

# 1 Отчет анализа стоимости ювелирных изделий у Эгэлгэ, дата парсинга - 20 апреля 2022

## 1.1 Парсинг

С сайта <https://egelge.com/> (<https://egelge.com/>) были спарсены данные ювелирных изделий предлагаемых на интернет магазине

Процесс парсинга, реализован в файле ParsingEgelge.ipynb, процесс очистки и расчет преysкуратной стоимости бриллиантов в файле Egelge\_1after\_parsing.ipynb

Всего колец с фильтром бриллиант на сайте более 145 изделий,

## 1.2 Информация о спарсенных данных

Ввод [61]:

1 svod\_0

Out[61]:

	изделий	price_old	price	weight	ШТУК бриллиантов	price_old_сред_за_грамм	price__сред_за_грамм
<b>nomenclatura</b>							
без вставок	23	792305	554611	83.55	0.00	9483.00	6638.07
бриллианты и ПДК	6	368250	147301	13.70	0.00	26879.56	10751.90
только ПДК	11	845063	591543	75.50	0.00	11192.89	7835.01
только с бриллиантами	105	7704990	3081987	226.80	588.00	33972.62	13589.01

_брилл_суммированы_эглгеЭг_14520-04-2022 - Microsoft Excel																			
расчет_склад без ДК_24-04-2022 *																			
РОЗНИЦА_СкладОстаткиТМЦсострочнойрасшифровкойБрп25-04-22 *																			
_брилл_суммированы_эглгеЭг_14520-04-2022 *																			
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
index_row	Unnamed: 1	h1	art	price	disc_price	action	describe	Бренд	а произво	Материал	gems	редний ве	url	мм	рублеамм	рубл	checklog	filename	gemsp
0	1 с бриллиантом	Золотое н Артикул:	21486	8594	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.1	<a href="https://eg">https://eg</a>	7812.727	6510.606	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	3	0.027	10.27			
1	2 с бриллиантом	Золотое н Артикул:	21486	8594	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.1	<a href="https://eg">https://eg</a>	7812.727	6510.606	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	3	0.027	10.27			
2	3 с бриллиантом	Золотое н Артикул:	24572	9829	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.2	<a href="https://eg">https://eg</a>	8190.833	6825.694	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.02	7			
3	4 с бриллиантом	Золотое н Артикул:	25645	10258	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1.2	<a href="https://eg">https://eg</a>	8548.333	7123.611	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.018	6.57			
4	5 с бриллиантом	Золотое н Артикул:	28128	11251	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.4	<a href="https://eg">https://eg</a>	8036.429	6697.024	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.042	14.07			
5	6 с бриллиантом	Золотое н Артикул:	28391	11356	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1	<a href="https://eg">https://eg</a>	11356	9463.333	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.061	20.74			
6	7 с бриллиантом	Золотое н Артикул:	29871	11948	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1.2	<a href="https://eg">https://eg</a>	9956.667	8297.222	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.045	15.52			
7	8 с бриллиантом	Золотое н Артикул:	30590	12236	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.7	<a href="https://eg">https://eg</a>	7197.647	5998.039	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.012	4.62			
8	9 с бриллиантом	Золотое н Артикул:	30660	12264	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1	<a href="https://eg">https://eg</a>	12264	10220	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.062	22.32			
9	10 с бриллиантом	Золотое н Артикул:	32959	13184	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1.2	<a href="https://eg">https://eg</a>	10986.67	9155.556	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.062	22.32			
10	11 с ПДК	Золотое н Артикул:	19326	13528	Акция: SP Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Хромди	1.6	<a href="https://eg">https://eg</a>	8455	7045.833	эглгеЭг_1 Хромди	1 Хромдиопс только ПДК						
11	12 с бриллиантом	Золотое н Артикул:	35043	14017	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.7	<a href="https://eg">https://eg</a>	8245.294	6871.078	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.046	15.41			
12	13 с ПДК	Золотое н Артикул:	21279	14895	Акция: SP Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Хромди	1.8	<a href="https://eg">https://eg</a>	8275	6895.833	эглгеЭг_1 Хромди	1 Хромдиопс только ПДК						
13	14 с бриллиантом	Золотое н Артикул:	37440	14976	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1.2	<a href="https://eg">https://eg</a>	12480	10400	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.08	35.2			
14	15 с бриллиантом	Золотое н Артикул:	39413	15765	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 8 Бр Кр 57	2	<a href="https://eg">https://eg</a>	7882.5	6568.75	эглгеЭг_8 Бр Кр 57	8 Бр Кр 57 0.0 только с бриллиа	8	0.05	1			
15	16 с бриллиантом	Золотое н Артикул:	39694	15878	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1	<a href="https://eg">https://eg</a>	15878	13231.67	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.09	39.6			
16	17 с бриллиантом	Золотое н Артикул:	39886	15954	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 17 Бр Кр 1	2.1	<a href="https://eg">https://eg</a>	7597.143	6330.952	эглгеЭг_17 Бр Кр 1	17 Бр Кр 17 0. только с бриллиа	17	0.059	21.83			
17	18 без вставок	Золотое н Артикул:	25352	17746	Акция: SP Кольцо из: ЭГЭЛГЭ	Россия	красное золото 585	2.4	<a href="https://eg">https://eg</a>	7394.167	6161.806	эглгеЭг_nan	['nan'] без вставок						
18	19 с бриллиантом	Золотое н Артикул:	44741	17896	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.4	<a href="https://eg">https://eg</a>	12782.86	10652.38	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.077	32.34			
19	20 с бриллиантом	Золотое н Артикул:	45031	18012	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	белое з 1 Бр Кр 57	1.4	<a href="https://eg">https://eg</a>	12865.71	10721.43	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.08	35.2			
20	21 с бриллиантом	Золотое н Артикул:	46184	18474	Акция: Ди: Кольцо из: ЭГЭЛГЭ	Россия	красное з 1 Бр Кр 57	1.5	<a href="https://eg">https://eg</a>	12316	10263.33	эглгеЭг_1 Бр Кр 57	1 Бр Кр 57 0.0 только с бриллиа	1	0.074	31.08			

## 1.3 Анализ датасета визуализация, определение коэффициентов

### 1.3.1 Считывание данных с файла

Для этого устанавливаем необходимые модули.

- загружаем данные с файла
- оставлены поля price : цена изделия, weight: масса изделия в граммах, prcost: прейскурант в долларах

```
Ввод [2]: 1 import numpy as np
          2 import pandas as pd
          3 import re
          4 import os
          5 import matplotlib.pyplot as plt
          6 %matplotlib inline
          7 %matplotlib notebook
          8 filename = 'jewels/egelge/_брилл_суммированы_эгэлгег_14520-04-2022.xlsx'
          9 #'jewels/miuz04-2022/_брилл_суммированы_mius140422_0-4431.xlsx' #'jewels/miuz/_брилл_суммированы_mius161121_0-4431.xlsx'
```

```
Ввод [46]: 1 pd.options.display.float_format = '{:.2f}'.format
```

```
Ввод [3]: 1 filename
```

```
Out[3]: 'jewels/egelge/_брилл_суммированы_эгэлгег_14520-04-2022.xlsx'
```

```
Ввод [4]: 1 raw = pd.read_excel('jewels/egelge/_брилл_суммированы_эгэлгег_14520-04-2022.xlsx') #filename, miuz, sheet_name
```

```
Ввод [5]: 1 raw.rename(columns={'price':'price_old', 'Средний вес':'weight'}, inplace=True)
```

```
Ввод [17]: 1 raw.rename(columns={'disc_price':'price', 'Unnamed: 1':'Номенклатура'}, inplace=True)
```

```
Ввод [26]: 1 raw['изделий'] = 1
```

```
Ввод [57]: 1 svod_0 = raw[['nomenklatura', 'изделий', 'price_old', 'price', 'weight', 'ШТУК']].groupby(
          2      ['nomenklatura']).sum().rename(columns = {'ШТУК':'ШТУК бриллиантов'})
```

```
Ввод [58]: 1 svod_0['price_old_сред_за_грамм'] = svod_0['price_old']/svod_0['weight']
```

```
Ввод [59]: 1 svod_0['price__сред_за_грамм'] = svod_0['price']/svod_0['weight']
```

Ввод [60]:

```
1 svod_0
```

Out[60]:

	изделий	price_old	price	weight	ШТУК бриллиантов	price_old_сред_за_грамм	price__сред_за_грамм
nomenklatura							
без вставок	23	792305	554611	83.55	0.00	9483.00	6638.07
бриллианты и ПДК	6	368250	147301	13.70	0.00	26879.56	10751.90
только ПДК	11	845063	591543	75.50	0.00	11192.89	7835.01
только с бриллиантами	105	7704990	3081987	226.80	588.00	33972.62	13589.01

Ввод [62]:

```
1 data = raw.loc[~raw.prcost.isna(),['price','weight','prcost']]
2 #data = raw.loc[(raw.gold.str.contains('585') ),['price','weight','prcost']].loc[(raw.price<500000),:].loc[
3 data.head()
```

Out[62]:

	price	weight	prcost
0	8594	1.10	10.27
1	8594	1.10	10.27
2	9829	1.20	7.00
3	10258	1.20	6.57
4	11251	1.40	14.07

Ввод [63]:

```
1 #raw.loc[~raw.prcost.isna(),['price','weight','prcost']]
```

Ввод [64]:

```
1 print(f'''После урезания получен датасет с {len(data)} данными
2 общая масса изделий {round(data.sum()['weight']/1000,2)} кг на сумму {round(data.sum()['price']/1000000)} м.
3 с преysкурaнтной стоимостью {round(data.sum()['prcost']/1000)} тыс долл''')
```

После урезания получен датасет с 105 данными  
общая масса изделий 0.23 кг на сумму 3 млн рублей  
с преysкурaнтной стоимостью 7 тыс долл

Ввод [65]:

```
1 #сумма датасета
2 data.describe()
```

Out[65]:

	price	weight	prcost
count	105.00	105.00	105.00
mean	29352.26	2.16	64.91
std	18704.53	1.08	70.61
min	3766.00	0.30	1.00
25%	15878.00	1.40	21.76
50%	23893.00	2.10	39.60
75%	38491.00	2.70	82.40
max	101385.00	5.60	449.86

Ввод [66]:

```
1 data.describe().loc['min','price']
```

Out[66]: 3766.0

### 1.3.2 Визуализация данных

Построим облако точек в трехмерной плоскости

- x вес золота в граммах

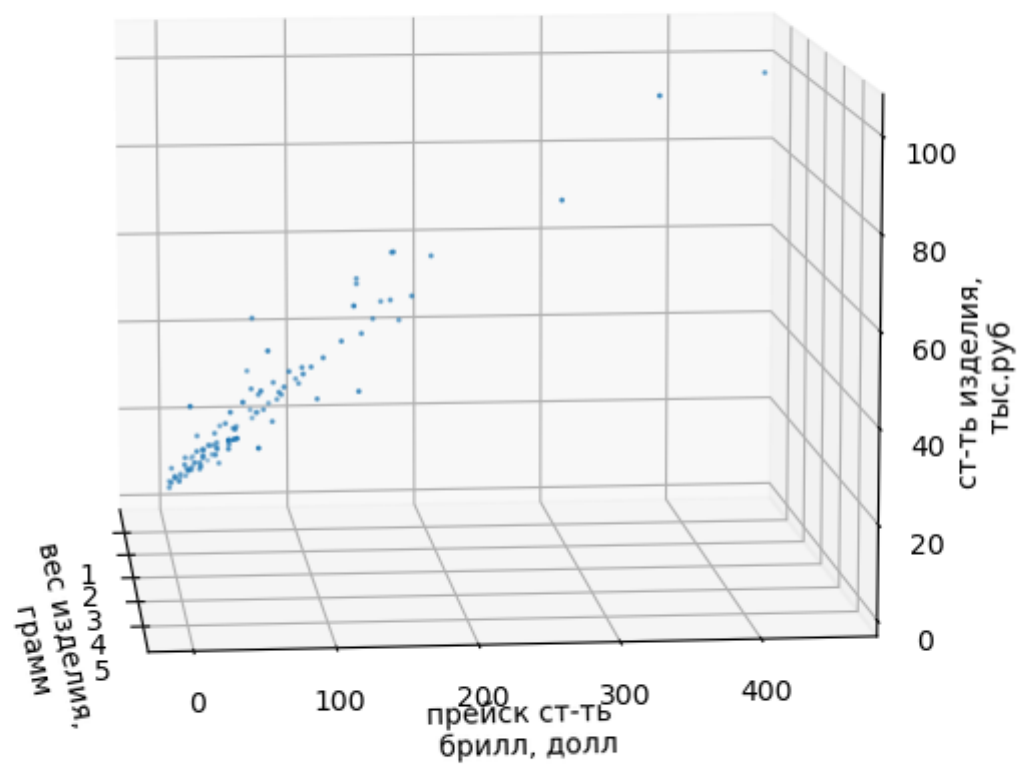
- y стоимость преискурантная бриллиантов
- z стоимость изделия в тысячах рублей

Если покрутить график, то видно что точки стремятся к плоскости

Ввод [67]: 1 `%matplotlib notebook`

Ввод [68]:

```
1 #
2 fig = plt.figure(figsize=(7,7))
3 ax = fig.gca(projection = "3d")
4 ax.scatter3D(data.weight,data.prcost,data.price/1000,s=1,c='#1f77b4')
5 ax.set_xlabel('вес изделия, \n грамм')
6 ax.set_ylabel('прейск ст-ть \n брилл, долл')
7 ax.set_zlabel('ст-ть изделия,\n тыс.руб')
8 ax.view_init(45,0)
9 #todo 3 projection need
```





### **1.3.3 Подготовка данных для определения коэффициентов**

Для того чтобы полученные коэффициенты имели интерпретируемый смысл, предлагается преysкурantную стоимость бриллиантов перевести в рубли, граммы золота перевести в биржевую стоимость золота, данного веса изделия в рублях.

Ввод [69]:

```

1  #Курс доллара и биржевая цена золота на 03-12-2021
2  #dollar_rate = 73.66
3  #gold_rate = 4191.05
4  #Курс доллара и биржевая цена золота на 19-04-2022
5  dollar_rate = 80.36
6  gold_rate = 5015.06
7  data['prcost_rub'] = data.prcost*dollar_rate
8  data['gold_rub'] = data.weight*0.585*gold_rate
9  pd.options.display.float_format = "{:.2f}".format
10 data

```

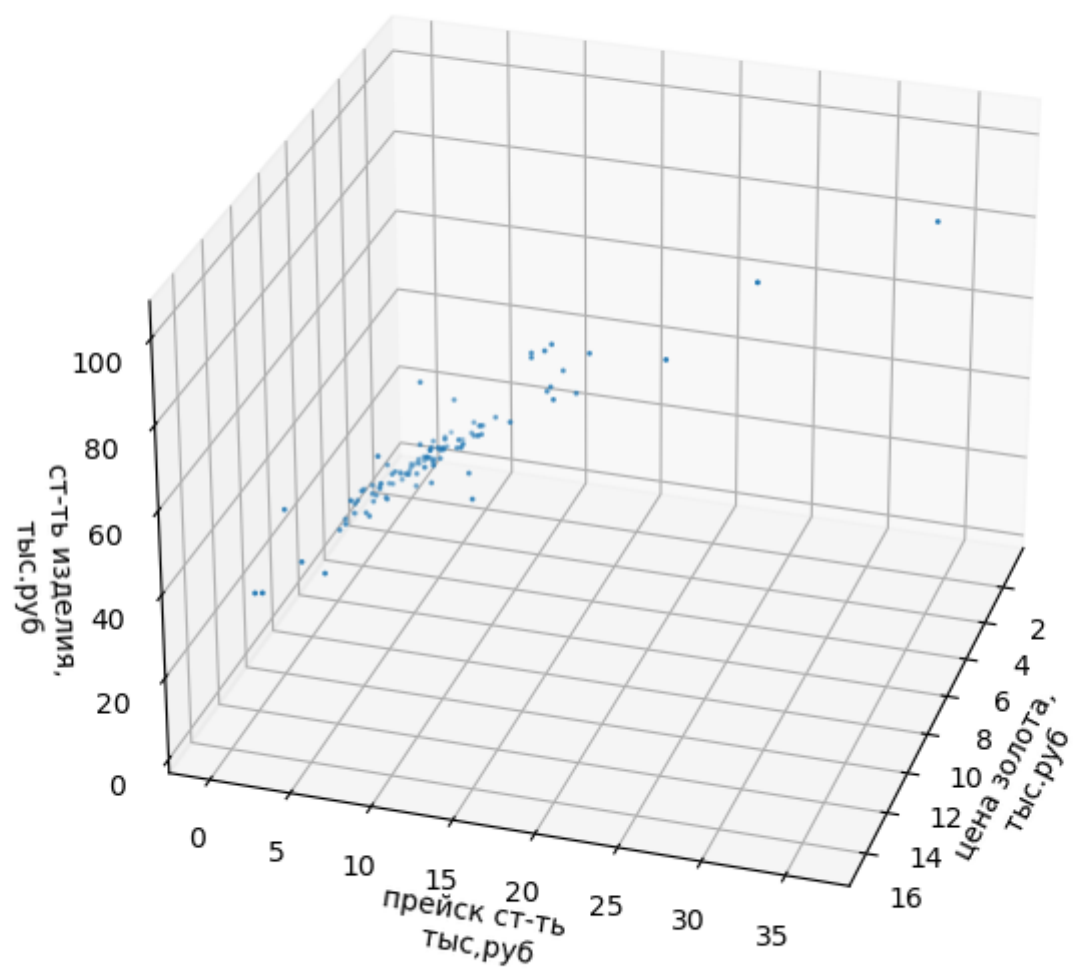
Out[69]:

	price	weight	prcost	prcost_rub	gold_rub
<b>0</b>	8594	1.10	10.27	825.30	3227.19
<b>1</b>	8594	1.10	10.27	825.30	3227.19
<b>2</b>	9829	1.20	7.00	562.52	3520.57
<b>3</b>	10258	1.20	6.57	527.97	3520.57
<b>4</b>	11251	1.40	14.07	1130.67	4107.33
...	...	...	...	...	...
<b>139</b>	11393	0.50	23.46	1885.25	1466.91
<b>140</b>	12577	1.60	9.36	752.17	4694.10
<b>141</b>	15795	1.80	37.64	3024.75	5280.86
<b>143</b>	21744	2.00	44.26	3556.73	5867.62
<b>144</b>	22734	2.90	30.10	2418.84	8508.05

105 rows × 5 columns

Ввод [70]:

```
1 #новое графическое представление будет иметь вид
2 fig2 = plt.figure(figsize=(7,7))
3 ax = plt.gca(projection = "3d")
4 ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.price/1000,s=1,c='#1f77b4')
5 ax.set_xlabel('цена золота, \n тыс.руб')
6 ax.set_ylabel('прейск ст-ть \n тыс,руб')
7 ax.set_zlabel('ст-ть изделия,\n тыс.руб')
8 ax.view_init(45,0)
```



Цена ювелирного изделия в первом приближении формируется линейной зависимостью, если обозначить за  $Z$  - цену изделия,  $X$  - масса золота,  $Y$  - преискуртная стоимость бриллианов то можно записать закономерностью вида:

$$Z = f(X, Y) = \text{НДС} * \text{СКИДКА} * \text{НАЦЕНКА} * \\ (\text{ПРОБА} * \text{БИРЖЕВАЯСТОИМОСТЬ} * \text{НАЦЕНКАЗОЛОТА} * \text{СЛОЖНОСТЬ} * X \\ + \text{НАЦЕНКАПРЕЙСКУРАНТ} * \text{СЛОЖНОСТЬ} * Y + \text{НАЦЕНКАБЕЗДКДМ})$$

можно раскрыть скобки, перемножить множители и мы получим уравнение вида:

$$Z = A * X + B * Y + C$$

линейное уравнение такого вида в трехмерной плоскости является плоскостью, поэтому необходимо усредненную плоскость и найти эти три коэффициента.

### 1.3.4 Вычисление коэффициентов усредненной по облаку данных плоскости

Для этого мы устанавливаем модуль отвечающий за расчеты линейной алгебры, и загружаем наши данные для расчета коэффициентов, и получаем коэффициенты

```
Ввод [71]: 1 from sklearn import linear_model
           2 #подготовка данных
           3 miuz = linear_model.LinearRegression()
           4 X_train = data.iloc[:, :].loc[:, ['gold_rub', 'prcost_rub']].to_numpy()
           5 Y_train = data.iloc[:, :].loc[:, 'price'].to_numpy()
           6 miuz.fit(X_train, Y_train)
           7 A, B = miuz.coef_
           8 C = miuz.intercept_
           9 A, B, C
```

```
Out[71]: (2.0239842264446803, 2.720397086093228, 2336.7445761091694)
```

### 1.3.5 Визуальная проверка полученных коэффициентов

```
Ввод [72]: 1 #проверка
2 data['predict'] = miuz.predict(X_train)
3 data['bycoef'] = data.gold_rub*miuz.coef_[0] + data.prcost_rub*miuz.coef_[1]+miuz.intercept_
4 data['diff'] = data.predict-data.bycoef
5 data.head()
```

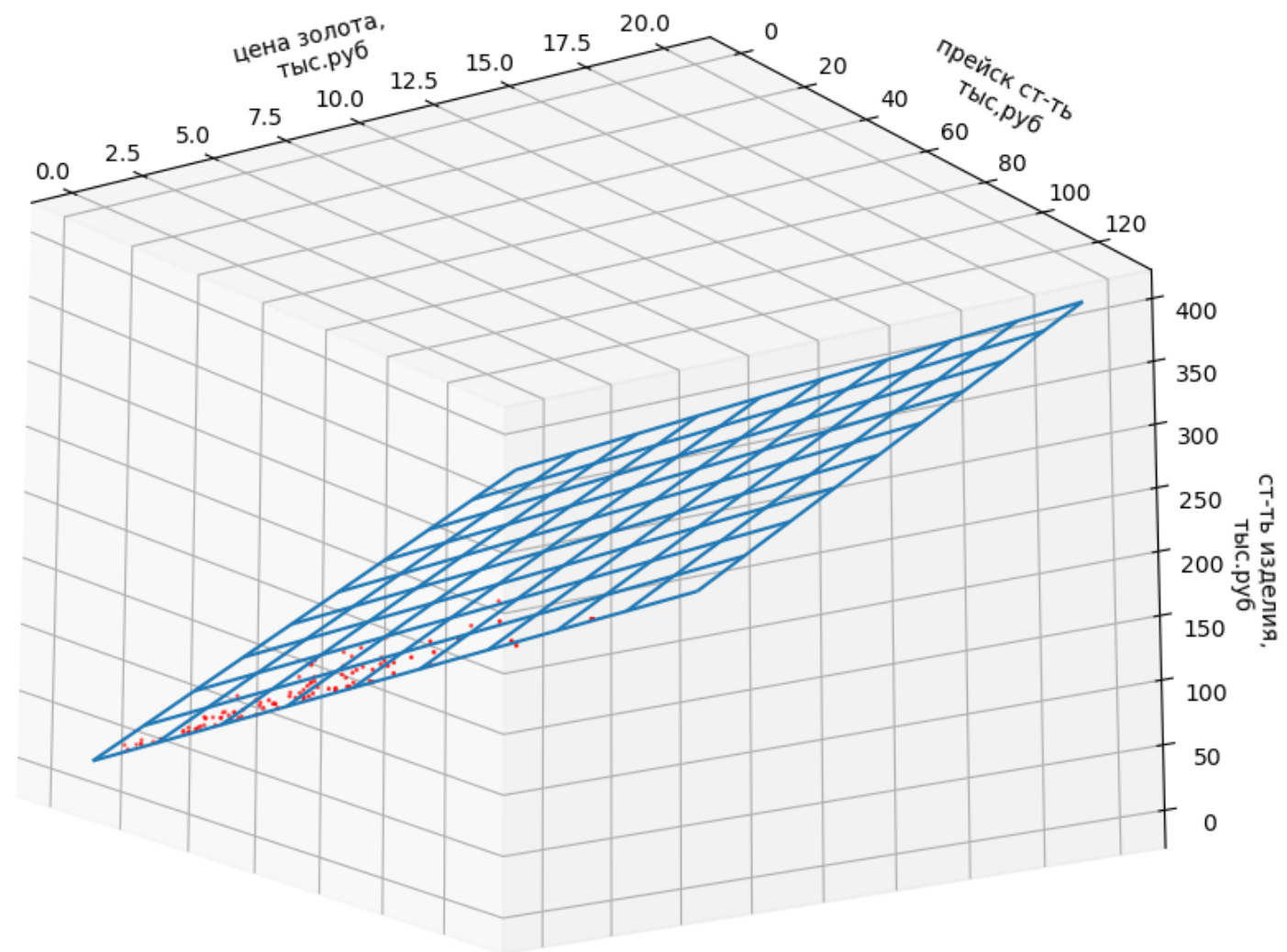
Out[72]:

	price	weight	prcost	prcost_rub	gold_rub	predict	bycoef	diff
0	8594	1.10	10.27	825.30	3227.19	11113.66	11113.66	0.00
1	8594	1.10	10.27	825.30	3227.19	11113.66	11113.66	0.00
2	9829	1.20	7.00	562.52	3520.57	10992.60	10992.60	0.00
3	10258	1.20	6.57	527.97	3520.57	10898.60	10898.60	0.00
4	11251	1.40	14.07	1130.67	4107.33	13725.78	13725.78	0.00

Усредненная по нашим данным плоскость и коэффициенты посчитаны модулем по методу наименьших квадратов, построим на плоскость построенной плоскости для проверки

Ввод [73]:

```
1 x_val = np.linspace(0,20000,10)# np.arange(0, 20000,5000)
2 y_val = np.linspace(0,130000,10)
3 xx,yy = np.meshgrid(x_val,y_val)
4 zz = A*xx + B*yy + C/1000
5 fig = plt.figure(figsize=(10,10))
6 ax = fig.gca(projection='3d')
7 ax.plot_wireframe(xx/1000, yy/1000, zz/1000 )
8 ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.price/1000,s=1,c='red')
9 #ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.predict/1000,s=1,c='red')
10 #ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.bycoef/1000,s=1,c='green')
11 ax.set_xlabel('цена золота, \n тыс.руб')
12 ax.set_ylabel('прейск ст-ть \n тыс,руб')
13 ax.set_zlabel('ст-ть изделия,\n тыс.руб')
14 #ax.plot_surface(xx/1000, yy/1000, zz/1000)
```

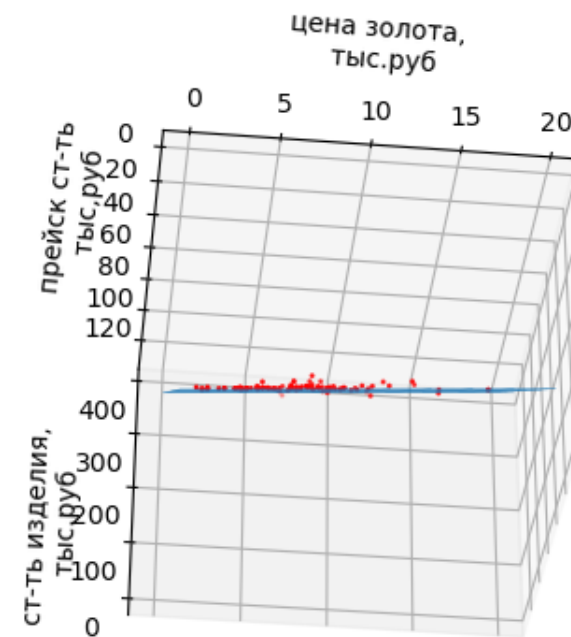
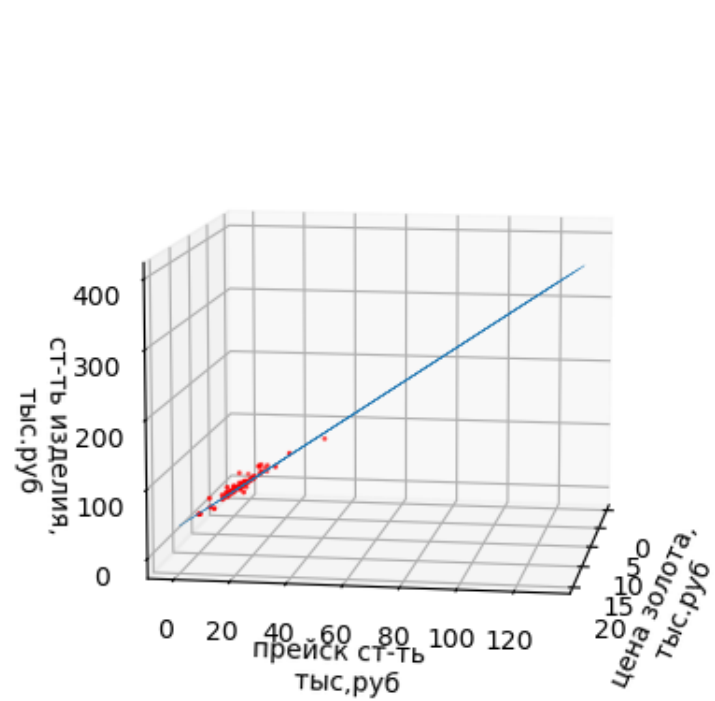




Out[73]: Text(0.5, 0, 'ст-ть изделия,\n тыс.руб')

Ввод [74]:

```
1 fig = plt.figure(figsize=plt.figaspect(0.5))
2
3 ax = fig.add_subplot(1, 2, 1, projection='3d')
4 ax.plot_wireframe(xx/1000, yy/1000, zz/1000, linewidth=0.2 )
5 ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.price/1000,s=1,c='red')
6 ax.set_xlabel('цена золота, \n тыс.руб')
7 ax.set_ylabel('прейск ст-ть \n тыс,руб')
8 ax.set_zlabel('ст-ть изделия,\n тыс.руб')
9 ax.view_init(-7,4)
10
11 ax = fig.add_subplot(1, 2, 2, projection='3d')
12 ax.plot_wireframe(xx/1000, yy/1000, zz/1000, linewidth=0.2 )
13 ax.scatter3D(data.gold_rub/1000,data.prcost_rub/1000,data.price/1000,s=1,c='red')
14 ax.set_xlabel('цена золота, \n тыс.руб')
15 ax.set_ylabel('прейск ст-ть \n тыс,руб')
16 ax.set_zlabel('ст-ть изделия,\n тыс.руб')
17 ax.view_init(-66,-166)
18
```



## 2 Выводы

Реализованная модель анализа цен ювелирных изделий с бриллиантами позволяет получить усредненную информацию по ценообразованию. В анализируемой партии изделий получены следующие коэффициенты цены изделия

Ввод [75]:

```
1 print(f'''
2
3
A = {A:2.2f}
B = {B:2.2f}
C = {C:2.2f}''')
```

A = 2.02  
B = 2.72  
C = 2336.74

где формула -

$$Z = A * X + B * Y + C$$

Z - стоимость изделия в рублях со всеми накрутками с учетом НДС и скидки X - биржевая стоимость золота, данного веса изделия в рублях Y - преysкурантная стоимость бриллиантов в рублях C - наценка в рублях не зависящая от граммов изделия и вставок

Для сравнения в прошлом анализе, подобной партии МЮЗ в (ноябрь 2021) были получены следующие коэффициенты:

A = 3.3  
B = 4.0  
C = 2205

Ввод [ ]:

1