# 1 Отчет анализа стоимости ювелирных изделий у МЮЗ, дата парсинга - 14 апреля 2022

## 1.1 Парсинг

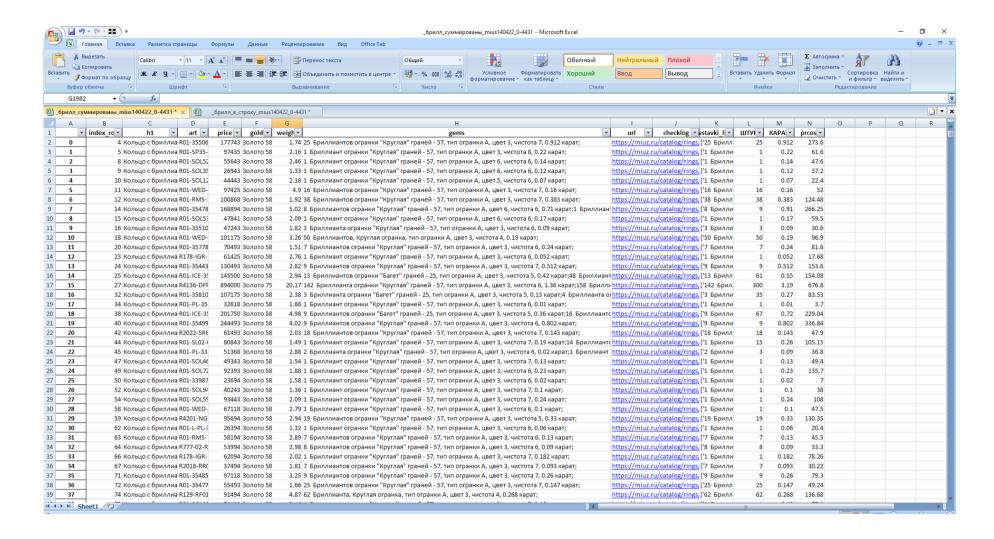
С сайта <a href="https://miuz.ru">https://miuz.ru</a> (<a href="https://miuz.ru">https://miuz.ru</a> (<a href="https://miuz.ru/catalog/rings/filter/stones\_main-is-brilliant/">https://miuz.ru/catalog/rings/filter/stones\_main-is-brilliant/</a> (<a href="https://miuz.ru/catalog/rings/filter/stones\_main-is-brilliant/">https://miuz.ru/catalog/rings/filter/stones\_main-is-brilliant/</a>) Процесс парсинга, реализован в файле ParsingMiuz.ipynb, процесс очистки и расчет прейскуратной стоимости бриллиантов в файле for parsing func.ipynb

Всего колец с фильтром бриллиант на сайте более 4431 изделий, часть данных было отброшено по причинам:

- если изделие содержало другие камни или жемчуг кроме бриллианта
- изделие не могло правильно считаться парсером
- характеристики бриллиантов содержало ошибки
- были использованы фантазийные по цвету бриллианты (возможно синтетика)

В результате получен датасет с ювелирными изделиями со следующими данными (параметрами):

- Наименование(кольца только с бриллиантами в количестве 1980 изделий),
- Артикул,
- Ссылка на карточку с изделием (кликабельная ссылка)
- Стоимость изделия в рублях с учетом скидки (цены от 10 тыс.руб до 981 тыс.руб, на общую сумму 223,3 млн рублей, средняя стоимость 113 тыс.руб),
- масса изделия в граммах (от 0,86 до 20 грамм, общая масса 5703 грамма средний вес 2,88 грамма),
- проба золота (585 проба 1807 шт, 750 140 шт, 375 33 шт),
- штук бриллиантов (всего 52748 штук, в среднем на изделие 26 бриллиантов)
- карат бриллиантов (всего 639 карат средняя масса камня 0,012 карат)
- прейскурантная стоимость бриллиантов в долларах (всего 289 тыс долларов, со средней прейскурантной стоимостью 452 доллара за карат) Датасет приведен в файле 'jewels/muiz04-2022/ брилл суммированы mius140422 0-4431.xlsx'



## 1.2 Анализ датасета МЮЗ, визуализация, определение коэфициентов

## 1.2.1 Считывание данных с файла

Для этого устанавливаем необходимые модули.

- загружаем данные с файла #- откыдываем изделия дороже 500т руб,
- оставляем изделия пробы 585, #- отбрасываем изделия с бриллиантами по прейскуранту свыше 3000 долларов
- оставлены поля price : цена изделя, weight: масса изделия в граммах, prcost: прейскурант в долларах

```
Ввод [11]:
            1 import numpy as np
            2 import pandas as pd
            3 import re
            4 import os
              import matplotlib.pyplot as plt
            6 #%matplotlib inline
            7 %matplotlib notebook
            8 filenamemiuz = 'jewels/muiz04-2022/ брилл суммированы mius140422 0-4431.xlsx'#'jewels/miuz/ брилл суммирова
            1 filenamemiuz
Ввод [12]:
           'jewels/muiz04-2022/ брилл суммированы mius140422 0-4431.xlsx'
            1 raw = pd.read excel(filenamemiuz, sheet name=0)
Ввод [13]:
Ввод [29]:
            data = raw.loc[(raw.gold.str.contains('585') ),['price','weight','prcost']].loc[(raw.price<500000),:].loc[(</pre>
            2 data.head()
 Out[29]:
```

|   | price  | weight | prcost |
|---|--------|--------|--------|
| 0 | 177743 | 1.74   | 273.60 |
| 1 | 97435  | 2.16   | 61.60  |
| 2 | 55643  | 2.46   | 47.60  |
| 3 | 26943  | 1.33   | 37.20  |
| 4 | 44443  | 2.18   | 22.40  |

После урезания получен датасет с 1780 данными общая масса изделий 4.73 кг на сумму 149 млн рублей с прейскурантной стоимостью 182 тыс долл

```
Ввод [31]:
```

- 1 #cymma ∂amacema
  2 data.describe()
- Out[31]:

|       | price     | weight  | prcost  |
|-------|-----------|---------|---------|
| count | 1780.00   | 1780.00 | 1780.00 |
| mean  | 83594.21  | 2.66    | 102.45  |
| std   | 71931.27  | 1.12    | 141.09  |
| min   | 14694.00  | 0.89    | 1.60    |
| 25%   | 42294.00  | 1.91    | 32.19   |
| 50%   | 62094.00  | 2.38    | 59.91   |
| 75%   | 94268.75  | 3.14    | 111.28  |
| max   | 494994.00 | 9.50    | 1484.00 |

#### 1.2.2 Визуализиация данных

Построим облако точек в трехмерной плоскости

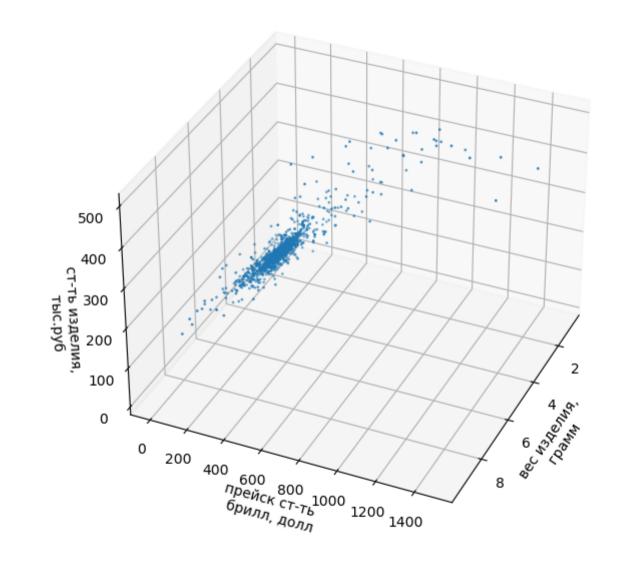
- х вес золота в граммах
- у стоимость прейскурантная бриллиантов
- z стоимость изделия в тысячах рублей

Если покрутить график, то видно что точки стремятся к плоскости

Ввод [33]: 1 %matplotlib notebook

```
BBOД [34]:

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
```



```
Ввод [ ]: 1
```

## 1.2.3 Подготовка данных для определения коэффициентов

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

```
Ввод [35]:

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.head()
```

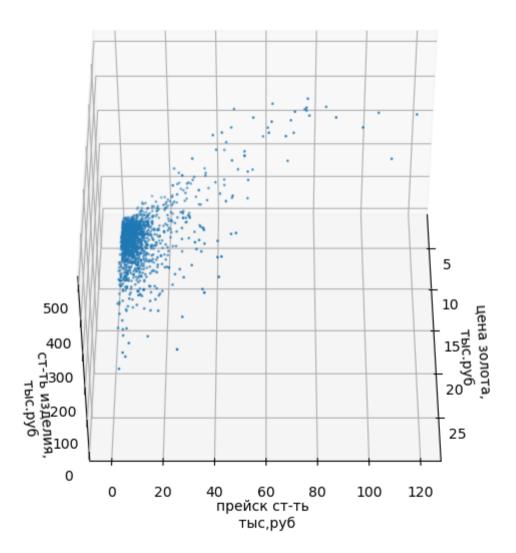
#### Out[35]:

|   | price  | weight | prcost | prcost_rub | gold_rub |
|---|--------|--------|--------|------------|----------|
| 0 | 177743 | 1.74   | 273.60 | 21986.50   | 5104.83  |
| 1 | 97435  | 2.16   | 61.60  | 4950.18    | 6337.03  |
| 2 | 55643  | 2.46   | 47.60  | 3825.14    | 7217.17  |
| 3 | 26943  | 1.33   | 37.20  | 2989.39    | 3901.97  |
| 4 | 44443  | 2.18   | 22.40  | 1800.06    | 6395.71  |

```
Ввод [36]:

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

-



Цена ювелирного изделия в первом приближении формируется линейной зависимостью, если обозначить за Z - цену издели, X - масса золота, Y - прейскурнтная стоимость бриллианов то можно записть закономерностю вида:

$$Z = f(X,Y) = HДС * СКИДКА * НАЦЕНКА *$$
 (ПРОБА \* БИРЖЕВАЯСТОИМОСТ \* НАЦЕНКАЗОЛОТА \* СЛОЖНОСТЬ \*  $X$  + НАЦЕНКАПРЕЙСКУРАНТ \* СЛОЖНОСТЬ \*  $Y$  + 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.

### 1.2.4 Вычисление коэффициентов усредненой по облаку данных плоскости

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

```
BBOД [37]:

1 from sklearn import linear_model
2 #no∂zomo6κα ∂αнных
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[37]: (5.1263482267010145, 5.436369408615513, -1152.3962490867125)

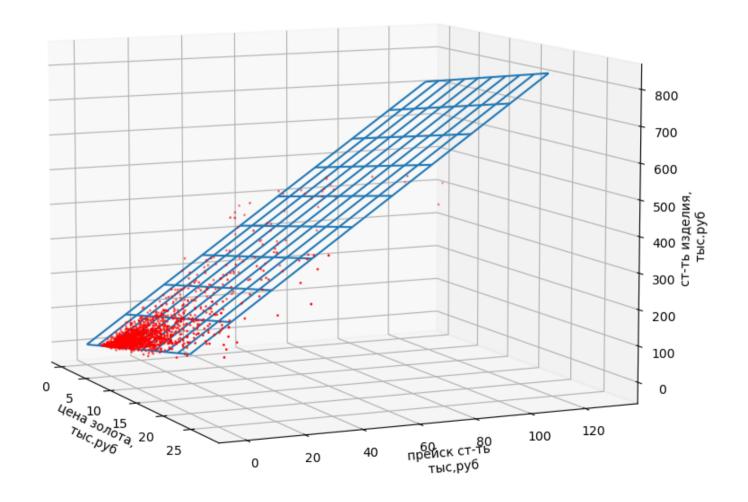
## 1.2.5 Визуальная проверка полученных коэффициентов

```
BBOД [38]: 1 #npoθepκα
data['predict'] = miuz.predict(X_train)
data['bycoef'] = data.gold_rub*miuz.coef_[0] + data.prcost_rub*miuz.coef_[1]+miuz.intercept_
data['diff'] = data.predict-data.bycoef
data.head()
```

#### Out[38]:

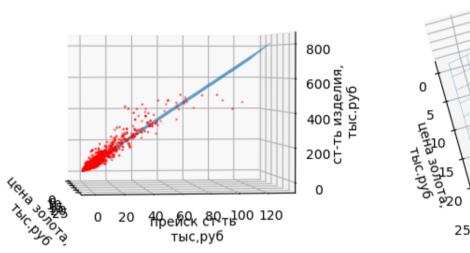
|   | price  | weight | prcost | prcost_rub | gold_rub | predict   | bycoef    | diff |
|---|--------|--------|--------|------------|----------|-----------|-----------|------|
| 0 | 177743 | 1.74   | 273.60 | 21986.50   | 5104.83  | 144543.45 | 144543.45 | 0.00 |
| 1 | 97435  | 2.16   | 61.60  | 4950.18    | 6337.03  | 58244.41  | 58244.41  | 0.00 |
| 2 | 55643  | 2.46   | 47.60  | 3825.14    | 7217.17  | 56640.20  | 56640.20  | 0.00 |
| 3 | 26943  | 1.33   | 37.20  | 2989.39    | 3901.97  | 35101.89  | 35101.89  | 0.00 |
| 4 | 44443  | 2.18   | 22.40  | 1800.06    | 6395.71  | 41420.03  | 41420.03  | 0.00 |

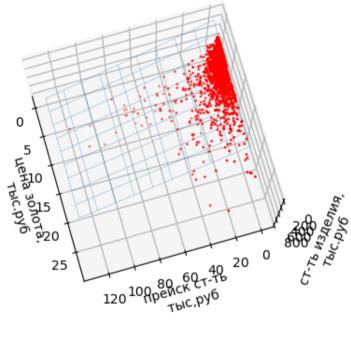
Усредненная по нашим данным плоскость и коэффициенты посчитаны модулем по методу наименьних квадратов, постороим на плоскость построенной плоскости для проверки



Out[39]: Text(0.5, 0, 'ст-ть изделия,\n тыс.руб')

```
Ввод [40]:
            1 fig = plt.figure(figsize=plt.figaspect(0.5))
            3 ax = fig.add subplot(1, 2, 1, projection='3d')
              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 [68]: 1 print(f''' A = {A:2.2f} B = {B:2.2f} C = {C:2.2f}''')

A = 5.13 B = 5.44 C = -1152.40

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 - прейскурантная 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