In [2]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing ,svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [3]:

```
df=pd.read_csv(r"C:\Users\91628\Downloads\data (1).csv")
print(df)
```

		da [.]	te	р	rice	bedrooms	bathrooms	saft	living
0	2014-05-02			130000		3.0	1.50	• -	1340
\									
1	2014-05-02	00:00:	aa 2.3	384000	e+06	5.0	2.50		3650
2	2014-05-02			120000		3.0	2.00		1930
3	2014-05-02			200000		3.0	2.25		2000
4	2014-05-02	00:00:	aa 5.5	500000	e+05	4.0	2.50		1940
			••		• • •	• • •	•••		
4595	2014-07-09			81667		3.0	1.75		1510
4596	2014-07-09			343333		3.0	2.50		1460
4597 4508	2014-07-09			169042		3.0	2.50		3010
4598 4599	2014-07-10 2014-07-10			934000 206000		4.0 3.0	2.00 2.50		2090 1490
4333	2014-07-10	00.00.	00 2.2	200000	E+03	3.0	2.50		1430
	sqft_lot	floors	wateri	front	view	conditio	n sqft_abo	N/E	
0	7912	1.5	water	0	0		. –	340 \	
1	9050	2.0		0	4			370	
2	11947	1.0		0	0			30	
3	8030	1.0		0	0			000	
4	10500	1.0		0	0			.40	
• • •	•••	•••		• • •				••	
4595	6360	1.0		0	0			10	
4596	7573	2.0		0	0			ŀ60	
4597	7014	2.0		0	0			10	
4598	6630	1.0		0	0		3 10	70	
4599	8102	2.0		0	0		4 14	190	
	sqft_basem		_built	yr_r	enovat			stre	
0	. –	0	1955	yr_r		905 1	8810 Densmo	re Ave	e N \
1	. –	0 280	1955 1921	yr_r		0 0 1	709 W B	ore Ave Blaine	e N ∖ St
1 2		0 280 0	1955 1921 1966	yr_r		005 1 0 0 26206	709 W B -26214 143r	ore Ave Slaine od Ave	e N \ St SE
1 2 3	1	0 280 0 .000	1955 1921 1966 1963	yr_r	26	005 1 0 0 26206 0	709 W B -26214 143r 857 170	ore Ave Blaine Pd Ave Oth Pl	e N \ St SE NE
1 2	1	0 280 0	1955 1921 1966 1963 1976	yr_r	26 19	905 1 0 0 0 26206 0 992	709 W B -26214 143r	ore Ave Blaine Pd Ave Oth Pl	e N \ St SE NE
1 2 3 4	1	0 280 0 .000 800	1955 1921 1966 1963 1976	yr_r	20 19	905 1 0 0 26206 0 992	709 W B -26214 143r 857 170 9105 170t	ore Ave Blaine d Ave Oth Pl ch Ave	e N \ St SE NE NE
1 2 3 4 4595	1	0 280 0 000 800 	1955 1921 1966 1963 1976 	yr_r	19	905 1 0 0 26206 0 992	709 W B -26214 143r 857 170 9105 170t 501 N	ore Ave Blaine od Ave Oth Pl Ch Ave	St SE NE NE
1 2 3 4 4595 4596	1	0 280 0 .000 800 0	1955 1921 1966 1963 1976 1954 1983	yr_r	19	005 1 0 26206 0 992 079	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE	ore Ave Blaine od Ave Oth Pl Ch Ave 143rd	e N \ St SE NE NE St St Pl
1 2 3 4 4595 4596 4597	1	0 280 0 .000 800 0 0	1955 1921 1966 1963 1976 1954 1983 2009	yr_r	19	005 1 0 26206 0 992 079 009	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa	ore Ave Blaine od Ave Oth Pl Ch Ave 143rd E 10th	e N \ St SE NE NE St St Pl NE
1 2 3 4 4595 4596 4597 4598	1	0 280 0 .000 800 0 0 0	1955 1921 1966 1963 1976 1954 1983 2009 1974	yr_r	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597	1	0 280 0 .000 800 0 0	1955 1921 1966 1963 1976 1954 1983 2009	yr_r	19	005 1 0 26206 0 992 079 009	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598	1	0 280 0 .000 800 0 0 0	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990		19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	1	0 280 0 .000 800 0 0 0	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990		19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598	r – 1 city	0 280 0 .000 800 0 0 0 .020 0	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour	ntry	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	1 city Shoreline	0 280 0 .000 800 0 0 .020 0 statez WA 981	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour	ntry USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	city Shoreline Seattle	0 280 0 .000 800 0 0 .020 0 statez WA 981	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour	ntry USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	city Shoreline Seattle Kent	0 280 0 .000 800 0 0 .020 0 statez: WA 981: WA 981:	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19	ntry USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	city Shoreline Seattle Kent Bellevue Redmond	0 280 0 .000 800 0 0 .020 0 statez WA 981 WA 980 WA 980 WA 980	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 208	ntry USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Preston	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599	city Shoreline Seattle Kent Bellevue Redmond Seattle	0 280 0 000 800 0 0 0 .020 0 statez: WA 981 WA 980 WA 980 WA 980 WA 980	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 28 52 	USA USA USA USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Pre Ave	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599 0 1 2 3 4 4595 4596	city Shoreline Seattle Kent Bellevue Redmond Seattle Bellevue	0 280 0 .000 800 0 0 .020 0 .020 0 Statez: WA 981: WA 980: WA 980: WA 980: WA 980: WA 980: WA 981: WA 981:	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 208 52 	USA USA USA USA USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Pre Ave	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599 0 1 2 3 4 4595 4596 4597	city Shoreline Seattle Kent Bellevue Redmond Seattle Bellevue Renton	0 280 0 .000 800 0 0 .020 0 statez WA 981 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 28 52 33 37 59	USA USA USA USA USA USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Pre Ave	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599 0 1 2 3 4 4595 4596 4597 4598	city Shoreline Seattle Kent Bellevue Redmond Seattle Bellevue Renton Seattle	0 280 0 .000 800 0 0 .020 0 .020 0 .020 WA 981 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 981	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 208 52 33 27 59 78	USA USA USA USA USA USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Pre Ave	St SE NE NE St St NE St Pl NE St St
1 2 3 4 4595 4596 4597 4598 4599 0 1 2 3 4 4595 4596 4597	city Shoreline Seattle Kent Bellevue Redmond Seattle Bellevue Renton	0 280 0 .000 800 0 0 .020 0 statez WA 981 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980 WA 980	1955 1921 1966 1963 1976 1954 1983 2009 1974 1990 ip cour 33 19 42 208 52 33 27 59 78	USA USA USA USA USA USA USA USA USA	19	005 1 0 26206 0 992 079 009 0 0	709 W B -26214 143r 857 170 9105 170t 501 N 14855 SE 759 Ilwa 5148 S Cr	Pre Ave	St SE NE NE St St NE St Pl NE St St

[4600 rows x 18 columns]

In [4]:

```
df=df[['sqft_living','yr_built']]
df.columns=['living','built']
```

In [5]:

```
df.head(10)
```

Out[5]:

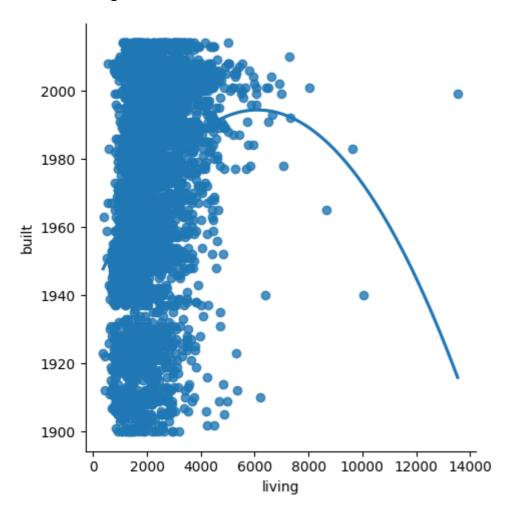
	living	built
0	1340	1955
1	3650	1921
2	1930	1966
3	2000	1963
4	1940	1976
5	880	1938
6	1350	1976
7	2710	1989
8	2430	1985
9	1520	1945

In [6]:

sns.lmplot(x="living",y="built",data=df,order=2,ci=None)

Out[6]:

<seaborn.axisgrid.FacetGrid at 0x24354b52e90>



In [7]:

df.describe()

Out[7]:

	living	built
count	4600.000000	4600.000000
mean	2139.346957	1970.786304
std	963.206916	29.731848
min	370.000000	1900.000000
25%	1460.000000	1951.000000
50%	1980.000000	1976.000000
75%	2620.000000	1997.000000
max	13540.000000	2014.000000

In [8]:

```
df.fillna(method='ffill',inplace=True)
```

In [9]:

```
x=np.array(df['living']).reshape(-1,1)
y=np.array(df['built']).reshape(-1,1)
```

In [10]:

```
df.dropna(inplace=True)
```

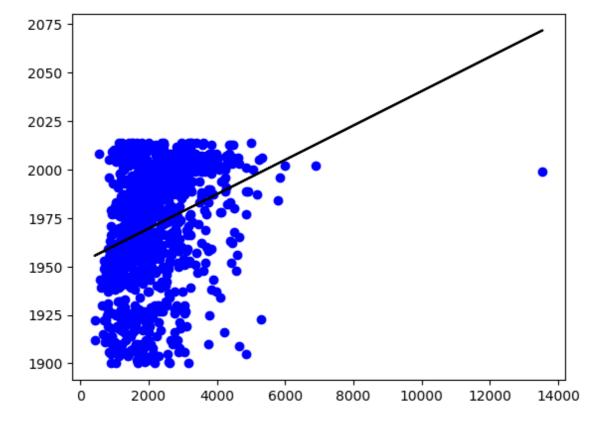
In [11]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
regr=LinearRegression()
regr.fit(x_train,y_train)
print(regr.score(x_test,y_test))
```

0.08578920038865767

In [12]:

```
y_pred=regr.predict(x_test)
plt.scatter (x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```

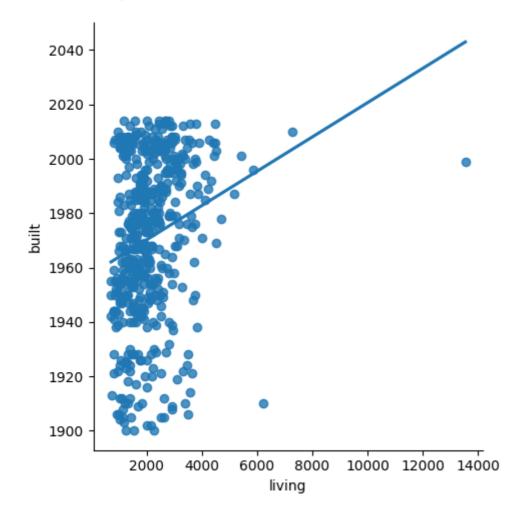


In [13]:

```
df500=df[:][:500]
sns.lmplot(x="living",y="built",data=df500,order=1,ci=None)
```

Out[13]:

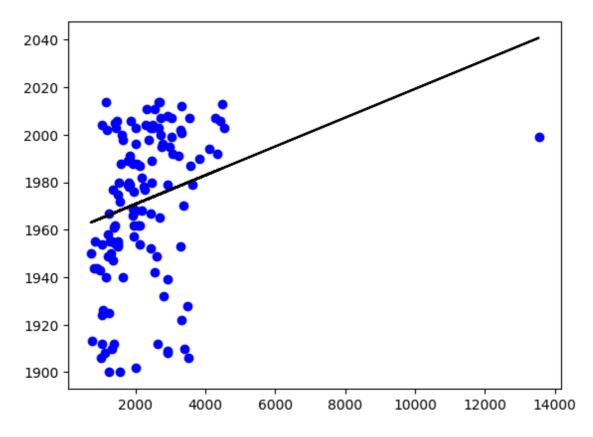
<seaborn.axisgrid.FacetGrid at 0x24339bbae90>



In [14]:

```
df500.fillna(method='ffill',inplace=True)
x=np.array(df500['living']).reshape(-1,1)
y=np.array(df500['built']).reshape(-1,1)
df500.dropna(inplace=True)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
regr=LinearRegression()
regr.fit(x_train,y_train)
print("Regression:",regr.score(x_test,y_test))
y_pred=regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
plt.show()
```

Regression: 0.07599677032427032



In [15]:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
mode1=LinearRegression()
mode1.fit(x_train,y_train)
y_pred=mode1.predict(x_test)
r2=r2_score(y_test,y_pred)
print("R2 score:",r2)
```

R2 score: 0.07599677032427032

In []: