

ml2

November 9, 2024

```
[2]: import pandas as pd
import matplotlib.pyplot as plt

import seaborn as sns
```

```
[4]: df= pd.read_csv('temperatures.csv')
df.head()
```

```
[4]:
```

| | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | \ |
|---|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 0 | 1901 | 22.40 | 24.14 | 29.07 | 31.91 | 33.41 | 33.18 | 31.21 | 30.39 | 30.47 | 29.97 | |
| 1 | 1902 | 24.93 | 26.58 | 29.77 | 31.78 | 33.73 | 32.91 | 30.92 | 30.73 | 29.80 | 29.12 | |
| 2 | 1903 | 23.44 | 25.03 | 27.83 | 31.39 | 32.91 | 33.00 | 31.34 | 29.98 | 29.85 | 29.04 | |
| 3 | 1904 | 22.50 | 24.73 | 28.21 | 32.02 | 32.64 | 32.07 | 30.36 | 30.09 | 30.04 | 29.20 | |
| 4 | 1905 | 22.00 | 22.83 | 26.68 | 30.01 | 33.32 | 33.25 | 31.44 | 30.68 | 30.12 | 30.67 | |

| | NOV | DEC | ANNUAL | JAN-FEB | MAR-MAY | JUN-SEP | OCT-DEC |
|---|-------|-------|--------|---------|---------|---------|---------|
| 0 | 27.31 | 24.49 | 28.96 | 23.27 | 31.46 | 31.27 | 27.25 |
| 1 | 26.31 | 24.04 | 29.22 | 25.75 | 31.76 | 31.09 | 26.49 |
| 2 | 26.08 | 23.65 | 28.47 | 24.24 | 30.71 | 30.92 | 26.26 |
| 3 | 26.36 | 23.63 | 28.49 | 23.62 | 30.95 | 30.66 | 26.40 |
| 4 | 27.52 | 23.82 | 28.30 | 22.25 | 30.00 | 31.33 | 26.57 |

```
[5]: df.isnull().sum()
```

```
[5]:
```

| | |
|------|---|
| YEAR | 0 |
| JAN | 0 |
| FEB | 0 |
| MAR | 0 |
| APR | 0 |
| MAY | 0 |
| JUN | 0 |
| JUL | 0 |
| AUG | 0 |
| SEP | 0 |
| OCT | 0 |
| NOV | 0 |
| DEC | 0 |

```
ANNUAL      0
JAN-FEB     0
MAR-MAY     0
JUN-SEP     0
OCT-DEC     0
dtype: int64
```

```
[6]: df.duplicated().sum()
```

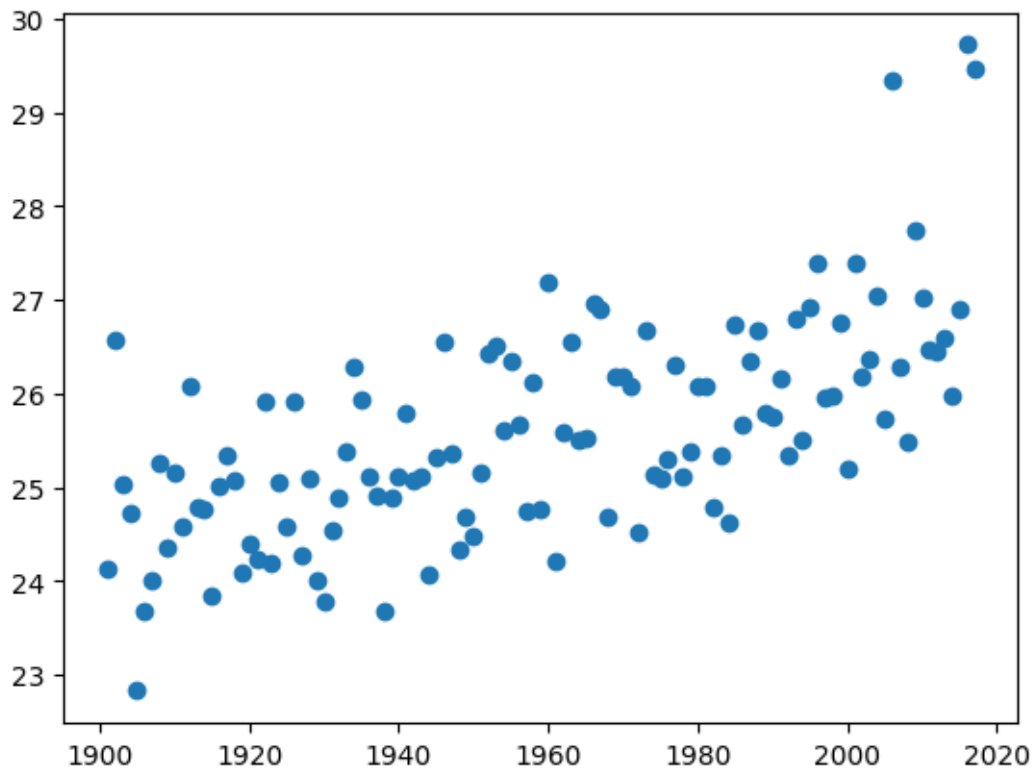
```
[6]: np.int64(0)
```

```
[7]: #for feb
X=df[['YEAR']] #independent
Y=df[['FEB']] #dependent
```

```
[8]: #visualise whole dataset

plt.scatter(X,Y)
```

```
[8]: <matplotlib.collections.PathCollection at 0x13611e730>
```

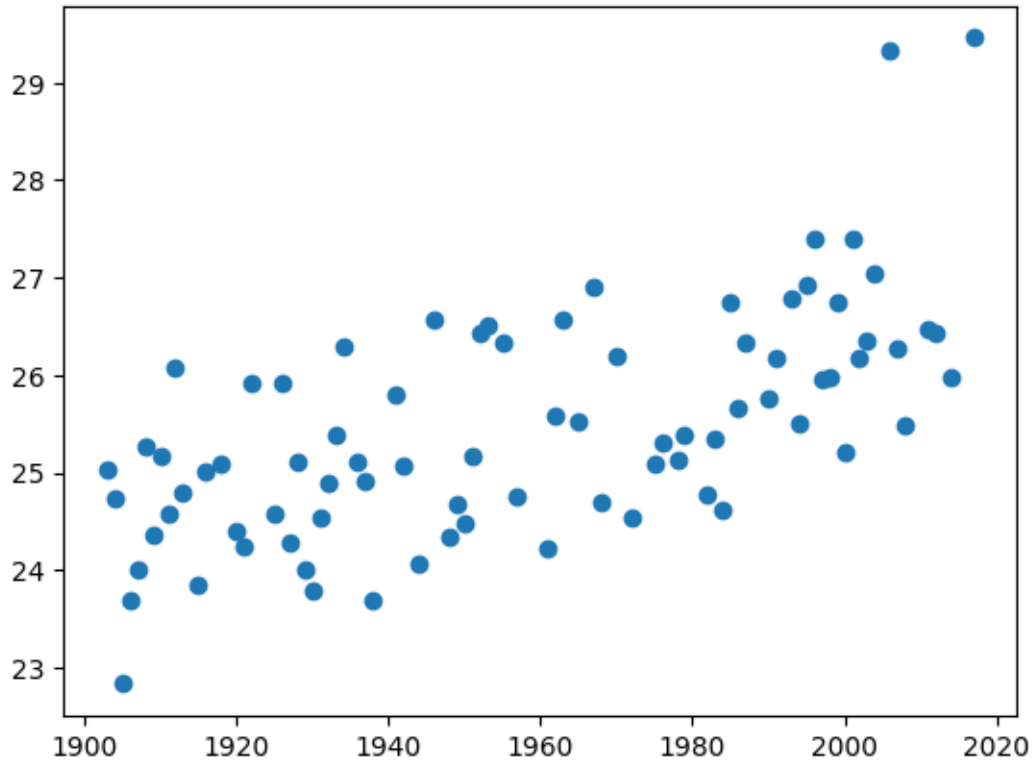


```
[11]: from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LinearRegression
```

```
[12]: Xtrain, Xtest, Ytrain, Ytest=train_test_split(X,Y,test_size=0.3)
```

```
[13]: plt.scatter(Xtrain, Ytrain)
```

```
[13]: <matplotlib.collections.PathCollection at 0x137c28700>
```



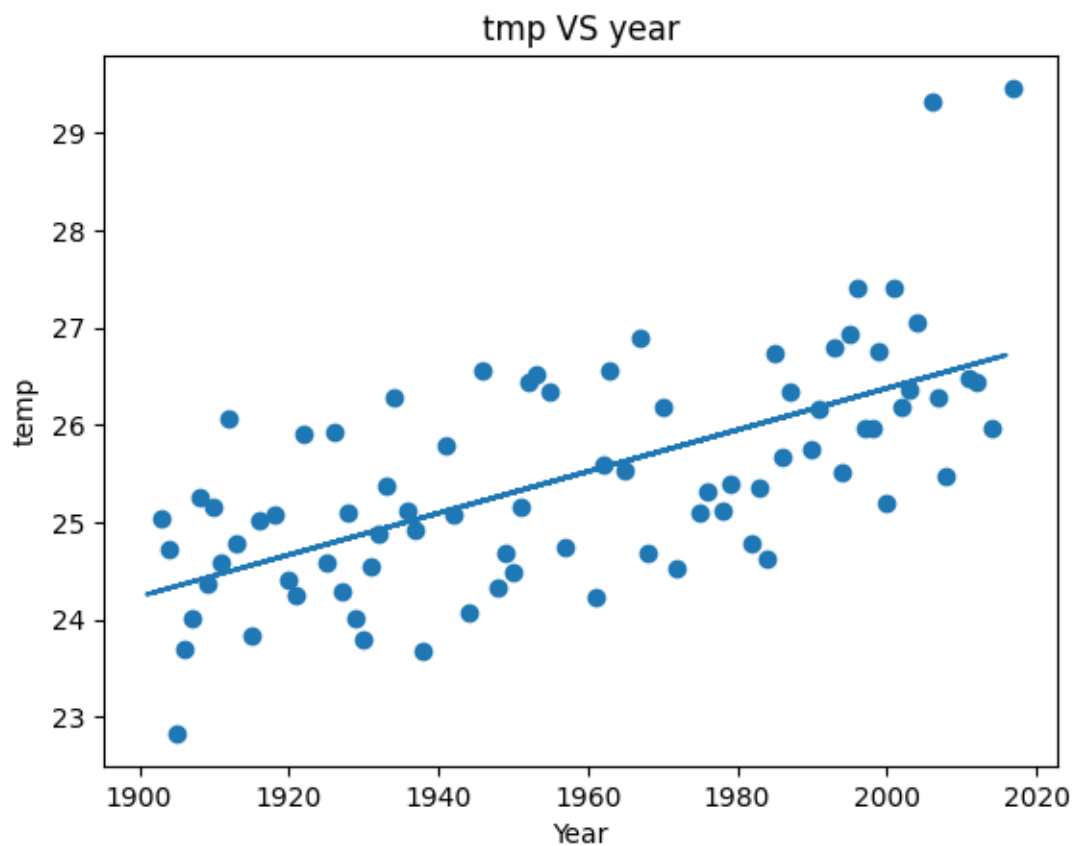
```
[14]: #train the model
      model = LinearRegression()
      model.fit(Xtrain,Ytrain)
```

```
[14]: LinearRegression()
```

```
[15]: #prediction on test data
      Ypred = model.predict(Xtest)
```

```
[18]: plt.xlabel("Year")
      plt.ylabel("temp")
      plt.title("tmp VS year")
      plt.scatter(Xtrain,Ytrain)
```

```
plt.plot(Xtest,Ypred)
plt.show()
```



```
[19]: from sklearn.metrics import mean_absolute_error, mean_squared_error
```

```
[24]: print({mean_absolute_error(Ytest,Ypred)})
      print({mean_squared_error(Ytest,Ypred)})
```

```
{np.float64(0.5979301066812404)}
{np.float64(0.7577909720982343)}
```

```
[ ]:
```