Housing Analysis

In this notebook, you will be replicating much of what you saw in this lesson using the housing data shown below.

After you complete this notebook and the set of quizzes that follow, you will be set to try out your skills on a couple new datasets to gain some additional practice. **Ignore the warning message**.

In [1]:
```
import numpy as np
import pandas as pd
import statsmodels.api as sm;

df = pd.read_csv('./house_price_area_only.csv')
df.head()
```

```
/opt/conda/lib/python3.6/site-packages/statsmodels/compat/pandas.py:56:
FutureWarning: The pandas.core.datetools module is deprecated and will
be removed in a future version. Please use the pandas.tseries module in
stead.
  from pandas.core import datetools
```

Out[1]:

	price	area
0	598291	1188
1	1744259	3512
2	571669	1134
3	493675	1940
4	1101539	2208

1. Use the documentation here (http://www.statsmodels.org/dev/regression.html) and the **statsmodels** library to fit a linear model to predict **price** based on **area**. Obtain a summary of the results, and use them to answer the following quiz questions. Don't forget to add an intercept.

In [2]:
```
df['intercept'] = 1

lm = sm.OLS(df['price'], df[['intercept', 'area']])
results = lm.fit()
results.summary()
```

Out[2]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.678
Model:	OLS	**Adj. R-squared:**	0.678
Method:	Least Squares	**F-statistic:**	1.269e+04
Date:	Tue, 14 Apr 2020	**Prob (F-statistic):**	0.00
Time:	02:36:12	**Log-Likelihood:**	-84517.
No. Observations:	6028	**AIC:**	1.690e+05
Df Residuals:	6026	**BIC:**	1.691e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	9587.8878	7637.479	1.255	0.209	-5384.303	2.46e+04
area	348.4664	3.093	112.662	0.000	342.403	354.530

Omnibus:	368.609	**Durbin-Watson:**	2.007
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	349.279
Skew:	0.534	**Prob(JB):**	1.43e-76
Kurtosis:	2.499	**Cond. No.**	4.93e+03

In [ ]:
```

```