```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model


df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/canada_per_capita_income.csv')
df
```

15	1985	11018.955850
16	1986	11482.891530
17	1987	12974.806620
18	1988	15080.283450
19	1989	16426.725480
20	1990	16838.673200
21	1991	17266.097690
22	1992	16412.083090
23	1993	15875.586730
24	1994	15755.820270
25	1995	16369.317250
26	1996	16699.826680
27	1997	17310.757750
28	1998	16622.671870
29	1999	17581.024140
30	2000	18987.382410
31	2001	18601.397240
32	2002	19232.175560
33	2003	22739.426280
34	2004	25719.147150
35	2005	29198.055690
36	2006	32738.262900
37	2007	36144.481220

```
%matplotlib inline
plt.xlabel('year')
plt.ylabel('per capita income (US$)')
plt.scatter(df['year'],df['per capita income (US$)'],color='red',marker='+')
```

```
<matplotlib.collections.PathCollection at 0x7fa1da879250>
```


```
year = df[['year']]
year
```

	year
0	1970
1	1971
2	1972
3	1973
4	1974
5	1975
6	1976
7	1977
8	1978
9	1979
10	1980
11	1981
12	1982
13	1983
14	1984
15	1985
16	1986
17	1987
18	1988
19	1989
20	1990
21	1991

```
income = df['per capita income (US$)']
income
```

```
0      3399.299037
1      3768.297935
2      4251.175484
3      4804.463248
4      5576.514583
5      5998.144346
6      7062.131392
7      7100.126170
8      7247.967035
9      7602.912681
10     8355.968120
```

```
11     9434.390652
12     9619.438377
13    10416.536590
14    10790.328720
15    11018.955850
16    11482.891530
17    12974.806620
18    15080.283450
19    16426.725480
20    16838.673200
21    17266.097690
22    16412.083090
23    15875.586730
24    15755.820270
25    16369.317250
26    16699.826680
27    17310.757750
28    16622.671870
29    17581.024140
30    18987.382410
31    18601.397240
32    19232.175560
33    22739.426280
34    25719.147150
35    29198.055690
36    32738.262900
37    36144.481220
38    37446.486090
39    32755.176820
40    38420.522890
41    42334.711210
42    42665.255970
43    42676.468370
44    41039.893600
45    35175.188980
46    34229.193630
Name: per capita income (US$), dtype: float64
```

```
reg = linear_model.LinearRegression()
reg.fit(year,income)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
reg.predict([[2020]])
```

```
array([41288.69409442])
```

```
reg.coef_ # m
```

```
array([828.46507522])
```

```
# Y = m * X + b
828.46507522*2020 + -1632210.7578554575
```

```
828.46507522*2020 + -1632210.7578554575
```

```
41288.694088942604
```

```
predicted_income = reg.predict(year)
predicted_income
```

```
array([ -134.55966672,   693.9054085 ,  1522.37048373,  2350.83555895,
        3179.30063417,  4007.7657094 ,  4836.23078462,  5664.69585984,
        6493.16093506,  7321.62601029,  8150.09108551,  8978.55616073,
        9807.02123595, 10635.48631118, 11463.9513864 , 12292.41646162,
       13120.88153685, 13949.34661207, 14777.81168729, 15606.27676251,
       16434.74183774, 17263.20691296, 18091.67198818, 18920.1370634 ,
       19748.60213863, 20577.06721385, 21405.53228907, 22233.9973643 ,
       23062.46243952, 23890.92751474, 24719.39258996, 25547.85766519,
       26376.32274041, 27204.78781563, 28033.25289085, 28861.71796608,
       29690.1830413 , 30518.64811652, 31347.11319175, 32175.57826697,
       33004.04334219, 33832.50841741, 34660.97349264, 35489.43856786,
       36317.90364308, 37146.3687183 , 37974.83379353])
```

```
predicted_df = pd.DataFrame(year)
predicted_df.head(3)
```

	year
0	1970
1	1971
2	1972

```
predicted_df['predicted income'] = predicted_income
predicted_df.head(3)
```

	year	predicted income
0	1970	-134.559667
1	1971	693.905409
2	1972	1522.370484

```
%matplotlib inline
plt.xlabel('year', fontsize=20)
plt.ylabel('per capita income (US$)', fontsize=20)
plt.scatter(df['year'],df['per capita income (US$)'],color='red',marker='+')
plt.plot(predicted_df['year'],predicted_df['predicted income'],color='blue')
```