

DS-ML Mini-Project - Diamonds

Rick Rongen - 316789 - 2502968

November 5, 2017

1 Dataset Choses

The dataset I'm planning to use is called Diamonds and can be find on Kaggle using the following link: <https://www.kaggle.com/shivam2503/diamonds>

This dataset describes all kinds of different diamonds with their quality (cut, color, carat and clarity) and size (x, y, z, dept and table). Also it describes the price of the diamond. I would like to use different machine learning techniques to calculate the price of a diamond given one or more of these measurements. Also I would like to be able to do the reverse and see what quality a diamond may be given it's size and price.

2 Example Data

In Pandas the cut, color and clarity column will be separated into separate columns or be represented as a number.

Example data:

Table 2.1: Data

id	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
2	0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
3	0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
4	0.29	Premium	I	VS2	62.4	58	334	4.2	4.23	2.63
5	0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
6	0.24	Very Good	J	VVS2	62.8	57	336	3.94	3.96	2.48
7	0.24	Very Good	I	VVS1	62.3	57	336	3.95	3.98	2.47
8	0.26	Very Good	H	SI1	61.9	55	337	4.07	4.11	2.53
9	0.22	Fair	E	VS2	65.1	61	337	3.87	3.78	2.49
10	0.23	Very Good	H	VS1	59.4	61	338	4	4.05	2.39
11	0.3	Good	J	SI1	64	55	339	4.25	4.28	2.73
12	0.23	Ideal	J	VS1	62.8	56	340	3.93	3.9	2.46
13	0.22	Premium	F	SI1	60.4	61	342	3.88	3.84	2.33
14	0.31	Ideal	J	SI2	62.2	54	344	4.35	4.37	2.71

2.1 Table Description

carat	Carat weight of the diamond
cut	The cut quality of the diamond. Quality in increasing order Fair, Good, Very Good, Premium, Ideal
color	Color of the diamond, with D being the best and J the worst
clarity	How obvious inclusions are within the diamond:(in order from best to worst, FL = flawless, I3= level 3 inclusions) FL,IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3
depth	Total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$
table	Width of top of diamond relative to widest point
price	The price of the diamond
x	Length mm
y	Width mm
z	Depth mm