

Визуализация данных

Гончаренко Д. Набор данных. Набор данных о кофе в магазинах.

Вывод первых 5 строк таблицы:

```
In [ ] : import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import pandas as pd
import seaborn as sns
df = pd.read_csv('coffee-listings-from-all-walmart-stores.csv')
df.drop(['thumbnail'], inplace=True, axis=1)
print(df.head())
sns.set_theme()

0      folgers classic roast ground coffee, 49.3-ounce
1      cafe bustelo, espresso style dark roast ground...
2      folgers classic roast ground coffee, medium ro...
3      Maxwell house original roast ground coffee, 42...
4      great value classic roast medium ground coffee...
```

Длина таблицы:

```
In [ ] : print(len(df))
```

1489

Удалю те, у кого есть неизвестные параметры:

```
In [ ] : df.dropna(inplace=True, how='any')
print(df.head())
print(len(df))
```

```
0      folgers classic roast ground coffee, 49.3-ounce
1      cafe bustelo, espresso style dark roast ground...
2      folgers classic roast ground coffee, medium ro...
4      great value classic roast medium ground coffee...
5      great value classic roast medium ground coffee...

0      coffee_type rating reviews seller_name price \
1      classic roast 3.80 93 walmart.com 13.92
2      espresso, dark roast 4.70 924 walmart.com 3.76
3      medium roast, classic roast 4.40 740 walmart.com 9.97
4      classic roast 4.70 1598 walmart.com 9.98
5      classic roast 4.20 263 walmart.com 7.98

0      weight weight_formatted_to_grams
1      49.3-ounce 1342.50
2      16 oz 283.50
3      25.9 ounce 734.30
4      48 oz 1368.80
5      30.5 oz 864.70
```

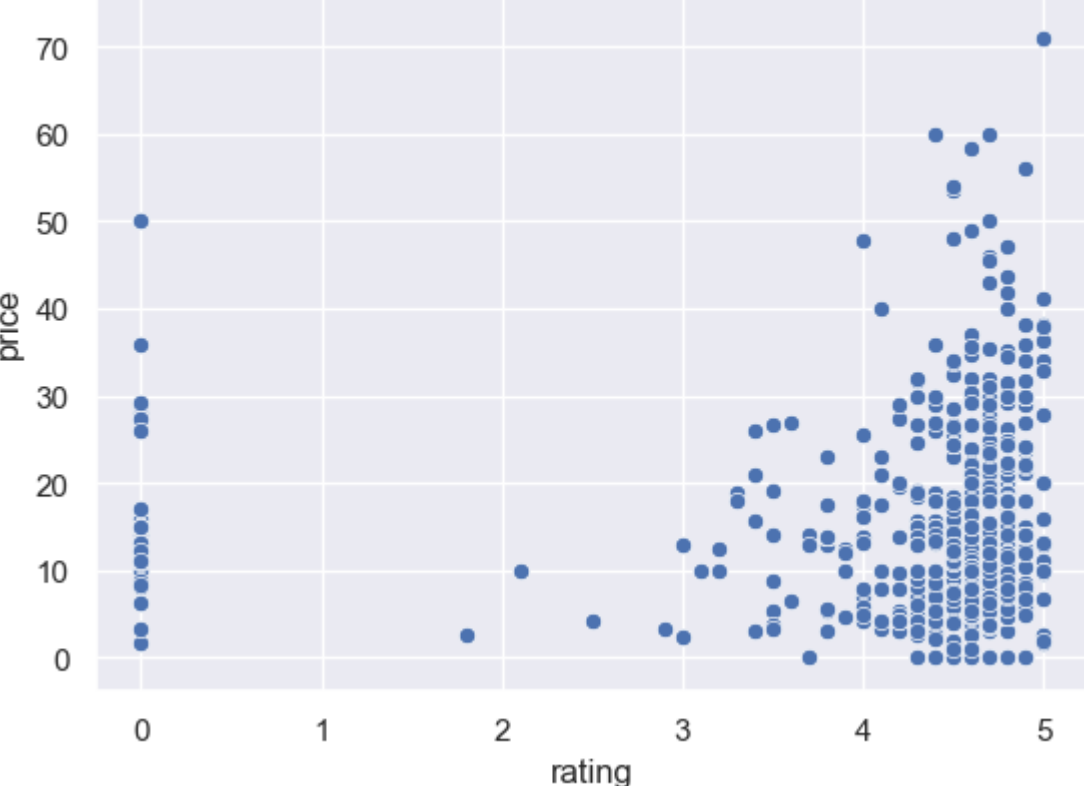
Сортировка по рейтингу:

```
In [ ] : df.sort_values('rating')
```

```
Out[ ] :
0      1080  royal kona coffee vanilla macadamia, light roa...  light roast  0.00  0  hawaii coffee company  15.95  12 coun  340.20
1      616  black rifle coffee beyond black, dark roast, g...  black rifle coffee, dark roast  0.00  0  walmart.com  12.98  12 oz  340.20
2      556  four sigmatic perform high caffeine organic gr...  dark roast  0.00  0  aa wholesale co.  35.90  12 oz  340.20
3      961  starbucks, nitro cold brew coffee drink, pumpk...  cold brew coffee  0.00  0  walmart.com  3.38  not mentioned  807.20
4      962  verena street jufers breakfast blend ground ...  medium roast  0.00  0  walmart.com  9.99  12 ounce  340.20
...
771  771  arabica ground 100% coffee 125 pk  arabica  5.00  1  www.hadotech.com  71.00  not mentioned  510.30
1007  777  cal bustelo ground coffee, dark roast, 6-ounce...  dark roast  5.00  20  sams online trading  11.10  6-ounce  170.10
1075  1005  ruta maya organic coffee dark roast 2.2 lbs.  dark roast  5.00  2  reliable & fast  36.44  not mentioned  340.20
633  582  black rifle coffee spirit of 76 single-serve...  medium roast, black rifle coffee  5.00  1  walmart.com  15.97  22 c  623.70
582  582  kawai coffee na pali coast k-cup coffee pods...  dark roast  5.00  4  walmart.com  12.94  24 c  680.40
```

Связь цены упаковки кофе и её рейтинга (чем дороже кофе тем выше его рейтинг):

```
In [ ] : sns.scatterplot(df, x='rating', y='price')
plt.show()
```



Группировка по типу кофе и магазину:

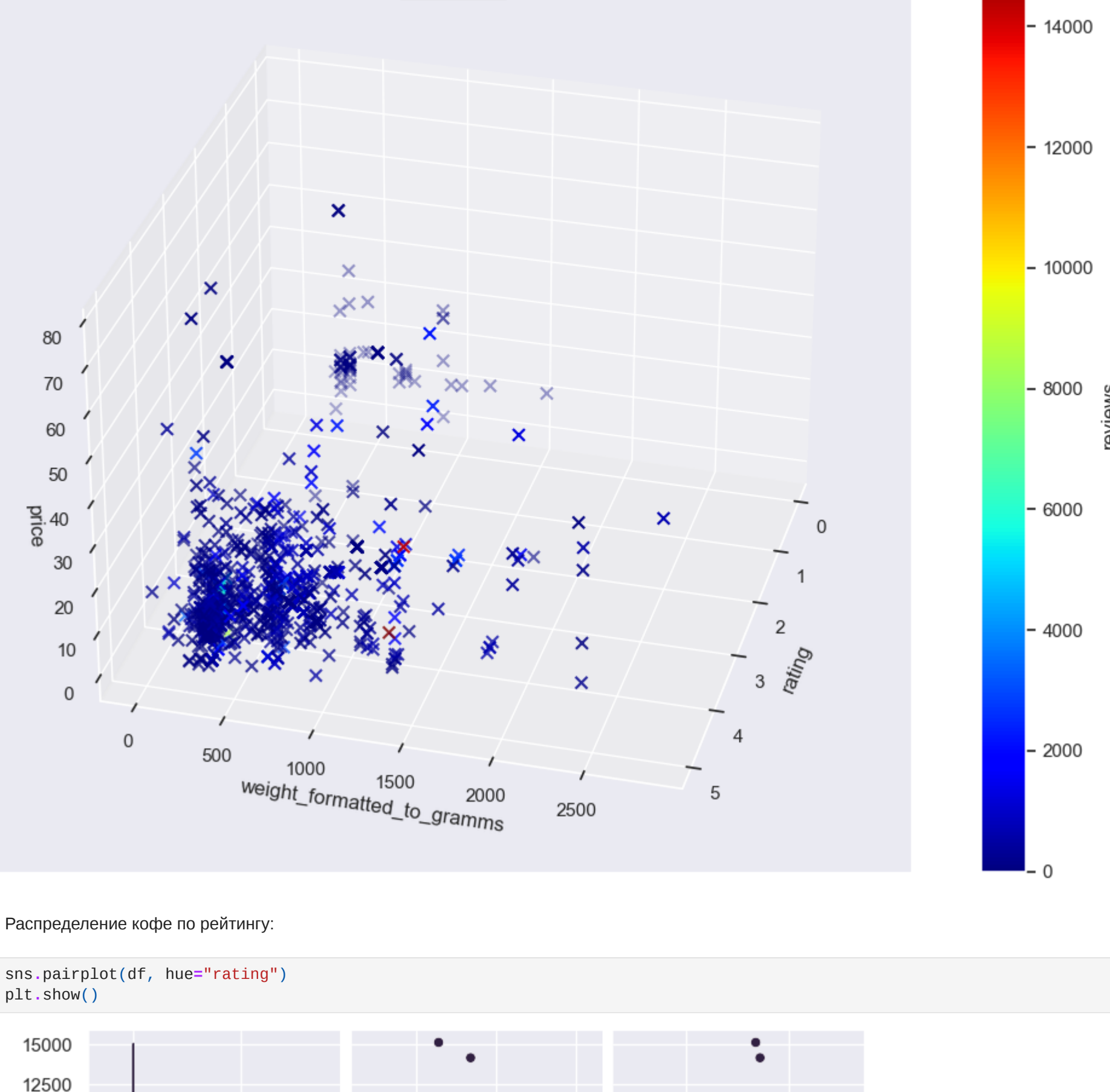
```
In [ ] : df.groupby(['coffee_type', 'seller_name'])['price'].mean()
```

```
Out[ ] :
coffee_type  seller_name  price
americano,cafe americano,medium dark,medium dark roast,dark roast  walmart.com  5.98
americano,cafe americano,medium dark,medium dark roast,dark roast  christopher bean coffee  16.99
arabica  unbeatablesale  12.53
arabica  walmart.com  11.21
arabica  www.hadotech.com  71.00
mocha,light roast  jrc e-commerce  11.86
mocha,medium roast  envision supply source  23.05
vienna  teccino  11.99
vienna  walmart.com  9.48
vienna  walmart.com  5.00
```

График распределения кофе по параметрам: цена, вес упаковки, рейтинг. Доп. параметр: количество отзывов.

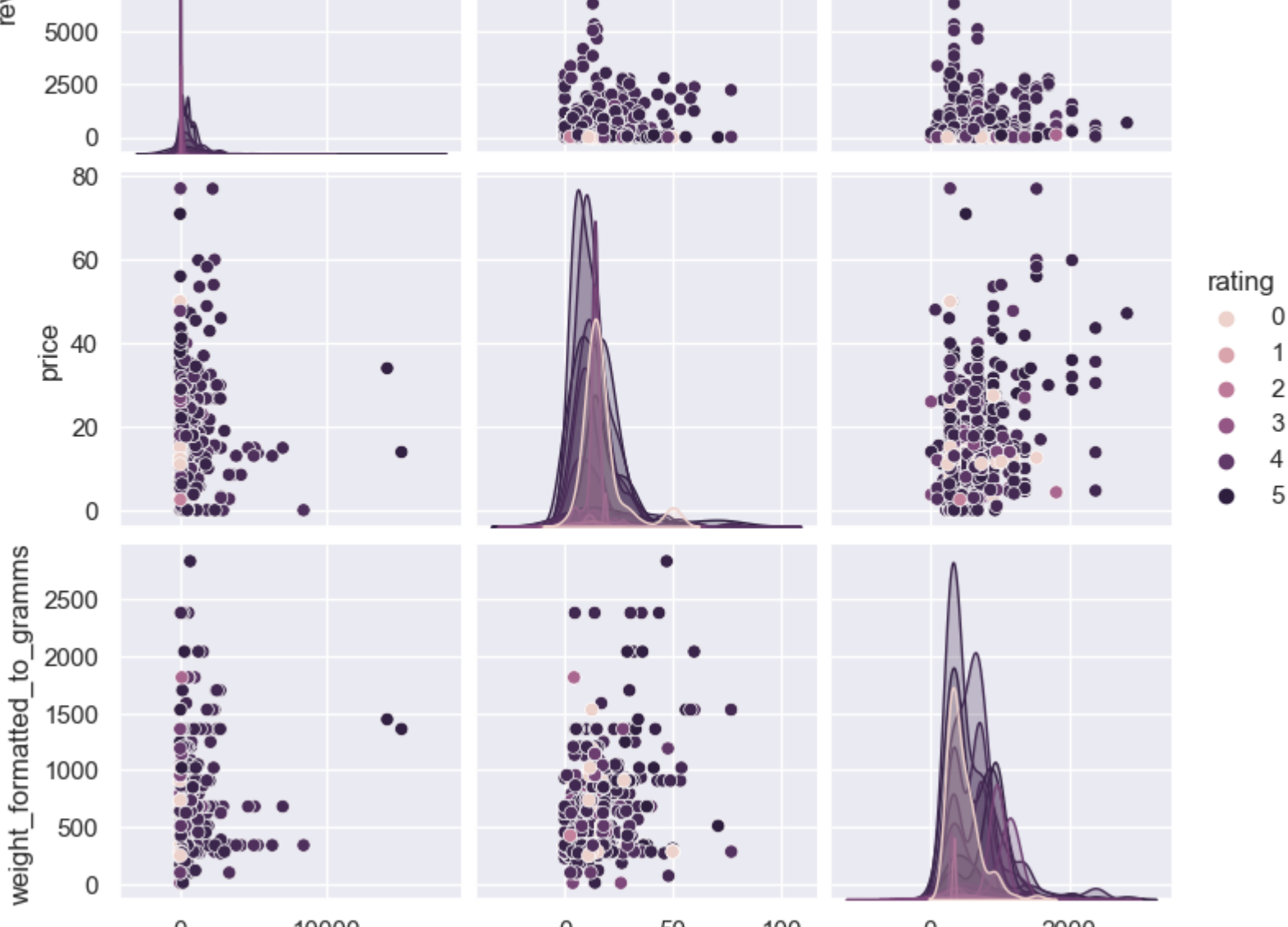
```
In [ ] : data = df
fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')
p, n = ax.scatter(data['rating'], data['weight_formatted_to_grams'], data['price'],
                  marker='x', s=50, c=data['reviews'], label='Coffe', cmap='jet', vmin=df['reviews'].min(), vmax=df['reviews'].max())

ax.set_xlabel('rating')
ax.set_ylabel('weight_formatted_to_grams')
ax.set_zlabel('price')
ax.axis.labelpad = 0
ax.view_init(30, 15)
plt.legend(ncol=3, loc=9)
fig.colorbar(p, n, ax=ax, label='reviews')
plt.tight_layout()
plt.show()
```



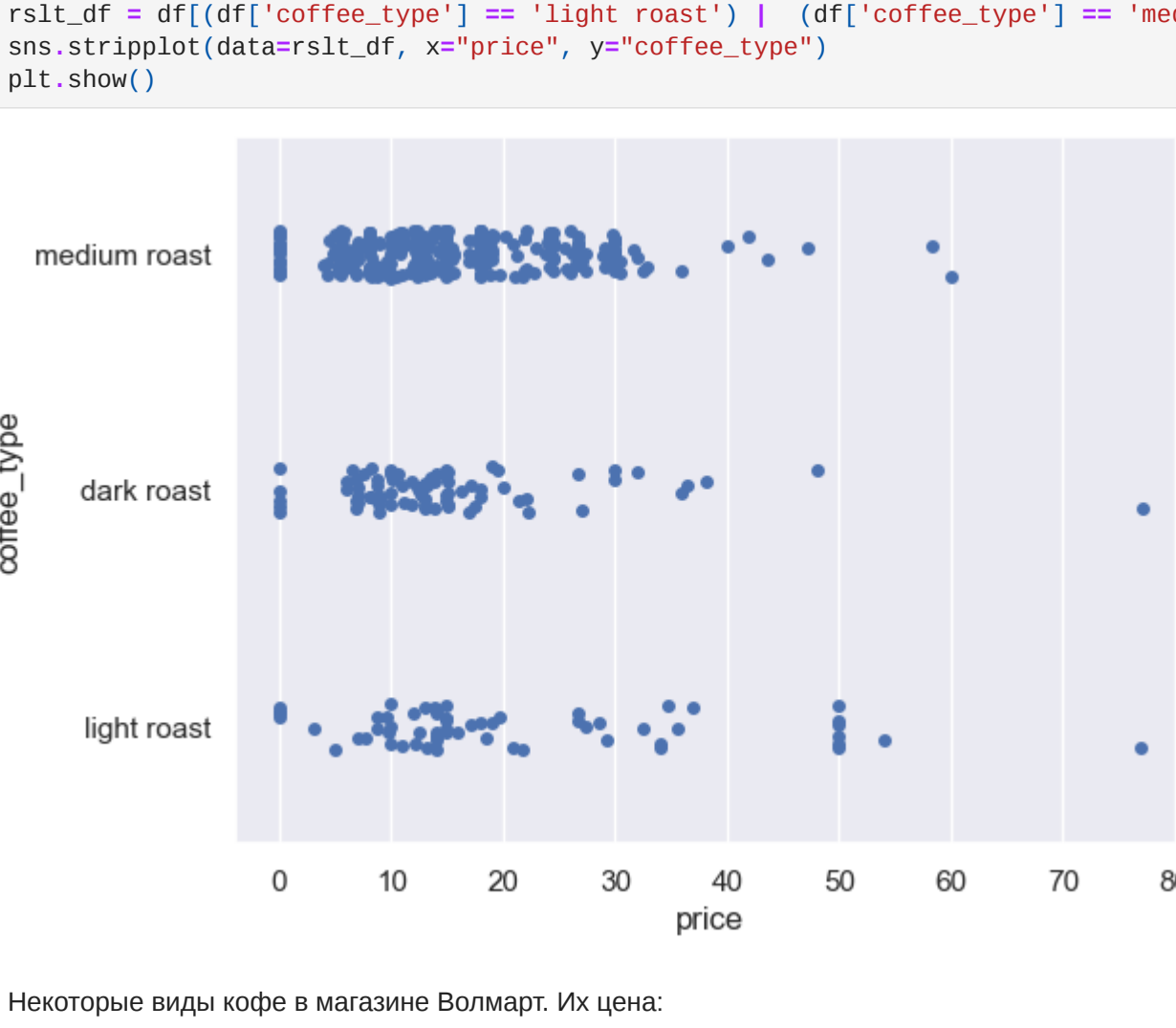
Распределение кофе по рейтингу:

```
In [ ] : sns.pairplot(df, hue='rating')
plt.show()
```



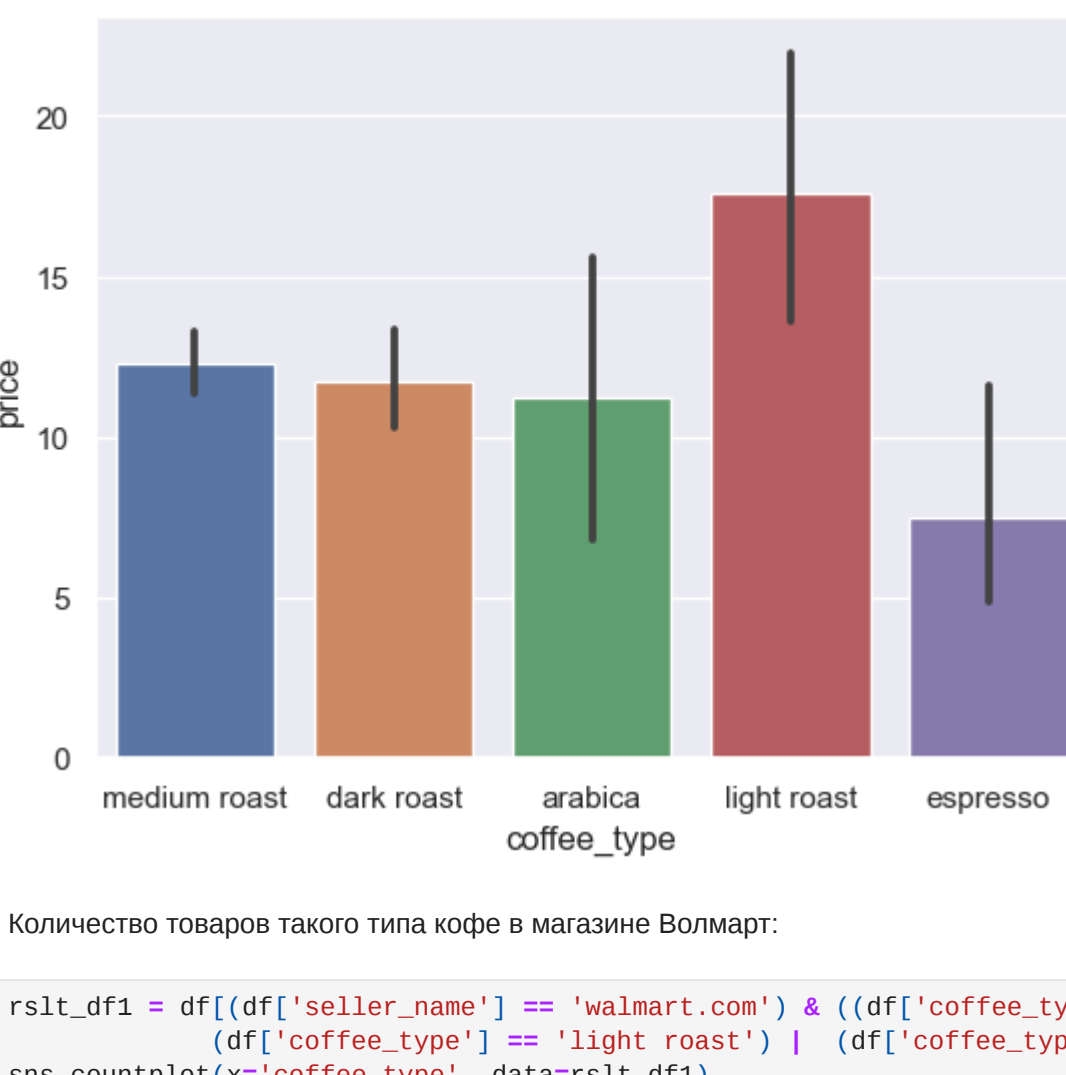
Распределение чистого кофе разного типа обжарки по цене:

```
In [ ] : rs1t_df = df[(df['seller_name'] == 'walmart.com') & ((df['coffee_type'] == 'medium roast')) & ((df['coffee_type'] == 'dark roast'))]
sns.stripplot(data=rs1t_df, x='price', y='coffee_type')
plt.show()
```



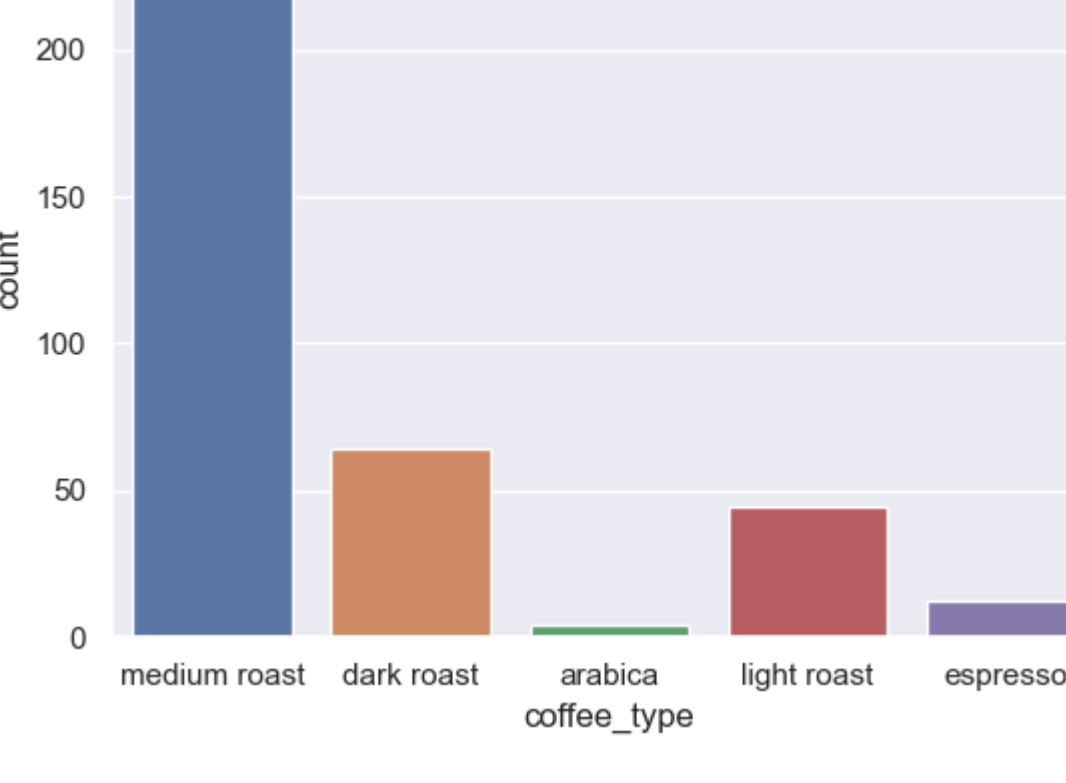
Некоторые виды кофе в магазине Walmart: их цена

```
In [ ] : rs1t_df1 = df[(df['seller_name'] == 'walmart.com') & ((df['coffee_type'] == 'arabica')) & ((df['coffee_type'] == 'espresso')) & ((df['coffee_type'] == 'light roast')) & ((df['coffee_type'] == 'medium roast')) & ((df['coffee_type'] == 'dark roast'))]
sns.barplot(data=rs1t_df1, x='coffee_type', y='price')
plt.show()
```



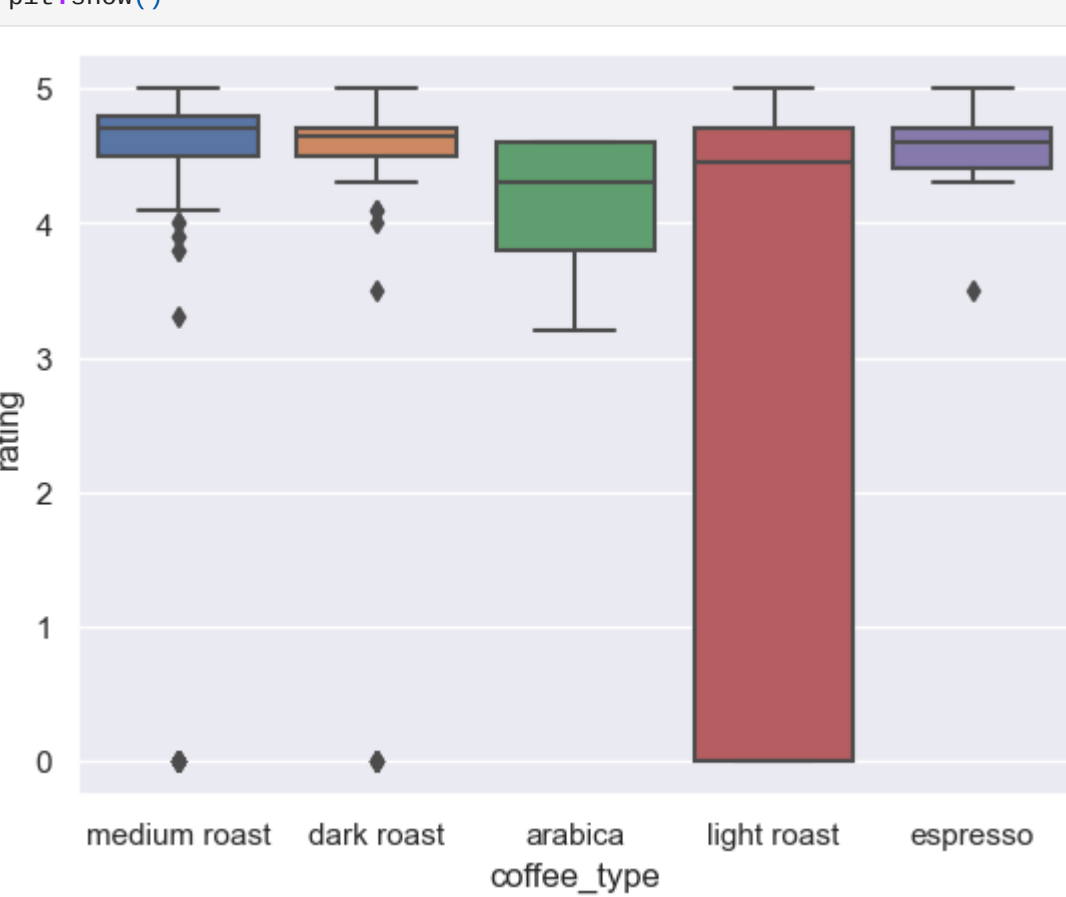
Количество товаров такого типа кофе в магазине Walmart:

```
In [ ] : rs1t_df1 = df[(df['seller_name'] == 'walmart.com') & ((df['coffee_type'] == 'arabica')) & ((df['coffee_type'] == 'espresso')) & ((df['coffee_type'] == 'light roast')) & ((df['coffee_type'] == 'medium roast')) & ((df['coffee_type'] == 'dark roast'))]
sns.countplot(x='coffee_type', data=rs1t_df1)
plt.show()
```



Распределение кофе из Волмарта по рейтингу:

```
In [ ] : sns.boxplot(x='coffee_type', y='rating', data=rs1t_df1)
plt.show()
```



Общая визуализация данных о кофе из Волмарта:

```
In [ ] : sns.pairplot(rs1t_df1, hue='coffee_type')
plt.show()
```

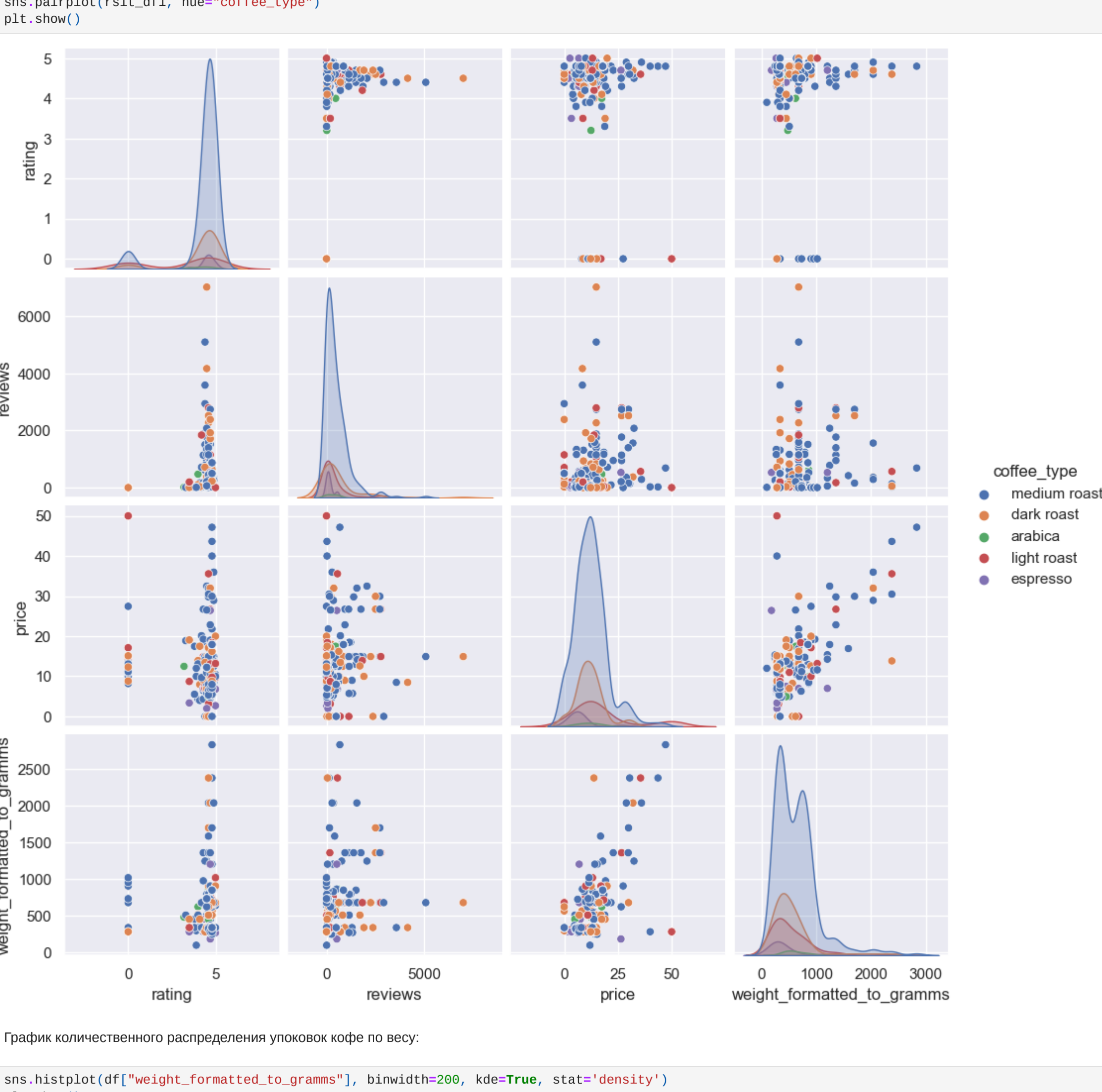
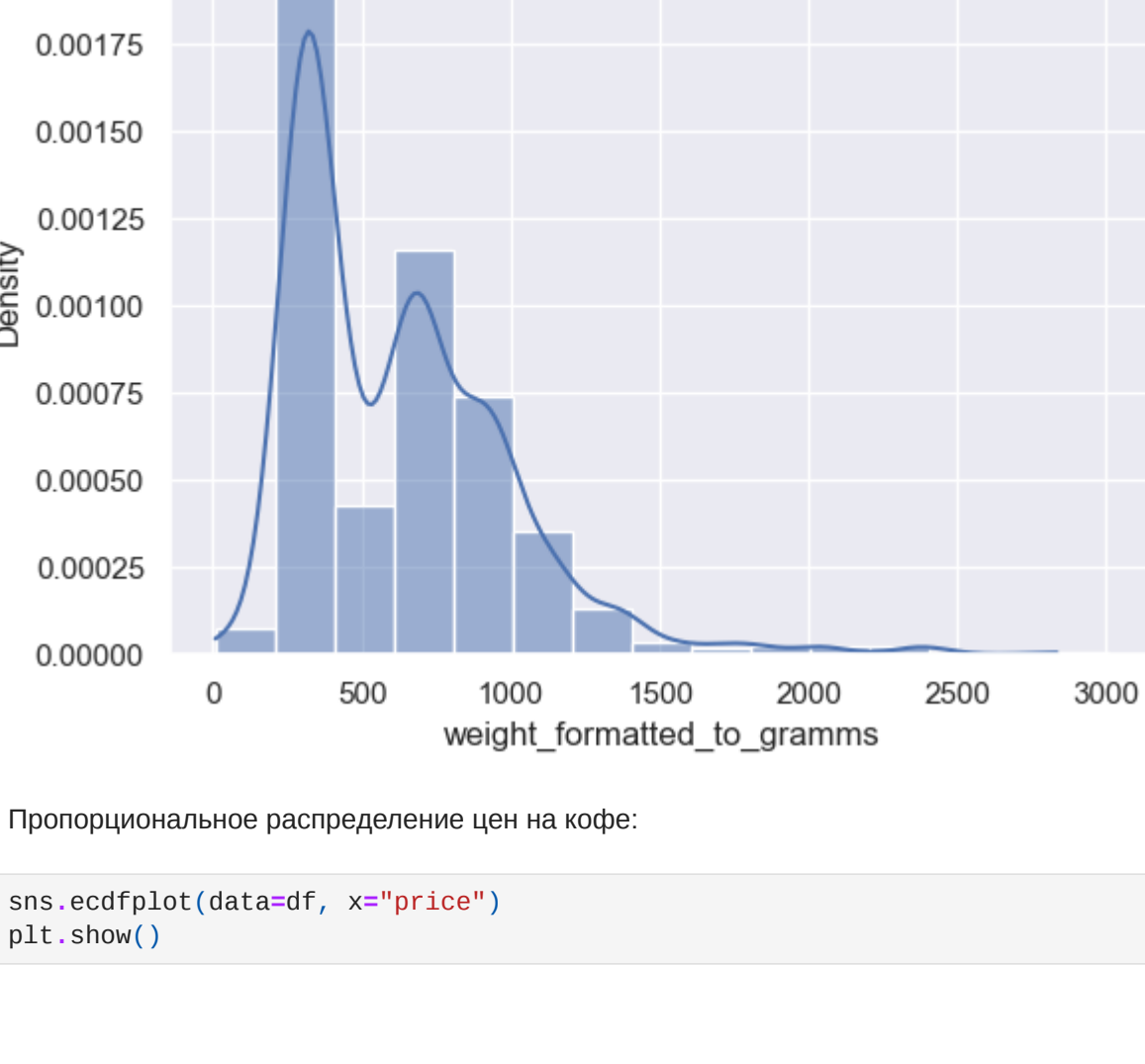


График количественного распределения упаковок кофе по весу:

```
In [ ] : sns.histplot(df['weight_formatted_to_grams'], binwidth=200, kde=True, stat='density')
plt.show()
```



Пропорциональное распределение цен на кофе:

```
In [ ] : sns.ecdfplot(data=df, x='price')
plt.show()
```