In [2]:

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

In [12]:

```python
# load the dataset

df = pd.read_csv("HousePrices.txt")
df
```

Out[12]:

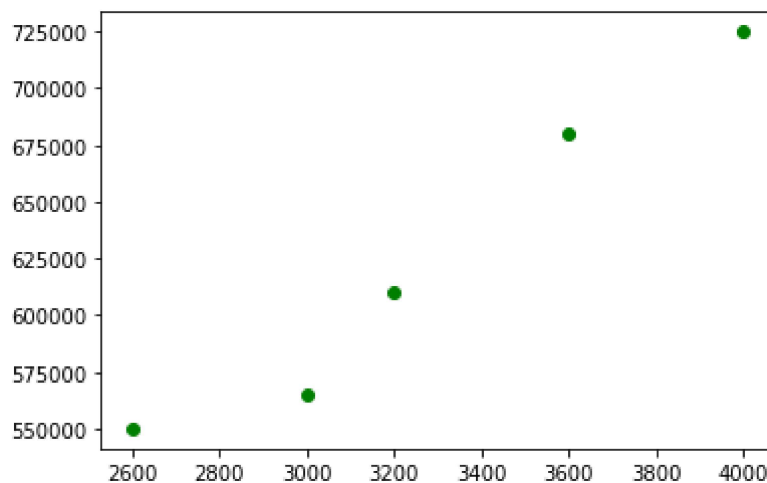|	area	price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

In [13]:

```python
# plot the values

df = pd.read_csv("HousePrices.txt")
plt.scatter(df["area"], df["price"], color="green")
```

Out[13]:

```
<matplotlib.collections.PathCollection at 0x7f114e330c18>
```

In [17]:

```python
reg = linear_model.LinearRegression()
X = df[['area']]
y =  df.price
model = reg.fit(X,y)          # first arg must be a 2D array
```

In [24]:

```python
# results
# coefficient of determination (R²) with .score(), parameters being predictor X and regres
sor y
# The attributes of model are .intercept_, which represents the coefficient, b₀ and .coef
_, which represents b₁
# To obtain the predicted response, use .predict(), parameter being the predictor
## .intercept_ is a one-dimensional array with the single element b₀
## .coef_ is a two-dimensional array with the single element b₁.
print('RESULTS\n')
print(f"r^2 : {model.score(X,y)}")
print(f"Line of regression equation: y = {model.coef_[0]}X + {model.intercept_}")
```
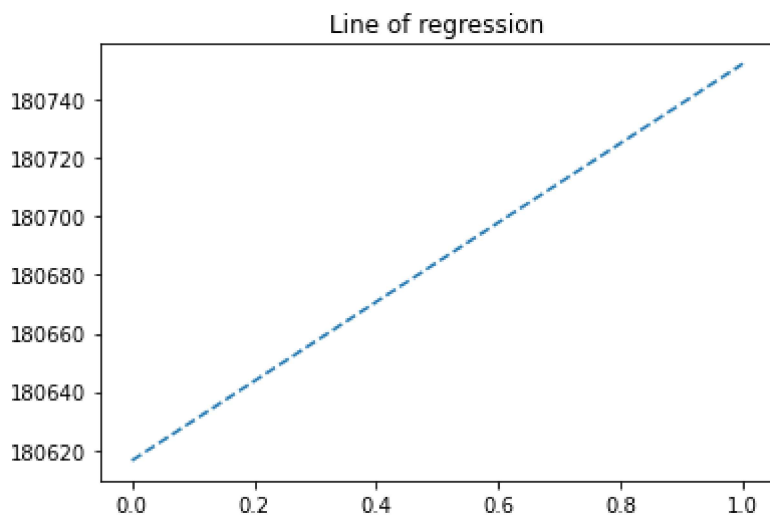
RESULTS

r^2 : 0.9584301138199486
Line of regression equation: y = 135.78767123287673X + 180616.43835616432

In [30]:

```python
axes = plt.gca()
x_vals = np.array(axes.get_xlim())
y_vals = model.intercept_ + model.coef_[0] * x_vals
plt.plot(x_vals, y_vals, '--')
plt.title("Line of regression")
```

Out[30]:

Text(0.5, 1.0, 'Line of regression')

In [47]:

```
a = pd.read_csv('areas.txt')
pred = reg.predict(a)
pred
a['predicted_prices'] = pred
a
```

Out[47]:

	areas	predicted_prices
0	1000	316404.109589
1	2000	452191.780822
2	3000	587979.452055
3	4200	750924.657534
4	4250	757714.041096
5	5000	859554.794521
6	5500	927448.630137
7	5750	961395.547945

results

In [48]:

```
print("GIVEN DATA\n")
df
```

GIVEN DATA

Out[48]:

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

In [50]:

```python
print("PREDICTED DATA\n")
a
```

PREDICTED DATA

Out[50]:

	areas	predicted_prices
0	1000	316404.109589
1	2000	452191.780822
2	3000	587979.452055
3	4200	750924.657534
4	4250	757714.041096
5	5000	859554.794521
6	5500	927448.630137
7	5750	961395.547945

In [ ]: