In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

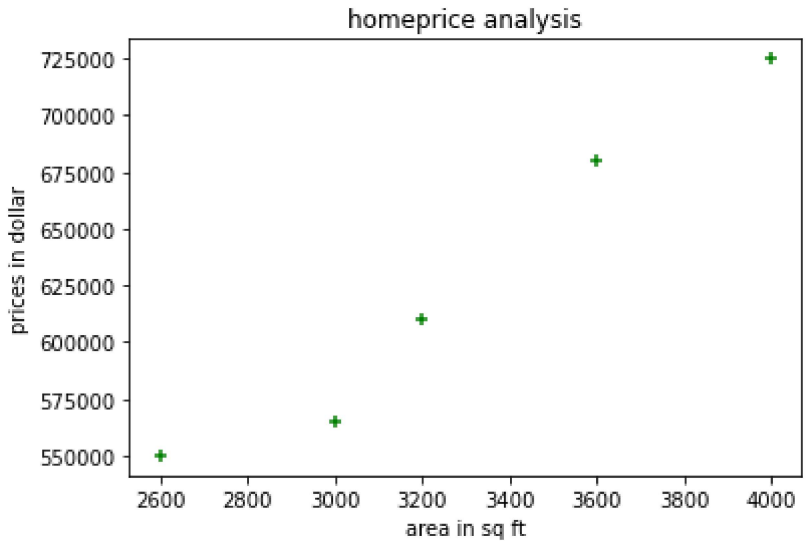
In [6]:
```
df = pd.read_csv("homeprices.csv")
df
```

Out[6]:

| | area | prices |
|---|---|---|
| 0 | 2600 | 550000 |
| 1 | 3000 | 565000 |
| 2 | 3200 | 610000 |
| 3 | 3600 | 680000 |
| 4 | 4000 | 725000 |

In [10]:
```
%matplotlib inline
plt.xlabel('area in sq ft')
plt.ylabel('prices in dollar')
plt.title('homeprice analysis')
plt.scatter(df.area , df.prices, color = 'g' , marker = '+' )
```

Out[10]:
```
<matplotlib.collections.PathCollection at 0x25ca6c40130>
```




In [11]:
```
reg = linear_model.LinearRegression()
```

In [13]:
```
reg.fit(df[['area']], df.prices)
```

Out[13]:
```
LinearRegression()
```

In [16]:
```
reg.predict([[3300]])
```

Out[16]:
```
array([628715.75342466])
```

In [17]:
```
reg.coef_
```

Out[17]:
```
array([135.78767123])
```

In [18]:
```
reg.intercept_
```

Out[18]:
```
180616.43835616432
```

In [39]:
```
df = pd.read_csv("canada.csv")
df
```

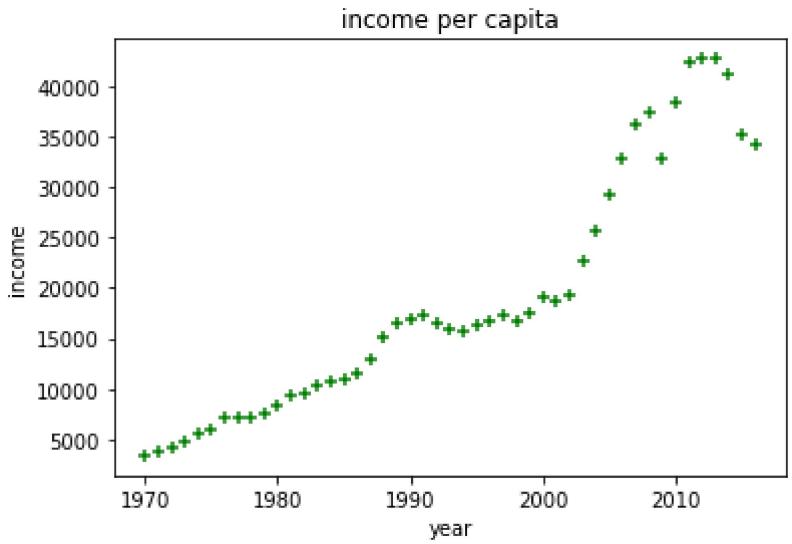
Out[39]:

| | year | income | Unnamed: 2 |
|---|---|---|---|
| **0** | 1970 | 3399.299037 | NaN |
| **1** | 1971 | 3768.297935 | NaN |
| **2** | 1972 | 4251.175484 | NaN |
| **3** | 1973 | 4804.463248 | NaN |
| **4** | 1974 | 5576.514583 | NaN |
| **5** | 1975 | 5998.144346 | NaN |
| **6** | 1976 | 7062.131392 | NaN |
| **7** | 1977 | 7100.126170 | NaN |
| **8** | 1978 | 7247.967035 | NaN |
| **9** | 1979 | 7602.912681 | NaN |
| **10** | 1980 | 8355.968120 | NaN |
| **11** | 1981 | 9434.390652 | NaN |
| **12** | 1982 | 9619.438377 | NaN |
| **13** | 1983 | 10416.536590 | NaN |
| **14** | 1984 | 10790.328720 | NaN |
| **15** | 1985 | 11018.955850 | NaN |
| **16** | 1986 | 11482.891530 | NaN |
| **17** | 1987 | 12974.806620 | NaN |
| **18** | 1988 | 15080.283450 | NaN |
| **19** | 1989 | 16426.725480 | NaN |
| **20** | 1990 | 16838.673200 | NaN |
| **21** | 1991 | 17266.097690 | NaN |
| **22** | 1992 | 16412.083090 | NaN |
| **23** | 1993 | 15875.586730 | NaN |
| **24** | 1994 | 15755.820270 | NaN |
| **25** | 1995 | 16369.317250 | NaN |
| **26** | 1996 | 16699.826680 | NaN |
| **27** | 1997 | 17310.757750 | NaN |
| **28** | 1998 | 16622.671870 | NaN |
| **29** | 1999 | 17581.024140 | NaN |
| **30** | 2000 | 18987.382410 | NaN |
| **31** | 2001 | 18601.397240 | NaN |
| **32** | 2002 | 19232.175560 | NaN |
| **33** | 2003 | 22739.426280 | NaN |

| | year | income | Unnamed: 2 |
|---|---|---|---|
| **34** | 2004 | 25719.147150 | NaN |
| **35** | 2005 | 29198.055690 | NaN |
| **36** | 2006 | 32738.262900 | NaN |
| **37** | 2007 | 36144.481220 | NaN |
| **38** | 2008 | 37446.486090 | NaN |
| **39** | 2009 | 32755.176820 | NaN |
| **40** | 2010 | 38420.522890 | NaN |
| **41** | 2011 | 42334.711210 | NaN |
| **42** | 2012 | 42665.255970 | NaN |
| **43** | 2013 | 42676.468370 | NaN |
| **44** | 2014 | 41039.893600 | NaN |
| **45** | 2015 | 35175.188980 | NaN |
| **46** | 2016 | 34229.193630 | NaN |

In [44]:
```
%matplotlib inline
plt.xlabel('year')
plt.ylabel('income')
plt.title('income per capita')
plt.rcParams["figure.figsize"] = (20,20)
plt.scatter(df.year , df.income, color = 'g' , marker = '+' )
```

Out[44]: <matplotlib.collections.PathCollection at 0x25ca829fcd0>




In [29]:
```
reg = linear_model.LinearRegression()
```

In [30]:
```
reg.fit(df[['year']], df.income)
```

Out[30]: LinearRegression()

In [49]:
```
reg.predict([[2018]])
```

Out[49]:
```
array([39631.76394397])
```

In [ ]:
```

```