coef_,g.columns,columns=['Co-effecient'])
coeff
```

Out[47]:

	Co-effecient
Id	-0.007059
SepalLengthCm	0.221447
SepalWidthCm	-0.212875
PetalLengthCm	1.030678

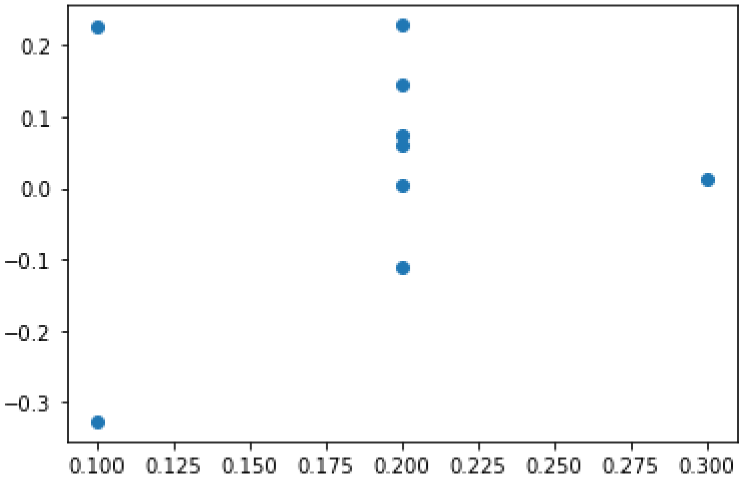
Best Fit line

In [48]:

```
prediction=lr.predict(g_test)
plt.scatter(h_test,prediction)
```

Out[48]:

```
<matplotlib.collections.PathCollection at 0x1a111ee2e50>
```



To find score

In [49]:

```
print(lr.score(g_test,h_test))
```

```
-14.762896693447123
```

Import Lasso and ridge

In [50]:

```
from sklearn.linear_model import Ridge,Lasso
```

Ridge

In [51]:

```
ri=Ridge(alpha=5)
ri.fit(g_train,h_train)
```

Out[51]:

```
Ridge(alpha=5)
```

In [52]:

```
ri.score(g_test,h_test)
```

Out[52]:

```
0.01483128654609589
```

In [53]:

```
ri.score(g_train,h_train)
```

Out[53]:

```
0.4047929916204691
```

Lasso

In [54]:

```
l=Lasso(alpha=6)
l.fit(g_train,h_train)
```

Out[54]:

```
Lasso(alpha=6)
```

In [55]:

```
l.score(g_test,h_test)
```

Out[55]:
```
-0.2403846153846152
```

In [56]:
```
ri.score(g_train,h_train)
```

Out[56]:
```
0.4047929916204691
```

In [ ]:
```

```