Regresja ceny telewizorów

Mateusz Jakubczak

Wydział Zarządzania, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

II rok Informatyka i Ekonometria

22 lutego 2021

Streszczenie

Celem projektu jest zbadanie, jakie parametry telewizora wpływają na jego cenę oraz w jakim stopniu. Główną hipotezą badawczą jest stwierdzenie czy cena telewizora jest wprost proporcjonalnie zależna od przekątnej ekranu. Drugą hipotezą jest stwierdzenie czy telewizor używany jest tańszy od nowego. Trzecią hipotezą jest stwierdzenie czy jakość ma wpływ na cenę telewizora.

Spis treści

1 Opis Danych i ich pochodzenie

1.1 Źródło pochodzenia

Źródłem pochodzenia danych są oferty na stronie allegro.pl dotyczące sprzedaży telewizorów, dane zostały zebrane w dniu 31-05-2020, wszystkie linki znajdują się w pliku łinki.csv". Możliwe, że cześć linków jest już nieaktywna, dlatego wszystkie informacje z tych linków zostały zapisane w pliku "parametry.csv".

1.2 Sposób pobierania

Sposób pobierania danych to webscaping informacji z linku za pomocą biblioteki BeautifulSoup w języku programowania Python. Metoda to znalezienie znacznika na cenę telewizora i zapis tej wartości. Potem wyszukuję tablice z parametrami telewizora i dla zbioru moich zmiennych, które wcześniej wybrałem. Jeśli w tabeli parametr zostanie znaleziony, to jego wartość jest zapisywana. W przypadku braku znalezienia określonego parametru zostaje przypisane -1 jako wartość parametru.

Parametry

Stan:	Powystawowy	Klasa efektywności energetycznej:	A
Faktura:	Wystawiam fakturę VAT	Pobór mocy w trybie czuwania:	0.5 W
Marka:	LG	Roczne zużycie energii:	212 kWh
Model:	OLED65B6V	Pobór mocy:	145 W
Kod producenta:	OLED65B6V	Zakrzywiony ekran (Curved):	Nie
Typ telewizora:	OLED	Standard VESA:	300 x 200
Kolor:	czarny	Szerokość produktu z podstawą:	145.1 cm
Przekątna ekranu (cale):	65"	Wysokość produktu z podstawą:	88.2 cm
Format HD:	4K UHD	Głębokość produktu z podstawą:	22.5 cm
Rozdzielczość ekranu (px):	3840 x 2160	Waga z podstawą:	25.4 kg
Smart TV:	inny system producenta	Szerokość produktu:	145.1 cm
Technologia 3D:	nie	Wysokość produktu:	83.8 cm
Technologia HDR:	Tak	Głębokość produktu:	4.86 cm
Liczba złączy HDMI:	4	Waga produktu:	22.1 kg
Liczba złączy USB:	3	Załączone wyposażenie:	pilot
Łączna moc głośników:	40 W		
Komunikacja:	Bluetooth, Wi-Fi Direct, Wi-Fi		
Złącza:	RJ-45, wyjście słuchawkowe, złącze antenowe, złącze komponentowe		
Tuner:	DVB-C, DVB-S2, DVB-T2		

Rysunek 1: Przykładowa tabela z parametrami

Wybrane przez mnie parametry to

- Stan
- Typ telewizora
- Marka
- Technologia 3D

- Przekątna ekranu
- Format HD
- Rozdzielczość ekranu
- Liczba złączy HDMI
- Technologia HDR
- Klasa efektywności
- Pobór mocy
- Waga produktu

2 Statystyki Opisowe

2.1 Preprocesing

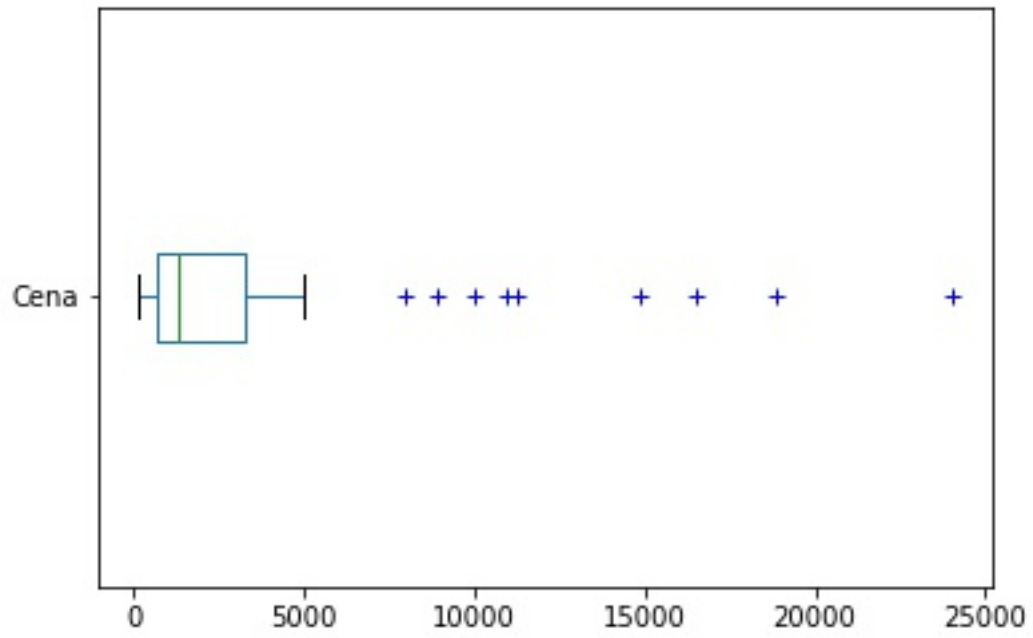
Wszystkie dane zostały wcześniej przygotowane, czyli nadano im odpowiedni typ oraz format, żeby były gotowe do użycia w modelu. Ważniejsza zmiana to przetransformowanie „rozdzielczość ekranu" oraz „format HD" w zmienną jakość. Każda marka z częstością występowania poniżej mediany została zamieniona na inną, żeby ograniczyć ilość kategorii o małej populacji w zmiennej „marka". Usunięta na starcie została zmienna „waga" ze względu na dużą liczbę brakujących danych.

2.2 Zmienna endogeniczna

Zmienną endogeniczną jest cena danego telewizora.

	Cena
count	81.00
mean	2894.43
std	4362.52
min	150.00
25%	649.00
50%	1299.00
75%	3299.00
max	24000.00

Mamy wysoką wartość odchyleń standardowych w porównaniu do wartości średniej, mówi to nam o dużej liczbie obserwacji odstających, które należałoby usunąć.



Rysunek 2: Wykres pudełkowy dla ceny

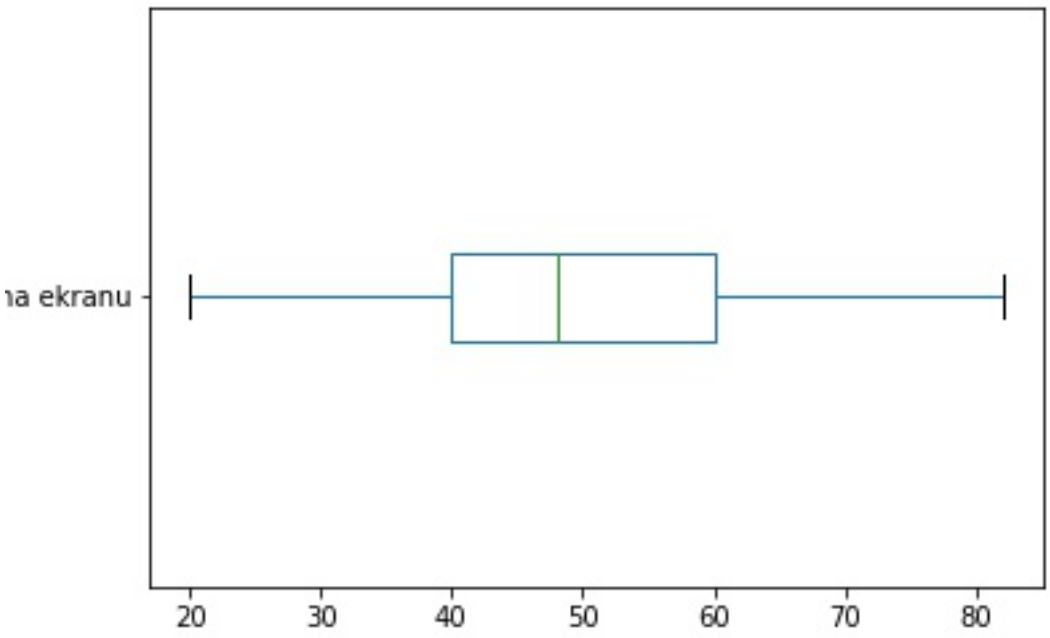
2.3 Zmienne ciągłe

2.3.1 Przekątna ekranu

Zmienna ta opisuje przekątną ekranu telewizora podaną w calach.

	Przekątna ekranu
count	81.00
mean	48.93
std	15.90
min	20.00
25%	40.00
50%	48.000
75%	60.00
max	82.00

Nie widać tu obserwacji odstających i wygląda, że wszystko z tą zmienną jest w porządku.



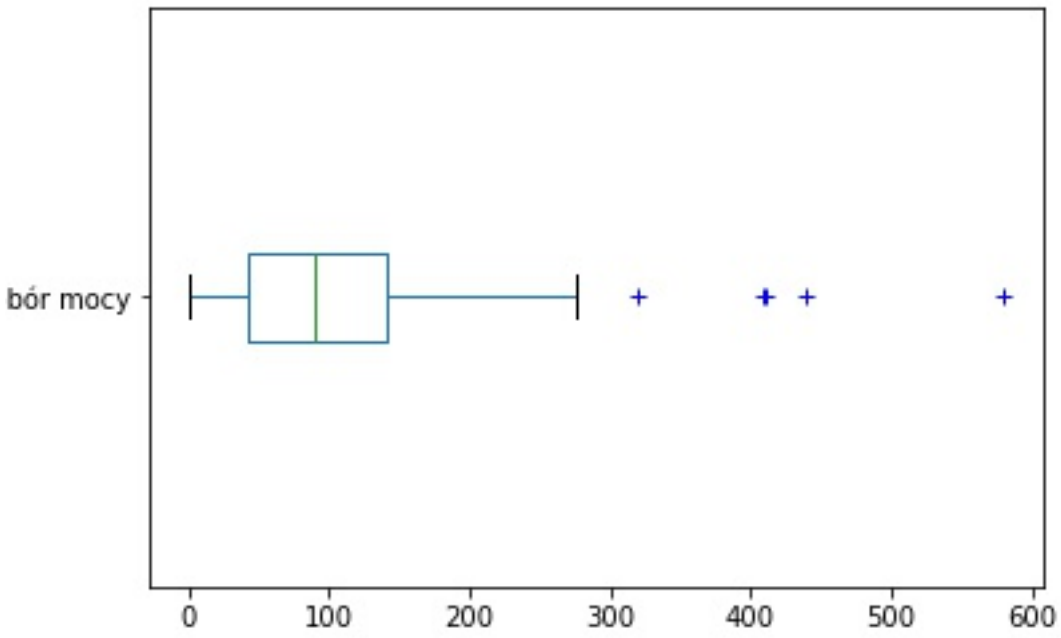
Rysunek 3: Wykres pudełkowy przekątnej ekranu

2.3.2 Pobór mocy

Zmienna ta opisuje ile mocy w wat pobiera telewizor podczas normalnej pracy.

	Pobór mocy
count	81.000000
mean	114.518519
std	107.063429
min	1.000000
25%	42.000000
50%	90.000000
75%	141.000000
max	579.000000

Mamy tutaj podobną sytuację co w cenie: dużo obserwacji odstających skierowanych w jedną stronę oraz wysokie odchylenie standardowe. Sugeruje to usunięcie obserwacji odstających.



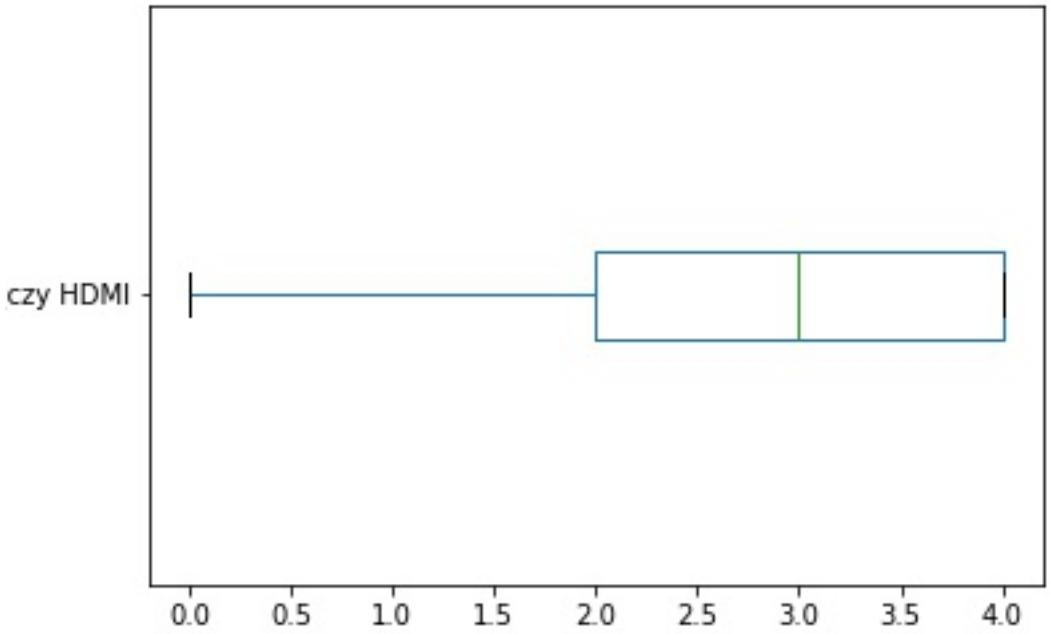
Rysunek 4: Wykres pudełkowy poboru mocy

2.3.3 złącza HDMI

Zmienna ta opisuje ile złączy HDMI posiada dany telewizor

	Liczba złączy HDMI
count	81.000000
mean	2.975309
std	0.987108
min	0.000000
25%	2.000000
50%	3.000000
75%	4.000000
max	4.000000

Zmienna ta nie powinna być obrazowana jako ciągła, ponieważ mamy tylko 5 możliwych wartości, ale możemy zobaczyć, że większość telewizorów ma 2, 3 lub 4 złącza HDMI.

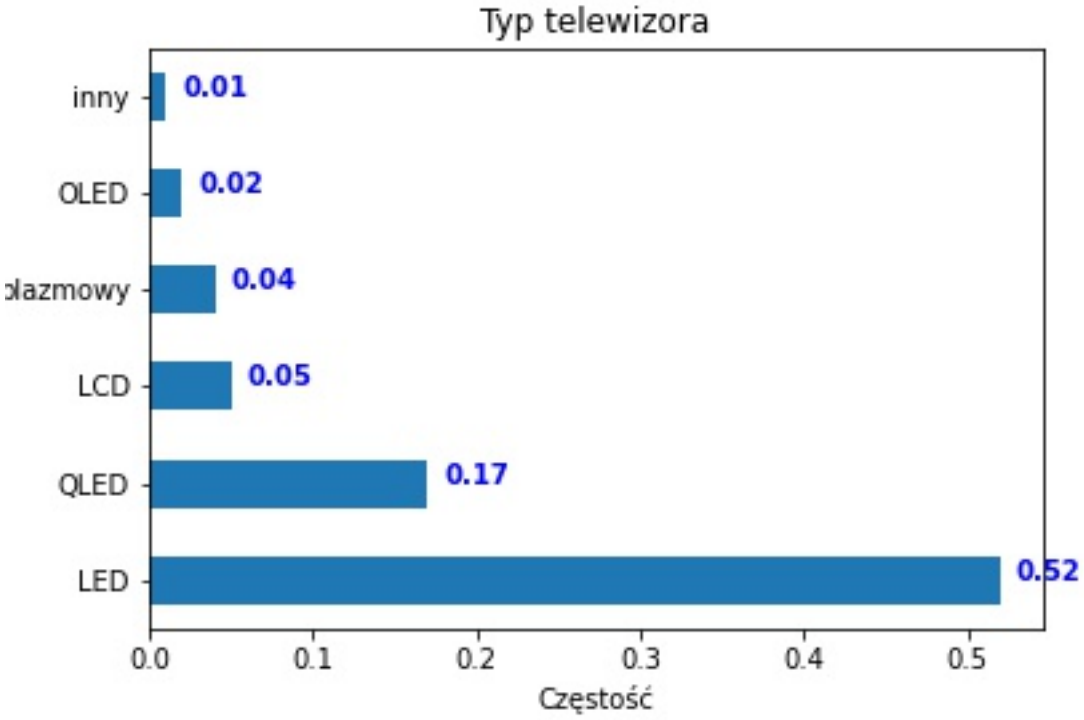


Rysunek 5: Wykres pudełkowy HDMI

2.4 Zmienne kategoryczne

2.4.1 Typ Telewizora

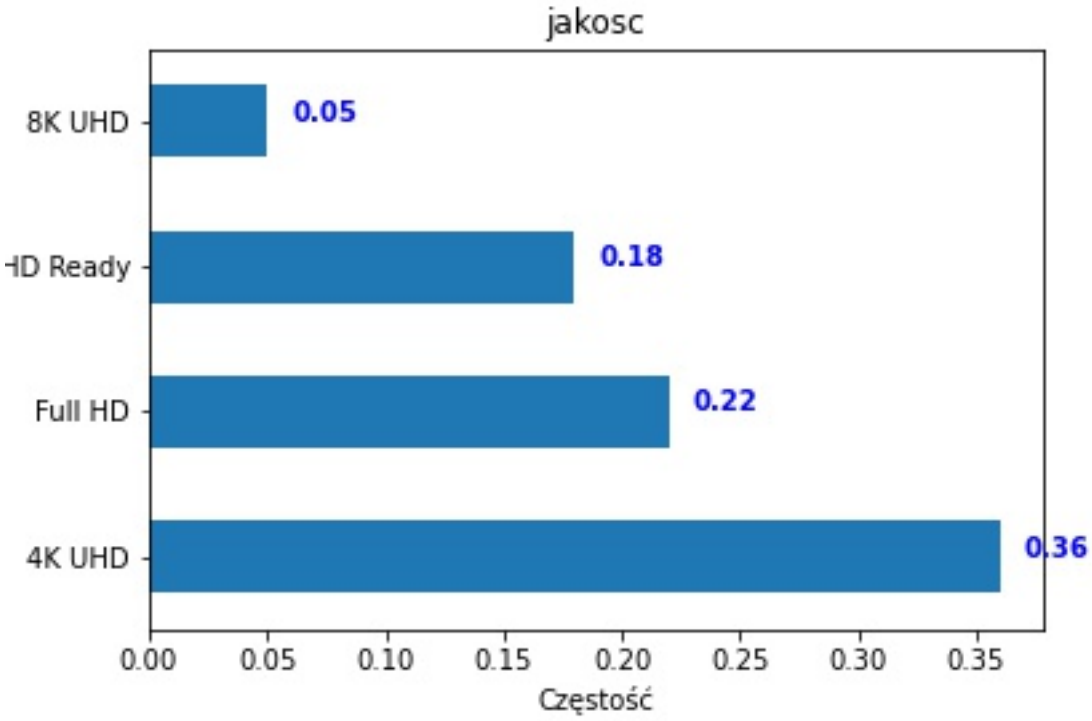
Zmienna ta opisuje, jaki jest typ wyświetlacza w telewizorze.



Rysunek 6: Częstość występowania poszczególnych typów

2.4.2 Jakość

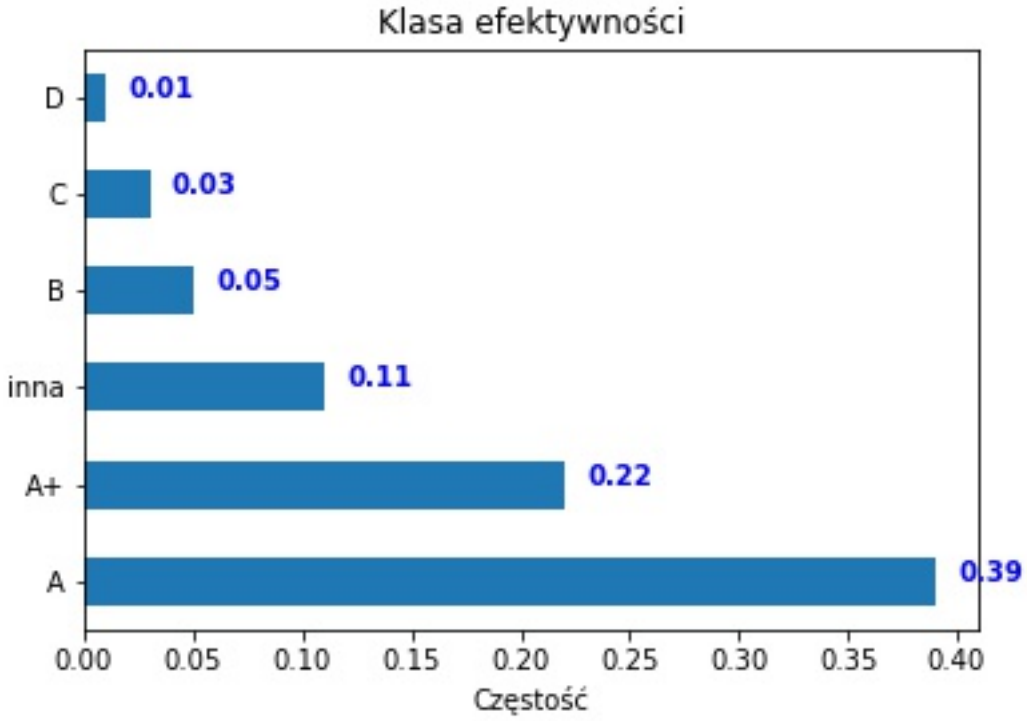
Zmienna ta składa się połączenia dwóch zmiennych(rozdzielczości ekranu oraz formatu obrazu) w procesie preprocesingu.



Rysunek 7: Częstość występowania poszczególnych jakości

2.4.3 Klasa efektywności

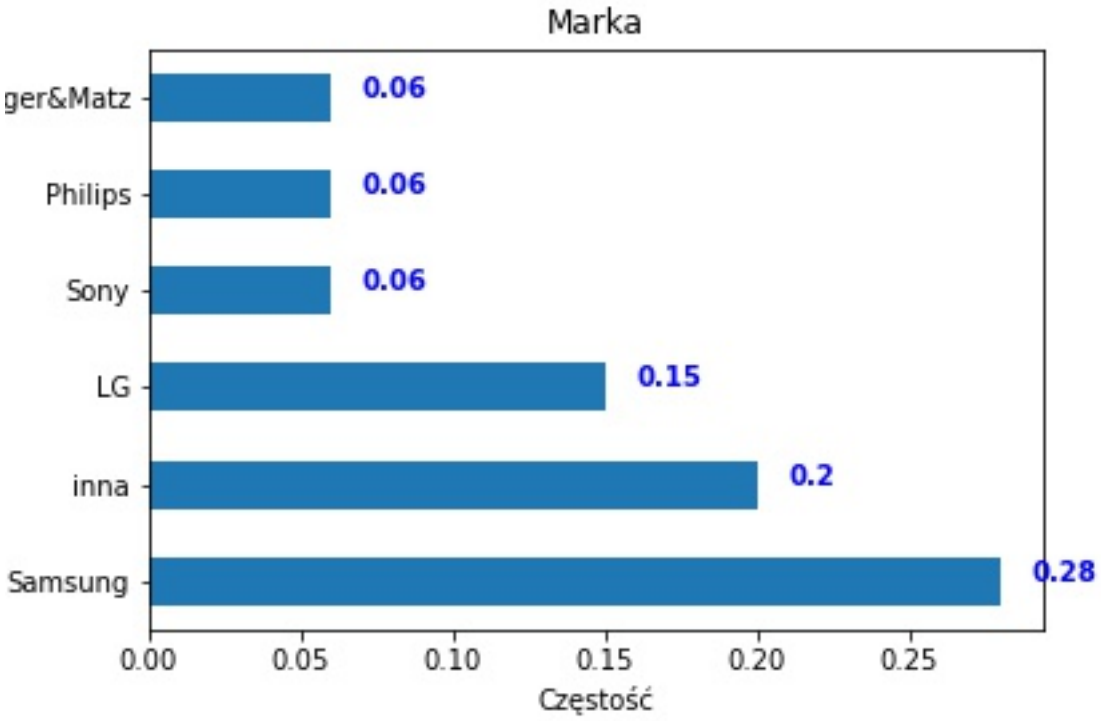
Zmienna ta opisuje, jakiej klasy efektywności jest telewizor.



Rysunek 8: Częstość występowania poszczególnych klas

2.4.4 Marka

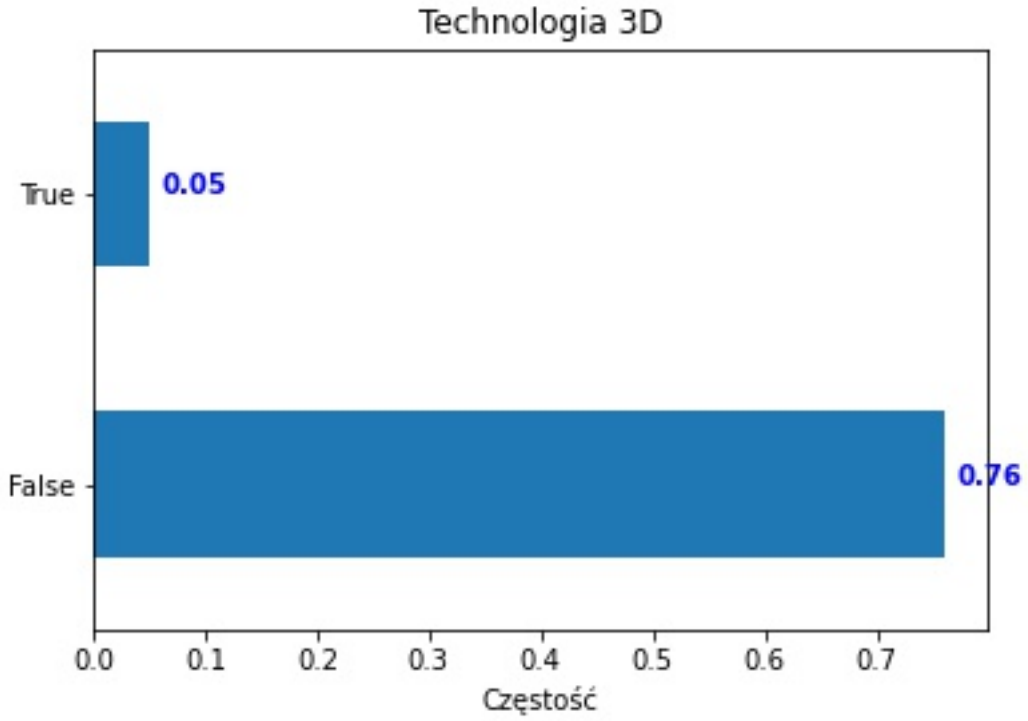
Zmienna ta opisuje, jakiej marki jest telewizor.



Rysunek 9: Częstość występowania poszczególnych marek

2.4.5 Technologia 3D

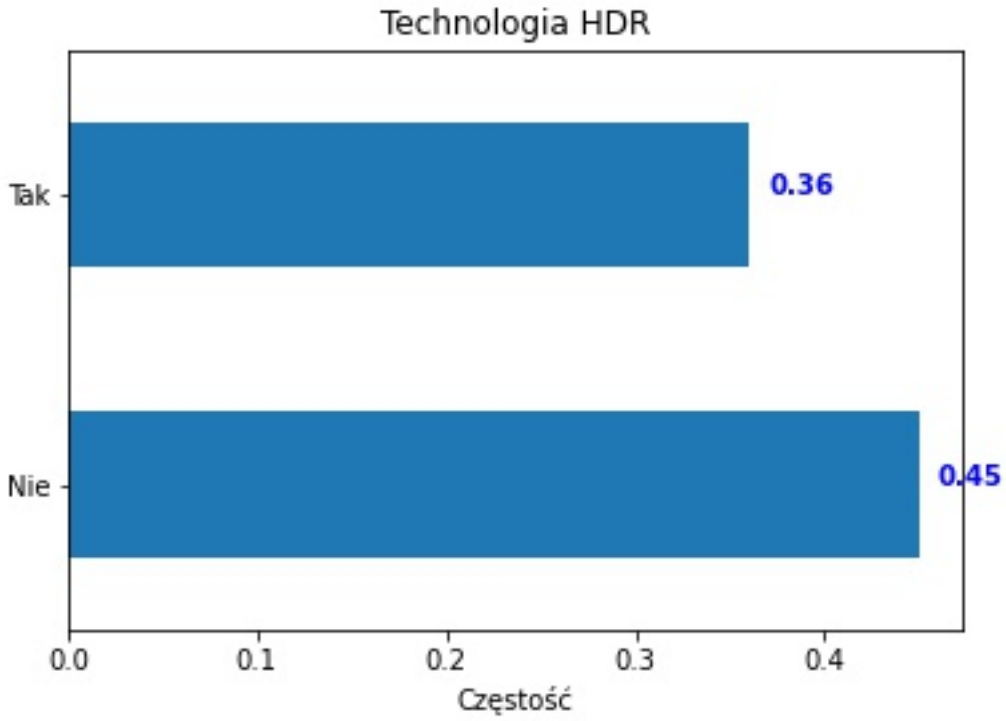
Zmienna ta mówi, czy telewizor może wyświetlać obraz w 3D.



Rysunek 10: Częstość występowania technologii 3D

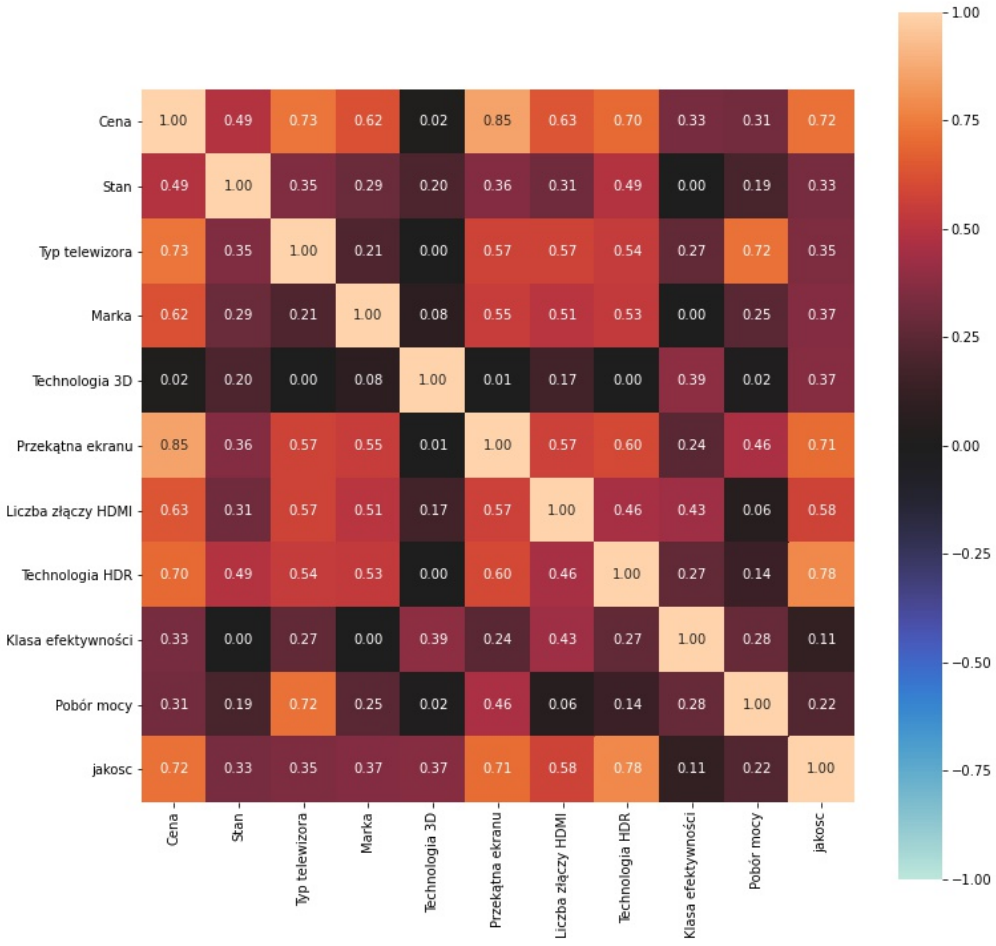
2.4.6 Technologia HDR

Zmienna ta mówi, czy telewizor posiada technologie HDR. Technologia HDR daje lepszą jakość obrazu bez zmiany rozdzielczości.



Rysunek 11: Częstość występowania technologii HDR

2.5 Macierz korelacji



Rysunek 12: Macierz korelacji

Macierz ta przedstawia korelacje Pearsona dla zmiennych ciągłych „Correlation Ratio" dla zmiennych kategoryczna z ciągłą Korelacje V Camera dla zmiennych kategorycznych. Z macierzy wynika duża korelacja między ceną, a przekątną, jakością oraz typem telewizora oraz brak korelacji technologii 3D z pozostałymi, co sugeruje niską istotność tego parametru. Inne ciekawe korelacje to między typem telewizora, a poborem mocy oraz między jakością, a przekątną, co może sugerować, że jedna z tych zmiennych może nie dostarczyć dotykowych informacji w modelu. Na podstawie tej macierzy możemy przewidzieć, że w modelu powinny znaleźć się zmienne: jakość lub przekątna (jedno z nich z uwagi na wysoką korelację między nimi), typ telewizora oraz marka.

3 Budowa modelu liniowego

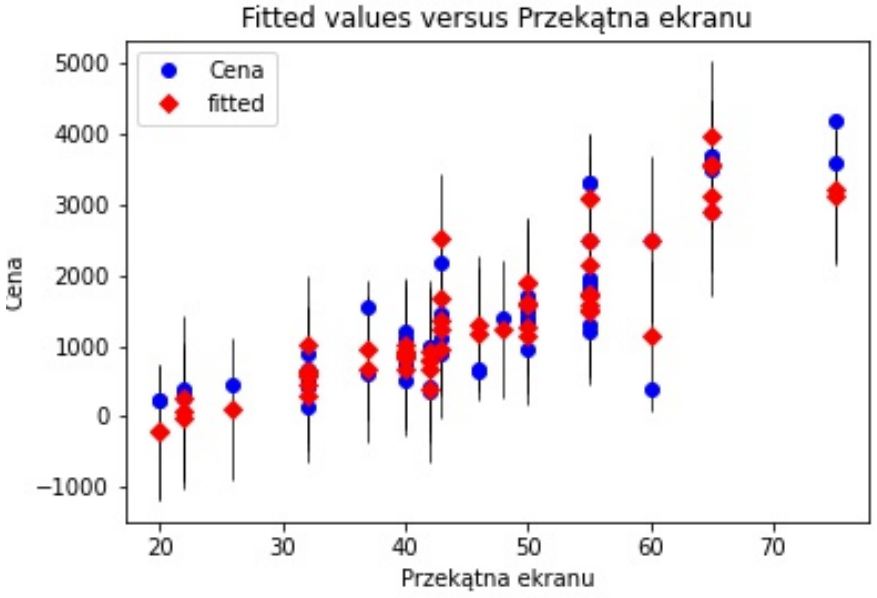
3.1 Wstępne szacowanie

Szacuję pierwszy model z usuniętymi obserwacjami odstającymi, które zostały usunięte na podstawie ceny. Zamienne kategoryczne są reprezentowane jako dummy.

Dep. Variable:	Cena	**R-squared:**	0.907
Model:	OLS	**Adj. R-squared:**	0.856
Method:	Least Squares	**F-statistic:**	17.91
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	1.37e-15
Time:	18:13:48	**Log-Likelihood:**	-498.25
No. Observations:	69	**AIC:**	1047.
Df Residuals:	44	**BIC:**	1102.
Df Model:	24		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-793.2401	502.552	-1.578	0.122	-1806.068	219.588
Przekątna ekranu	46.2102	8.122	5.690	0.000	29.842	62.579
Liczba złączy HDMI	-2.8332	88.866	-0.032	0.975	-181.932	176.265
Pobór mocy	0.4069	1.287	0.316	0.753	-2.187	3.001
Stan_Po zwrocie	-379.1746	348.611	-1.088	0.283	-1081.755	323.406
Stan_Powystawowy	391.6052	327.285	1.197	0.238	-267.995	1051.205
Stan_U¿ywany	-337.6844	190.170	-1.776	0.083	-720.946	45.577
Typ telewizora_LED	-41.0985	246.994	-0.166	0.869	-538.882	456.686
Typ telewizora_OLED	-256.0298	581.007	-0.441	0.662	-1426.973	914.913
Typ telewizora_QLED	908.7689	365.592	2.486	0.017	171.966	1645.572
Typ telewizora_inny	391.9251	509.150	0.770	0.446	-634.199	1418.049
Typ telewizora_plazmowy	-457.9512	437.322	-1.047	0.301	-1339.317	423.414
Marka_LG	442.5958	282.264	1.568	0.124	-126.271	1011.462
Marka_Philips	99.4449	356.995	0.279	0.782	-620.030	818.920
Marka_Samsung	333.8736	296.659	1.125	0.267	-264.002	931.750
Marka_Sony	292.2040	321.864	0.908	0.369	-356.471	940.879
Marka_inna	86.5933	228.325	0.379	0.706	-373.566	546.753
Technologiia 3D_True	123.1922	296.478	0.416	0.680	-474.320	720.705
Technologiia HDR_Tak	62.7285	252.602	0.248	0.805	-446.357	571.814
x _A+	27.3171	134.129	0.204	0.840	-243.002	297.636
Klasa efektywnosci_B	-263.0567	491.918	-0.535	0.596	-1254.451	728.338
Klasa efektywnosci_C	782.4736	587.331	1.332	0.190	-401.214	1966.161
Klasa efektywnosci_D	-1.208e-12	6.09e-12	-0.198	0.844	-1.35e-11	1.11e-11
Klasa efektywnosci_inna	-190.0021	210.011	-0.905	0.371	-613.251	233.246
jakosc_Full HD	-120.3698	227.939	-0.528	0.600	-579.750	339.010
jakosc_HD Ready	-110.9307	263.460	-0.421	0.676	-641.900	420.038

Omnibus:	2.845	**Durbin-Watson:**	2.125
Prob(Omnibus):	0.241	**Jarque-Bera (JB):**	2.030
Skew:	0.329	**Prob(JB):**	0.362
Kurtosis:	3.523	**Cond. No.**	1.23e+16



Rysunek 13: Wykres dopasowania modelu do ceny na postawie przekątnej ekranu

Wysokie wartości p-value dla zmiennych wskazują na ich niepotrzebność w modelu. Na podstawie wartości statystki Durbin-Watson nie występuje autokorelacja. P-value Jarque-Bera wskazuje, że nie mamy spełnionego założenia o normalności reszt. Obecna postać modelu jest niepoprawna i musimy ją odrzucić.

	Zmienna	VIF
0	const	101.425957
1	Przekątna ekranu	4.551214
2	Liczba złączy HDMI	2.966977
3	Pobór mocy	3.756607
4	Stan_Po zwrocie	1.373651
5	Stan_Powystawowy	1.788984
6	Stan_Używany	3.623989
7	Typ telewizora_LED	5.042987
8	Typ telewizora_OLED	1.936244
9	Typ telewizora_QLED	6.088026
10	Typ telewizora_inny	1.486924
11	Typ telewizora_plazmowy	4.194356
12	Marka_LG	4.596809
13	Marka_Philips	3.440026
14	Marka_Samsung	7.274918
15	Marka_Sony	3.303129
16	Marka_inna	4.309457
17	Technologiia 3D_True	2.372596
18	Technologiia HDR_Tak	5.812804
19	Klasa efektywnosci_A+	1.441642
20	Klasa efektywnosci_B	1.387976
21	Klasa efektywnosci_C	1.978623
22	Klasa efektywnosci_D	NaN
23	Klasa efektywnosci_inna	2.194941
24	jakosc_Full HD	4.531532
25	jakosc_HD Ready	5.374781

Na podstawie statystyki VIF nie mamy rażącej współliniowości między zmiennymi, wiec nie ma podstaw do odrzucenia zmiennych na podstawie wysokiej współliniowości.

3.2 Iteratywne poprawnie modelu

Do wyboru zmiennych używam metody krokowej wstecznej, przyjęta wartość α to $0,1$. Kolejne zmienne usunięte ze zbioru danych to:
Liczba złączy HDMI, z p-value równe 0.97
jakość, z p-value równe 0.86
Technologiia 3D, z p-value równe 0.79
Pobór mocy, z p-value równe 0.72
Technologiia HDR, z p-value równe 0.51
Marka, z p-value równe 0.20
Klasa efektywności, z p-value równe 0.16
Wartości p-value to p-value F statystki wszystkich dummy zmiennych z ka-

tegorii lub wartość t-testu dla zmiennej.

Otrzymany w ten sposób model wygląda tak:

Dep. Variable:	Cena	**R-squared:**	0.878
Model:	OLS	**Adj. R-squared:**	0.859
Method:	Least Squares	**F-statistic:**	47.02
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	1.21e-23
Time:	18:13:49	**Log-Likelihood:**	-507.77
No. Observations:	69	**AIC:**	1036.
Df Residuals:	59	**BIC:**	1058.
Df Model:	9		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-1150.0660	262.728	-4.377	0.000	-1675.784	-624.348
Przekątna ekranu	56.5314	4.794	11.793	0.000	46.939	66.123
Stan_Po zwrocie	-265.3156	306.228	-0.866	0.390	-878.077	347.446
Stan_Powystawowy	480.0764	311.772	1.540	0.129	-143.778	1103.931
Stan_Używany	-305.4451	116.852	-2.614	0.011	-539.266	-71.624
Typ telewizora_LED	21.4830	209.644	0.102	0.919	-398.014	440.980
Typ telewizora_OLED	-105.5515	554.635	-0.190	0.850	-1215.374	1004.271
Typ telewizora_QLED	1058.4140	282.623	3.745	0.000	492.887	1623.941
Typ telewizora_inny	461.8203	452.754	1.020	0.312	-444.138	1367.779
Typ telewizora_plazmowy	-215.3403	289.991	-0.743	0.461	-795.611	364.931

Omnibus:	5.021	**Durbin-Watson:**	2.038
Prob(Omnibus):	0.081	**Jarque-Bera (JB):**	6.855
Skew:	-0.016	**Prob(JB):**	0.0325
Kurtosis:	4.544	**Cond. No.**	591.

Model ten spełnia podstawowe założenia. W modelu nie ma katalizatorów.
Model nie spełnia testu serii z wynikiem p-value = 0.82, co z dużą pewnością odrzuca H_0 o liniowości modelu.

Oszacujmy nowy model dla logarytmu ceny

Dep. Variable:	Cena	**R-squared:**	0.825
Model:	OLS	**Adj. R-squared:**	0.798
Method:	Least Squares	**F-statistic:**	30.81
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	4.06e-19
Time:	18:13:49	**Log-Likelihood:**	-22.273
No. Observations:	69	**AIC:**	64.55
Df Residuals:	59	**BIC:**	86.89
Df Model:	9		

	coef	std err	t	P> |t|	[0.025	0.975]
const	4.5693	0.231	19.775	0.000	4.107	5.032
Przekątna ekranu	0.0427	0.004	10.132	0.000	0.034	0.051
Stan_Po zwrocie	0.0405	0.269	0.150	0.881	-0.498	0.579
Stan_Powystawowy	0.0863	0.274	0.315	0.754	-0.462	0.635
Stan_Używany	-0.1141	0.103	-1.110	0.272	-0.320	0.092
Typ telewizora_LED	0.4988	0.184	2.705	0.009	0.130	0.868
Typ telewizora_OLED	0.5401	0.488	1.107	0.273	-0.436	1.516
Typ telewizora_QLED	0.9832	0.249	3.956	0.000	0.486	1.481
Typ telewizora_inny	0.1265	0.398	0.318	0.752	-0.670	0.923
Typ telewizora_plazmowy	0.2606	0.255	1.022	0.311	-0.250	0.771

Omnibus:	31.663	**Durbin-Watson:**	1.856
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	86.469
Skew:	-1.394	**Prob(JB):**	1.67e-19
Kurtosis:	7.723	**Cond. No.**	591.

Model ten spełnia poprzednie założenia.

Otrzymujemy wysokie p-value dla testu RESET, p-value = 0.67, prowadzi to do odrzucenia H_0 o poprawności obecnego modelu.

Test ten przyjmuje tylko zmienne ciągłe a jedyną zmienną ciągłą w modelu, jest przekątna ekranu. Więc dokonamy transformacji tej zmiennej, podnosząc ją do kwadratu.

Oszacujmy nowy model z kwadratem zmiennej przekątna ekranu

Dep. Variable:	Cena	**R-squared:**	0.805
Model:	OLS	**Adj. R-squared:**	0.775
Method:	Least Squares	**F-statistic:**	27.08
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	8.33e-18
Time:	18:13:49	**Log-Likelihood:**	-25.902
No. Observations:	69	**AIC:**	71.80
Df Residuals:	59	**BIC:**	94.14
Df Model:	9		

	coef	std err	t	P> |t|	[0.025	0.975]
const	5.3786	0.209	25.754	0.000	4.961	5.797
Przek¹tna ekranu	0.0004	4.7e-05	9.302	0.000	0.000	0.001
Stan_Po zwrocie	0.1391	0.282	0.492	0.624	-0.426	0.704
Stan_Powystawowy	0.0632	0.290	0.218	0.828	-0.517	0.644
Stan_U¿ywany	-0.0465	0.108	-0.430	0.669	-0.263	0.170
Typ telewizora_LED	0.6247	0.192	3.259	0.002	0.241	1.008
Typ telewizora_OLED	0.6817	0.512	1.331	0.188	-0.343	1.707
Typ telewizora_QLED	1.1406	0.256	4.449	0.000	0.628	1.654
Typ telewizora_inny	-0.0224	0.418	-0.054	0.957	-0.860	0.815
Typ telewizora_plazmowy	0.3887	0.266	1.464	0.149	-0.143	0.920

Omnibus:	28.917	**Durbin-Watson:**	1.829
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	70.967
Skew:	-1.308	**Prob(JB):**	3.89e-16
Kurtosis:	7.224	**Cond. No.**	3.17e+04

Model ten spełnia poprzednie założenia, w tym nie ma podstaw do odrzucenia H_0 w teście RESET, co oznacza poprawną postać modelu.

Zarówno test Breusch–Pagana jak i test Whita daje podstawy do odrzucenia H_0 o homdeskatyczności modelu.

Wykonujemy więc standardową korektę na heteroskedastyczność i otrzymujemy model w postaci:

Dep. Variable:	y	**R-squared:**	0.809
Model:	OLS	**Adj. R-squared:**	0.780
Method:	Least Squares	**F-statistic:**	27.76
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	4.69e-18
Time:	18:13:49	**Log-Likelihood:**	-17.733
No. Observations:	69	**AIC:**	55.47
Df Residuals:	59	**BIC:**	77.81
Df Model:	9		

	coef	std err	t	P> |t|	[0.025	0.975]
const	5.6993	0.181	31.483	0.000	5.337	6.061
Przek¹tna ekranu	0.0002	2.36e-05	7.116	0.000	0.000	0.000
Stan_Po zwrocie	0.3620	0.249	1.452	0.152	-0.137	0.861
Stan_Powystawowy	0.4769	0.252	1.891	0.064	-0.028	0.982
Stan_U¿ywany	-0.0655	0.096	-0.681	0.498	-0.258	0.127
Typ telewizora_LED	0.8279	0.168	4.921	0.000	0.491	1.165
Typ telewizora_OLED	1.0879	0.452	2.406	0.019	0.183	1.992
Typ telewizora_QLED	1.6965	0.216	7.853	0.000	1.264	2.129
Typ telewizora_inny	-0.1934	0.371	-0.521	0.604	-0.936	0.549
Typ telewizora_plazmowy	0.5140	0.240	2.144	0.036	0.034	0.994

Omnibus:	6.372	**Durbin-Watson:**	1.969
Prob(Omnibus):	0.041	**Jarque-Bera (JB):**	10.243
Skew:	-0.117	**Prob(JB):**	0.00597
Kurtosis:	4.873	**Cond. No.**	3.86e+04

Model ten spełnia poprzednie założenia i jest homdeskatyczny.

Przeprowadzając test Chowa, dzieląc losowa dane na połowę, zauważamy, że model, jest wrażliwy na dobór danych.

Numer próby	P-value dla testu Chowa
0	$2.045\,944 \times 10^{-1}$
1	$6.442\,818 \times 10^{-4}$
2	$4.177\,916 \times 10^{-2}$
3	$2.064\,195 \times 10^{-3}$
4	$7.591\,984 \times 10^{-3}$
5	$3.833\,329 \times 10^{-1}$
6	$3.330\,146 \times 10^{-8}$
7	$1.805\,170 \times 10^{-1}$
8	$5.618\,481 \times 10^{-3}$
9	$9.585\,642 \times 10^{-1}$
10	$4.304\,746 \times 10^{-1}$
11	$6.677\,985 \times 10^{-3}$
12	$1.630\,223 \times 10^{-1}$
13	$5.255\,308 \times 10^{-2}$
14	$7.709\,834 \times 10^{-1}$
15	$2.007\,603 \times 10^{-1}$
16	$1.769\,388 \times 10^{-1}$
17	$4.549\,207 \times 10^{-1}$
18	$2.890\,782 \times 10^{-1}$
19	$2.451\,896 \times 10^{-2}$

Widać tu że mamy wyniki mówiące zarówno o stabilności modelu, jak i jego braku. Wynika to z dużej ilości obserwacji rzadkich , które mają wysoki wpłych na wynik tego testu. Usuwanie danych, żeby doprowadzić do stabilnośći modelu, jest nieuzasadnione i zmniejszy tylko możliwości generalizacji modelu.

3.3 Wybór postaci modelu

Ostateczna postać modelu to

$$log(\text{Cena}) \sim \alpha_0 + \alpha_1 * \text{Przekątna ekranu}^2 + \alpha_2 * \text{Stan}_{dummy} + \alpha_2 \text{Typ telewizora}_{dummy}$$

4 Testowanie własności modelu

4.1 Testy Modelu

Ostateczna wersja modelu to:

Dep. Variable:	y	**R-squared:**	0.809
Model:	OLS	**Adj. R-squared:**	0.780
Method:	Least Squares	**F-statistic:**	27.76
Date:	Sat, 13 Jun 2020	**Prob (F-statistic):**	4.69e-18
Time:	18:13:50	**Log-Likelihood:**	-17.733
No. Observations:	69	**AIC:**	55.47
Df Residuals:	59	**BIC:**	77.81
Df Model:	9		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	5.6993	0.181	31.483	0.000	5.337	6.061
Przekątna ekranu	0.0002	2.36e-05	7.116	0.000	0.000	0.000
Stan_Po zwrocie	0.3620	0.249	1.452	0.152	-0.137	0.861
Stan_Powystawowy	0.4769	0.252	1.891	0.064	-0.028	0.982
Stan_Używany	-0.0655	0.096	-0.681	0.498	-0.258	0.127
Typ telewizora_LED	0.8279	0.168	4.921	0.000	0.491	1.165
Typ telewizora_OLED	1.0879	0.452	2.406	0.019	0.183	1.992
Typ telewizora_QLED	1.6965	0.216	7.853	0.000	1.264	2.129
Typ telewizora_inny	-0.1934	0.371	-0.521	0.604	-0.936	0.549
Typ telewizora_plazmowy	0.5140	0.240	2.144	0.036	0.034	0.994

Omnibus:	6.372	**Durbin-Watson:**	1.969
Prob(Omnibus):	0.041	**Jarque-Bera (JB):**	10.243
Skew:	-0.117	**Prob(JB):**	0.00597
Kurtosis:	4.873	**Cond. No.**	3.86e+04

Statystyka/Test	P-value	Komentarz	Decyzja
Jarque-Bera	0.005	Rozkład reszt jes normalny	Brak podstaw do odrzucenia H_0
F-statistic	4.69×10^{-18}	R^2 jest statystycznie znaczący	Brak podstaw do odrzucenia H_0
Durbin-Watson	1.969	Nie ma autokorelacji	Brak podstaw do odrzucenia H_0
Istotnośći zmiennych	——	——	——
F Stan	0.04	Jest znacząca	Brak podstaw do odrzucenia H_0
F Typ telewizora	$5.886\,497\,88 \times 10^{-8}$	Jest znacząca	Brak podstaw do odrzucenia H_0
t-testPrzekątna ekranu	$1.723\,455\,164\,413\,816 \times 10^{-9}$	Jest znacząca	Brak podstaw do odrzucenia H_0
t-test const	$1.379\,804\,996\,591\,566\,9 \times 10^{-38}$	Jest znacząca	Brak podstaw do odrzucenia H_0
Kataliza	——	Brak par katalizatorów	——
Test serii	0.00	Model jest liniowy	Brak podstaw do odrzucenia H_0
RESET	0.0046	Postać modelu jest poprawna	Brak podstaw do odrzucenia H_0
Breusch-Pagan	$3.140\,388\,461\,885\,039 \times 10^{-7}$	Model jest homodeksatyczny	Brak podstaw do odrzucenia H_0
White	$2.294\,558\,828\,256\,874\,8 \times 10^{-5}$	Model jest homodeksatyczny	Brak podstaw do odrzucenia H_0
Chow	——	Model nie jest stabliny	——-
Współność	——	VIF i test współiniowsci wskazuje na jej brak	——-

4.2 Interpretacja parametrów modelu.

Interpretacja dla Stanu i Typu jest taka sama, z uwagą żeby otrzymać o ile zmienni to wartość w cenie należy użyć ich jako argument w funkcji e^x.

W Stanie porównuje się do sytuacji, kiedy Stan to nowy, a w typie kiedy typ to LCD.

Interpretacja Przekątnej wraz ze wzrostem przekątnej następuje wzrost ceny. Dokładna wartość nie możliwa do podania ponieważ zależność jest nie liniowa, wzrost ceny o jeden cal z 20 do 21 nie jest równe zmiany o jeden cal z 70 do 71.

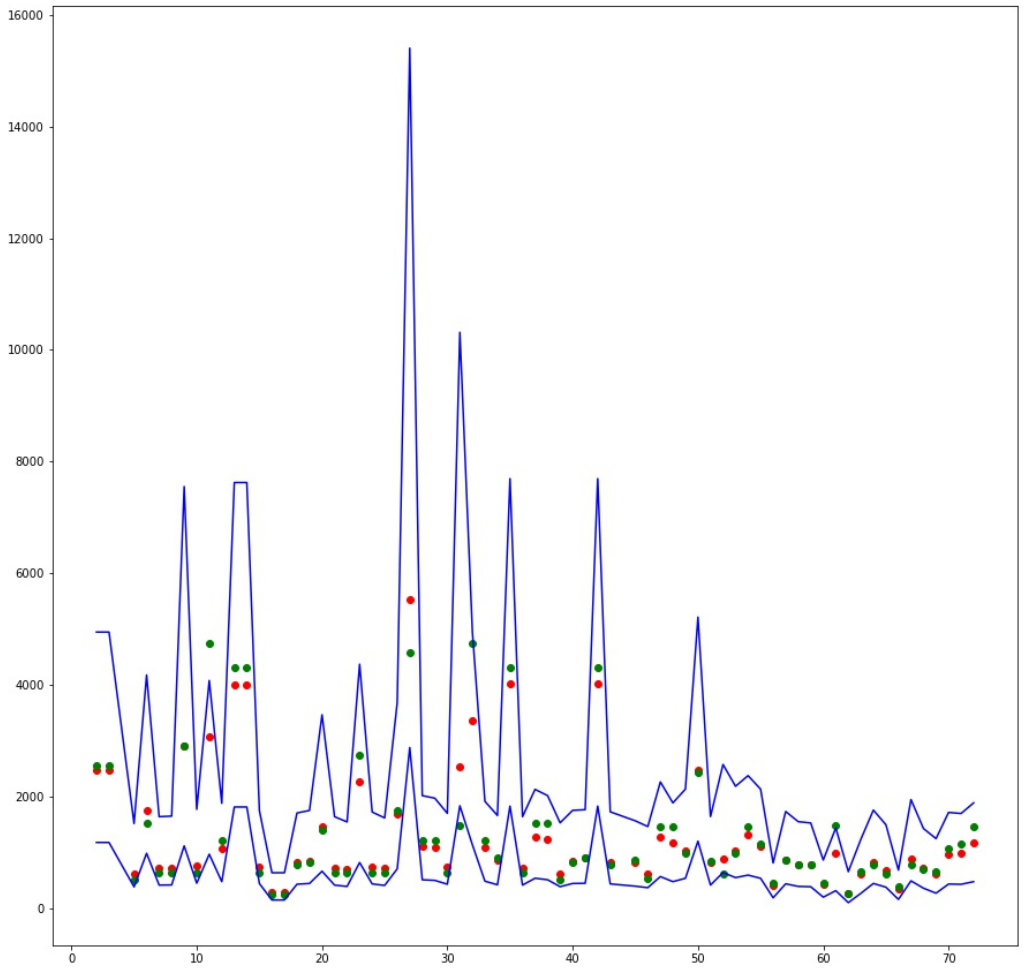
4.3 predykcja wraz z 95% przedziałem ufności.

	średnia wartosc	Dolna	Górna	Prawdziwa
0	2477.346083	1174.160943	4944.613420	3299.0
1	2477.346083	1174.160943	4944.613420	3299.0
2	615.199526	378.295259	1513.229797	399.0
3	1755.097975	980.177779	4175.716966	2178.0
4	720.531276	410.408866	1636.248821	599.0
5	722.793677	413.031360	1646.376215	579.0
6	2899.000000	1112.795004	7552.335309	2899.0
7	749.178958	444.101065	1767.320037	419.0
8	3066.151654	963.761964	4076.936438	4199.0
9	1066.199624	471.656930	1876.099749	1585.0
10	3999.332996	1809.964927	7623.517395	3699.0
11	3999.332996	1809.964927	7623.517395	3699.0
12	744.763607	438.840188	1746.715312	439.0
13	272.200996	143.085099	631.713298	229.0
14	272.200996	143.085099	631.713298	229.0
15	814.179933	426.512161	1701.673946	730.0
16	845.200028	439.281724	1748.442639	879.0
17	1448.601606	659.347273	3462.556155	1349.0
18	720.531276	410.408866	1636.248821	599.0
19	699.036214	385.826670	1541.917682	899.0
20	2273.167078	815.195432	4368.230180	3599.0
21	738.887435	431.876923	1719.522486	469.0
22	715.510799	404.613073	1613.911042	649.0
23	1675.306026	700.884229	3680.687097	1799.0
24	5525.555125	2875.423108	15407.973975	3490.0
25	1104.362366	505.899087	2013.309502	1299.0
26	1091.946134	494.657771	1968.013483	1379.0
27	733.756165	425.832241	1695.989835	499.0
28	2531.585577	1831.725541	10316.271365	399.0
29	3350.748625	1132.729444	4947.324527	3599.0
30	1076.015115	480.376057	1910.821768	1499.0
31	868.620061	415.809568	1657.118554	1449.0
32	4017.392554	1825.944816	7694.222180	3659.0
33	720.531276	410.408866	1636.248821	599.0
34	1272.514533	533.530526	2125.705859	1738.0
35	1238.973706	506.064055	2013.976073	1976.0
36	617.933146	381.748610	1526.371734	367.0

	średnia wartosc	Dolna	Górna	Prawdziwa
37	845.439714	439.533038	1749.425963	877.0
38	895.927552	442.672925	1761.721464	1099.0
39	4017.392554	1825.944816	7694.222180	3659.0
40	819.356651	431.982467	1723.264747	699.0
41	823.364341	390.792879	1562.139778	1349.0
42	615.191045	363.972238	1458.972751	454.0
43	1279.535999	563.134829	2259.325292	1299.0
44	1170.000553	472.174385	1883.748679	1899.0
45	1018.861246	532.145719	2129.410888	680.0
46	2464.866227	1197.858648	5212.556440	2499.0
47	827.861431	410.303421	1638.047400	999.0
48	879.389087	635.765770	2571.885062	150.0
49	1030.889548	544.520344	2181.043383	650.0
50	1310.123036	589.584036	2371.830952	1199.0
51	1108.262847	532.253140	2129.857701	950.0
52	414.534153	182.718383	806.587467	418.0
53	866.789114	433.550867	1729.466087	890.0
54	775.967442	386.985278	1547.411235	1100.0
55	769.914551	380.864490	1523.792456	1199.0
56	427.712536	194.324963	859.551237	349.0
57	992.075369	310.675252	1434.447694	2500.0
58	250.000000	95.963695	651.287971	250.0
59	612.680606	262.527908	1231.152460	1000.0
60	826.813141	439.909407	1754.657774	659.0
61	671.617433	372.729999	1492.512250	679.0
62	348.825142	153.886683	678.476830	599.0
63	869.866399	486.755660	1942.787954	499.0
64	706.851938	354.604947	1423.254348	1549.0
65	615.836034	265.391430	1243.152119	950.0
66	952.694245	429.182524	1712.206178	1400.0
67	988.969597	424.452393	1693.559656	1720.0
68	1170.000553	472.174385	1883.748679	1899.0

5 Podsumowanie

Mimo że model spełnia wszystkie testy i potwierdził początkową hipotezę o zależności ceny od przekątnej, to pokazał on również pewne nieliniowe zachowania wpływające na cenę telewizorów. Zaskakujące okazało się, że jakość nie ma statystyczne znaczącego wpływy na cenę i 80% można wyjaśnić z wiedzą o stanie i typie telewizora. Same też predykcje z dużym absolutnym odchyłom wskazują na niską użyteczność tego w realnym świecie. Błędy te wynikają po części z małej ilości danych.



Rysunek 14: Predykcje z przedzialem ufnosci

6 Lireratura