

Water Usage

November 9, 2019

```
In [129]: #importing libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

In [130]: #importing dataset
dataset=pd.read_csv('water.csv')
x=dataset.iloc[:, :-1].values#years
y=dataset.iloc[:, 1].values#water usage

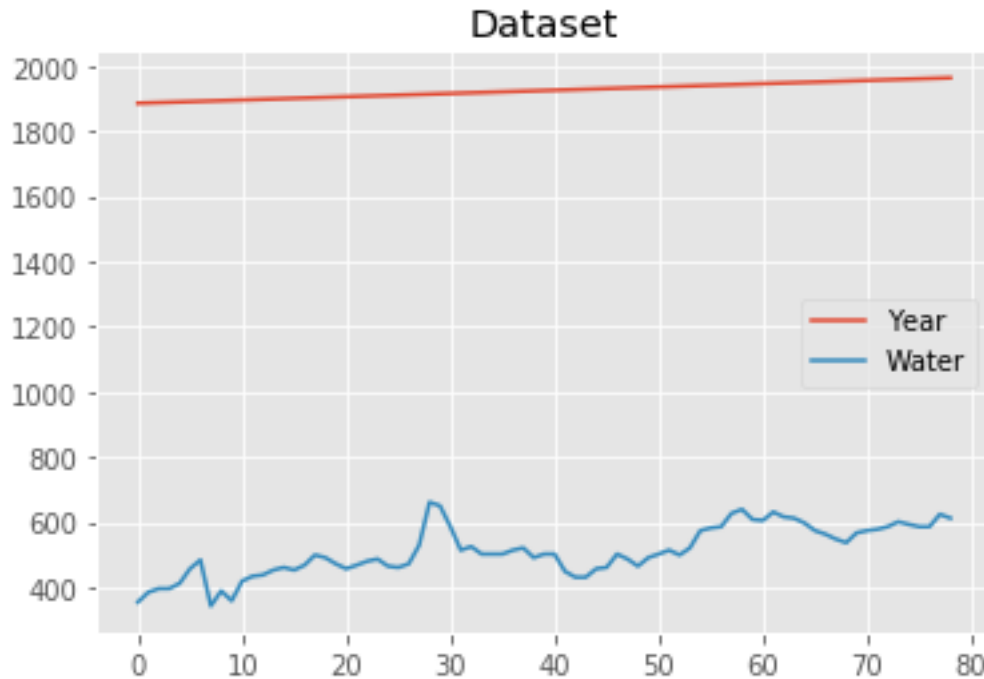
In [131]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=1/3,random_state=0)

In [132]: print(dataset.describe())
```

| | Year | Water |
|-------|-------------|------------|
| count | 79.000000 | 79.000000 |
| mean | 1924.000000 | 512.025316 |
| std | 22.949219 | 75.685072 |
| min | 1885.000000 | 344.000000 |
| 25% | 1904.500000 | 462.000000 |
| 50% | 1924.000000 | 503.000000 |
| 75% | 1943.500000 | 581.000000 |
| max | 1963.000000 | 662.000000 |

```
In [140]: dataset.plot()

plt.title('Dataset')
plt.show()
```



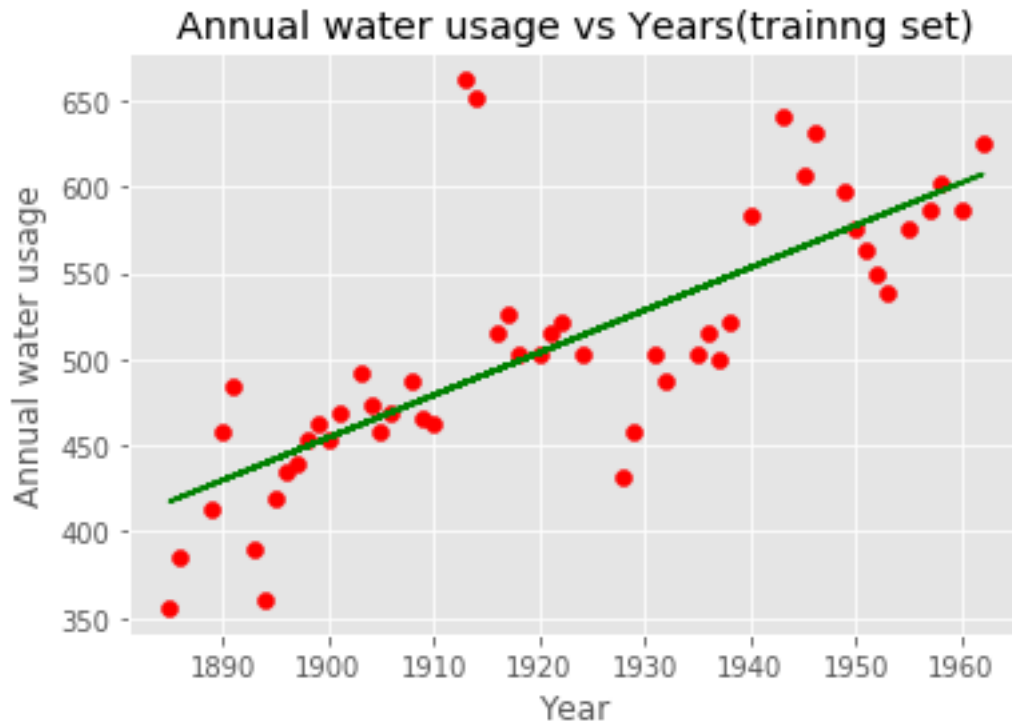
```
In [134]: #fitting simple linear regression to the trainng set
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

```
Out[134]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

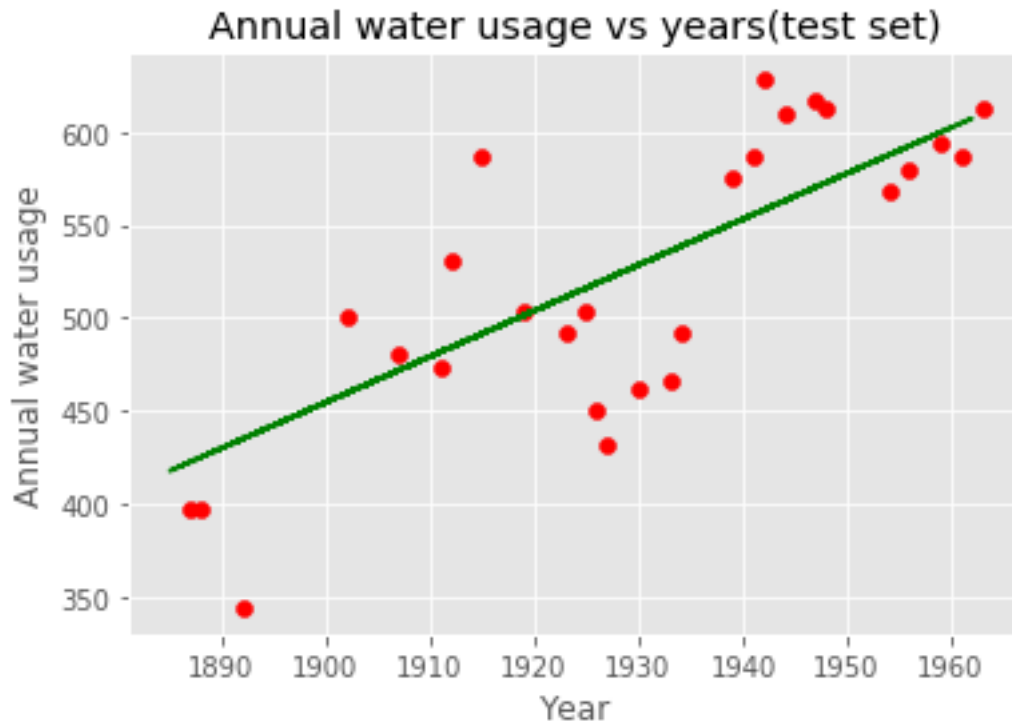
```
In [135]: y_pred=lr.predict(x_test)
y_pred
```

```
Out[135]: array([538.26090753, 484.06231056, 491.45302832, 572.75092378,
528.40661717, 587.53235932, 599.85022227, 604.77736744,
535.79733494, 557.96948824, 516.08875422, 570.28735119,
422.47299582, 592.4595045 , 555.50591565, 501.30731868,
424.93656841, 471.74444761, 434.79085876, 481.59873797,
521.0158994 , 511.16160904, 518.55232681, 609.70451262,
550.57877047, 459.42658466, 562.89663342])
```

```
In [136]: #plotting regression line
plt.scatter(x_train,y_train,color='red')
plt.plot(x_train,lr.predict(x_train),color='green")
plt.title('Annual water usage vs Years(trainng set)')
plt.xlabel('Year')
plt.ylabel('Annual water usage')
plt.show()
```



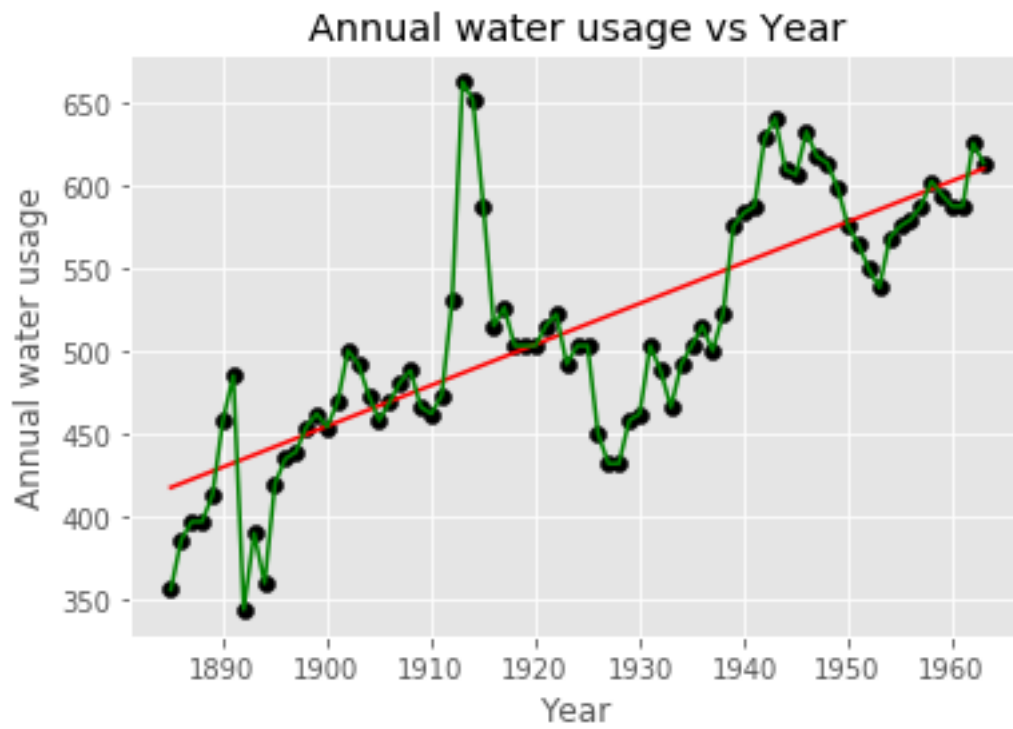
```
In [137]: #plotting test
plt.scatter(x_test,y_test,color='red')
plt.plot(x_train,lr.predict(x_train),color="green")#no change because lr is trained
plt.title('Annual water usage vs years(test set)')
plt.xlabel('Year')
plt.ylabel('Annual water usage')
plt.show()
```



```
In [138]: from sklearn import metrics
from sklearn.metrics import r2_score
print("Rmse:",np.sqrt(metrics.mean_absolute_error(y_test,y_pred)))
print("Mse:",metrics.mean_absolute_error(y_test,y_pred))
r2 = r2_score(y_test,y_pred)
print("R2_score=",r2)
```

```
Rmse: 6.197415167154058
Mse: 38.40795475407116
R2_score= 0.6221860538899177
```

```
In [139]: plt.scatter(x,y,color='k')
plt.plot(x,lr.predict(x),color="red")
plt.plot(x,y,color="green")
plt.title('Annual water usage vs Year')
plt.xlabel('Year')
plt.ylabel('Annual water usage')
plt.show()
```



In []: