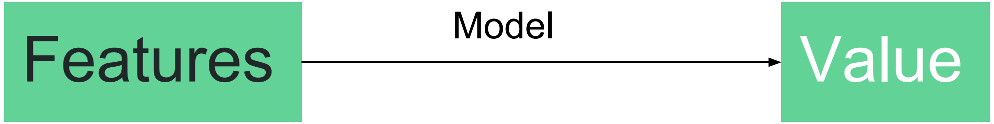
Ski Resort Lift Pass Prices

Benjamin Leung

Objective

- ➢ Find ski resorts with good value



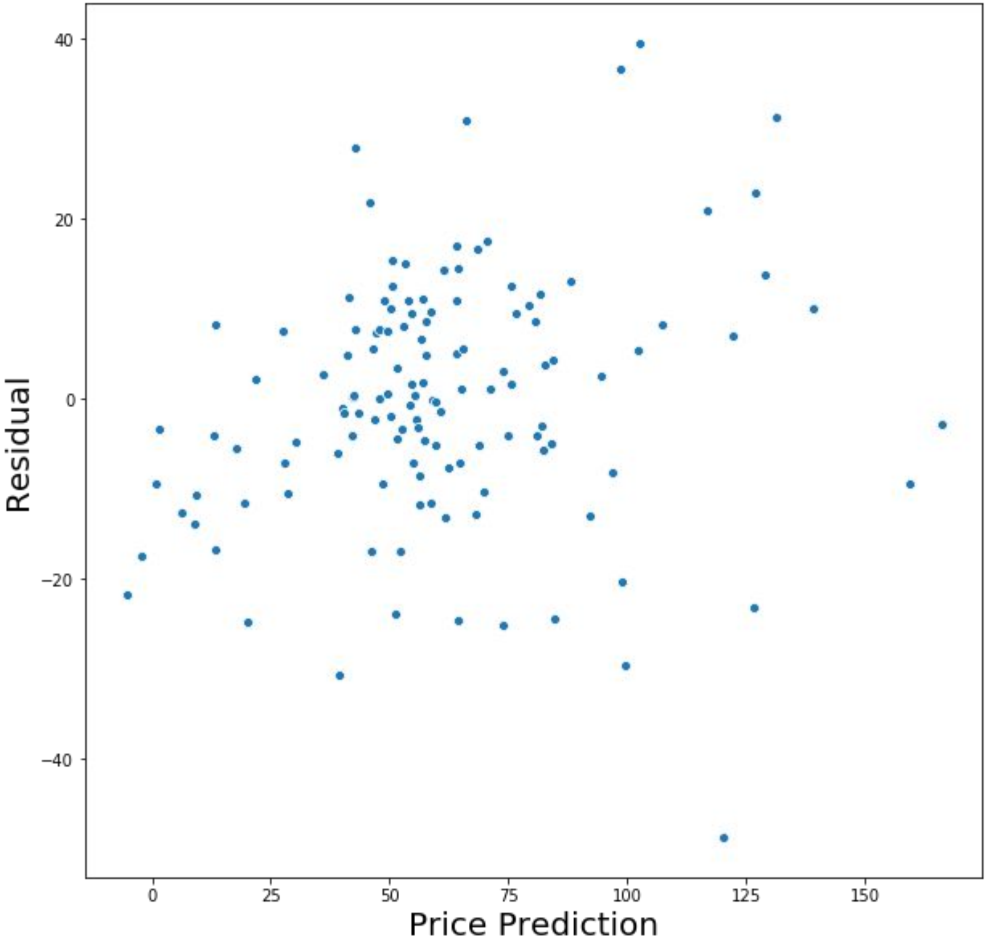
- ➢ Find a model predicting actual value($) based on features
- ➢ Compare actual prices with model output actual values

Tools and Data

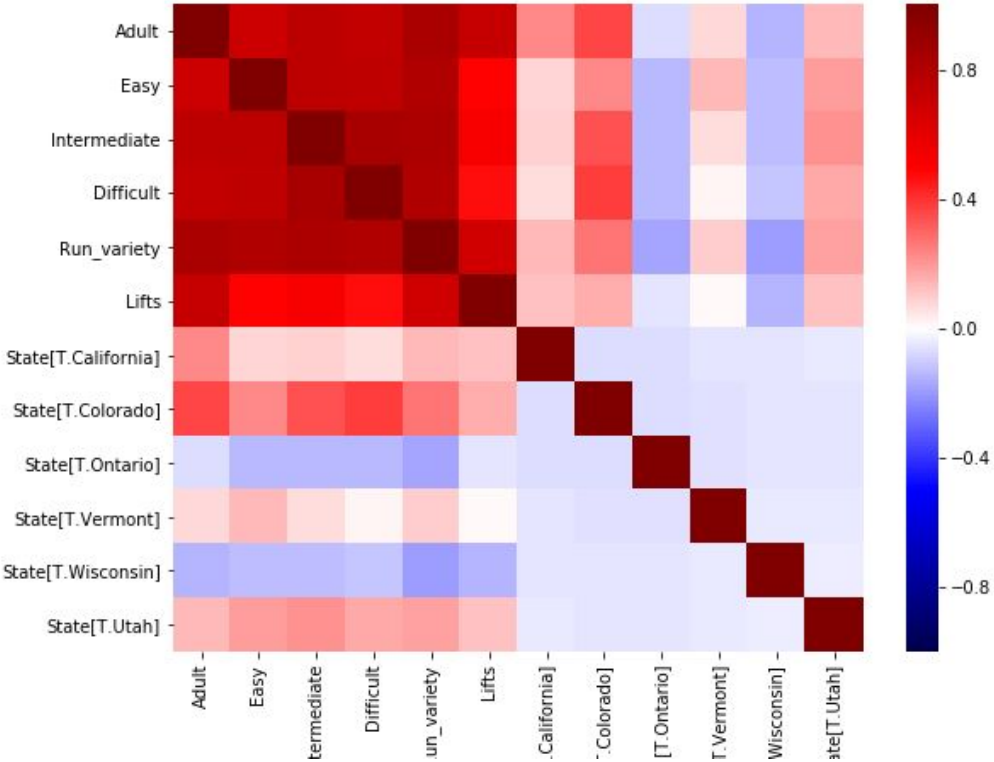
Modules: Numpy, Pandas, Statsmodel, Sklearn, Seaborn, Matplotlib, Beautifulsoup, Selenium, Scipy

Initial Fitting

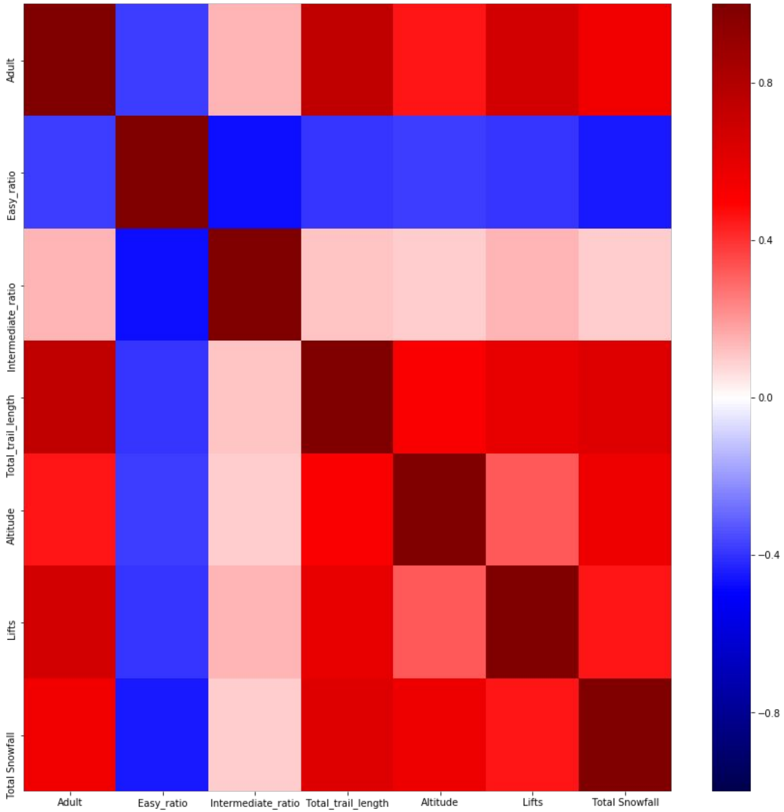
$r^2 = 0.442$



Colinearity



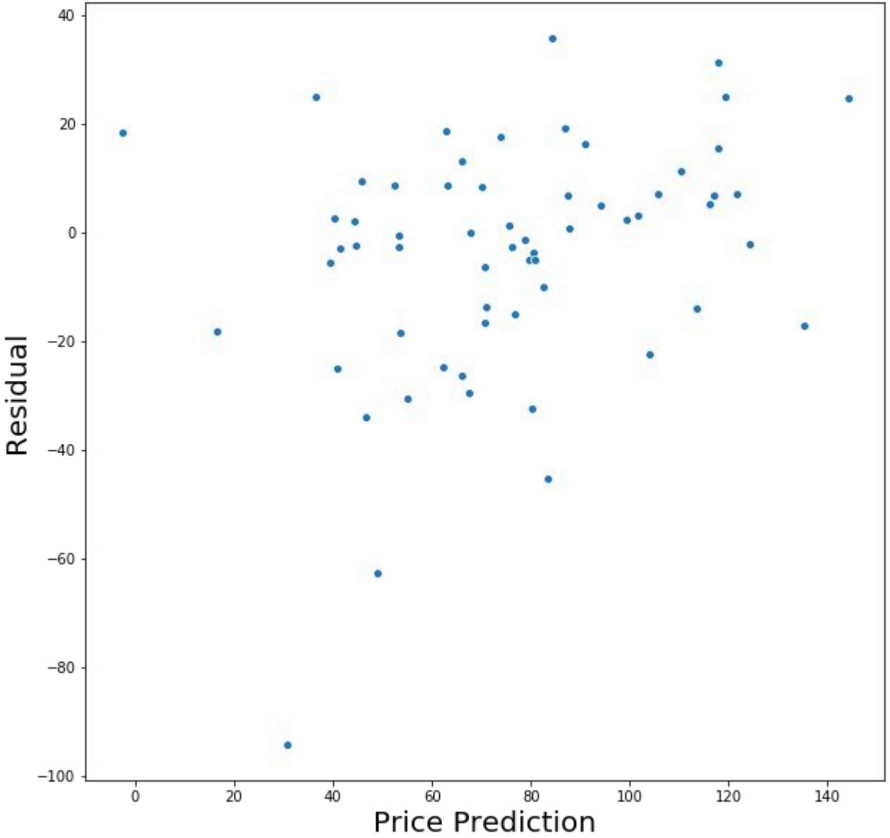
New Heat Map with FE



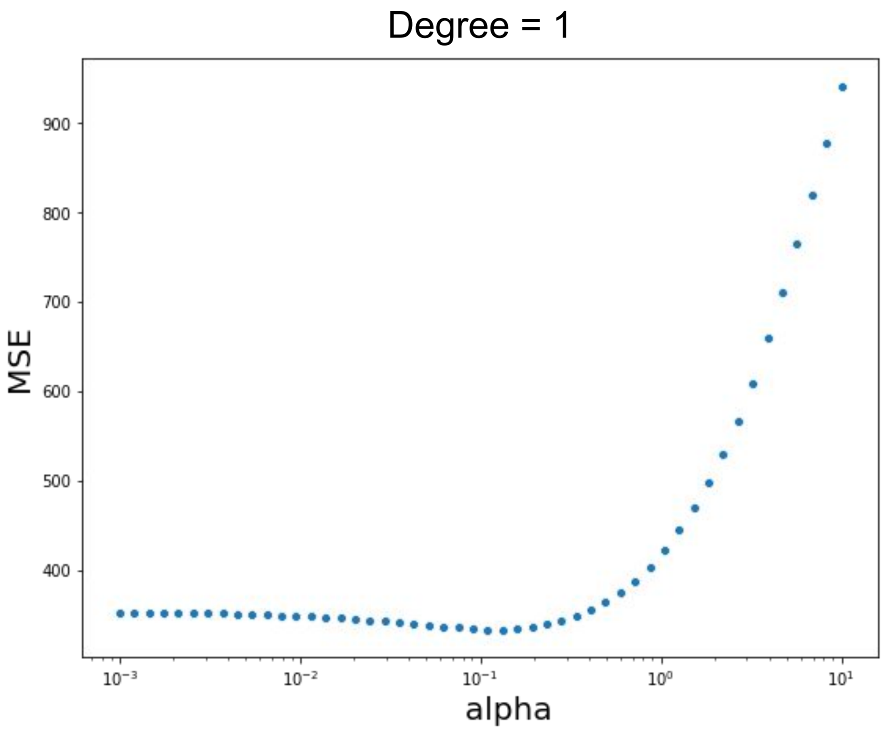
Statsmodel After FE

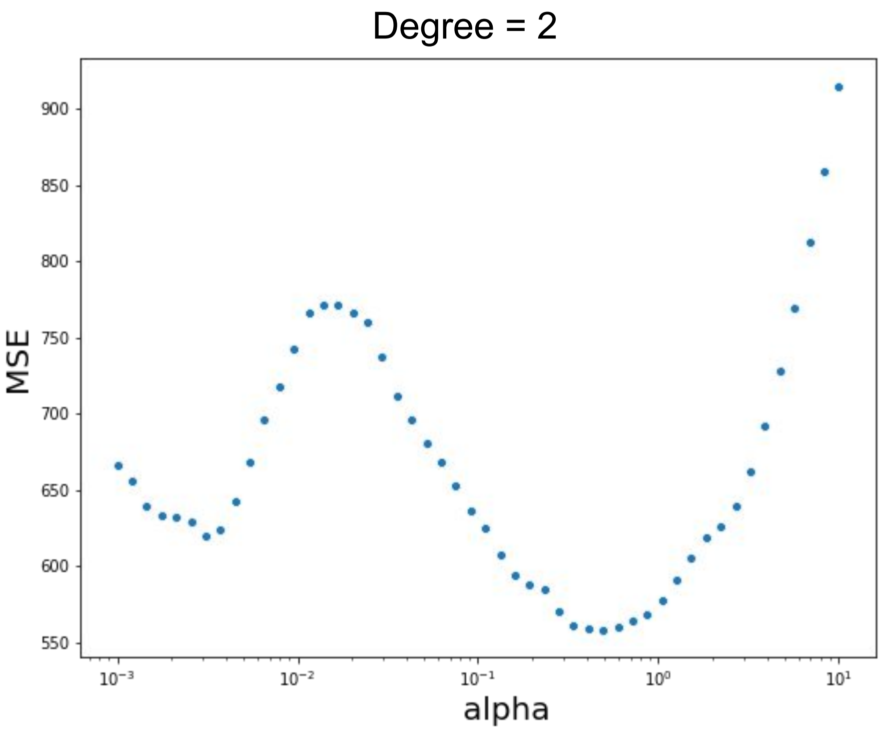
Train r^2 = 0.803

Test r^2 = 0.659



SKLearn Polynomial Regression

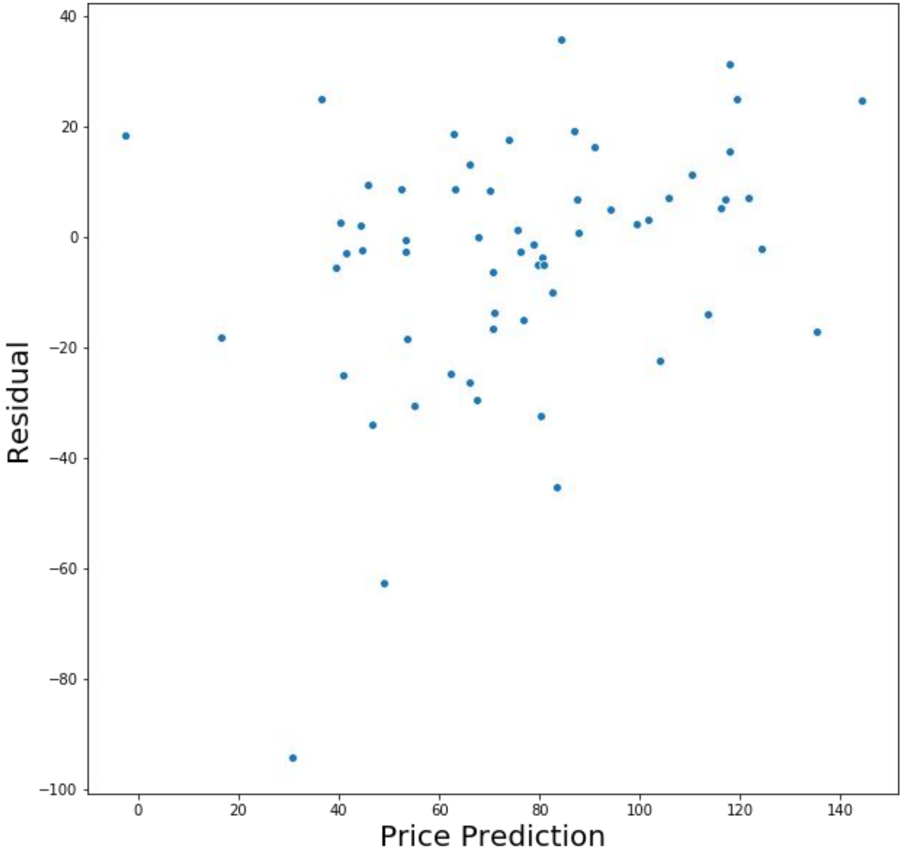




Final Residual Plot

Test r^2 = 0.640

Similar to SM!



Final Model Chosen

Statsmodel OLS Linear Regression with FE, Test r^2 = **0.659**

- Easy Ratio
- Intermediate Ratio
- Total Trail Length
- Altitude
- Lifts
- Total Snowfall
- 16 States

Best Lift Pass Value

Resort Name	% Lower than Model (%)
Kendall Mountain Silverton	230
Howelsen Hill Steamboat Springs	112
Wolf Creek	78
Soda Springs	70
Wenatchee Mission Ridge	68
Beaver Mountain	60
Marquette Mountain	56
Brian Head	51
Caberfae Peaks	44
Badger Pass	43

Best Lift Pass Value

Resort Name	$ Lower than Model ($)
Wolf Creek	47
Kendall Mountain Silverton	46
Wenatchee Mission Ridge	43
Brain Head	35
Loveland	32
Lake Louise	31
Beaver Mountain	29
Park City	29
Howelsen Hill Steamboat Springs	28
Purgatory Durango	27

Observations

- ➢ Small resorts have the best value
- ➢ California ski resorts are expensive!

Next Steps

- Need more data or features?
- More data cleaning
- Scrape more features
  - Proximity
- Engineering more features
  - Bang = (Resort size + Run Variety) / Price

Conclusion

Linear Regression OLS with Feature Engineering

Thank you!

Road to Fixing SM

	coef	std err	t	P>|t|
Altitude	-0.0019	0.003	-0.584	0.560
Easy	0.1636	0.167	0.983	0.327
Intermediate	0.3167	0.097	3.251	0.001
Difficult	0.3167	0.095	3.317	0.001
Resort_size	-5.5442	3.947	-1.405	0.161
Run_variety	16.7584	2.905	5.769	0.000
Lifts	11.1280	1.700	6.545	0.000
Intercept	-20.3857	9.443	-2.159	0.032
State[T.Alberta]	0.8503	8.898	0.096	0.924
State[T.Arizona]	16.7861	14.594	1.150	0.251
State[T.British Columbia]	1.8723	8.777	0.213	0.831
State[T.California]	27.7371	9.999	2.774	0.006
State[T.Colorado]	27.6001	11.321	2.438	0.015
State[T.Connecticut]	18.1575	11.224	1.618	0.107

Road to Fixing SM

	coef	std err	t	P>|t|
Altitude	-0.0019	0.003	-0.584	0.560
Easy	0.1636	0.167	0.983	0.327
Intermediate	0.3167	0.097	3.251	0.001
Difficult	0.3167	0.095	3.317	0.001
Resort_size	-5.5442	3.947	-1.405	0.161
Run_variety	16.7584	2.905	5.769	0.000
Lifts	11.1280	1.700	6.545	0.000
Intercept	-20.3857	9.443	-2.159	0.032
State[T.Alberta]	0.8503	8.898	0.096	0.924
State[T.Arizona]	16.7861	14.594	1.150	0.251
State[T.British Columbia]	1.8723	8.777	0.213	0.831
State[T.California]	27.7371	9.999	2.774	0.006
State[T.Colorado]	27.6001	11.321	2.438	0.015
State[T.Connecticut]	18.1575	11.224	1.618	0.107

Road to Fixing SM

Omnibus:	37.129	**Durbin-Watson:**	1.840
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	111.484
Skew:	0.512	**Prob(JB):**	6.19e-25
Kurtosis:	5.790	**Cond. No.**	1.00e+16

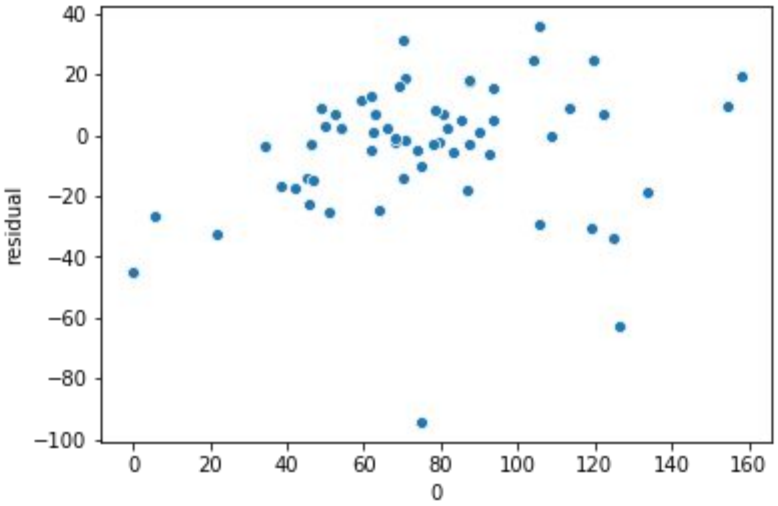
Road to Fixing SM

Omnibus:	37.129	**Durbin-Watson:**	1.840
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	111.484
Skew:	0.512	**Prob(JB):**	6.19e-25
Kurtosis:	5.790	**Cond. No.**	1.00e+16

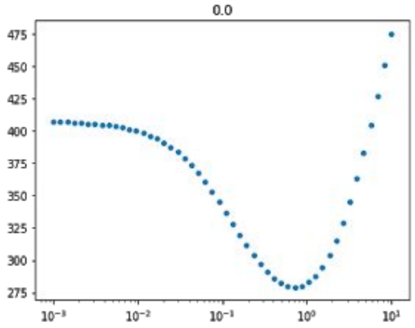
Road to Fixing SM

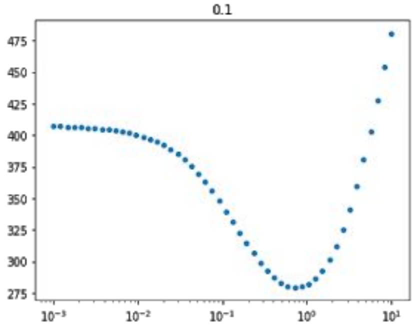
$r^2 = 0.659$

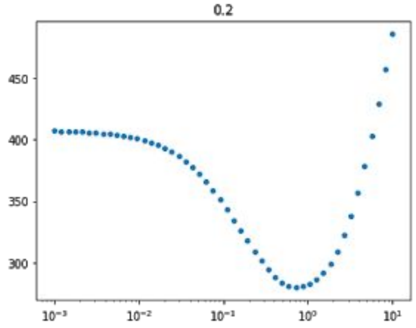
Omnibus:	1.419	**Durbin-Watson:**	1.869
Prob(Omnibus):	0.492	**Jarque-Bera (JB):**	0.939
Skew:	0.070	**Prob(JB):**	0.625
Kurtosis:	3.424	**Cond. No.**	1.84e+21

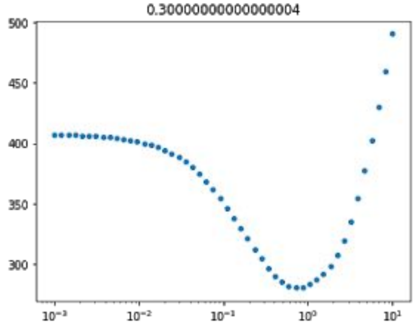


Optimizing SKLearn Elastic Net PR









Bang Formula Here!

Also: Bang = (Resort size + Run Variety) / Price