jupyter **phase-2-jupyter-notebook** Last Checkpoint: Last Monday at 12:02 PM Autosave Failed!

Logout

File   Edit   View   Insert   Cell   Kernel   Navigate   Widgets   Help

Not Connected   error   Trusted   Python 3

Contents ⟳ ⚙

1	538000.0	-0.275883	-0.409901	0.453143	1.260322	1	1	1	1	0	0
2	180000.0	-1.530739	-1.516093	-1.474146	-1.733591	0	0	1	0	0	0
3	604000.0	0.978974	-0.409901	1.609516	0.245718	1	0	1	1	0	0
4	510000.0	-0.275883	0.696291	0.067685	-0.220002	1	0	1	0	0	0

In [170]:
```python
outcome = baseline_df['price']
x_cols = baseline_df.drop(['price'], axis=1 )
predictors = sm.add_constant(x_cols)
model = sm.OLS(endog=outcome, exog=predictors).fit()
model.summary()
```

Out[170]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.389
Model:	OLS	**Adj. R-squared:**	0.388
Method:	Least Squares	**F-statistic:**	1159.
Date:	Tue, 04 Jan 2022	**Prob (F-statistic):**	0.00
Time:	15:56:33	**Log-Likelihood:**	-2.4102e+05
No. Observations:	18235	**AIC:**	4.821e+05
Df Residuals:	18224	**BIC:**	4.822e+05
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]

4.5 Fourth Iteration

Let's add the final variables and remove the high p-score from the last model.