Q- Predict the per capita income of canada in the year 2020

In [1]:
```
import pandas as pd
```

In [2]:
```
df=pd.read_csv("https://raw.githubusercontent.com/codebasics/py/master/ML/1_linear_reg/Exercise/canada_per_capita_income.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	year	per capita income (US$)
0	1970	3399.299037
1	1971	3768.297935
2	1972	4251.175484
3	1973	4804.463248
4	1974	5576.514583

In [17]:
```
df.shape
```

Out[17]:
```
(47, 2)
```

In [18]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 47 entries, 0 to 46
Data columns (total 2 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   year                    47 non-null     int64  
 1   per capita income (US$)  47 non-null     float64
dtypes: float64(1), int64(1)
memory usage: 880.0 bytes
```

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

Out[19]:

	year	per capita income (US$)
42	2012	42665.25597
43	2013	42676.46837
44	2014	41039.89360
45	2015	35175.18898
46	2016	34229.19363

In [ ]:
```

```

In [4]:
```
!pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\users\rohit\appdata\local\programs\python\python39\lib\site-packages (1.0.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\rohit\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn) (3.1.0)
Requirement already satisfied: scipy>=1.1.0 in c:\users\rohit\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn) (1.8.0)
Requirement already satisfied: joblib>=0.11 in c:\users\rohit\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn) (1.1.0)
Requirement already satisfied: numpy>=1.14.6 in c:\users\rohit\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn) (1.22.3)
WARNING: You are using pip version 21.2.3; however, version 22.0.4 is available.
You should consider upgrading via the 'C:\Users\Rohit\AppData\Local\Programs\Python\Python39\python.exe -m pip install --upgrade pip' command.
```

In [5]:
```
import matplotlib.pyplot as plt
```

In [6]:
```
from sklearn import linear_model
```

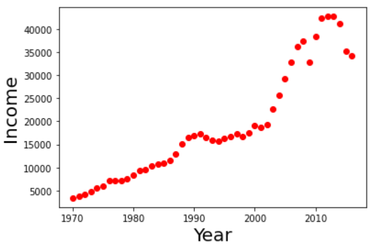
In [7]:
```
df.head()
```

Out[7]:

	year	per capita income (US$)
0	1970	3399.299037
1	1971	3768.297935
2	1972	4251.175484
3	1973	4804.463248
4	1974	5576.514583

In [9]:
```
plt.xlabel("Year", fontsize=20)
plt.ylabel("Income", fontsize=20)
plt.scatter(df["year"],df["per capita income (US$)"],color="red")
```

Out[9]:
```
<matplotlib.collections.PathCollection at 0x1906aa52760>
```



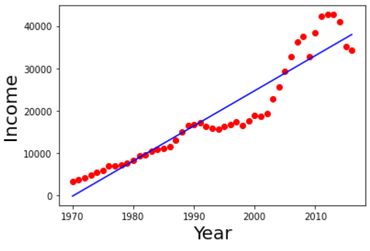
In [10]:
```
reg=linear_model.LinearRegression()
```

In [11]:
```
reg.fit(df[["year"]],df["per capita income (US$)"])
```

Out[11]:
```
LinearRegression()
```

In [12]:
```
%matplotlib inline
plt.xlabel("Year", fontsize=20)
plt.ylabel("Income", fontsize=20)
plt.scatter(df["year"],df["per capita income (US$)"],color="red")
plt.plot(df.year,reg.predict(df[["year"]]),color="blue")
```

Out[12]:
```
[<matplotlib.lines.Line2D at 0x1906aae9820>]
```



In [13]:
```
reg.coef_
```

Out[13]:
```
array([828.46507522])
```

In [14]:
```
reg.intercept_
```

Out[14]:
```
-1632210.7578554575
```

In [15]:
```
reg.predict([[2020]])
```

```
C:\Users\Rohit\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

Out[15]:
```
array([41288.69409442])
```

The Predicted per capita income of canada in 2020 is 41288.694 US dollars