

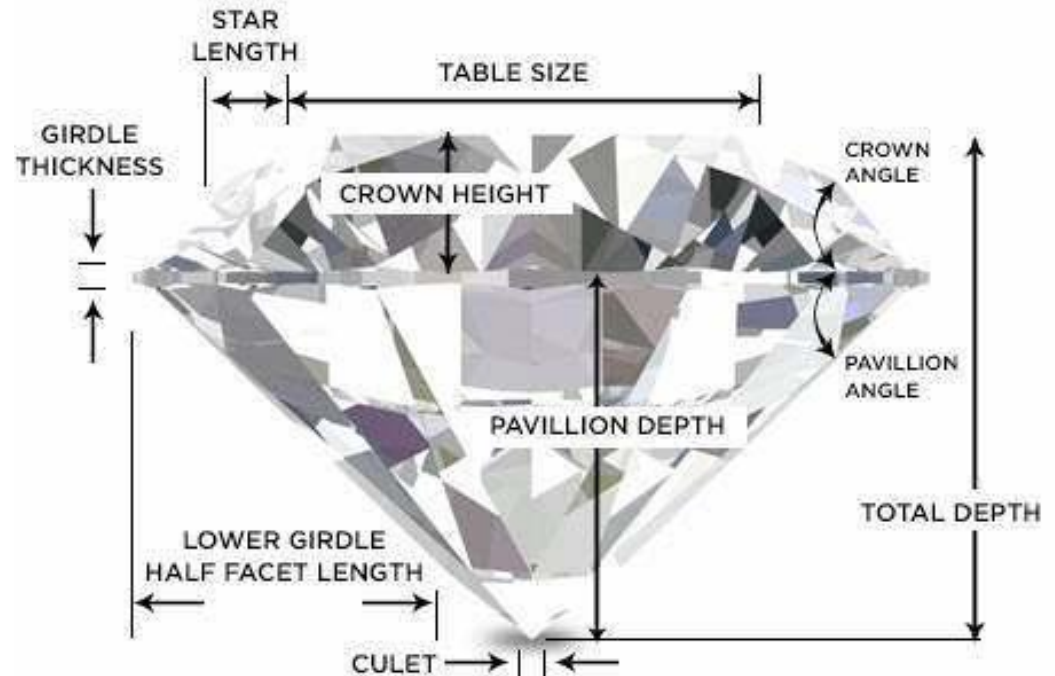
# Model Selection: Diamonds and Life Expectancy

Joe Read and Ashray Shetty



# Diamonds: The Data

- Target Price
- Carat
- Depth
- Table
- Individual Dimensions
- 54,000 rows



**ANATOMY OF A DIAMOND**

# Diamonds: The Models

	Model	Power	Hyper Parameters	Root Mean Squared Error (RMSE)	R-squared (training)	Adjusted R-squared (training)	R-squared (test)	Adjusted R-squared (test)	5-Fold Cross Validation
0	Multivariate Linear Regression	1.0	None	1486.771	0.859	0.859	0.859	0.859	0.854
1	Multivariate Polynomial Regression	2.0	None	1886.163	0.876	0.876	0.774	0.773	-42.203
2	Multivariate Polynomial Regression	3.0	None	1909.963	0.883	0.883	0.768	0.768	-452794.283
3	Lasso Linear Regression	1.0	Iterations: 100000 and Alpha: 3.685019951033...	1488.173	0.859	0.859	0.859	0.859	0.859
4	Lasso Polynomial Regression	2.0	Iterations: 100000 and Alpha: 352.612582868472	1537.234	0.849	0.849	0.850	0.850	-4.154
5	Lasso Polynomial Regression	3.0	Iterations: 1000000 and Alpha: 1653.54684242...	2582.831	0.571	0.570	0.575	0.575	-416.423
6	Ridge Linear Regression	1.0	Alpha: 92.0	1489.417	0.859	0.859	0.859	0.859	0.858
7	Ridge Polynomial Regression	2.0	Alpha: 95.0	1520.130	0.875	0.875	0.853	0.853	-11.428
8	Ridge Polynomial Regression	2.0	Alpha: 160.0	1482.440	0.874	0.874	0.860	0.860	-7.425
9	Ridge Polynomial Regression	2.0	Alpha: 8462.0	1446.352	0.865	0.865	0.867	0.867	0.321
10	Ridge Polynomial Regression	2.0	Alpha: 10000.0	1453.248	0.864	0.864	0.866	0.866	-4.147

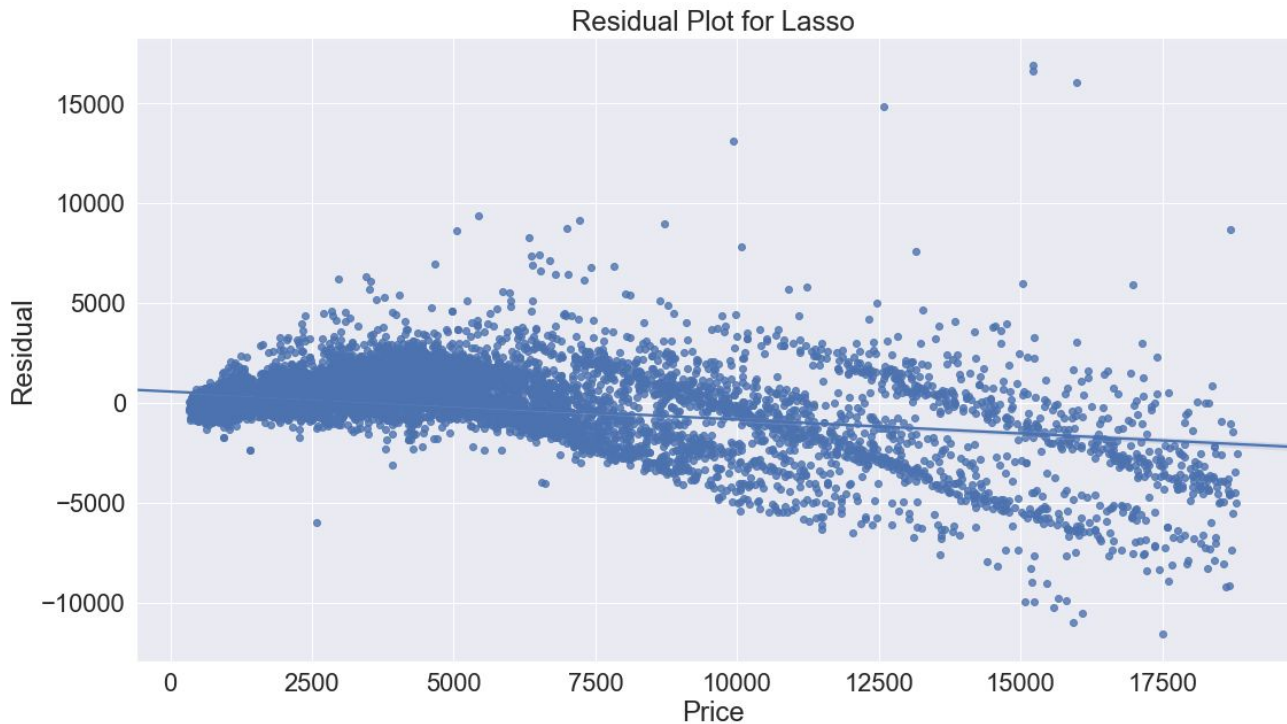
# Diamonds: The Best Model Residuals Plot

LassoCV linear

regression model

100,000 iterations

$\alpha$  value of 3.68



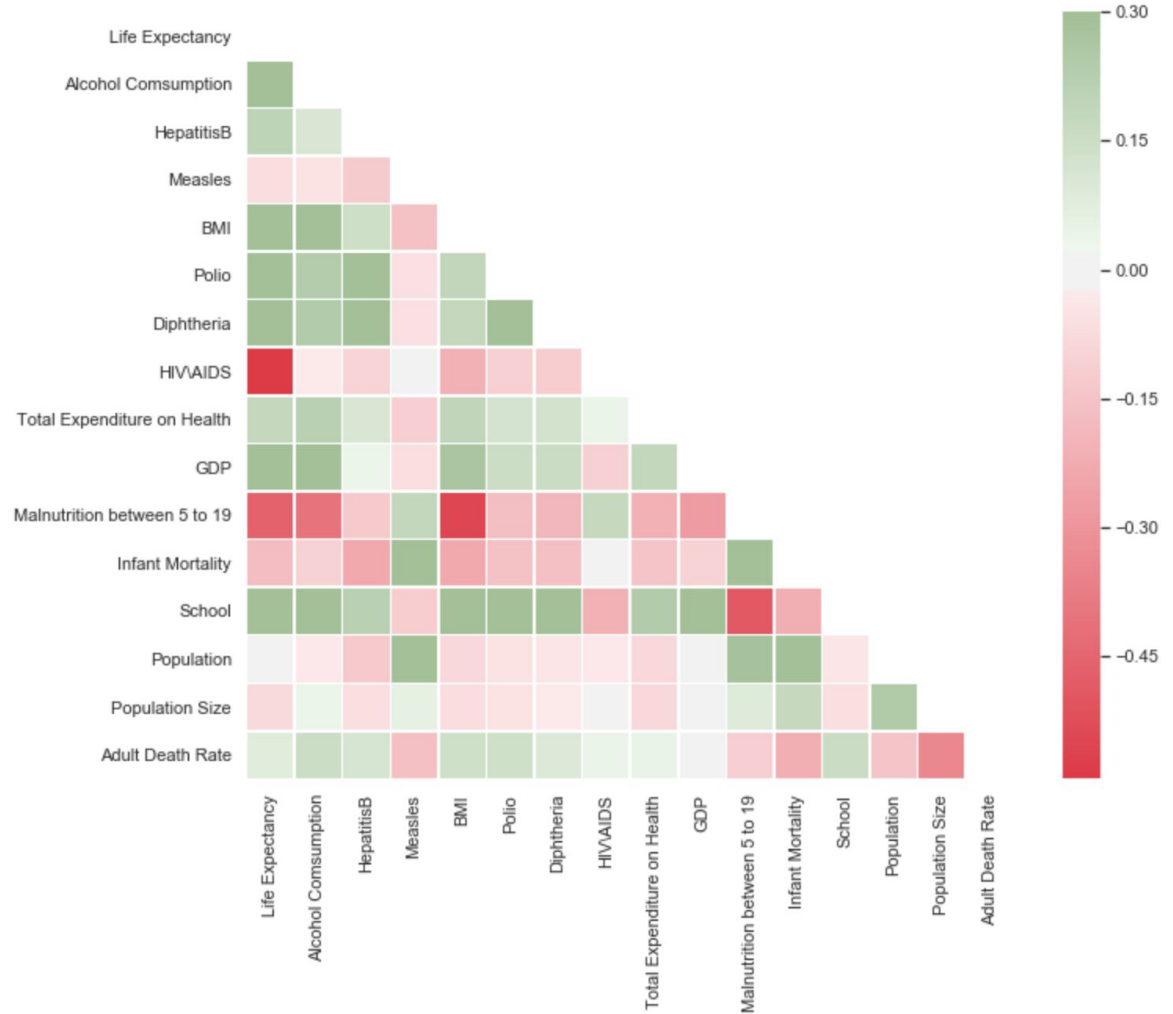
# Diamonds: Key Takeaways

**To make more money sell:**

1. Higher carat diamonds
2. Larger diamonds

# Life Expectancy

- BMI
- HIV/AIDS
- GDP
- School (Education)
- Adult Death Rate



# Life Expectancy: The Models

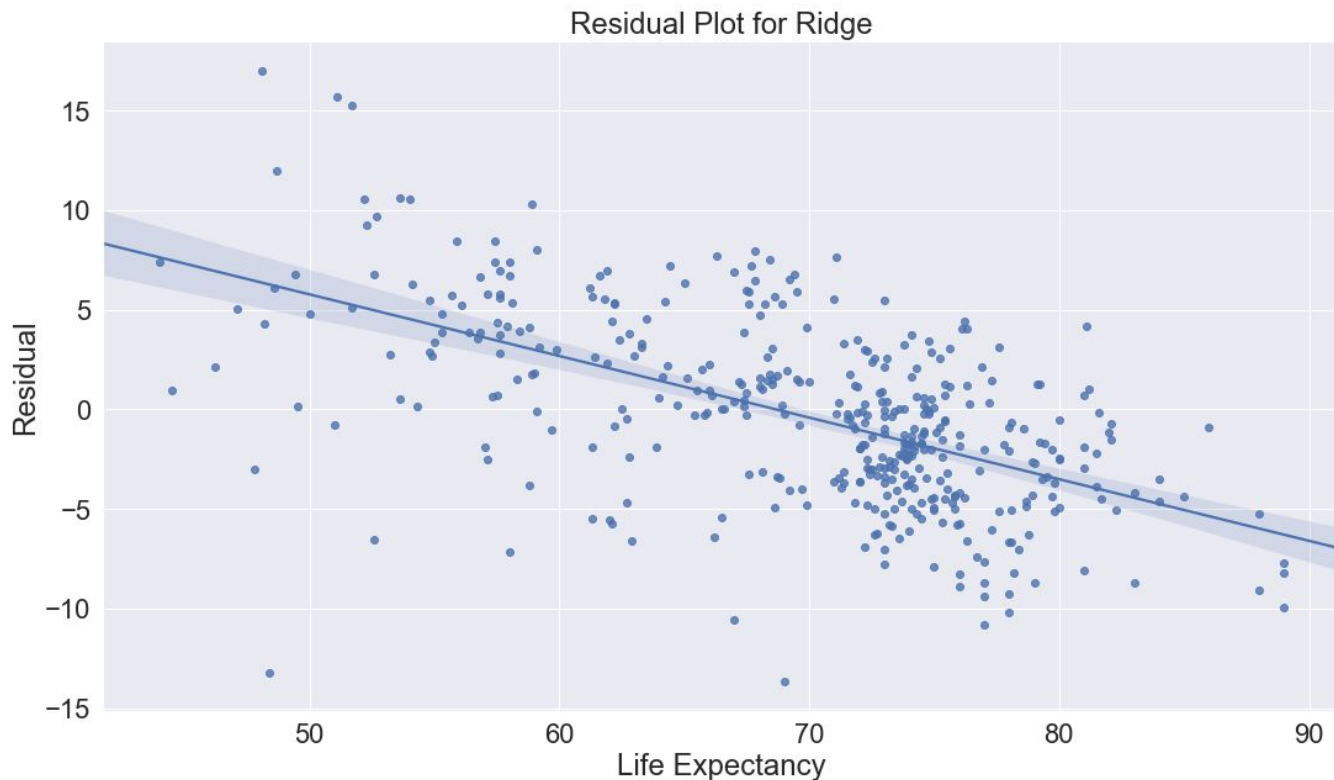
	Model	Power	Hyper Parameters	Root Mean Squared Error (RMSE)	R-squared (training)	Adjusted R-squared (training)	R-squared (test)	Adjusted R-squared (test)	5-Fold Cross Validation
0	Multivariate Linear Regression	1.0	None	4.646	0.751	0.750	0.711	0.708	0.746
1	Multivariate Polynomial Regression	2.0	None	3.996	0.831	0.830	0.787	0.784	0.823
2	Multivariate Polynomial Regression	3.0	None	3.734	0.854	0.853	0.814	0.811	0.819
3	Multivariate Polynomial Regression	4.0	None	3.540	0.877	0.877	0.832	0.830	-0.074
4	Multivariate Polynomial Regression	5.0	None	23.292	0.902	0.902	-6.252	-6.341	-75.038
5	Lasso Linear Regression	1.0	Iterations: 100000 and Alpha: 0.017256881441...	4.644	0.751	0.750	0.712	0.708	0.746
6	Lasso Polynomial Regression	2.0	Iterations: 100000 and Alpha: 0.019167177496...	3.986	0.830	0.829	0.788	0.785	0.823
7	Lasso Polynomial Regression	3.0	Iterations: 100000 and Alpha: 0.108389123248...	3.934	0.835	0.834	0.793	0.791	0.826
8	Lasso Polynomial Regression	4.0	Iterations: 100000 and Alpha: 0.667883866400...	4.477	0.773	0.772	0.732	0.729	0.755
9	Ridge Linear Regression	1.0	Alpha: 1.0	4.646	0.751	0.750	0.712	0.708	0.746
10	Ridge Linear Regression	1.0	Alpha: 5.0	4.643	0.751	0.750	0.712	0.708	0.746
11	Ridge Linear Regression	1.0	Alpha: 7.0	4.643	0.751	0.750	0.712	0.708	0.746
12	Ridge Polynomial Regression	2.0	Alpha: 6.0	3.996	0.830	0.830	0.787	0.784	0.823
13	Ridge Polynomial Regression	2.0	Alpha: 5.0	3.996	0.830	0.830	0.787	0.784	0.823
14	Ridge Polynomial Regression	3.0	Alpha: 6.0	3.755	0.851	0.851	0.811	0.809	0.828
15	Ridge Polynomial Regression	4.0	Alpha: 9.0	3.547	0.867	0.867	0.832	0.830	0.836

# Life Expectancy: The Best Model Residuals Plot

RidgeCV

4th order polynomial

$\alpha$  value of 9





# Life Expectancy: Key Takeaways

**Focus resources mainly towards:**

1. Increase HIV awareness
2. Promote education
3. Decrease Malnutrition



Thanks For  
Listening

Any Questions?

