

# «به نام خدا»

تکلیف ششم – سوال اول – مرضیه علیدادی – 9631983

(کد های مربوط، در دو فرمت py و .ipynb. ضمیمه شده اند.)

## 1.

(b) در این دیتاست missing value به صورت مقادیر 0 ذخیره شده اند. آن ها را جایگزین کردم.

پس از اینکه داده های عددی را normalize کردم، و داده های دسته ای را اینکد کردم، دیتاست به این صورت شد:

	Model	Release date	Max resolution	Low resolution	Effective pixels	Zoom wide (W)	Zoom tele (T)	Normal focus range	Macro focus range	Storage included	Weight (inc. batteries)	Dimensions	Price
0	0.0	0.839952	0.430701	0.269188	0.002001	0.015983	0.047949	0.029442	0.016824	0.001682	0.176655	0.039958	179.0
1	1.0	0.798663	0.511656	0.255828	0.000400	0.015190	0.045569	0.019987	0.003548	0.001599	0.167887	0.063158	179.0
2	2.0	0.705536	0.225772	0.660130	0.001678	0.015875	0.015875	0.017941	0.003131	0.000706	0.114957	0.037678	179.0
3	3.0	0.826099	0.476071	0.264484	0.001966	0.014464	0.014464	0.021017	0.003668	0.001653	0.134668	0.044139	269.0
4	4.0	0.826698	0.476416	0.264676	0.001967	0.017783	0.017783	0.020678	0.003670	0.016542	0.124067	0.052935	1299.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1033	1033.0	0.653406	0.668753	0.334376	0.000980	0.012409	0.037226	0.003265	0.003265	0.002612	0.104493	0.039185	62.0
1034	1034.0	0.651378	0.667011	0.333506	0.000977	0.011399	0.034197	0.026055	0.002931	0.005211	0.127019	0.037780	62.0
1035	1035.0	0.652955	0.668292	0.334146	0.000979	0.011421	0.031979	0.026105	0.003263	0.002611	0.110947	0.034916	62.0
1036	1036.0	0.593984	0.712425	0.356213	0.000891	0.010390	0.029091	0.023748	0.002968	0.004750	0.100927	0.031762	62.0
1037	1037.0	0.743383	0.594112	0.297056	0.000371	0.014110	0.014110	0.014853	0.007426	0.002971	0.066838	0.031934	129.0

1038 rows × 13 columns

(c) داده های پرت، همان 0 هایی بود که به عنوان missing value حذف کردم. تا حدی که در این زمینه سرچ کردم، بقیه ی داده ها می توانند معتبر باشند.

(d) دو پارامتر Release date و Zoom wide (W) به هم وابستگی دارند.

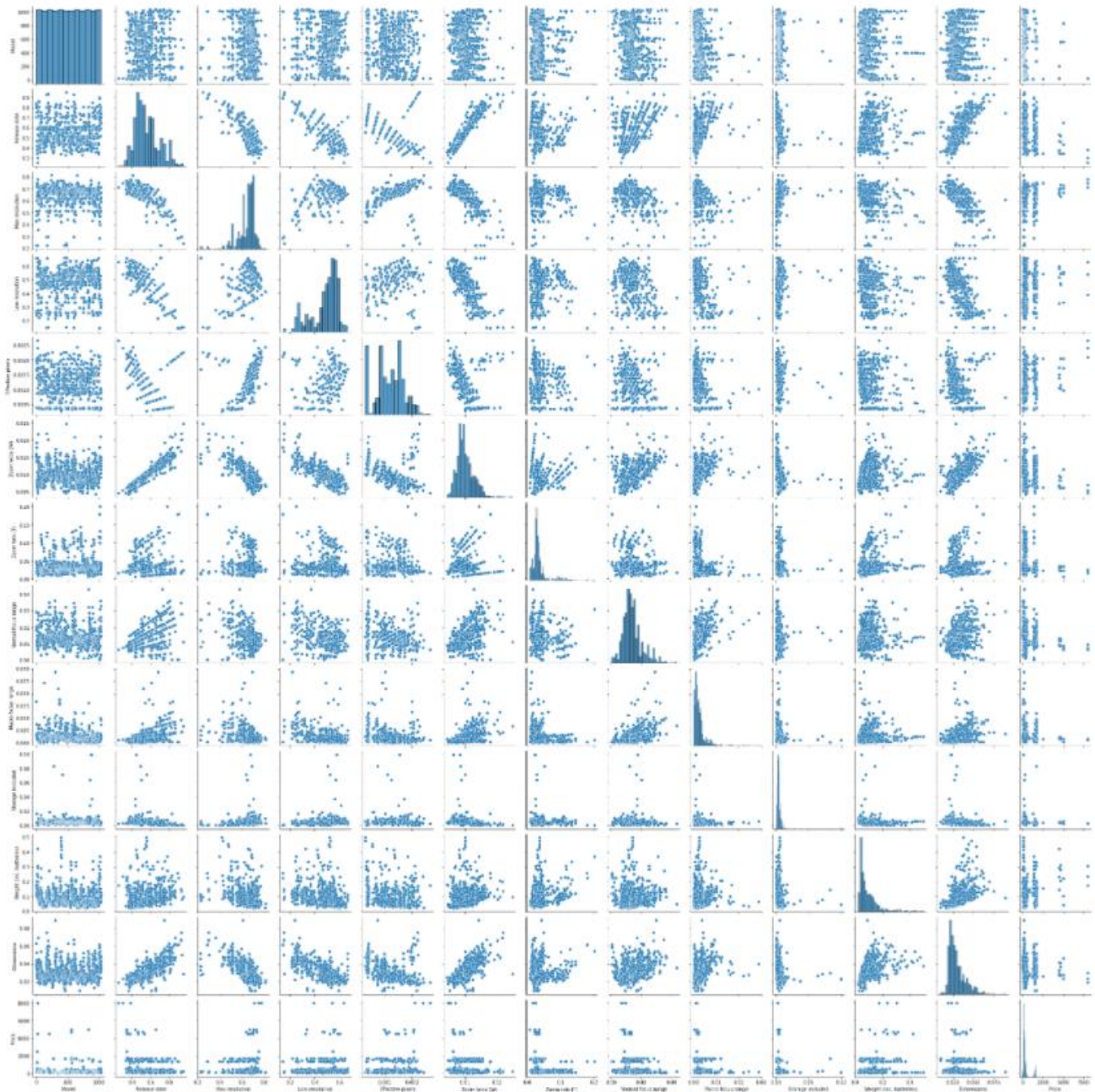
دو پارامتر Release date و Dimension به طور نسبی به هم وابستگی دارند.

دو پارامتر Release date و Max resolution به طور نسبی به هم وابستگی دارند.

دو پارامتر Low resolution و Zoom wide (W) به هم وابستگی دارند.

دو پارامتر Zoom wide (W) و Dimension به طور نسبی به هم وابستگی دارند.

دو پارامتر Dimension و Max resolution به طور نسبی به هم وابستگی دارند.



(g) یکبار مدلی با استفاده از OLS تولید کردم. که جدول زیر را حاصل شد:

با توجه به مقدار R-squared، این مدل 23 درصد تخمین دیتاست را پوشش میدهد.

ضرایب هر کدام از متغیرها نیز به صورت نشان داده شده در جدول زیر به دست آمده است: مثلاً متغیر Normal focus ضرایب range بیشترین تاثیر را در تخمین دارد، و تاثیرش به صورت عکس با Price است. و کمترین تاثیر برای Macro focus range است. بقیه ی coef ها نیز به همین صورت تفسیر می شوند.

یکبار دیگر با کلاس linearRegression این مدل را تولید کردم. و در بخش بعدی از آن استفاده می کنم.

#### OLS Regression Results

Dep. Variable:	Price	R-squared:	0.237
Model:	OLS	Adj. R-squared:	0.226
Method:	Least Squares	F-statistic:	21.14
Date:	Fri, 28 May 2021	Prob (F-statistic):	8.12e-41
Time:	00:40:01	Log-Likelihood:	-6570.4
No. Observations:	830	AIC:	1.317e+04
Df Residuals:	817	BIC:	1.323e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	4716.9816	1843.978	2.558	0.011	1097.490	8336.473
Model	0.0283	0.079	0.359	0.720	-0.127	0.183
Release date	-4152.5242	1327.154	-3.129	0.002	-6757.557	-1547.491
Max resolution	-1821.0261	1027.152	-1.773	0.077	-3837.193	195.141
Low resolution	-2350.5286	892.854	-2.633	0.009	-4103.086	-597.971
Effective pixels	3.419e+04	8.37e+04	0.409	0.683	-1.3e+05	1.98e+05
Zoom wide (W)	1.721e+04	2.56e+04	0.672	0.502	-3.3e+04	6.75e+04
Zoom tele (T)	-3922.0813	1043.874	-3.757	0.000	-5971.073	-1873.090
Normal focus range	-6612.1906	4638.180	-1.426	0.154	-1.57e+04	2491.963
Macro focus range	1.046e+04	1.06e+04	0.986	0.325	-1.04e+04	3.13e+04
Storage included	-3639.3744	3655.100	-0.996	0.320	-1.08e+04	3535.118
Weight (inc. batteries)	4831.0650	667.142	7.241	0.000	3521.551	6140.579
Dimensions	-3279.4045	5523.847	-0.594	0.553	-1.41e+04	7563.200

Omnibus:	709.756	Durbin-Watson:	2.089
Prob(Omnibus):	0.000	Jarque-Bera (JB):	29458.808
Skew:	3.621	Prob(JB):	0.00
Kurtosis:	31.273	Cond. No.	2.15e+06

(h) مقادیر مورد نظر به شکل زیر به دست آمد:

Coef های بدست آمده، مشابه مدل OLS بخش قبل به دست آمدند. و همان تحلیل را می توان اینجا هم داشت. Intercept نشان می دهد که خط رگرسیون تخمینی، از چه نقطه ای از محور y عبور می کند. درواقع همان coef مربوط به const در OLS regression است. و اینجا نیز مقداری مشابه با قبل را حاصل شده است.

```

: print('intercept:', model.intercept_)

intercept: 4716.981592193721

: print('coef:', model.coef_)

coef: [ 0.00000000e+00  2.83356913e-02 -4.15252420e+03 -1.82102606e+03
 -2.35052859e+03  3.41918509e+04  1.72128719e+04 -3.92208127e+03
 -6.61219058e+03  1.04611132e+04 -3.63937444e+03  4.83106502e+03
 -3.27940454e+03]
```